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

Regulation of α-ketoglutarate dependent dioxygenase TET3 – novel insights into epigenetic mechanisms in neurons

Franziska Regine Traube

aus

Ulm, Deutschland

2020

Erklärung

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

Eidesstattliche Versicherung

Diese Dissertation wurde selbstständig und ohne unerlaubte Hilfe erarbeitet.

München, den 10.02.2020

……..

Franziska R. Traube

Dissertation eingereicht am: 13.01.2020

1. Gutachter: Prof. Dr. Thomas Carell

2. Gutachter: Prof. Dr. Stylianos Michalakis

Mündliche Prüfung am: 03.02.2020

« La lutte elle-même vers les sommets suffit à remplir un cœur d'homme. Il faut imaginer Sisyphe heureux. «

Albert Camus "Le mythe de Sisyphe"

Published Work

Parts of this thesis are already published in scientific journals or were presented on scientific conferences.

Published in scientific journals:

Schröder AS*, Parsa E*, Iwan K, Traube FR, Wallner M, Serdjukow S, Carell T. "2'-(R)-Fluorinated mC, hmC, fC and caC triphosphates are substrates for DNA polymerases and TET-enzymes" *Chem. Commun.* **52** (2016), 14361 – 14364.

Traube FR, Carell T. "The chemistries and consequences of DNA and RNA methylation and demethylation" *RNA Biol.* **14** (2017), 1099 – 1107.

Traube FR*, Schiffers S*, Iwan K*, Kellner S, Spada F, Müller M, Carell T. "Isotope-dilution mass spectrometry for exact quantification of noncanonical DNA nucleosides" *Nat. Protoc.* **14** (2019), 283 – 312.

Wildenhof TM, Schiffers S, Traube FR, Mayer P, Carell T. "Influencing Epigenetic Information with a Hydrolytically Stable Carbocyclic 5-Aza-2'-deoxycytidine" *Angew. Chem. Int. Ed. Engl.* **58** (2019), 12984 – 12987.

Presented on scientific conferences:

Künzel AF, Traube FR, Carell T "Tet interaction partners – identification, validation and characterization" Poster Presentation at the EmergMed **2015**, Stockholm (Sweden).

Künzel AF*, Traube FR*, Parsa E, Spada F, Carell T "Proteomics-based investigation of the different functional contexts of the Tet family members" Poster Presentation at the Keystone Symposia on Chromatin and Epigenetics **2016**, Whistler (Canada).

Traube FR, Künzel A, Parsa E, Michalakis S, Carell T "Epigenetics, metabolism and neural activity – how a network of metabolic enzymes regulates Tet3 activity in the brain" Poster Presentation at the Cold Spring Harbor Meeting on Mechanisms of Metabolic Signaling **2017**, Cold Spring Harbor (USA).

Traube FR, Özdemir D, Michalakis S, Carell T "Epigenetics, metabolism and neural activity – metabolic regulation of Tet3 activity in the brain" Poster Presentation at the Conference Nucleic Acids, Immunity and Genome Defense **2018**, Brno (Czech Republic).

Traube FR, Özdemir D, Scheel C, Michalakis S, Carell T "Gdh interacts with Tet3 to supply α-ketoglutarate, linking 5mdC oxidation to neuronal function and metabolic state" Poster Presentation at the EMBL Symposium Metabolism meets Epigenetics **2019**, Heidelberg (Germany).

* These authors contributed equally.

Danksagung

Die vorliegende Arbeit wäre niemals ohne die Unterstützung zahlreicher Menschen innerhalb und außerhalb des Labors zustande gekommen. Dafür möchte ich mich bei allen, die am Gelingen dieser Arbeit direkt oder indirekt beteiligt waren, bedanken!

Meinem Doktorvater *Prof. Dr. Thomas Carell* danke ich herzlich, dass er mich in seine Arbeitsgruppe aufgenommen und vom ersten Tag an gefördert und gefordert hat. Das spannende Thema Epigenetik und Metabolismus im Gehirn hat mich manchmal fast in den Wahnsinn getrieben, aber ich habe mir dabei stets dein Motto „Es gibt keine Probleme, nur Herausforderungen" zu Herzen genommen. Vielen Dank für das entgegengebrachte Vertrauen, die guten Ideen und die kontroversen Diskussionen. Auch wenn du am Ende meistens Recht behalten hast, habe ich unsere Auseinandersetzungen immer genossen. Besonders bedanken möchte ich mich auch für deine Unterstützung während und nach der Schwangerschaft, sodass ich fast normal während dieser Zeit im Labor arbeiten konnte.

Prof. Dr. Stylianos Michalakis möchte ich nicht nur für die Übernahme des Zweitgutachtens, sondern auch für die fantastische Zusammenarbeit während meiner gesamten Promotion bedanken. Deine Gruppe in der Pharmazie war nach dem Carell-Labor mein zweites Zuhause und dein Input zu jeder Zeit wertvoll.

Den weiteren Mitgliedern meiner Prüfungskommission, *Prof. Dr. Philip Tinnefeld*, *Dr. Stefanie Kellner*, *Dr. Sabine Schneider* und *Prof. Dr. Klaus Förstemann* danke ich für die Mitwirkung an der mündlichen Prüfung.

Dem Boehringer Ingelheim Fonds und den (ehemaligen) Stipendiat*innen möchte ich für die finanzielle und ideelle Unterstützung danken! Jedes Zusammentreffen mit euch ist eine bereichernde Erfahrung und ich konnte so viele inspirierende Menschen kennen lernen. Dem ganzen BIF Team in Mainz danke ich für das große Engagement. Ihr habt immer schnell und unkompliziert geholfen und habt dieses Stipendium zu so viel mehr als nur einer Finanzierungshilfe gemacht!

Dr. Markus Müller, *Angie Kirchner*, *Dilara Özdemir*, *Dr. Victoria Splith* und *Constanze Scheel* danke ich für das Korrekturlesen dieser Arbeit.

Vielen Dank an das gesamte Carell-Labor in allen Zusammensetzungen über die Jahre – ihr habt es mir leicht gemacht, dass ich wirklich sehr viel Spaß an dieser Doktorarbeit hatte.

Besonders bedanken möchte ich mich bei *Dr. Markus Müller* für viele gute Ideen, wertvolle Ratschläge und all die Dinge, die du für mich und die gesamte Gruppe machst, ohne dass wir wirklich davon mitbekommen! Ich habe sehr viel von dir gelernt – wie man Experimente angeht und hinterfragt, welche Kontrollen man gleich mitdenken muss, usw. Vielen Dank dafür und auch dafür, dass ich so jemanden im Arbeitskreis hatte, der meine Begeisterung für Fahrräder teilt.

Frau Gärtner danke ich herzlich für all die organisatorischen Dinge, ohne die ich nicht hätte arbeiten können!

Großer Dank geht an mein Lab-Girls Team (das Trio Infernal) *Dr. Katharina Iwan* und *Angie Kirchner*. Kathi, danke für all die Stunden, die du wegen mir am Triple-Quad zubringen musstest, damit ich meine vielen Proben doch noch über das Wochenende messen konnte. Angie, vielen Dank für alles, was du mir in der Zellkultur beigebracht hast. Und vielen Dank euch beiden für all das, was man nicht richtig in Worte fassen kann! Egal, ob es um das Labor oder um Privates geht, ihr wart und seid immer für mich da und ich schätze mich glücklich, mich euch befreundet zu sein! Das Ziel Weltherrschaft steht noch!

Dr. Andrea Glück danke ich für die Einführung in das Thema und unseren gemeinsamen Kampf gegen die Windmühlen!

Dem Minion Team, *Dilara Özdemir*, *Sabine Oganesian* und *Hanife Sahin* danke ich für die tolle Zusammenarbeit! Wir sind alle sehr unterschiedliche Charaktere, doch gerade darin liegt die Stärke unseres Teams. Herzlichen Dank für eure Unterstützung, Motivation und Arbeit! Ohne euch würde das Projekt nicht vorangehen und ich bin euch sehr dankbar, dass ihr bei allen Schwierigkeiten, mit denen wir zu kämpfen haben, trotzdem nicht aufgebt!

Kristof Hufnagel und *Luis de la Osa de la Rosa* danke ich herzlich für die Unterstützung bei allen möglichen und unmöglichen Biochemie- und Zellkulturexperimenten. Ohne euch und eure Hilfe hätte ich diese Doktorarbeit vermutlich nicht überlebt. Immer, wenn ich kurz vor der Schnappatmung war, weil ich wieder zu viel auf einmal machen wollte, oder wenn mich einfache Alltagsaufgaben überfordert haben, wart und seid ihr für mich da. Diese Motivation, dieser Elan und diese Zuverlässigkeit sind nicht selbstverständlich und ich kann euch nicht genug dafür danken!

Kerstin Kurz danke ich herzlich für die Unterstützung am Triple-Quad. Danke, dass du die Maschine pflegst, unzählige Yolos misst und an der Maschine alles versuchst, bis es endlich passt! Vielen Dank auch, dass du einem immer helfend zur Seite stehst, wenn die Organisation des Labors einen mal wieder überfordert.

Dr. Michael Stadlmeier und *Leander Runtsch* danke ich für zahlreiche Proteomics Experimente, die Diskussionen, die Hilfe und den Spaß, den ich mit euch hatte und habe.

Dr. Fabio Spada danke ich für sein unerschöpfliches Wissen rund um die Zellkultur und Stammzellen. Dein Erfahrungsschatz lässt jedes Lehrbuch dazu schlecht aussehen! Ich habe viel von dir gelernt – allen voran, wie man iNGNs kultiviert.

Dem ganzen weiteren Biochemie-Team, *Johanna Bretzler*, *Dr. Nina Simon*, *Olga Kamerer*, *Timm Ensfelder* und *Ewelina Kamińska* möchte ich für die Unterstützung, aber vor allem auch den Spaß danken, den wir zusammen haben und hatten. Danke, dass ihr es mir oft nachgesehen habt, wenn ich mal wieder in der BC gewütet habe.

Dr. René Rahimoff, *Dr. Christina Schneider*, *Dr. Sarah Schiffers*, *Dr. Mirko Wagner*, *Dr. Edris Parsa* und *Clemens Dialer* danke ich für Spaß im Labor, beim Sport oder bei politischen Diskussionen.

Eva Korytiaková danke ich für ihre Seelenverwandtschaft in so vielen Dingen. Niemand kann besser verstehen, dass man in allen Bereichen entweder 200% gibt oder es gleich lässt. Vielen Dank für die Vier-Augen-Gespräche im Äkta-Raum, der Zellkultur im dritten Stock oder an sonstigen Orten, an die sich garantiert niemand zufällig verirrt.

Besonderen Dank auch an den gesamten 4. Stock des C Gebäudes (aka meine Pharmazie). Immer, wenn ich Abwechslung brauchte, konnte ich zu euch kommen und wurde herzlich aufgenommen. Ganz besonderer Dank gebührt hier *Dr. Anna Geserich*, *Constanze Scheel* und *Dr. Victoria Splith*! Ihr seid die besten Kooperationspartnerinnen aller Zeiten. Mit euch habe ich nicht nur alle Themen bearbeitet, unzählige Stunden, in denen man über alles reden konnte, vor dem Covaris oder am Mikroskop verbracht, sondern immer sehr viel Spaß gehabt. Ohne euch wäre diese Doktorarbeit niemals gelungen!

Dem Arbeitskreis von *Dr. Sabine Schneider* danke ich herzlich für das Mitbenutzen verschiedener Geräte. *Dr. Raffaella Villa* aus dem Arbeitskreis von *Prof. Becker* danke ich für die tolle Einführung in ChIPseq und *Dr. Mario Raab* aus dem Arbeitskreis von *Prof. Tinnefeld* danke ich für die Durchführung

von Super-Resolution Microscopy Versuchen. Beides war am Ende leider erst einmal nicht erfolgreich, aber was noch nicht ist, kann ja noch werden.

Ina Kugler und *Lisa Engelbrecht* danke ich für inzwischen über zehn Jahre BC-Freundschaft! Ohne euch hätte das Studium nur halb so viel Spaß gemacht und auch wenn wir jetzt nicht mehr an der gleichen Laborbank stehen, tut es gut, dass ich bis heute Probleme, Erfolge und alles andere mit euch teilen und diskutieren kann.

Meinen Eltern, *Michael* und *Christa Traube*, danke ich für ihre Unterstützung in allen Lebenslagen. Danke, dass ihr mir beigebracht habt, Dinge zu hinterfragen, neugierig und ausdauernd zu sein!

Den größten Dank möchte ich am Schluss meiner Münchner Homebase aussprechen. *Rafael Hefele*, *Malin Hefele*, *Matthias Traube*, *Ria Spallek*, *Vincent Hefele* und *Vera Wanka* – ein Leben ohne euch möchte ich mir nicht vorstellen! *Matthias*, wir konnten uns immer gegenseitig unterstützen, verstehen und motivieren und du bist Ansporn für mich, weiter an mir zu arbeiten. *Ria*, du warst die beste Mitbewohnerin, die mir passieren konnte, und bist die beste Freundin, die ich mir vorstellen kann. Gemeinsam bringen wir die Dodos an die Macht! (Und ich verstehe bis heute nicht, wie man das nicht verstehen und nicht witzig finden kann!) *Malin*, du bist der Sonnenschein und Frechdachs, für den allein es sich schon lohnen würde, jeden Morgen aufzustehen. Ich hoffe, ich kann dir in vielem ein Vorbild sein und dir helfen, nach und nach deinen eigenen Weg in dieser Welt zu finden. *Rafael*, ohne dich wäre ich nicht ich. Du hast mich immer unterstützt, nie enttäuscht. Egal, ob es um die Arbeit geht oder die gemeinsame Bezwingung des Timmelsjochs – dank dir habe ich nicht aufgegeben, wenn es weh tat und auch bei starkem Gegenwind den Spaß nicht verloren. Danke für deinen grenzenlosen Rückhalt, ohne den diese Doktorarbeit, unsere Familie und alles, was die Zukunft bringt, nicht möglich (gewesen) wäre!

Table of Content

Summary

The mammalian brain is a highly complex organ that controls and regulates the organism's vital functions, such as breathing or body temperature, and non-vital functions, such as movement, learning and instinct. Millions of neurons and glia, the two major cell types of the brain, are linked to each other to receive, process and eventually store information at any time of the organism's lifespan from the moment on they develop during embryogenesis.[1, 2] To fulfill its functions, the brain needs to adapt constantly on a molecular level and dynamic changes of the gene expression pattern of individual brain cells are one essential part of this adaption.[3-6] Changes of gene function without changing the DNA sequence itself are epigenetic changes. The DNA sequence is the universal blueprint of an organism that does not change except for random mutations in individual cells. In contrast, epigenetic changes, which include modifications of DNA nucleobases, histone modifications and changes of the chromatin status by non-coding RNAs and chromatin remodelers, are highly orchestrated and dynamic.[7] This epigenetic plasticity allows short-term or permanent changes of gene expression. In the brain, 5-methyl-2'-deoxycytidine (5mdC) and 5-hydroxymethyl-2'-deoxycytidine (5hmdC) are important epigenetic marks.[8] Genes that are important for neural function and whose expression levels depend on and are regulated by neural activity show dynamic changes of the 5mdC and 5hmdC pattern in their promoters and intragenic sites.[9-12] 5mdC is oxidized to 5hmdC by ten-eleven translocation (TET, Tet1-3) enzymes. TET enzymes belong to the family of α-ketoglutarate (αKG)-dependent dioxygenases and in neurons, Tet3 has the highest expression levels among the three paralogues.[13, 14] In mice, Tet3 has been shown to be important for synaptic transmission and homeostatic plasticity and in humans, TET3 mutations within its catalytic domain could be recently linked to developmental delay, severe intellectual disability and growth retardation.[15, 16] This TET3-linked disorder shows similarities with a well-studied developmental disorder known as Rett syndrome (RTT) that is caused by mutations within the methyl-CpG-binding protein 2 (MeCP2). MeCP2 is also highly expressed in neurons and glia and binds 5mdC and 5hmdC thereby attracting additional proteins and influencing gene expression.[17-19]

This PhD thesis focuses on the metabolic control of Tet3 activity in the brain, compares the phenotypes of TET3-deficient and MeCP2-deficient neurons and studies the interplay of TET3 and MeCP2 in neurons.

We discovered that many metabolic enzymes are enriched by Tet3 co-immunoprecipitation (coIP) in nuclear brain lysate of adult mice. We could identify glutamate dehydrogenase (*Glud1*, Gdh), which converts abundant glutamate to αKG, among the interactors. I confirmed and characterized this interaction between Tet3 and Gdh in hippocampal neurons. *Glud1* is nuclear encoded but possesses a mitochondria-targeting sequence (MTS) and the expressed protein (Gdh) is therefore normally transported to the mitochondria, where the MTS is eventually cleaved off. I found out that in hippocampal neurons, Gdh is predominantly nuclear located to supply Tet3 with αKG, which is rate limiting for the Tet3-catalyzed 5mdC oxidation reaction. Nuclear Gdh still possesses the MTS, which seems to be important for the interaction with Tet3 (Figure I). In contrast, in hepatocytes of the liver, where TET expression and 5hmdC levels are very low in a non-fasted state[20, 21], Gdh is in the mitochondria.

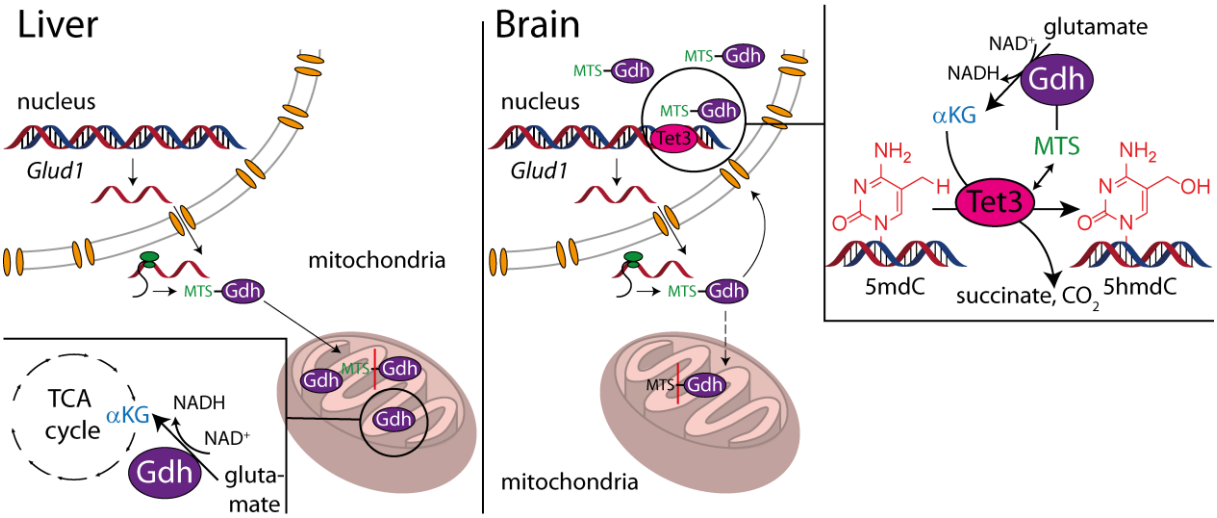


Figure I: **Interaction of Gdh with Tet3.** In liver, Gdh is transported to the mitochondria, where its MTS is cleaved off and Gdh supplies the TCA cycle with αKG. In hippocampal neurons, Gdh is mainly transported to the nucleus, where it interacts via its MTS with Tet3 to supply αKG for 5mdC to 5hmdC oxidation. 5hmdC = 5-hydroxymethyl-2'-deoxycytidine, 5mdC = 5-methyl-2'-deoxycytidine, αKG = α-ketoglutarate, Gdh = glutamate dehydrogenase (protein), *Glud1* = glutamate dehydrogenase (gene), MTS = mitochondria-targeting sequence, Tet3 = ten eleven translocation enzyme 3.

During neuronal activation, Gdh activity regulates TET activity. When hippocampal neurons are depolarized, the 5hmdC levels are increased compared to the unstimulated control. However, when Gdh is inhibited, this increase in global 5hmdC cannot be observed anymore. The increase in 5hmdC goes along with increased gene expression levels of neuronal PAS domain-containing protein 4 (*Npas4*) and brain-derived neurotrophic factor (*Bdnf*). Npas4 and Bdnf are two major players in neurons for synaptic plasticity and memory formation[22, 23] and when Gdh is inhibited, both genes show lower expression levels compared to stimulated neurons without Gdh-inhibition. Furthermore, I observed that neuronal TET activity also depends on the general metabolic state of the cell and when glycolysis is impaired during stimulation, TET activity is lower. Taken together, my results show that Gdh activity controls Tet3-dependent 5mdC oxidation, and thereby gene expression, in hippocampal neurons. As αKG is important for energy metabolism and a co-substrate for more than 60 αKG-dependent oxygenases[24, 25], the increase of the effective αKG molarity only in the environment of Tet3 is an elegant way to avoid globally changing αKG levels that would otherwise go along with various side effects.

Not only Gdh was enriched in the Tet3-coIP in nuclear brain lysate, but also the malate dehydrogenases (MDH, Mdh1/2) and the aspartate aminotransferases (GOT, Got1/2), which are part of the malate aspartate shuttle (MAS), and lactate dehydrogenase (LDH, Ldhb). The MAS and LDH are essential to regulate the nicotinamide adenine dinucleotide (oxidized NAD^+/reduced NADH) levels between the cytosol and the mitochondria.[26] MDH and LDH use NAD^+/NADH as redox equivalents and GOT performs a pyridoxal-dependent transaminase reaction to convert glutamate to αKG, thereby converting oxaloacetate, which can be subsequently reduced to malate, to aspartate or vice versa. Using *in vitro* assays, I could show that oxaloacetate is a Tet3-inhibitor and when I fed Tet3/Got2-transfected HEK293T cells with aspartate, I observed drastically reduced 5hmdC levels compared to the unfed control. These preliminary results indicate that Tet3-activity is controlled by a whole metabolic network, which includes the MAS, but further studies are needed to decipher the individual contribution of these metabolic enzymes to the network.

In my second main project, I focused on the interplay between TET3 and MeCP2 in iNGN-derived neurons. iNGNs are a small molecule-inducible Neurogenin induced pluripotent stem cells (iPSCs)

line.[27] Neurogenin 1 and 2 overexpression in iPSCs leads to rapid differentiation of the cells to neurons with immature neurons four days and mature neurons two weeks after induction. I found out that TET3 and MeCP2 interact in the iNGN-derived neurons in a phase of ongoing synaptogenesis (eight days post-induction) and that MeCP2, but not two common RTT-associated mutants thereof, reduce TET3 activity. Our study revealed that *MeCP2*$^{-/Y}$ and *TET3*$^{-/-}$ iNGN-derived neurons are very similar on the DNA methylation, transcriptome and proteome level and different from the wildtype (wt). Neither of the knockout cell lines shows obvious differences during the differentiation and early maturation process. However, eight days post-induction, genes that belong to the category of neuronal signaling are massively upregulated, whereas genes that belong to the category cell migration are downregulated. The main difference between *MeCP2*$^{-/Y}$ and *TET3*$^{-/-}$ iNGN-derived neurons are the global 5hmdC levels, which are higher in the *MeCP2*$^{-/Y}$, but lower in the *TET3*$^{-/-}$ compared to the wt. These results indicate that in the resting neuron, global 5hmdC levels are not essential for gene expression regulation. Rather, the interplay between TET3 and MeCP2 during ongoing synaptogenesis seems to be important for transcription regulation. In the developed brain, the situation is probably different since I failed to confirm the interaction between both proteins in adult murine brain and fully matured neuronal networks in culture. These results fit to the clinical progression of RTT in humans, which starts six to 18 months after birth when synaptogenesis is high but enters afterwards a stabilization period. Taken together, our results suggest that TET3 might play an important role in the clinical progression of RTT by influencing MeCP2 and help to explain why TET3 mutations cause a neurodevelopmental disorder reminiscent of RTT.

Abbreviations

2DG	2-deoxy-D-glucose
2HG	2-hydroxyglutarate
2HGDH	2-hydroxyglutarate dehydrogenase
2i	medium supplemented with PD0325901 and CHIR99021
5cadC	5-carboxy-2'-deoxycytidine
5fdC	5-formyl-2'-deoxycytidine
5hmdC	5-hydroxymethyl-2'deoxycytidine
5mdC	5-methyl-2'deoxycytidine
6mdA	N6-methyl-2'-deoxyadenosine
acetyl-CoA	acetyl-coenzyme A
Aco2	aconitase 2
AcOH	acetic acid
ADP	adenosine diphosphate
αKG	alpha-ketoglutarate
AMPAR	α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic receptors
ANLS	astrocyte-neuron lactate shuttle
APS	ammonium persulfate
ATP	adenosine triphosphate
BAT	brown adipose tissue
Bdnf	brain-derived neurotrophic factor
BER	base-excision repair
bp	base pairs
C	carbon
Ca^{2+}	calcium ions
cd	catalytic domain
cDNA	complementary DNA
ChIP	chromatin immunoprecipitation
cKO	conditional knockout
CNS	central nervous system
CO_2	carbon dioxide
CoA	coenzyme A
coIP	co-immunoprecipitation
coREST	REST corepressor
CpA	cytidine/adenosine dinucleotide
CpG	cytidine/guanosine dinucleotide
dA	2'-deoxyadenosine
dC	2'-deoxycytidine
ddH_2O	ultrapure water
dG	2'-deoxyguanosine
DG	dentate gyrus
DKO	double knockout
DM-αKG	dimethyl-αKG
DMEM	Dulbecco's Modified Eagle's Medium
DNA	2'-deoxyribonucleic acid
DNMT	DNA methyltransferase
DTT	dithiothreitol
EAAT	excitatory amino acid transporter

EC	entorhinal cortex
E. coli	Escherichia coli
EDTA	ethylenediaminetetraacetic acid
EpiSC	epiblast-like stem cell
EtOH	ethanol
FA	formaldehyde
FAD	flavin adenine dinucleotide (oxidized form)
$FADH_2$	flavin adenine dinucleotide (reduced form)
FDR	false discovery rate
Fe^{2+}	$iron^{2+}$
FBS	fetal bovine serum
FH	fumarate hydratase
GABA	γ-aminobutyric acid
Gadd45b	growth arrest and DNA-damage-inducible, beta
Gdh	glutamate dehydrogenase 1 (protein)
gDNA	genomic DNA
Glud1	glutamate dehydrogenase 1 (gene)
GluR1	glutamate receptor 1
GO	gene ontology
GOT	aspartate aminotransferase (formerly glutamate-oxaloacetate transaminase)
GTP	guanosine triphosphate
HAT	histone acetyltransferase
HDAC	histone deacetylase
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
hESC	human embryonic stem cell
HF	hippocampal formation
HIF	hypoxia-inducible factor
H_2O	water
hPSC	human pluripotent stem cell
HRP	horseradish peroxidase
iBKM	immortalized baby mouse kidney epithelial cells
ICC	immunocytochemistry
IHC	immunohistochemistry
IDH	isocitrate dehydrogenase
IF	immunofluorescence
IgG	immunoglobulin G
iNGNs	small molecule-inducible Neurogenin iPSCs
iPSC	induced pluripotent stem cell
JHDM	Jumonji C domain-containing histone demethylase
KCl	potassium chloride
K_i	inhibition constant
K_m	Michaelis constant
KO	knockout
LC-MS/MS	liquid chromatography tandem mass spectrometry
LDH	lactate dehydrogenase
LIF	leukemia inhibitory factor
lncRNA	long non-coding RNA
LTD	long-term depression

LTP	long-term potentiation
MALDI-TOF-MS	matrix assisted laser desorption ionization – time of flight mass spectrometry
Map2	microtubule-associated protein 2
MAS	malate-aspartate-shuttle
MBD	methyl-CpG-binding domain (of MeCP2)
MDH	malate dehydrogenase
MeCP2	methyl-CpG-binding protein 2
MeOH	methanol
mEPSC	miniature glutamatergic excitatory postsynaptic current
mESC	mouse embryonic stem cell
$MgCl_2$	magnesium chloride
miRNA	micro RNA
MS	mass spectrometry
NaCl	sodium chloride
$NAD(P)^+$	nicotinamide adenine dinucleotide (phosphate) (oxidized form)
NAD(P)H	nicotinamide adenine dinucleotide (phosphate) (reduced form)
NALS	neuron-astrocyte lactate shuttle
NCP	nucleosome core particle
ncRNA	non-coding RNA
NCOR	nuclear receptor corepressor
NeuN	neuronal nuclear protein
Npas4	neuronal PAS domain protein 4
NPC	neural progenitor cell
NP complex	nuclear pore complex
NMDR	N-methyl D-aspartate receptors
O_2	oxygen
OAA	oxaloacetate
Oct4	octamer binding transcription factor 4
OxPhos	oxidative phosphorylation
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDC	pyruvate dehydrogenase complex
PIC	protease inhibitor cocktail
pK_a	acid dissociation constant (logarithmic scale)
PLA	proximity ligation assay
PLP	pyridoxal phosphate
PPP	pentose phosphate pathway
Psat1	phosphoserine aminotransferase 1
PTM	post-translational modification
REST	repressor element 1-silencing transcription factor
RFU	relative fluorescent unit
RNA	ribonucleic acid
RNA-seq	RNA-sequencing
RT-qPCR	quantitative real-time PCR
RTT	Rett syndrome
SAM	*S*-adenosyl-methionine
SDH	succinate dehydrogenase

SDS	sodium dodecyl sulfate
Sirt1	sirtuin 1
SMRT	silencing mediator of retinoic acid and thyroid hormone receptor
SYN1	synapsin 1
T	thymidine
TAE	Tris-acetate-EDTA
TBS	Tris-buffered saline
TCA cycle	tricarboxylic acid cycle
TEMED	N,N,N',N'-tetramethyl ethylenediamine
TET (enzyme)	ten-eleven translocation enzyme
TRD	transcriptional repressor domain (of MeCP2)
Tris	2-amino-2-(hydroxymethyl)propane-1,3-diol
TSS	transcription starting site
TTX	tetrodotoxin
Ucp2	uncoupling protein 2

Introduction

Life is based on an enormously complex interaction network of molecules and a major goal of biological science is to understand these networks and ultimately to manipulate them in a desired way. A classical approach towards these aims is to change one parameter experimentally, e.g. to knock-out a single gene or change an important nutrient, and subsequently study the effects of this manipulation. However, this approach is often time and material consuming and ideally it should be possible to simulate these experiments for any kind of cell in any kind of environment in advance. But the state of a single cell is already so complex that not even using the computer power nowadays it is possible to simulate all processes within one cell. In principle, all the information about the cell is encoded in its genome, which is the entity of 2'-deoxyribonucleic acid (DNA). DNA is a molecule that forms a double-stranded helix with thymidine (T), 2'-deoxyadenosine (dA), 2'-deoxycytidine (dC) and 2'-deoxyguanosine (dG) as the principle building blocks. These nucleotides are connected by phosphodiester bonds, resulting in a coding DNA-sequence. The sense and anti-sense strand in the DNA are held together by hydrogen bonds resulting from T/dA and dC/dG base pairing (Figure 1A).[28] However, the DNA sequence is only the basic software, which is essential, but not sufficient for proper cell functioning. In general, all cells of a multicellular organism carry identical genomic information, but random mutations at every step of development cause a certain variability of the DNA sequence – a phenomenon which is called mosaicism.[29] However, mosaicism cannot explain the highly orchestrated and heritable differences that cells at different developmental stages and tissues show. The importance of additional information layers beyond the DNA sequence also becomes clear when one recalls that even terminally differentiated cells in culture can be reprogrammed to be pluripotent by changing the culturing conditions.[30, 31] The heritable changes in gene function that do not depend on the DNA sequence are called epigenetic changes by the original definition.[7] Nowadays, the definition of "epigenetics" has broadened and is applied in sense of the Greek prefix "epi", which means "over". Therefore, the term "epigenetic changes" often describes any changes in gene function that do not depend on DNA sequence alterations.

Genomic DNA (gDNA) is not naked in eukaryotic cells but is organized in a highly conserved DNA/protein polymer that is called chromatin (Figure 1).[32] Chromatin is very dynamic and epigenetic layers define chromatin organization, which in turn decides whether the DNA is accessible for ribonucleic acid (RNA) polymerases to start gene transcription or not.[7, 33] Modifications of the DNA nucleobases (Figure 1A) which can be recognized by specific reader proteins[34] are the first layer of epigenetic information and are described in more detail in the following subchapter *The Chemistries and Consequences of DNA and RNA methylation and demethylation*[8]. The second epigenetic information layer consists of post-translational modifications (PTMs) of histone proteins.[35, 36] Histones are small, positively charged proteins being able to bind DNA very efficiently. In complex with DNA, they form the basic organization unit of chromatin – the nucleosome core particle (NCP). The NCP consists of a disc-shaped histone octamer formed by the histone proteins H2A, H2B, H3 and H4 with approximately 145 – 147 base-pairs (bp) of DNA wrapped 1.65 times around.[37] Several residues of the histone proteins can become heavily modified post-translationally (Figure 1B), which in turn can be recognized by specific proteins (readers). Most of these modifications were shown to be dynamic, and enzymes for their introduction (writers) and their removal (erasers) exist.[38] Arginine residues were found to be (mono-, di- or tri-) methylated and deiminated to citrulline. Furthermore, phosphorylation of serine and threonine residues was discovered as well as adenosine-diphosphate (ADP)-ribosylation of glutamate residues and isomerization of proline residues. However, most PTMs of histones are introduced on lysine residues. They were found to be (mono-, di- or tri-) methylated, acetylated, ubiquitylated and sumoylated.[33] All modifications influence transcription, whereby methylation, ubiquitination, ADP-ribosylation, acetylation and phosphorylation are also associated with DNA

repair.[33] One well studied example how chromatin condensation is influenced by histone PTMs are acetylation and phosphorylation. Under physiological conditions, with a pH environment of 7.4, lysine residues are positively charged, since the acid dissociation constant (pK_a, logarithmic scale) is 10.53 for the side chain.[39] Acetylation removes this positive charge and therefore weakens the binding of histone proteins to DNA, resulting in a more open chromatin state (euchromatin or permissive chromatin) with enhanced transcription as a consequence.[40] Phosphorylation has the same effect, since it introduces a negative charge to a previously neutral hydroxyl group.[41] Deacetylation and dephosphorylation have the opposite effect, resulting in denser chromatin. To further condense chromatin to repressive chromatin (heterochromatin), where gene transcription is significantly slowed down or even completely prevented, certain DNA and histone modifications attract histone H1 and additional non-histone proteins that can function as writers, readers or erasers of histone PTMs (Figure 1C).[42] Histone H1 binds the linker DNA between the NCPs and can therefore stabilize the chromatin structure.[43] As a third layer of epigenetic information, long non-coding RNAs (lncRNAs) and other types of non-coding RNA (ncRNA) are often enriched in the chromatin fraction, implying that they play a role in epigenetic processes and chromatin remodeling by recruiting remodeling enzymes.[42, 44-46] Interestingly, lncRNAs that are involved in heterochromatin formation and maintenance are often transcribed from heterochromatin loci that become accessible during DNA replication. The lncRNAs act through base pairing with the DNA and can attract further chromatin remodeling enzymes in case of heterochromatin formation or proteins that shield chromatin modifications from removal in case of heterochromatin maintenance.[45] For example, differentiation and self-renewal of hematopoietic stem cells depend on lncRNAs that control the expression of many transcription factors[47] and X-chromosome inactivation is mediated by the lncRNA Xist[48].

When a cell is dividing, both genetic and epigenetic information must be passed on to the daughter cells. If only genetic information was passed on, lineage preservation would not be possible. However, in contrast to genetic information, epigenetic information is dynamic. For genetic information, semi-conservative DNA replication before division ensures that all the genetic information is divided equally. But how can this be achieved for epigenetic information that is more complex and depends on many factors? Especially during DNA replication prior to mitosis, chromatin is remodeled extensively, allowing to modify epigenetic marks at the replication fork.[7] The transmission of the epigenetically highly relevant dC modification 5-methyl-2'deoxycytidine (5mdC) to the daughter cells is well studied. DNA methyltransferase (DNMT) 1 recognizes hemi-methylated DNA strands after replication and adds the methyl-group to cytosines on the newly synthesized DNA strand accordingly to maintain the methylation pattern.[49] If preservation of the DNA-methylation pattern is not desired, DNA can be either shielded from Dnmt1 activity when ncRNAs are interacting with Dnmt1 and thereby blocking its binding to DNA[50] or Dnmt1 activity can be downregulated by certain PTMs or Dnmt1 can be even excluded from the nucleus.[51, 52] If a similar mechanism exists for other DNA modifications is unknown to date. For histone modifications, several mechanisms are discussed. During DNA replication, the NCP disassembles and reassembles afterwards with the assistance of histone chaperons. Not only the DNA strand has to be replicated, but also the histone octamer must be copied. When the histone octamer reassembles, half of the new octamer can consist of parental histone proteins that already carry the epigenetic information, which in turn can be established on the new histone proteins.[7, 53] Moreover, the histone modifying enzymes often act in big chromatin remodeling complexes and can stay close to the DNA to reestablish the epigenetic marks right after DNA replication.[7, 53]

Taken together, epigenetic modifications allow the cells within one organism to be very different regarding morphology and function, while having the same set of genes. The dynamic character of epigenetic information ensures not only proper differentiation, but also fast and nevertheless precise

adaption when required. Semiconservative replication of DNA facilitates the inheritance of epigenetic modification patterns, while the situation on histone modifications is more complex.

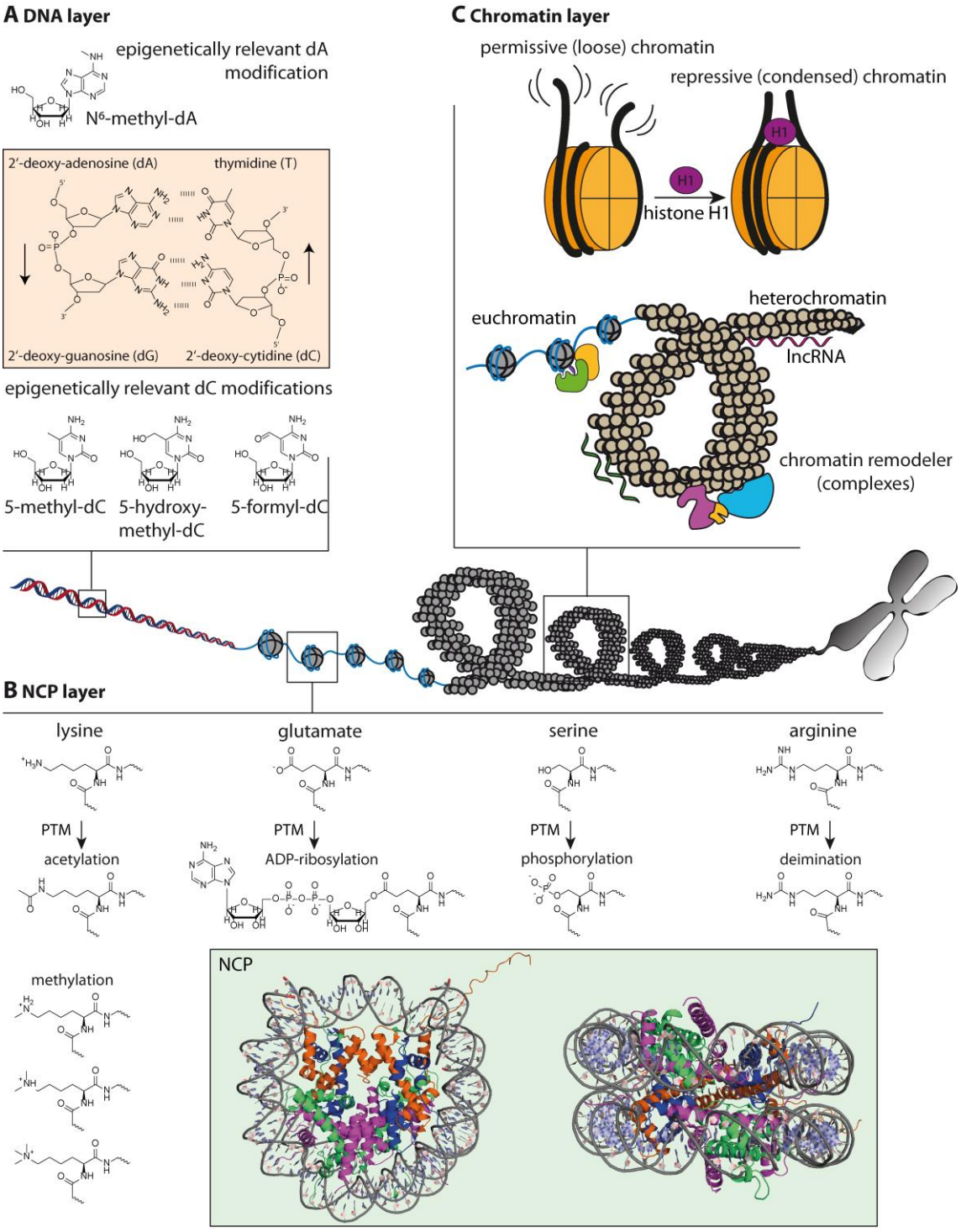


Figure 1: **Organization of DNA as chromatin and epigenetic modifications on chromatin.** Organizational units of DNA, starting from naked DNA that is wrapped around histone octamers to form NCPs. NCPs form euchromatin that can be further condensed to heterochromatin, which eventually leads to chromosome formation during mitosis. (A) Depiction of the canonical DNA building blocks dA, dG, dC and T and epigenetically relevant modifications of the nucleobases. (B) Crystal structure of the NCP[37] and typical PTMs of histone residues. (C) Representation of heterochromatin formation by histone H1, chromatin remodeling enzymes and lncRNA.

dA = 2′-deoxyadenosine, dC = 2′-deoxycytidine, dG = 2′-deoxyguanosine, DNA = 2′-deoxyribonucleic acid, lncRNA = long-noncoding RNA, NCP = nucleosome core particle, PTM = post-translational modification, T = thymidine.

RNA BIOLOGY
2017, VOL. 0, NO. 0, 1–9
https://doi.org/10.1080/15476286.2017.1318241

Taylor & Francis
Taylor & Francis Group

POINT OF VIEW

The chemistries and consequences of DNA and RNA methylation and demethylation

Franziska R. Traube and Thomas Carell

Department of Chemistry, Ludwig-Maximilians-Universität München, Butenandtstrasse, Munich, Germany

ABSTRACT

Chemical modification of nucleobases plays an important role for the control of gene expression on different levels. That includes the modulation of translation by modified tRNA-bases or silencing and reactivation of genes by methylation and demethylation of cytosine in promoter regions. Especially dynamic methylation of adenine and cytosine is essential for cells to adapt to their environment or for the development of complex organisms from a single cell. Errors in the cytosine methylation pattern are associated with most types of cancer and bacteria use methylated nucleobases to resist antibiotics. This Point of View wants to shed light on the known and potential chemistry of DNA and RNA methylation and demethylation. Understanding the chemistry of these processes on a molecular level is the first step towards a deeper knowledge about their regulation and function and will help us to find ways how nucleobase methylation can be manipulated to treat diseases.

ARTICLE HISTORY
Received 10 January 2017
Revised 4 April 2017
Accepted 6 April 2017

KEYWORDS
Cytosine modifications; DNA modifications; Epigenetics; methyltransferases; RNA modifications; TET enzymes

Since the discovery of (deoxy)adenosine (dA, A), (deoxy)cytidine (dC, C), (deoxy)guanosine (dG, G), (deoxy)thymidine (dT, T) and uracil (U) in the early 20th century as the information carrying building blocks, which form the basis for RNA and DNA, various modifications of these nucleosides were discovered (Fig. 1).[1,2] Particularly in transfer-RNA (tRNA) but also in rRNA (rRNA), modified bases are central elements, needed to fine tune the translation of the genetic code.[3-6] In rRNA of bacterial pathogens, many methylated bases are present to block binding of small molecules that work as translation inhibitors, resulting in a resistance against antibiotics such as aminoglycosides.[7] More recently it was discovered that also mRNA (mRNA) contains modified bases. Although it is not yet fully understood what the function of these bases are, it was revealed that the modification chemistry is to some extent reversible. This suggests that the modification and de-modification chemistry has a novel and yet unexplored regulatory function. In this regard N6-methylated adenine (m6A) is the best analyzed modification, but most recently also the reversible formation of N6,C2′-dimethyl adenine (m6Am) was discovered. According to current knowledge, reversible chemistry on modified RNA bases is limited to methyl groups, which are introduced by methyltransferases and removed by demethylases. DNA, in contrast, as the prime carrier of genetic information in the biosphere, is structurally less complex and only few modified bases are known. Most prominent is the methylated base 5-methyl deoxycytosine (5mdC). Ideas about the potential chemistry of methylations and demethylation are the focus of this review. For other aspects, the following excellent reviews can be consulted.[1,2]

5mdC is the most abundant modified base in genomic DNA of eukaryotes and also present in the DNA of prokaryotes.[8-10] In mammals, 5mdC typically reaches global levels between 1 and 5% in genomic DNA.[11,12] Methylated adenine (6mA), which is the DNA equivalent to m6A in RNA, is another DNA modification that is under intensive investigation at the moment. Whereas 6mA is a well-characterized modification in bacterial DNA, its presence was only recently shown in several higher eukaryotic organisms.[13,14] In *Caenorhabditis elegans*, where 5mdC is not detectable, 6mA is dynamically regulated and linked to other epigenetic marks[15] and in early embryos of *Drosophila melanogaster*, 6mA levels are high, but decrease fast during development, resulting in very low 6mA levels in adult tissue.[16] In the unicellular green alga *Chlamydomonas reinhardtii*, 6mA was discovered in 84% of the genes, where it is mainly located at transcription start sites.[17] Recently, it was reported that mammalian DNA, including human and mouse, also contains 6mA.[13,18] There, 6mA seems to be distributed across the genome, but absent in gene exons,[19] and 6mA-demethylation in mouse embryonic stem cell (mESC) DNA was shown to correlate with ALKBH1 depletion.[18] These findings question the previous paradigm that DNA modifications in mammalian genome are limited to cytosine residues. However, when our group tried to confirm these results by a novel ultrasensitive UHPLC-MS method, we were not able to detect 6mA in mESC DNA or DNA from mouse tissue, whereas *Chlamydomonas* DNA, which served as a positive control, delivered the expected positive result.[20] These observations suggest that 6mA might be present at defined time points in mammalian DNA, but is not an epigenetic mark. In the coming years, the question whether 6mA is a relevant modification in mammalian DNA or not will thus certainly be under intensive investigation.

CONTACT Franziska R. Traube franziska.traube@cup.uni-muenchen.de; Thomas Carell Thomas.Carell@cup.uni-muenchen.de LMU Munich, Butenandtstr 5 -13, Munich 81377, Germany.

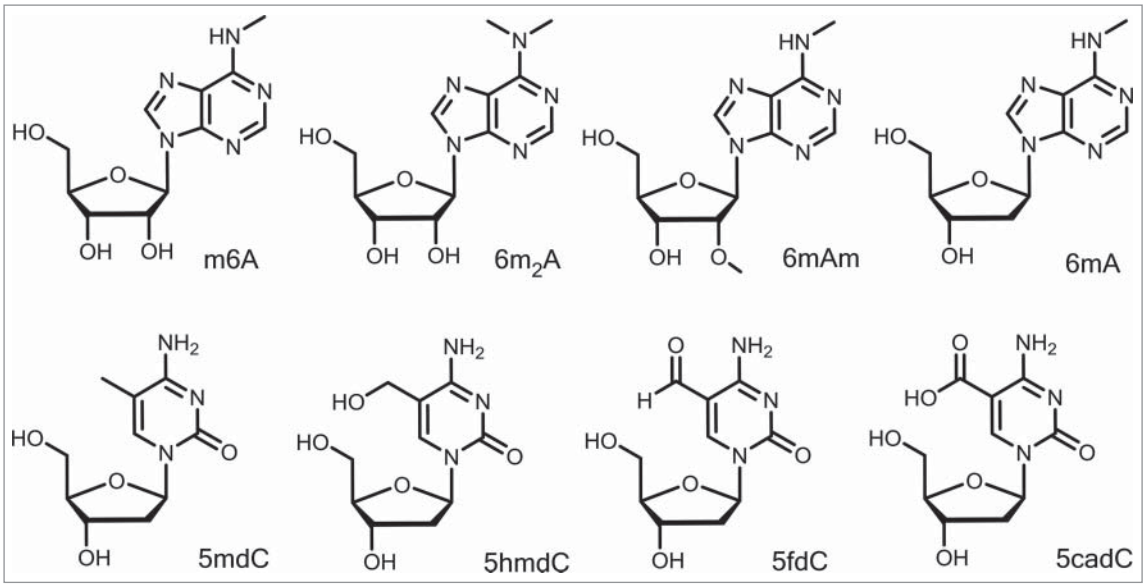


Figure 1. Examples of methylated and oxidized bases found in RNA and DNA.

Chemistry of RNA and DNA base methylation

The addition of the methyl-group to DNA and RNA bases (Fig. 2) is catalyzed by DNA- and RNA-methyltransferases that use *S*-adenosyL-methionine (SAM) as an active methyl-group donor.[21-23] While the methyltransferases that methylate RNA bases are now under extensive investigations, the enzymes that catalyze the methylation of dC in DNA are well characterized. In mammalian cells, 3 active DNA-methyltransferases (DNMTs: DNMT1, DNMT3a and DNMT3b) exist.[24,25] DNMT3a and 3b are *de novo* DNMTs, which methylate canonical dC bases.[26] In contrast, DNMT1 maintains the methylation status during cell division. DNMT1 operates on hemi-methylated DNA during replication, where the template strand is already methylated, but the newly synthesized strand is lacking methylation.[27] As such, DNMT1 converts the methylation of dC into an inheritable modification that can be transferred during reproduction.[28,29]

DNMTs and thus cytosine methylation is essential in those multicellular organisms, where it exists. The presence or absence of 5mdC is associated with various important cellular functions, such as transcription control, X-chromosome silencing and genomic imprinting.[28] A global deletion of only one of the 3 DNMTs leads to severe cellular aberrations and is therefore lethal in early embryogenesis (DNMT1 and 3b) or postnatal (DNMT3a).[26,30] During differentiation the "methylome" is highly dynamic and a celltype-characteristic 5mdC pattern is established during this process.[31] While 5mdC is located to a CpG-dinucleotide context in the majority of somatic cells, non-CpG methylation is also present in embryonic stem cells, many pluripotent progenitor cells and adult brain. However, CpG-methylation is also dominating here.[32-34] Cytosine-methylation in vertebrates occurs in all types of DNA sequence contexts, including repetitive and regulatory sequences, genes and transposable elements; in contrast to invertebrates, where mostly repetitive sequences are methylated.[35] The majority of cytosines in a CpG-context, depending on the cell type up to 80%, are methylated, leaving so-called CpG islands (CGI) of actively transcribed genes as unmethylated patterns in a CpG-context.[36,37]

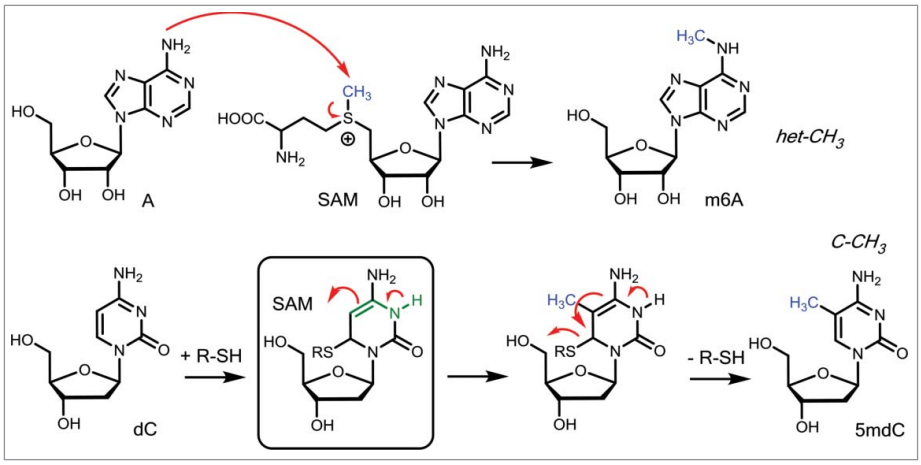


Figure 2. Mechanism of methylation leading to the formation of *het-CH3* and *C-CH3* connectivities in RNA and DNA.

CGIs are regions of high CpG frequency over a length of at least 500 base pairs compared with the bulk genomic DNA and found in 40% of promoter regions in the mammalian genome, with even higher levels (60%) in the human genome.[38,39] Symmetric methylation of CpG:GpC islands is consequently a hallmark of silenced genes.[40,41]

The enzymatic mechanism of how methyltransferases methylate DNA and RNA bases is shown in Fig. 2. Centers with a certain nucleophilicity like the amino group of the RNA base A can attack the SAM coenzyme directly leading to immediate methylation. This type of direct methylation is certainly operating for the formation of $6m_2A$, 4mC or m6Am. SAM as nature's "methyl iodide" is hence reactive enough to methylate even weak nucleophilic centers such as the exocyclic amino groups of A, which feature, as an sp^2-hybridized N-atom only a very weak nucleophilic lone pair at the *N*-atom. This type of direct methylation creates bases, which possess the methyl group attached to a heteroatom establishing a *het-CH_3* system. This will be important in the context of active demethylation (*vide infra*).

In contrast to the formation of *het-CH_3* connections, methylation of the dC base in DNA at position C5 is far more complex. The C5-center features no nucleophilicity at all, making direct methylation impossible. Nature solves this problem by exploiting a helper nucleophile (R-SH, Fig. 2). The DNMT enzymes attack the dC base first with a nucleophilic thiol in a 1,6 addition reaction. This establishes a nucleophilic enamine substructure (green in Fig. 2), which can subsequently be methylated with the SAM cofactor. Importantly, the helper nucleophile is subsequently eliminated, thereby re-establishing the aromatic system. This more complex enzymatic transformation allows nature to methylate non-nucleophilic carbon atoms to create *C-CH_3* connectivities which feature a strong and stable C-C single bond.

Chemistry of demethylation

To establish the reversibility needed for switching biochemical processes, nature requires to remove the attached methyl groups. Removal of het-CH_3 groups found predominantly in RNA was found to occur with the help of α-ketoglutarate (α-KG) dependent oxidases. These proteins contain a reactive Fe(II) center, which reacts to a strongly oxidizing Fe(IV) = O species with oxygen under concomitant decarboxylation of α-KG to succinate (Fig. 3).[42] The Fe(IV) = O species is able to abstract a H-atom from the het-CH_3 group to form a het-stabilized het-CH_2• radical, which reacts with the Fe-bound hydroxylradical to form a *het-CH_2-OH* hemiaminal/acetal functionality.

However, these structures are unstable. In water, they decompose in a spontaneous reaction under loss of formaldehyde to give the unmethylated compound. It is interesting that formaldehyde is formed as a byproduct of this reaction because it is typically a rather toxic compound. It needs to be seen how this molecule is detoxified in the context of the demethylation reaction. Particularly well studied is the removal of the N6-methyl group from m6A to revert into the canonical RNA base A. So far 2 α-KG dependent oxidases were found to catalyze the oxidation. One is the fat mass and obesity-associated protein (FTO) protein and the second is ALKBH5. It was shown, that knockdown of FTO led to increased amounts of m6A and in turn overexpression of FTO resulted in decreased m6A levels.[43] Alkbh5-deficient mice had a similar effect as FTO knockdown in human cells and resulted in increased m6A levels of the mRNA.[44] The demethylation activity of both proteins is comparable, although ALKBH5 shows direct demethylation, whereas FTO-mediated demethylation is supposed to create hm6A and f6A as intermediates.[42,44]

In 2009 it was found that also 5mdC is further enzymatically oxidized in a stepwise fashion to give first 5-hydroxymethyldesoxycytosine (5hmdC), followed by 5-formyldesoxycytosine (5fdC) and 5-carboxydesoxycytosine (5cadC). "Ten-11 translocation" (TET) enzymes, which are Fe^{2+}/α-KG dependent dioxygenases, were discovered to catalyze this iterative 5mdC oxidation reaction.[45,46] Regarding the first oxidation step that transforms 5mdC to 5hmdC, the Fe^{2+}/α-KG catalyzed reaction generates a stable *C-CH_2-OH* connectivity, which is as a primary alcohol stable in water (Fig. 3). 5hmdC is consequently a

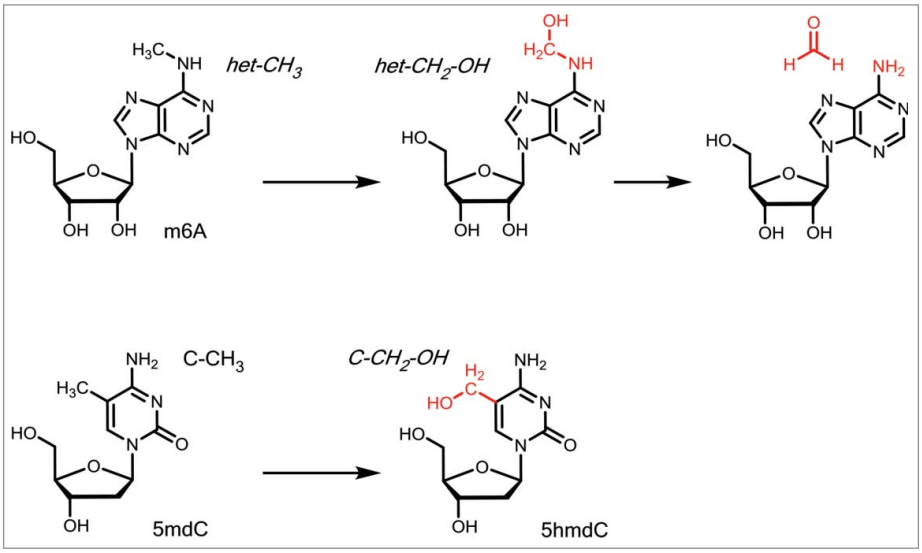


Figure 3. Oxidation of m6A followed by decomposition of the hemiaminal to A and oxidation if 5mdC to stable 5hmdC.

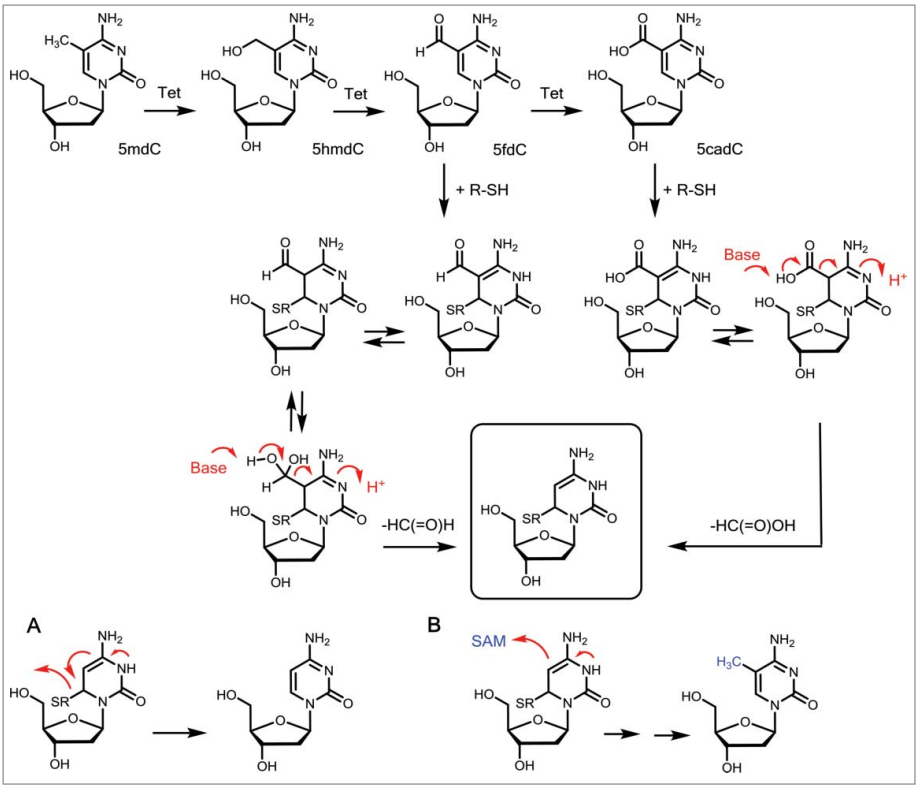


Figure 4. Potential mechanism of chemically induced active demethylation with a potential immediate re-methylation.

stable DNA base modification and it was suggested that the base has indeed epigenetic functions. For example, 5hmdC constitutes 0.6% of all nucleotides in Purkinje neurons, a special neural cell type of the cerebellum, and 0.032% of all nucleotides in embryonic stem (ES) cells.[45,47] The highest 5hmdC levels in fully differentiated tissues were found in the brain with up to 1% of all cytosines.[48,49] Evidence accumulates that 5hmdC in a given gene is able to accelerate transcription and it is not surprising that 5hmdC is mainly present in the promoter of actively transcribed genes.[50,51]

TET enzymes are in this sense required to orchestrate the transcriptional activity of genes. In vertebrates, TET proteins exist in 3 different types (TET1 – TET3) that do not differ regarding their chemistry, but seem to have different spatiotemporal activity. Whereas TET1 is mostly expressed in stem cells, TET3 is upregulated during differentiation and the most abundant TET enzyme in fully differentiated cells.[52-54] A global TET3-knockout is lethal in embryogenesis, because it prevents epigenetic reprogramming during differentiation.[55] It is interesting, that the presence of 5hmdC in mammalian DNA was described first already in 1972.[56] It took more than 30 y to confirm that 5hmdC is really present in substantial amounts that are highly depending on the cell and tissue type.[57]

The further oxidized bases 5fdC and especially 5cadC (Fig. 4) could not be associated yet with distinct cellular functions, but for 5fdC it was reported that it might have regulatory purposes and is also a stable epigenetic mark.[58] In accordance with these previous findings, a recently reported single-cell 5fdC-sequencing method called CLEVER-seq revealed that the generation of 5fdC in promoter regions precedes the upregulation of gene expression.[59] Despite this faint evidence for epigenetic functions, 5fdC and 5cadC are currently mainly considered to be intermediates on the way of an active DNA demethylation process. DNA demethylation is a crucial process of cell development. Especially during fertilization (paternal part of the genome), early embryogenesis (maternal part of the genome) and the development of germ cells, DNA demethylation takes place in a genome-wide manner, allowing a broad reprogramming of the fertilized oocyte and the cells in the early embryo.[60-63] But not only during development, also in fully differentiated cells, it occurs at specific sites of the genome. In brain, for example, locus-specific DNA demethylation and *de novo* methylation is induced by neural activation, arguing that DNA demethylation is important for normal brain function, including memory formation and learning.[64-66] DNA demethylation can take place either actively, which means replication-independent, or passively when DNMT1 does not methylate the nascent DNA strand in hemi-methylated DNA after replication. Passive demethylation occurs, when DNMT1 is absent or blocked during the replication process, which happens for example during early embryogenesis to ensure the demethylation of the maternal genome.[67] Interestingly, 6mA demethylation in *Drosophila* is catalyzed by *Drosophila's* TET homolog (DMAD or dTet). DMAD depletion results in higher 6mA levels, but unchanged 5mdC patterns, and is lethal at pupa stage or shortly after.[16] DMAD and TET possess similar catalytic active Cys-rich and DSBH domains, however, 6mA-demethylation activity was not observed yet for mammalian TET enzymes.[16]

Although oxidation of 5hmdC to 5fdC and 5cadC creates stable molecules due to the lack of a het-atom in β-position, it is discussed that both could be turned into unstable structures

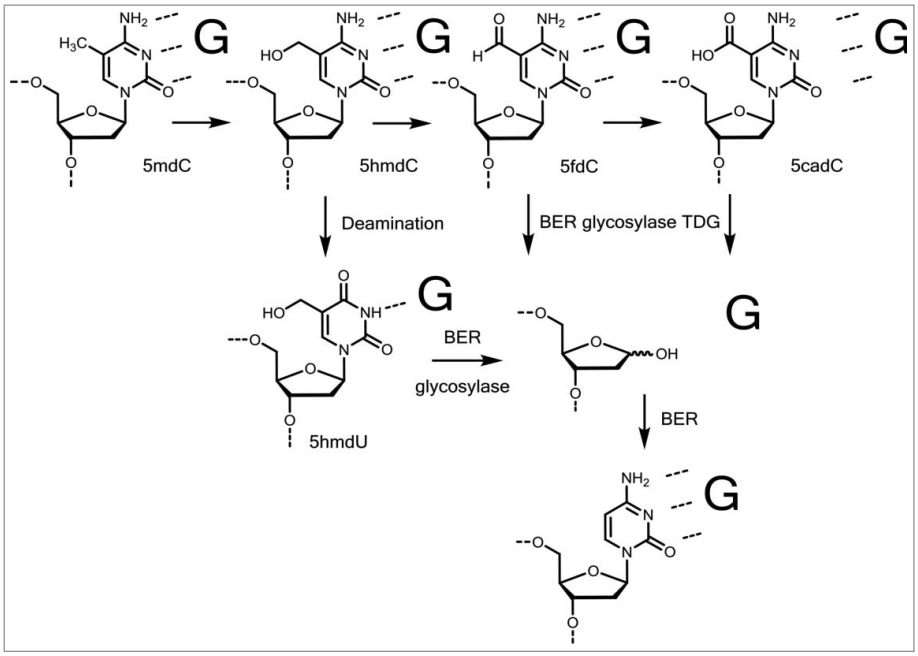


Figure 5. Active demethylation via base excision repair. Two possibilities are discussed: A direct removal of 5fdC and 5cadC in xdC:dG base pairs or removal of a deaminated 5hmdU in a 5hmdU:dG mismatch by BER glycosylase.

upon further chemical manipulation. A chemically attractive mechanism requires that 5fdC and 5cadC are attacked by a helper nucleophile, preferentially a thiol group at the C6 position, in a Michael-type reaction (Fig. 4). Hydratization of 5fdC and tautomerization of the reacted 5fdC and 5cadC allows us to formulate a "β-imino-type" substructure that is prone to deformylation and decarboxylation (red arrows in Fig. 4). Indeed, we could show that reaction of 5fdC and 5cadC with a thiol-nucleophile leads to spontaneous deformylation and decarboxylation showing that the suggested chemistry is feasible. There is currently no evidence that this type of chemistry occurs *in vivo* but we could show that stem cell lysates feature a decarboxylating activity.[68] Interesting is the observation that deformylation and decarboxylation of 5fdC and 5cadC after reaction with a thiol nucleophile leads to a reaction intermediate (boxed in Fig. 2 and 4) that is the key intermediate observed already during methylation of dC to 5mdC by the DNMTs. It is therefore tempting to speculate that DNMT enzymes are involved in the deformylation and decarboxylation maybe followed by immediate re-methylation. Although this reaction sequence would follow chemical logic, it needs to clarified in the near future, if such reactions occur indeed in nature. It was, however, shown that C5-DNA-methyltransferases are indeed able to remove formaldehyde from 5hmdC, converting 5hmdC directly to dC, therefore supporting these ideas.[69]

In this context, it is interesting to note that 5hmC and 5fC were also discovered in RNA. In human cells at tRNA position C34, the oxidation of the corresponding RNA base 5mrC to 5frC is catalyzed by the Fe^{2+}/α-KG dependent enzyme ALKBH1, which is also responsible for m1 A demethylation in mammalian tRNA.[70,71] Interestingly, 5hmrC was not detected as an intermediate in the ALKBH1-dependent 5mrC oxidation.[70] In *Drosophila*, 5hmrC was discovered in polyadenylated RNA and is associated with enhanced mRNA-translation efficiency back to normal level, when 5mrC has lowered the efficiency.[72] Surprisingly, the oxidation reaction is catalyzed by *Drosophila's* TET homolog dTet that is also responsible for 6mA demethylation, but does not oxidize 5mdC.[16,72] Moreover, there is evidence that TET enzymes are also responsible for 5mrC oxidation,[73,74] but at the moment it is not clear whether TET-mediated 5hmrC or 5frC formation are stable or rather transient modifications.

In contrast to the chemical mechanism of active demethylation discussed above, strong evidence exists that active demethylation via formation of 5fdC and 5cadC is also linked to base excision repair (BER), which repairs also mismatches caused by deamination of 5hmdC to 5hmdU (Fig. 5). This mechanism includes excision of 5dfC or 5cadC and subsequent activation of BER. The dG/dT mismatch specific thymine DNA glycosylase (TDG) recognizes dG/dT mismatches, but with an even higher activity it excises 5fdC or 5cadC, but not 5mdC and 5hmdC, *in vitro*.[75] This reactivity was not observed for other DNA glycosylases. Evidence that TDG excises 5fdC and 5cadC also *in vivo* is given by the fact that 5fdC and 5cadC levels are 5–10 times increased in TDG-deficient ES cells compared with the wildtype.[76] However, TET/TDG-mediated demethylation is very unlikely to be the only demethylation mechanism. It rather occurs at defined promoter regions in the genome than in a genome-wide manner. First, TDG-activity causes abasic sites.[77] If this happened genome-wide, it may impair genomic stability, which is crucial for correct development. Second, TDG knockout starts to be lethal not before embryonic day 12.5 and TDG levels are very low in the zygote, where the paternal genome is demethylated.[78,79]

Most recently it was suggested that nature may not need to oxidize 5mdC to 5fdC and 5cadC for demethylation and that a third TET-independent pathway has to exist. In the zygote, the most drastic demethylation occurs when 5mdC is globally erased from the paternal part of the genome, while the maternal part is shielded from demethylation. DNA-demethylation of

the paternal pro-nuclei is replication- and TET-independent, since 5hmdC levels increase after 5mdC levels have dropped and global demethylation can be detected in Tet3-deficient zygotes.[80] It might be that deamination of genomic 5mdC to dT and subsequent dT/dG mismatch repair are the mechanism behind this observation.[81] However, this would also impair genomic stability.

Implication of misguided methylation and demethylation

Whereas the distribution of 5mdC and 5hmdC is tightly regulated to ensure the anticipated functionality of a cell and its response to DNA damage, one hallmark of cancer cells is their completely different methylation and hydroxyL-methylation pattern.[82,83] In many cancer types, the global methylation levels are decreased, while promoter regions of important regulatory and tumor suppressor genes are hypermethylated and therefore silenced.[84] One example is the hypermethylation of the promoter region of HIC1, which is a transcriptional repressor of SIRT1, a survival protein (proto-oncogene) that is consequently upregulated.[85] It was also shown that certain CGIs are coordinately methylated in some tumor cells, which is called "CpG island methylator phenotype."[86-88] Since TET1 was initially discovered in 2002 as a fusion protein to MLL1 H3K4 methyltransferase in patients with acute lymphoblastic leukemia, which is characterized by mutations in the MLL1 protein, it was considered to be an oncogene.[89] Only when the biologic function of the TET enzymes was elucidated in 2009, it was proven a few years later that TET enzymes are actually tumor suppressors that are silenced in various types of tumors. Decreased levels and activity of TET1 and therefore reduced 5hmdC levels are associated with haematopoietic malignancies, colon, breast, prostate, liver and lung cancers, which show greater levels of proliferation and for breast cancer increased invasion rates as a direct consequence of TET1 downregulation.[90-93] TET2 mutations and consequently decreased 5hmdC levels occur in various myeloid malignancies, including chronic myelomonocytic leukemia, myeloid proliferative neoplasm and acute myeloid leukemia.[82,94,95] Mutations in the genes of isocitrate dehydrogenase (IDH) 1 and 2 lead to the production of D-2-hydroxyglutarate (D-2HG), a metabolite that inhibits TET2 activity for example in AML and MPN, but also in malignant gliomas, resulting in a dramatic decrease of 5hmdC levels.[96] Interestingly, IDH1/2 and TET2 mutations seem to be mutually exclusive in these types of tumor, with IDH1/2 mutations being the ones with higher oncogenic potential.[97,98] Recent results show that not only mutations in TET genes or their inhibition by cancer metabolites are important for tumorigenesis, but also tumor hypoxia is responsible for reduced TET activity.[99]

There is more and more evidence that epigenetics and metabolism are closely connected not only via D-2HG in cancer metabolism, but also in normal cells.[100,101] As an intermediate in the tricitric acid (TCA) cycle and part of nitrogen catabolism through deamination of glutamate, α-KG is one of the key metabolites. Since it is the co-substrate of TET enzymes and other dioxygenases involved in epigenetic regulation, such as histone lysine demethylase, it links epigenetics directly to metabolism. Levels of α-KG are rate limiting for TET activity and higher α-KG levels result in higher TET activity with direct impact on differentiation processes.[102] Depending on the cell type and status, α-KG can either promote self-renewal or induce differentiation.[103] In brown adipose tissue (BAT) development, for example, TET3 mediates cell commitment to BAT by demethylating the Prdm16 promoter. AMP activated protein kinase α1 (AMPKα1) influences α-KG levels positively and therefore increase TET3 activity.[104] Glutamate and glutamine metabolism increases α-KG levels, leading to self-renewal in pluripotent mouse embryonic stem cells, while succinate supply leads to differentiation.[102] Additionally, succinate and also fumarate, another 2 intermediates of TCA cycle, show an inhibitory effect on TET enzymes *in vitro*.[105]

In the future, it will be challenging not only to prove the existence, but to reveal the distinct biologic functions of the various DNA and RNA modifications that exist. The role of the modified bases in mRNA are currently under extensive investigation and for DNA, especially the functions of 5hmdC in regulatory and learning processes in brain, but also during development and in cancer cells are of great interest.

Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.

Acknowledgments

FRT thanks the Boehringer Ingelheim Fonds for a PhD fellowship. We thank the Deutsche Forschungsgemeinschaft for financial support through the programs: SFBs 646, 749 and 1032, as well as the SPP1784. Further support is acknowledged from the Excellence Cluster CiPS^{M} (Center for Integrated Protein Science).

References

1. Breiling A, Lyko F. Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. Epigenetics Chromatin 2015; 8:24; PMID:26195987; https://doi.org/10.1186/s13072-015-0016-6
2. Chen K, Zhao BS, He C. Nucleic acid modifications in regulation of gene expression. Cell Chem Biol 2016; 23:74-85; PMID:26933737; https://doi.org/10.1016/j.chembiol.2015.11.007
3. Duechler M, Leszczynska G, Sochacka E, Nawrot B. Nucleoside modifications in the regulation of gene expression: Focus on tRNA. Cell Mol Life Sci 2016; 73:3075-95; PMID:27094388; https://doi.org/10.1007/s00018-016-2217-y
4. Penzo M, Galbiati A, Trere D, Montanaro L. The importance of being (slightly) modified: The role of rRNA editing on gene expression control and its connections with cancer. Biochim Biophys Acta 2016; 1866:330-8; PMID:27815156; https://doi.org/10.1016/j.bbcan.2016.10.007
5. Sloan KE, Warda AS, Sharma S, Entian KD, Lafontaine DL, Bohnsack MT. Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. RNA Biol 2016:1-16; PMID:27911188; https://doi.org/10.1080/15476286.2016.1259781
6. El Yacoubi B, Bailly M, de Crecy-Lagard V. Biosynthesis and function of posttranscriptional modifications of transfer RNAs. Annu Rev Genet 2012; 46:69-95; PMID:22905870; https://doi.org/10.1146/annurev-genet-110711-155641
7. Lioy VS, Goussard S, Guerineau V, Yoon E-J, Courvalin P, Galimand M, Grillot-Courvalin C. Aminoglycoside resistance 16S rRNA methyltransferases block endogenous methylation, affect translation efficiency and fitness of the host. RNA 2014; 20:382-91; PMID:24398977; https://doi.org/10.1261/rna.042572.113

8. Ehrlich M, Wang RY-H. 5-Methylcytosine in eukaryotic DNA. Science 1981; 212:1350-7; PMID:6262918; https://doi.org/10.1126/science.6262918
9. Kahramanoglou C, Prieto AI, Khedkar S, Haase B, Gupta A, Benes V, Fraser GM, Luscombe NM, Seshasayee AS. Genomics of DNA cytosine methylation in escherichia coli reveals its role in stationary phase transcription. Nat Commun 2012; 3:886; PMID:22673913; https://doi.org/10.1038/ncomms1878
10. Ehrlich M, Gama-Sosa MA, Carreira LH, Ljungdahl LG, Kuo KC, Gehrke CW. DNA methylation in thermophylic bacteria: N^4-methylcytosine, 5-methylcytosine, and N^6-methyladenine. Nucleic Acids Res 1985; 13:1399-412; PMID:4000939; https://doi.org/10.1093/nar/13.4.1399
11. Ehrlich M, Gama-Sosa MA, Huang L-H, Midgett RM, Kuo KC, McCune RA, Gehrke C. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. Nucleic Acids Res 1982; 10:2709-21; PMID:7079182; https://doi.org/10.1093/nar/10.8.2709
12. Globisch D, Münzel M, Müller M, Michalakis S, Wagner M, Koch S, Brückl T, Biel M, Carell T. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PLoS One 2010; 5:e15367; PMID:21203455; https://doi.org/10.1371/journal.pone.0015367
13. Koziol MJ, Bradshaw CR, Allen GE, Costa AS, Frezza C, Gurdon JB. Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications. Nat Struct Mol Biol 2016; 23:24-30; PMID:26689968; https://doi.org/10.1038/nsmb.3145
14. Liu J, Zhu Y, Luo GZ, Wang X, Yue Y, Wang X, Zong X, Chen K, Yin H, Fu Y. Abundant DNA 6mA methylation during early embryogenesis of zebrafish and pig. Nat Commun 2016; 7:13052; PMID:27713410; https://doi.org/10.1038/ncomms13052
15. Greer EL, Blanco MA, Gu L, Sendinc E, Liu J, Aristizabal-Corrales D, Hsu CH, Aravind L, He C, Shi Y. DNA methylation on N6-Adenine in C. elegans. Cell 2015; 161:868-78; PMID:25936839; https://doi.org/10.1016/j.cell.2015.04.005
16. Zhang G, Huang H, Liu D, Cheng Y, Liu X, Zhang W, Yin R, Zhang D, Zhang P, Liu J, et al. N6-methyladenine DNA modification in drosophila. Cell 2015; 161:893-906; PMID:25936838; https://doi.org/10.1016/j.cell.2015.04.018
17. Fu Y, Luo GZ, Chen K, Deng X, Yu M, Han D, Hao Z, Liu J, Lu X, Doré LC, et al. N6-methyldeoxyadenosine marks active transcription start sites in chlamydomonas. Cell 2015; 161:879-92; PMID:25936837; https://doi.org/10.1016/j.cell.2015.04.010
18. Wu TP, Wang T, Seetin MG, Lai Y, Zhu S, Lin K, Liu Y, Byrum SD, Mackintosh SG, Zhong M, et al. DNA methylation on N(6)-adenine in mammalian embryonic stem cells. Nature 2016; 532:329-33; PMID:27027282; https://doi.org/10.1038/nature17640
19. Koziol MJ, Bradshaw CR, Allen GE, Costa AS, Frezza C, Gurdon JB. Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications. Nat Struct Mol Biol 2016; 23:24-30; PMID:26689968; https://doi.org/10.1038/nsmb.3145
20. Schiffers S, Ebert C, Rahimoff R, Kosmatchev O, Steinbacher J, Bohne A-V, Spada F, Michalakis S, Nickelsen J, Müller M, et al. Quantitative LC–MS provides no evidence for m6da or m4dc in the genome of mouse embryonic stem cells and tissues. Angew Chem Int Ed 2017; PMID:28371147; https://doi.org/10.1002/anie.201700424
21. Kumar S, Cheng X, Klimasauskas S, Mi S, Postfai J, Roberts RJ, Wilson GG. The DNA (cytosine-5) methyltransferases. Nucleic Acids Res 1993; 22:1-10; PMID:8127644; https://doi.org/10.1093/nar/22.1.1
22. Smith SS, Kaplan BE, Sowers LC, Newman EM. Mechanism of human methyl-directed DNA methyltransferases and the fidelity of cytosine methylation. Proc Natl Acad Sci USA 1992; 89:4744-8; PMID:1584813; https://doi.org/10.1073/pnas.89.10.4744
23. Hori H. Methylated nucleosides in tRNA and tRNA methyltransferases. Front Genet 2014; 5:144; PMID:24904644; https://doi.org/10.3389/fgene.2014.00144
24. Bestor T, Laudano A, Mattaliano R, Ingram V. Cloning and sequencing of a cDNA encoding dna methyltransferase of mouse cells. J Mol Biol 1988; 203:971-83; PMID:3210246; https://doi.org/10.1016/0022-2836(88)90122-2
25. Okano M, Xie S, Li E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat Genet 1998; 19:219-20; PMID:9662389; https://doi.org/10.1038/890
26. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 1999; 99:247-57; PMID:10555141; https://doi.org/10.1016/S0092-8674(00)81656-6
27. Pradhan S, Bacolla A, Wells RD, Roberts RJ. Recombinant human DNA (Cytosine-5) methyltransferase. J Biol Chem 1999; 274:33002-10; PMID:10551868; https://doi.org/10.1074/jbc.274.46.33002
28. Klose RJ, Bird AP. Genomic DNA methylation: The mark and its mediators. Trends Biochem Sci 2006; 31:89-97; PMID:16403636; https://doi.org/10.1016/j.tibs.2005.12.008
29. Spada F, Haemmer A, Kuch D, Rothbauer U, Schermelleh L, Kremmer E, Carell T, Längst G, Leonhardt H. DNMT1 but not its interaction with the replication machinery is required for maintenance of DNA methylation in human cells. J Cell Biol 2007; 176:565-71; PMID:17312023; https://doi.org/10.1083/jcb.200610062
30. Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell 1992; 69:915-26; PMID:1606615; https://doi.org/10.1016/0092-8674(92)90611-F
31. Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Scholer A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D, et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature 2011; 480:490-5; PMID:22170606; https://doi.org/10.1038/nature10716
32. Ramsahoye BH, Biniszkiewicz D, Lyko F, Clark V, Bird AP, Jaenisch R. Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc Natl Acad Sci USA 2000; 97:5237-42; PMID:10805783; https://doi.org/10.1073/pnas.97.10.5237
33. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 2009; 462:315-22; PMID:19829295; https://doi.org/10.1038/nature08514
34. Xie W, Barr CL, Kim A, Yue F, Lee AY, Eubanks J, Dempster EL, Ren B. Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. Cell 2012; 148:816-31; PMID:22341451; https://doi.org/10.1016/j.cell.2011.12.035
35. Suzuki MM, Bird A. DNA methylation landscapes: Provocative insights from epigenomics. Nat Rev Genet 2008; 9:465-76; PMID:18463664; https://doi.org/10.1038/nrg2341
36. Bird A, Taggart M, Frommer M, Miller OJ, Macleod D. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell 1985; 40:91-9; PMID:2981636; https://doi.org/10.1016/0092-8674(85)90312-5
37. Bird AP. CpG-rich islands and the function of DNA methylation. Nature 1986; 321:209-13; PMID:2423876; https://doi.org/10.1038/321209a0
38. Antequera F, Bird A. Number of CpG islands and genes in human and mouse. Proc Natl Acad Sci USA 1993; 90:11995-9; PMID:7505451; https://doi.org/10.1073/pnas.90.24.11995
39. Fatemi M, Pao MM, Jeong S, Gal-Yam EN, Egger G, Weisenberger DJ, Jones PA. Footprinting of mammalian promoters: Use of a CpG DNA methyltransferase revealing nucleosome positions at a single molecule level. Nucleic Acids Res 2005; 33:e176; PMID:16314307; https://doi.org/10.1093/nar/gni180
40. Boyes J, Bird A. DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. Cell 1991; 64:1123-34; PMID:2004419; https://doi.org/10.1016/0092-8674(91)90267-3
41. Siegfried Z, Eden S, Mendelsohn M, Feng X, Tsuberi B-Z, Cedar H. DNA methylation represses transcription in vivo. Nat Genet 1999; 22:203-6; PMID:10369268; https://doi.org/10.1038/9727
42. Zou S, Toh JDW, Wong KHQ, Gao Y-G, Hong W, Woon ECY. N(6)-methyladenosine: A conformational marker that regulates the

substrate specificity of human demethylases FTO and ALKBH5. Sci Rep 2016; 6:25677; PMID:27156733; https://doi.org/10.1038/srep25677
43. Jia G, Fu Y, Zhao X, Dai Q, Zheng G, Yang Y, Yi C, Lindahl T, Pan T, Yang YG, et al. N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat Chem Biol 2011; 7:885-7; PMID:22002720; https://doi.org/10.1038/nchembio.687
44. Zheng G, Dahl JA, Niu Y, Fedorcsak P, Huang CM, Li CJ, Vågbø CB, Shi Y, Wang WL, Song SH, et al. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Mol Cell 2013; 49:18-29; PMID:23177736; https://doi.org/10.1016/j.molcel.2012.10.015
45. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 2009; 324:930-5; PMID:19372391; https://doi.org/10.1126/science.1170116
46. Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science 2011; 333:1300-3; PMID:21778364; https://doi.org/10.1126/science.1210597
47. Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in purkinje neurons and the brain. Science 2009; 324:929-30; PMID:19372393; https://doi.org/10.1126/science.1169786
48. Bachman M, Uribe-Lewis S, Yang X, Williams M, Murrell A, Balasubramanian S. 5-Hydroxymethylcytosine is a predominantly stable DNA modification. Nat Chem 2014; 6:1049-55; PMID:25411882; https://doi.org/10.1038/nchem.2064
49. Wagner M, Steinbacher J, Kraus TF, Michalakis S, Hackner B, Pfaffeneder T, Perera A, Müller M, Giese A, Kretzschmar HA, et al. Age-dependent levels of 5-methyl-, 5-hydroxymethyl-, and 5-formylcytosine in human and mouse brain tissues. Angew Chem Int Ed Engl 2015; 54:12511-4; PMID:26137924; https://doi.org/10.1002/anie.201502722
50. Mellen M, Ayata P, Dewell S, Kriaucionis S, Heintz N. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. Cell 2012; 151:1417-30; PMID:23260135; https://doi.org/10.1016/j.cell.2012.11.022
51. Perera A, Eisen D, Wagner M, Laube SK, Künzel AF, Koch S, Steinbacher J, Schulze E, Splith V, Mittermeier N, et al. TET3 is recruited by REST for context-specific hydroxymethylation and induction of gene expression. Cell Rep 2015; 11:283-94; PMID:25843715; https://doi.org/10.1016/j.celrep.2015.03.020
52. Costa Y, Ding J, Theunissen TW, Faiola F, Hore TA, Shliaha PV, Fidalgo M, Saunders A, Lawrence M, Dietmann S, et al. NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. Nature 2013; 495:370-4; PMID:23395962; https://doi.org/10.1038/nature11925
53. Koh KP, Yabuuchi A, Rao S, Huang Y, Cunniff K, Nardone J, Laiho A, Tahiliani M, Sommer CA, Mostoslavsky G, et al. Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. Cell Stem Cell 2011; 8:200-13; PMID:21295276; https://doi.org/10.1016/j.stem.2011.01.008
54. Szwagierczak A, Bultmann S, Schmidt CS, Spada F, Leonhardt H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. Nucleic Acids Res 2010; 38:e181; PMID:20685817; https://doi.org/10.1093/nar/gkq684
55. Gu TP, Guo F, Yang H, Wu HP, Xu GF, Liu W, Xie ZG, Shi L, He X, Jin SG, et al. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature 2011; 477:606-10; PMID:21892189; https://doi.org/10.1038/nature10443
56. Penn NW, Suwalski R, O'Riley C, Bojanowski K, Yura R. The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. Biochem J 1972; 126:781-90; PMID:4538516; https://doi.org/10.1042/bj1260781
57. Globisch D, Münzel M, Müller M, Michalakis S, Wagner M, Koch S, Brückl T, Biel M, Carell T. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PLoS One 2010; 5:e15367; PMID:21203455; https://doi.org/10.1371/journal.pone.0015367
58. Bachman M, Uribe-Lewis S, Yang X, Burgess HE, Iurlaro M, Reik W, Murrell A, Balasubramanian S. 5-Formylcytosine can be a stable DNA modification in mammals. Nat Chem Biol 2015; 11:555-7; PMID:26098680; https://doi.org/10.1038/nchembio.1848
59. Zhu C, Gao Y, Guo H, Xia B, Song J, Wu X, Zeng H, Kee K, Tang F, Yi C. Single-cell 5-formylcytosine landscapes of mammalian early embryos and escs at single-base resolution. Cell Stem Cell 2017; 20:720-731.e5; PMID:28343982; https://doi.org/10.1016/j.stem.2017.02.013
60. Mayer W, Niveleau A, Walter J, Fundele R, Haaf T. Embryogenesis: Demethylation of the zygotic paternal genome. Nature 2000; 403:501-2; PMID:10676950; https://doi.org/10.1038/35000656
61. Oswald J, Engemann S, Lane N, Mayer W, Olek A, Fundele R, Dean W, Reik W, Walter J. Active demethylation of the paternal genome in the mouse zygote. Curr Biol 2000; 10:475-8; PMID:10801417; https://doi.org/10.1016/S0960-9822(00)00448-6
62. Feng S, Jacobsen SE, Reik W. Epigenetic reprogramming in plant and animal development. Science 2010; 330:622-7; PMID:21030646; https://doi.org/10.1126/science.1190614
63. Sasaki H, Matsui Y. Epigenetic events in mammalian germ-cell development: Reprogramming and beyond. Nat Rev Genet 2008; 9:129-40; PMID:18197165; https://doi.org/10.1038/nrg2295
64. Guo JU, Ma DK, Mo H, Ball MP, Jang M-H, Bonaguidi MA, Balazer JA, Eaves HL, Xie B, Ford E, et al. Neuronal activity modifies the DNA methylation landscape in the adult brain. Nat Neurosci 2011; 14:1345-51; PMID:21874013; https://doi.org/10.1038/nn.2900
65. Morris MJ, Monteggia LM. Role of DNA methylation and the DNA methyltransferases in learning and memory. Dialogues Clin Neurosci 2014; 16:359-71; PMID:25364286
66. Ma DK, Jang M-H, Guo JU, Kitabatake Y, Chang M-l, Pow-anpongkul N, Flavell RA, Lu B, Ming GL, Song H. Neuronal activity–induced gadd45b promotes epigenetic DNA demethylation and adult neurogenesis. Science 2009; 323:1074-7; PMID:19119186; https://doi.org/10.1126/science.1166859
67. Seisenberger S, Peat JR, Reik W. Conceptual links between DNA methylation reprogramming in the early embryo and primordial germ cells. Curr Opin Cell Biol 2013; 25:281-8; PMID:23510682; https://doi.org/10.1016/j.ceb.2013.02.013
68. Schiesser S, Hackner B, Pfaffeneder T, Müller M, Hagemeier C, Truss M, Carell T. Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing. Angew Chem Int Ed Engl 2012; 51:6516-20; PMID:22644704; https://doi.org/10.1002/anie.201202583
69. Liutkevičiūtė Z, Lukinavičius G, Masevičius V, Daujotytė D, Klimašauskas S. Cytosine-5-methyltransferases add aldehydes to DNA. Nat Chem Biol 2009; 5:400-2; PMID:19430486; https://doi.org/10.1038/nchembio.172
70. Haag S, Sloan KE, Ranjan N, Warda AS, Kretschmer J, Blessing C, Hübner B, Seikowski J, Dennerlein S, Rehling P, et al. NSUN3 and ABH1 modify the wobble position of mt-tRNA Met to expand codon recognition in mitochondrial translation. EMBO J 2016; 35:2104-19; PMID:27497299; https://doi.org/10.15252/embj.201694885
71. Liu F, Clark W, Luo G, Wang X, Fu Y, Wei J, Wang X, Hao Z, Dai Q, Zheng G, et al. ALKBH1-mediated tRNA demethylation regulates translation. Cell 2016; 167:816-28.e16; PMID:27745969; https://doi.org/10.1016/j.cell.2016.09.038
72. Delatte B, Wang F, Ngoc LV, Collignon E, Bonvin E, Deplus R, Calonne E, Hassabi B, Putmans P, Awe S, et al. Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. Science 2016; 351:282-5; PMID:26816380; https://doi.org/10.1126/science.aac5253
73. Fu L, Guerrero CR, Zhong N, Amato NJ, Liu Y, Liu S, Cai Q, Ji D, Jin SG, Niedernhofer LJ, et al. Tet-mediated formation of 5-hydroxymethylcytosine in RNA. J Am Chem Soc 2014; 136:11582-5; PMID:25073028; https://doi.org/10.1021/ja505305z
74. Zhang H-Y, Xiong J, Qi B-L, Feng Y-Q, Yuan B-F. The existence of 5-hydroxymethylcytosine and 5-formylcytosine in both DNA and RNA in mammals. Chem Commun 2016; 52:737-40; PMID:26562407; https://doi.org/10.1039/C5CC07354E
75. Maiti A, Drohat AC. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: Potential implications for

active demethylation of CpG sites. J Biol Chem 2011; 286:35334-8; PMID:21862836; https://doi.org/10.1074/jbc.C111.284620
76. He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science 2011; 333:1303-7; PMID:21817016; https://doi.org/10.1126/science.1210944
77. Schomacher L, Han D, Musheev MU, Arab K, Kienhofer S, von Seggern A, Niehrs C. Neil DNA glycosylases promote substrate turnover by Tdg during DNA demethylation. Nat Struct Mol Biol 2016; 23:116-24; PMID:26751644; https://doi.org/10.1038/nsmb.3151
78. Cortellino S, Xu J, Sannai M, Moore R, Caretti E, Cigliano A, Le Coz M, Devarajan K, Wessels A, Soprano D, et al. Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. Cell 2011; 146:67-79; PMID:21722948; https://doi.org/10.1016/j.cell.2011.06.020
79. Cortazar D, Kunz C, Selfridge J, Lettieri T, Saito Y, MacDougall E, Wirz A, Schuermann D, Jacobs AL, Siegrist F, et al. Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. Nature 2011; 470:419-23; PMID:21278727; https://doi.org/10.1038/nature09672
80. Amouroux R, Nashun B, Shirane K, Nakagawa S, Hill PW, D'Souza Z, Nakayama M, Matsuda M, Turp A, Ndjetehe E, et al. De novo DNA methylation drives 5hmC accumulation in mouse zygotes. Nat Cell Biol 2016; 18:225-33; PMID:26751286; https://doi.org/10.1038/ncb3296
81. Santos F, Peat J, Burgess H, Rada C, Reik W, Dean W. Active demethylation in mouse zygotes involves cytosine deamination and base excision repair. Epigenetics Chromatin 2013; 6:39; PMID:24279473; https://doi.org/10.1186/1756-8935-6-39
82. Ko M, Huang Y, Jankowska AM, Pape UJ, Tahiliani M, Bandukwala HS, An J, Lamperti ED, Koh KP, Ganetzky R, et al. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature 2010; 468:839-43; PMID:21057493; https://doi.org/10.1038/nature09586
83. Esteller M. Epigenetics in Cancer. N Engl J Med 2008; 358:1148-59; PMID:18337604; https://doi.org/10.1056/NEJMra072067
84. Jones PA, Baylin SB. The epigenomics of cancer. Cell 2007; 128:683-92; PMID:17320506; https://doi.org/10.1016/j.cell.2007.01.029
85. Chen WY, Wang DH, Yen RC, Luo J, Gu W, Baylin SB. Tumor suppressor HIC1 directly regulates SIRT1 to modulate p53-dependent DNA-damage responses. Cell 2005; 123:437-48; PMID:16269335; https://doi.org/10.1016/j.cell.2005.08.011
86. Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB. Issa J-PJ. CpG island methylator phenotype in colorectal cancer. Proc Natl Acad Sci USA 1999; 96:8681-6; PMID:10411935; https://doi.org/10.1073/pnas.96.15.8681
87. Weisenberger DJ, Siegmund KD, Campan M, Young J, Long TI, Faasse MA, Kang GH, Widschwendter M, Weener D, Buchanan D, et al. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. Nat Genet 2006; 38:787-93; PMID:16804544; https://doi.org/10.1038/ng1834
88. Issa J-P. CpG island methylator phenotype in cancer. Nat Rev Cancer 2004; 4:988-93; PMID:15573120; https://doi.org/10.1038/nrc1507
89. Ono R, Taki T, Taketani T, Taniwaki M, Kobayashi H, Hayashi Y. 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; 62:4075-80; PMID:12124344
90. Yang H, Liu Y, Bai F, Zhang JY, Ma SH, Liu J, Xu ZD, Zhu HG, Ling ZQ, Ye D, et al. Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. Oncogene 2013; 32:663-9; PMID:22391558; https://doi.org/10.1038/onc.2012.67
91. Neri F, Dettori D, Incarnato D, Krepelova A, Rapelli S, Maldotti M, Parlato C, Paliogiannis P, Oliviero S. TET1 is a tumour suppressor that inhibits colon cancer growth by derepressing inhibitors of the WNT pathway. Oncogene 2015; 34:4168-76; PMID:25362856; https://doi.org/10.1038/onc.2014.356
92. Hsu C-H, Peng K-L, Kang M-L, Chen Y-R, Yang Y-C, Tsai C-H, Chu CS, Jeng YM, Chen YT, Lin FM, et al. TET1 suppresses cancer invasion by activating the tissue inhibitors of metalloproteinases. Cell Rep 2012; 2:568-79; PMID:22999938; https://doi.org/10.1016/j.celrep.2012.08.030
93. Liu C, Liu L, Chen X, Shen J, Shan J, Xu Y, Yang Z, Wu L, Xia F, Bie P, et al. Decrease of 5-hydroxymethylcytosine is associated with progression of hepatocellular carcinoma through downregulation of TET1. PLoS One 2013; 8:e62828; PMID:23671639; https://doi.org/10.1371/journal.pone.0062828
94. Langemeijer SMC, Kuiper RP, Berends M, Knops R, Aslanyan MG, Massop M, Stevens-Linders E, van Hoogen P, van Kessel AG, Raymakers RA, et al. Acquired mutations in TET2 are common in myelodysplastic syndromes. Nat Genet 2009; 41:838-42; PMID:19483684; https://doi.org/10.1038/ng.391
95. Lemonnier F, Couronné L, Parrens M, Jaïs J-P, Travert M, Lamant L, Tournillac O, Rousset T, Fabiani B, Cairns RA, et al. Recurrent TET2 mutations in peripheral T-cell lymphomas correlate with TFH-like features and adverse clinical parameters. Blood 2012; 120:1466-9; PMID:22760778; https://doi.org/10.1182/blood-2012-02-408542
96. Xu W, Yang H, Liu Y, Yang Y, Wang P, Kim S-H, Ito S, Yang C, Wang P, Xiao MT, et al. Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α-ketoglutarate-dependent dioxygenases. Cancer Cell 2011; 19:17-30; PMID:21251613; https://doi.org/10.1016/j.ccr.2010.12.014
97. Inoue S, Lemonnier F, Mak TW. Roles of IDH1/2 and TET2 mutations in myeloid disorders. Int J Hematol 2016; 103:627-33; PMID:26980223; https://doi.org/10.1007/s12185-016-1973-7
98. Shih AH, Abdel-Wahab O, Patel JP, Levine RL. The role of mutations in epigenetic regulators in myeloid malignancies. Nat Rev Cancer 2012; 12:599-612; PMID:22898539; https://doi.org/10.1038/nrc3343
99. Thienpont B, Steinbacher J, Zhao H, D'Anna F, Kuchnio A, Ploumakis A, Ghesquière B, Van Dyck L, Boeckx B, Schoonjans L, et al. Tumour hypoxia causes DNA hypermethylation by reducing TET activity. Nature 2016; 537:63-8; PMID:27533040; https://doi.org/10.1038/nature19081
100. Yang H, Lin H, Xu H, Zhang L, Cheng L, Wen B, Shou J, Guan K, Xiong Y, Ye D. TET-catalyzed 5-methylcytosine hydroxylation is dynamically regulated by metabolites. Cell Res 2014; 24:1017-20; PMID:24971736; https://doi.org/10.1038/cr.2014.81
101. Sciacovelli M, Goncalves E, Johnson TI, Zecchini VR, da Costa AS, Gaude E, Drubbel AV, Theobald SJ, Abbo SR, Tran MG, et al. Fumarate is an epigenetic modifier that elicits epithelial-to-mesenchymal transition. Nature 2016; 537:544-7; PMID:27580029; https://doi.org/10.1038/nature19353
102. Carey BW, Finley LW, Cross JR, Allis CD, Thompson CB. Intracellular alpha-ketoglutarate maintains the pluripotency of embryonic stem cells. Nature 2015; 518:413-6; PMID:25487152; https://doi.org/10.1038/nature13981
103. TeSlaa T, Chaikovsky AC, Lipchina I, Escobar SL, Hochedlinger K, Huang J, Graeber TG, Braas D, Teitell MA. alpha-ketoglutarate accelerates the initial differentiation of primed human pluripotent stem cells. Cell Metab 2016; 24:485-93; PMID:27476976; https://doi.org/10.1016/j.cmet.2016.07.002
104. Yang Q, Liang X, Sun X, Zhang L, Fu X, Rogers CJ, Berim A, Zhang S, Wang S, Wang B, et al. AMPK/alpha-ketoglutarate axis dynamically mediates dna demethylation in the prdm16 promoter and brown adipogenesis. Cell Metab 2016; 24:542-54; PMID:27641099; https://doi.org/10.1016/j.cmet.2016.08.010
105. Laukka T, Mariani CJ, Ihantola T, Cao JZ, Hokkanen J, Kaelin WG Jr., Godley LA, Koivunen P. Fumarate and succinate regulate expression of hypoxia-inducible genes via TET enzymes. J Biol Chem 2016; 291:4256-65; PMID:26703470; https://doi.org/10.1074/jbc.M115.688762

Metabolism, Epigenetics and the Nuclear-Mitochondrial Crosstalk

Metabolic processes do not only ensure the energy supply of the cell but also control every cellular aspect, including self-renewal, proliferation and differentiation.[54] Within the metabolic network of the cell, however, energy metabolism has a pivotal role. Living cells in any situation rely on constant supply with adenosine triphosphate (ATP), which is the universal carrier of energy in biology.[55] Interestingly, basal cellular maintenance even needs more energy than mammalian cells need for proliferation.[56] Glucose is an energy carrier that can be quickly taken up and catabolized by any kind of cell, even under hypoxic conditions, and therefore plays a major role in energy supply.[56] Metabolically very active cells or tissues like tumors or brain mainly catabolize glucose to meet their energy demands.[57-59] For example, the human brain uses 25% of the total amount of consumed glucose, although it only accounts for 2% of the total cell mass.[57]

In principle, glucose metabolism in higher eukaryotes is divided in three basic processes: glycolysis, pentose phosphate pathway (PPP) and glycogenesis. Glycogenesis is an anabolic process for glucose storage as glycogen.[60] The PPP can be anabolic or catabolic and its final products depend on the stage of the cell. When the cell is proliferating, and DNA building blocks are needed, ribose-5-phosphate is the main product. In contrast, when the energy demand is high, the PPP is used for the generation of the reduced nicotinamide adenine dinucleotide phosphate (NADPH) and pyruvate, of which pyruvate is also the final product of glycolysis (Figure 2).[61] The main pathway for glucose catabolism, however, is the interplay of glycolysis, tricarboxylic acid cycle (TCA cycle) and oxidative phosphorylation (OxPhos) (Figure 2). Glycolysis takes place in the cytosol and does not require oxygen (O_2).[62] The TCA cycle takes place in the mitochondrial matrix and the inner mitochondrial membrane. Although the TCA cycle itself does not directly depend on O_2, it is suppressed under hypoxic conditions, because without O_2 it would not be possible to maintain the nicotinamide adenine dinucleotide (NAD^+) pool that is required to keep the TCA cycle running.[63] Finally, OxPhos, which takes place at the inner mitochondrial matrix, provides most of the energy. As the name suggests, OxPhos is strictly O_2-dependent.[64]

Glycolysis catabolizes one molecule of glucose to two molecules of pyruvate under consumption of two molecules of NAD^+, thereby generating two molecules of ATP and NADH net (Figure 2A). Under hypoxic conditions, energy supply relies on glycolysis and the TCA cycle is suppressed. In this situation, the glycolytic rate is increased and pyruvate is reduced to lactate, thereby recovering the cytosolic NAD^+ pool (Figure 2A).[65] However, if O_2 is available, glycolysis is normally linked to the TCA and subsequently to OxPhos. Alternatively, the TCA cycle can be directly fueled by amino acid and fatty acid metabolism (Figure 2B). If glycolysis is the major source of ATP although O_2 is available, the cell is on aerobic glycolysis. This effect is known as the “Warburg effect”, which is a metabolic hallmark of many cancer and other fast proliferating cells. It seems counterintuitive that fast proliferating cells restrict themselves to the, in terms of ATP production, less efficient glycolytic pathway. However, not ATP is the limiting factor for proliferation, but the availability of other biomolecules such as nucleotides, proteins or fatty acids for membranes. Aerobic glycolysis still provides ATP, but nevertheless keeps six carbons available for the production of biomolecules, which is not the case when glucose is finally oxidized to carbon dioxide (CO_2).[59]

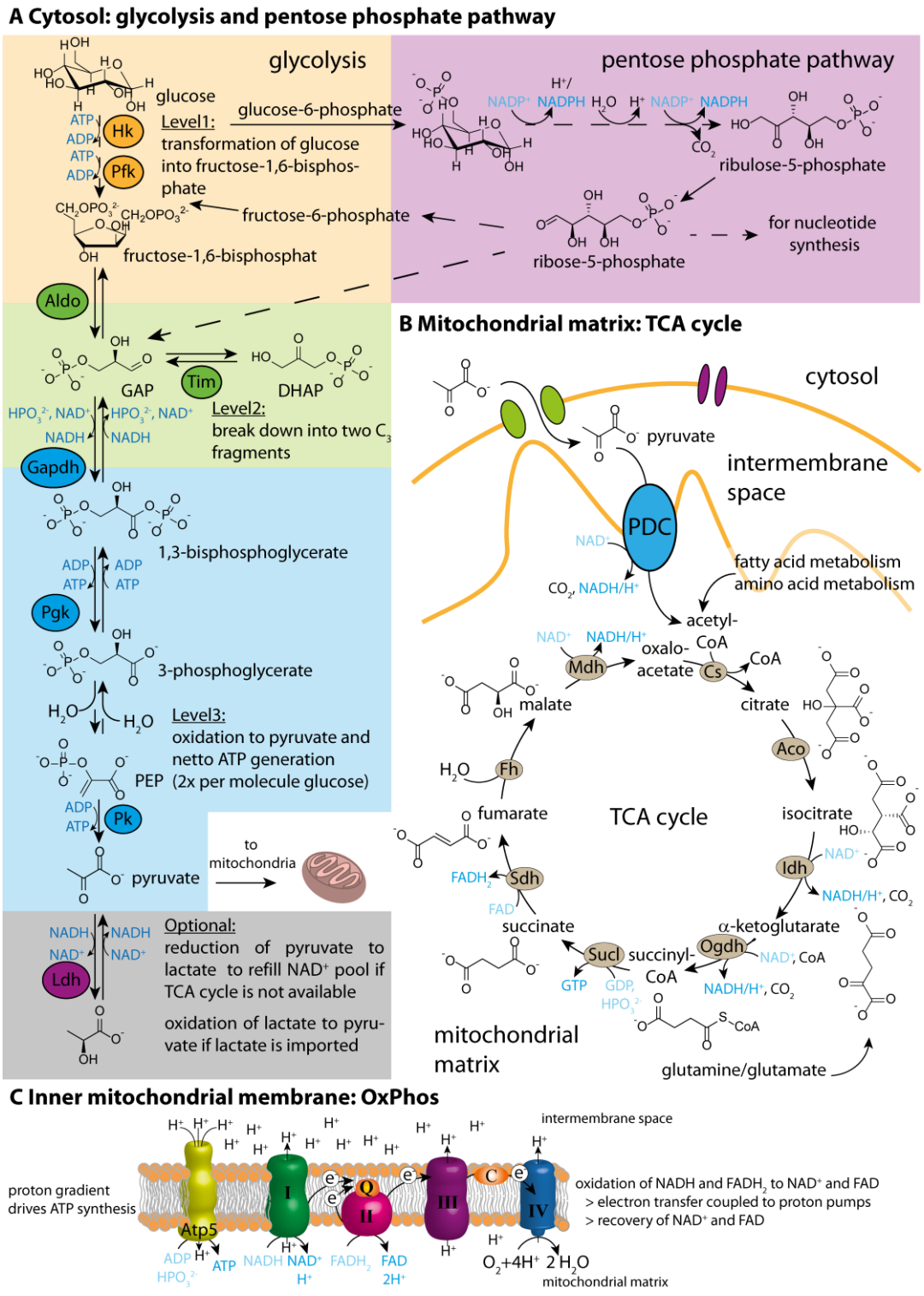


Figure 2: **Depiction of glucose catabolism.** (A) Glycolysis and pentose phosphate pathway, (B) TCA cycle and (C) OxPhos. Dashed arrows indicate more than one reaction step. I = NADH-coenzyme Q oxidoreductase, II = succinate-Q oxidoreductase, III = Q-cytochrome c oxidoreductase, IV = cytochrome c oxidase, Aco = aconitase, ADP = adenosine diphosphate, Aldo = aldolase, ATP = adenosine triphosphate, Atp5 = ATP synthase, C = cytochrome C, CoA = coenzyme A, Cs = citrate synthase, DHAP = dihydroxyacetone phosphate, e^- = electron, Fh = fumarate hydratase, GAP = glyceraldehyde 3-phosphate, Gapdh = glyceraldehyde 3-phosphate dehydrogenase, GDP = guanosine diphosphate, GTP = guanosine triphosphate, Hk = hexokinase, Idh = isocitrate dehydrogenase, Ldh = lactate dehydrogenase, Mdh = malate dehydrogenase, NAD^+/NADH = nicotinamide adenine dinucleotide (oxidized/reduced), Ogdh = α-ketoglutarate dehydrogenase complex, OxPhos = oxidative phosphorylation, Pdh = pyruvate dehydrogenase complex, PEP = phosphoenolpyruvate, Pfk = 6-phosphofructokinase, Pgk = phosphoglycerate kinase, Pk = pyruvate kinase, Q = flavoprotein-Q oxidoreductase, Sdh = succinate dehydrogenase, Sucl = succinyl-CoA-synthetase, TCA cycle = tricarboxylic acid cycle, Tim = triosephosphate isomerase.

The TCA cycle is the core of cellular metabolism, because it is the central common oxidative pathway for carbohydrates, amino acids and fatty acids, where they are eventually oxidized to CO_2 (Figure 2B). Furthermore, it provides precursors for many biomolecules and intermediates of the TCA cycle are components of many other pathways. Carbohydrates, amino acids and fatty acids usually enter the TCA cycle as acetyl-coenzyme A (acetyl-CoA), except for glutamate and glutamine, which enter the TCA cycle as α-ketoglutarate (αKG) after deamination. The link between glycolysis and the TCA cycle is the pyruvate dehydrogenase complex (PDC), which is located in the inner mitochondrial membrane to irreversibly decarboxylate pyruvate and generate acetyl-CoA and NADH.[66] For the final oxidation to CO_2, which is performed by dehydrogenases, three molecules of NAD^+ and one molecule of oxidized flavin adenine dinucleotide (FAD) are reduced as redox equivalents to NADH and $FADH_2$ per molecule acetyl-CoA entering the TCA cycle. Additionally, one molecule of guanosine triphosphate (GTP) is produced.[63]

In the last step, OxPhos, electrons from NADH and $FADH_2$ are transferred to O_2 by a sequence of electron carriers in the inner mitochondrial matrix, eventually resulting in water (H_2O) (Figure 2C). The reduction of O_2 to H_2O is in summary an exothermal reaction and the energy that is released in this process drives proton pumps in the inner mitochondrial membrane. The so established proton gradient between the intermembrane space (high proton concentration) and the mitochondrial matrix (low proton concentration) finally enables the phosphorylation of 26 molecules of ADP to ATP per molecule glucose.[64]

The interplay of glycolysis, TCA cycle and OxPhos can dictate cellular fate. Many intermediates of these fundamental cellular processes are used in other metabolic pathways. Changes in the basic epigenetic program of the cell, e.g. from self-renewal to differentiation, are often accompanied or even induced by metabolic changes. As mentioned above, fast proliferating cells are more limited by the availability of carbon building blocks than by the absolute amount of ATP and therefore often use aerobic glycolysis to balance their need for ATP and for precursor molecules. Differentiated cells, in contrast, need more ATP, e.g. neurons to maintain a membrane potential, but less precursor molecules for DNA or membrane synthesis.[59] Consequently, differentiated cells use TCA and OxPhos to obtain the maximal amount of ATP. A prominent example how cellular metabolism adapts to meet the requirements of the cell and how these adaptions induce further changes, is the metabolic switch that stem cells show upon differentiation. Pluripotent stem cells in culture show a shift from alternatively aerobic glycolysis or OxPhos to exclusively aerobic glycolysis upon priming, and eventually to almost exclusively OxPhos upon further differentiation.[67-70]

In mouse embryonic stem cells (mESCs) that are kept in a naïve state in culture, the steady-state level of the intermediates of the TCA cycle are usually low, except for αKG, indicating that the cells mainly use aerobic glycolysis.[24] In addition to their need of carbon precursor building blocks, naïve mESCs have low DNA methylation and a very open chromatin structure.[71] Aerobic glycolysis ensures high acetate and thereby high acetyl-CoA levels.[72] Acetyl-CoA in turn is a co-substrate for histone acetyl-transferases (HAT), which are responsible for a preferable open chromatin structure in pluripotent cells by establishing an acetyl-group on lysine residues of histones.[73] In contrast, *S*-adenosyl-methionine (SAM) that is a substrate for various methyltransferases, including the DNMTs, is kept at low levels in the pluripotent state, but the levels increase in early differentiation.[74] However, in the naïve state, the cells can switch between aerobic glycolysis and OxPhos on demand.[74, 75] When mESCs undergo priming to become epiblast-like stem cells (EpiSCs), which are still pluripotent, but more committed than naïve cells, they exclusively depend on aerobic glycolysis.[70] EpiSCs proliferate even faster than naïve ESC and still have a very open chromatin structure and therefore heavily rely on carbon precursor molecules for nucleotides, amino and fatty acids and co-factors.[73] Although the mitochondria are in principal functional in the pluripotent state of primed cells despite their different

morphology, uncoupling protein 2 (Ucp2), which is located in the mitochondrial inner membrane[76], facilitates glycolysis and prevents OxPhos. The fact that stimulation of aerobic glycolysis in naïve mESCs drives the cells into an epiblast-like cell fate[75] and in the opposite direction, promotion of OxPhos in EpiSCs allows reestablishment of naïve pluripotency[77], elucidates that metabolic switches are not only a consequence of cell-fate decisions, but also induce them. When Ucp2 is repressed and energy metabolism switches from aerobic glycolysis to OxPhos, EpiSCs further differentiate.[67] However, in the developing organism *in vivo*, this switch only happens in endo- and mesoderm differentiation, but not for ectoderm differentiation which still requires glycolytic flux.[78]

A key metabolite that links these metabolic changes to epigenetic changes, is αKG. Naïve mESCs in culture require glucose and glutamine catabolism to maintain a high intracellular αKG to succinate ratio. Manipulation on that ratio is directly linked to chromatin changes regarding DNA methylation and histone modifications and can therefore force cells under cultured conditions to stay in the naïve state when αKG is high or to start differentiation when succinate is increased.[24, 25] To preserve the naïve pluripotent ground state, mESCs are often kept in a medium that is supplemented with leukemia inhibitory factor (LIF) and two inhibitors (PD0325901 and CHIR99021 (2i)) against glycogen synthase kinase-3 beta and the mitogen-activated protein kinases pathway, respectively.[79, 80] When mESCs are transferred from 2i/LIF conditions into medium supplemented with fetal bovine serum (FBS) and LIF, they change from the naïve to a more committed, in short, a primed phenotype.[24, 81] But the effect of αKG on the ability to keep cultured mESCs in the naïve state is so strong that even when the culturing conditions change from 2i/LIF to FBS/LIF, the shift to a primed phenotype is inefficiently induced when a high αKG to succinate ratio is maintained by external αKG supply.[24] The high αKG levels in naïve mESC are maintained by the phosphoserine aminotransferase 1 (Psat1).[82] Psat1 generates αKG from glutamate by transferring the amino group of glutamate on 3-phospho-oxypyruvate thereby producing αKG and *O*-phospho-L-serine, and is down-regulated during the priming process. This already indicates that when it comes to the regulation of epigenetic processes by metabolites, the control of intracellular αKG levels is crucial.[70, 83-85] However, it is oversimplification to claim that αKG in general promotes self-renewal and succinate induces differentiation. The impacts of αKG and succinate on epigenetic plasticity are different depending on the cell's state. Therefore, local αKG and succinate amounts must be tightly controlled. Whereas a high concentration of cytosolic and nuclear αKG promotes self-renewal when the cells are in a naïve state, it accelerates early differentiation once the cells are in a primed state. Succinate has the opposite effect, both on naïve and primed stem cells.[25] Transition of naïve to primed to differentiated cells is therefore induced by controlled fluctuations of αKG levels.[24, 25, 70, 82] The central roles for αKG and succinate for epigenetic plasticity seem to be based on the fact that αKG is a co-substrate for many epigenome-modifying enzymes such as Jumonji C domain-containing histone demethylases (JHDMs) (Figure 3A) and ten-eleven translocation (TET) enzymes (Figure 3B), which both belong to the family of αKG-dependent dioxygenases.[24, 25, 83] To perform the oxidation reaction, those enzymes decarboxylate αKG to succinate, which in turn is a competitive inhibitor for them.[86, 87]

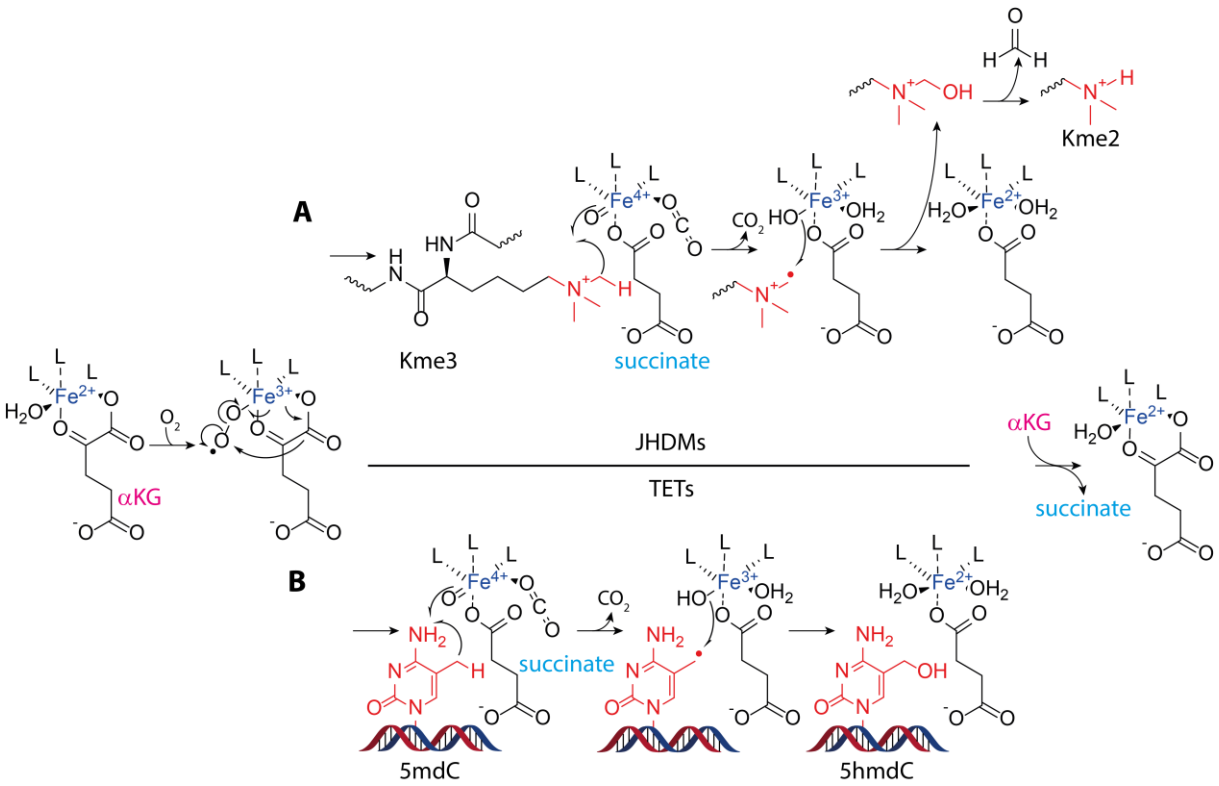


Figure 3: **Reaction mechanisms of αKG-dependent dioxygenases.** (A) Reaction mechanism of JHDMs. (B) Reaction mechanism of TET enzymes. αKG = α-ketoglutarate, JHDMs = Jumonji C domain-containing histone demethylases, TET = ten-eleven translocation enzyme.

In mammals, more than 60 αKG-dependent oxygenases exist and when O_2 is not limited, the availability of αKG is rate limiting for those enzymes.[24, 83, 88, 89] In cultured cells, cell-permeable dimethyl-αKG (DM-αKG) can be added to the medium in millimolar concentrations to increase the intracellular αKG-concentration simply by diffusion through the cell membrane.[24, 25] Furthermore, some cell types, such as human pluripotent stem cells (hPSCs), have membrane transporters for αKG.[25] However, intracellular αKG levels, especially in tissue *in vivo*, can be modulated without external αKG-supply (Figure 4). But how are intracellular αKG levels regulated and balanced between the mitochondria and the cytosol/the nucleus? Every cell can take up glucose or lactate and glutamine or glutamate via membrane transporters and the catabolism of these nutrients balances many metabolite levels in the cell, including αKG.[24, 59, 90] αKG can be generated by aminotransferases, e.g. by aspartate aminotransferases (GOT) or Psat1, which transfer the amino group of glutamate to another substrate and thereby produce αKG (Figure 4). An additional group of αKG-generating enzymes are certain dehydrogenases that perform a $NAD(P)^+$ dependent oxidation reaction. Isocitrate dehydrogenases (IDH), glutamate dehydrogenase 1 (*Glud1*, Gdh) or 2-hydroxyglutarate dehydrogenases (2HGDH) belong to this category (Figure 4). In the cell, glucose/lactate and glutamine/glutamate catabolism and therefore also αKG metabolism are linked to redox chemistry through $NAD(P)^+$ and its reduced form NAD(P)H.[24] $NAD(P)^+$/NAD(P)H are essential redox equivalents, $NAD(P)^+$ for oxidations and NAD(P)H for reductions, to regulate cell homeostasis on every level.[26, 91] A dedicated shuttling system, the malate aspartate shuttle (MAS), exists to maintain redox homeostasis between cytosol and mitochondria. It consists of a malate/αKG and a glutamate/aspartate carrier in the inner mitochondrial membrane, Got1 (cytosolic) and Got2 (mitochondrial) and the malate dehydrogenases (MDH) Mdh1 (cytosolic) and Mdh2 (mitochondrial) (Figure 4).[92, 93] Thereby it also regulates the αKG amount between the cytosol/nucleus, where αKG is needed as cofactor for the αKG-dependent oxygenases, and the mitochondria, where αKG is an intermediate of the TCA cycle (Figure 4).[25, 92] Malfunction of the MAS

by inhibition of Got1/Got2 leads to lower αKG levels in the cytosol, which impairs differentiation of cultured hPSCs, showing the importance of cytosolic αKG concentration control by the MAS for epigenetic plasticity.[25]

Absolute quantification of αKG levels in immortalized baby mouse kidney epithelial (iBKM) and HeLa cells revealed that its steady-state concentration is in the high micromolar to even low millimolar range in those cell lines.[85, 94] Cytosol and nucleus of eukaryotes are separated by a double membrane, called the nuclear envelope, which allows the exchange of small molecules and macromolecules between the two compartments via nuclear pore complexes.[95] These pores allow diffusion of proteins bigger than 60 kDa through them and thus are not expected to be a barrier for αKG. Changes in the gene expression profile presumably result in a concentration gradient of αKG between the cytosol and the nucleus, where αKG is consumed by TET enzymes and JHDMs. The more drastic the change in the gene expression profile is, the more αKG is consequently needed. In contrast, the αKG in a cell that is in a steady state, which means that its gene expression profile is stable, should be equally distributed between the cytosol and the nucleus. Assuming that a typical mammalian cell has a diameter of 20 µm[96] and the nucleus accounts for 10% of the cellular volume and has a round shape[97, 98], the nuclear volume (I) is

$$V_{nucl.} = 0.1 \times \frac{4}{3} \times \pi \times r_{cell}^3 = 0.1 \times \frac{4}{3} \times \pi \times (10 \times 10^{-6} m)^3 \approx 4.2 \times 10^{-16}\, m^3 = 420\, \mu m^3 \text{ (I)}.$$

In the steady state, an αKG concentration of 1 mM means a total nuclear αKG amount (II) of

$$n_{\alpha KG(nucl.)} = c_{\alpha KG(nucl.)} \times V_{nucl.} = 10^{-3} \frac{mol}{l} \times 10^3 \frac{l}{m^3} \times 4.2 \times 10^{-16}\, m^3 = 4.2 \times 10^{-16}\, mol \text{ (II)},$$

which is more than 250 million molecules of αKG in the nucleus (III)

$$molecules_{\alpha KG(nucl.)} = n_{\alpha KG(nucl.)} \times NA \approx 4.2 \times 10^{-16}\, mol \times 6.022 \times 10^{23}\, mol^{-1} \approx 253 \times 10^6 \text{ (III)}.$$

The mouse genome has a size of roughly 2.5 billion bp[99] with almost 20 million of NCPs when one NCP consists of 150 bp and the average dG/dC content is 0.42[100]. If 5'-hydroxymethyl-2'-deoxycytidine (5hmdC) levels are very high, almost 20% of 5mdC, which is in total about 1% of dC, are hydroxymethylated[20], resulting in roughly 10 million 5hmdC. With 250 million molecules of αKG in the nucleus more than ten methyl-groups could be potentially removed from histone lysine residues per NCP in an αKG-dependent process catalyzed by JHDMs and 20% of 5mdC could be oxidized by TET enzymes. As these processes never happen simultaneously, this amount of αKG in the nucleus exceeds by far the amount that is needed by αKG-dependent epigenetic enzymes, even if their αKG affinity is low, in any situation. Moreover, if αKG is massively consumed in the nucleus there will be an αKG gradient in the cell and therefore flux of αKG from the mitochondria, where αKG is mainly produced, to the nucleus, where it is used, to maintain an overall αKG level of 1 mM in the cell. Consequently, αKG concentrations of about 1 mM cannot be rate-determining for any αKG-dependent enzyme. Nevertheless, manipulation of the αKG level, *e.g.* in stem cells, results in multiple consequences for the epigenetic landscape and cell fate.[24, 25] Consequently, one must assume that the cytosolic intracellular αKG level is substantially lower than 1 mM in cellular systems that are still under proliferation control and therefore are different from cancer or immortalized cell lines like HeLa or iBKM. Evidence for this assumption comes from a study from *Yang et al.* on brown adipose tissue (BAT) development, where they reported that αKG concentrations in brown stromal vascular cells and mouse embryonic fibroblasts are in the low micromolar range of about 15 µM and are doubled during BAT development.[89] Whereas a concentration of 1 mM of αKG in the nucleus means more than 250 million molecules thereof, 15 µM results in less than four million molecules. The Michaelis constant K_m, which describes the substrate concentration that is needed for half-maximal reaction

rate, was shown to be between 50 - 60 µM for Tet1 and Tet2 regarding αKG.[87] In BAT development, the elevation of the αKG level towards the K_m is directly linked to changes in the epigenetic landscape by increasing TET activity, which is needed for active DNA demethylation of promoter regions of relevant BAT-developmental genes.[89] Under physiological conditions, the majority of αKG seems to be located in the mitochondria, whereas in HeLa cells cytosolic αKG levels exceed the mitochondrial ones by almost ten times.[94, 101] These observations indicate that cancer cell lines or more general, cell lines that lack proliferation control, are not an appropriate model system to study the effects of metabolism on epigenetic changes, unless cancer-specific implications are desired to be drawn.

In addition to αKG and succinate, there are many more metabolites that influence the activity of αKG-dependent oxygenases with at least potential epigenetic relevance. Vitamin C, which is a reducing agent that is able to restore the catalytic iron^{2+} (Fe^{2+}) in TET enzymes, was discovered to enhance TET activity and thereby promote active DNA demethylation. Interestingly, other reducing agents do not show this effect, suggesting that vitamin C is a specific accessory factor for TET enzymes.[102, 103] Oxaloacetate (OAA), which has one hydrocarbon moiety less than αKG, could be in theory an additional co-substrate for αKG-dependent oxygenases that does not bind as efficiently as αKG, but is also accepted. However, since it is bound in the wrong orientation in the active site, it is a competitive inhibitor with an inhibition constant K_i around 50 µM, which is similar to the K_i of succinate.[86] Whereas αKG is bound with its carbon (C) 1 carboxylic/C-2 keto-group, OAA is bound with its C-4 carboxylic group, making it impossible for the enzyme to perform the oxidation.[86] Whether this inhibitory potential is physiologically relevant is under debate as the intracellular concentrations of mitochondrial OAA are usually in the low micromolar range and not even quantifiable in the cytosol.[85] Fumarate, isocitrate and L-malate are additional TCA cycle intermediates with less potent inhibitory potential regarding αKG-dependent oxygenases.[86, 87] Under normal physiological conditions, this is probably of no physiological relevance, since the cytosolic concentrations of these metabolites are not even quantifiable either.[85] Many types of tumor cells, however, show an accumulation of succinate and fumarate to millimolar concentrations due to loss-of-function mutations of the fumarate hydratase (FH) or the succinate dehydrogenase (SDH) and in these tumors, the function of the αKG-dependent oxygenases is impaired.[87] An additional metabolite that is a competitive inhibitor of αKG-dependent oxygenases is 2-hydroxyglutarate (2HG), but again is hardly present in the cell under physiological conditions (Figure 4). If low amounts of αKG are reduced to 2HG under physiological conditions, 2Hgdh immediately reconvert 2HG to αKG.[84, 104] However, 2HG can accumulate in the cell to high micromolar or even low millimolar concentrations under pathological or hypoxic conditions. *R*-2HG is generated by reduction of αKG by malfunctional Idh1 or 2 in cancer cells and is therefore often referred to as oncometabolite.[84, 105, 106] *S*-2HG is mainly generated by lactate dehydrogenase A (Ldh A), but also by Mdh1 and 2, in response to hypoxic conditions.[107] Both enantiomers show inhibitory potential.

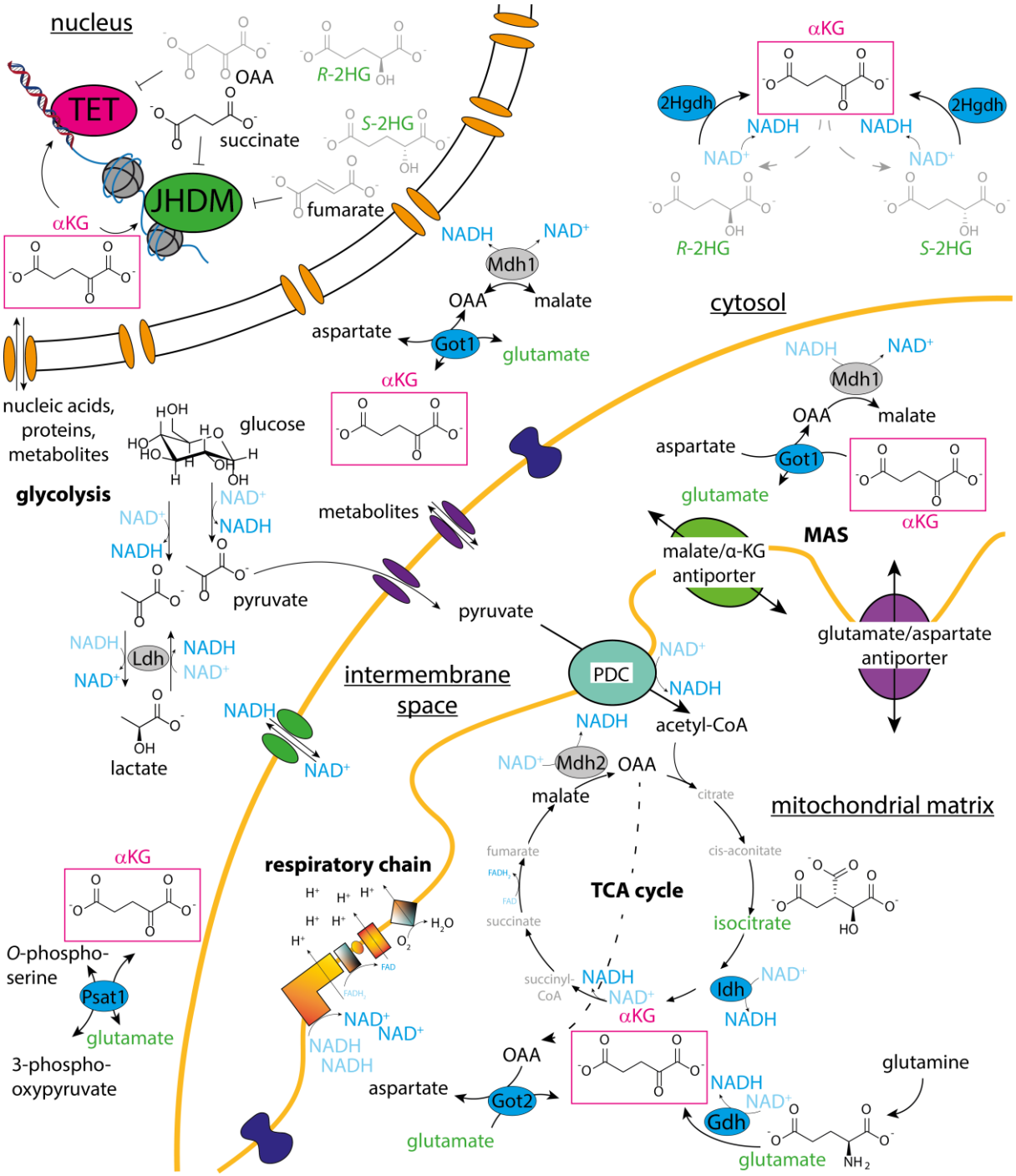


Figure 4: **Cellular mechanisms to adjust the cytosolic/nuclear and the mitochondrial αKG and NAD⁺/NADH concentrations.** Depiction of the MAS and of metabolites that inhibit αKG-dependent dioxygenases in the nucleus. Molecules in grey are only present at very little concentrations (< 1 µM) in the respective compartment of mammalian cells under physiological conditions. 2HG = 2-hydroxyglutarate, 2Hgdh = 2-hydroxyglutarate dehydrogenase, Acetyl-CoA = acetyl-coenzyme A, αKG = α-ketoglutarate, $FAD/FADH_2$ = flavin adenine dinucleotide, Gdh = glutamate dehydrogenase, Got = aspartate aminotransferase, Idh = isocitrate dehydrogenase, JHDM = Jumonji C-domain containing histone demethylase, Ldh = lactate dehydrogenase, MAS = malate-aspartate-shuttle, Mdh = malate dehydrogenase, NAD⁺/NADH = nicotinamide adenine dinucleotide (oxidized/reduced), PDC = pyruvate dehydrogenase complex, OAA = oxaloacetate, Psat1 = phosphoserine aminotransferase 1, TCA cycle = tricarboxylic acid cycle, TET = ten-eleven translocation enzyme.

Metabolism in mammals does not only rely on the availability of metabolites, but also on the availability of O_2. Whereas most differentiated cells strictly rely on sufficient O_2 supply, the situation is different for stem cells and cancer cells. It was shown that local O_2 availability directly regulates self-renewal and differentiation in stem cells.[68, 108, 109] In culture, stem cells are often kept under

atmospheric O_2 levels (21%), but this does not reflect the situation *in vivo*, where physiological levels of O_2 are much lower (5%).[109] The microenvironment, where stem cells are located in the embryo during development and in the adult organism, e.g. in the bone marrow, are even considered as hypoxic (1.5 – 5.3% O_2).[110-112] Stem cells that are grown under physiological or lower O_2 level show similar growth rates, but less spontaneous differentiation compared to stem cells that are grown under atmospheric O_2 level.[113] Lower O_2 levels result in the expression of hypoxia-inducible factors (HIFs) and Hif2α and Hif3α for example ensure upregulated expression of the octamer binding transcription factor 4 (Oct4), which is essential for early embryonal development and a hallmark of pluripotency.[109, 112, 114] However, it could be shown that a hypoxic environment does not only maintain pluripotency, but also facilitates proper differentiation of stem cells.[109] Due to the O_2-dependance of αKG-dependent oxygenases, O_2 availability also influences their activity and thereby epigenetic processes. *In vitro* studies on Tet1 and Tet2 activity revealed that O_2 is bound very efficiently and therefore the authors assumed that TET activity is not impaired under hypoxic conditions.[87] However, studies on human and mouse cancer cells showed that tumor hypoxia significantly reduces TET activity although the 2HG levels are not increased.[115]

The data indicate that there are tight connections between metabolic processes and epigenetic plasticity. Many metabolites, among them αKG, acetyl-CoA or SAM as the most prominent ones, are often co-substrates for chromatin modifying enzymes. The concept that these metabolites are provided "*in situ*" by metabolic enzymes directly in the nucleus is now established. It has been discovered that all enzymes involved in glycolysis, but also many proteins that were assumed to be exclusively located in the mitochondria are present in the nucleus.[116] These proteins are often referred to as "*moonlighting proteins*" since they fulfill a different purpose at their new location, which is not directly connected to their original function. Regarding metabolic enzymes in the nucleus, their new functions are providing co-substrates for chromatin-changing enzymes, balancing local NAD^+/NADH levels and interacting with DNA or DNA-binding proteins such as HIFs rather than direct energy supply.[116] For LDH, post-translational modifications trigger the nuclear import[117], where LDH interacts directly with DNA and several DNA-polymerases to modulate their activity[118, 119]. Furthermore, it was shown that LDH co-immunoprecipitates with sirtuin 1 (Sirt1) and that the activity of Ldh in the nucleus provides NAD^+ that is used by Sirt1 to increase the deacetylation activity.[120] Not only cytosolic enzymes, but also mitochondrial ones are moonlighting the nucleus. This is even more surprising since these enzymes have a mitochondrial import sequence. How these enzymes can end up in the nucleus is still under debate and mechanisms like co-import with other proteins upon stimuli like stress-response and mitochondria-derived vesicles are discussed.[116] The mitochondrial enzymes FH[121], SDH[122], Aconitase 2 (Aco2)[123], Idh2[124] and Mdh2[125] were reported to be present in the nucleus, with often unknown functions.[116] During zygotic genome activation (2-cell stadium in the murine embryo), the PDC and the TCA enzymes citrate synthase, Aco2 and Idh3a are transiently located to the nucleus to provide acetyl-CoA and αKG for epigenetic enzymes.[126] In more developed and fully differentiated cells, the best studied mitochondrial protein (complex) in the nucleus is the PDC.[127] Like the mitochondrial PDC, the nuclear PDC also generates acetyl-CoA that is in this case used by HATs.[128] *Sutendra et al*. could show that nuclear PDC activity modulates histone-acetylation levels and depends on the cell-cycle. Interestingly, the translocation seems to be dynamic and happens from the mitochondria to the nucleus and vice versa in response to signaling pathways. This is indicated by the fact that increasing nuclear PDC levels are accompanied by decreasing mitochondrial PDC levels at the same time.[128]

Taken together, metabolic changes ensure a rapid and adequate response of the cell to extra and intracellular signaling. The metabolic control of epigenetic plasticity preserves this response in the short-term and the long-term perspective and αKG plays a pivotal role in these processes. Epigenetic changes must be very precise regarding their location on the genome and one way to avoid

uncontrolled chromatin modifications is to provide the co-substrate, which is needed for the modification, only in the direct environment of the modifying enzyme at a certain location in the nucleus. The interaction of metabolic enzymes with the chromatin-modifying enzymes directly in the nucleus makes the required co-substrates available, whenever and wherever they are needed, and is therefore an important regulation mechanism for epigenetic changes.

The Mammalian Brain – Overview

Vertebrates that belong to the class *Mammalia* have in common a brain region called neocortex that is essential for higher-order brain functions and shows uniform characteristics regarding neuron number and proportion among all mammals.[129, 130] Although the neocortex is a common and special feature of mammals, mammalian brains are very diverse regarding anatomy and functional capacity, but on the molecular and cellular level many features are shared. There are two main types of neural cells – glia and neurons (Figure 5). The term "neural" thereby refers to both neuronal and non-neuronal cell types in the nervous system, whereas the term "neuronal" is exclusively used for neurons. The group of glial cells, which includes mainly astrocytes, the oligodendrocyte lineage cells and microglia (Figure 5), account for one up to two third of the total cell mass of mammalian brains.[2] Initially, glial cells were thought to be just structural and housekeeping support for neurons as suggested by the ancient Greek word "glía", which translates to glue. Nowadays, it is unquestionable that an intact network between neurons and glial cells is essential for proper brain functioning[57] and without glial support, vertebrate neurons are neither functional nor viable[131]. Astrocytes are the most common glia with widespread functions ranging from ion and water homeostasis to neurotransmitter recycling.[132] Oligodendrocytes lineage cells include mature oligodendrocytes and their progenitors NG2-glia. Mature oligodendrocytes produce myelin in the central nervous system (CNS) similar to the Schwann cells in the peripheral nervous system to allow rapid saltatory conduction and support of axons[133] and NG2-glia renew the pool of mature oligodendrocytes throughout lifespan and can even form synapses with neurons in the hippocampus and additional brain areas[134, 135]. Finally, microglia are in a simplified form the immunocompetent phagocytic cells that clean up in the CNS and are the only type of glia that do not develop from the ectoderm, but from progenitor cells of the yolk-sac, which, among other functions, supplies the early embryo with nutrients and is a source multiple types of progenitor cells[136].[2]

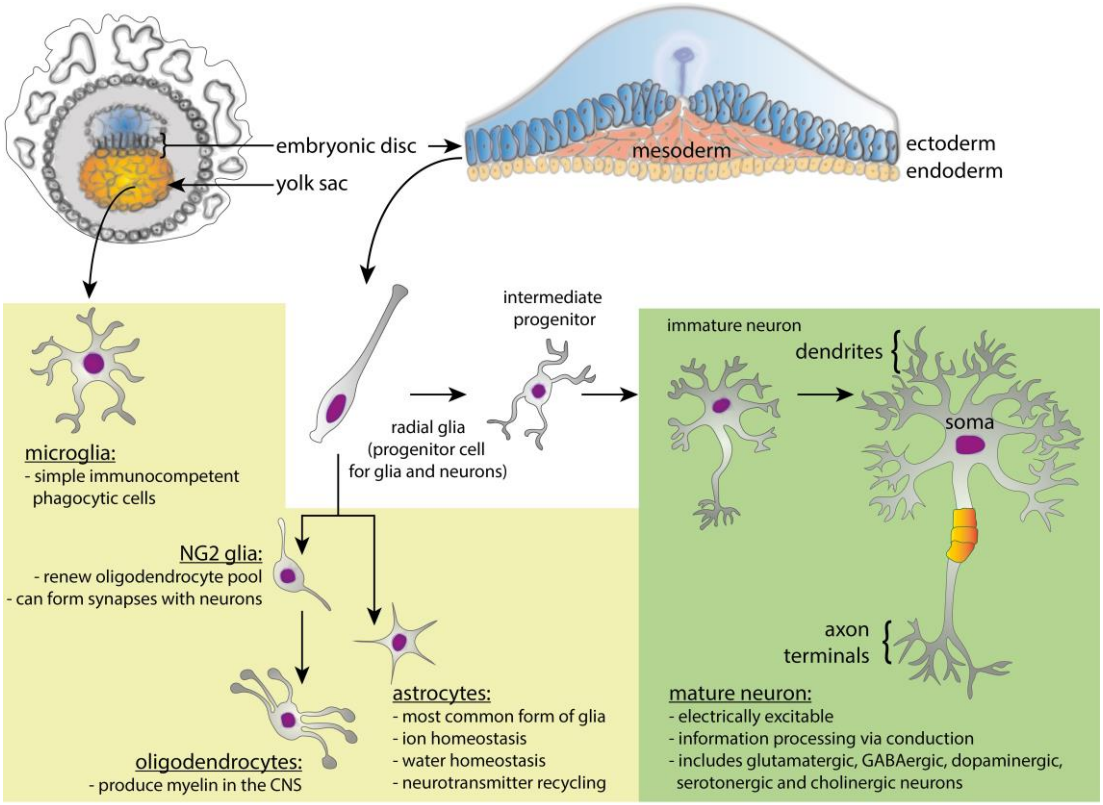


Figure 5: **Overview of differentiation of neural progenitor cells to neurons and glia, including their typical features and function.** Glial cells are depicted in the yellow box and neurons are depicted in the green box. CNS = central nervous system, GABA = γ-aminobutyric acid.

Neurons are a morphological and functional highly diverse group of electrically excitable cells that origin from neural progenitor cells (NPCs) and even though they have been extensively studied for more than 100 years, this diversity still prevents a universal definition of neurons.[1, 137-139] Protein markers that are applied to distinguish neurons from glia, such as the neuronal nuclear protein (NeuN) or the microtubule-associated protein 2 (Map2), are present in most types of neurons and are absent in glia, but there are exceptions.[140, 141] A "typical" neuronal cell that is depicted in most textbooks about neuroscience has a bipolar shape with a cell body (soma), including the nucleus, and two types of cellular extensions extruding from it – the dendrites and a single axon (Figure 5). Dendrites that receive signals at the postsynaptic side and an axon that transmit the signal on the presynaptic side are defined cellular compartments that are essential for unidirectional signal transmission.[142]

Neural activity, even in very simple organisms, involves communication between neurons via synapses. The synapse is an intercellular junction that allows intercellular communication.[143] Two main types of synapses exist – electrical and chemical synapses and both are widespread in the CNS.[144] Electrical synapses enable direct and therefore very fast metabolic and electrical communication between most neural cell types, in particular between neurons. The distance between the membranes in an electrical synapse is only 3.5 nm and the membranes are in most cases connected by gap junctions (Figure 6A).[145-147] Gap junctions consist of two hemichannels, also called connexons, and mediate intercellular contacts in many different cell types.[148] Neuronal connexons have a pore size of 16 – 20 Å and therefore allow the transmission of information via passive diffusion of molecules with a maximum size of 1 kDa.[147] Importantly, the composition of the hemichannels of each gap junction defines whether the flow of current and small molecules through this particular gap junction is bidirectional or unidirectional, whereby unidirectional flow requires specific different connexon compositions. By expressing different types of connexons, a neuron can modulate its transmission via electric synapses very precisely.[149] Furthermore, only a small number of gap junctions is usually open at the same time to avoid uncontrolled flux of ion and metabolites between the neurons.[145] Electric synapses are essential for electric coupling of neuron and neuron/glia clusters, so called neural circuits that process specific information. For example, they mediate the spiking activity of clustered neurons to ensure that incoming and outgoing signals are synchronized to maximize their effect and prevent unwanted extinction.[150]

Whereas electrical synapses can be both, symmetrical or asymmetrical, chemical synapses are always asymmetrical and have a defined pre- and postsynaptic membrane (Figure 6B). Chemical synapses are the dominant synapses in the CNS and include many different types regarding neurotransmitters that are released on the presynaptic (axonal) side, and their corresponding receptors on the postsynaptic (dendritic) side. This functional and molecular diversity is essential for the function of neural circuits, synaptic plasticity and transmission.[144] The synaptic cleft of a chemical synapse is between 12 - 20 nm distance. This distance is optimal regarding neurotransmitter concentration, which increases with narrower clefts, and local receptor currents, which decreases with narrower clefts due to higher electric resistance of the intracleft medium.[151] To remove the neurotransmitter from the synaptic cleft and therefore prevent overexcitation, the transmitter is either taken up by astrocytes or neurons and recycled or is directly degraded by enzymes in the synaptic cleft.[152] In the resting neuron, the majority of neurotransmitter molecules are stored in vesicles in the axon terminal and the exocytotic machinery prevents spontaneous transmitter release.[153] In this state, the intracellular concentration of calcium ions (Ca^{2+}) is extremely low, usually between 50 – 100 nM and a Ca^{2+} concentration of 10 µM is sufficient for neurotransmitter release.[154] In contrast, extracellular Ca^{2+} in the brain reaches normally levels of about 1 mM, therefore providing a big concentration gradient between the extra- and the intracellular milieu.[155] When the action potential reaches the axon terminal of a chemical synapse, leading to a depolarization of the neuron in this area, voltage-dependent Ca^{2+} channels are opened,

resulting in a rapid influx of Ca^{2+} to increase the intracellular Ca^{2+} concentration. The interaction of the exocytotic machinery with Ca^{2+} evokes fusion of the neurotransmitter vesicles with the plasma membrane, eventually leading to a massive release of neurotransmitters into the synaptic cleft (Figure 6B).[156] On the postsynaptic side, neurotransmitter-specific receptors are responsible for propagating the received signal. Those receptors can be either ionotropic, which means that they are ion channels, which are activated upon transmitter binding, or they can be metabotropic, which means that they usually activate G-proteins that subsequently transmit the signal directly or indirectly to an ionotropic receptor. The various combinatorial possibilities of neurotransmitters and corresponding receptors allow the very specific transmission of information. The type of neurotransmitter and the type of the receptor define whether a signal is excitatory, e.g. for glutamatergic synapses, or inhibitory, e.g. for γ-aminobutyric acid (GABA)ergic synapses, whether the signal propagation is fast or slow. In general, Ionotropic receptors are fast reacting receptors, whereas signal transduction of metabotropic receptors can even take several hours.[157]

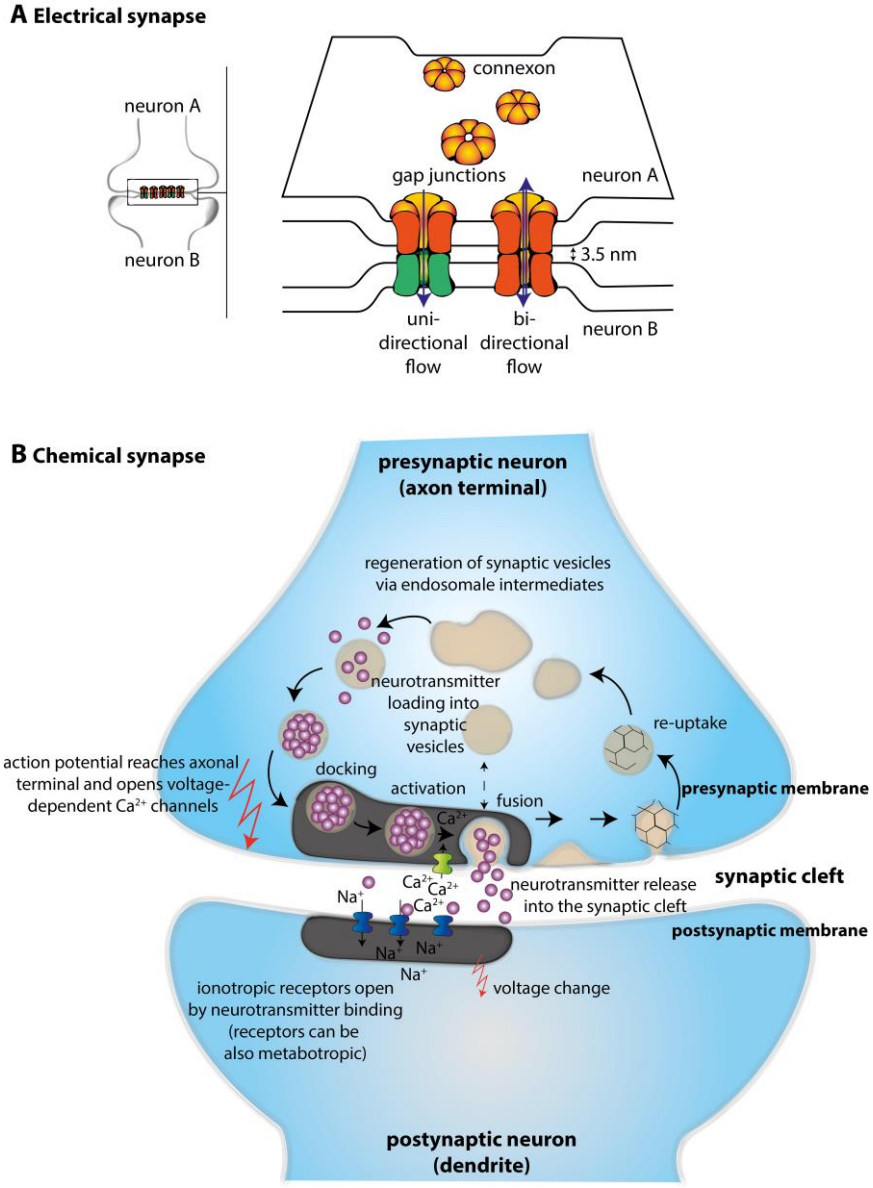


Figure 6: **Electrical and chemical synapses.** (A) Schematic depiction of an electrical synapse and the gap junction that connect the pre- and the postsynaptic membrane. (B) Schematic depiction of a chemical synapse and its mode of function.

A major characteristic of both types of synapses is their ability to adapt to the signals that they receive and transmit, which is called synaptic plasticity.[150, 158, 159] Synaptic plasticity is the cellular basis for memory and learning. It affects for example spiking activity[159, 160] and epigenetic plasticity is pivotal for synaptic plasticity to conserve the changes on the gene expression level. Exposure of synapses to certain stimuli or cellular changes often results in long-term changes of synaptic strength. Long-term plasticity is characterized by long-term potentiation (LTP) and long-term depression (LTD) and appears in the hippocampus among other brain areas.[161] LTP is an activity-dependent increase in the transmission efficacy of synapses after a persistent sequence of action potentials that lasts hours, whereas LTD is the activity-dependent reduction in the transmission efficacy. LTP and LTD are hallmarks of synaptic plasticity and crucial for learning and memory formation.[161]

Most data on brain function are obtained from rodents, since they are easier to access from an experimental and ethical point of view than primates, especially than humans. Recent studies suggest, that the major differences of human brain to other mammals is not based in neurons, but rather in glia. Human astrocytes are larger and more complex than their rodent equivalent and can propagate Ca^{2+} waves four times faster.[162] However, also human neurons have unique gene expression profiles even compared to chimpanzee and differentially expressed genes include genes that are relevant for neuronal differentiation, synaptic transmission and metabolic processes – in summary genes that are relevant for every aspect of neuronal development and functioning.[163, 164] Furthermore, adult neurogenesis in the subgranular zone of the dentate gyrus (DG, Figure 7) of the adult mammalian hippocampus, which is associated with learning and memory, response to stress and exercise (vide infra), is well established in rodents and even monkeys[165-169] and was also reported in humans[170, 171]. However, recently the event of neurogenesis in the adult human hippocampus was questioned by a study from *Sorrells et al.*[172]. Altogether, the data indicate that the human brain is functionally very different from other mammals. The focus of this thesis is on epigenetics, where many observations regarding DNA modifications in rodents can be indeed also observed in humans. For example, the 5mdC and 5hmdC patterns in mouse and human brain correlate[10] and the age-dependent increase of 5hmdC levels can be also observed in both[173-175]. However, implications from mouse data regarding epigenetic processes in the brain, which are described in this thesis, cannot be necessarily transferred to humans.

The Hippocampal Formation

The hippocampal formation (HF) is a neuron-dense brain area that is important for the encoding, consolidation and retrieval of long-term episodic memory and spatial navigation among other cognitive functions.[176-178] It describes the bilateral neuronal system that is an integral part of the cerebral cortex of the mammalian brain[179], but the definitions of the subregions that are comprised by the HF vary. By one definition, which is applied in this thesis, the HF includes three distinct subregions – the DG, the hippocampus proper, which consists of the *Cornu Ammonis* areas CA3, CA2 and CA1,[180] and the subiculum as a link between the hippocampus proper and the entorhinal cortex (EC) (Figure 7A).[176, 181] Other definitions of the HF exclude the subiculum or additionally include the EC, which is part of the parahippocampal region by the first definition.[182] In the narrower sense, the term "hippocampus" refers to the hippocampus proper. However, it is often used to describe the entire HF.

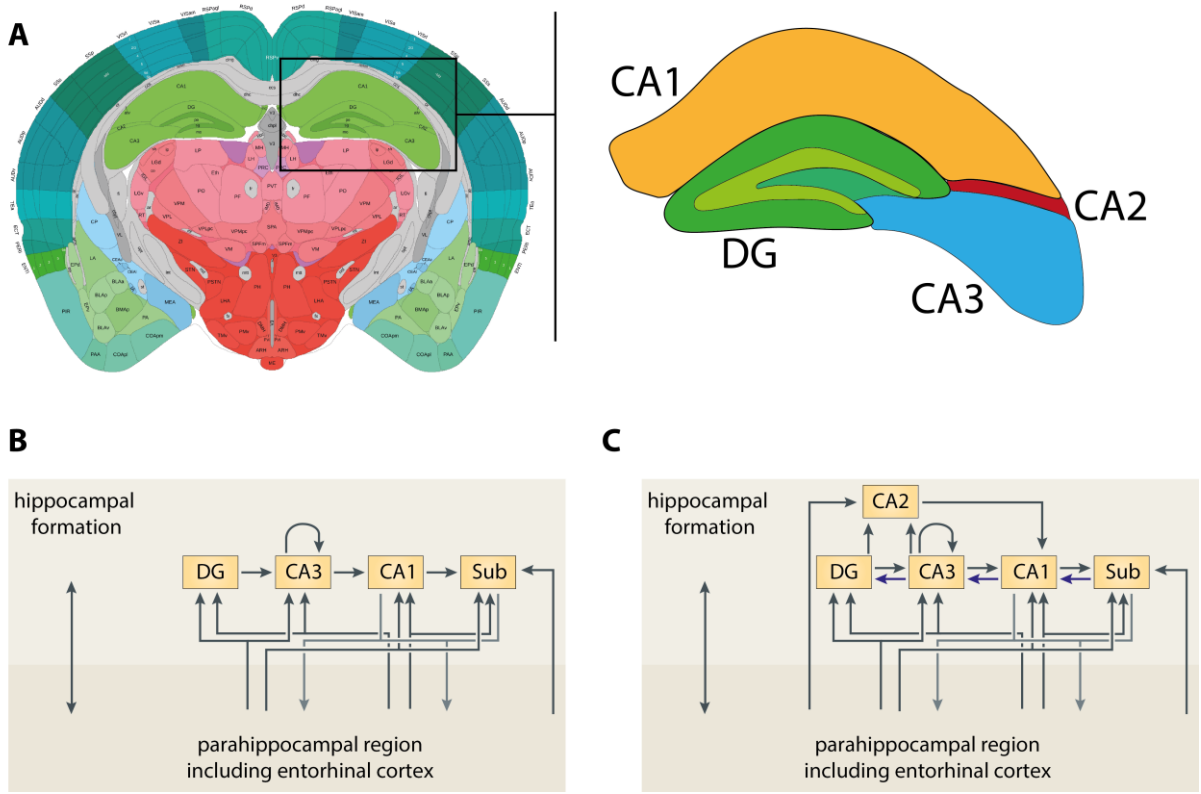


Figure 7*:* **The hippocampal formation in mouse.** (A) Coronal (section 75 of 132) of the mouse brain (P56) showing the hippocampus, which consists of the *Cornu Ammonis* areas CA1, CA2 and C3 and the DG (Image credit: Allen Institute). (B) Standard model of the hippocampal circuitry. Figure adapted from *van Strien et al*[181]. (C) Extended model of the hippocampal circuitry, including CA2 and several recently reported back projections (blue). DG = dentate gyrus, P56 = postnatal day 56, Sub = subiculum.

The HF consists of three different layers.[181] In the DG, the first layer, which is also sometimes considered as the CA4 region, is referred to as *hilus*, whereas it is called *stratum oriens* in the hippocampus proper. This layer is polymorph and consists of afferent (arriving in the layer) and efferent (exiting the layer) nerve fibers and interneurons.[181, 183] Interneurons, in contrast to receptors or motor neurons, are neurons that receive signals from and send signals to other neurons and show great diversity regarding morphology, connectivity and physiological properties.[184, 185] The next layer comprises interneurons for local circuits and principal cells, which are neurons with long axons that can transmit information over long distances from one brain region to another. This layer is referred to as *granule layer* in the DG and as *stratum pyramidale* in the hippocampus proper and the subiculum.[181] The most superficial layer, which is further divided in sublayers in the CA regions, is the

stratum moleculare and contains neurons and the apical dendrites of the neurons from the *stratum pyramidale*.[181] DG, CA3, CA1 and subiculum form the *polysynaptic HF pathway* (Figure 7B).[181, 186-189] In this standard model of hippocampal circuitry, *mossy fibers* of the DG granule cells, which are characterized by high concentrations of zinc, project (form synapses) unidirectionally to the pyramidal neurons of the CA3 region[189, 190] and the axons of the pyramidal neurons of the CA3 region, called *Schaffer collaterals*, project to the CA1 region[186], which in turn projects to the subiculum[187]. Furthermore, the CA3 region has recurrent collaterals that form synapses within the CA3 region.[186, 188] In addition to the standard model, several back projections have been reported: from CA3 to DG[191, 192], from CA1 to CA3[187, 193] and also from the subiculum to CA1[188, 194]. The extension of the polysynaptic HF pathway is the *performant pathway* that describes the EC-HF circuitry. Axons from nerve cells of the EC project to all fields of the hippocampus, whereas mainly the CA1 region and the subiculum project back to the EC.[181] Although the CA2 region was already characterized in 1934 by *Lorento de Nó* as a distinct region that does not receive input from *mossy fibers* of the DG, it was still under debate for many decades whether CA2 is a subregion of CA3 or not[195-197] and it is often not included in the standard presentation of hippocampal circuitry[176]. However, in recent years it was shown that CA2 is a distinct region with a gene expression pattern that is clearly different from CA3 and CA1[195, 198, 199] and that the molecular defined CA2 region is integral part of the hippocampal circuitry, which receives intrahippocampal input from the DG, the *Schaffer collaterals* of the CA3 region and from the EC and mainly projects to the CA1 region (Figure 7C).[195, 200]

As part of this thesis, the HF, especially the DG, was chosen as a model brain region for studying the activity of Tet3 and the distribution of 5hmdC for several reasons. First, of all brain regions, the adult neurons of the DG show the highest 5hmdC content.[20] Second, it is already known that Tet3 is not only expressed in the HF, but plays a key role there for epigenetic plasticity.[16] Furthermore, the HF plays a pivotal role in memory formation and therefore, a better understanding of the epigenetic processes that underlie hippocampal functioning are of great interest.

Metabolism and Neural Function

The brain is a metabolic very active tissue that not only needs a high amount of energy, but also uses metabolites as neurotransmitters.[55, 201] Glucose is the main source for energy supply, but glutamate, acetate, lactate, pyruvate or glutamine are also sources that are used for oxidative metabolism.[202] Despite its high energy demand, brain tissue does not have big glycogen storages such as the liver and therefore relies on constant nutrient supply. Although the amount of glycogen in brain is too low to bridge states of hypoglycemia more than a few minutes, proper glycogen metabolism is important for brain function, including synaptic activity, learning and memory storage.[203, 204] The small glycogen storages in brain tissue are mainly found in astrocytes and can be mobilized quickly when there is a local demand in glucose increase.

In principle, both neurons and astrocytes are capable of using glucose and lactate to fulfill their energy demands.[202] During neuronal differentiation, aerobic glycolysis is important for proper differentiation, but in addition mitochondrial biogenesis peaks up and it was shown that the transition from NPCs to immature neurons requires downregulation of key enzymes of aerobic glycolysis, indicating the important role of oxidative metabolism for mature neurons.[205, 206] In recent years, there were controversial discussions whether in the developed brain, neurons or astrocytes are predominant in taking up glucose upon neuronal stimulation.[57, 207] The supporters of the astrocyte-neuron lactate shuttle (ANLS) hypothesis claim that especially during neuronal activation, astrocytes are responsible for glucose uptake and lactate release, which in turn is then used by neurons for OxPhos.[57, 208-210] In contrast, the latest data favor the opposite direction of glucose flux (neuron-astrocyte lactate shuttle (NALS) hypothesis) and show that neurons increase their glycolytic activity upon stimulation and export lactate, which is then taken up by astrocytes.[58, 207, 211, 212] The direction of glucose flux is important for estimating the alteration of cytosolic NAD^+/NADH ratio during neuronal stimulation. In principle, there are three main mechanisms to control the NAD^+/NADH levels (Figure 4).[26] Glycolysis lowers the cytosolic NAD^+/NADH levels (mechanism 1), but reduction of the glycolysis end product pyruvate to lactate by LDH can restore the cytosolic NAD^+ pool to some extent (mechanism 2). In addition to glycolysis and the activity of the LDH, the MAS is essential for NAD^+/NADH homeostasis (mechanism 3). The reduction of pyruvate to lactate by the LDH is not fast enough to compensate the NAD^+ consumption by glycolysis and the MAS can only restore the cytosolic NAD^+ pool when OxPhos provides sufficient mitochondrial $NAD(P)^+$. Although OxPhos is the main energy pathway for activated neurons, it starts delayed after neuronal stimulation.[58, 213] Therefore, when neuronal glycolysis is increased upon neuronal stimulation, the previously high cytosolic NAD^+/NADH level of the resting neurons decrease until OxPhos starts.[58, 214, 215]

Metabolites in brain are not only parts of anabolic and catabolic processes but are often also neuroactive compounds. Glutamate for example is a very abundant amino acid with millimolar concentrations in mammalian cells[85], which can be used either for protein biosynthesis or oxidative metabolism, but in brain it is also a neurotransmitter with 80% of the neurons in the cortex being glutamatergic neurons.[57, 216] There are two aspects that have to be considered when it comes to glutamate for transporting neuronal information. Glutamate is the major excitatory neurotransmitter, which is essential for learning, memory and cognition by forwarding nerve impulses in the short term and modulating synapse plasticity in the long term.[216, 217] However, the sensitivity of neurons towards glutamate makes it also highly neurotoxic since it induces hyperexcitability when its concentrations in the extracellular space are not tightly controlled.[218-220] When the electrical action potential reaches the pre-synaptic terminal of a excitatory glutamatergic synapse, it is translated into a chemical signal to bridge the synaptic cleft. The action potential opens the Ca^{2+} channels on the cell membrane of the pre-synaptic neuron, which in turn releases glutamate of the intracellular vesicles into the synaptic cleft via exocytosis. The vesicles are also recycled via endocytosis and refilled with glutamate in the

pre-synaptic end using a proton/glutamate antiporter system.[221, 222] On the other side of the cleft, at the postsynaptic membrane, glutamate receptors are located and binding of glutamate to these receptors induce a conformational change. This allows, depending on the receptor type, either the influx of ions to the neuron to start the action potential (ionotropic receptors) or to the activation of membrane-coupled G-proteins (metabotropic receptors) to start a signaling cascade that alters the epigenome, which influences gene expression and as a consequence synaptic plasticity.[216, 223, 224] Typical ionotropic glutamate receptors are N-methyl D-aspartate receptors (NMDAR), α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic receptors (AMPAR) or kainate receptors. Interestingly, NMDAR expression is not limited to neurons, but was also discovered in astrocytes.[216, 225] Whereas there are controversial discussions about the role of astrocytes for neuronal energy metabolism, it is unquestionable that they play a crucial role in neurotransmitter recycling. Once glutamate molecules are released into the synaptic cleft via exocytosis and have bound to the glutamatergic receptors on the post-synaptic side, they must be removed to avoid hyperexcitation. However, in contrast to acetylcholine and other neurotransmitters, there is no enzyme present in the synaptic cleft to degrade glutamate to limit its excitatory potential. Instead, glutamate is removed and recycled by uptake.[216, 226] Excitatory amino acid transporters (EAATs), which allow the uptake of glutamate, are located on the outer membrane of the pre-synaptic and the post-synaptic neuron and on the outer membrane of the surrounding glial cells, mainly astrocytes.[226] Most of the glutamate is taken up by the astrocytes, where it can be either converted to glutamine and be re-transported to the neurons or used for other metabolic processes.[227] As glutamine is not a neuroactive compound, it does not cause excitation and back in neurons it can be reconverted to glutamate.

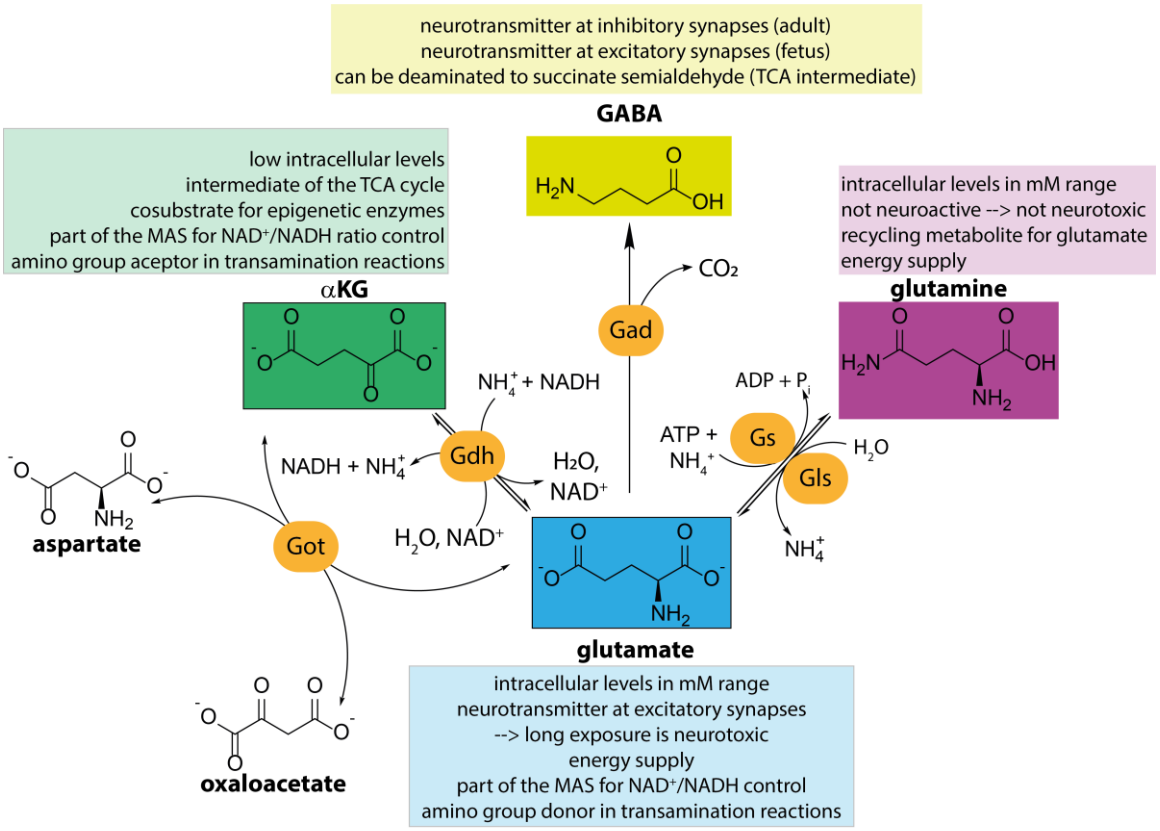


Figure 8: **Neural glutamate metabolism.** Glutamate is a key metabolite in neurons and glia that functions as a neurotransmitter but is also a precursor for αKG and GABA.
ADP = adenosine diphosphate, αKG = α-ketoglutarate, ATP = adenosine triphosphate, GABA = γ-butyric acid, Gad = glutamate decarboxylase, Gdh = glutamate dehydrogenase, Gls = glutaminase, Got = aspartate transaminase, Gs = glutamine synthetase, MAS = malate-aspartate shuttle, NAD+/NADH = nicotinamide adenine dinucleotide (oxidized/reduced), P_i = phosphate, TCA cycle = tricarboxylic acid cycle.

Regarding the interplay of metabolism, epigenetics and neural function, glutamate metabolism is a pivotal point (Figure 8).[228] As described above, glutamate is a major neurotransmitter at excitatory synapses, but it is also the precursor for GABA, which is in contrast a major neurotransmitter at inhibitory synapses (Figure 8) in the adult mammalian brain.[229] Glutamate decarboxylases use pyridoxal phosphate (PLP) as a co-factor to perform the decarboxylation reaction on the α-carboxyl-group, yielding GABA. Deamination of glutamate instead of decarboxylation results in αKG. This reaction can be performed by multiple enzymes, such as Got1 and 2, which perform a transamination with PLP as co-factor, or Gdh. Gdh ablation in murine brain does not show a specific phenotype when glucose is not limited, whereas its overexpression or Gdh gain-of-function mutations change synaptic plasticity and induces neuronal loss in specific brain regions, e.g. in the CA1 region of the hippocampus.[230-232] A reason for that might be that in affected brain regions, which suffer from neuronal loss and altered synaptic plasticity upon Gdh overexpression, mitochondrial αKG that is produced due to higher Gdh activity is shuttled to the cytosol of the nerve terminals, where it is reconverted to glutamate and immediately used as a neurotransmitter. This leads to an increased glutamate release at the pre-synaptic end upon neuronal stimulation.[230] In contrast, glutamate levels are unaffected in the brain-specific Gdh-deficient mice.[231] This indicates that other enzymes such as Got1 and 2 can compensate the loss of Gdh but are not able to handle increased Gdh activity. Interestingly, although Gdh-ablation seems to be easily compensated, transient inhibition of Gdh aggravates glutamate uptake by astrocytes.[233] Although there is strong link via αKG between epigenetics and glutamate metabolism, the effects of changes in glutamate metabolism on epigenetic modifications in neurons are only poorly understood.

Epigenetic Processes in the Developing and the Adult Brain

Neural activity and epigenetic processes are tightly coupled and influence each other.[3, 4, 6, 9, 10] Especially in the developing brain, when the epigenome changes dynamically on a global level, epigenetic processes ensure correct neural differentiation and brain maturation.[173, 234, 235]

Regarding DNA methylation, Dnmt1 and Dnmt3a are the dominant DNMTs in the postnatal mammalian brain and their high expression in post-mitotic neurons points towards a functional role of DNMTs for brain functioning that goes way beyond preservation of 5mdC patterns during cell division.[173, 236-238] Dnmt3b, which plays an important role in the first stage of embryonal neurogenesis[239], is also detectable in the early postnatal brain, but is significantly lower expressed.[173] During embryonal development, the conditional knockout (cKO) of *Dnmt1* has severe consequences for brain maturation and results in degenerated cortex and hippocampal areas and abnormal morphology and electrophysical properties of neurons.[240, 241] A knockout (KO) of *Dnmt3a* in postnatal neural stem cells impairs neurogenesis, although the global cytosine methylation pattern remains unchanged and the expression levels of Dnmt1 and Dnmt3b are not influenced either.[242] In the adult brain, Dnmt1 and Dnmt3a have overlapping function and are required for synaptic plasticity, learning and memory.[243] Mice with a *Dnmt1*/*Dnmt3a* double knockout (DKO) in post-mitotic forebrain neurons have altered long-term plasticity in the hippocampal *Schaffer collateral*/CA1 synapses, resulting in learning and memory deficiencies.[243] There is no neural loss in the hippocampus of the *Dnmt1*/*Dnmt3a* DKO mice. However, the *Dnmt1*/*Dnmt3a* DKO hippocampal neurons are smaller, show decreased DNA methylation and a down-regulation of genes being associated with synaptic plasticity; whereas the *Dnmt1* and *Dnmt3a* single KOs in the post-mitotic neurons do not show these effects.[243] In contrast to other cell types, including glia cells, where 5mdC is almost exclusively found in CpG regions, the DNA of mature neurons shows 25% of dC methylation in a non-CpG context.[173, 244, 245] The biggest changes in dC methylation patterns occur within genes and their promoter regions that are important for neuronal development and happen in mice between week two and four postnatal, especially in the non-CpG context.[173] This global change correlates with an increase of Dnmt3a expression and interestingly, also with synaptogenesis.[173, 242, 246] In the developing brain, not only cytosine methylation patterns, but also cytosine hydroxymethylation patterns are dynamically changed.[173] The global levels of 5hmdC in the brain are increasing during neural differentiation and during synaptogenesis, resulting in the highest 5hmdC levels in the adult brain.[174, 175, 234 247, 248] Neuronal cells have higher 5hmdC levels compared to non-neuronal cells, *e.g.* glia, and 5hmdC was especially found to be enriched in synapse-related genes.[11 19, 173] In the mouse embryo, these cell-type specific differences in the 5hmdC level of NPCs manifest between embryonic day 12.5 (E12.5) and E13.5[249], which is an important developmental stage for early neural differentiation with early neuronal differentiation taking place between E9.5 and E12.5 and early glial differentiation taking place from E13.5 on.[250, 251] Tet2 and Tet3 seem to be important for establishing 5hmdC patterns during neurogenesis, since knockdown of both impairs the progression of differentiation.[247] Although 5hmdC levels are low in the fetal brain compared to the adult one, the adult 5hmdC pattern is already established in the developing brain, where 5hmdC marks particularly regulatory CpG sites of active genes that are demethylated in the adult brain[19, 173] and in the gene body of developmentally activated genes[234, 252]. Regarding dC modifications in the adult murine brain, about 4.5% of all dC are methylated and up to 0.6% are hydroxymethylated (Figure 9). In the adult human brain, some areas even show more than 1% hydroxymethylation of dC.[20, 253]

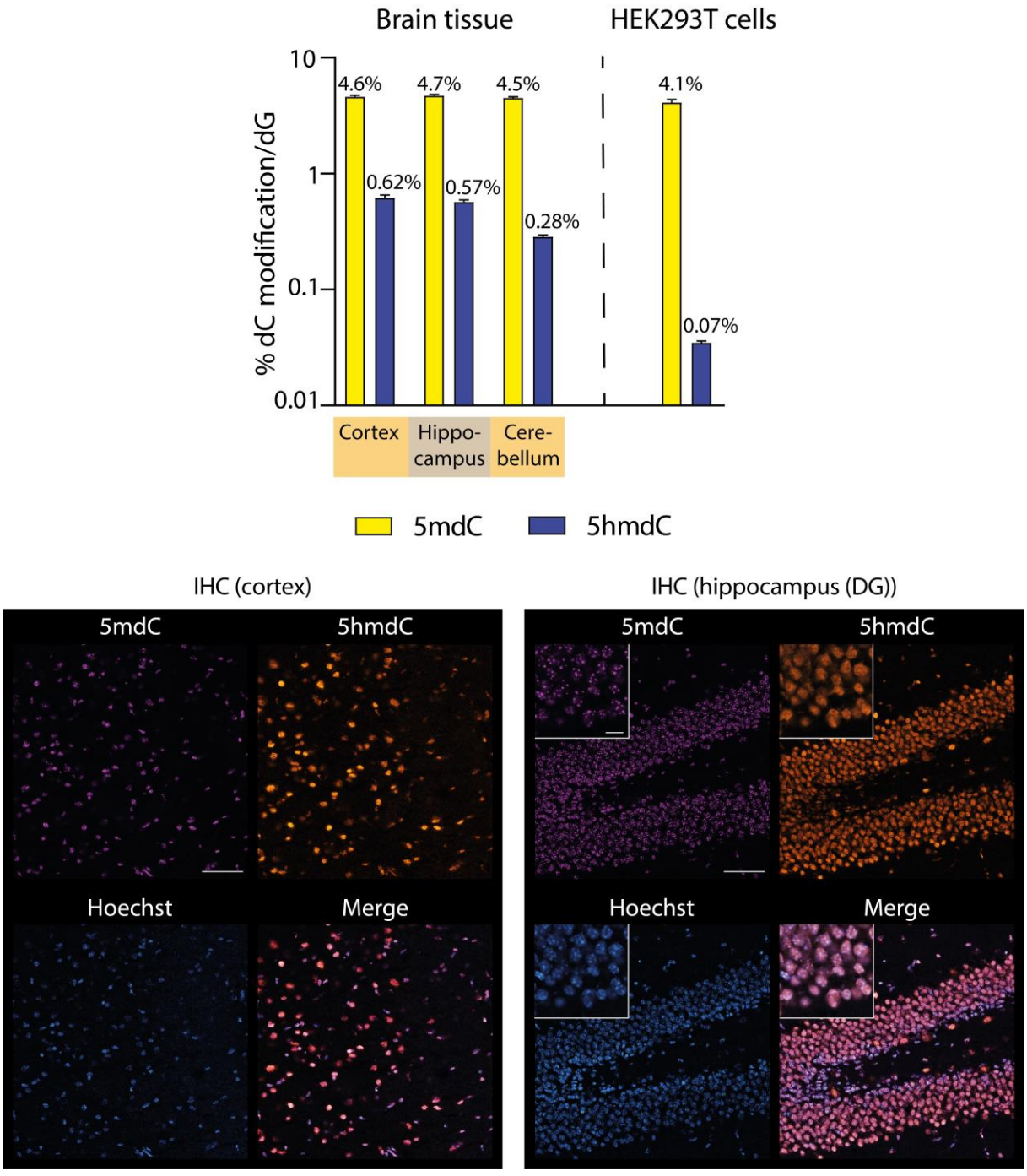


Figure 9: **dC modifications levels in the brain.** Comparison of 5mdC and 5hmdC levels in different brain regions compared to HEK293T cells. IHC data of 5mdC and 5hmdC in the cortex and the DG. Images show overlaid *Z*-stacks, scale bar zoom out is 50 µm; scale bar zoom in is 10 µm.
5hmdC = 5-hydroxymethyl-2'-deoxycytidine, 5mdC = 5-methyl-2'-deoxycytidine, dC = 2'-deoxycytidine, dG = 2'-deoxyguanosine, DG = dentate gyrus, IHC = immunohistochemistry.

More than 99.9% of 5hmdC is found in a CpG context, resulting in 13% cytosine hydroxymethylated CpGs in the adult human brain.[173, 175] Interestingly, the occurrence of 5hmdC in a CpG context has different consequences for chromatin organization and gene expression when compared to hydroxymethylation in a CpA context. Whereas 5hmdC in CpG correlates positively with active transcription, this was not observed for 5hmdC in a CpA context.[12] The majority of 5hmdC in brain is enriched in gene bodies and distal, which means upstream of the gene, regulatory regions, but depleted from transcription starting sites (TSS).[173, 175, 234] Whereas 5mdC is usually correlated with repressive histone marks like H3K9 methylation[254], 5hmdC in neurons shows negative correlation with

the repressive histone marks H3K27me3[255] and H3K9me3[256]. In contrast, 5hmdC shows positive correlation with H3K4me1, which marks poised enhancers or active enhancers when in combination with H3K27ac[257],[175] and is associated with active transcription when it occurs within gene bodies of neurons.[12, 252, 258] About 70% of 5hmdC in neurons is intragenic with an enrichment in exons compared to introns.[234] Although splicing processes take place on the mRNA level, 5hmdC seems to play a role in alternative splicing in neurons, as 5hmdC levels within the 20 bp exon/intron boundary positively correlate with constitutively spliced exons compared to alternatively spliced exons.[11, 175, 259] This correlation was not observed in hepatocytes of the liver.[11]

Cytosine modifications were not the only DNA modifications that were discovered to play a role for modulation of and response to neural activity. In 2017, *Yao et al.* observed accumulation of N6-methyl-2'-deoxyadenosine (6mdA) in the prefrontal cortex upon environmental stress.[260] Recently, 6mdA, most likely introduced in neurons by N6-adenine-specific DNA methyltransferase 1, was reported to positively regulate neural activity-induced gene expression. Depolarization of primary cortical neurons by potassium chloride (KCl) leads to a more than three-fold increase of global 6mdA and 6mdA was shown to accumulate upon extinction learning. [260, 261] However, attempts of our group to detect 6mdA as an epigenetic mark in gDNA from whole mouse brain failed.[262] This indicates that 6mdA might be only abundant and physiologically relevant as an epigenetic mark in distinct neurons or glia.

Despite this latest discovery of an adenine DNA nucleobase modification, cytosine modifications are the most important regarding neural activity and development in mammals. Methylation and hydroxymethylation of dC play a pivotal role in neurogenesis[263, 264], brain function and neuronal plasticity[3, 16]. The synchronous activation of neurons in the DG results in a changing methylation landscape in CpGs with 1.4% of the studied 200000 CpGs to be affected.[264] Especially genes that are linked to neuronal plasticity show both, demethylation of 5mdC and *de novo* methylation of dC.[264] For example, depolarization of post-mitotic neurons leads to increased transcription of the brain-derived neurotrophic factor (*Bdnf*) caused by demethylation of the *Bdnf* promoter.[265] Bdnf is expressed in many brain areas including the hippocampus and is important for synaptic function and plasticity, neuronal survival[266] and long-term memory formation in the hippocampus[267]. In the adult brain, all three TET enzymes are expressed with increasing expression levels from Tet1 to Tet3.[21] However, compared to Tet1 only few cell biology data about the role of Tet3 in neurons exist, among them a first hint that Tet3 might help to prevent neurodegenerative diseases.[10, 268] Most of the studies regarding TET functions in the brain focus on Tet1, especially on its role in the hippocampus. *Tet1* KO mice are viable with overall normal health, locomotion and brain morphology.[269] In those mice, the neuron number and synaptic connectivity is not different from the corresponding control and also hippocampus-dependent learning and memory formation are not affected.[270] Regarding effects on spatial memory, confounding results have been published.[270, 271] Tet1 ablation leads to a statistically significant reduction of neuronal 5hmdC and down-regulation of several neuronal activity-regulated genes, such as neuronal PAS domain protein 4 (*Npas4*), by hypermethylation of their promoter region.[270] Npas4 is a highly expressed transcription factor in neurons, which regulates the formation and maintenance of inhibitory synapses.[272] Moreover, the hippocampal LTD is increased and memory extinction is impaired in Tet1-deficient mice, while the basal synaptic transmission and the LTP remain unchanged.[270] Interestingly, while long-term memory does not seem to be affected in *Tet1* KO mice, these mice show significant deficiencies in short-term memory.[271] *Kaas et al*., found out that the transcription of the *Tet1* gene itself is neuronal-activity dependent in the hippocampus, where Tet1 is mostly expressed in neurons and only little in astrocytes.[273] They showed in their study that in response to four hours of depolarization with 25 mM KCl, the levels of Tet1 are down-regulated in primary cultures of hippocampal neurons and that 5mdC and 5hmdC levels change in CA1 tissue after having induced seizures with variable hours for recovery.[273] Furthermore, activity-induced region-specific DNA

demethylation in the adult brain by Tet1 and the cytidine deaminase Apobec1 via an oxidation/deamination mechanism was reported.[274] This leads to enhanced gene expression of *Bdnf* among other genes in the DG and can be prevented by ablation of the growth arrest and DNA-damage-inducible, beta (Gadd45b) protein.[263, 274] Accordingly, knockdown of Tet1 expression in the hippocampus results in hypermethylation of the *Bdnf* promoter IV, which consequently reduces expression of *Bdnf* from this promoter.[16] Interestingly, the effects of Tet1 in the hippocampus seem to be not only dependent on the protein's catalytic activity. When Tet1 or a catalytic inactive mutant thereof is overexpressed after Adeno-associated virus-mediated induction in mice, several memory-associated and synaptic plasticity-associated genes in the hippocampus are upregulated and long-term memory formation is impaired.[273] The in part contradictory results about the importance of Tet1 for neural functioning and memory formation might be explained by the role of Tet1 in adult neurogenesis. Tet1, which is highly expressed in NPCs, is important for the self-renewal and proliferation capability of the NPCs in four-months-old mice. Tet1 ablation therefore results in impaired neurogenesis due to a depleted NPCs pool in the subgranular zone of the DG. The differentiation capability of the cells in contrast is not affected.[271] Additionally, Tet1 and Tet2 expression correlates with higher 5hmdC content in the promoter region of micro RNA (miRNA) 137 and thus higher expression level. In hippocampal neurons, miRNA-137 is also involved in adult neurogenesis[275] and adult neurogenesis in mice is linked to better memory formation, but as mentioned before, it depends on many other parameters, including exercise and stress.[168, 169] Moreover, the activity and expression of Tet1 is also influenced by environmental parameters, such as physical exercise[275], suggesting that to some extent the effects of Tet1-deficiency and the extent of compensation by other TET enzymes depend on the keeping conditions of the mice.

Tet2 activity is important for hydroxymethylation of CpGs of developmentally differentially expressed genes, but its role in brain functioning is not well studied yet.[173, 259] DKO of *Tet1* and *Tet2* leads to increased perinatal deaths of mouse embryos, but the surviving animals are viable and fertile and do not show severe defects. Although, the 5hmdC levels are decreased during development and in the adult tissues, including brain, a substantial amount of 5hmdC is still detectable.[276]

In contrast, global Tet3-deficient mice are not viable[277] and neither murine Tet3 nor human TET3 is tolerant to haploinsufficiency[15]. These findings in addition to the relatively high expression levels of Tet3 in differentiated cells indicate that Tet3 can compensate the function of Tet1 and Tet2 in case of deficiency and is essential for cellular development and function. In primary cortical neurons, expression of Tet3 is regulated by neuronal activity, whereas this is not the case for Tet1.[278] Upon extinction learning, which means the gradual decrease in response to a conditioned stimulus[279], a Tet3-dependent genome-wide redistribution of 5hmdC in the neocortex is observed that goes along with transient reduction of the repressive histone mark H3K9me3[256] and a delayed increase of the transcriptional activating histone mark H3K27ac[257,278]. In hippocampal neurons, Tet3 is important for synaptic transmission and the maintenance of the homeostatic synaptic plasticity.[16] Synaptic transmission, which is the communication of two neurons via the synaptic cleft, and synaptic plasticity, which is the adaption capacity of the synapse to different stimuli, describe the fundamental properties of a synapse.[280] When synapses change in number and strength, individual neurons can adapt (synaptic scaling), e.g. by changing the number of glutamatergic receptors at the synapse. This is considered as homeostatic synaptic plasticity and is crucial for stable neural circuits.[281] Tet3 seems to regulate these processes by influencing the expression levels of the surface glutamate receptor 1 (GluR1) that plays a key role in synaptic transmission and scaling.[16, 282] One parameter to characterize synaptic transmission processes at glutamatergic synapses is the miniature glutamatergic excitatory postsynaptic current (mEPSC). Although releasing of neurotransmitters into the synaptic cleft is strictly controlled and only happens massively in the healthy brain when the neurons are stimulated, there is spontaneous release

of single vesicles, which is commonly referred to as mEPSC. This event even happens when the voltage-gated sodium channels are blocked by tetrodotoxin (TTX) and action potential can neither be formed nor propagated.[283, 284] When Tet3 is knocked down in hippocampal neurons, the global neuron properties do not change, but the mEPSC is increased.[16] An increase of the mEPSC amplitude is also observed for knockdown of either Tet1 or Tet2, however the effects are less pronounced. One factor that controls excitatory synaptic transmission and synaptic scaling is Bdnf.[285, 286] As shown for Tet1, also Tet3 regulates *Bdnf* expression by DNA demethylation of promoter IV. When central components of base-excision repair pathway (BER) are inhibited, demethylation of *Bdnf* promoter IV is prevented and subsequent effects as the change of mEPSC in case of Tet3-knockdown are not detectable anymore. This indicates that DNA demethylation is mediated by TET-dependent oxidation of 5mdC and subsequent DNA repair by BER.[16] Additionally, Tet3 expression levels are increased with elevated synaptic activity induced by bicuculline, whereas reduced synaptic activity induced by TTX[287], reduces Tet3 expression levels.[16] Interestingly, the expression levels of neither Tet1 nor Tet2 were reported to be affected by this treatment, although Tet1 expression levels were reported to be decreased in hippocampal neurons upon depolarization with KCl.[273]

Taken together, TET activity in the brain is important to establish 5hmdC as an epigenetic mark, but also to initiate DNA demethylation. So far, neuronal activity-dependent DNA demethylation was reported to be obtained by TET-induced oxidation of 5mdC in combination with subsequent DNA repair.[16, 274] Whether DNA demethylation is also mediated by deformylation of 5-formyl-2'-deoxycytidine (5fdC)[288] or decarboxylation of 5-carboxy-2'-deoxycytidine (5cadC)[268] in brain, is not elucidated yet. Although the TET expression levels of different brain areas are similar, there are remarkable differences in the 5hmdC levels, albeit the 5hdmC levels of all brain areas are high when compared to other tissues.[20, 21] For example, despite lower *TET* gene expression in the cortex compared to the cerebellum, the 5hmdC levels in the cortex are two times higher.[20] In both brain regions, the global amount of 5mdC is the same.[20] This indicates that neither substrate availability nor *TET* gene expression are determining the 5hmdC levels in the brain[259], but other factors such as translation efficiency, PTMs of TET enzymes, TET interaction partners, the availability of co-factors and other metabolites are the underlying reason.

As explained before, DNA modifications are the first layer of epigenetic information and histone modifications are the second one. Histone acetylation in general is linked to improved learning and facilitated long-term memory formation and retrieval[289, 290] and HATs and histone deacetylases (HDACs) are the best characterized chromatin-modifying enzymes in neurons[4, 291]. The application of sodium butyrate, a non-specific HDAC inhibitor[292], to mice improves long-term memory formation in the hippocampus even after massive neural loss.[290] Yet, the effect of HDACs in brain depends on the HDAC class.[3] In a further study on the role of class I HDACs in brain, *Guan et al.* reported that specifically Hdac2 negatively regulates synaptogenesis by suppressing the expression of neuronal genes such as *Bdnf*, synaptic plasticity and long-term memory formation in the hippocampus.[289] Mice with neuron-specific overexpression of Hdac2 have normal brain anatomy, locomotion and short-term memory, but impaired function of the hippocampus due to decreased dendric spine density, decreased synapse number and impaired synaptic plasticity. This, in turns, results in deficits regarding long-term memory formation and spatial learning.[289] Conversely, mice with a constitutive *Hdac2* KO also have normal brain anatomy, but show an increased synapse number and enhanced associative learning and memory formation. In the same study, Hdac1 showed neither in overexpression nor in ablation a significant effect compared to the wildtype (wt).[289] However, HDACs are not exclusively negative regulators of memory formation. In mice, neuron-specific deletion of *Hdac4*, which belongs to the class IIa HDACs, showed the inverse effect compared to *Hdac2* deletion. While the basal synaptic transmission remains unchanged, the LTP induction and therefore synaptic plasticity is impaired in the

neuron-specific *Hdac4*-deficient mice, resulting in aggravated hippocampus-dependent long-term memory formation and learning.[293] In addition to (de-)methylation and (de-)acetylation of lysine residues, also phosphorylation on serine is induced by synaptic activity.[294] The histone mark H3S10phos is associated with histone H3 acetylation and activated gene expression[295] and in hippocampal neurons also with big rearrangements of the nuclear geometry upon synaptic activity.[294]

Last, in addition to direct chromatin modifications and chromatin remodeling enzymes, ncRNAs are also part of the epigenetic control mechanisms in brain. They can regulate the expression and the activity of DNA-binding proteins that are involved in epigenetic pathway regulation.[296] For example, the expression of the methyl-CpG-binding protein 2 (MeCP2) is silenced by the ncRNA miRNA-132.[297, 298]Moreover, especially lncRNAs can recruit chromatin remodeling enzymes and transcription factors to their target on the DNA. In brain, the transcription levels of lncRNAs are high compared to other tissues and the transcribed lncRNAs are often evolutionary conserved.[44, 296, 299] They show a specific dynamical expression pattern depending on cell type and cell differentiation during development[44, 296, 300-303] and are essential for neurogenesis[303, 304]. For example, in the CNS, lncRNAs are involved in the recruitment of the repressor element 1-silencing transcription factor (REST) and the REST corepressor (coREST), both being master regulators of neural genes and neural cell fate decisions.[296, 305]

Taken together, the highly complex and diverse tasks of the brain demand a complex regulation of the cellular processes in all type of neural cells. Epigenetic mechanisms are a crucial part of the regulatory network to ensure correct brain maturation and function. Even small changes to this epigenetic network can have fatal effects on brain function. One well-studied example for this are several known point mutations in the *MeCP2* gene, with each of them causing independently a developmental disease known as the Rett syndrome (RTT).

MeCP2 and Neural Function

RTT is a severe neurological disorder with a prevalence of 1/10000 in female living births, which is mainly caused by loss-of function mutations within the *MeCP2* gene.[306] MeCP2 is not exclusively expressed in mature neurons, but the highest expression levels were found in postmitotic neurons, whereas the protein is not needed in early neurogenesis.[307] Since MeCP2 is encoded on the X-chromosome and prevalence is almost 100% for certain MeCP2 mutations in heterozygote females, much more cases are reported for women than for men. Men can also suffer from RTT. However, they are homozygote for MeCP2 mutations and therefore many of them die prenatally or in early childhood due to severe neonatal encephalopathy. Nevertheless, there are reported cases in males caused by *de novo* mutations in MeCP2 later in development, which result in somatic mosaicism. Depending on the type of mutation these cases can show similarities to RTT-affected females.[306, 308] Clinical criteria for RTT are based on affected females and RTT can manifest in a wide range of clinical phenotypes. All reported cases of RTT have in common that the disorder starts with a period of mental regression between six and 18 months after birth, followed by a period of stabilization, before the cognitive decline continues. The four main clinical criteria for typical RTT patients are “Partial or complete loss of acquired purposeful hand skills”, “Partial or complete loss of acquired spoken language”, “Gait abnormalities” and “Stereotypic hand movements”.[309] Atypical cases, which can be less or more severe than typical cases,[306] must at least fulfill two of the four main criteria and additionally at least five of eleven supporting criteria such as “Abnormal muscle tone” or “Growth retardation” to be diagnosed as RTT.[309] Overall, MeCP2 mutations were found in about 75% of RTT patients, with 95% in typical cases and 50% in atypical cases.[17, 306, 308] However, as RTT can also manifest without MeCP2 mutations, RTT is a clinical and not a genetic diagnosis.[308]

Several RTT-mouse models have been established.[18, 310-318] In mice, Mecp2 is also encoded on the X-chromosome and the murine Mecp2 shares more than 95% sequence identity with the human MeCP2. Mecp2-deficient mice (*Mecp2*$^{-/-}$ (females) or *Mecp2*$^{-/Y}$ (males)) develop as unaffected mice until approximately six weeks after birth. After that period, they develop a severe phenotype with uncoordinated motor behavior and breathing difficulties and therefore die within the next weeks. Neurons of those mice are smaller in size, but apart from that do not show major structural abnormalities or signs of neurodegeneration, which is in line with examinations of postmortem brains of RTT patients.[18, 306, 311] For heterozygote female mice (*Mecp2*$^{-/+}$) it was reported, that they have an apparently normal phenotype until young adulthood (six months of age) and start to show RTT-like neurological symptoms afterwards.[18, 306] Recent studies, however, suggest that motor dysfunctions start already at about six weeks of age, which is more reminiscent of RTT-development in humans than the previously reported six months.[318] The important role of Mecp2 especially for neuronal functioning is indicated by the fact that cKO of Mecp2 in neurons of the CNS during embryonal development results in the same phenotype as global Mecp2 ablation and cKO in postmitotic neurons leads to a delayed, but eventually similar, RTT-phenotype.[311] However, not only MeCP2 loss-of-function results in neurological dysfunctions. Mild overexpression of Mecp2 in mice leads to higher susceptibility towards epileptic seizures with increasing amplitude and frequency when endogenous Mecp2 is functional but rescues RTT-phenotype of Mecp2-deficient mice.[310, 313] *De novo* expression of functional Mecp2 in neurons can reverse the RTT neurological phenotype even in mature heterozygote female mice, whereas induced loss of Mecp2 in the adult murine brain leads to the RTT phenotype, showing that Mecp2 function is required not only for brain development, but for the entire lifespan.[312, 319, 320]

MeCP2 was originally discovered as a binding protein of 5mdC in a CpG dinucleotide context, with highly increased binding affinity when dA or T are adjacent to 5mdC.[321, 322] In many mammals, two isoforms exist through alternative splicing of the N-terminus. In humans, MeCP2B (Mecp2alpha or Mecp2E1 in mice) and MeCP2A (Mecp2beta or Mecp2E2 in mice) exist (Figure 10). MeCP2B is the

longer isoform (498 in B vs. 486 amino acids in A) and in neurons, lung and thymus ten times higher expressed than isoform MeCP2A, although both isoforms are highly expressed in those cells.[323, 324] MeCP2 shows transcriptional repression activity *in vitro* and *in vivo* and recruits a histone-deacetylase complex, including the corepressor mSin3a[325] and the nuclear receptor corepressor (Ncor)/ silencing mediator of retinoic acid and thyroid hormone receptor (SMRT) corepressor complex[326], via its transcriptional repressor domain (TRD).[306, 322, 327, 328] At the N-terminus, MeCP2 features the methyl-CpG binding domain (MBD), which is sufficient for DNA-binding and the C-terminus provides a domain for additional protein-protein-interactions.[306] Most mutations within the *MeCP2* gene are not inherited, but happen *de novo* during spermatogenesis and almost 70% of the mutations resulting in the RTT phenotype are dC to T transitions (Figure 10).[329] Point mutations that are located upstream or within the TRD result in general in a more severe phenotype than mutations in the C-terminal protein-protein-interaction domain. Mutations that are exclusively in the TRD lead to a MeCP2 protein that is still able to bind heterochromatin, but not to recruit the histone-deacetylation complex.[315, 330, 331]

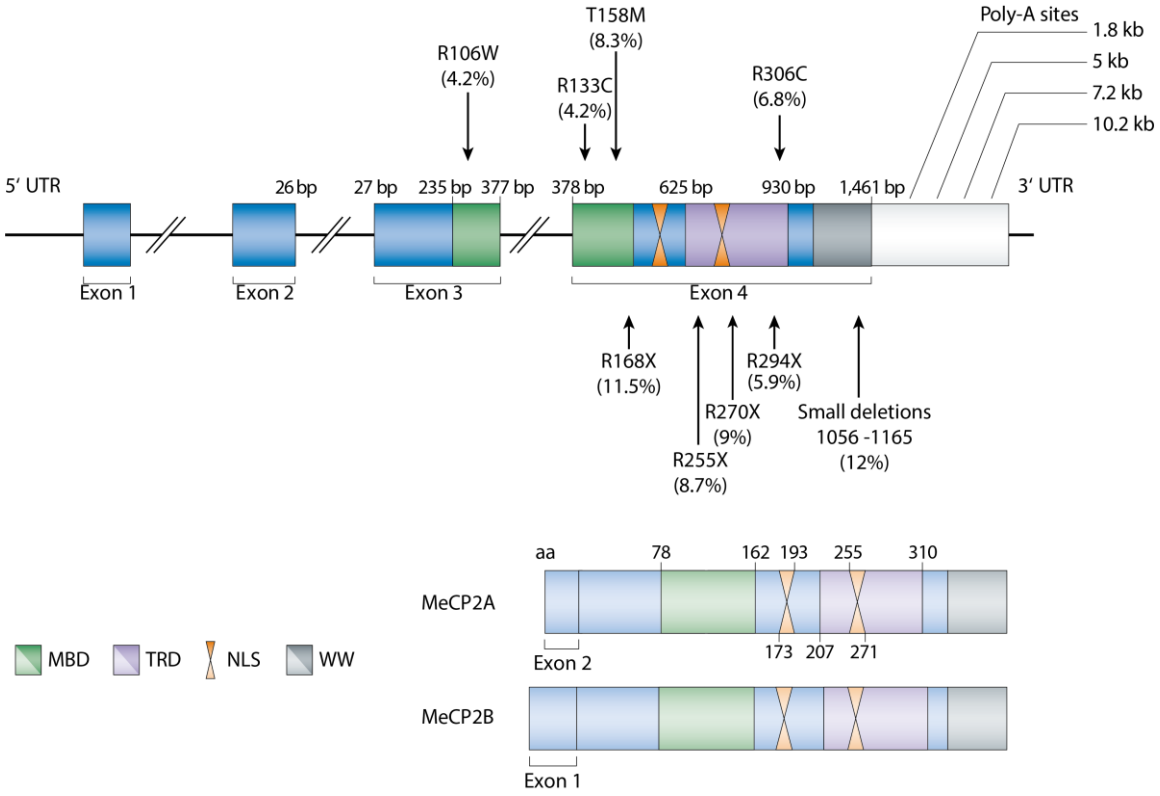


Figure 10: **MeCP2 isoforms and Rett syndrome-linked mutations thereof.** Figure adapted from *Bienvenu et al., 2006*.[306] aa = amino acids, bp = base pairs, C = cysteine, kb = kilo base pairs, M = methionine, MBD = methyl-CpG binding domain, MeCP2 = methyl-CpG binding protein 2, NLS = nuclear localization signal, Poly-A-site = polyadenylation site, R = arginine, T = threonine, TRD = transcription repression domain, UTR = untranslated region, W = tryptophan, WW = RNA splicing binding factor region, X = stop codon.

Recently, it was reported, that MeCP2 does not only recognize and bind 5mdC, but also 5hmdC with similar affinity and that point mutation R133C, which has been reported in many RTT patients, preferentially disrupts the 5hmdC binding.[19] However, the affinity of MeCP2 towards 5hmdC is contentiously discussed. Whereas *Mellén et al.* observed similar binding affinities of MeCP2 towards 5mdC and 5hmdC and identified MeCP2 as the major 5hmdC binding protein in brain[19], other groups reported lower binding affinity towards 5hmdC when compared to 5mdC and identified additional specific 5hmdC readers [34, 332]. Interestingly, five years after the initial study, *Mellén et al*. recently proposed that hydroxymethylation of 5mdC in a CpA context does not alter MeCP2 binding, in contrast to 5hmdC in a CpG context where MeCP2 binding is reduced upon 5mdC oxidation.[12] MeCP2 does not

seem to be involved in establishing the 5hmdC pattern in neurons[19] and data of the effect of MeCP2 ablation in neurons on the global 5hmdC levels are inconsistent. While *Mellén et al.* reported a small, but significant decrease of intragenic 5hmdC of actively transcribed genes in Mecp2-deficient murine neurons[19], *Szulwach et al.* observed a negative correlation between Mecp2 dosage and 5hmdC[234]. Moreover, the existing data about effects of MeCP2 ablation on gene expression are also confounding. Although MeCP2 was first reported to be a global transcriptional repressor, Mecp2-deficient mice only show little changes in global gene expression in brain.[317, 333] In human RTT-patients, in contrast, changes seem to be more dramatic.[334, 335] Furthermore, MeCP2 contributes directly and indirectly to gene activation in the hypothalamus[336] and seem to act as a splicing regulator in brain[337]. Although changes in gene expression levels do not seem to be dramatic in absolute numbers in Mecp2-deficient mice, even small changes can cause severe neurological effect when genes that are important for synaptic functioning, as for example *Bdnf*, are affected. Transcription of *Bdnf* is known to be repressed by Mecp2, which binds at *Bdnf* promoter III. Upon neuronal stimulation, the influx of Ca^{2+} leads to the phosphorylation of Mecp2 and consequently to its release from the *Bdnf* promoter III and to *Bdnf* expression. However, compared to membrane depolarization in neurons that can result in 100-times higher expression of *Bdnf*, the effect of Mecp2 release from *Bdnf* promoter III is small with only two times higher expression of *Bdnf* as a result.[265, 338] Latest data indicate that the mechanism of MeCP2 to influence gene expression works by slowing down elongating RNA polymerase II, which could also explain the reversibility of RTT in mouse models.[339] Taken together these data suggest, that the importance of MeCP2 for proper neuronal functioning goes beyond simple global expression control as a transcription factor. For example, additional findings point towards a role of MeCP2 in exon recognition and exon skipping events.[340]

Although there are no major structural changes in the neurons, small changes of neuronal dendrites structures and dendritic spines were observed for RTT patients and in the existing mouse models.[341] Dendritic spines are crucial for synapse functioning[342] and manifestation of RTT occurs in a period of enhanced synaptogenesis.[341, 343] These findings suggest that synaptic transmission is impaired in RTT patients, which eventually leads to the RTT phenotype. More evidence for this hypothesis came from numerous studies of synaptic transmission processes in mouse models.[329, 341] Basal synaptic transmission and short-term plasticity were shown to be affected by Mecp2 mutations and the balance between excitatory and inhibitory brain activity, which is an integral part of neural functioning[344], is disturbed in Mecp2-mutant mice.[341, 345, 346] Mecp2-mutant mice show altered LTP and LTD in the hippocampus. These changes in synaptic plasticity eventually lead to impaired memory formation and learning difficulties.[347, 348] Regarding the manifestation of these changes the data are not overall consistent, but they seem to occur before RTT symptoms are observed.[348] Furthermore, Mecp2 regulates the number of glutamatergic synapses[349] and dysfunction of GABAergic Mecp2-deficient neurons is sufficient to trigger many symptoms of RTT[350]. Initially, MeCP2 was thought to play an exclusive role in neurons, not in glia cells.[306] This view changed within the last years. In 2009, *Maezawa et al.* showed that astrocytes express Mecp2 as well and astrocyte-specific deficiency of Mecp2 leads to aberrant *Bdnf* expression regulation and neuronal dendritic induction.[351] Importantly, they observed in heterozygous *Mecp2*$^{+/-}$ female mice that the Mecp2-deficient astrocytes impose their state partially via gap junction communication[352] on astrocytes with functional Mecp2. These findings might be one piece of the puzzle to explain why prevalence of RTT is almost 100% for certain MeCP2 mutations. Within the last years, more evidence for a role of glia cells in RTT was reported. Mecp2-deficient microglia show fivefold higher glutamate release compared to the wt that damages dendrites and synapses [353] and excitatory synaptic signaling depends on Mecp2 expression in astrocytes[354].

In summary, the role of MeCP2 and the consequences when MeCP2 is not functional are an excellent example to display the importance of epigenetic mechanisms for neural function. Interventions in the

epigenetic landscape can result in severe neurological defects. In case of MeCP2, which is important for brain functioning, not initial development, these defects seem to be potentially reversible given that expression of functional MeCP2 could be induced in MeCP2-mutation affected neural cells. However, the mechanisms behind the effects of MeCP2 on the epigenetic landscape and vice versa have still to be elucidated, as many contentious results on the molecular biology level, especially on the interplay of MeCP2 and 5hmdC, were reported.

Results

Manuscript draft

Interaction of Gdh with Tet3 to supply α-ketoglutarate links neuronal metabolism to 5mdC oxidation

Neural activity induces changes on the epigenetic landscape of neurons and glia[1, 2] Dynamic changes of DNA (hydroxy-)methylation patterns are an integral part of this regulatory mechanism to which ten-eleven translocation enzymes (TET, Tet1-3) contribute by oxidizing 5'-methyl-2'-deoxycytidine (5mdC) to 5'-hydroxymethyl-dC (5hmdC) and further on to 5'-formyl-dC (5fdC) and 5'-carboxy-dC (5cadC), which are finally replaced by unmodified dC.[3-5] TET activity depends on α-ketoglutarate (αKG)[6] that plays a pivotal role in energy metabolism and cellular functioning.[7, 8] A recent study on TET3 deficiency in humans shows that the affected patients suffer from intellectual disability and growth retardation[9], which emphasizes the importance of Tet3 for neural function. We show that in hippocampal neurons Tet3 interacts with glutamate dehydrogenase 1 (*Glud1*, Gdh), which converts ubiquitous glutamate into the rate-limiting αKG co-substrate. Gdh is a mitochondrial enzyme, but in hippocampal neurons, Gdh is transported into the nucleus instead. In contrast, we do not observe this effect in liver where *Glud1* expression is high, but *TET* expression and activity are low in a non-fasted state.[10, 11] When hippocampal neurons are depolarized and we apply the Gdh inhibitor R162[12], we observe lower levels of global 5hmdC, along with a decrease of the gene expression levels of the neural activity-dependent genes neuronal PAS domain containing protein 4 (*Npas4*)[13] and brain-derived neurotrophic factor (*Bdnf*)[14] compared to the depolarized control without R162. Our results show that Gdh activity controls Tet3-dependent 5mdC oxidation in hippocampal neurons and thereby activation of neural activity-dependent genes, linking neuronal metabolism to 5mdC oxidation and thereby epigenetic plasticity and memory formation.

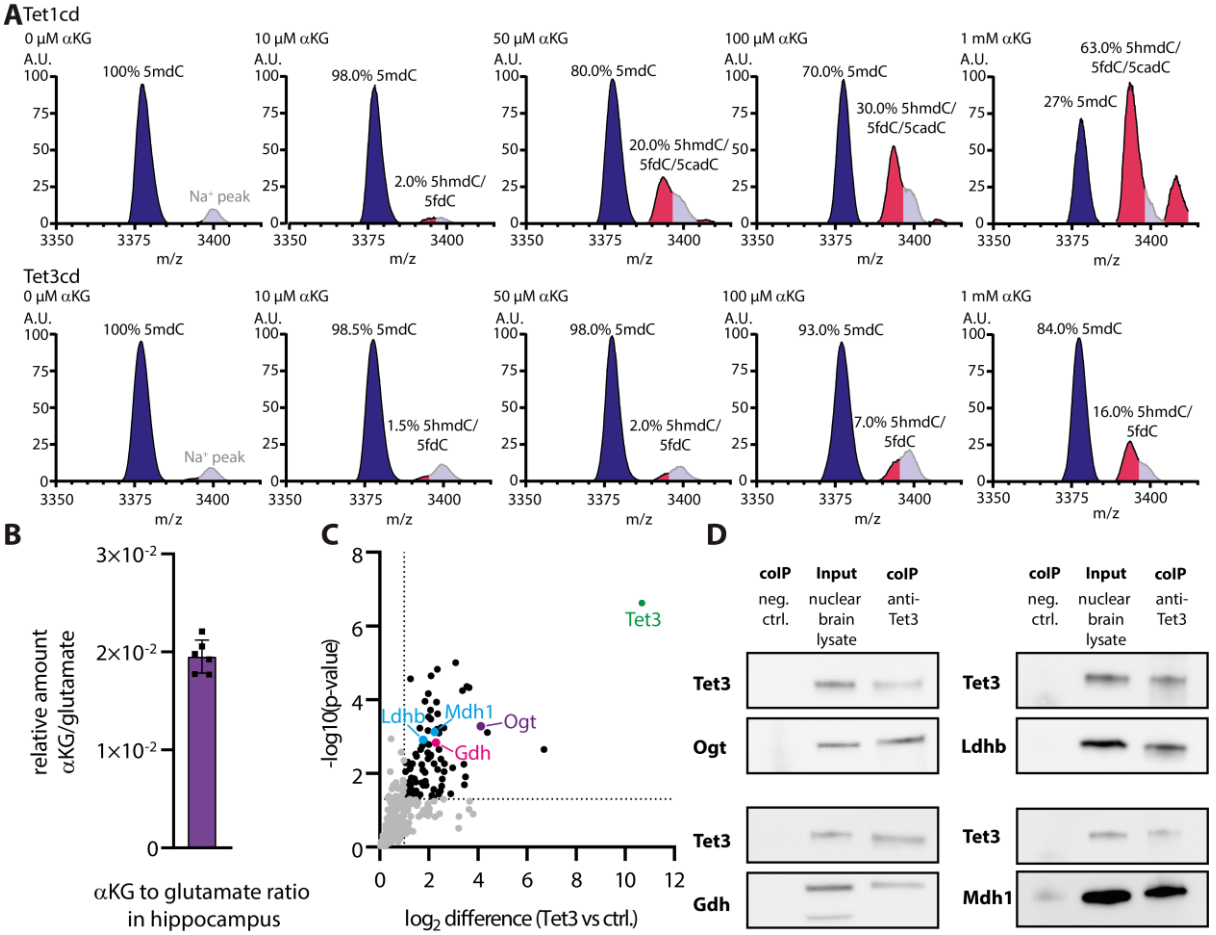
Introduction

DNA methylation, hydroxymethylation and demethylation processes influence gene expression and in case of neural activity-dependent genes consequently synaptic plasticity, learning processes and memory formation.[15, 16] In promoter regions, 5mdC correlates with low gene expression; whereas 5mdC oxidation by TET enzymes correlates with reactivation of gene expression by either establishing 5hmdC as a new epigenetic mark or by further oxidation of 5hmdC to 5fdC and 5cadC and their subsequent removal from the genome.[17, 18] In fully differentiated cells, the highest expression levels of TET enzymes and of 5hmdC are found in adult brain, where Tet3 is the most abundant among them[11, 19] (Extended Data Figure 1A). In hippocampal neurons, Tet3 is important for synaptic transmission and the maintenance of the homeostatic synaptic plasticity.[4] The concept that cellular metabolism dictates cellular fate by influencing the epigenetic landscape is already well established for differentiating cells and cancer cells.[20] In both cases, the amount of intracellular αKG plays a pivotal role as it is a key intermediate of the citric acid cycle (TCA cycle), important for balancing nicotinamide adenine dinucleotide (NAD^+/NADH) levels between the mitochondria and the cytosol and a co-substrate for TET enzymes and other αKG-dependent dioxygenases that are involved in epigenetic changes.[20, 12, 8, 21] In brown adipose tissue (BAT) development an increase of the cellular αKG concentration towards the Michaelis constant K_m of Tet1 could be directly linked to increased Tet1 activity and consequently demethylation of promoter regions of BAT-development specific genes.[21] If a similar regulatory mechanism for TET enzymes exists in neurons is unknown. Brain is a metabolically very active tissue.[22] However, the dependencies of the epigenetic landscape on the DNA modification level on neural metabolism are only poorly understood. Here we show that in the nucleus of hippocampal neurons, Tet3 interacts with Gdh, which supplies αKG from abundant glutamate[23] for the oxidation reaction. The transport of Gdh to the nucleus instead to the mitochondria, where the enzyme is normally located, seems to be Tet3-dependent. This interaction in neuronal nuclei allows to increase the effective molarity of αKG around Tet3 to specifically boost its activity without changing the global αKG level in the cell.

Results

Reported *in vitro* assays with Tet1, Tet2 and Tet3 showed significantly lower activity for Tet3 than for Tet1 or Tet2, which were similar in activity.[24, 25] In a first step, we compared the activities of the catalytic domains of Tet1 and Tet3 in an *in vitro* assay with increasing αKG concentrations (Data Figure 1A, Extended Data Figure 1B). At 10 µM αKG, both enzymes, Tet1 and Tet3, showed low activity and oxidized between 1.5 and 2% of 5mdC. At 50 µM concentration, however, about 20% of the 5mdC substrate was already oxidized to 5hmdC and 5fdC by Tet1, whereas Tet3 only converted about 2%. Both enzymes showed increased activity towards higher αKG concentrations. But in contrast to Tet1, which oxidized at 1 mM αKG more than 60% of 5mdC to 5hmdC, 5fdC and even to 5cadC, Tet3 reached at 1 mM αKG just a similar conversion rate of almost 20% as Tet1 did at 50 µM αKG, indicating a lower turnover rate for Tet3 than for Tet1 under the same conditions. Compared to the glutamate concentration, which was reported to be in the low millimolar range[23], the measured αKG concentration was 50 times lower in the hippocampus (Data Figure 1B). Based on these results, we were wondering whether there is a direct αKG supply system for Tet3 in neurons. Therefore, we checked for Tet3 interaction partners using a Tet3-enriched co-immunoprecipitation (coIP) (Data Figure 1C) and a Tet3-endogenous coIP approach (Data Figure 1D, Extended Data Figure 1C). Mass spectrometric (MS) analysis of the Tet3-enriched coIP in total murine nuclear brain lysate revealed several possible interaction partners that were significantly enriched compared to the control, among them the previously identified Tet3 interaction partner *O*-linked β-*N*-acetylglucosamine transferase (Ogt)[26] and three metabolic dehydrogenases – Gdh, lactate dehydrogenase b (Ldhb) and malate dehydrogenase 1 (Mdh1) (Data Figure 1C). All of them could be confirmed as Tet3 interactors in the

Tet3-endogenous coIP that was analyzed by MS and western blotting (Data Figure 1D, Extended Data Figure 1C). Gdh directly converts glutamate, which is ubiquitous in the cell, to the rate-limiting αKG co-substrate, making it a perfect match for Tet3 to specifically control and stimulate Tet3-activity. This interaction allows to establish a very high αKG concentration in the direct environment of Tet3 without changing the global αKG concentration in the cell, which might have various undesired side effects. Ldhb and Mdh1 are not directly associated with αKG turnover but are especially important to control cytosolic and nuclear NAD^+ levels[27, 28] and could therefore provide NAD^+ for Gdh. In accordance with this hypothesis, Ldhb was already reported to be in the nucleus to provide NAD^+ for NAD^+-dependent histone deacetylases.[29] Immunofluorescence staining confirmed the presence of Ldhb and Mdh1 in the nucleus of hippocampal neurons (Extended Data Figure 2A).



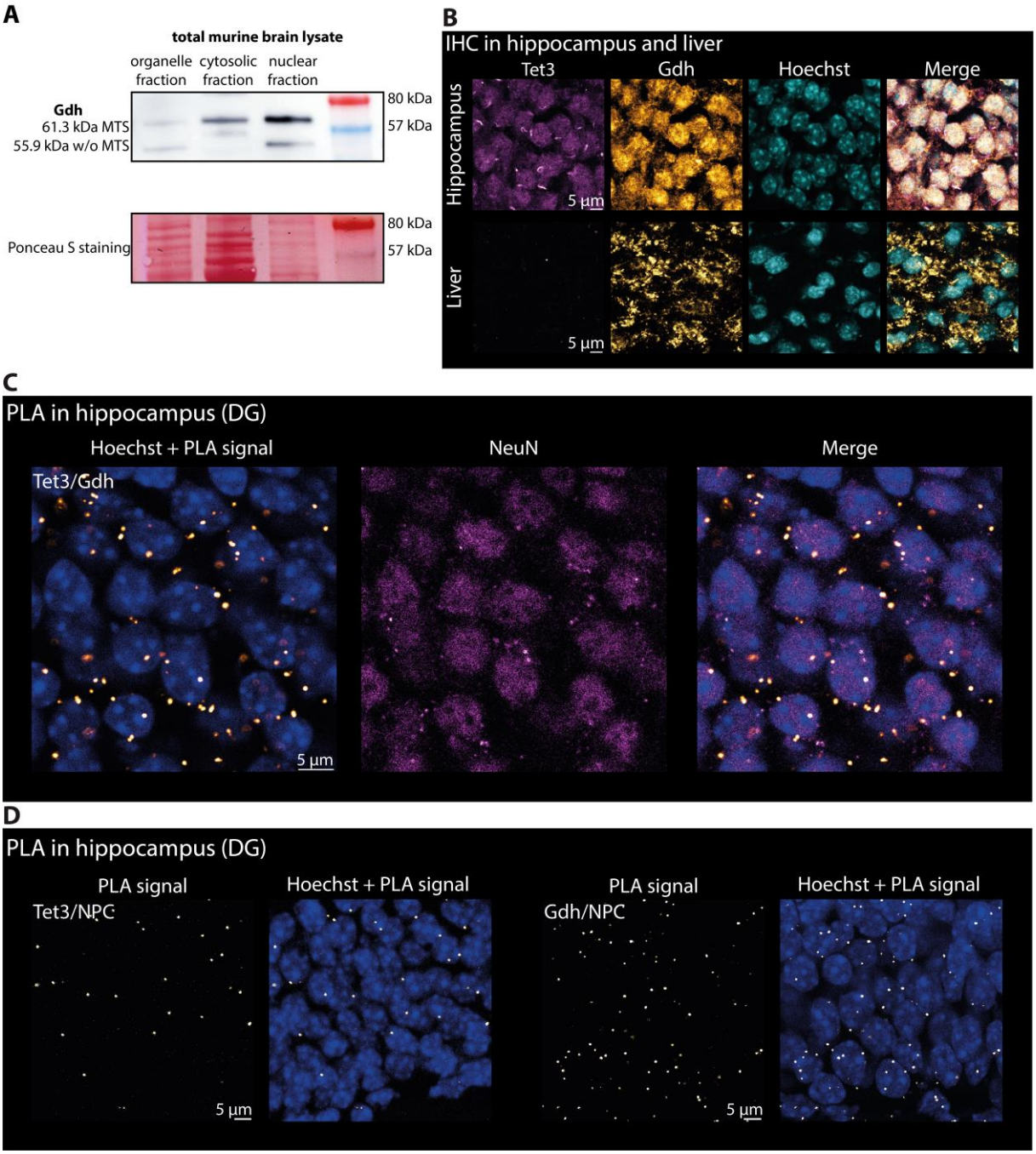
Data Figure 1: **Kinetics of Tet1cd and Tet3cd, αKG to glutamate ratio in the hippocampus and Tet3 interactors in brain.** (A) Mean MALDI-TOF MS spectra (n=3) of a single stranded DNA oligonucleotide containing 5mdC (3381.6 m/z (blue) and 3403.6 m/z (grey, Na^+ peak of 5mdC)) that is oxidized *in vitro* to 5hmdC (3397.6 m/z), 5fdC (3395.6 m/z) and 5cadC (3411.6 m/z) (all magenta) by Tet1cd or Tet3cd in presence of different concentrations of αKG. (B) Ratio of αKG to glutamate in murine hippocampus. (C) Volcano plot of the Tet3-interaction partners found in the Tet3-enriched coIP in nuclear brain lysate of adult mice (n=4; FDR=0.05; s0=2). Significantly enriched interactors ($log_2 > 1.0$; adj. p-value < 0.05) are highlighted in black or colored. (D) Western blots detecting Tet3, Ogt, Gdh, Ldhb and Mdh1 after Tet3-endogenous coIP in nuclear brain lysate.
5cadC = 5'-carboxy-2'-deoxycytidine, 5fdC = 5'-fomyl-dC, 5hmdC = 5'-hydroxymethyl-dC, 5mdC = 5'-methyl-dC, αKG = α-ketoglutarate, A.U. = arbitrary units, cd = catalytic domain, ctrl. = control, coIP = co-immunoprecipitation, FDR = false discovery rate, Gdh = glutamate dehydrogenase 1, Ldhb = lactate dehydrogenase b, MALDI-TOF-MS = matrix-assisted laser desorption ionization-time of flight mass spectrometry; Mdh1 = malate dehydrogenase 1, neg. = negative, Ogt = O-linked β-N-acetylglucosamine transferase, Tet = ten-eleven translocation enzyme.

In contrast, Gdh was previously reported to be predominantly located in the mitochondria.[30, 31] To clarify the question of localization, we analyzed the amount of Gdh in different cellular compartments of total murine brain lysate using western blot (Data Figure 2A, Extended Data Figure 2B) and immunohistochemistry (IHC) (Data Figure 2B). We detected a low amount of Gdh in the organelle

fraction, including the mitochondria, and the cytosolic fraction. However, the majority of Gdh was detected in the nuclear fraction (Data Figure 2A, Extended Data Figure 2B). Interestingly, we observed in the organelle fraction two equally prominent signals for Gdh, one above and one below the 57 kDa marker band. In the cytosolic fraction one weak signal above 80 kDa appeared, but the most prominent signal was the one above 57 kDa and there were some very weak signals below 57 kDa that probably represented aborted fragments. In the nuclear fraction, we predominantly detected the signal above 57 kDa. *Glud1* is nuclear encoded and possesses a 53 amino acids long mitochondrial targeting sequence (MTS) that is cleaved off after the protein has been successfully imported into the mitochondria.[30, 31] In terms of size, the >57 kDa band fits to Gdh including MTS (61.3 kDa), while the <57 kDa band fits to Gdh without MTS (55.9 kDa). The organelle fraction contains both versions as the cleavage takes place in the mitochondrial matrix after import. It is expected that in the cytosol only the version including the MTS can be found. The small amount of the short version in the nuclear fraction was therefore most likely a minor contamination of the nuclear fraction with the organelle fraction. Interestingly, Tet3-endogenous coIP showed that Tet3 is only interacting with Gdh including the MTS (Data Figure 1D), supporting the conclusion that the MTS is relevant for the Tet3-Gdh interaction and Gdh is transported to the nucleus before its mitochondrial import.

IHC confirmed a predominantly nuclear localization for Tet3 and Gdh in hippocampal neurons (Data Figure 2B, Extended Data Figure 2C/D). In contrast, in liver from non-fasted adult mice, where we failed to detect Tet3, we observed Gdh only in the mitochondria (Data Figure 2B, Extended Data Figure 2C/D). The different localization of neuronal Gdh compared to Gdh in the liver raised the question how Gdh is transported to the nucleus despite its MTS. When (murine) *Tet3* and *Glud1* were ectopically expressed simultaneously in HEK293T, we observed that almost 40% of the transfected cells were positive for Gdh in the nucleus and the mitochondria compared to those that exclusively had Gdh in the mitochondria (Extended Data Figure 3A/B). *Glud1* expression without parallel expression of *Tet3* resulted in exclusive mitochondrial localization of Gdh (Extended Data Figure 3B). These results indicated that Gdh localization is directly influenced by Tet3 and that the MTS is important for the interaction.

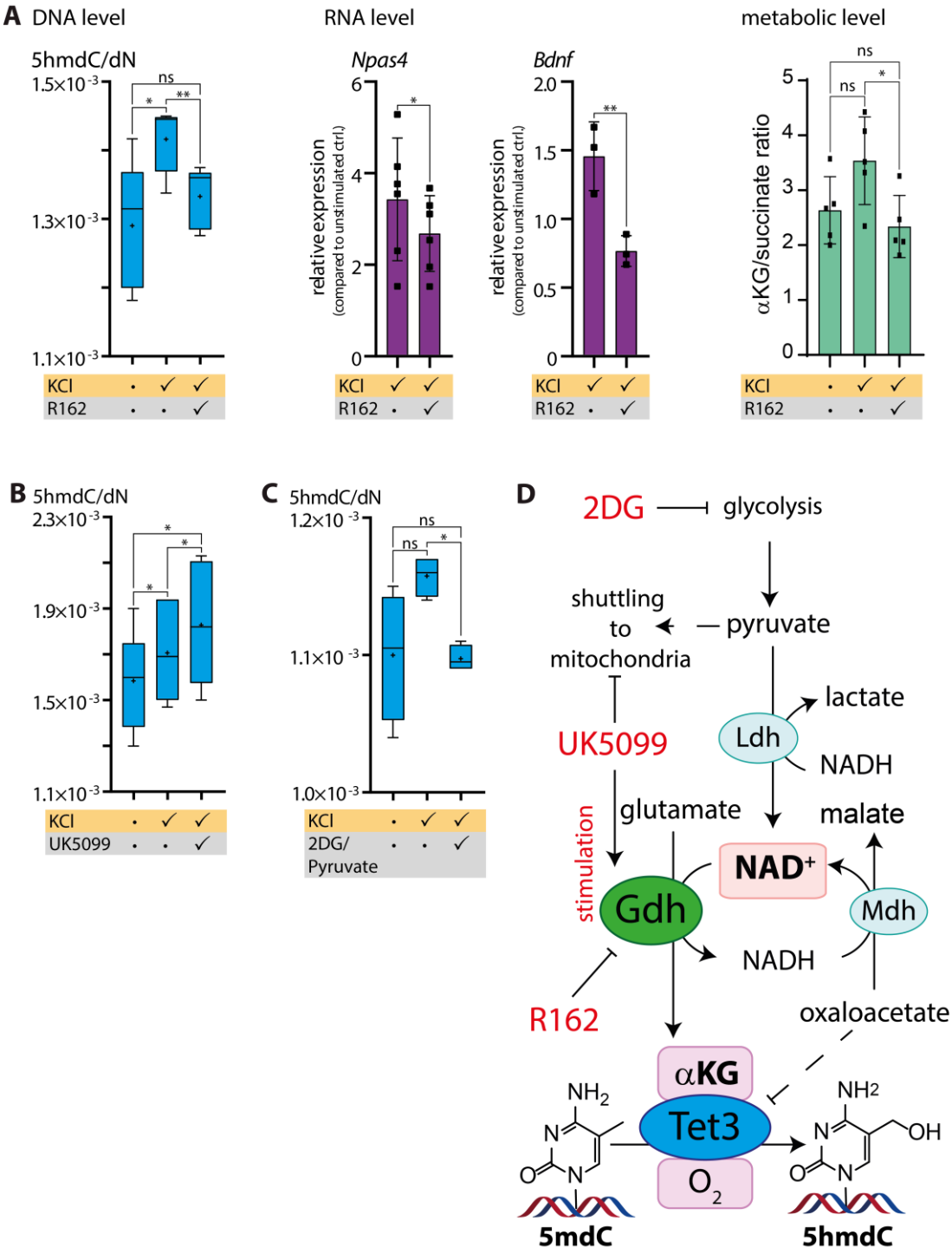
Next, we performed a proximity ligation assay (PLA) to check for co-localization of Tet3 and Gdh interaction and obtained a positive PLA signal in the dentatae gyrus region (DG) of the murine hippocampus (Data Figure 2C, Extended Data Figure 4A). Interestingly, the PLA-signal was mostly located around the nucleus and not in the nucleus. Spatial organization of the chromatin in the nucleus was shown to have a direct effect on gene expression. Whereas localization of a gene at the nuclear laminar is associated with gene expression, proximity to the nuclear pore complexes (NPCs) is linked to active gene expression.[32] We tested in an additional PLA whether Tet3 and Gdh are located near the NPCs and obtained a positive signal for both (Data Figure 2D). This finding supports the idea that Tet3 and Gdh are mainly interacting in areas of active gene expression in the DG. Further evidence for this hypothesis was provided by our proteomics data after Tet3-enriched and Tet3-endogenous coIP. The data showed significant enrichment of several RNA-processing proteins, among them for example Fus, a DNA/RNA binding protein with several functions including pre-mRNA binding and transport and distinct functions in neuronal homeostasis.[33]



Data Figure 2: **Localization of Gdh in total murine brain and in the DG region of the murine hippocampus and its interaction with Tet3.** (A) Western blot against Gdh in different cellular compartments of total murine brain. Ponceau S staining shows total protein load. (B) Localization of Gdh and Tet3 in the DG of the murine hippocampus compared to liver. (C) PLA of the Tet3/Gdh interaction in the DG of the murine hippocampus. NeuN staining shows neuronal nuclei. (D) PLA of the Tet3/NPC and Gdh/NPC interaction in the DG of the murine hippocampus. (B) – (D) Images show Z-stacks.
DG = dentate gyrus, Gdh = glutamate dehydrogenase 1, IHC = immunohistochemistry, MTS = mitochondrial targeting sequence, NPC = nuclear pore complex, PLA = proximity ligation assay, Tet3 = ten-eleven translocation enzyme 3.

To investigate the role of Gdh activity for 5mdC oxidation during neuronal activation, we depolarized mouse acute hippocampal slices for six hours using 25 mM of potassium chloride (KCl) in absence and presence of several inhibitors and measured dC modifications levels. Depolarization resulted in slightly increased global 5mdC levels, but overall the global 5mdC levels did not differ significantly among all the tested conditions (Extended Data Figure 4B). However, we consistently detected an increase of global 5hmdC in response to the KCl stimulus, whereas simultaneous application of the Gdh inhibitor R162[12] did not change the 5hmdC levels compared to the unstimulated control (Data Figure 3A). This result confirmed that Gdh activity is important to control TET activity during neuronal activation. One consequence of neuronal stimulation is the activation by DNA demethylation of promoter genes of several neural activity-dependent genes that modulate the response of the neuron towards present and future stimuli. Among those genes, *Npas4* and *Bdnf* play a pivotal role for synaptic homeostasis, learning and memory formation, whereby the transcription factor Npas4 controls the expression of *Bdnf*.[13] When hippocampal neurons were depolarized, we observed an increase of *Npas4* and *Bdnf* gene expression. Applying R162, *Npas4* was 25% less expressed than without R162 and for *Bdnf* expression, we did not see an increase at all when R162 was applied compared to the unstimulated control (Data Figure 3A). On a metabolic level, depolarization of hippocampal neurons resulted in an increased αKG/succinate ratio, but when we applied R162, this ratio was significantly decreased (Data Figure 3A).

To gain further insight how metabolic changes influence TET activity during neural activation, we applied UK5099 during depolarization. UK5099 inhibits the mitochondrial pyruvate carrier (MPC) and therefore attenuates TCA and oxidative phosphorylation (OxPhos) supplied by glucose catabolism, but positively stimulates Gdh activity and aerobic glycolysis to produce lactate.[34] In agreement with our model, that Tet3 activity is enhanced through interaction with Gdh, which receives NAD^+ from Ldhb activity, we observed an increase of global 5hmdC beyond the increase that we observed with the KCl stimulus alone (Data Figure 3B). In a complementary experiment, we blocked glycolysis and thereby conversion of pyruvate to lactate by Ldhb by addition of 2-deoxy-D-glucose (2DG) and simultaneously fed pyruvate to maintain TCA and OxPhos. Under these conditions, we did not observe an increase of the 5hmdC levels above the levels of the unstimulated control (Data Figure 3C). This indicates that the metabolic state, especially whether the cells are in a glycolytic state or not, directly influences 5mdC oxidation by TET enzymes.



Data Figure 3: **Changes of global 5hmdC level and of neural activity dependent genes expression levels in mouse acute hippocampal slices upon depolarization with 25 mM KCl and application of different inhibitors and schematic overview of interaction network.** (A) Global 5hdmC levels in absence and presence of 20 µM of Gdh inhibitor R162. Tukey boxplot, n = 6 (measured in technical duplicates). Relative gene expression levels of *Npas4* (Tukey boxplot, n = 6, measured in technical duplicates) and *Bdnf* (n = 3, measured in technical duplicates) compared to the unstimulated control. αKG to succinate ratio in absence and presence of 20 µM of Gdh inhibitor R162 (n = 5). (B) Global 5hdmC levels in absence and presence of 40 µM of mitochondrial pyruvate carrier inhibitor UK5099. Tukey boxplot, n = 6 (measured in technical triplicates). (C) Global 5hdmC levels after depolarization of hippocampal slices using 25 mM KCl in absence and presence of 10 mM glycolysis inhibitor 2DG/2 mM pyruvate. Tukey boxplot, n = 4 (measured in technical triplicates). (D) Schematic overview of the proposed Tet3/Gdh/Ldhb/Mdh1 interaction network. ns p-value ≥ 0.05, * p-value < 0.05, ** p-value < 0.01.

2DG = 2'-deoxyglucose, 5hmdC = 5'-hydroxymethyl-2'-deoxycytidine, αKG = α-ketoglutarate, Bdnf = brain-derived neurotrophic factor, dN = nucleotide, Gdh = glutamate dehydrogenase 1, Ldhb = lactate dehydrogenase b, Mdh1 = malate dehydrogenase 1, NAD^+ = nicotinamide adenine dinucleotide (oxidized), NADH = nicotinamide adenine dinucleotide (reduced), Npas4 = neuronal PAS 4 domain, Tet3 = ten-eleven translocation enzyme 3.

Discussion

The concept that rate-limiting metabolites that are needed for the activity of chromatin-modifying enzymes, are provided by the manufacturing enzymes directly in the nucleus was reported before, but not for DNA modifying enzymes.[35] The best studied example of a mitochondrial protein complex in the nucleus is the pyruvate dehydrogenase complex (PDC), which is typically located in the mitochondrial matrix to decarboxylate pyruvate and generate acetyl coenzyme A (acetyl-CoA) to enter the TCA.[36] In the nucleus, PDC generates acetyl-CoA for histone acetylation.[37] In our study we present a novel model for metabolic control of Tet3 activity in neurons by Gdh (Data Figure 3D). Neuronal activity is energy-demanding and induces metabolic shifts first to glycolysis and later to OxPhos to meet the high energy demands of the firing neuron.[28] Gdh activity towards oxidative deamination of glutamate to αKG is heavily controlled by various parameters, including NAD^+ availability and low ATP levels. Dysregulation of Gdh activity has severe consequences for neuronal health and is linked to many neurodegenerative disorders.[38, 39] Our data suggest that not only the impaired glutamate and energy metabolism are the underlying cause for these findings, but also the direct impact of Gdh activity on Tet3-dependent 5mdC-oxidation and therefore gene expression. When neurons are firing and the energy consumption and the glycolytic rate are high[28], Gdh activity is upregulated, which stimulates Tet3 to change DNA methylation patterns with broad consequences for neural function. The increase of the effective αKG molarity only in the environment of Tet3 is an elegant way to avoid globally changing αKG levels when not desired that would impact other αKG-dependent enzymes and energy metabolism in general.

Materials and Methods

All procedures concerning animals conform to the German animal protection laws and were approved by the local authority (Regierung von Oberbayern).

All experiments were at least repeated twice to ensure reproducibility.

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Antibodies

Cytochrome C antibody (Santa Cruz Biotechnology sc-13560, clone 7H8, mouse monoclonal IgG): western blotting (1:1000)

Histone H3 antibody (Cell Signaling Technology 4499S, clone D1H2, rabbit monoclonal IgG): western blotting 1:1000

Gdh antibody (Invitrogen PA5-19267, goat polyclonal IgG, Lot #74422112): western blotting (1:1000), immunocytochemistry (ICC) (1:200), IHC (1:200)
Validation:

- Western Blotting: signal at 56 kDa for haploid mESC, no signal for haploid mESC$^{Glud1\,-/}$ (Extended Data Figure 5A)
- ICC: strong signal for haploid mESC, no signal for haploid mESC$^{Glud1\,-/}$ (Extended Data Figure 5A)

Ldhb antibody (Santa Cruz Biotechnology sc-100775, clone 431.1, mouse monoclonal IgG): western blotting (1:1000), IHC (1:10)

Mdh1 antibody (Santa Cruz Biotechnology sc-166879, clone H-6, mouse monoclonal IgG): western blotting (1:1000), IHC (1:10)

NeuN antibody (EMD Millipore MAB377X, clone A60, Alexa488 conjugated, mouse monoclonal IgG): IHC (1:100)

Nuclear Pore Complex antibody (Sigma-Aldrich N8786, clone 414, mouse monoclonal IgG): IHC (1:250)

Ogt antibody (Invitrogen PA5-22071, rabbit polyclonal IgG, Lot #RH2258725): western blotting (1:1000)

Tet1 antibody (Active Motif 61741, clone 5D6, rat monoclonal IgG): western blotting (1:1000)
Validation:

- Western blotting: one strong signal above 200 kDa for mESC, but not for mESC TET triple knockout (TKO) cells (Extended Data Figure 1A)

Tet2 antibody (ptglab 21207-1-AP, rabbit polyclonal IgG): western blotting (1:1000)

Tet3 antibody (Abiocode N1 R1092-1, rabbit polyclonal IgG, Lot #7063 and #9013):

western blotting (1:1000), coIP (1 µL for MS analysis, 2 µL for western blot analysis), ICC (1:500), IHC (1:500)
Validation:

- Western blotting: one strong signal at 190 kDa for nuclear brain lysate, but not for mESC TET TKO cells (Extended Data Figure 1)
- ICC: total signal overlap with GFP in GFP-Tet3 transfected HEK293T, but not with GFP in GFP only transfected HEK293T (Extended Data Figure 5B)
- ICC: no signal in mESC TET TKO (Extended Data Figure 5B)
- IP: validation for IP has already been reported[40]

Secondary antibodies
Anti-goat IgG (Sigma-Aldrich G4018, rabbit polyclonal): IP (1 µg for MS analysis, 2 µg for western blot analysis)

Cy2-anti-goat IgG (Jackson ImmunoResearch 705-225-147): IHC (1:200)

Cy3-anti-goat IgG (Jackson ImmunoResearch 805-165-180): IHC (1:400)

HRP-conjugated anti-goat IgG (Sigma-Aldrich A5420): western blotting (1:5000)

Alexa488-anti-mouse IgG (Cell Signaling Technologies 4408): IHC (1:800)

Alexa555-anti-mouse IgG (Cell Signaling Technologies 4409): IHC (1:800)

HRP-conjugated anti-mouse IgG (Sigma-Aldrich AP130P): western blotting (1:5000)

Cy3-anti-mouse IgG (Jackson ImmunoResearch 715-165-150): IHC (1:400)

Alexa488-anti-rabbit IgG (Cell Signaling Technologies 4412): IHC (1:800)

Alexa555-anti-rabbit IgG (Cell Signaling Technologies 4413): IHC (1:800)

Cy3-anti-rabbit IgG (Jackson ImmunoResearch 711-165-152): IHC (1:400)

HRP-conjugated anti-rabbit IgG (Sigma-Aldrich A0545): western blotting (1:5000)

Expression plasmids
GFP-Tet1cd (plasmid with ampicillin resistance, CAG promoter):
(N-terminus) eGFP – TEV cleavage site – Tet1cd (C-terminus); Tet1cd is the shortened version of murine Tet1 (UniprotKB – Q3URK3) starting from amino acid 1367 of the original sequence

GFP-Tet3cd (plasmid with ampicillin resistance, CAG promoter):
(N-terminus) eGFP – TEV cleavage site – Tet3cd (C-terminus); Tet3cd is the shortened version of murine Tet3 (UniprotKB – Q8BG87-1) starting from amino acid 696 of the original sequence

Glud1-FLAG (plasmid with ampicillin resistance, CMV promoter):
(N-terminus) Glud1-FLAG (C-terminus); murine Glud1 (UniprotKB – P26443) starting from amino acid 1 of the original sequence

GFP-Tet3 (plasmid with ampicillin resistence, CAG promoter):
(N-terminus) eGFP – TEV cleavage site – Tet3; murine Tet3 (UniprotKB – Q8BG87-4) starting from amino acid 1 of the original sequence

Cell culture HEK293T cells

HEK293T cells (ATCC) were cultivated at 37°C in water saturated, CO_2-enriched (5%) atmosphere. DMEM (Sigma-Aldrich D6546) or RPMI 1640 (Sigma-Aldrich R0883), containing 10% (v/v) fetal bovine serum (FBS) (Invitrogen 10500-064), 1% (v/v) L-alanyl-L-glutamine (Sigma-Aldrich G8541) and 1% (v/v) penicillin-streptomycin (Sigma-Aldrich P0781), were used as growing medium. When reaching a confluence of 70% to 80%, the cells were routinely passaged. Cells were tested once in two months for Mycoplasma contamination using Mycoplasma Detection Kit (JenaBioscience PP-401L).

Transfection of HEK293T cells

Transfection of HEK293T cells for high protein expression:
The transfection was performed in four p150 petri dishes (Sarstedt 83.3903.300). Five to six million cells per p150 were seeded in 25 mL of medium. After seeding, the cells were incubated for 24 h to reach a confluence of 40% to 80%. 10 µg of expression plasmid DNA and 30 µL of the transfection reagent jetPRIME (Polyplus Transfection VWR 114-15) were used as described by the manufacturer. 4 h and 28 h after transfection the medium was changed and sodium butyrate (final conc. 4 mM) was added. 48 h after transfection, the cells were harvested by trypsinization and immediately used for protein extract preparation.

Transfection of HEK293T cells for immunofluorescence staining:
The transfection was performed in 15 µ-slide 8 well plates (ibidi 80826). 3×10^4 cells were seeded per well in 200 µL of medium. 24 h after seeding, the cells were transfected using 150 ng of DNA per expression plasmid, 0.5 µL of jetPRIME and 15 µL of jetPRIME buffer. 24 h after transfection, the cells were washed once with PBS supplemented with $MgCl_2$ and $CaCl_2$ (Dulbecco's phosphate buffered saline, Sigma-Aldrich D8662) and immunofluorescence staining were performed.

Protein extract preparation

Protein extract preparation from transfected HEK293T cells:
The harvested HEK293T cells and the resulting lysate were kept on ice during the preparation all time. Per eight million cells, 1 mL of RIPA buffer (10 mM Tris (pH = 7.5), 150 mM NaCl, 0.5 mM EDTA, 0.1% (w/v) SDS, 1 % (v/v) Triton X-100, 1% (w/v) deoxycholate), supplemented with 2.5 mM $MgCl_2$, 100 U/mL benzonase (Merck Millipore 70746-3) and 1 x protease inhibitor cocktail (PIC, Roche 05056489001) on the day of preparation was used for the lysis. Cells were resuspended in RIPA buffer and lysed for one hour at 4 °C on a tube rotator. Afterwards, the lysate was centrifuged (10000 x g, 15 min, 4 °C) and the supernatant containing the proteins were transferred into a new tube. To enrich GFP-tagged proteins, the lysate was immediately incubated with GFP Nano-Traps either on agarose beads (Chromotek gta-20, for *in vitro* activity assay) or on magnetic agarose beads (Chromotek gtma-20).

Protein extract preparations from murine brain:
Protein extracts from whole murine brain (Mus musculus, C57-BL6/J wild type, both genders, 110 days old; Charles River, Sulzfeld, Germany), including separation into the organelle, the cytosolic and the nuclear fraction, were prepared according to a previously published protocol.[41] The nuclear extract

was then treated with 25 U/mL benzonase for 30 min on ice and subsequently centrifuged (21000 x g, 15 min, 4 °C). The supernatant containing the nuclear lysate was transferred to a new tube. A Bradford protein assay (Bio-Rad 5000006) was performed according to the manufacturer´s instructions to determine the protein concentration. To check whether the nuclear fraction showed specific enrichment for nuclear proteins in comparison to the combined organelle/cytosolic fraction and was not heavily contaminated with proteins from other compartments, a western blot against histone H3 (nuclear) and cytochrome c (mitochondrial) was performed. The western blot confirmed the specific enrichment of nuclear proteins in the nuclear fraction (Extended Data Figure 5C).

Protein extracts from mouse embryonic stem cells (mESC):
Protein extracts from mESC and mESC with TET TKO that were primed for 96 h according to a previously published protocol[42] were prepared as previously described.[43]

***In vitro* activity assay**

For the *in vitro* activity assay on GFP Nano-Trap on agarose beads, GFP-Tet1cd or GFP-Tet3cd bound to the trap was used. The proteins were ectopically expressed in HEK293T (per *in vitro* assay 1 x P150 culture dishes for GFP-Tet1cd, 3 x P150 culture dishes for GFP-Tet3cd) and the resulting protein extracts were incubated for 1 h at 4 °C on a tube rotator with 140 µL of GFP Nano-Traps on agarose beads per extract. Afterwards, the beads were washed twice with coIP wash buffer 1 (10 mM HEPES pH = 7.5, 150 mM NaCl, 0.5 mM EDTA), twice with coIP wash buffer 2 (10 mM HEPES pH = 7.5, 1 M NaCl, 0.5 mM EDTA) and twice with coIP wash buffer 1. The fluorescence of the GFP-tagged proteins bound to the GFP Nano-Trap was checked on a Tecan Plate Reader (Tecan GENios Pro, fluorescence intensity excitation 400 nm, emission 535 nm) and per reaction, the amount of Nano-Trap was adjusted that the fluorescence signal was 25000 per reaction. Per assay, seven reactions in TET 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) with different concentrations of αKG (0 µM; 10 µM; 50 µM; 100 µM; 250 µM; 500 µM; 1000 µM) and 4 µM of DNA oligonucleotide (5'-TTTTG[5mdC]GGTTG-3') were set up in 50 µL reaction volume/sample. The samples were incubated at 35 °C for 4 h under shaking. For Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) measurements, 1 µL of the reaction supernatant was desalted on a 0.025 µm ø VSWP filter membrane against ddH_2O for at least one hour, co-crystallized in a 3-hydroxypicolinic acid matrix (HPA) and mass spectra were recorded on a Bruker Autoflex II in a m/z range of 1500 to 6000. Spectra were normalized to the recorded maximum intensity. For each condition (Tet1cd or Tet3cd, defined αKG concentration) three independent experiments were set up and MALDI-TOF spectra recorded. Afterwards the mean was calculated. The area under the curve (AUC) was calculated using GraphPad Prism (version 8.0.0 or higher) for 5mdC in the m/z range 3373.5 to 3389.0 and 3397.5 to 3404.5 (Na^+ peak), 5hmdC and 5fdC in the m/z range 3389.0 to 3397.5 and 5cadC in the m/z range 3404.5 to 3410.

Tet3-enriched coIP

20 µl of magnetic anti-GFP beads (Chromotek gtma-20) were washed three times with GFP wash buffer (10 mM Tris pH 7.5, 150 mM NaCl, 0.5 mM EDTA) and then incubated for 15 min on ice with nuclear extract of GFP-Tet3 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-Tet3 loaded beads were then washed twice with coIP wash buffer 1, twice with coIP wash buffer 2 and twice 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). The GFP-Tet3 beads were subsequently incubated with 200 µg of nuclear brain extract for 15 min on ice. Following, they were washed twice with GFP wash buffer (10 mM Tris pH = 7.5, 150 mM NaCl, 0.5 mM EDTA). To elute the bound proteins, 50 µl of 200 mM glycine pH 2.5 were added and the solution was vortexed for 30 s. To gain more yield, the elution step was repeated. For the negative control, the same procedure was followed using GFP instead of GFP-Tet3.

Tet3 endogenous coIP

CoIP 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 and 10 µL of Dynabeads Protein G (ThermoFisher 10003D) were used per replicate; for the analysis by western blot the doubled amount of nuclear brain extract, antibodies and Dynabeads Protein G were used. Anti-goat IgG was used for the negative control.

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 GFP wash buffer. Afterwards, the Dynabeads were added to the lysate, PBS was added to a final volume of 500 µL and the suspension was incubated for 1 h at 4°C on a tube rotator. After incubation, the beads were washed three times with coIP wash buffer 1. Last, proteins were eluted with 30 µL of 1% (v/v) formic acid for 15 min at room temperature (MS 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) bromophenol blue) for 10 min at 70 °C (subsequent western blotting).

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 V5113) at 37°C for 16 h. Afterwards, they were incubated for 5 min at 100°C and subsequently 1 mM phenylmethylsulphonyl fluoride was added. StageTips were utilized to purify the samples for MS.[44]

The samples were analysed 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; 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 60 000. 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. 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.

LFQ data processing

The MaxQuant software (version 1.5.0.25) was used for LFQ. Quantification was performed with four biological replicates for Tet3-enriched coIP. GFP alone (four biological replicates) served here as control. Three biological replicates were used for the Tet3 endogenous coIP and the coIP (three biological replicates) 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 PRIDE[45] partner repository with the dataset identifier PXD004518.

LFQ data was analysed 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).

Western blotting

Samples were loaded on a 4-15% precast polyacrylamide gel (Bio-Rad) and MagicMark XP Standard (ThermoFisher LC5603) and Blue Prestained Protein Standard, Broad Range (11-190 kDa) (New England Biolabs P7706S) or Color-coded Prestained Protein Marker, Broad Range (11-250 kDa) (New England Biolabs 14208) were used as protein standards. The gel was run at constant 150 V for 60 min in SDS running buffer (25 mM Tris, 192 mM glycine, 0.1% (w/v) SDS). For blotting, we used a PVDF blotting membrane (GE Healthcare Amersham Hybond P0.45 PVDG membrane 10600023) 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 1 min in methanol, washed with water and equilibrated for additional 1 min in Towbin blotting buffer; the Whatman gel blotting papers (Sigma-Aldrich WHA 10426981) were equilibrated for 15 min in Towbin buffer and the precast gel was equilibrated for 5 min in Towbin buffer after the run. Western blotting (tank (wet) electro transfer) was performed at 4°C for 10 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 pH = 7.5, 150 mM NaCl, 0.1% (v/v) Tween-20). The primary antibodies were diluted in 5 mL of 5% (w/v) milk powder in TBS-T. The blocking suspension was discarded, and the diluted primary antibodies were added for 12 h at 4°C and shaking. After incubation, the primary antibodies were discarded, and the membrane was washed three times five minutes with TBS-T. HRP-conjugated secondary antibodies were diluted in 5% (w/v) milk powder in TBS-T and added for 1 h at room temperature under shaking. In case of Tet3-endogenous coIP samples, TidyBlot HRP conjugated Western blot detection reagent (Biorad STAR209P) was used instead of HRP-conjugated secondary antibodies if the primary antibodies against the potential interactor were produced in rabbit to avoid detection of the antibody fragments from the coIP. Afterwards, the membrane was washed two times with TBS-T and one time with TBS (TBS-T without Tween-20) before SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific 34077) was used for imaging.

IHC

All steps were performed in in a humidity chamber and at room temperature when not otherwise specified. 12 µm thick coronar cryo-sections of snap-frozen adult mouse brain and liver were incubated (if applicable) with MitoTracker Deep Red FM (Invitrogen M22426) before fixation. Cryo-sections were fixed on slides using 4% paraformaldehyde (4% PFA, ThermoScientific 28908) in 0.1 M phosphate buffered solution, pH 7.4 (0.1 M PB). After three times washing with 0.1 M PB, the slices were permeabilized and blocked for 30 min using 0.3% (v/v) Triton X-100 and 5% (v/v) blocking reagent CB (Chemiblocker, Merck Millipore 2170) in 0.1 M PB. The primary antibodies were diluted in 0.1 M PB, containing 5% (v/v) CB and 0.3% (v/v) Triton X-100 and applied for 12 h at 4°C. For the negative controls, no primary antibodies were added. After incubation, slices were washed three times with 0.1 M PB containing 2% (v/v) CB. For secondary detection, the fluorescent labelled secondary antibodies were diluted in 0.1 M PB, containing 3% (v/v) CB and applied the antibodies for 1 h in the dark, followed by three times washing with 0.1 M PB. Cell nuclei were stained with Hoechst 33342 (5 µg/mL), which was applied for 10 min in the dark, followed by one washing step with 0.1 M PB. After mounting (Mountant Permafluor ThermoScientific TA-030-FM), the slices were analysed using a Leica SP8 confocal laser scanning microscope (Leica, Wetzlar).

ICC

ICC experiments were performed as IHC experiments, but instead of 0.1 M PB, PBS supplemented with $MgCl_2$ and $CaCl_2$ (PBS^+) was used. When ICC of transfected HEK293T cells was analysed, Image J (version 2.0.0) was used.

PLA

For the PLA experiments, Duolink InSitu Orange Starter Kit (Sigma-Aldrich DUO92102 and DUO92106) were used. Until the application of primary antibodies, the procedure followed the IHC and ICC procedures described above. PLA probes without primary antibodies served as a negative control. After incubation with primary antibodies, slices/cells were washed once with 0.1 M PB/PBS^+. The following steps were carried out as described by the manufacturer with modifications after the last washing step as described in the manual with 0.01 wash buffer B. Before staining of cell nuclei and mounting, slices/cells were washed with 0.1 M PB/PBS^+. Where applicable, Alexa488-NeuN antibody was applied for 2 h in 2% (v/v) CB in 0.1M PB and slices were washed afterwards three times with 0.1 M PB. However, Alexa488-NeuN application impaired the intensity of the PLA signal. Cell nuclei were stained with Hoechst 33342 (5 µg/mL), which was applied for 10 min in the dark, followed by one washing step with 0.1 M PB/PBS^+. After mounting (Mountant Permafluor Thermo Scientific TA-030-FM), the slices were analysed using a Leica SP8 confocal laser scanning microscope (Leica, Wetzlar).

Isolation of genomic DNA (gDNA), RNA and triple-quadrupole (QQQ) MS/MS

Isolation of gDNA, RNA and QQQ-MS/MS experiments were performed according to a previously published protocol.[46]

Synthesis of the Glud1 inhibitor R162

Synthesis of the inhibitor R162 was done according to the literature.[47, 48] Prior to use for the cell culture, the compound was purified via preparative HPLC (Nucleosil VP 250/10 C18 column Macherey Nagel, 100% MeCN for 10 minutes).

Depolarization of hippocampal neurons followed by UHPLC-QQQ-MS analysis

Animals:

4 – 5 weeks old C57-BL6/J (Charles River, Sulzfeld, Germany) wildtype mice of both genders were used.

Hippocampal slices:

For the experiment with R162, hippocampi of six mice were used. For the experiment with UK5099 (Sigma-Aldrich PZ0160-5MG), hippocampi of six mice were used. For the experiment with 2DG (Santa Cruz Biotechnology sc-202010) and pyruvate (Sigma-Aldrich S8636-100ML), hippocampi of four mice were used.

Acute transverse hippocampal slices (400 µm thick) were prepared as described previously.[49, 50] 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% (v/v) 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% (v/v) DMSO) and oxygenated 25 mM KCl solution supplemented with 20 µM inhibitor R162 or 40 µM UK5099 or 10 mM 2DG and 2 mM pyruvate. 6 – 10 slices were pooled for each replicate. After 6 h incubation time, the slices were transferred into reaction tubes, snap frozen in liquid nitrogen and stored at -80°C until use. We excluded mice from further analysis, when *Npas4* expression was not upregulated in the depolarized hippocampal slices compared to the unstimulated control. This correlated with unchanged 5hmdC levels.

RT-qPCR

For cDNA synthesis, the RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific K1621) was used according to the manufacturer's instructions. 190 ng of total RNA was used for cDNA synthesis. Real-time quantitative PCR (RT-qPCR) was performed on StepOnePlus Real-Time PCR system (Applied Biosystems Thermo Fisher Scientific) using PowerUp™ SYBR™ Green qPCR Master Mix (Applied Biosystems Thermo Fisher Scientific A25742). C_T values of each sample were determined in technical duplicates by the StepOne™ Software (Applied Biosystems Thermo Fisher Scientific) using the fast cycle protocol. The relative expression levels of target genes were then quantified from seven biological replicates (n = 7) according to *Pfaffl et al*.[51, 52]

The following primer sets were used:

Alas (mouse):	5′ TCG CCG ATG CCC ATT CTT ATC 3′	(forward)
	5′ GGC CCC AAC TTC CAT CAT CT 3′	(reverse)
Npas4 (mouse):[53]	5′ CTG CAT CTA CAC TCG CAA GG 3′	(forward)
	5′ GCC ACA ATG TCT TCA AGC TCT 3′	(reverse)
Bdnf total (mouse):[54]	5′ GCC TTT GGA GCC TCC TCT 3′	(forward)
	5′ CTG TCA CAC ACG CTC AGC TC 3′	(reverse)

Metabolomics

Pre-weighed frozen hippocampal tissues from five mice (n = 5) per condition (physiological solution, 25 mM KCl, 25 mM KCl + R162) were homogenized using a bullet blender and five 1.4 mm stainless steel beads in a mixture of 1.05 ml ice-cold methanol/chloroform 1:2 (v/v). Following sonication of supernatants for 1 min, an additional 0.7 ml mixture of chloroform/LC grade water, 1:1 (v/v) were added, samples were vortexed for 20 sec and incubated on wet ice for 30 min. Extracts were then centrifuged for 5 min at 16000 rcf and 4 °C and 0.6 ml from the upper aqueous phase were collected and dried under nitrogen gas. For NMR measurements, the dried extracts were dissolved in 0.25 ml 0.15 M phosphate buffer (pH 7.4) in deuterated water containing 0.04 mM sodium trimethylsilylpropionate-d_4 (TSP) as internal standard. One 1D proton NMR was collected for each sample in a 14.1 T Bruker Avance II NMR under standardized conditions as described in *Kostidis et al., 2017*[55]. The NMR buffer was degassed, and NMR tubes were flushed with nitrogen to prevent oxidation. Metabolites were quantified in Chenomx NMR suit 8.4 (Chenomx Inc, Canada) and concentrations were corrected according to the tissue mass per sample.

Statistical analysis

Statistical analysis, except for LFQ data processing as described above, was performed using GraphPad Prism (version 8.0.0 or higher) and only biological replicates were considered.

Data Figure 3A

5hmdC level: for each mouse (n = 6) the prepared hippocampal slices were distributed to three different conditions – (A) physiological buffer, (B) 25 mM KCl, (C) 25 mM KCl + 20 µM R162.
RM one-way ANOVA with Geisser-Greenhouse correction p-value 0.0016; matching p-value <0.0001.
Tukey's multiple comparisons test with individual variances computed for each comparison: A/B p-value 0.0043, A/C p-value 0.7424, B/C p-value 0.0140.

Npas4 expression: n = 6, ratio paired t-test p-value (two-tailed) 0.0248; pairing correlation coefficient 0.9500, p-value 0.0018.

Bdnf expression: n = 3, ratio paired t-test p-value (two-tailed) 0.0042; pairing correlation coefficient 0.9212, p-value 0.1272.

αKG/succinate ratio: n= 5, ordinary one-way ANOVA p-value 0.0368. Tukey's multiple comparisons test A/B p-value 0.1215, A/C p-value 0.7684, B/C p-value 0.0364.

Data Figure 3B

5hmdC level: for each mouse (n = 6) the prepared hippocampal slices were distributed to three different conditions – (A) physiological buffer, (B) 25 mM KCl, (C) 25 mM KCl + 40 µM UK5099.
RM one-way ANOVA with Geisser-Greenhouse correction p-value 0.0034; matching p-value <0.0001.
Tukey's multiple comparisons test with individual variances computed for each comparison: A/B p-value 0.0410, A/C p-value 0.0123, B/C p-value 0.0130.

Data Figure 3C

5hmdC level: for each mouse (n = 4) the prepared hippocampal slices were distributed to three different conditions – (A) physiological buffer, (B) 25 mM KCl, (C) 25 mM KCl + 10 mM 2DG/2 mM pyruvate.

RM one-way ANOVA with Geisser-Greenhouse correction p-value 0.0957; matching p-value 0.4735.
Tukey's multiple comparisons test with individual variances computed for each comparison: A/B p-value 0.1147, A/C p-value 0.9957, B/C p-value 0.0233.

Extended Data Figure 4B
5mdC level: for each mouse (n = 6) the prepared hippocampal slices were distributed to three different conditions – (A) physiological buffer, (B) 25 mM KCl, (C) 25 mM KCl + 20 µM R162.
RM one-way ANOVA with Geisser-Greenhouse correction p-value 0.4896; matching p-value 0.0023.
Tukey's multiple comparisons test with individual variances computed for each comparison: A/B p-value 0.5547, A/C p-value 0.4637, B/C p-value 0.9964.

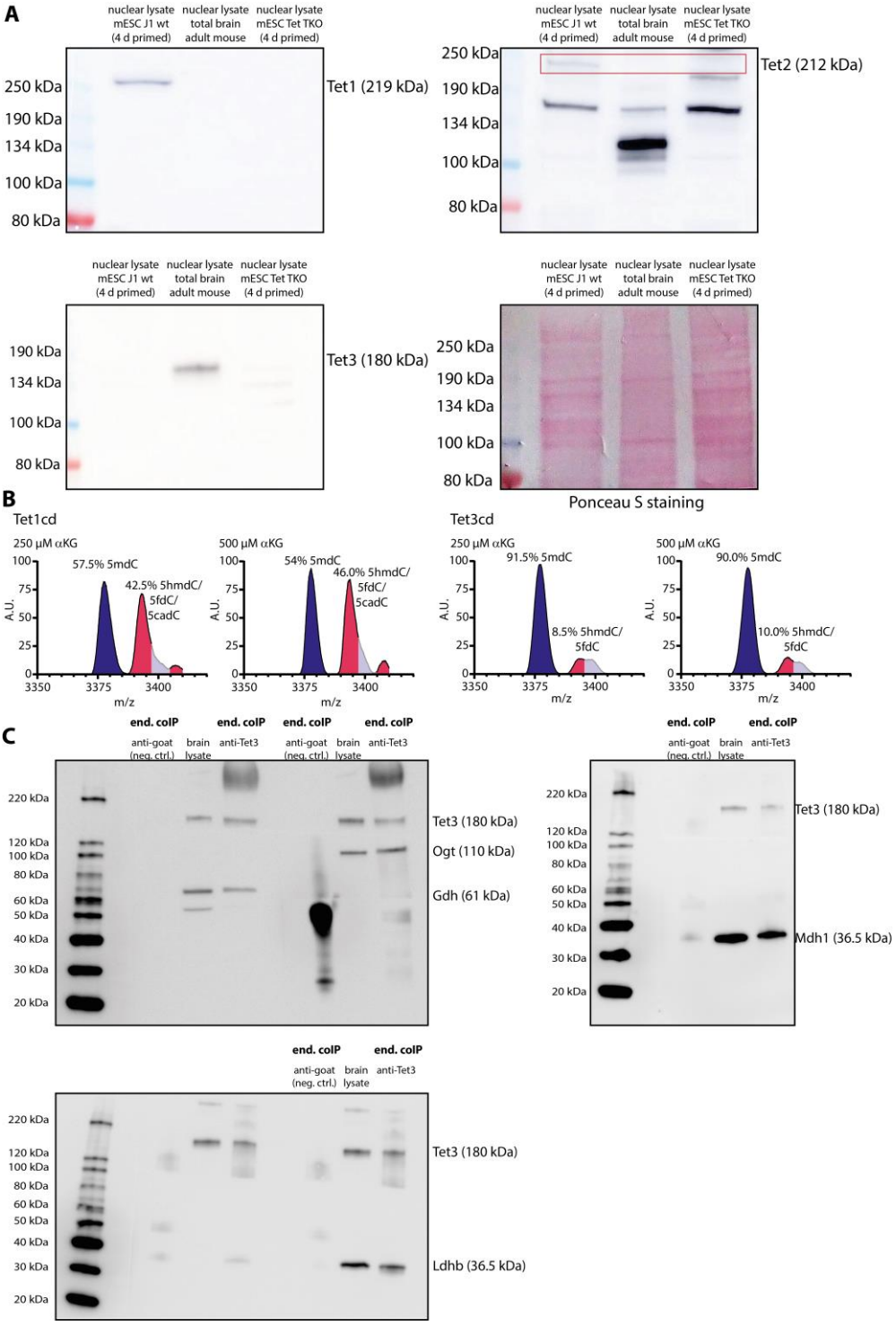
Extended Data Figure 4C
5mdC level: for each mouse (n = 6) the prepared hippocampal slices were distributed to three different conditions – (A) physiological buffer, (B) 25 mM KCl, (C) 25 mM KCl + 40 µM UK5099.
RM one-way ANOVA with Geisser-Greenhouse correction p-value 0.2164; matching p-value <0.0001.
Tukey's multiple comparisons test with individual variances computed for each comparison: A/B p-value 0.6278, A/C p-value 0.2213, B/C p-value 0.5375.

Extended Data Figure 4D
5mdC level: for each mouse (n = 4) the prepared hippocampal slices were distributed to three different conditions – (A) physiological buffer, (B) 25 mM KCl, (C) 25 mM KCl + 10 mM 2DG/2 mM pyruvate.
RM one-way ANOVA with Geisser-Greenhouse correction p-value 0.4941; matching p-value 0.3876.
Tukey's multiple comparisons test with individual variances computed for each comparison: A/B p-value 0.7322, A/C p-value 0.5803, B/C p-value 0.9246.

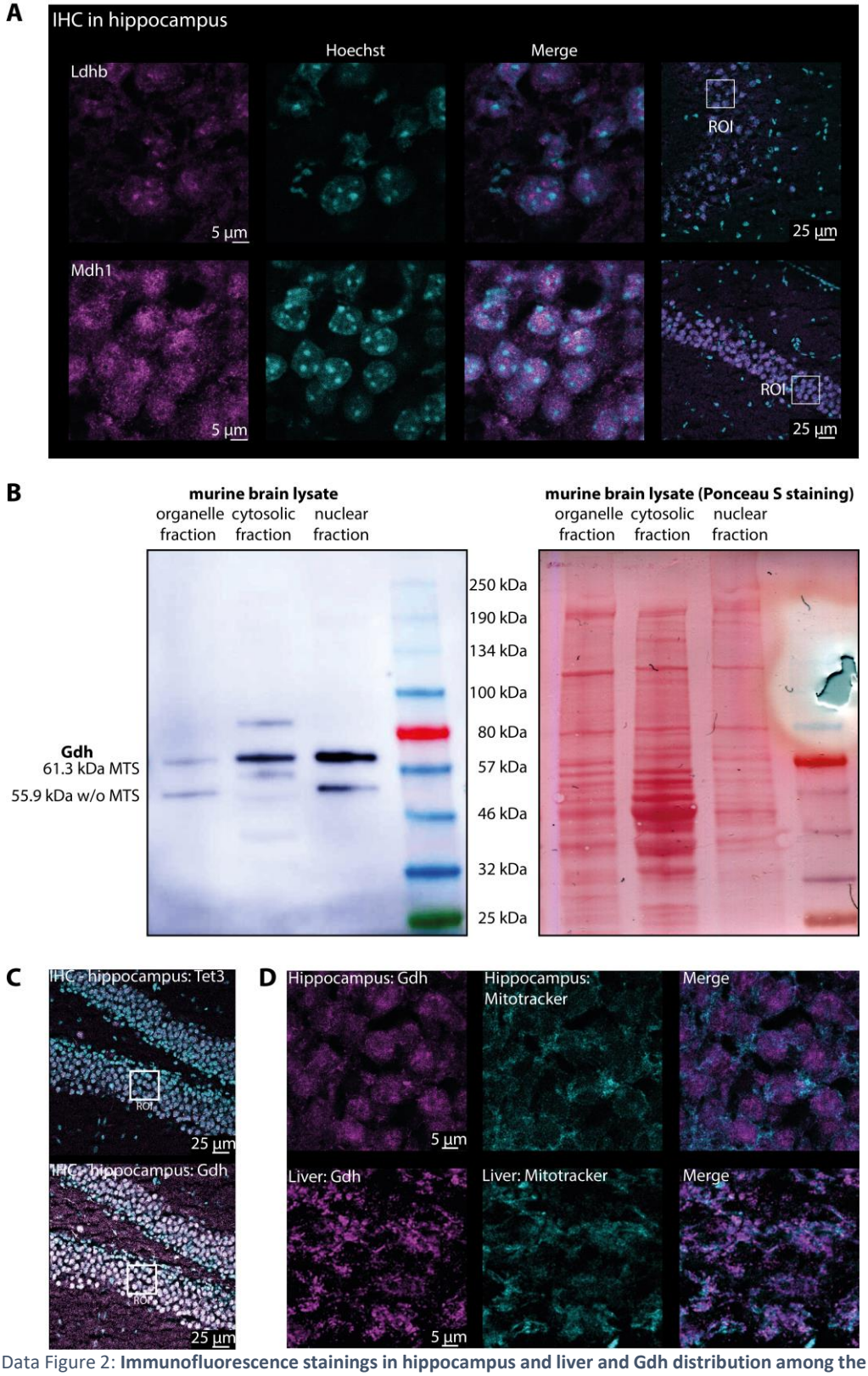
Contributions:
I designed experiments, performed and analysed Tet3 endogenous coIPs followed by western blotting, performed and analysed western blots, IHC/ICC experiments, PLAs, UHPLC-QQQ-MS experiments, *in vitro* activity assays, helped with depolarization experiments of hippocampal slices, interpreted data and wrote the manuscript. *Dilara Özdemir*[1] performed and analysed IHC/ICC experiments and PLAs and helped with western blots and *in vitro* activity assays. *Dr. Andrea Glück*[1] designed proteomics experiments, developed the GFP-Tet3 enriched coIP workflow, performed and analysed GFP-Tet3 enriched and Tet3 endogenous coIPs followed by LFQ MS/MS and interpreted proteomics data. *Constanze Scheel*[2] performed depolarization experiments of acute hippocampal slices, dissected mouse brain and liver and prepared cryosections thereof, analysed IHC/ICC experiments and performed and analysed RT-qPCR experiments. *Dr. Anna Geserich*[2] helped performing and analysing IHC/ICC experiments and PLAs. *Dr. Katharina Iwan*[1] helped performing and analysing UHPLC-QQQ-MS experiments. *Dr. René Rahimoff*[1] synthesized R162. *Sabine Oganesian*[1] helped with *in vitro* activity assays. *Dr. Sarantos Kostidis*[3] and *Dr. Martin Giera*[3] performed and analysed metabolomics experiments. *Dr. Markus Müller*[1] and *Dr. Fabio Spada*[1] helped designing experiments and interpreted data. *Prof. Dr. Stylianos Michalakis*[2] designed and supervised experiments, performed depolarization experiments of acute hippocampal slices and interpreted data. *Prof. Dr. Thomas Carell*[1] designed and supervised the whole study and interpreted data.

[1] Carell Group, Department of Chemistry, LMU München
[2] Michalakis Group, Department of Pharmacy, LMU München
[3] Giera Group, Leids Universitair Medisch Centrum

Extended Data Figures



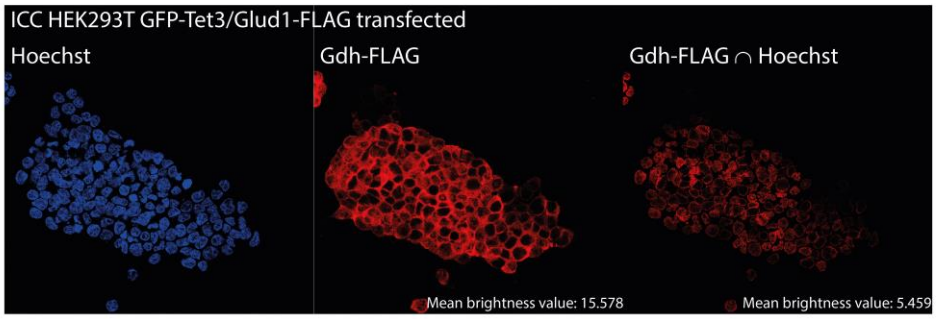
Extended Data Figure 1: **TET protein levels in different cellular systems, Tet1cd and Tet3cd kinetics and full western blots of Tet3-endogenous coIP.** (A) TET protein levels in nuclear lysate of 4 d primed mESC compared to nuclear brain lysate. Nuclear lysate of mESC TET TKO served as a control. Ponceau S staining indicates protein load. (B) MALDI-TOF MS spectra (n=3) of a single stranded DNA oligonucleotide containing 5mdC (3381.6 m/z (blue) and 3403.6 m/z (grey, Na^+ peak of 5mdC)) that is oxidized *in vitro* to 5hmdC (3397.6 m/z), 5fdC (3395.6 m/z) and 5cadC (3411.6 m/z) (all magenta) by Tet1cd or Tet3cd in presence of different concentrations of αKG. (C) Complete western blot images of Tet3-endogenous coIP.
5cadC = 5'-carboxy-2'-deoxycytidine, 5fdC = 5'-fomyl-dC, 5hmdC = 5'-hydroxymethyl-dC, 5mdC = 5'-methyl-dC, αKG = α-ketoglutarate, A.U. = arbitrary unit, cd = catalytic domain, coIP = co-immunoprecipitation, ctrl. = control, end. = endogenous, FDR = false discovery rate, Gdh = glutamate dehydrogenase 1, Ldhb = lactate dehydrogenase b, MALDI-TOF-MS = matrix assisted laser desorption ionization-time of flight mass spectrometry; Mdh1 = malate dehydrogenase 1, neg. = negative, Ogt = O-linked β-N-acetylglucosamine transferase, Tet = ten-eleven translocation enzyme.

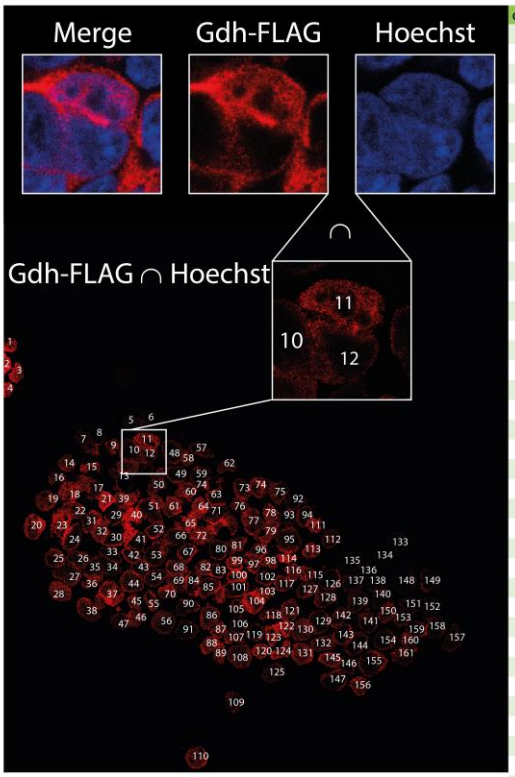


Extended Data Figure 2: **Immunofluorescence stainings in hippocampus and liver and Gdh distribution among the different cellular compartments.** (A) Localization of Ldhb and Mdh1 in the murine hippocampus. Images show Z-stacks. ROI shows area of close-up image. (B) Complete western blot against Gdh in different cellular compartments of total murine brain. Ponceau S staining shows total protein load. (C) Overview of the hippocampal regions for IHC of Tet3 and Gdh. ROI marks the region that was chosen for close-up images in Figure 2B/Extended Data Figure 2D. (D) IHC of Figure 2B showing the corresponding mitotracker signal.

Gdh = glutamate dehydrogenase 1, IHC = immunohistochemistry, Ldhb = lactate dehydrogenase b, Mdh1 = malate dehydrogenase 1, MTS = mitochondrial targeting sequence, ROI = region of interest, Tet3 = ten-eleven translocation enzyme 3.

A





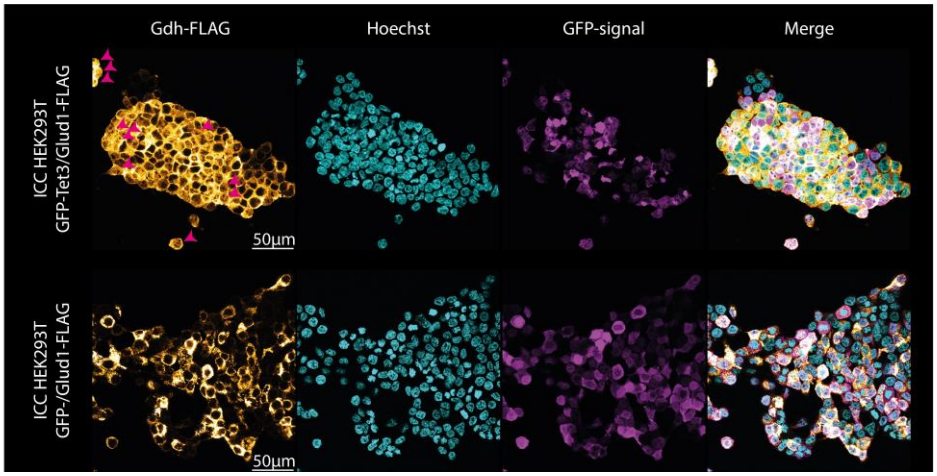
Cell Nr.	Mean	Mean>30
1	57.818	1
2	95.768	2
3	66.690	3
4	55.741	4
5	7.710	
6	3.778	
7	10.863	
8	3.541	
9	22.918	
10	17.565	
11	48.705	5
12	17.385	
13	26.781	
14	18.992	
15	23.450	
16	14.843	
17	27.751	
18	58.579	6
19	33.141	7
20	44.983	8
21	83.030	9
22	36.572	10
23	69.166	11
24	20.393	
25	27.052	
26	24.704	
27	23.746	
28	38.750	12
29	11.673	
30	21.636	
31	26.483	
32	48.907	13
33	14.375	
34	14.103	
35	18.382	
36	28.604	
37	65.294	14
38	33.803	15
39	56.326	16
40	104.994	17
41	21.119	

Cell Nr.	Mean	Mean>30
42	24.000	
43	15.229	
44	28.252	
45	27.240	
46	39.609	18
47	16.941	
48	11.649	
49	14.000	
50	22.420	
51	27.195	
52	33.724	19
53	16.925	
54	18.873	
55	33.556	20
56	19.893	
57	15.369	
58	36.316	21
59	14.018	
60	42.625	22
61	51.190	23
62	19.611	
63	14.847	
64	75.033	24
65	61.961	25
66	12.094	
67	20.128	
68	30.774	26
69	31.852	27
70	40.417	28
71	18.788	
72	57.331	29
73	27.558	
74	44.340	30
75	26.690	
76	21.180	
77	34.593	31
78	47.927	32
79	41.163	33
80	11.630	
81	37.053	34
82	42.173	35

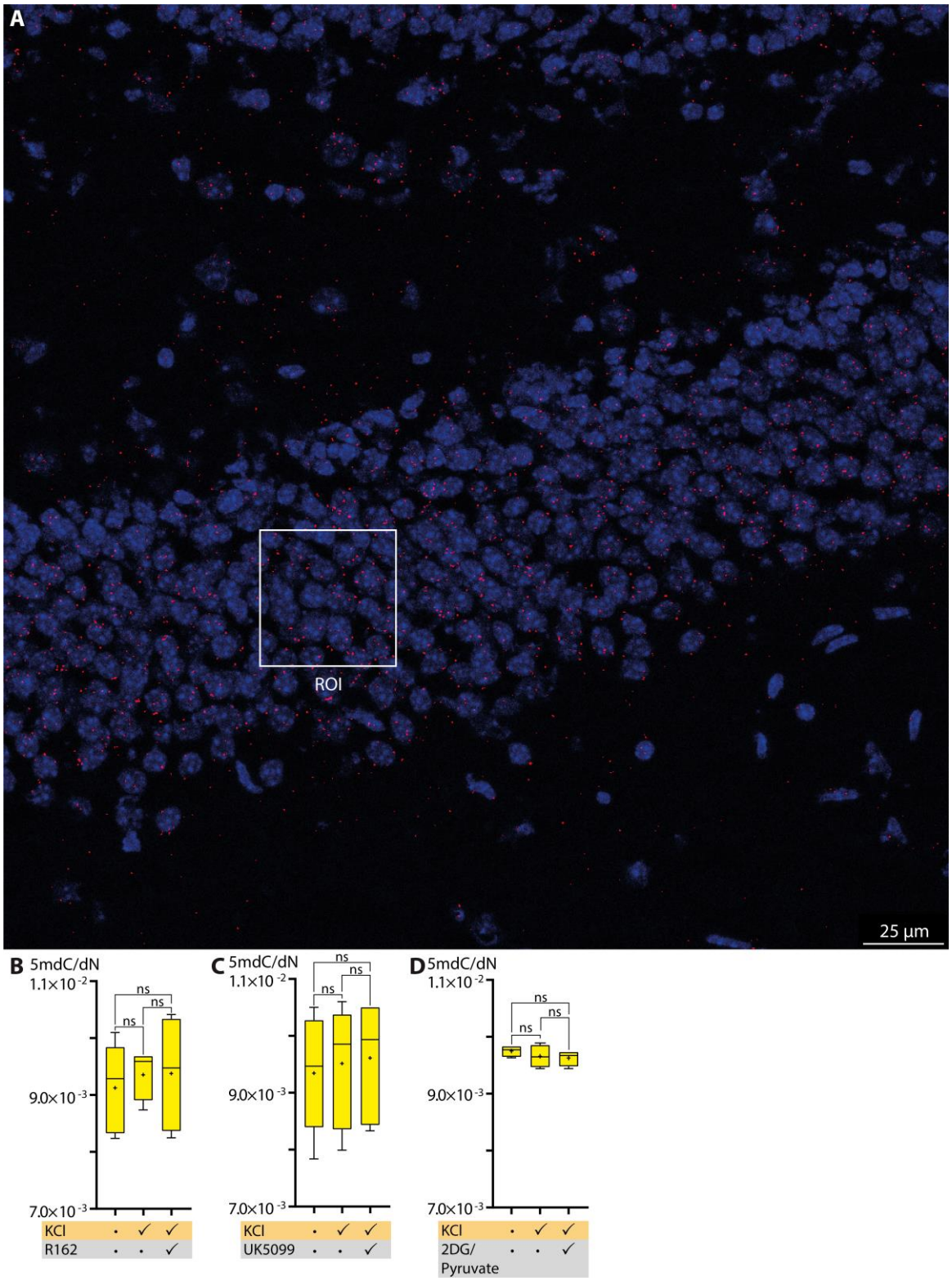
Cell Nr.	Mean	Mean>30
83	27.314	
84	25.225	
85	26.904	
86	26.942	
87	38.653	36
88	42.127	37
89	41.710	38
90	10.120	
91	6.014	
92	7.251	
93	8.906	
94	16.853	
95	16.846	
96	33.373	39
97	43.169	40
98	40.590	41
99	70.788	42
100	57.693	43
101	13.549	
102	17.576	
103	65.114	44
104	67.286	45
105	20.477	
106	12.904	
107	33.258	46
108	25.531	
109	12.871	
110	34.976	47
111	52.907	48
112	20.405	
113	47.895	49
114	88.783	50
115	38.455	51
116	44.215	52
117	29.234	
118	30.418	53
119	11.148	
120	52.574	54
121	35.155	55
122	79.326	56
123	54.708	57

Cell Nr.	Mean	Mean>30
124	44.940	58
125	13.799	
126	8.081	
127	25.457	
128	21.061	
129	15.496	
130	30.709	59
131	43.049	60
132	21.719	
133	3.981	
134	1.697	
135	2.931	
136	5.242	
137	10.433	
138	6.358	
139	13.927	
140	8.548	
141	16.710	
142	22.253	
143	17.672	
144	13.539	
145	39.268	61
146	14.563	
147	14.221	
148	3.000	
149	9.613	
150	24.421	
151	8.707	
152	6.649	
153	14.261	
154	11.018	
155	16.538	
156	31.652	62
157	9.579	
158	9.705	
159	9.939	
160	23.568	
161	19928	

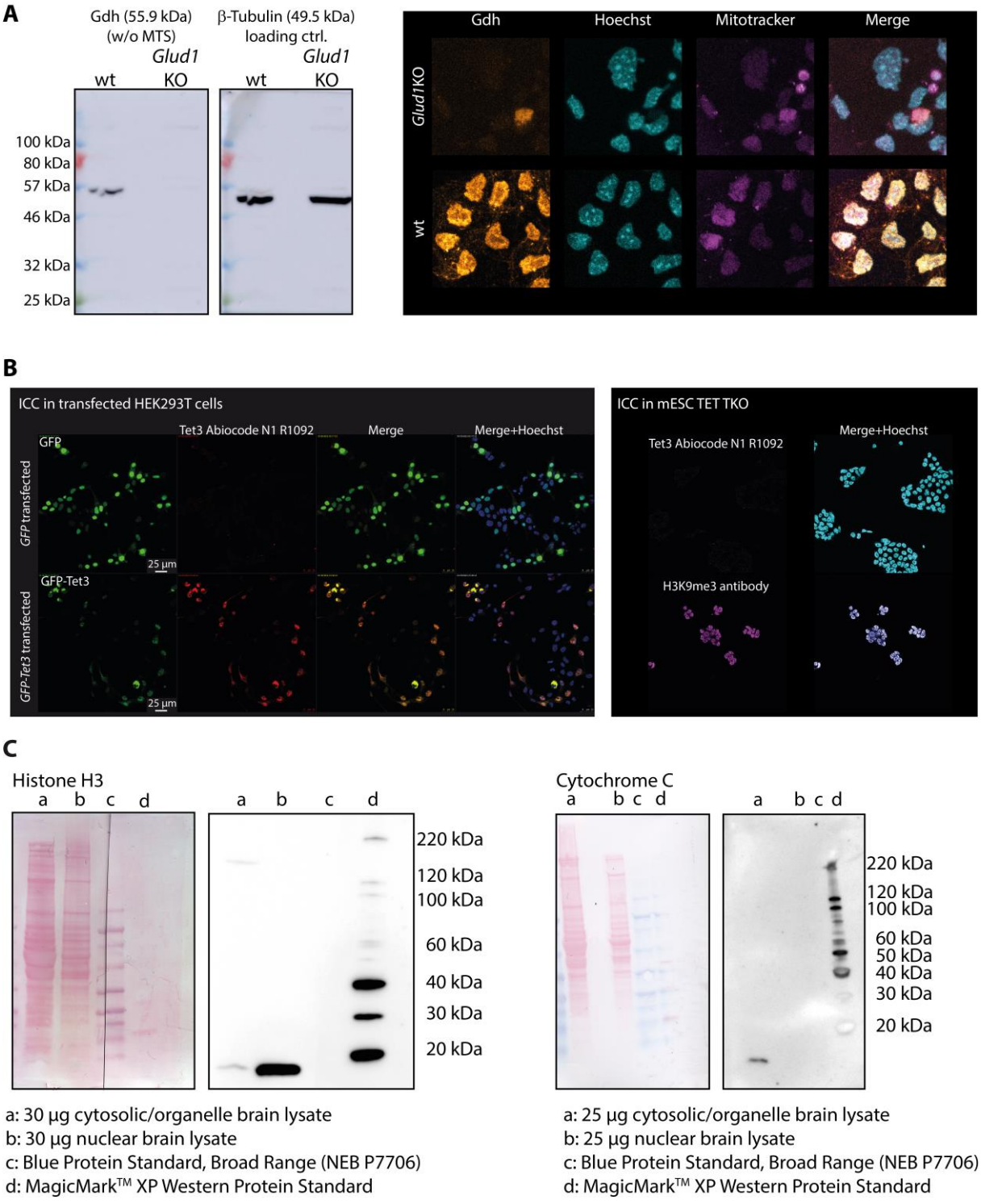
B



Extended Data Figure 3: **Ectopic co-expression of murine Tet3 and Gdh in HEK293T.** (A) ICC of *GFP-Tet3*/*Glud1*-FLAG transfected HEK293T cells. Calculation of the brightness value per cell of the overlap of Gdh-FLAG and Hoechst from the ICC using image J. A mean brightness value above 30 was considered as a nuclear signal. Approximately 40% of the cells show Gdh-FLAG signal in the nucleus, when co-expressed with GFP-Tet3. (B) ICC of GFP-Tet3 and Gdh-FLAG compared to GFP and Gdh-FLAG in *GFP-Tet3*/*Glud1*-FLAG and *GFP*/*Glud1*-FLAG-transfected HEK293T cells.
Gdh (protein)/*Glud1* (gene) = glutamate dehydrogenase 1, GFP = green fluorescent protein, ICC = immunocytochemistry, Tet3 = ten-eleven translocation enzyme 3.



Extended Data Figure 4: **PLA of Gdh and Tet3 in murine hippocampus and 5mdC level of murine acute hippocampal slices after depolarization in absence and presence of various inhibitors.** (A) PLA Tet3/Gdh in murine hippocampus with the ROI that marks the region that was chosen for the close-up image in Figure 2C. PLA signal is shown in red, nuclear staining (Hoechst) is shown in blue. (B) Global 5dmC levels in absence and presence of 20 µM of Gdh inhibitor R162. Tukey boxplot, n = 6 (measured in technical duplicates). (C) Global 5dmC levels in absence and presence of 40 µM of mitochondrial pyruvate carrier inhibitor UK5099. Tukey boxplot, n = 6 (measured in technical triplicates). (D) Global 5hdmC levels after depolarization of hippocampal slices using 25 mM KCl in absence and presence of 10 mM glycolysis inhibitor 2DG/2 mM pyruvate. Tukey boxplot, n = 4 (measured in technical triplicates). ns p-value ≥ 0.05.
2DG = 2'-deoxyglucose, 5mdC = 5'-methyl-2'-deoxycytidine, dN = nucleotide, Gdh = glutamate dehydrogenase 1, ROI = region of interest, Tet3 = ten-eleven translocation enzyme 3.



a: 30 µg cytosolic/organelle brain lysate
b: 30 µg nuclear brain lysate
c: Blue Protein Standard, Broad Range (NEB P7706)
d: MagicMark™ XP Western Protein Standard

a: 25 µg cytosolic/organelle brain lysate
b: 25 µg nuclear brain lysate
c: Blue Protein Standard, Broad Range (NEB P7706)
d: MagicMark™ XP Western Protein Standard

Extended Data Figure 5: **Antibody and subcellular prefractionation validation.** (A) Validation of the used anti-Gdh antibody using haploid *Glud1*$^{-/}$ (mESC). Western blot and ICC confirmed the specifity of the antibody. (B) Validation of the used anti-Tet3 antibody using transfected HEK293T cells and TET TKO mESC. The antibody showed no signal overlap with GFP in GFP-transfected HEK293T cells, but total signal overlap with GFP in *GFP-Tet3* transfected HEK293T cells. TET TKO mESC did not show a signal for Tet3 either when using this antibody. Histone modification H3K9me3 was tested in the same ICC as a control. (C) Western blotting against histone H3 and cytochrome C confirmed that the nuclear brain extract is specifically enriched for nuclear proteins.
GFP = green fluorescent protein, Gdh (protein)/*Glud1* (gene) = glutamate dehydrogenase 1, ICC = immunocytochemistry, mESC = mouse embryonic stem cells, TET = ten eleven translocation enzyme, TKO = triple knockout.

References

1. Cortes-Mendoza, J., Diaz de Leon-Guerrero, S., Pedraza-Alva, G. & Perez-Martinez, L. Shaping synaptic plasticity: the role of activity-mediated epigenetic regulation on gene transcription. *Int. J. Dev. Neurosci.* **31**, 359-369 (2013).
2. Guan, J.S., Xie, H. & Ding, X. The role of epigenetic regulation in learning and memory. *Exp. Neurol.* **268**, 30-36 (2015).
3. Guo, J.U., Su, Y., Zhong, C., Ming, G.L. & Song, H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145**, 423-434 (2011).
4. Yu, H. *et al.* Tet3 regulates synaptic transmission and homeostatic plasticity via DNA oxidation and repair. *Nat. Neurosci.* **18**, 836-843 (2015).
5. Jin, S.G. *et al.* Tet3 Reads 5-Carboxylcytosine through Its CXXC Domain and Is a Potential Guardian against Neurodegeneration. *Cell Rep.* **14**, 493-505 (2016).
6. Tahiliani, M. *et al.* Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science* **324**, 930-935 (2009).
7. Carey, B.W., Finley, L.W., Cross, J.R., Allis, C.D. & Thompson, C.B. Intracellular alpha-ketoglutarate maintains the pluripotency of embryonic stem cells. *Nature* **518**, 413-416 (2015).
8. TeSlaa, T. *et al.* alpha-Ketoglutarate Accelerates the Initial Differentiation of Primed Human Pluripotent Stem Cells. *Cell Metab.* **24**, 485-493 (2016).
9. Beck, D.B. *et al.* Delineation of a Human Mendelian Disorder of the DNA Demethylation Machinery: TET3 Deficiency. *Am. J. Hum. Genet.* **106**, 234-245 (2020).
10. Globisch, D. *et al.* Tissue Distribution of 5-Hydroxymethylcytosine and Search for Active Demethylation Intermediates. *PLoS One* **5**, e15367 (2010).
11. Szwagierczak, A., Bultmann, S., Schmidt, C.S., Spada, F. & Leonhardt, H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.* **38**, e181 (2010).
12. Jin, L. *et al.* Glutamate Dehydrogenase 1 Signals through Antioxidant Glutathione Peroxidase 1 to Regulate Redox Homeostasis and Tumor Growth. *Cancer Cell* **27**, 257-270 (2015).
13. Sun, X. & Lin, Y. Npas4: Linking Neuronal Activity to Memory. *Trends Neurosci.* **39**, 264-275 (2016).
14. Kowiański, P. *et al.* BDNF: A Key Factor with Multipotent Impact on Brain Signaling and Synaptic Plasticity. *Cell. Mol. Neurobiol.* **38**, 579-593 (2018).
15. Guo, J.U. *et al.* Neuronal activity modifies DNA methylation landscape in the adult brain. *Nat. Neurosci.* **14**, 1345-1351 (2011).
16. Morris, M.J. & Monteggia, L.M. Role of DNA methylation and the DNA methyltransferases in learning and memory. *Dialogues Clin. Neurosci.* **16**, 359-371 (2014).
17. Breiling, A. & Lyko, F. Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. *Epigenetics Chromatin* **8**, 24 (2015).
18. Traube, F.R. & Carell, T. The chemistries and consequences of DNA and RNA methylation and demethylation. *RNA Biol.* **14**, 1099-1107 (2017).
19. Wagner, M. *et al.* Age-Dependent Levels of 5-Methyl-, 5-Hydroxymethyl-, and 5-Formylcytosine in Human and Mouse Brain Tissues. *Angew. Chem. Int. Ed. Engl.* **54**, 12511-12514 (2015).
20. Kaelin, W.G., Jr. & McKnight, S.L. Influence of metabolism on epigenetics and disease. *Cell* **153**, 56-69 (2013).
21. Yang, Q. *et al.* AMPK/α-Ketoglutarate Axis Dynamically Mediates DNA Demethylation in the Prdm16 Promoter and Brown Adipogenesis. *Cell Metab.* **24**, 542-554 (2016).
22. Rolfe, D.F. & Brown, G.C. Cellular energy utilization and molecular origin of standard metabolic rate in mammals. *Physiol. Rev.* **77**, 731-758 (1997).
23. Featherstone, D.E. Intercellular glutamate signaling in the nervous system and beyond. *ACS Chem. Neurosci.* **1**, 4-12 (2010).
24. Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129 - 1133 (2010).

25. Laukka, T. *et al.* Fumarate and Succinate Regulate Expression of Hypoxia-inducible Genes via TET Enzymes. *J. Biol. Chem.* **291**, 4256-4265 (2016).
26. Bauer, C. *et al.* Phosphorylation of TET Proteins Is Regulated via O-GlcNAcylation by the O-Linked N-Acetylglucosamine Transferase (OGT). *J. Biol. Chem.* **290**, 4801-4812 (2015).
27. Yudkoff, M., Nelson, D., Daikhin, Y. & Erecińska, M. Tricarboxylic acid cycle in rat brain synaptosomes. Fluxes and interactions with aspartate aminotransferase and malate/aspartate shuttle. *J. Biol. Chem.* **269**, 27414-27420 (1994).
28. Díaz-García, C.M. *et al.* Neuronal Stimulation Triggers Neuronal Glycolysis and Not Lactate Uptake. *Cell Metab.* **26**, 361-374 (2017).
29. Castonguay, Z., Auger, C., Thomas, S.C., Chahma, M.h. & Appanna, V.D. Nuclear lactate dehydrogenase modulates histone modification in human hepatocytes. *Biochem. Biophys. Res. Commun.* **454**, 172-177 (2014).
30. Mastorodemos, V. *et al.* Human GLUD1 and GLUD2 glutamate dehydrogenase localize to mitochondria and endoplasmic reticulum. *Biochem. Cell. Biol.* **87**, 505-516 (2009).
31. Plaitakis, A., Kalef-Ezra, E., Kotzamani, D., Zaganas, I. & Spanaki, C. The Glutamate Dehydrogenase Pathway and Its Roles in Cell and Tissue Biology in Health and Disease. *Biology (Basel)* **6**, 11 (2017).
32. Sexton, T., Schober, H., Fraser, P. & Gasser, S.M. Gene regulation through nuclear organization. *Nat. Struct. Mol. Biol.* **14**, 1049 (2007).
33. Chen, C., Ding, X., Akram, N., Xue, S. & Luo, S.-Z. Fused in Sarcoma: Properties, Self-Assembly and Correlation with Neurodegenerative Diseases. *Molecules* **24**, 1622 (2019).
34. Yang, C. *et al.* Glutamine Oxidation Maintains the TCA Cycle and Cell Survival during Impaired Mitochondrial Pyruvate Transport. *Mol. Cell* **56**, 414-424 (2014).
35. Boukouris, A.E., Zervopoulos, S.D. & Michelakis, E.D. Metabolic Enzymes Moonlighting in the Nucleus: Metabolic Regulation of Gene Transcription. *Trends Biochem. Sci.* **41**, 712-730 (2016).
36. Halestrap, A.P. The mitochondrial pyruvate carrier. Kinetics and specificity for substrates and inhibitors. *Biochem. J.* **148**, 85-96 (1975).
37. Sutendra, G. *et al.* A nuclear pyruvate dehydrogenase complex is important for the generation of acetyl-CoA and histone acetylation. *Cell* **158**, 84-97 (2014).
38. Bahi-Buisson, N. *et al.* Neurological aspects of hyperinsulinism–hyperammonaemia syndrome. *Dev. Med. Child Neurol.* **50**, 945-949 (2008).
39. Kim, A.Y. & Baik, E.J. Glutamate Dehydrogenase as a Neuroprotective Target Against Neurodegeneration. *Neurochem. Res.* **44**, 147-153 (2019).
40. Perera, A. *et al.* TET3 Is Recruited by REST for Context-Specific Hydroxymethylation and Induction of Gene Expression. *Cell Rep.* **11**, 283-294 (2015).
41. Guillemin, I., Becker, M., Ociepka, K., Friauf, E. & Nothwang, H.G. A subcellular prefractionation protocol for minute amounts of mammalian cell cultures and tissue. *Proteomics* **5**, 35-45 (2005).
42. Iwan, K. *et al.* 5-Formylcytosine to cytosine conversion by C–C bond cleavage in vivo. *Nat. Chem. Biol.* **14**, 72 - 78 (2017).
43. Dignam, J.D., Lebovitz, R.M. & Roeder, R.G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11**, 1475-1489 (1983).
44. Kulak, N.A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **11**, 319 - 324 (2014).
45. Vizcaíno, J.A. *et al.* 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **44**, D447-D456 (2016).
46. Traube, F.R. *et al.* Isotope-dilution mass spectrometry for exact quantification of noncanonical DNA nucleosides. *Nat. Protoc.* **14**, 283-312 (2019).
47. Murty, K.V.S.N., Pal, R., Datta, K. & Mal, D. Glucose Promoted Claisen Rearrangement of 1-Allyloxy Anthraquinones. *Syn. Comm.* **24**, 1287-1292 (1994).

48. Sharghi, H. & Aghapour, G. Claisen Rearrangement of Allyloxyanthraquinones with Silver/Potassium Iodide in Acetic Acid as a New and Efficient Reagent. *J. Org. Chem.* **65**, 2813-2815 (2000).
49. Kleppisch, T. *et al.* 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.* **19**, 48-55 (1999).
50. Michalakis, S. *et al.* Altered synaptic plasticity and behavioral abnormalities in CNGA3-deficient mice. *Genes, Brain Behav.* **10**, 137-148 (2011).
51. Pfaffl, M.W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, e45 (2001).
52. Pfaffl, M.W., Horgan, G.W. & Dempfle, L. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res.* **30**, e36 (2002).
53. Ramamoorthi, K. *et al.* Npas4 regulates a transcriptional program in CA3 required for contextual memory formation. *Science* **334**, 1669-1675 (2011).
54. Ma, D.K. *et al.* Neuronal activity-induced Gadd45b promotes epigenetic DNA demethylation and adult neurogenesis. *Science* **323**, 1074-1077 (2009).
55. Kostidis, S., Addie, R.D., Morreau, H., Mayboroda, O.A. & Giera, M. Quantitative NMR analysis of intra- and extracellular metabolism of mammalian cells: A tutorial. *Anal. Chim. Acta* **980**, 1-24 (2017).

End of the manuscript draft

Influence of metabolites on TET activity

Introduction

Metabolites control the activity of a variety of epigenetic enzymes.[24, 25, 87, 116, 120, 122] Whereas αKG stimulates TET activity, fumarate, succinate and the oncometabolite 2HG are TET-inhibitors[87, 105] and OAA was also shown to inhibit αKG-dependent dioxygenases[83]. We studied in detail the interaction between Gdh and Tet3 as they share a natural link via αKG, but we are only at the beginning to understand the metabolic network that controls TET activity. Hippocampal depolarization in presence of 2-deoxy-D-glucose (2DG) and pyruvate resulted in lower 5hmdC levels. 2DG inhibits glycolysis but has no direct effect on Gdh when pyruvate is not limited. This indicates that global metabolic changes do effect TET activity that seem to be activated under glycolytic conditions (addition of UK5099) and downregulated when glycolysis does not take place (addition of 2DG/pyruvate). The NAD^+/NADH machinery might play a role, but other metabolites could be also important.

Results

In our initial proteomics study using HEK293T-expressed GFP-Tet3 as a fishing bait for interaction partners in murine nuclear brain lysate, not only Gdh, Mdh1/2 and Ldhb were identified as metabolic enzymes that interact with Tet3, but many more (Figure 11A). Interestingly, of the enzymes of the TCA cycle, neither SDH nor FH nor IDH were identified among them. SDH, FH and IDH are often found mutated in cancers, which results in loss of function or altered activity of these enzymes and therefore succinate, fumarate and 2HG accumulate to millimolar concentrations that consequently inhibit TET enzymes.[106, 355, 356] Under physiological conditions, the concentrations of these metabolites are not high enough for TET-inhibition and our Tet3 co-immunoprecipitation (coIP) results indicated that the effective molarity is not increased as the producing enzymes are not interacting. Among the identified metabolic enzymes was Aco2, which could be confirmed as a Tet3-interaction partner in a Tet3-endogenous coIP analyzed via liquid chromatography tandem mass spectrometry (LC-MS/MS) (p-value for the enrichment 0.0398) and western blotting (Figure 11A). Aco2 is a mitochondrial, iron-dependent enzyme of the TCA cycle that catalyzes the isomerization of citrate to isocitrate.[357] Surprisingly, only Aco2, which is the mitochondrial form and not Aco1, which is the cytoplasmatic form, was found enriched. Conversely to the co-IP results with nuclear lysate, immunohistochemistry (IHC) of Aco2 in the hippocampus pointed towards a mitochondrial localization instead of a nuclear one (Figure 11B). Nevertheless, there were already reports that Aco2 can also be in the nucleus.[116] Since we found Aco2, but not any enzyme of the IDH protein family as a Tet3-interactor, we assumed that isocitrate might accumulate close to Tet3 in brain. Ligand docking simulations using Maestro 10.7 and the available crystal structure of the Tet2 catalytic domain (cd, PDB ID: 4NM6) indicated indeed that isocitrate can possibly bind in the catalytic pocket of the TET enzymes. Therefore, we wanted to investigate the inhibitory potential of isocitrate regarding Tet3. In an *in vitro* assay using a 5mdC-containing oligonucleotide and GFP-Tet3cd immobilized on agarose beads and which we analyzed by matrix assisted laser desorption ionization – time of flight mass spectrometry (MALDI-TOF-MS), we confirmed that isocitrate can competitively inhibit TET-activity (Figure 11C). Taken together, these results suggest that isocitrate can be a TET-inhibitor, however the possibility does not imply physiological relevance.

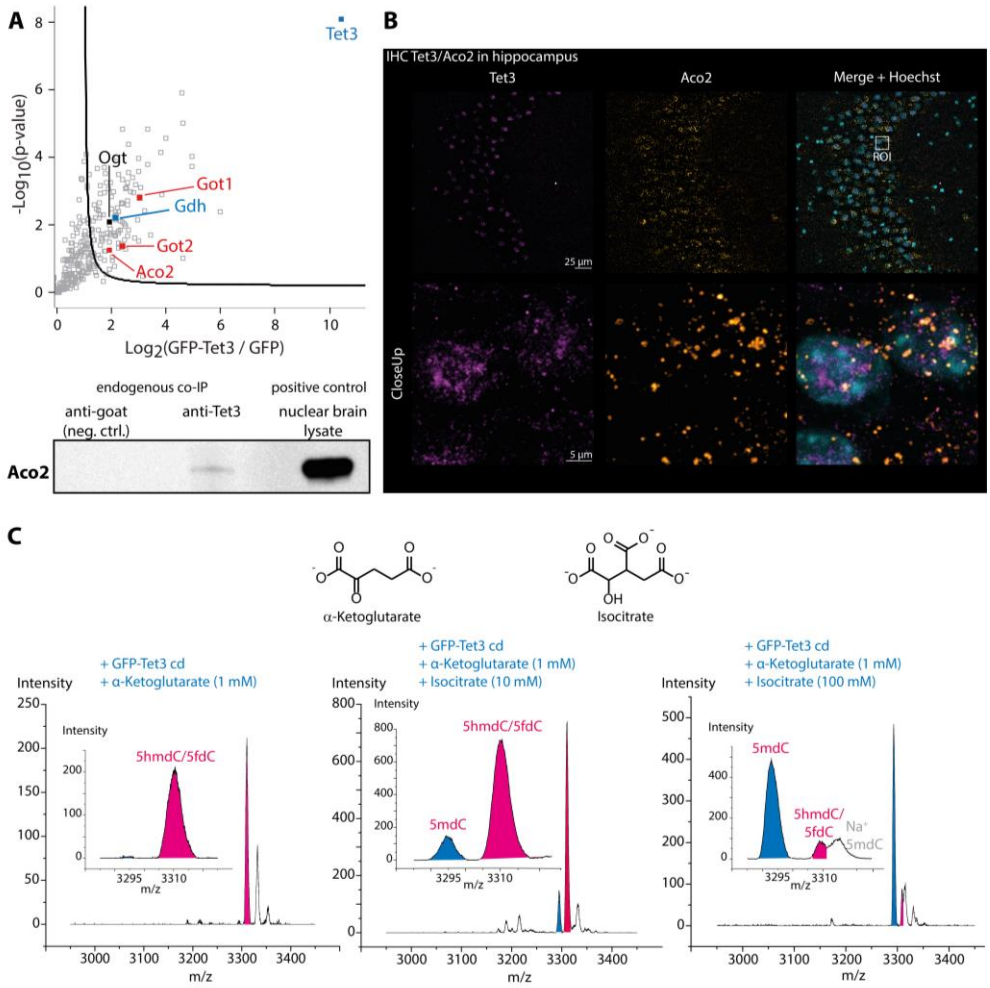
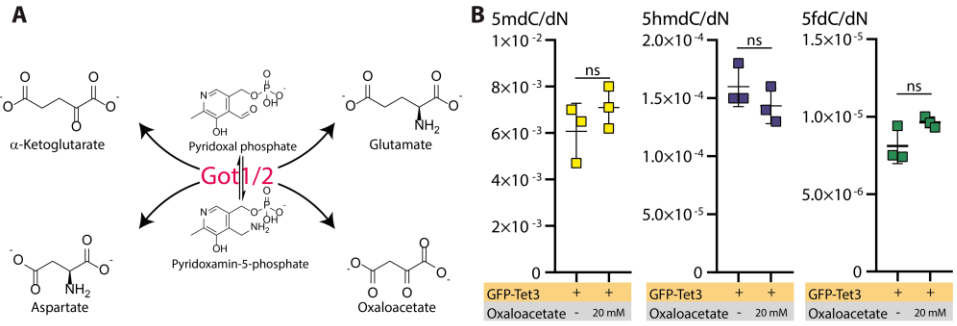


Figure 11: **Aco2 as a Tet3 interaction partner and effect of isocitrate on TET-activity.** (A) Volcano blot after mass spectrometry-based analysis of Tet3-enriched co-IP and western blot after Tet3-endogenous co-IP in murine brain lysate showing that Aco2, Got1 and Got2 interact with Tet3. Ogt (black) and Gdh (blue) are shown as comparison. (B) IHC of Tet3 and Aco2 in the murine hippocampus. Hoechst staining shows the nuclei. The ROI that was used for the close-up image is marked. (C) Molecular structure of αKG and isocitrate and MALDI-TOF-MS spectra of a single stranded DNA oligonucleotide containing 5mdC (3294.5 m/z*)* that is oxidized *in vitro* to 5hmdC and 5fdC (3310.8 m/z) by the catalytic domain of GFP-Tet3cd in presence of 1 mM αKG and increased amounts of isocitrate.
5fdC = 5-formyl-2'-deoxycytidine, 5hmdC = 5-hydroxymethyl-2'-deoxycytidine, 5mdC = 5-methyl-2'-deoxycytidine, Aco2 = aconitase 2, αKG = α-ketoglutarate, cd = catalytic domain, co-IP = co-immunoprecipitation, Gdh = glutamate dehydrogenase 1, GFP = green fluorescent protein, Got = aspartate aminotransferase, IHC = immunohistochemistry, MALDI-TOF-MS = matrix assisted laser desorption ionization – time of flight mass spectrometry, Ogt = *O*-linked β-*N*-acetylglucosamine transferase, ROI = region of interest, Tet3 = ten-eleven translocation enzyme 3.

In addition of Aco2, all members of the MAS, which consists of Mdh1 and 2 and Got1 and 2, were found to be enriched in the GFP-Tet3 coIP in nuclear brain lysate. In contrast to dehydrogenases, which use NAD^+ as a redox equivalent, Got1 and 2 are regulating the NAD^+ levels, but do not depend on them. They use PLP as a cofactor to transfer the amino group from glutamate to OAA, thereby generating αKG and aspartate or vice versa (Figure 12A)[358] and could in principle still supply Tet3 with αKG when NAD^+ is limited and therefore not available for Gdh-mediated αKG-supply. We found both, Got1 and 2 significantly enriched in the Tet3-enriched coIP in murine nuclear brain lysate (Figure 11A). Although, Got2 was also significantly enriched in the Tet3-endogenous coIP (p-value for the enrichment 0.0102), attempts using proximity ligation assay (PLA) in brain tissue and GFP-Tet3/Got2-expressing HEK293T cells failed to show a direct interaction of Tet3 and Got1 or 2. However, the direct interaction is not necessarily needed to influence Tet3-activity by metabolites – proximity is enough to increase the effective molarity. However, Gdh already supplies Tet3 with αKG in brain and this raises the question

why proximity of Got2, which does not influence Gdh activity, is also beneficial. One explanation could be that in case of glucose-deprivation, Gdh is essential in the mitochondria to fuel the TCA cycle with αKG[359] and in this case Got2 would step in to secure the αKG-supply for Tet3. To test this hypothesis, mice should be kept on a low-glucose diet and Gdh localization and Tet3 interaction in the hippocampus should be checked. Furthermore, Got1/2 can produce αKG and at the same time reduce the amount of the potential Tet-inhibitor OAA[83] when TET-activity is desired or reduce the amount of αKG co-substrate and produce OAA when TET-activity should be decreased. Therefore, the activity of Got1/2 could influence TET-activity very efficiently. It has been shown for other αKG-dependent dioxygenases than TET enzymes that OAA is an inhibitor, since it binds to the active center with the δ-carboxyl group instead of the α-carboxyl group.[83] Feeding of 20 mM of OAA to GFP-Tet3 transfected HEK293T cells did not influence the 5hmdC or 5fdC levels in a certain direction (Figure 12B). Since OAA is doubly negatively charged, it might be that the molecule cannot enter the cells and it might be that feeding of the dimethyl-OAA ester (the corresponding molecule to DM-αKG) would have an effect. MALDI-TOF-MS analysis of an *in vitro* assay with GFP-Tet3cd and a 5mdC-containing oligonucleotide confirmed that OAA is a competitive (Figure 12C) and compared to isocitrate a better inhibitor of TET enzymes (Figure 12D).



D

	-	Oxaloacetate		Isocitrate	
		10 mM	100 mM	10 mM	100 mM
5mdC	0 %	79.2 %	100 %	14.7 %	86.6 %
5hmdC/ 5fdC	100 %	20.8 %	0 %	85.3 %	13.4 %

Figure 12*:* **Got1/2 activity and the effects of OAA on Tet3 activity.** (A) Transaminase reactions that are catalyzed by Got1/2. Pyridoxal phosphate serves as a cofactor. (B) Cytosine modification levels of GFP-Tet3-transfected HEK293T cells in the presence of 0 mM or 20 mM OAA. Graph presents biological replicates. The bar represents the mean; error bars represent standard deviation. Statistical analysis unpaired t-test; p-values 0.6872 (5mdC ctrl./20 mM OAA), 0.1161 (5hmdC ctrl./20 mM OAA) and 0.0182 (5fdC ctrl./20 mM OAA). (C) MALDI-TOF-MS spectra of a single stranded DNA oligonucleotide containing 5mdC (3294.5 m/z*)* that is oxidized *in vitro* to 5hmdC and 5fdC (3310.8 m/z) by the GFP-Tet3cd in presence of 1 mM αKG and increased amounts of OAA. (D) Direct comparison of Tet3 inhibition by OAA and isocitrate at different concentrations.
5fdC = 5-formyl-2'-deoxycytidine, 5hmdC = 5-hydroxymethyl-2'-deoxycytidine, 5mdC = 5-methyl-2'-deoxycytidine, αKG = α-ketoglutarate, cd = catalytic domain, ctrl. = control, GFP = green fluorescent protein, Got = aspartate aminotransferase, MALDI-TOF-MS = matrix assisted laser desorption ionization – time of flight mass spectrometry, OAA = oxaloacetate, Tet3 = ten-eleven translocation enzyme 3.

To investigate the role of Got1/2 for TET activity, the enzymes should be directly addressed by inhibition. However, ectopic expression of TET enzymes in HEK293T cells is an artificial approach, with different transfection and expression efficiencies possibly perturbating the results. Therefore, we wanted to address the question of metabolic influence on TET enzymes in an appropriate, which means neuronal, cellular system. Our model of choice was a small molecule-inducible Neurogenin induced pluripotent stem cells (iPSCs) line (iNGNs). This cell line was created by *Busskamp et al*. by lentiviral transfection of human PGP1 iPSCs[360] with a bicistronic construct (two loci for protein generation) for the doxycycline-inducible expression of Neurogenin 1 and 2.[27] Neurogenin 1 and 2 are key players in neuronal differentiation [361, 362] which upon overexpression results in rapid differentiation of iPSCs into neurons [27]. Before induction, the cells showed a typical stem cell morphology with round cells and colonies with clear margins (Figure 13A, uninduced). Already one day post-induction of Neurogenin 1 and 2 expression by application of doxycycline, the cells were dissociated from the colonies and started to adapt the bipolar-shaped morphology (Figure 13A, 1 day). Two days post-induction, the cells showed neuronal progenitor morphology with ongoing neuron projection development, which describes the formation of neurites (Figure 13A, 2 days). At this stage, the cells were already post-mitotic.[27] Four days post-induction, the iNGNs were fully differentiated to neurons with axons and dendrites, however not fully maturated.[27] The cells have developed a bipolar-shaped morphology with a small cell body and a big nucleus (Figure 13A, 4 days). The iNGN-derived neurons are fully maturated after two weeks with electrophysiological properties reminiscent of normally differentiated neurons.[27]

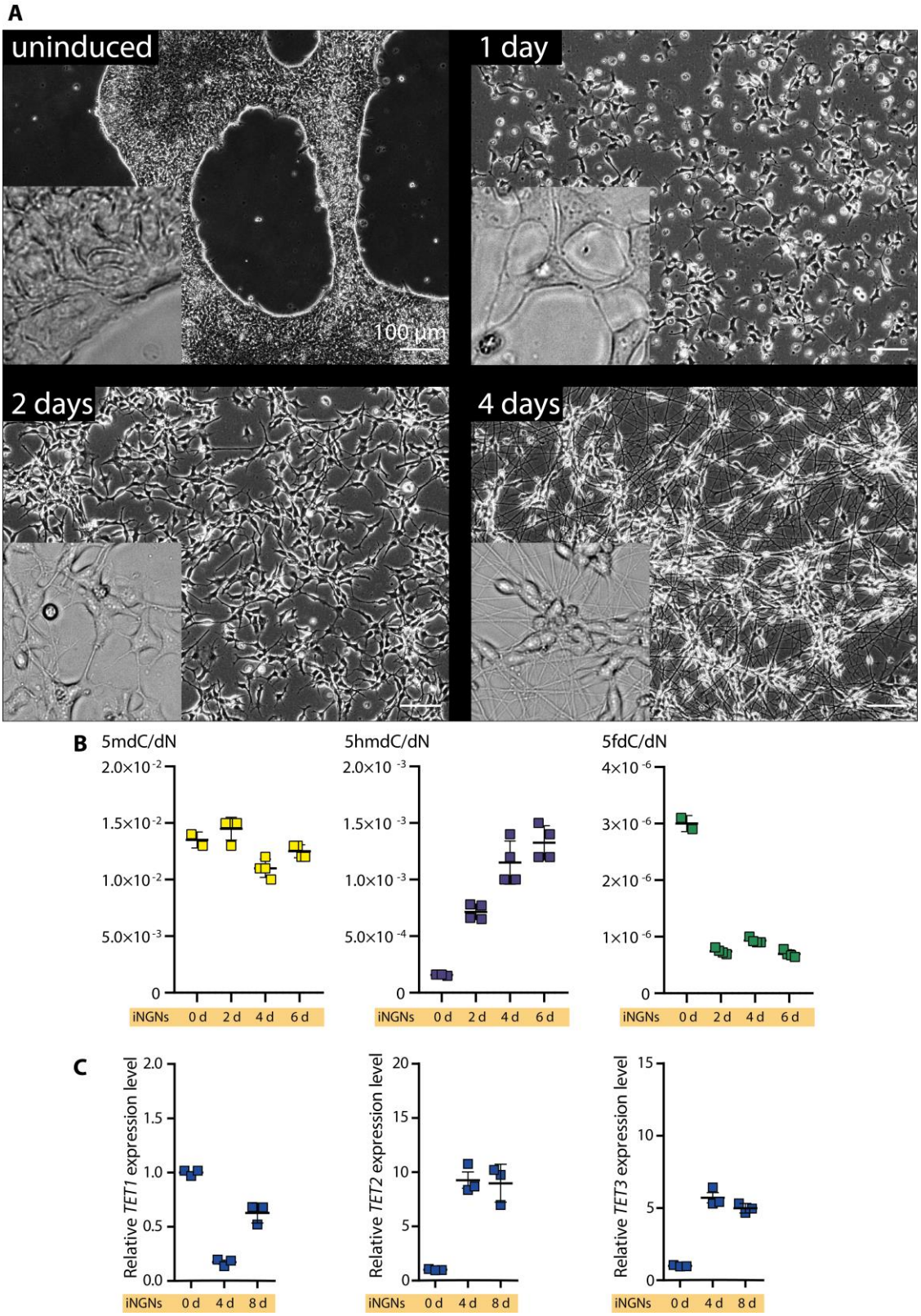


Figure 13*:* **Differentiation of iNGNs.** (A) Brightfield microscopy images (10x and 40x magnification) of iNGNs without induction, one day, two days and four days post-induction of neuronal differentiation by application of doxycycline. (B) Cytosine modification levels and (C) different relative *TET* expression levels of iNGNs at different stages of differentiation. (B) and (C) represent biological replicates. The bar represents the mean; error bars represent standard deviation.
5fdC = 5-formyl-2'-deoxycytidine, 5hmdC = 5-hydroxymethyl-2'-deoxycytidine, 5mdC = 5-methyl-2'-deoxycytidine, d = days post-induction, dN = nucleotide, iNGNs = small molecule-inducible Neurogenin induced pluripotent stem cells line, TET = ten-eleven translocation enzyme.

Since lentiviral transfection was not 100% efficient when the cell line was created, a few cells in the culture (< 5%) do not differentiate to neurons. Consequently, they increasingly contaminate the culture, as they keep proliferating while neurons do not. Therefore, and to avoid the necessity to co-culture the iNGNs with astrocytes, our experiments on iNGNs were carried out between six or eight days post-induction and not 14 days after. Furthermore, wherever possible, cells were harvested by applying phosphate buffered saline (PBS). Washing with PBS was sufficient to detach the neuronal network from the culturing plate but left the undifferentiated cells, which require harsher conditions for removal, on the plate. However, for further experiments with this cell line, when fully matured neurons are desired, the cells should be sorted using flow-cytometry to identify those with successful lentiviral integration.

During the differentiation process, the 5mdC levels stayed within the same range of ~ $1.2x10^{-2}$ 5mdC/dN, whereas the 5hmdC levels increased from day zero to day six and the 5fdC levels decreased (Figure 13B). Quantitative real-time polymerase chain reaction (RT-qPCR) revealed that the gene expression level of *TET1* decreased slightly until four days post-induction, whereas the *TET2* and *TET3* levels strongly increased with beginning of differentiation (Figure 13C). The decreasing 5fdC levels might therefore be a result of lower TET1 expression. Eight days post-induction, *TET1* expression level was higher than at four days post-induction, but lower than in the uninduced iNGNs and for *TET2* and *TET3* the levels did not change compared to four days post-induction. However, the changes in *TET2* and *TET3* gene expression were much higher than for *TET1* expression.

Regarding the TET3/GDH or the TET3/GOT2 interaction, neither the TET3-enriched nor the endogenous TET3 coIP approach with nuclear lysate from iNGN-derived neurons revealed metabolic enzymes as TET3 interactors, indicating that the iNGN system eight days post-induction was unsuitable for studying the previously identified connection between TET3 and metabolic enzymes in neurons. The culturing medium might provide enough αKG or precursors thereof, making the interaction of TET3 with GDH obsolete. However, as we detected positive PLA signal between Tet3 and Gdh in primary cultures of murine hippocampal neurons, which are cultured under similar conditions, but not in the iNGN-derived neurons, the culturing conditions are probably not the underlying cause. The brain consists of many different types of neurons and glia, whereas iNGN-derived neurons are a much more homogenous. Therefore, iNGN-derived neurons might simply be the wrong neuronal type to look at. Another reason might be that iNGN-derived neurons at eight days post-induction are not fully matured or that without neuronal stimulation this interaction network is not established at all.

Although the iNGNs did not seem to be right model system to study the direct interaction of TET3 with metabolic enzymes, they could be used to test the effects of metabolites on TET activity. For the initial feeding trial, I tested DM-αKG, isocitrate, OAA, aspartate and glutamate as metabolites (10 mM concentration each) and started the feeding with the induction of Neurogenin 1 and 2 until six days post-induction. As described, the iNGNs showed already full neuronal morphology at day six but were not fully maturated yet. Feeding of DM-αKG at 10 mM concentration was toxic for the cells (Figure 14A), probably as it perturbed the differentiation process. Only some cells survived until three days post-induction and those who did, showed heavily impaired neurogenesis, if any, compared to the control. Therefore, lower concentrations were tested. At 2 mM DM-αKG concentration, we observed a mild increase in the 5hmdC levels after eight days of feeding (eight days post-induction), whereas with 5 mM concentration we saw a decrease (Figure 14B). We expected to see elevated 5hmdC levels upon DM-αKG feeding if αKG was rate-limiting for TET enzymes in cultures of iNGNs as reported for other systems.[89] However, higher concentrations of DM-αKG seemed to put selection pressure on the cells. The data did not reveal whether the higher concentration of available αKG or the application of DM-αKG, which is hydrolyzed to αKG and methanol (MeOH) by esterases in the cell, was the problem.

With 5 mM concentration it is, however, in both cases possible that we selected those cells that did not differentiate as fast as the others and therefore had lower 5hmdC levels.

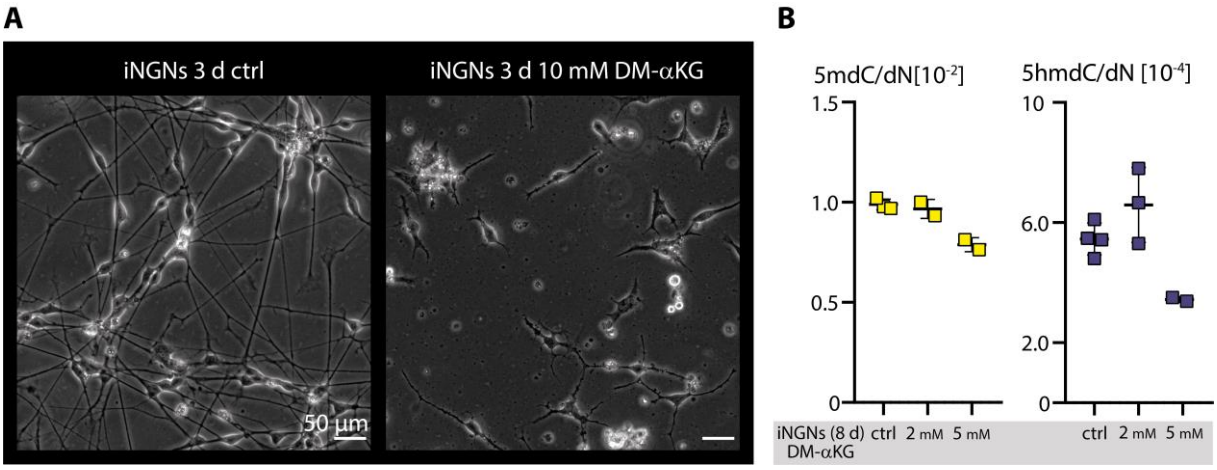


Figure 14: **Feeding of DM-αKG to iNGNs during differentiation.** (A) Brightfield microscopy images ctrl versus 10 mM DM-αKG feeding of iNGNs three days after Neurogenin 1 and 2 induction by doxycycline. (B) Cytosine modification levels of iNGNs eight days post-induction of neurogenesis. Graph presents biological replicates. The bar represents the mean; error bars represent standard deviation.
5hdmC = 5-hydroxymethyl-2'-deoxycytidine, 5mdC = 5-methyl-2'-deoxycytidine, ctrl = control, d = days post-induction, DM-αKG = dimethyl-α-ketoglutarate, iNGNs = small molecule-inducible Neurogenin induced pluripotent stem cells line.

Feeding of aspartate, glutamate, OAA or isocitrate until six days post-induction did not show any effect on the 5mdC and the 5hmdC levels. For 5fdC, the overall difference was significant (p-value 0.0088 for ordinary one-way ANOVA), but the individual differences compared to the unfed control were not. Nevertheless, we observed a decrease of the 5fdC levels for aspartate and glutamate feeding and an increase for OAA and isocitrate feeding. However, without quantification of the intracellular metabolite levels, we could not judge whether the metabolites were taken up by the cells or not. In principle, glutamate and aspartate should be taken up by excitatory amino acid transporters (EAATs)[363, 364], but it might be that only mature iNGN-derived neurons have functional EAATs. This would be in line with the fact that four days after Neurogenin 1 and 2 induction, iNGN-derived neurons are not able to fire action potentials whereas at 14 days post-induction they are.[27] EAATs expression could be checked by RT-qPCR and immunofluorescence/immunocytochemistry (IF/ICC) at different time points. We repeated the feeding but extended the feeding period to eight days post-induction. This time, the 5fdC levels were below the lower limit of quantification, but we observed significant effects on the 5hmdC levels. Again, feeding of OAA only showed a marginal effect on the 5mdC and 5hmdC levels and feeding of isocitrate did not show any effect, most likely because those metabolites were not taken up by the cells. However, feeding of aspartate resulted in significantly lower 5hmdC levels (p-value 0.0403, - 30 %) and feeding of glutamate resulted in significantly lower 5mdC (p-value 0.0420, - 20 %) and 5hmdC levels (p-value 0.0008, - 50 %). Maturated iNGN-derived neurons have glutamatergic synapses.[365] Consequently, long-time exposure to glutamate is toxic for them[218-220] and apparently their susceptibility towards glutamate develops between six and eight days post-induction. The toxic effects of glutamate could explain the almost two third lower 5hmdC levels, which correlated with the maturation. One possibility is that only cells that were not as mature as the others, and therefore had lower 5hmdC levels, survived the glutamate treatment. Another possibility is that 5mdC and 5hmdC levels change in iNGN-derived neurons when the cell initiates the programmed cell death (apoptosis). This was already reported for developmental and degeneration processes in the murine retina, where changes in the 5mdC and 5hmdC patterns seem to be part of apoptotic progression.[366] Furthermore, the changes in the 5mdC and 5hmdC levels could also be explained by neuronal activity that is induced by glutamate application. To test whether the observed effects were based on contamination by

undifferentiated cells that survived glutamate treatment, on cell death events or on neuronal activity one should first exclude the contamination of the culture by flow-cytometry based cell sorting as mentioned above. Apoptosis can be checked by immunostaining of cytochrome C.[367] This protein is located in the inner mitochondrial membrane and is released to the cytosol when the cell is dying, which initiates apoptosis.[368] Additionally, RT-qPCR for certain genes or RNA-sequencing (RNA-seq) for a global comparison of the expression levels should reveal, which genes are up- and downregulated upon glutamate treatment. If apoptosis was the main event happening, we would expect that genes that are associated with cell death are upregulated.[369-371] In contrast, if neuronal activity was the reason for changing levels of 5mdC and 5hmdC, expression of genes associated with neuronal function would be expected to be upregulated.[372, 373] In combination with analysis of the expression profiles, 5mdC and 5hmdC sequencing of pre-defined regions, e.g. *BDNF*, would be a powerful tool to study the effects of glutamate on iNGN-derived neurons during maturation.

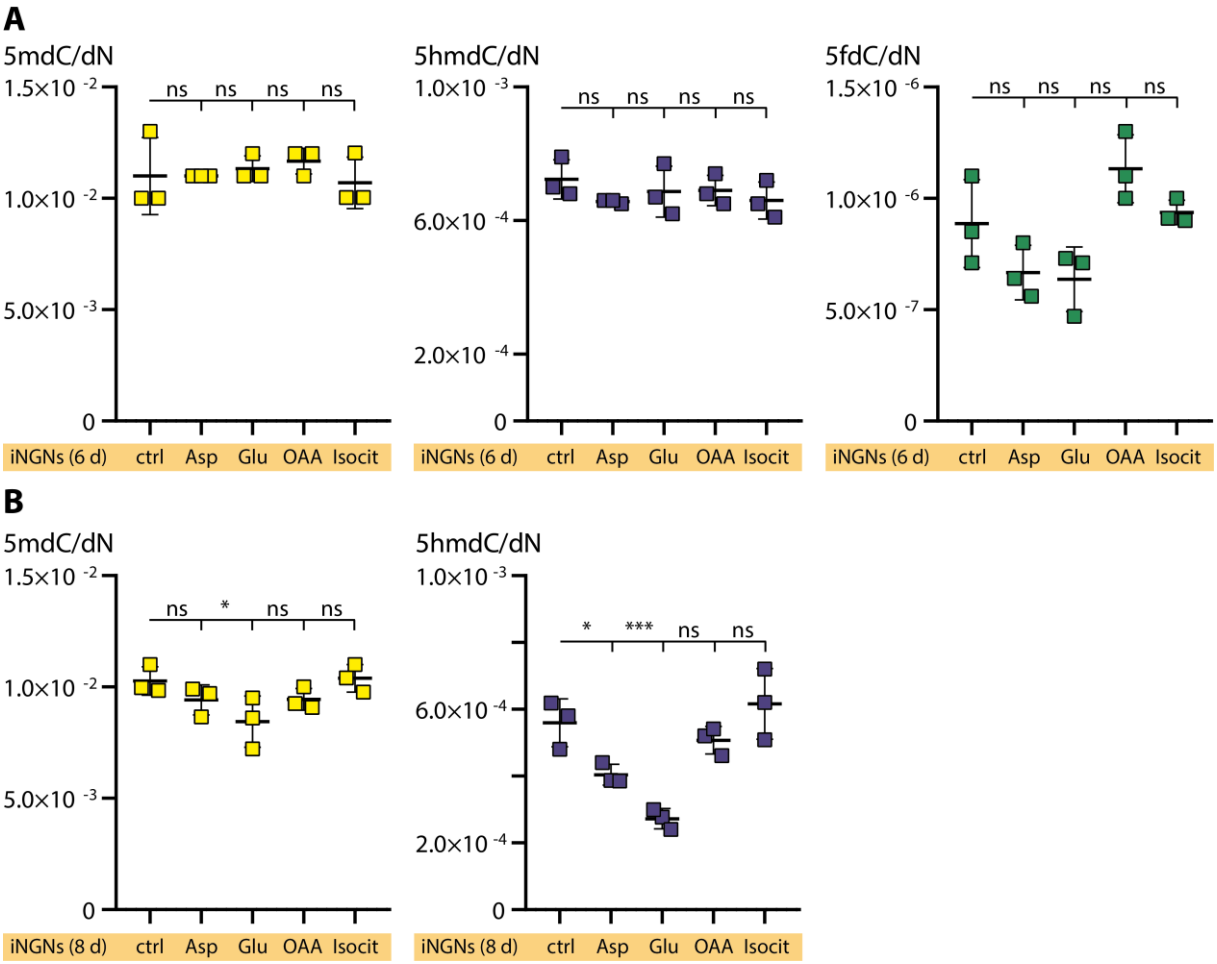


Figure 15: **Cytosine modification levels of iNGNs under different feeding conditions.** (A) Cytosine modification levels for feeding of Asp, Glu, OAA and Isocit (10 mM each) to iNGNs for six days post-induction of differentiation. Unfed iNGNs served as ctrl. Statistical analysis ordinary one-way ANOVA (p-value 0.7818 for 5mdC, 0.5784 for 5hmdC, 0.0088 for 5fdC) with Dunnett's multiple comparisons test against the ctrl column (5mdC: p-value 1.0 (Asp), 0.9809 (Glu), 0.8289 (OAA), 0.9809 (Isocit); 5hmdC: p-value 0.4061 (Asp), 0.8192 (Glu), 0.8601 (OAA), 0.4472 (Isocit); 5fdC: p-value 0.2409 (Asp), 0.1633 (Glu), 0.1706 (OAA), 0.9770 (Isocit)). (B) Cytosine modification levels for feeding of Asp, Glu, OAA and Isocit (10 mM each) to iNGNs for eight days post-induction of differentiation. Unfed iNGNs served as ctrl. Statistical analysis ordinary one-way ANOVA (p-value 0.05639 for 5mdC, 0.0004 for 5hmdC) with Dunnett's multiple comparisons test against the ctrl column (5mdC: p-value 0.4742 (Asp), 0.0420 (Glu), 0.4966 (OAA), 0.9988 (Isocit); 5hmdC: p-value 0.0403 (Asp), 0.0008 (Glu), 0.7085 (OAA), 0.6548 (Isocit)). (A) and (B) The p-values were adjusted for multiple testing. ns p-value ≥ 0.05, * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001.
5fdC = 5-formyl-2'-deoxycytidine, 5hmdC = 5-hydroxymethyl-2'-deoxycytidine, 5mdC = 5-methyl-2'deoxycytidine, Asp = aspartate, ctrl = control, d = days post-induction, dN = nucleotide, Glu = glutamate, iNGNs = small molecule-inducible Neurogenin induced pluripotent stem cells line, Isocit = isocitrate, OAA = oxaloacetate.

Aspartate showed a similar, but milder effect than glutamate. Aspartate is also neurotoxic[374], which might explain the lowered 5hmdC levels. However, lower 5hmdC levels could also be the result of GOT1/2 activity that converted aspartate to OAA, thereby depleting the αKG pool (Figure 12A). Evidence for this hypothesis came from GFP-Tet3-transfected HEK293T cells, which showed lower 5hmdC (- 33 %) and 5fdC (- 30 %) levels when the cells were fed in parallel with 20 mM of aspartate (Figure 16). This effect was multiplied when the cells were not only transfected for Tet3, but also for Got2. Upon aspartate feeding, the 5hmdC levels even dropped by - 70 %. This result could be explained under the assumption that Got2 metabolized aspartate to OAA. Interestingly, the 5fdC levels only dropped by - 35 % in this case.

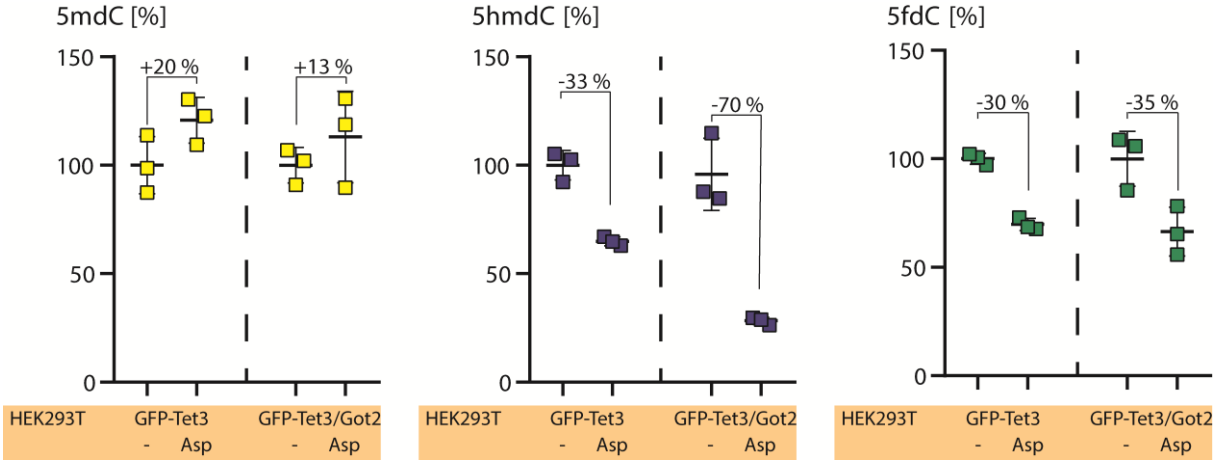


Figure 16: **Cytosine modification levels upon feeding of 20 mM of Asp to GFP-Tet3 only or GFP-Tet3/Got2 transfected HEK293T cells.** The modification levels were normalized to the unfed control. Graph presents biological replicates. The bar represents the mean; error bars represent standard deviation. 5fdC = 5-formyl-2'-deoxycytidine, 5hmdC = 5-hydroxymethyl-2'-deoxycytidine, 5mdC = 5-methyl-2'-deoxycytidine, Asp = aspartate, Got2 = aspartate aminotransferase 2, Tet3 = ten-eleven translocation enzyme 3.

To test whether GOT1/2 activity was the reason for the decreased 5hmdC levels in iNGN-derived neurons after aspartate feeding, metabolomics and inhibition of GOT1/2 by aminooxyacetate[375] should be performed.

In summary, cytosine modification levels in differentiating iNGNs can be influenced by metabolites. However, many parameters, as cell viability and differences in differentiation due to contaminating cells, were not controlled in these preliminary studies. The experiments suggested above would help elucidating the mechanisms behind the observed changes in 5mdC and 5hmdC.

Material and Methods

All experiments were at least repeated twice to ensure reproducibility.

Antibodies

Aco2 antibody (abcam ab110321, clone 6F12BD9, mouse monoclonal immunoglobulin G (IgG)): IHC (1:100), western blotting (1:1000)

Tet3 antibody (Abiocode, R1092-N1, rabbit polyclonal IgG): IHC (1:500), coIP (1 µg)

Secondary antibodies

Alexa555-anti-mouse IgG (Cell Signaling Technologies 4409): IHC (1:800)

Horseradish peroxidase (HRP)-conjugated anti-mouse IgG (Sigma-Aldrich AP130P): western blotting (1:5000)

Alexa488-anti-rabbit IgG (Cell Signaling Technologies 4412): IHC (1:800)

HRP-conjugated anti-rabbit IgG (Sigma-Aldrich A0545): western blotting (1:5000)

Expression plasmids

GFP-Tet1cd (plasmid with ampicillin resistance, CAG promoter):
(N-terminus) eGFP – TEV cleavage site – Tet1cd (C-terminus); Tet1cd is the shortened version of murine Tet1 (UniprotKB – Q3URK3) starting from amino acid 1367 of the original sequence

GFP-Tet3cd (plasmid with ampicillin resistance, CAG promoter):
(N-terminus) eGFP – TEV cleavage site – Tet3cd (C-terminus); Tet3cd is the shortened version of murine Tet3 (UniprotKB – Q8BG87-1) starting from amino acid 696 of the original sequence

GFP-Tet3 (plasmid with ampicillin resistance, CAG promoter):
(N-terminus) eGFP – TEV cleavage site – Tet3; murine Tet3 (UniprotKB – Q8BG87-4) starting from amino acid 1 of the original sequence

Got2-FLAG (plasmid with ampicillin resistance, CMV promoter):
(N-terminus) Got2-FLAG (C-terminus); murine Got2 (UniprotKB – P05202) starting from amino acid 1 of the original sequence

Cell culture HEK293T cells

HEK293T cells (ATCC) were cultivated at 37°C in water saturated, CO_2-enriched (5%) atmosphere. Dulbecco's Modified Eagle's Medium (DMEM, Sigma-Aldrich D6546) or RPMI 1640 (Sigma-Aldrich R0883), containing 10% (v/v) fetal bovine serum (FBS) (Invitrogen 10500-064), 1% (v/v) L-alanyl-L-glutamine (Sigma-Aldrich G8541) and 1% (v/v) penicillin-streptomycin (Sigma-Aldrich P0781), were used as growing medium. When reaching a confluence of 70% to 80%, the cells were routinely passaged. Cells were tested once in two months for Mycoplasma contamination using Mycoplasma Detection Kit (JenaBioscience PP-401L).

Transfection of HEK293T cells

Transfection of HEK293T cells for high protein expression:

The transfection was performed in four p150 petri dishes (Sarstedt 83.3903.300). Five to six million cells per p150 were seeded in 25 mL of medium. After seeding, the cells were incubated for 24 h to reach a confluence of 40% to 80%. 10 µg of expression plasmid DNA and 30 µL of the transfection reagent jetPRIME (Polyplus Transfection, VWR 114-15) were used as described by the manufacturer. 4 h and 28 h after transfection the medium was changed and sodium butyrate (Sigma-Aldrich B5887, final conc. 4 mM) was added. 48 h after transfection, the cells were harvested by trypsinization (Life Technologies 12604013) and immediately used for protein extract preparation.

Transfection of HEK293T cells for QQQ experiments:
The transfection was performed in 6 well plates (Sarstedt 83.3920.300). 2.5×10^5 cells were seeded per well in 3 mL of medium. 24 h after seeding, the cells were transfected using 1.5 µg of DNA per expression plasmid, 4 µL of jetPRIME and 150 µL of jetPRIME buffer. Six hours after transfection, the medium was changed and 24 h after transfection, the cells were lysed and gDNA was isolated as described previously.[376]

Protein extract preparation
Protein extract preparation from transfected HEK293T cells:
The harvested HEK293T cells and the resulting lysate were kept on ice during the preparation all time. Per eight million cells, 1 mL of RIPA buffer (10 mM 2-amino-2-(hydroxymethyl)propane-1,3-diol (Tris, Thermo Fisher Scientific 10724344) pH = 7.5, 150 mM sodium chloride (NaCl, BerndKraft 04160.3600), 0.5 mM ethylenediaminetetraacetic acid (EDTA, AppliChem A1104.0500), 0.1% (w/v) sodium dodecyl sulfate (SDS, AppliChem A2572.0500), 1 % (v/v) Triton X-100 (AppliChem A4975.0500), 1% (w/v) deoxycholate (Sigma-Aldrich 30970)), supplemented with 1 mM dithiothreitol (DTT, AppliChem A2948.0025), 2.5 mM magnesium chloride ($MgCl_2$, Sigma-Aldrich M8266), 100 U/mL benzonase (Merck Millipore 70746-3) and 1 x protease inhibitor cocktail (PIC, Roche 05056489001) on the day of preparation was used for the lysis. Cells were resuspended in RIPA buffer and lysed for one hour at 4 °C on a tube rotator. Afterwards, the lysate was centrifuged (10000 x g, 15 min, 4 °C) and the supernatant containing the proteins were transferred into a new tube. To enrich GFP-tagged proteins, the lysate was immediately incubated with GFP Nano-Traps either on agarose beads (Chromotek gta-20, for *in vitro* activity assay) or on magnetic agarose beads (Chromotek gtma-20). A Bradford protein assay (Bio-Rad 5000006) was performed according to the manufacturer´s instructions to determine the protein concentration.

Protein extract preparations from murine brain:
Protein extracts from whole murine brain (Mus musculus, C57-BL6/J wild type, both genders, 110 days old; Charles River, Sulzfeld, Germany), including separation into the organelle, the cytosolic and the nuclear fraction, were prepared according to a previously published protocol.[377] The nuclear extract was then treated with 25 U/mL benzonase for 30 min on ice and subsequently centrifuged (21000 x g, 15 min, 4 °C). The supernatant containing the nuclear lysate was transferred to a new tube. A Bradford protein assay was performed according to the manufacturer´s instructions to determine the protein concentration.

***In vitro* activity assay**
For the *in vitro* activity assay on GFP Nano-Trap on agarose beads, GFP-Tet3cd bound to the trap was used. The proteins were ectopically expressed in HEK293T (per *in vitro* assay 1 x P150 culture dishes for GFP-Tet1cd, 3 x P150 culture dishes for GFP-Tet3cd) and the resulting protein extracts were

incubated for 1 h at 4 °C on a tube rotator with 140 µL of GFP Nano-Traps on agarose beads per extract. Afterwards, the beads were washed twice with coIP wash buffer 1 (10 mM 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (HEPES, Roth HN77.5) pH = 7.5, 150 mM NaCl, 0.5 mM EDTA), twice with coIP wash buffer 2 (10 mM HEPES pH = 7.5, 1 M NaCl, 0.5 mM EDTA) and twice with coIP wash buffer 1. The fluorescence of the GFP-tagged proteins bound to the GFP Nano-Trap was checked on a Tecan Plate Reader (Tecan GENios Pro, fluorescence intensity excitation 400 nm, emission 535 nm) and per reaction, the amount of Nano-Trap was adjusted that the fluorescence signal was 25000 per reaction. Per assay, three reactions in TET reaction buffer (50 mM HEPES pH = 7.5, 100 mM NaCl, 2 mM Vitamin C (Sigma-Aldrich A5960), 1.2 mM ATP (Sigma-Aldrich A26209), 2.5 mM DTT, 1 mM αKG (Sigma-Aldrich 75890), 0.1 mM ammonium iron(II) sulfate hexahydrate (Sigma-Aldrich 215406) with different concentrations (0 mM; 10 mM; 100 mM) of OAA (Sigma-Aldrich O7753) or isocitrate (Sigma-Aldrich I1252) and 4 µM of DNA oligonucleotide (5'-UUUUG[5mdC]GGUUG-3') were set up in 50 µL reaction volume/sample. The samples were incubated at 35 °C for 4 h under shaking. For MALDI-TOF-MS measurements, 1 µL of the reaction supernatant was desalted on a 0.025 µm ø VSWP filter membrane (Thermo Fisher Scientific VSWP02500) against ultrapure water (ddH_2O) for at least one hour, co-crystallized in a 3-hydroxypicolinic acid matrix and mass spectra were recorded on a Bruker Autoflex II in a m/z range of 1500 to 6000. For each condition (Tet3cd, defined concentrations of OAA or isocitrate) three independent experiments were set up and MALDI-TOF spectra recorded. Afterwards the mean was calculated. The area under the curve was calculated using GraphPad Prism (version 8.0.0 or higher).

Tet3-enriched coIP

20 µl of magnetic anti-GFP beads (Chromotek gtma-20) were washed three times with GFP wash buffer (10 mM Tris pH 7.5, 150 mM NaCl, 0.5 mM EDTA) and then incubated for 15 min on ice with nuclear extract of GFP-Tet3 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-Tet3 loaded beads were then washed twice with coIP wash buffer 1, twice with coIP wash buffer 2 and twice 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). The GFP-Tet3 beads were subsequently incubated with 200 µg nuclear brain extract for 15 min on ice. Following, they were washed twice with GFP wash buffer (10 mM Tris pH = 7.5, 150 mM NaCl, 0.5 mM EDTA). To elute the bound proteins, 50 µl of 200 mM glycine pH 2.5 were added and the solution was vortexed for 30 s. To gain more yield, the elution step was repeated. For the negative control, the same procedure was followed using GFP instead of GFP-Tet3.

Tet3 endogenous coIP

CoIP of endogenous Tet3 was performed using nuclear brain extract. When the interactors were subsequently analysed by MS, 250 µg of nuclear brain extract, 1 µg of antibody and 10 µL of Dynabeads Protein G (Thermo Fisher Scientific 10003D) were used per replicate; for the analysis by western blot the doubled amount of nuclear brain extract, antibodies and Dynabeads Protein G were used. Anti-goat IgG was used for the negative control.

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 GFP wash buffer. Afterwards, the Dynabeads were added to the lysate, PBS was added to a final volume of 500 µL and the suspension was incubated for 1 h at 4°C on a tube rotator. After incubation, the beads were washed three times with coIP wash

buffer 1. Last, proteins were eluted with 30 µL of 1% (v/v) formic acid for 15 min at room temperature (MS 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) bromophenol blue) for 10 min at 70 °C (subsequent western blotting).

LC-MS/MS analysis

Samples for the mass spectrometer were reduced by the addition of 100 mM Tris(2-carboxyethyl)phosphine (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 V5113) at 37 °C for 16 h. Afterwards, they were incubated for 5 min at 100 °C and subsequently 1 mM phenylmethylsulphonyl fluoride was added. StageTips were utilized to purify the samples for MS.[378]

The samples were analysed 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; 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 60 000. 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. 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.

LFQ data processing

The MaxQuant software (version 1.5.0.25) was used for LFQ. Quantification was performed with four biological replicates for Tet3-enriched coIP. GFP alone (four biological replicates) served here as control. Three biological replicates were used for the Tet3 endogenous coIP and the coIP (three biological replicates) 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 PRIDE[379] partner repository with the dataset identifier PXD004518.

LFQ data was analysed 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).

Western blotting

For SDS polyacrylamide gel electrophoresis (PAGE) with subsequent western blotting, samples were loaded on a 10% polyacrylamide gel and MagicMark XP Standard (ThermoFisher LC5603) and Blue Prestained Protein Standard, Broad Range (11-190 kDa) (New England Biolabs P7706S) or Color-coded Prestained Protein Marker, Broad Range (11-250 kDa) (New England Biolabs 14208) were used as protein standards. The gel was run at constant 150 V for 60 min in SDS running buffer (25 mM Tris, 192 mM glycine, 0.1% (w/v) SDS). For blotting, we used a PVDF blotting membrane (GE Healthcare Amersham Hybond P0.45 PVDG membrane 10600023) and pre-cooled Towbin blotting buffer (25 mM Tris, 192 mM glycine, 20% (v/v) MeOH, 0.038% (w/v) SDS). The membrane was activated for 1 min in MeOH, washed with water and equilibrated for additional 1 min in Towbin blotting buffer; the Whatman gel blotting papers (Sigma-Aldrich WHA 10426981) were equilibrated for 15 min in Towbin buffer and the precast gel was equilibrated for 5 min in Towbin buffer after the run. Western blotting (tank (wet) electro transfer) was performed at 4 °C for 10 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 Tris-buffered saline including Tween-20 (TBS-T; 20 mM Tris pH = 7.5, 150 mM NaCl, 0.1% (v/v) Tween-20). The primary antibodies were diluted in 5 mL of 5% (w/v) milk powder in TBS-T. The blocking suspension was discarded, and the diluted primary antibodies were added for 12 h at 4°C and shaking. After incubation, the primary antibodies were discarded, and the membrane was washed three times five minutes with TBS-T. HRP-conjugated secondary antibodies were diluted in 5% (w/v) milk powder in TBS-T and added for 1 h at room temperature under shaking. Afterwards, the membrane was washed two times with TBS-T and one time with TBS (TBS-T without Tween-20) before SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific 34077) was used for imaging.

IHC

All steps were performed in in a humidity chamber and at room temperature when not otherwise specified. 12 µm thick coronar cryo-sections of snap-frozen adult mouse brain and liver were incubated (if applicable) with MitoTracker Deep Red FM (Invitrogen M22426) before fixation. Cryo-sections were fixed on slides using 4% formaldehyde (4% FA, ThermoScientific 28908) in 0.1 M phosphate buffered solution, pH 7.4 (0.1 M PB). After three times washing with 0.1 M PB, the slices were permeabilized and blocked for 30 min using 0.3% (v/v) Triton X-100 and 5% (v/v) blocking reagent CB (Chemiblocker, Merck Millipore 2170) in 0.1 M PB. The primary antibodies were diluted in 0.1 M PB, containing 5% (v/v) CB and 0.3% (v/v) Triton X-100 and applied for 12 h at 4°C. For the negative controls, no primary antibodies were added. After incubation, slices were washed three times with 0.1 M PB containing 2% (v/v) CB. For secondary detection, the fluorescent labelled secondary antibodies were diluted in 0.1 M PB, containing 3% (v/v) CB and applied the antibodies for 1 h in the dark, followed by three times washing with 0.1 M PB. Cell nuclei were stained with Hoechst 33342 (5 µg/mL), which was applied for 10 min in the dark, followed by one washing step with 0.1 M PB. After mounting (Mountant Permafluor, ThermoScientific TA-030-FM), the slices were analysed using a Leica SP8 confocal laser scanning microscope (Leica, Wetzlar).

Isolation of gDNA, RNA and triple-quadrupole (QQQ) MS/MS

Isolation of gDNA, RNA and QQQ-MS/MS experiments were performed according to a previously published protocol.[376]

Cell culture iNGNs

iNGNs were cultivated at 37°C in water saturated, CO_2-enriched (5%) atmosphere. iNGNs were grown on Geltrex (ThermoScientific A1413202)-coated tissue plates. For coating, Geltrex was thawed on ice and diluted 1:1000 in ice-cold DMEM/Ham's F-12 (F-12, Sigma-Aldrich N4888) (1:1) to a final concentration of 15 µg/mL. Per 6 well plate at least 4 mL, per p60 (Sarstedt 83.3901.300) at least 6 mL, per p100 (Sarstedt 83.3902.300) at least 15 mL and per p150 at least 25 mL of coating medium were used. Plates with coating medium were kept for two to 16 hours in the incubator. For storage of up to four weeks, Geltrex-coated plates were sealed with parafilm and stored in the fridge. Uninduced iNGNs were cultured in E8 medium (1:1 DMEM: F-12, supplemented with 2 mM L-alanyl-glutamine, 0.1 mg/mL penicillin-streptomycin, 0.2 mM L-ascorbic acid 2-phosphate, 77.6 nM sodium selenite, 10.90 mM NaCl, 10 µg/mL hHOLO-Transferrin (Merck Millipore 616424), 20 µg/mL hrInsulin (Sigma-Aldrich I9278), 100 ng/mL hrFGF-2 (PeproTech AF-100-18B), 2.0 ng/mL hrTGF-β1 (PeproTech 100-21C)) and iNGNs were passaged in the uninduced state, when they reached about 60 % confluency. For passaging, 2 µM (final concentration) of Thiazovivin (Merck Millipore 420220) was added to the medium for 24 h. For induction of neuronal differentiation, doxycycline (Sigma-Aldrich D9891) final concentration 500 µM) was added for three days and the medium was changed to E6 (E8 without hrFGF2 and hrTGF-β1). Four days post-induction, the medium was changed to E6/modified Neurobasal A (Neurobasal A (Thermo Fisher 10888022, supplemented with 2 % NS21 (PAN Biotech P07-20021)[380] and 2 mM L-alanyl-glutamine) and six days post-induction the medium was changed to modified Neurobasal A. When iNGNs were fed, the feeding started at the day of induction and medium, supplemented with the metabolite in the desired concentration, was changed every day. Metabolites were added from a freshly prepared aqueous 1 M stock solution; the pH was adjusted to 7.4 before addition.

RT-qPCR

For complementary DNA (cDNA) synthesis, the RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific K1621) was used according to the manufacturer's instructions. 190 ng of total RNA was used for cDNA synthesis. RT-qPCR was performed on StepOnePlus Real-Time PCR system (Applied Biosystems, Thermo Fisher Scientific) using PowerUp™ SYBR™ Green qPCR Master Mix (Applied Biosystems, Thermo Fisher Scientific A25742). C_T values of each sample were determined in technical duplicates by the StepOne™ Software (Applied Biosystems, Thermo Fisher Scientific) using the fast cycle protocol. The relative expression levels of target genes were then quantified from three biological replicates (n = 3) according to *Pfaffl et al*.[381, 382]

The following primer sets were used:

hACTB:	5′ GCC GCC AGC TCA CCA T 3′	(forward)
	5′ CAC GAT GGA GGG GAA GAC G 3′	(reverse)
hTET1:	5′ GCT CTC ATG GGT GTC CAA TTG CT 3′	(forward)
	5′ ATG AGC ACC ACC ATC ACA GCA G 3′	(reverse)
hTET2:	5′ AAG GCT GAG GGA CGA GAA CGA 3′	(forward)
	5′ TGA GCC CAT CTC CTG CTT CCA 3′	(reverse)
hTET3:	5′ GCA AGA CAC CTC GCA AGT TC 3′	(forward)
	5′ CCT CGT TGG TCA CCT GGT TC 3′	(reverse)

Contributions

I designed experiments, performed and analysed Tet3 endogenous coIPs followed by LC-MS/MS and western blotting, performed and analysed western blots, IHC/ICC experiments, UHPLC-QQQ-MS experiments, *in vitro* activity assays, RT-qPCR, cultured and differentiated iNGNs and did metabolite feeding experiments. *Dr. Andrea Glück*[1] designed proteomics experiments, developed the GFP-Tet3 enriched coIP workflow, performed and analysed GFP-Tet3 enriched and Tet3 endogenous coIPs followed by LFQ MS/MS and interpreted proteomics data. *Dr. Anna Geserich*[2] helped performing and analysing IHC experiments, performed and analysed RT-qPCR experiments. *Dr. Katharina Iwan*[1] helped performing and analysing UHPLC-QQQ-MS experiments. *Angie Kirchner*[1] helped culturing, differentiating and feeding of iNGNs. *Dr. Markus Müller*[1] and *Dr. Fabio Spada*[1] helped designing experiments and interpreted data. *Prof. Dr. Stylianos Michalakis*[2] provided iNGNs, designed experiments and interpreted data. *Prof. Dr. Thomas Carell*[1] designed and supervised the whole study and interpreted data.

[1] Carell Group, Department of Chemistry, LMU München

[2] Michalakis Group, Department of Pharmacy, LMU München

The interplay of TET3 and MeCP2 in iNGN-derived neurons

Introduction

The approach using TET3-coIPs in iNGN-derived neurons was not suitable for studying the interaction of TET3 with metabolic enzymes, but it revealed other potentially interesting TET3 interaction partners, among them MeCP2. We decided to have a closer look at the MeCP2/TET3 interaction for several reasons. First, although MeCP2 has already been studied intensively, there are many open questions regarding the interplay with TET3. Both proteins are epigenetically acting and are known to be important for neural functioning[15, 17] and both proteins recognize and bind 5mdC and 5hmdC.[14, 19, 322] Furthermore, chromatin immunoprecipitation with subsequent DNA sequencing (ChIPseq) in the IMR90 cell line (derived from human lung fibroblasts) and the HCT-116 (derived from human colon) revealed that the genomic locus of *TET3* is a putative target of MeCP2.[383] Interestingly, the genomic locus of *Tet3* does not appear as a putative target for Mecp2 in murine brain samples.[384-387] Based on those findings, MeCP2 and TET3 do not only have the same target (5mdC and 5hmdC), but additionally MeCP2 seems to be a potential regulator of *TET3* expression in human cells. Moreover, MeCP2 might influence TET3 activity, which could consequently alter 5mdC and 5hmdC patterns of common target genes and thereby influence their expression levels. Consequently, RTT patients might show altered TET3 activity in neurons with unknown consequences for clinical progression of RTT. On the other side, TET3 might influence the binding affinity of MeCP2 towards 5mdC and 5hmdC. In this case, TET3 mutations might play a role in untypical RTT. To decipher the contribution of the TET3/MeCP2 interplay to neurological developmental disorders, we created TET3-deficient ($TET3^{-/-}$) and MeCP2-deficient ($MeCP2^{-/Y}$) iNGN cell lines and characterized the resulting phenotype and the consequences of MeCP2-deficiency on TET3-activity and vice versa. Our results show that $MeCP2^{-/Y}$ and $TET3^{-/-}$ iNGN-derived neurons have a distinct, but very similar phenotype on the DNA methylation, transcriptional and proteomics level that is very different from the wt. Additionally, MeCP2 regulates TET3 on the transcriptional and on the protein activity level. Whether TET3 influences MeCP2 binding affinity towards specific genetic loci is not clarified yet but experiments to elucidate this question are ongoing.

Results

MeCP2 was identified as a TET3-interactor using TET3-enriched coIP (Figure 17A) and TET3-endogeneous coIP (Figure 17B) in iNGN-derived neurons eight days post-induction (d8). The MS-analysis after TET3-enriched coIP and the western blot after TET3-endogenous IP both revealed MeCP2 as a specific TET3 interactor in iNGN-derived neurons. Although MeCP2 is only about 53 kDa, the signal for MeCP2 in the western blot appeared just below 80 kDa. However, the phenomenon that MeCP2 gives inconsistent bands regarding molecular weight in polyacrylamide gel electrophoresis because of various posttranslational modifications was already reported in numerous previous studies.[388, 389] Interestingly, we did not detect Mecp2 as a Tet3 interactor in nuclear extracts from total brain of adult mice. Mecp2 was present in the data set, yet not significantly enriched. Human and murine TET3 share 89% sequence identity and human and murine MeCP2 even show more than 95% sequence identity.[390] Therefore, we wanted to test whether the small sequence differences between human and murine TET3 and MeCP2 were pivotal for the interaction. To this end, we ectopically expressed murine GFP-tagged Tet3 (GFP-Tet3) and Myc-tagged Mecp2 (either isoform 1 or 2, Myc-Mecp2) in HEK293T cells and did a coIP against Myc to check by western blotting if we could enrich GFP-Tet3 (Figure 17C). We successfully identified GFP-Tet3 as an interactor of both isoforms of Myc-Mecp2. To rule out that this result was only based on the artificial ectopic co-expression of both proteins in HEK293T, we tested if we also obtain a positive signal in primary cultures of (murine) hippocampal neurons using the PLA. Indeed, we also obtained a specific positive PLA signal for the Tet3/Mecp2 interaction in these cells, which were cultured for seven days at the day of the assay (Figure 17D).

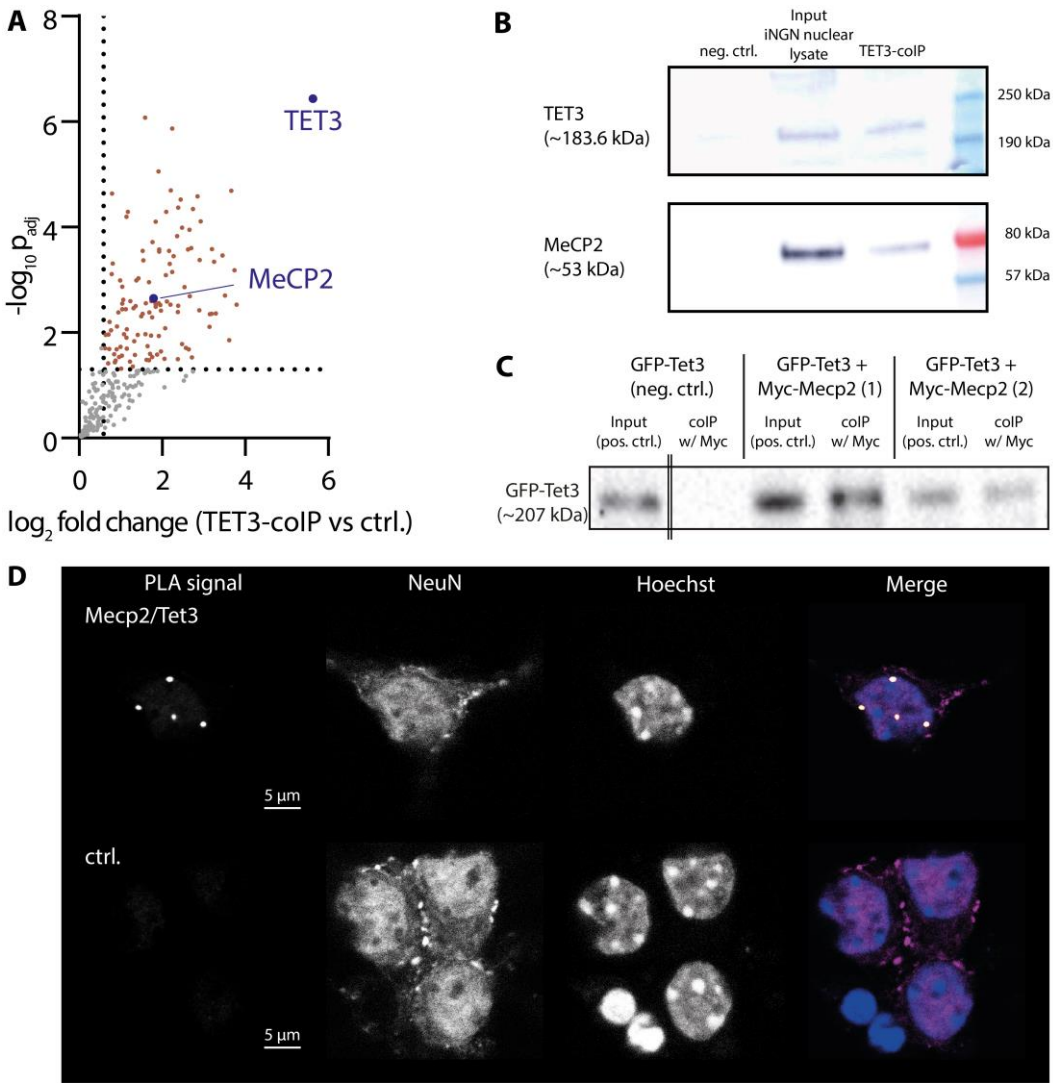


Figure 17: **Interaction of TET3 and MeCP2 in different cellular systems.** (A) TET3 enriched coIP revealed MeCP2 as a TET3 interactor in d8 iNGN-derived neurons (n=4; FDR=0.05; s0=2). (B) Western blot after TET3-endogenous coIP confirmed MeCP2 as TET3 interactor in d8 iNGN-derived neurons. (C) Western blot against GFP after coIP in HEK293T using an anti-Myc-tag antibody. HEK293T cells ectopically expressed GFP-tagged murine Tet3 and either Myc-tagged murine Mecp2 isoform 1 or 2. *GFP-Tet3* single transfected HEK293T served as a negative control. (D) PLA in primary cultures of murine hippocampal neurons also confirmed interaction of Tet3 and Mecp2 in murine neurons.
coIP = co-immunoprecipitation, ctrl = control, d8 = eight days post-induction, FDR = false discovery rate, GFP = green fluorescent protein, iNGNs = small molecule-inducible Neurogenin induced human pluripotent stem cell line, MeCP2/Mecp2 = methyl-CpG binding protein 2, PLA = proximity ligation assay, TET3/Tet3 = ten-eleven translocation enzyme 3.

Based on these results, Mecp2 is apparently also a Tet3 interactor in murine neurons. The fact that it was not enriched in Tet3-coIP in nuclear extracts of total brain of adult mice might have several causes. Nuclear lysate from total brain includes the extracts of various neural cell types. The Mecp2/Tet3 interaction might be limited to neurons or even to neurons in a certain brain area. In contrast, cultures of iNGN-derived or primary hippocampal neurons are more homogenous and therefore MeCP2 might be enriched above background in the TET3-coIP in iNGN-derived neurons, but not in the Tet3-coIP in total brain. However, we did not obtain a positive PLA signal for Tet3 and Mecp2 in the hippocampal region of adult mice either, which does not support the idea that the use of total brain for the Tet3-coIP was the problem. Rather, the TET3/MeCP2 interaction could be only relevant at a certain point in time of neural and especially synaptic development. Evidence for this hypothesis comes from clinical progression of RTT, where the cognitive decline starts months after birth in a time period of pronounced synaptogenesis, but is followed by a period of stabilization.[309] At d8, iNGN-derived

neurons are already mature by morphology, but their electrophysical properties are still developing as well as the synaptic network they form. For primary cultures of hippocampal neurons, it was reported that the number of synapses steeply increases in the second and third week of culturing.[391] Additionally, we were unable to reproduce the positive PLA signal in the primary cultures of hippocampal neurons that were already three weeks old at the time of the assay, supporting the hypothesis that the TET3 and MeCP2 interaction mainly takes place during synaptogenesis.

To further study the interplay between MeCP2 and TET3 in neurons, we created TET3-deficient and MeCP2-deficient iNGN cell lines. Live cell imaging showed that regarding differentiation there were no obvious defects in any of the KO cell lines compared to the wt (Figure 18A). *TET3*$^{-/-}$ and *MeCP2*$^{-/Y}$ iNGNs showed the same speed of differentiation as the wt with neural progenitors two days after doxycycline-induced differentiation and neurons four days post-induction (d4). At d8, all samples showed an established neuronal network with no obvious abnormalities regarding neuronal morphology or number of synapses. However, this result does not imply that *in vivo* neither of the two proteins is important for neuronal differentiation. The strong expression of Neurogenin 1 and 2 under an artificial promoter may simply overrule TET3- or MeCP2-dependent effects on neuronal differentiation.

Next, we analyzed the 5mdC and 5hmdC levels of those cell lines compared to the wt before induction with doxycycline (d0), at d4 and at d8 (Figure 17B). Mixed-effects analysis was chosen for statistical testing, which is a modified form of two-way ANOVA that allows missing values. In summary, for 5mdC and 5hmdC the overall differences between the time points and the cell lines as well as the combination of time points and cell lines were significant (5mdC p-values: time point < 0.0001, cell line 0.0031, time point x cell line 0.0024; 5hmdC p-values: time point < 0.0001, cell line < 0.0001, time point x cell line < 0.0001). Before induction, the 5mdC levels of the *TET3*$^{-/-}$ iNGNs were significantly increased compared to the wt (p-value <0.0001). Surprisingly, the 5hmdC levels in the *TET3*$^{-/-}$ iNGNs were not different compared to the wt, ruling out the possibility that the increase in 5mdC at d0 in the *TET3*$^{-/-}$ iNGNs was a result of less 5mdC to 5hmdC oxidation. At d4 and d8, the 5mdC levels of the *TET3*$^{-/-}$ iNGNs were not significantly different from the wt levels. In contrast, the post-induction 5hmdC levels were decreased in the *TET3*$^{-/-}$ iNGNs compared to the wt. At d4, the decrease was not significant (p-value 0.1132), but at d8 the effect was prominent (p-value 0.0004). This was expected as the gene expression level of *TET3* and therefore its contribution to 5mdC oxidation increases from a low level in the uninduced pluripotent state to a high level in the neuronal state (Figure 13C). The increased 5mdC levels at d0 when the cells were in a pluripotent state and *TET3* expression was low, indicated that the influence of TET3 on the DNA methylation pattern might go beyond simple 5mdC to 5hmdC oxidation. A comparison of the *MeCP2*$^{-/Y}$ iNGNs and the wt showed that regarding 5mdC the levels were not different at d0 and d8, but at d4, the 5mdC levels were significantly increased in the *MeCP2*$^{-/Y}$ iNGNs (p-value 0.0227). For 5hmdC, the situation was different. Whereas the 5hmdC levels in the *TET3*$^{-/-}$ iNGNs dropped after induction compared to the wt, the 5hmdC levels in the *MeCP2*$^{-/Y}$ iNGNs were increased at d4 (p-value < 0.0001) and d8 (p-value < 0.0001), which is in accordance with the literature[234]. RT-qPCR in *TET3*$^{-/-}$ iNGNs showed that *TET1* and *TET2* expression levels were not different from the expression levels in the wt at any timepoint [*PhD Thesis Dr. Victoria Splith*]. In contrast, MeCP2-ablation changed *TET* expression at d0, d4 and d8 (Figure 18C). At d0, the *MeCP2*$^{-/Y}$ iNGNs showed slightly higher *TET1* expression and at d8, *TET2* and *TET3* expression were highly increased compared to the wt, which could explain the higher 5hmdC levels.

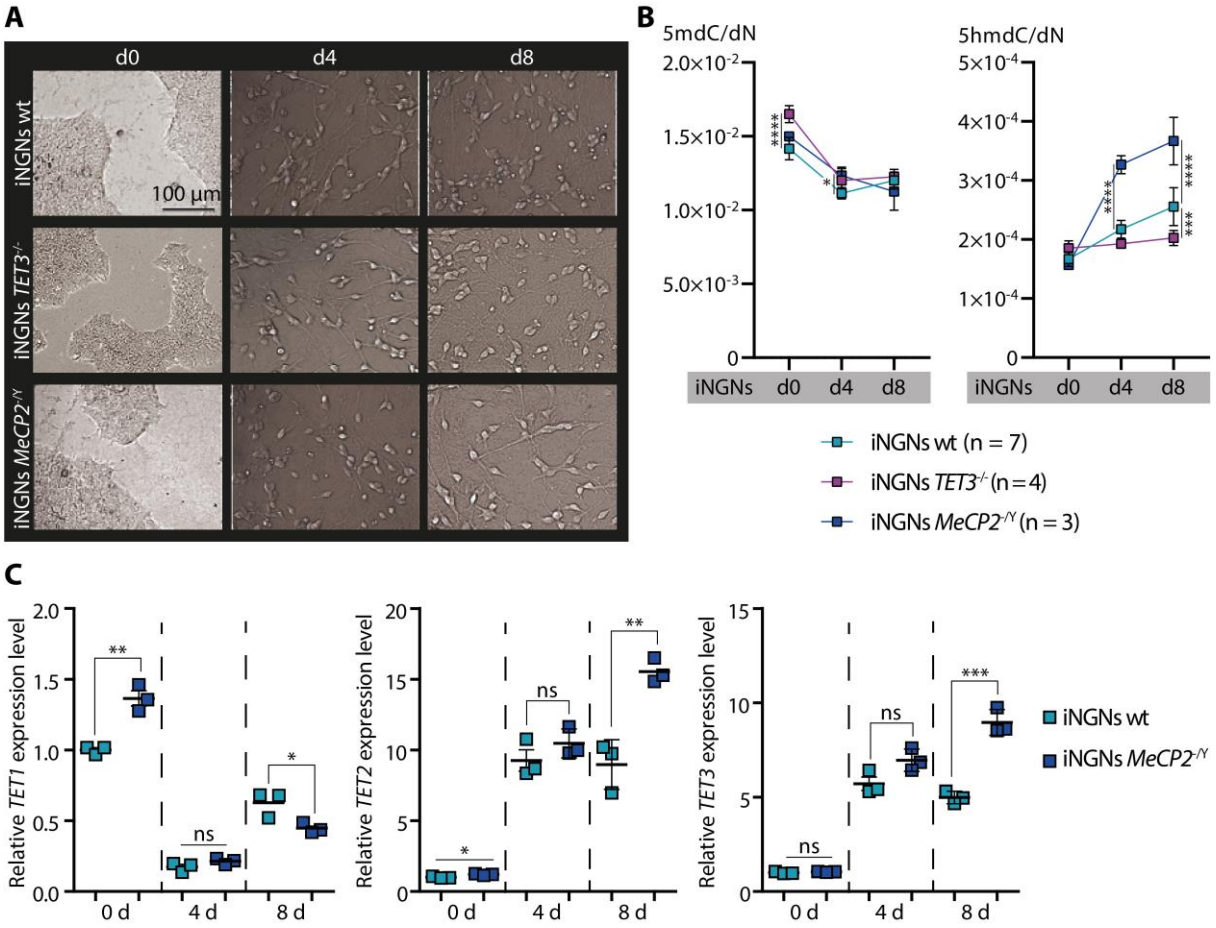


Figure 18: **Characterization of the differentiation of *TET3*$^{-/-}$ iNGNs and *MeCP2*$^{-/Y}$ iNGNs after induction with doxycycline.** (A) Live-cell imaging pictures of iNGNs wt, *TET3*$^{-/-}$ and *MeCP2*$^{-/Y}$ at d0, d4 and d8. (B) Cytosine modifications of iNGNs wt, *TET3*$^{-/-}$ and *MeCP2*$^{-/Y}$ at d0, d4 and d8. Boxes represent the mean of at least three biological replicates. Error bars represent standard deviation. Statistical analysis Mixed-effects analysis (modified two-way ANOVA) with Dunnett's multiple comparisons test. (C) Relative *TET* expression levels of iNGNs wt and *MeCP2*$^{-/Y}$ at d0, d4 and d8. Boxes represent biological replicates. The bars represent the mean of the respective biological replicates; error bars represent standard deviation. Statistical analysis Student's t-test.
5hmdC = 5-hydroxymethyl-2'-deoxycytidine, 5mdC = 5-methyl-2'-deoxycytidine, d0 = before doxycycline induction, d4 = four days after doxycycline induction, d8 = eight days after doxycycline induction, iNGNs = small molecule-inducible Neurogenin induced human pluripotent stem cell line, MeCP2 = methyl-CpG binding protein 2, TET = ten-eleven translocation enzyme, wt = wildtype.
ns p-value ≥ 0.05, * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001, **** p-value < 0.0001.

We decided to study the consequences of TET3 and MeCP2-deficiency in d8 iNGN-derived neurons, since the neurons are already mature by morphology at this state, but their electrophysiological properties still develop and synaptogenesis is ongoing.[392] We did an immunofluorescence staining of three typical neuronal markers – synapsin 1 (SYN1), MAP2 and NeuN – in *TET3*$^{-/-}$ and *MeCP2*$^{-/Y}$ and compared the staining pattern to the corresponding wt (Figure 19A). SYN1 belongs to the family of synapsins, which are neuronal phosphoproteins that are essential for modulating neurotransmitter release and therefore for proper synaptic function and plasticity.[393] MAP2 is part of the MAP2/Tau family of microtubule-associated proteins and plays a pivotal role in stabilizing the microtubules and F-actin in the dendrites and axons of neurons.[141] The function of NeuN, however, is still elusive, but it is known as an important marker of mature neurons.[140] In the wt, SYN1 immunostaining showed the expected very bright and condensed signal around the synapses. In the *TET3*$^{-/-}$ and the *MeCP2*$^{-/Y}$ iNGN-derived neurons, we also detected a signal for SYN1, but it was not as bright and condensed around the synapses compared to the wt, indicating that synaptic activity might be altered in both KO cell lines. We observed a similar staining pattern for MAP2, with very bright immunofluorescence signal in

the wt, but not in the TET3- and MeCP2-deficient iNGN-derived neurons. For NeuN, in contrast, no obvious differences between the three cell lines could be observed. Based on these observations, we aimed to quantify the differences in *SYN1* and *MAP2* gene expression at d8 using RT-qPCR. For both genes, the expression levels were significantly lower in both KO genotypes compared to the wt (p-values *SYN1*: *TET3*$^{-/-}$/wt 0.0046, *MeCP2*$^{-/Y}$/wt 0.0014; p-values *MAP2*: *TET3*$^{-/-}$/wt 0.0279, *MeCP2*$^{-/Y}$/wt 0.0384). Since MAP2 is a member of the MAP2/TAU protein family, we also checked the expression levels of *TAU*, but could not find a significant difference between the three genotypes (Figure 19B). Next, we studied the global changes of TET3- or MeCP2-deficiency in d8 iNGN-derived neurons on the DNA methylation, the gene expression and the protein expression level (Figure 19C – E). To characterize differentially methylated CpGs, the Infinium® MethylationEPIC BeadChip Array was used and all methylation sites with a differential methylation $\log_2$ fold change >|0.2| and a false discovery rate (FDR)-corrected p-value < 0.05 were considered as significant. *TET3*$^{-/-}$ and *MeCP2*$^{-/Y}$ d8 iNGN-derived neurons showed massive changes in CpG-methylation compared to the wt (Figure 19C). In the *TET3*$^{-/-}$, we detected in total 2127 differentially methylated CpGs, whereby 1613 CpGs showed less methylation and 514 showed more methylation than the wt. In the *MeCP2*$^{-/Y}$, we detected 1993 CpGs to be differentially methylated compared to the wt, with 1499 being downregulated and 494 being upregulated. Most striking, however, was the discovery that *TET3*$^{-/-}$ and *MeCP2*$^{-/Y}$ shared 1399 differentially methylated CpGs. Compared to all CpGs that were differentially methylated, this overlap was highly significant (Fisher's exact test p-value 0.0045) and there was a strong positive correlation (Pearson correlation coefficient 0.9748) regarding the overlapping differentially methylated CpGs between *TET3*$^{-/-}$ and *MeCP2*$^{-/Y}$ d8 iNGN-derived neurons. Only one of the overlapping differentially methylated CpGs showed higher methylation in the *TET3*$^{-/-}$ and lower methylation in the *MeCP2*$^{-/Y}$ compared to the wt, all the others had the same direction of methylation changes. These results indicated that the CpG methylation pattern in *TET3*$^{-/-}$ and *MeCP2*$^{-/Y}$ d8 iNGN-derived neurons is very similar and different from the wt. As CpG methylation directly influences gene expression[8], we examined whether the similarities of the DNA methylome also resulted in a similar transcriptome, which was analyzed by RNA-seq. Differential gene expression with a $\log_2$ fold change > |1| and an FDR-corrected p-value < 0.05 were considered as significant and our result shows that TET3- and MeCP2-ablation in iNGN-derived neurons have a huge effect on gene transcription. In the *TET3*$^{-/-}$, 2486 transcripts were found to be differentially expressed with 1630 downregulated and 856 upregulated transcripts. In the *MeCP2*$^{-/Y}$, 1347 transcripts were downregulated and 571 were upregulated, resulting in total in 1918 differentially expressed transcripts. Again, we observed that the majority of differentially expressed transcripts (1510) were shared among the *TET3*$^{-/-}$ and the *MeCP2*$^{-/Y}$ d8 iNGN-derived neurons (p-value Fisher's exact test < 0.0001) and differential expression of overlapped transcripts showed a strong positive correlation (Pearson correlation coefficient 0.9626). Experiments that allow us to directly link between DNA methylation and the transcriptome are ongoing. Finally, we aimed to study the differences of the proteome between the different genotypes and we considered a $\log_2$ fold change > |0.58| and an FDR-corrected p-value < 0.05 as significant. On the proteome level, however, we only detected 56 proteins that were differentially expressed in the *TET3*$^{-/-}$ d8 iNGN-derived neurons and only 54 proteins that were differentially expressed in the *MeCP2*$^{-/Y}$ d8 iNGN-derived neurons compared to the wt (Figure 19E). This finding has presumably several reasons. First, whereas transcriptome analysis via RNA-seq does not discriminate between highly abundant and less abundant transcripts, proteome analysis via LC-MS/MS does. In RNA-seq, a single mRNA is in principle sufficient for the analysis as it is amplified before the analysis.[394] This is not possible in proteomics and consequently, current proteomics techniques based on LC-MS/MS, which are becoming more sensitive with every year, still face the problem of mining low abundant proteins.[395] Moreover, gene expression is only the first part of protein biosynthesis and gives only limited information about translation efficiencies. It was reported before that protein levels are more stable than transcript levels and that

cells can compensate higher or lower transcription levels with adapted translation efficiency.[396, 397] Altered transcription levels are especially important when dynamic and fast changes of the proteome are required[396], for example in the case of neuronal activity. In this scenario, the cell with a higher transcript level of a certain protein can synthesize the protein faster when more mRNA is present. Therefore, the heavily altered transcription levels in the KO iNGN cell lines do not seem to have immediate consequences on the phenotype of the resting neurons but will probably become important when the neurons are activated, and synaptic plasticity is required. Although the changes of the proteome compared to the wt were marginal, the phenotype-copying of the *TET3*$^{-/-}$ and the *MeCP2*$^{-/Y}$ d8 iNGN-derived neurons was again obvious. Both genotypes shared 39 differentially expressed proteins and the expression of those proteins was highly positively correlated (Pearson correlation coefficient 0.9937). Among the differentially expressed proteins were several laminin subunits (beta-1 and gamma-1) that were depleted in the KOs compared to the wt. Despite the minor changes on the proteome level, several proteins that are important for neuronal functioning were differentially expressed. The cannabinoid receptor 1 (Uniprot ID P21554), which is important for neuronal development regarding neurite outgrowth, axonal pathfinding and synaptogenesis,[398] was enriched in both KOs compared to the wt. In contrast, the protein piccolo (Uniprot ID Q9Y6V0), which acts as a scaffolding protein in the active zone where neurotransmitters are released,[399] was depleted in both. The limited number of differentially expressed proteins did not allow to perform a gene ontology (GO) analysis, but these results indicate that synaptic functioning is affected in the same direction in the *TET3*$^{-/-}$ and the *MeCP2*$^{-/Y}$ d8 iNGN-derived neurons even in a state of rest.

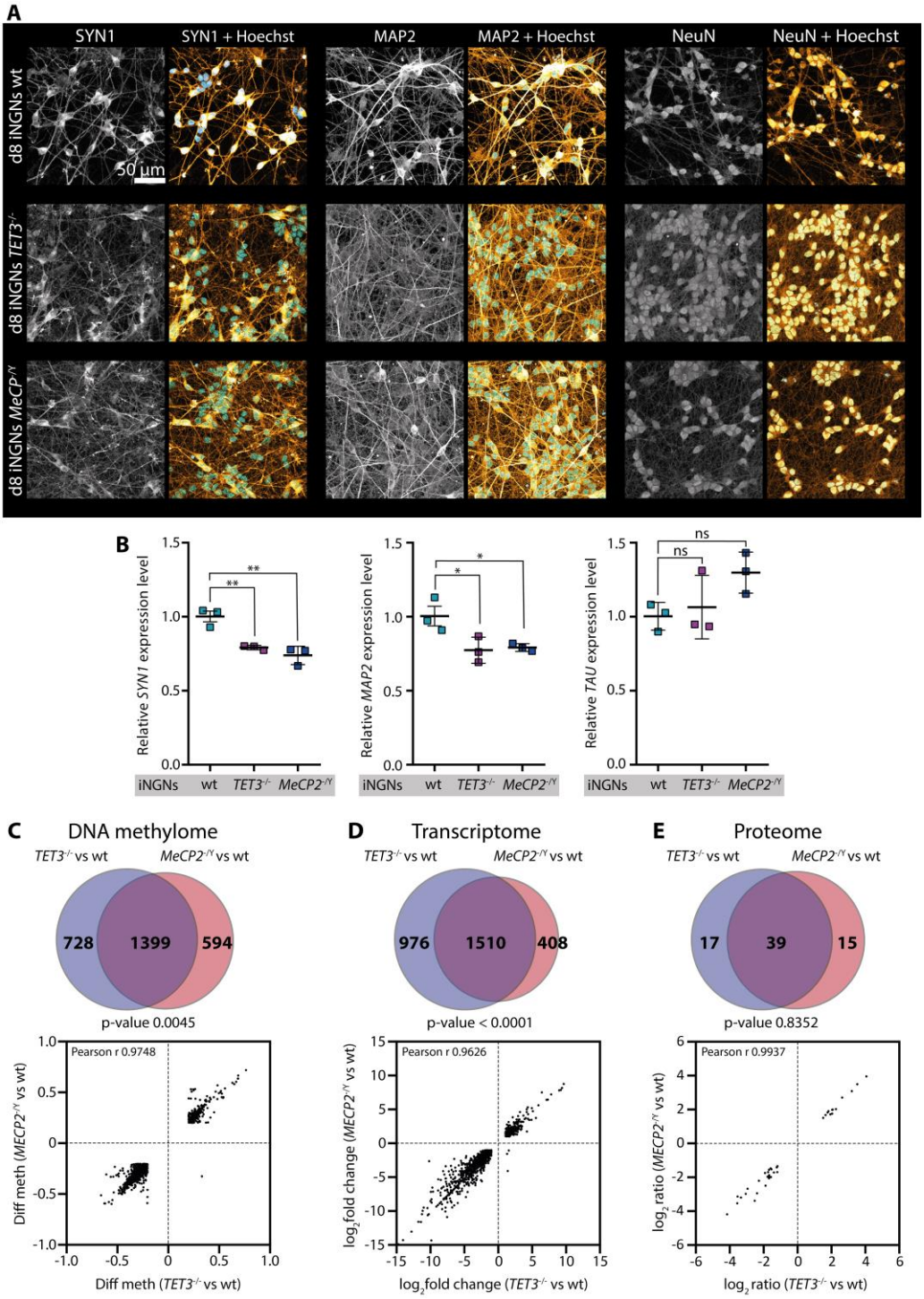


Figure 19*:* **Characterization of TET3-deficient and MeCP2-deficient iNGN-derived neurons at d8.** (A) Immunofluorescence staining against the typical neuronal markers SYN1, MAP2 and NeuN. Images show Z-stacks. (B) Comparison of the gene expression levels of *SYN1*, *MAP2* and *TAU* using RT-qPCR. Statistical analysis Ordinary one-way ANOVA with Dunnett's multiple comparisons test. (C) Venn diagram depicting the overlap between differentially methylated CpGs in *TET3*$^{-/-}$ and *MeCP2*$^{-/Y}$ as determined using Infinium MethylationEPIC BeadCHip (Fisher's Exact Test) and correlation (Pearson correlation coefficient) of overlapping differential methylated CpGs between both. (D) Venn diagram depicting the overlap between differentially expressed transcripts in *TET3*$^{-/-}$ and *MeCP2*$^{-/Y}$ (Fisher's Exact Test) and correlation (Pearson correlation coefficient) of differentially expressed transcripts between both. (E) Venn diagram depicting the overlap between differentially expressed proteins in *TET3*$^{-/-}$ and *MeCP2*$^{-/Y}$ (Fisher's Exact Test) and correlation (Pearson correlation coefficient) of differentially expressed proteins between both.

d8 = eight days after doxycycline induction, iNGNs = small molecule-inducible Neurogenin induced human pluripotent stem cell line, MAP2 = microtubule-associated protein 2, MeCP2 = methyl-CpG binding protein 2, *SYN*1 = synapsin 1, TET 3 = ten-eleven translocation enzyme 3, wt = wildtype.

ns p-value ≥ 0.05, * p-value < 0.05, ** p-value < 0.01.

GO analysis on the transcriptome level revealed that the expression of genes involved in neuronal signaling and cell migration were mainly changed in the KOs compared to the wt (Figure 20A). Upregulated genes were associated with G-protein coupled receptor signaling, action potential, neuropeptide signaling and neurotrophin production. Downregulated genes were associated with anatomical structure morphogenesis, cell adhesion and cell migration.

To learn more about the existing differences between the two KO genotypes, we directly compared the *TET3*$^{-/-}$ to the *MeCP2*$^{-/Y}$ d8 iNGN-derived neurons. As expected, the differences between them were much smaller than their individual differences compared to the wt (Figure 20B – D). We found 534 differentially methylated CpGs (Figure 20B), 449 differentially expressed genes (Figure 20C) and 68 differentially expressed proteins (Figure 20D). Interestingly, we could not find any enriched GO terms for neither the upregulated transcripts nor the upregulated proteins in the *MeCP2*$^{-/Y}$. For the differentially upregulated transcripts in the *TET3*$^{-/-}$, GO terms associated with protein targeting to the membrane and cellular compartments and RNA processing were enriched. Interestingly, on the protein level, the *TET3*$^{-/-}$ d8 iNGN-derived neurons showed altered histone protein composition compared to the *MeCP2*$^{-/Y}$. Consequently, GO terms that are all connected to chromatin organization (Figure 20E) were enriched. However, although there are changes in nucleosome and chromatin organization in the *TET3*$^{-/-}$ compared to the *MeCP2*$^{-/Y}$, this effect has only minor consequences for CpG methylation and gene expression as indicated by the massive phenotype-coping of *TET3*$^{-/-}$ and *MeCP2*$^{-/Y}$ on these levels.

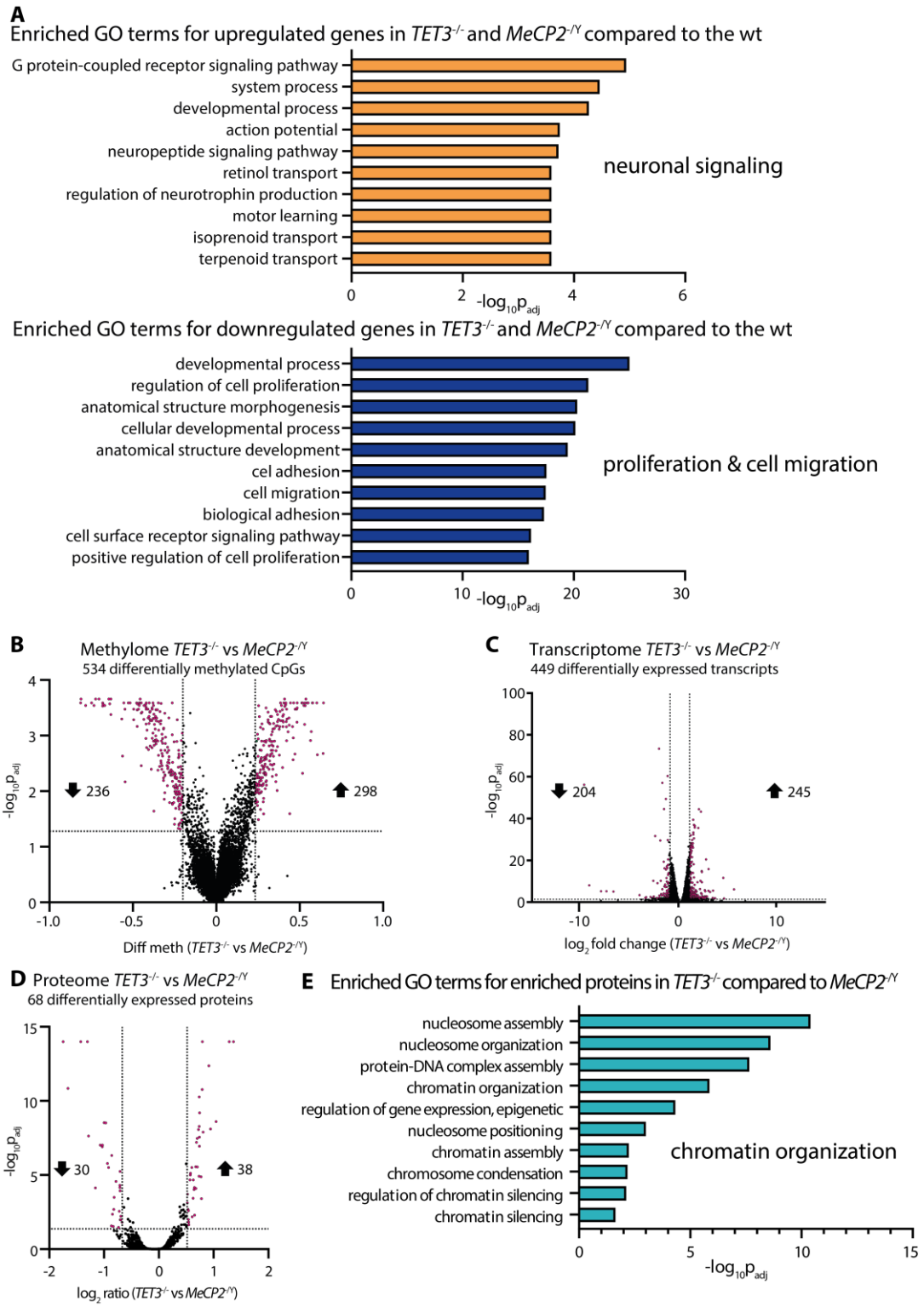


Figure 20: **GO enrichment analysis and direct comparison between *TET3*$^{-/-}$ and *MeCP2*$^{-/Y}$ d8 iNGN-derived neurons**. (A) Top enriched GO terms for the common up- and the common downregulated genes in *TET3*$^{-/-}$ and *MeCP2*$^{-/Y}$ compared to the wildtype. (B) Comparison between *TET3*$^{-/-}$ and *MeCP2*$^{-/Y}$ regarding differentially methylated CpGs. (C) Comparison between TET3$^{-/-}$ and MeCP2$^{-/Y}$ regarding differentially expressed genes. (D) Comparison between *TET3*$^{-/-}$ and *MeCP2*$^{-/Y}$ regarding differentially expressed proteins. (E) GO enrichment analysis of proteins that were enriched in the *TET3*$^{-/-}$ compared to the *MeCP2*$^{-/Y}$.

d8 = eight days after doxycycline induction, CpG = 2'-deoxy-cytidine-2'deoxy-guanosine dinucleotide, GO = gene ontology, iNGNs = small molecule-inducible Neurogenin induced human pluripotent stem cell line, MeCP2 = methyl-CpG binding protein 2, TET 3 = ten-eleven translocation enzyme 3, wt = wildtype.

After analyzing the effects of TET3-ablation or MeCP2-ablation in d8 iNGN-derived neurons separately, we wanted to learn more about their interplay on different levels. As it was already reported for two human cell lines that MeCP2 binds to the *TET3* locus, we checked if this was also the case in d8 iNGN-derived neurons. We performed MeCP2-chromatin immunoprecipitation (ChIP) in the wt and used the MeCP2-deficient neurons as a control. In the subsequent RT-qPCR, where we checked enrichment above background (= *MeCP2*$^{-/Y}$ control), we focused on three different parts of the *TET3* locus that were already reported as MeCP2 binding sites[383]. Two of them (peak 1 and 2) covered an intronic positions, one (peak 3) covered an exonic position and for all three positions, we confirmed MeCP2-binding in d8 iNGN-derived neurons (Figure 21A). Next, we studied the effects of Mecp2 on Tet3 activity using HEK293T cells for ectopic expression. In this experiment, we transfected HEK293T cells with either a plasmid coding for GFP-Tet3 or did a co-transfection of a plasmid coding for GFP-Tet3 and of a plasmid coding for Mecp2. The presence of both isoforms of Mecp2, significantly reduced Tet3-activity as indicated by the lower detected 5hmdC levels (Figure 21B – C). Differences in the GFP-Tet3 expression levels were normalized to the detected GFP signal. Interestingly, when GFP-Tet3 was co-expressed with Mecp2 that carried one of the typical RTT-associated mutations[306], we could not detect a significant reduction of the 5hmdC levels. This effect was more prominent for Mecp2E1, which is the dominant isoform in neurons. In presence of the T158M and the R133C mutations, the 5hmdC levels were on the same level as for the GFP-Tet3 only transfected HEK293T (Figure 21B). For the Mecp2E2 isoform, the 5hmdC levels were still lower compared to the GFP-Tet3 only transfected HEK293T when Mecp2E2 was mutated, but this effect was not significant (Figure 21C).

Taken together, our results show an immense phenotype-copying of the *TET3*$^{-/-}$ and the *MeCP2*$^{-/Y}$ d8 iNGN-derived neurons on the DNA methylation, the transcriptome and the proteome level. MeCP2 and TET3 interact on the protein level and MeCP2 influences TET3 activity. Additionally, MeCP2 binds the *TET3* gene locus and influences *TET3* transcription. MeCP2 downregulates TET3-activity when both proteins are co-expressed in HEK293T, possibly by occupying 5mdC. Typical mutations of MeCP2 that are found in almost 20% of RTT patients seem to abolish this effect. These findings support the idea that the interplay of TET3 and MeCP2, rather than one of the proteins alone, is essential for neuronal functioning, especially during synaptogenesis and synaptic maturation. So far, the influence of MeCP2 mutations and TET3 mutations on brain development were studied isolated from each other. Our results suggest that TET3 plays an important role in the clinical progression of RTT or might even be the cause for atypical RTT.

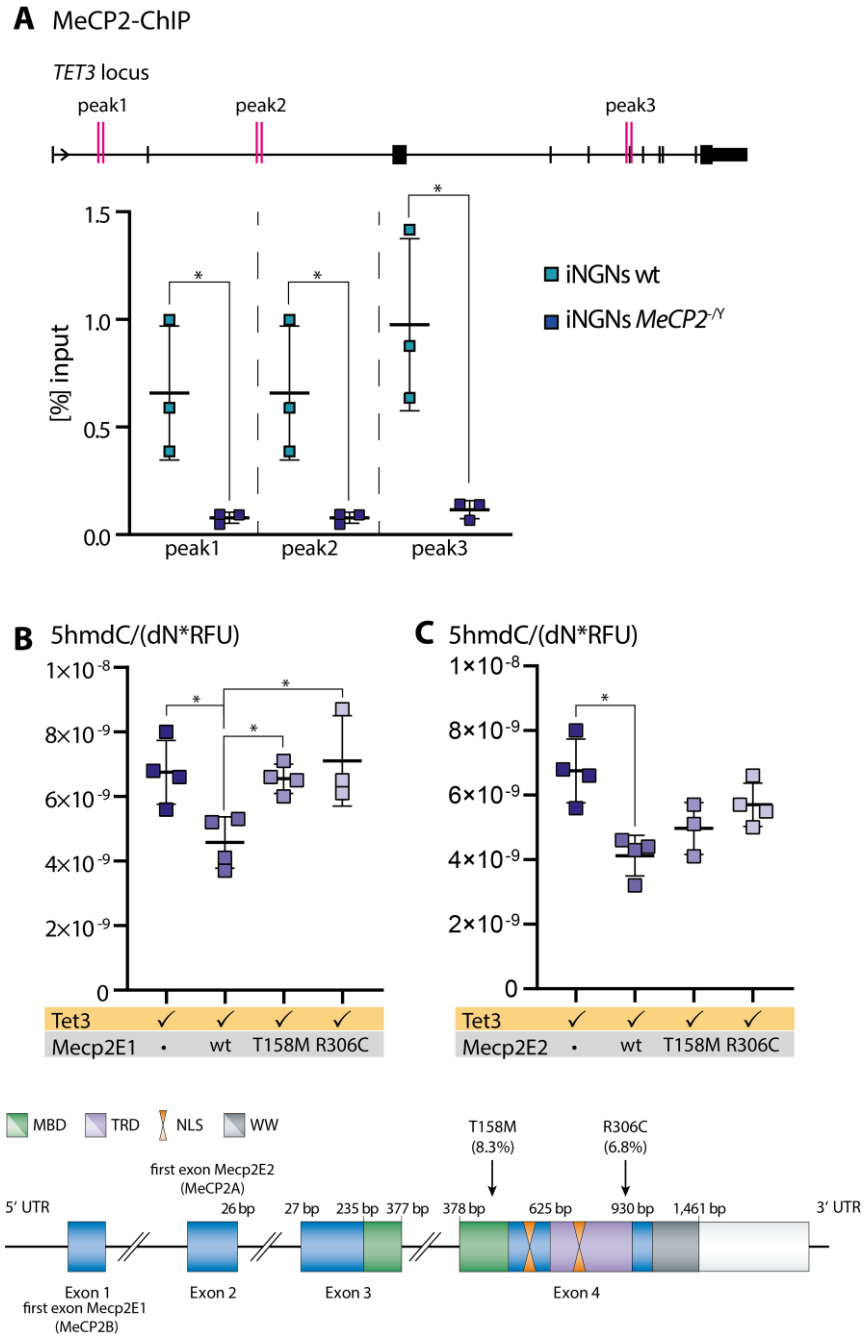


Figure 21: **MeCP2 binding to the *TET3* locus and influence of Mecp2 of Tet3 activity.** (A) ChIP-PCR results showing that MeCP2 binds the *TET3* locus. (B) To the GFP-signal normalized 5hmdC levels in GFP-Tet3 transfected HEK293T cells. The cells were either transfected with GFP-Tet3 alone or in combination with Mecp2 (isoform E1) or a typical RTT-mutant thereof. (C) To the GFP-signal normalized 5hmdC levels in GFP-Tet3 transfected HEK293T cells. The cells were either transfected with GFP-Tet3 alone or in combination with Mecp2 (isoform E2) or a typical RTT-mutant thereof. (B) – (C) Statistical analysis: Ordinary one-way ANOVA with Dunnett's multiple comparison test, only significant differences between the samples are indicated by *.

bp = base pairs, ChIP = chromatin immunoprecipitation, iNGNs = small molecule-inducible Neurogenin induced human pluripotent stem cell line, MBD = methyl-CpG binding domain, Mecp2/MeCP2 = methyl-CpG binding protein 2, R306C = arginine to cysteine at position 306, NLS = nuclear localization signal, RFU = relative fluorescent unit, T158M = threonine to methionine at position 158, Tet3/TET3 = ten-eleven translocation enzyme 3, TRD = transcription repression domain, UTR = untranslated region, wt = wildtype, WW = RNA splicing binding factor region.

ns p-value ≥ 0.05, * p-value < 0.05.

Materials and Methods

All experiments were at least repeated twice to ensure reproducibility.

Unless otherwise mentioned, culturing and transfection of HEK293T for high protein expression, culturing and differentiation of iNGNs, RT-qPCR, western blotting, gDNA and RNA isolation, UHPLC-QQQ-MS, TET3-enriched and TET3-endogenous coIP were in principle performed as described in the previous subchapter (Influence of metabolites on TET activity – Materials and Methods).

Antibodies

MAP2 antibody (Sigma-Aldrich M9942, clone HM-2, mouse monoclonal IgG): ICC (1:400)

MeCP2 antibody (Diagenode C15410052, Lot A20-0042, rabbit polyclonal IgG): western blotting (1:1000), ChIP 6 µg/reaction

MeCP2 antibody (Santa Cruz Biotechnology sc-137070, clone G-6, mouse monoclonal IgG): western blotting (1:250)

NeuN-Alexa488 antibody (Merck Millipore MAB377X, mouse monoclonal IgG): ICC (1:100)

SYN1 antibody (Cell Signaling 5297, clone D1205, rabbit monoclonal IgG): ICC (1:200)

TET3 antibody (Abiocode N1-R1092, Lot 7083, rabbit polyclonal IgG): coIP (1 µL/reaction for MS analysis, 2 µL/reaction for western blot analysis)

TET3 antibody (made by Prof. Ammer's laboratory (Tierärztliche Fakultät LMU München), C-terminal #5454, Serum 5): western blotting (1:400)

Secondary antibodies

Alexa488-anti-mouse IgG (Cell Signaling Technologies 4408): IHC (1:800)

Alexa555-anti-mouse IgG (Cell Signaling Technologies 4409): IHC (1:800)

Cy3-anti-mouse IgG (Jackson ImmunoResearch 715-165-150): IHC (1:400)

HRP-conjugated anti-mouse IgG (Sigma-Aldrich AP130P): western blotting (1:5000)

Alexa488-anti-rabbit IgG (Cell Signaling Technologies 4412): IHC (1:800)

Alexa555-anti-rabbit IgG (Cell Signaling Technologies 4413): IHC (1:800)

Cy3-anti-rabbit IgG (Jackson ImmunoResearch 711-165-152): IHC (1:400)

HRP-conjugated anti-rabbit IgG (Sigma-Aldrich A0545): western blotting (1:5000)

Expression plasmids

Empty vector pESG-iba45 (iba 5-4445-001)

Mecp2E1-MycDDK (plasmid with kanamycin resistance, CMV promoter, OriGene MR226839) (N-terminus) Mecp2-MycDDK; murine Mecp2 (UniprotKB – Q9Z2D6-2) starting from amino acid 1 of the original sequence. Mecp2E1 T158M threonine to methionine mutation at position 175; Mecp2E1 R306C arginine to cysteine mutation at position 323

Mecp2E2-MycDDK (plasmid with kanamycin resistance, CMV promoter, OriGene MR207745)

(N-terminus) Mecp2-MycDDK; murine Mecp2 (UniprotKB – Q9Z2D6-1) starting from amino acid 1 of the original sequence. Mecp2E2 T158M threonine to methionine mutation at position 158; Mecp2E2 R306C arginine to cysteine mutation at position 306

GFP-Tet3 (plasmid with ampicillin resistance, CAG promoter):
(N-terminus) eGFP – TEV cleavage site – Tet3; murine Tet3 (UniprotKB – Q8BG87-4) starting from amino acid 1 of the original sequence

Site-directed mutagenesis
To introduce the T148M or R306C mutations in the Mecp2 expression plasmids, PCR site-directed mutagenesis was used.

Primer for site-directed mutagenesis:
T158M 5′ CTC TCC CAG TTA CCA TGA AGT CAA AAT CAT TAG GGT CC 3′ (forward)
5′ GGA CCC TAA TGA TTT TGA CTT CAT GGT AAC TGG GAG AG 3′ (reverse)

R306C 5′ CTC CCG GGT CTT GCA CTT CTT GAT GGG GA 3′(forward)
5′ TCC CCA TCA AGA AGT GCA AGA CCC GGG AG 3′(reverse)

PCR reaction set up per sample:

30.34 µL	ddH_2O
10.00 µL	5x GC Buffer (New England Biolabs B0519S)
1.50 µL	DMSO
5.00 µL	dNTP mix (New England Biolabs N0447S)
0.83 µL	forward primer (from 10 µM stock solution)
0.83 µL	reverse primer (from 10 µM stock solution)
1.00 µL	Template (= 50 ng)
0.50 µL	Phusion DNA Polymerase (New England Biolabs M0530S)

Step 1	98 °C	2:00 min
Step 2	95 °C	0:30 min
Step 3	62 °C	0:20 min
Step 4	72 °C	3:00 min
- repeat Steps 2-4 22x –		
Step 5	72 °C	5:00 min

After the PCR, 10 µL of the PCR product was checked on an 1% agarose gel. The rest was stored on ice while the agarose gel was running. Agarose gel electrophoresis confirmed that in any case only one PCR product at the correct bp size was obtained. Subsequently, 1 µL of DpnI (New England Biolabs R0176S) was added per sample and samples were incubated for 1 h at 37 °C. Afterwards, the mix was used completely for transformation of chemical competent Escherichia coli (E. coli, New England Biolabs C2987I). Plasmids were isolated from positive clones and checked by sequencing for having the correct mutation.

Transfection of HEK293T cells and QQQ analysis thereof

Transfection of HEK293T cells for QQQ experiments:

The transfection was performed in 6 well plates (Sarstedt 83.3920.300). 2.5×10^5 cells were seeded per well in 3 mL of medium. 24 h after seeding, the cells were transfected using 1.5 µg of DNA per expression plasmid (combination of GFP-Tet3 and one of the Mecp2-Myc expression plasmids), 4 µL of jetPRIME and 150 µL of jetPRIME buffer. In case of GFP-Tet3 only expressing cells, 1.5 µg of pESG-iba45 were co-transfected to keep the amount of transfected DNA constant. Six hours after transfection, the medium was changed and 24 h after transfection, the cells were harvested and washed once with 1 mL of ice-cold PBS. 100 µL were used for GFP signal quantification of living cells using FACS (BD LSRFortessa; FSC 130 V, SSC 300 V, GFP 370 V log, GFP-A > 10^3, 10000 events per measurement), the rest was lysed and gDNA was isolated and analyzed by UHPLC-QQQ-MS as described previously.[376] The 5hmdC/dN were subsequently divided by the GFP signal to obtain 5hmdC/(dN*RFU).

Protein extract preparation

Whole protein extract preparations from iNGN-derived neurons:

3.5 Mio cells were plated in a P100 and differentiated as described above to obtain d8 iNGN-derived neurons. Cells were harvested using PBS to detach them from the plate and lysed using 1 mL of RIPA buffer as described in the previous subchapter.

Nuclear extract preparations from iNGN-derived neurons:

Nuclear extract from d8 iNGN-derived neurons was prepared as described in a previously published protocol.[400] A Bradford protein assay was performed according to the manufacturer´s instructions to determine the protein concentration.

Tet3-enriched coIP

Tet3-enriched coIP was performed as described above, except that 200 µg of nuclear protein extract from d8 iNGN-derived neurons were used instead of nuclear protein extract from murine brain. Tet3-enriched coIP was subsequently analyzed by LC-MS/MS.

Tet3 endogenous coIP

Tet3-endogenous coIP was performed as described above, except that 500 µg of nuclear protein extract from d8 iNGN-derived neurons were used instead of nuclear protein extract from murine brain. Tet3 endogenous coIP was subsequently analyzed by western blotting.

ICC

For ICC, 30000 iNGN stem cells were seeded in 15 µ-slide 8 well plates and differentiated as described before. Eight days post-induction, the medium was removed, and cells were washed once with PBS^+ (PBS^+ supplemented with $MgCl_2$ and $CaCl_2$, Sigma-Aldrich D8662) and ICC experiments were performed. All steps were performed in a humidity chamber and at room temperature if not otherwise stated. Cells were fixed with 4 % FA in PBS^+ for 15 min and subsequently washed three times with PBS^+. After washing, cells were permeabilized in 0.1% (v/v) Triton X-100 in PBS^+ for 10 min, washed three times with PBS^+ and subsequently blocked for 30 min with 5% (v/v) blocking reagent CB in PBS^+. The primary antibodies were diluted in PBS^+, containing 5% (v/v) CB and 0.3% (v/v) Triton X-100 and applied for 12 h at 4 °C. To remove unbound primary antibodies, cells were washed three times with 3% (v/v) CB in PBS^+ and secondary antibodies were diluted in 3% (v/v) in PBS^+ and applied for 1 h in the dark. After the incubation, cells were washed three times with PBS^+ followed by nuclei staining with Hoechst

33342 (5 µg/mL). Hoechst was applied for 10 min in the dark and cells were subsequently washed once with PBS^+. A few drops of mounting medium were applied, and ICC experiments were finally analyzed using a Leica SP8 confocal laser scanning microscope (Leica, Wetzlar).

RT-qPCR

cDNA synthesis and RT-qPCR were performed as described in the previous subchapter.

The following primer sets were used:

hACTB:	5′ GCC GCC AGC TCA CCA T 3′ (forward)
	5′ CAC GAT GGA GGG GAA GAC G 3′ (reverse)
hTET1:	5′ GCT CTC ATG GGT GTC CAA TTG CT 3′ (forward)
	5′ ATG AGC ACC ACC ATC ACA GCA G 3' (reverse)
hTET2:	5′ AAG GCT GAG GGA CGA GAA CGA 3′ (forward)
	5′ TGA GCC CAT CTC CTG CTT CCA 3′ (reverse)
hTET3:	5′ GCA AGA CAC CTC GCA AGT TC 3′ (forward)
	5′ CCT CGT TGG TCA CCT GGT TC 3′ (reverse)
hMAP2:	5′ CTG AGG CTG TAG CAG TCC TG 3′ (forward)
	5′ TGG GAG CCA GAG CTG ATT CC 3′ (reverse)
hSYN1:	5′ AGC CCT GGG TGT TTG CCC A 3′ (forward)
	5′ CCT TGA CCT TGC CCA TCC CA 3′ (reverse)
hTAU:	5′ GTT CGA AGT GAT GGA AGA TCA CG 3′ (forward)
	5′ CAG AGC CCG GTT CCT CAG 3′ (reverse)

ChIP

For crosslinking-ChIP, 3 Mio iNGN stem cells (wt or *MeCP2*$^{-/Y}$) were plated per p100 and differentiated as described in the previous subchapter. Eight days-post induction, proteins were crosslinked to DNA by dropwise addition of 0.75% FA (final concentration) in PBS to the medium. The plates were rotated gently for 10 min at room temperature and afterwards glycine was added from a 1.25 M stock solution in ddH_2O to a final concentration of 125 mM. The plates were gently rotated for additional 5 min at room temperature After incubation, cells were rinsed twice with 10 ml of ice-cold PBS. Then, 5 ml of ice-cold PBS, containing PIC, were added and dishes were scraped thoroughly with a cell scraper. Cells of the same genotype from two p100 were pooled and the cell suspension was transferred into a 15 ml tube and centrifuged for 5 min, 4 °C and 1.000 x g. The supernatant was discarded by careful aspiration, the pellet was resuspended in 500 µl of ChIP Lysis Buffer (50 mM HEPES-KOH pH = 7.5, 140 mM NaCl, 1 mM EDTA, 1% (v/v) Triton X-100, 0.1% (w/v) sodium deoxycholate, 0.1% (w/v) SDS, supplemented with PIC), transferred to 1.5 mL Eppendorf tubes and incubated for 10 min on ice. To shear DNA to an average fragment size of 200 - 1000 bp, the lysate was sonicated (Branson). 30 consecutive fragmentation steps were applied each with an amplitude of 10 % for 10 s and 10 s pause after each pulse. Samples were cooled down on ice for 20 min and again 12 consecutive fragmentation steps were applied under the same conditions. After sonication, cell debris was pelleted by

centrifugation for 10 min, 4 °C and 8.000 g. The supernatant was transferred to a new tube. To determine DNA concentration and fragment size, 50 µl of each sonicated sample was removed. The rest of the sample was snap frozen in liquid nitrogen and stored at -80 °C until ChIP.

To determine DNA concentration and fragment size, 4.8 µL of 5 M NaCl and 0.5 µL of 100 mg/mL RNase A (Qiagen 19101) were added to the 50 µl of chromatin and incubated by shaking at 65 °C overnight. To each sample 2 µL of 20 mg/mL proteinase K (Qiagen 19131) were added and incubated while shaking at 60 °C for 1 h. The DNA was purified using a Chromatin DNA IP Purification Kit (Active Motif 58002) according to the manufacturer's protocol. The purified DNA was run in a 1.5 % agarose gel with a DNA marker to determine the fragment size.

For ChIP, the prepared chromatin was thawed on ice. Per ChIP, 3.5 µg of DNA was used. Each sample was diluted with ice-cold ChIP RIPA buffer (50 mM Tris pH = 7.5, 150 mM NaCl, 2 mM EDTA, 1% (v/v) NP-40, 0.5% (w/v) sodium deoxycholate, 0.1% (w/v) SDS, supplemented with PIC) to a total volume of 550 µl and pre-cleared by incubating with 10 µl Dynabeads Protein A (ThermoFisher 10001D) for 1 h at 4 °C. The supernatant was transferred to a new tube and 5.4 µl were removed to serve as input sample. The input was stored at -20 °C until further use. The pre-cleared chromatin was incubated with 3 µl of anti-MeCP2 antibody for 1 h at 4 °C by end-over-end rotation. Per ChIP experiment, 50 µl of Dynabeads Protein A were washed three times in ChIP RIPA Buffer. After washing, the beads were resuspended in 100 µL of ChIP RIPA Buffer and single stranded salmon sperm DNA (Sigma-Aldrich D7656) was added to a final concentration of 75 ng/µl and BSA (New England Biolabs B9000S) was added to a final concentration of 0.1 µg/µl. The beads were incubated for 30 min under rotation at room temperature. After incubation, the beads were washed once with ChIP, resuspended in 100 µL of ChIP RIPA Buffer and added to the pre-cleared chromatin. To perform ChIP, the samples were incubated for 16 h at 4° C under end-over-end rotation. The next day, the samples were washed (15 min per washing step at 4° C under rotation) once with ice-cold ChIP Low Salt Wash Buffer (20 mM Tris pH = 8.0, 150 mM NaCl, 2 mM EDTA, 0.1% (w/v) SDS, 1% (v/v) Triton X-100), once with ice-cold ChIP High Salt Wash Buffer (20 mM Tris pH = 8.0, 500 mM NaCl, 2 mM EDTA, 0.1% (w/v) SDS, 1% (v/v) Triton X-100) and once with ice-cold ChIP LiCl Wash Buffer (10 mM Tris pH = 8.0, 250 mM LiCl, 1 mM EDTA, 1% (v/v) NP-40, 1% (w/v) sodium deoxycholate). For elution, 120 µl of ChIP Elution Buffer (100 mM $NaHCO_3$, 1% (w/v) SDS) was added to the samples and samples were incubated for 15 min at 30 °C under fast shaking. The supernatant was transferred into a new tube.

The input samples were thawed, and ChIP Elution Buffer was added to a final volume of 120 µl. To the eluted samples and the input samples, 4.8 µL of 5 M NaCl and 0.5 µL RNase A were added and incubated while shaking at 65 °C overnight. To each sample 2 µL proteinase K were added and incubated while shaking at 60 °C for 1 h. The DNA was purified using a Chromatin DNA IP Purification Kit according to the manufacturer´s protocol. ChIP samples and input samples were eluted in 55 µl. DNA levels were quantitatively measured by RT-qPCR as described in the previous subchapter. For each reaction 4 µl of the purified DNA was used. Data were analyzed by the Percent Input method.

GO Analysis

GO analysis was performed using GOrilla[401, 402]. "Homo sapiens" was chosen for organism, "Two unranked lists of genes (target and background lists)" was chosen for running mode. The target set was the list of genes or proteins that were significantly enriched or depleted in RNA-seq or proteomics experiments. The background set was the list of all genes or proteins detected in the experiment. All ontologies (process, function and component) were analyzed. The p-value threshold was set to 10^{-3}. Only results that had an FDR q-value < 0.05 were considered as significant.

Contributions

I designed experiments, cultured and differentiated iNGNs, performed and analyzed TET3-enriched and TET3-endogenous coIP, western blots and immunofluorescence experiments, did the cloning of MeCP2_T158M and MeCP2_R133C, performed HEK293T assays, helped with proteomics experiments, isolated gDNA, RNA and proteins, performed and analyzed UHPLC-MS-QQQ experiments, performed RT-qPCR and MeCP2-ChIP, did the statistical analysis for various experiments, the GO-analysis of the proteomics data, analyzed and interpreted data. *Dr. Victoria Splith*[2] designed experiments, designed and created the *TET3*$^{-/-}$ iNGN cell line, helped with the design of the *MeCP2*$^{-/Y}$ iNGN cell line, cultured and differentiated iNGNs, did the live cell imaging, isolated gDNA, RNA and proteins, did the statistical analysis of various experiments and the GO analysis of the RNAseq data, analyzed the DNA methylome data, performed RT-qPCR, prepared Figure 23 C-E, Figure 24 B and C, analyzed and interpreted data. *Dr. Anna Geserich*[2] designed experiments, designed and created the *MeCP2*$^{-/Y}$ iNGN cell line, cultured and differentiated iNGNs, isolated gDNA, RNA and proteins, performed immunofluorescence experiments, did the live cell imaging, did the statistical analysis of various experiments, analyzed the DNA methylome data, performed RT-qPCR and MeCP2-ChIP, analyzed and interpreted data. *Constanze Scheel*[2] helped culturing primary hippocampal neurons, cultured and differentiated iNGNs, isolated gDNA, RNA and proteins and analyzed and interpreted data. *Dilara Özdemir*[1] performed the PLA experiment and cultured primary hippocampal neurons. *Dr. Michael Stadlmeier*[1] performed and analyzed proteomics experiments. *Leander Runtsch*[1] performed and analyzed proteomics experiments. *Dr. Verena Hammelmann*[3] prepared the primary hippocampal neuron culture. *Dr. Gilles Gasparoni*[3] and *Dr. Karl Nordström*[3] performed and analyzed RNAseq experiments. *Prof. Dr. Thomas Carell*[1] and *Prof. Dr. Jörn Walter*[3] helped with study design and data interpretation. *Prof. Dr. Stylianos Michalakis*[2] supervised the whole story, designed experiments and interpreted data.

[1] Carell Group, Department of Chemistry, LMU München

[2] Michalakis Group, Department of Pharmacy, LMU München

[3] Walter Group, Genetik/Epigenetik, Universität des Saarlandes

General Materials and Methods

Standard Molecular Biology Techniques

Agarose Gel Electrophoresis

For one agarose gel, 0.75 g (1.5% gel, used for expected band size < 1000 bp) or 0.5 g (1% gel, used for expected band size between 1000 and 3000 bp) or 0.25 g (0.5% gel, used for expected band size > 3000 bp) of agarose (Eurogentec EP-0010-05) were added to 50 mL of 1x Tris-acetate-EDTA buffer (TAE buffer; 40 mM Tris pH = 8.3, 20 mM acetic acid (AcOH, 1.21 mL/L), 1 mM EDTA) and brought to the boil in a microwave until the agarose was completely dissolved. The solution was cooled down to 50 °C before 2.5 µL of peqGREEN DNA/RNA binding dye (VWR 732-3196) were added. The solution was poured into a mini horizontal electrophoresis chamber and an appropriate comb was inserted. After the gel had been developed, the comb was removed, and 0.5 x TAE buffer was added to the chamber to cover the gel completely. An appropriate amount of 6 x DNA loading buffer (New England Biolabs B7021S) was added to the samples before loading and Quick-Load 100 bp DNA ladder (New England Biolabs N0467S) or Quick-Load 1 kb DNA ladder (New England Biolabs N0468S) was used as a reference. The gel was run at 90 V for 50 min before imaging using blue light.

SDS-PAGE

Self-made polyacrylamide gels consisted of a running gel (10% acrylamide) and of a stacking gel (5% acrylamide) on top of the running gel. For pouring and running gels, the Mini-PROTEAN Tetra Cell from BioRad (0.75 mM spacer) was used. Six gels were prepared as listed in Table 1 and 2. First, the running gel was prepared and quickly poured between the glass plates, leaving ~3 cm space to the top. The solution was covered with isopropanol to provide a straight gel end. After polymerization had been completed, the isopropanol was completely removed, and the stacking gel was poured on top. The appropriate 10well comb was inserted before polymerization could occur. Gels were used immediately or stored for maximal three weeks at 4 °C in a humified atmosphere.

Table 1: **Running gel for SDS-PAGE**.

Running gel 10%	
Rotiphorese Gel 30 (Roth 3029.1)	6.7 mL
ddH_2O	7.9 mL
1.5 M Tris pH = 8.8	5.0 mL
10% (w/v) SDS	0.2 mL
10% (w/v) ammonium persulfate (APS, Sigma-Aldrich 09913)	0.2 mL
N,N,N',N'-Tetramethyl ethylenediamine (TEMED, Sigma-Aldrich T9281)	20 µL

Table 2: **Stacking gel for SDS-PAGE**.

Stacking gel 5%	
Rotiphorese Gel 30	1.7 mL
ddH_2O	6.8 mL
1.0 M Tris pH = 6.8	1.25 mL
10% (w/v) SDS	0.1 mL
10% (w/v) APS	0.1 mL
TEMED	10 µL

An appropriate amount of 4 x SDS loading buffer (Table 3) was added to the samples before they were loaded, except for the coIP samples that were already eluted in SDS loading buffer. Samples were incubated for 5 min at 92 °C and loaded directly or stored at -20 °C. After storage, the samples were incubated for additional 2 min at 92 °C before loading. SDS PAGE was performed at constant 150 V for approximately 1 h in SDS running buffer. After the run, the stacking gel was removed and when the gel was not used for western blotting, it was stained using either silver staining following a previously published protocol[403] or Coomassie staining. For Coomassie staining, the gel was incubated for 5 – 16 h in aqueous Coomassie Staining Solution (45.4 % (v/v) ethanol (EtOH), 9.2 % (v/v) AcOH, 0.25 % (w/v) Coomassie Brilliant Blue R250 (Fluka 27816); filtered before usage) and unstained in aqueous Coomassie Unstaining Solution (20% (v/v) EtOH, 10 % (v/v) AcOH) until only protein bands remained stained.

Table 3: **4 x SDS loading buffer**.

4 x SDS loading buffer	
Tris pH = 6.8	500 mg
Glycerol	5 mL
SDS	1 g
Bromophenol Blue (Sigma-Aldrich B0126)	1.25 mg
2-Mercaptoethanol (Sigma-Aldrich M6250)	2.5 mL
ddH_2O	to 50 mL

Standard Microbiology Techniques

For plasmid amplification, E. coli strains NEB 5-alpha (New England Biolabs C2987H) or NEB 10-beta (New England Biolabs C3019H) were used.

Transformation of Chemical Competent E. coli

For transformation of intact plasmids, 1 – 10 ng of plasmid DNA was used. The cells were thawed on ice and plasmid DNA was added. After additional 30 min on ice, cells were heat-shocked for exactly 45 seconds in a 42 °C warm water bath. Afterwards cells were placed on ice for 5 min and afterwards 750 µL of SOC medium (Table 4) were added. When the plasmid encoded a carbenicillin resistance cassette, the cells were incubated for 60 min at 37 °C under shaking. In case of a kanamycin resistance cassette, the cells were incubated for 90 min. Afterwards 50 – 100 µL of the cell suspension were plated on LB-agar plates (Table 5) including either carbenicillin (Duchefa C0109.0025, final concentration 100 µg/mL) or kanamycin (Carbosynth K-0250, final concentration 25 µg/mL). The plates were incubated overnight at 37 °C and were afterwards sealed with Parafilm M (Wagner & Munz 9170002) and stored at 4 °C for up to four weeks.

Table 4: **SOC medium**.

SOC medium pH = 7.5	
Pepton (Roth 8986.3)	20 g/L
Yeast extract (Roth 2363.2)	5 g/L
$MgCl_2$	0.96 g/L
Glucose (AppliChem A1422.1000)	3.6 g/L
NaCl	0.5 g/L
KCl (Grüssing 12008)	0.186 g/L

Table 5: **Recipe for LB agar without antibiotics**.

LB agar pH = 7.5	
Pepton	10 g/L
Yeast extract	5 g/L
NaCl	5 g/L
Agar-agar (Roth 5210.2)	15 g/L

All media were autoclaved before usage. In case of SOC the medium was autoclaved without glucose and afterwards, sterile filtered glucose from a 1 M stock solution was added. For the preparation of LB agar plates, the agar was dissolved by heating up the solution. The antibiotics were added after the agar was dissolved and the solution cooled down to 55 °C and afterwards the plates were poured quickly in sterile 10 cm petri dishes.

DNA Plasmid Preparations

For overnight cultures and plasmid amplification, one successfully transformed clone was picked and added to 4 mL of antibiotics-containing LB medium (Table 6) for small-scale amplification or 100 mL of antibiotics-containing LB medium for large-scale amplification. For carbenicillin a final concentration of 100 µg/mL and for kanamycin a final concentration of 25 µg/mL were used. If the plasmids were only needed for re-transformation of E. coli small-scale amplification was chosen. If the plasmids were needed for transfection of mammalian cells, large-scale amplification was chosen. The cultures were incubated for ~15 h at 37 °C under shaking. Afterwards, cells were harvested at 6000 x g for 5 – 10 min. The pellets were either used directly for plasmid isolation using Monarch Plasmid Miniprep Kit (New England Biolabs T1010L) for small-scale applications or ZymoPURE II Plasmid Midiprep Kit (Zymo Research D4201) for large-scale applications. Plasmid isolation was performed following the instructor's manual. If pellets were stored, they were washed with PBS and stored at -20 °C.

Table 6: **Recipe for LB medium without antibiotics**.

LB medium pH = 7.5	
Pepton	10 g/L
Yeast extract	5 g/L
NaCl	5 g/L

All media were autoclaved before usage. Antibiotics were added directly before usage.

nature protocols

PROTOCOL

https://doi.org/10.1038/s41596-018-0094-6

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

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

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

Introduction

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

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

Determination of the exact amount of all noncanonical DNA nucleosides (epigenetic nucleosides or DNA lesions) is consequently an important task and requires a fast, ultrasensitive and highly reproducible approach.

[1]Center for Integrated Protein Science Munich (CiPSM), Department of Chemistry, Ludwig-Maximilians-Universität München, Munich, Germany. [2]These authors contributed equally: Franziska R. Traube, Sarah Schiffers, Katharina Iwan. *e-mail: thomas.carell@lmu.de

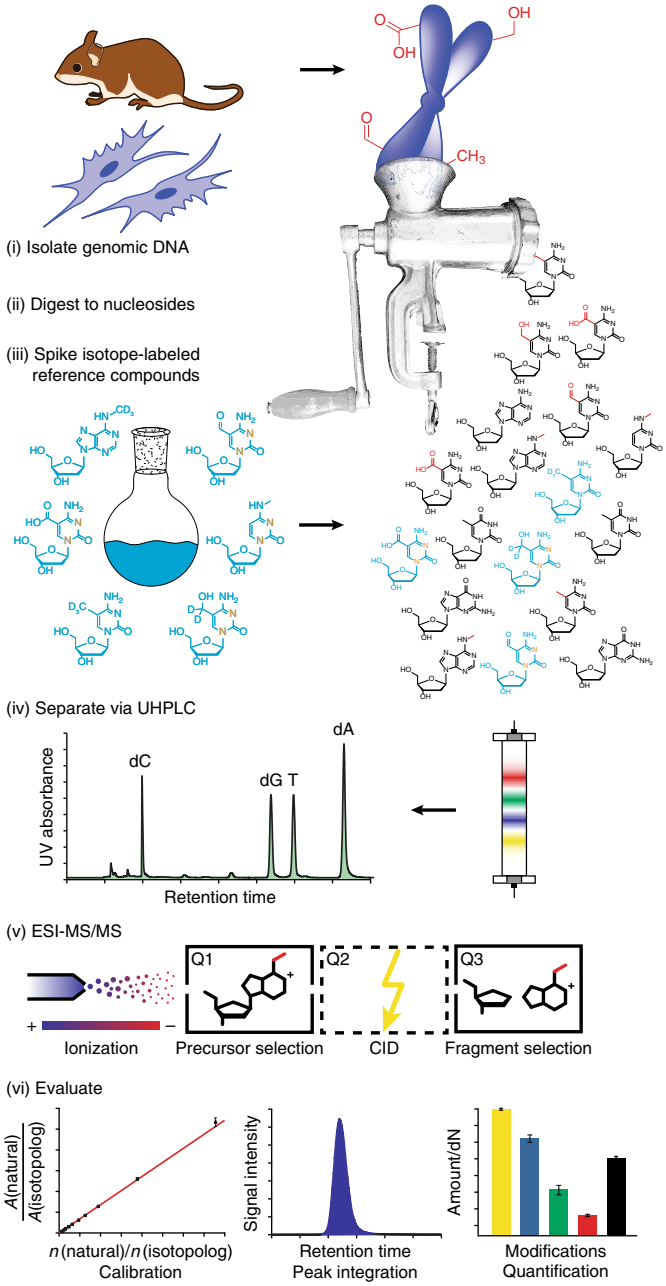


Fig. 1 | Procedure overview. General workflow for rapid DNA extraction, digestion and ultrasensitive quantification of the obtained nucleosides via UHPLC-QQQ-MS. T, deoxythymidine.

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

Development of the protocol

Quantification of very abundant DNA modifications, such as m^5dC in vertebrates, has been possible for decades without the need for ultrasensitive quantification methods[31]. Owing to their abundance, small quantification errors do not affect interpretation of the data. However, accurate quantification of DNA nucleosides can be challenging when the modification is very rare. Contaminations at the cell culture level or during tissue dissection are a particular problem if they go unnoticed. These contaminations include bacteria and fungi, which often carry DNA modifications that might be rare or not present in mammals and can have a substantial effect on the resulting values. Therefore, the biological material that is investigated must be carefully tested for the presence of such contaminations. With respect to the investigation of DNA lesions, it must be noted that lesions are also formed during DNA isolation, for example, by releasing reactive oxygen species in the course of cell lysis. Most problematic are unspecific oxidations and deaminations. To overcome these problems, addition of butylated hydroxytoluene (BHT) and deferoxamine (DFOA) during isolation is essential. BHT is a radical scavenger that prevents DNA from being oxidized, and DFOA inhibits deaminases during the isolation process[11,32].

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

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

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

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

Applications of the method

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

Comparison with other methods

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

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

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

Chromatographic separation

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

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

Detectors

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

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

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

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

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

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

Experimental design

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

The Procedure provides detailed steps for two methods:

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

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

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

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

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

Internal calibration

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

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

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

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

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

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

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

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

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

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

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

The limit of detection

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

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

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

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

Calibration curves

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

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

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

Evaluation of UHPLC–MS/MS data

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

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

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

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

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

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

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

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

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

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

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

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

$$\begin{gathered}
\mathrm{dC[pmol]/dG[pmol]} \times 100 = \%\mathrm{dC}, \\
\mathrm{m^5dC[pmol]/dG[pmol]} \times 100 = \%\mathrm{m^5dC}, \\
\mathrm{hm^5dC[pmol]/dG[pmol]} \times 100 = \%\mathrm{hm^5dC}, \\
\cdots \\
\sum(\%\mathrm{dC};\ \%\mathrm{m^5dC};\ \%\mathrm{hm^5dC};\ldots) = \%\mathrm{dC^{total}}.
\end{gathered}$$

$$\begin{gathered}
\%\mathrm{dC}/\%\mathrm{dC^{total}} = \mathrm{dC}/\mathrm{dG^{correct}}, \\
\%\mathrm{m^5dC}/\%dC^{\mathrm{total}} = \mathrm{m^5dC}/\mathrm{dG^{correct}}, \\
\%\mathrm{hm^5dC}/\%\mathrm{dC^{total}} = \mathrm{hm^5dC}/\mathrm{dG^{correct}}, \\
\cdots \\
\mathrm{m^5dC}/\mathrm{dG^{correct}} \times 0.21 = \mathrm{m^5dC}/\mathrm{dN}, \\
\mathrm{hm^5dC}/\mathrm{dG^{correct}} \times 0.21 = \mathrm{hm^5dC}/\mathrm{dN}, \\
\cdots
\end{gathered}$$

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

Level of expertise needed to implement the protocol

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

Limitations

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

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

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

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

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

There are also some challenges when performing the measurements, as not every rare nucleoside of interest can be quantified with every setup. Sometimes it may not be possible to measure all nucleosides with the same setup and conditions. If no further optimization is possible, one needs to process the sample with two different methods, but this of course requires increased instrument time and it also requires more material. For nucleosides that are quantified using the UV trace (dC and dG), it is critical that the peaks be baseline separated. For MS analysis, chromatographic separation of nucleosides is mostly not necessary because of the separation in the mass spectrometer according to their unique *m/z* values and fragmentation patterns. Only when analyzing samples that might contain isomers, e.g., m^5dC and m^4dC, which have identical precursor and product ions, chromatographic separation becomes necessary in order to clearly determine the identity of the detected signal.

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

Materials

Biological materials

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

Reagents

- 2-Mercaptoethanol (βME, CAS no. 60-24-2; Sigma-Aldrich, cat. no. M3148-25mL) **! CAUTION** 2-Mercaptoethanol is toxic, so avoid exposure.
- Acetonitrile (MeCN, 99.95% (vol/vol), LC–MS grade; Roth, cat. no. AE70.2) **▲ CRITICAL** Sensitivity of the mass spectrometer might vary when a different supplier is used. Check new suppliers carefully.
- Antarctic phosphatase (New England Biolabs, cat. no. M0289L)
- BHT (CAS no. 128-37-0; Sigma-Aldrich, cat. no. B1378-100G) **▲ CRITICAL** To keep background oxidation at a minimum, it is recommended to store BHT powder under vacuum.
- Benzonase nuclease, 10 KU (VWR, cat. no. 70746-3)
- Blood & Cell Culture DNA Midi Kit (Qiagen, cat. no. 13343)
- Degradase Plus (Zymo Research, cat. no. E2021)
- DFOA (Sigma-Aldrich, CAS 138-14-7, cat. no. D9533-1G)
- Dimethyl sulfoxide (DMSO, CAS no. 67-68-5; Acros, cat. no. 327182500)
- DNA pre-wash buffer (Zymo Research, cat. no. D3004-5)
- Dulbecco's PBS without $MgCl_2$, $CaCl_2$ (DPBS; sterile filtered, suitable for cell culture; Sigma-Aldrich, cat. no. D8537-500 mL)
- EDTA disodium salt (NA_2[EDTA]; CAS no. 6381-92-6; VWR, cat. no. 33600.267)

- Formic acid (CAS no. 64-18-6; Fluka Honeywell Chemicals, cat. no. 94318-50 mL-F) **! CAUTION** Formic acid is highly corrosive and can lead to severe burns when inhaled. Use only in a fume hood or in a highly ventilated area and protect your skin and eyes carefully.
- gDNA wash buffer (Zymo Research, cat. no. D3004-2)
- Genomic lysis buffer (Zymo Research, cat. no. D3004-1) **▲ CRITICAL** Do not add βME in advance.
- Glycerol (CAS no. 56-81-5; Roth, cat. no. 3783.2)
- Magnesium chloride hexahydrate ($MgCl_2 \cdot H_2O$, CAS no. 7786-30-3; Merck, cat. no. M8266-100g)
- Methanol (MeOH; 99.9% (vol/vol), LC–MS grade VWR, cat. no. HONC34966-1L) **! CAUTION** Methanol is toxic, so avoid exposure. **▲ CRITICAL** Mass sensitivity might vary when a different supplier is used. Check new suppliers carefully.
- Nuclease S1 from *Aspergillus oryzae* (Merck, cat. no. N5661-50 KU) **▲ CRITICAL** Stock solution should be stored at −20 °C and kept on ice while making aliquots for usage.
- Parafilm M (4 × 125 inches, clear; Bemis, cat. no. 52858-000)
- Phosphodiesterase I from *Crotalus adamanteus* venom (Abnova, cat. no. P5263)
- PicoGreen dsDNA Assay Kit (Thermo Fisher Scientific, cat. no. P7589)
- Buffer RLT (Qiagen, cat. no. 79216)
- RNase A (100 mg/mL, 7,000 U/mL; Qiagen, cat. no. 19101)
- Sodium chloride (NaCl; CAS no. 77-86-1; Bernd Kraft, cat. no. 10724344)
- Tetrahydrouridine (THU; CAS no. 18771-50-1; Merck Millipore, cat. no. 584222)
- Tris(hydroxymethyl)aminomethane (Tris base; CAS no. 77-86-1; Fisher Scientific, cat. no. BP152-5)
- Water, LC–MS grade (Honeywell, cat no. 39253-1L) **▲ CRITICAL** Mass sensitivity might vary when a different supplier is used. Check new suppliers carefully.
- Zinc sulfate (ZnSO , CAS no. 7446-20-0; Grüssing, cat. no. 14039)

Nucleosides

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

Equipment

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

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

Software

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

Reagent setup

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

Internal standard mastermix

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

1,000× BHT

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

1,000× DFOA

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

Lysis buffer

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

RNase wash buffer

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

Washing of stainless-steel beads

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

Phosphodiesterase I buffer

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

Snake venom phosphodiesterase I

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

$ZnSO_4$ stock solution

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

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

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

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

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

Procedure

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

1 Dissolve a small amount (typically 1-2 mg in 1,000 µL) of each desired nucleoside (natural and isotopically labeled) in a 1.5-mL centrifuge tube in water.

2 Measure the absorption of the solution at the respective wavelength for the maximum absorption of the nucleoside on a photometer.
▲ **CRITICAL STEP** The extinction coefficient must be known for the respective nucleoside (see Supplementary Table 6 for the extinction coefficients of m^5dC, m^4dC, hm^5dC, f^5dC, ca^5dC and m^6dA). If it is unknown, use microscales (Box 1) instead. When using a photometer, make sure that your absorption is within the linear range.

3 Determine the concentration c of the nucleoside using the Beer–Lambert law:

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

with wavelength-dependent extinction E_λ, wavelength-dependent extinction coefficient ε_λ and path length d.

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

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

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

Preparation of ISTD mastermix ● Timing 1-1.5 h

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

5 Thaw all labeled nucleosides needed for the ISTD mastermix for at least 30 min at room temperature, vortex rigorously and spin down (5,000*g*, room temperature, 3 s).
▲ **CRITICAL STEP** If the concentrations of the labeled nucleosides do not match the concentration needed for the ISTD mastermix, dilute them further (or if the concentration is too low, repeat Steps 1-4 for the desired nucleoside).

6 Prepare the ISTD mastermix.
▲ **CRITICAL STEP** If your ISTD mastermix deviates from the one given in the 'Reagent setup' section, consider adjusting the solvent, column and method for the UHPLC–MS/MS in Steps 7 and 8.
■ **PAUSE POINT** If the ISTD mastermix is not evaluated at once, store it at −20 °C for up to 1 year.

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

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

7 Set up the UHPLC–MS/MS system by installing the Poroshell 120 SB C8 column and preparing new buffers for LC. Therefore, add 75 µL of formic acid to a full 1.0-L bottle of MS-grade water (solvent A) and 187.5 µL of formic acid to a full 2.5-L bottle of MS-grade acetonitrile (solvent B). (Alternatively, if solvent B is MeOH, use a full 1.0-L bottle of MS-grade MeOH and add 75 µL of formic acid.)

8 Attach the bottles to the UHPLC system (solvent A to port A, solvent B to port B) and purge at 50% A/50% (vol/vol) B for 5 min with a flow of 5 mL/min. To provide reproducible separation efficiency, the columns must be equilibrated. Therefore, you need to perform 20 chromatographic runs using the method for epigenetic modifications (Table 1) without injection for new columns and three runs before measurement of the first sample per set. The flowrate is 0.35 mL/min, the pressure is 600 bar, the temperature of the column oven of the UHPLC is 35 °C and the gas temperature is 80 °C. The UV detector monitors an absorption signal at the 260-nm wavelength. (Alternatively, if the method for m^6dA and m^4dC is used, see Supplementary Table 7. The temperature of the column oven of the UHPLC is 30 °C; flowrate, pressure and gas temperature are not changed.)

9 If this is the first ever produced ISTD mastermix, add 12 µL to an HPLC vial and analyze it three times, each with an injection volume of 3 µL. Evaluate the ISTD mastermix according to Steps 49–53 and Steps 55 and 56. If a previous ISTD mastermix is available, add 12 µL of the previous one and 12 µL of the new one each to separate HPLC vials. Measure both three times, alternating with an injection volume of 3 µL, and evaluate the respective data. Determine the resulting areas for each labeled nucleoside, calculate the mean of this technical triplicate and evaluate the deviation of the mean as a percentage between the previous ISTD mastermix and the new one.
▲ **CRITICAL STEP** The deviation of the resulting areas must be smaller than 5% for each labeled nucleoside in order for this ISTD mastermix to qualify for exact quantification.

10 For the sensitivity check, add a few microliters ($3 + 3 \times n$ µL, n = number of performance checks) of the ISTD mastermix to an HPLC vial.

11 Measure 3 µL of the ISTD mastermix with the provided measuring method, integrate the areas of each labeled nucleoside and compare the areas of the labeled nucleosides and the corresponding SNRs from this measurement with those of the measurements from Step 9.
? TROUBLESHOOTING

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

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

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

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

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

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

Calculation of calibration curves ● Timing 1 d

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

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

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

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

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

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

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

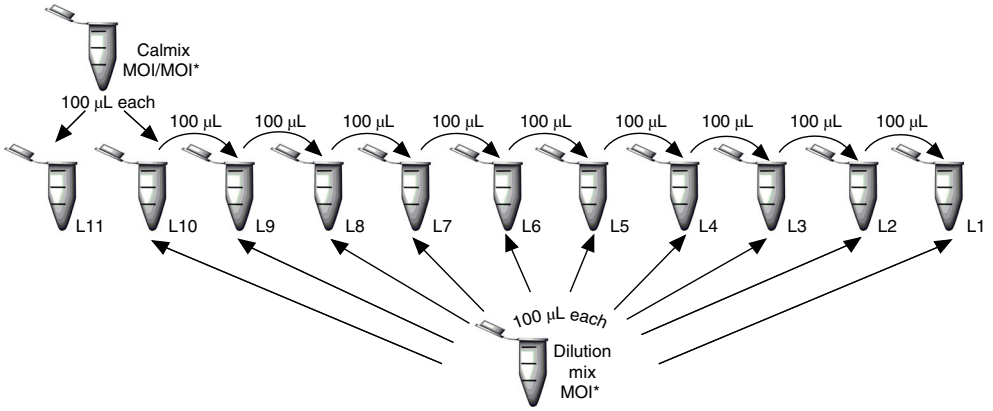


Fig. 2 | Calibration curve preparation. Pipetting scheme for generating a calibration curve of MOI and MOI* with a fixed specific amount of MOI*. L, level.

Table 4 | Example of evaluation of a calibration curve for m^4dC

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

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

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

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

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

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

24 If the accuracy check fails, remove the highest or lowest data point and repeat Steps 22 and 23 with only the remaining data points. Repeat this step until only data points with sufficient accuracy make up the linear equation. A valid linear equation must consist of at least five consecutive levels.

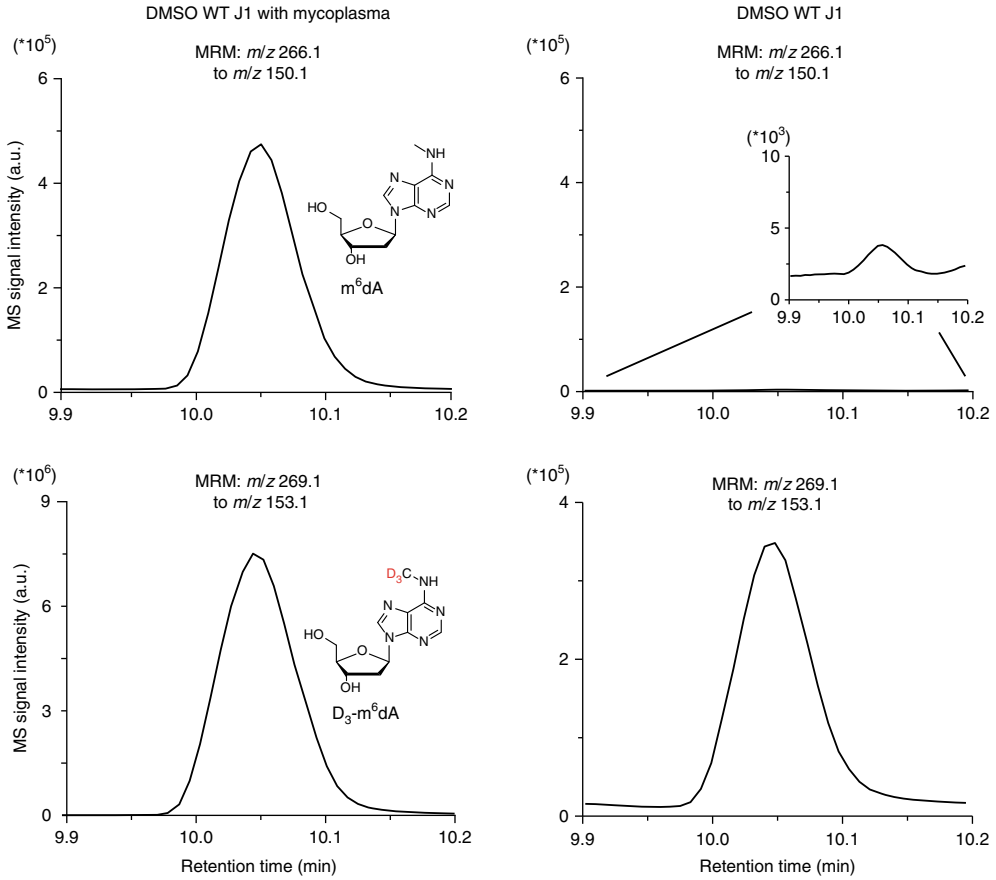


Fig. 3 | Exemplary impairment of results by mycoplasma contamination. Chromatograms with nucleoside signals for the internal standard D_3-m^6dA (lower row) and natural m^6dA (upper row) in 100 ng of DNA from WT J1 mouse embryonic stem cells (mESCs) with a mycoplasma contamination (left column) or 10 µg from mycoplasma-free DNA of WT J1 cultures (right column). a.u., arbitrary units; WT, wild-type.

Isolation of gDNA ● Timing 1.5-3 h

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

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

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

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

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

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

? TROUBLESHOOTING

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

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

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

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

? TROUBLESHOOTING

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

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

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

? TROUBLESHOOTING

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

? TROUBLESHOOTING

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

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

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

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

? TROUBLESHOOTING

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

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

Sample ID	*c* (ng/µL)	*m* (µg)	*V* (H_2O) (µL)	*V* (DNA) (µL)	Sample no.
1_Blank 1	—	—	35.0	0	1
2_Blank 2					2
3_Blank 3					3
4_1.DNA_1	444.61	10.0	12.5	22.5	4
5_1.DNA_2					5
6_1.DNA_3					6
7_2.DNA_1	405.34	10.0	10.3	24.7	7
8_2.DNA_2					8
9_2.DNA_3					9

V, volume.

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

1 Produce a single technical replicate, instead of a technical triplicate as described in Step 37, by digesting 1–5 µg of DNA. Then continue with Step 38.
2 Check if all requirements of the protocol are fulfilled:
(i) Complete digestion of DNA,
(ii) All MOIs and MOI*s are detectable,
(iii) All MOIs can be evaluated with the existing calibration curves.

If there are problems with (i), your sample contains impurities (e.g., salt) or shows an unfavorable pH. See the Troubleshooting section (incomplete digestion).
If there are problems with (ii):

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

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

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

DNA digestion ● Timing 7 h or overnight

37 Create a digestion sheet (Supplementary Methods 3) to easily calculate the required volume of DNA and water per sample. Each DNA sample should be processed from now on as a technical triplicate. Your sheet starts with samples 1–3 as digestion blanks, which contain no DNA, but to which mastermixes 1 and 2 will be added. For all the other samples, DNA and water should have a total volume of 35 µL (Table 5). Number the samples consecutively, as these will help you to identify your samples afterwards.
▲ **CRITICAL STEP** The volume of DNA must not exceed 35 µL per sample.

38 Start with the calculated volume of water, and add the DNA according to the digestion sheet.
▲ **CRITICAL STEP** Use fresh water for each digestion. Autoclaved water or water that has been stored for a long time might contain dust or additional contaminations that could impair the digestion or the measurements.
▲ **CRITICAL STEP** If this is the first digestion of your DNA sample and/or you are interested in a new DNA modification, you should consider testing your digestion protocol (Box 2). This may help to determine the appropriate addition of labeled nucleosides, e.g., spiking and the proper amount of DNA in general for fulfilling the requirements of the calibration curves.

39 Create a pipetting scheme for your mastermixes 1 and 2, using our provided Excel template (Supplementary Methods 3). Calculate the number of DNA samples (Step 37) you want to digest and add three samples more as spares.
▲ **CRITICAL STEP** Instead of following the two-step digestion protocol, which was designed to give an optimal SNR in the MS, you can use Degradase Plus from Zymo Research, which requires only one digestion step. See Box 3 for details.

Box 3 | DNA digestion with Degradase Plus (replaces Steps 39–42) ● Timing 5 h

1 Create a pipetting scheme for the Degradase mastermix using our provided Excel template for Degradase digestion (see the tabs for 'Degradase digestion' in Supplementary Methods 3). Calculate the number of DNA samples and add at least one more sample as a spare.
2 Prepare the Degradase mastermix, add 15.0 µL to each sample and incubate for 4 h at 37 °C.

When using Degradase Plus, adjust the enzyme concentration if >5 µg of DNA/sample must be digested (Supplementary Methods 3). The additional use of benzonase is required for residual RNA content within your DNA sample. The enzyme mixture in Degradase Plus cannot digest RNA. If benzonase is not used, this will lead to an overpressure in the LC device due to residual undigested RNA, and subsequent clogging of the column. However, there is no need to add the enzyme benzonase in the case of digestion of pure DNA, e.g., synthetic DNA oligonucleotides.

40 Prepare mastermix 1, then add 7.5 µL of it to each sample and incubate for 3 h at 37 °C.
41 Before the incubation time for Step 40 runs out, start preparing mastermix 2.
42 Add 7.5 µL of mastermix 2 to each sample and incubate for 3 h at 37 °C.
▲ **CRITICAL STEP** If >5 µg of DNA/sample is digested, incubate overnight at 37 °C.
■ **PAUSE POINT** Samples can be stored at 4 °C (overnight) or at −20 °C for long-term (up to 5 years) storage.

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

43 Tape an AcroPrep Advance 96-well Supor plate (0.2 µm) to a skirted 96-well PCR plate.
44 After the digestion, pipette the full 50 µL of digestion mixture for each sample into a separate well of the assembled filtration plate.
45 Cover the filtration plate with Parafilm M and secure it with tape.
▲ **CRITICAL STEP** Water should not enter the plate from an external source, but the Parafilm M should not seal the plate airtight.
46 Centrifuge the filtration plate, using a plate rotor for 35 min at 3,220*g* at 4 °C.
? TROUBLESHOOTING
47 Transfer 43 µL of the filtered digestion mixture to an HPLC vial and close the vial with an HPLC vial cap. If the volume of the filtered digestion mixture is <43 µL, note which volume is transferred to the vial.
▲ **CRITICAL STEP** The unused wells of the filtration plate can be used for later filtrations. For storage, seal the filtration plate, which is taped to the 96-well PCR tube plate, with Parafilm M and store it at room temperature for an infinite amount of time.
48 For each sample from Step 47, inject 39 µL of sample into the UHPLC–MS/MS system when the machine is ready. Record the signal of the MOIs with the relevant method for your MOIs (Supplementary Methods 4 for epigenetic modifications or Supplementary Methods 5 for m^6dA- and m^4dC-sensitive or Supplementary Methods 6 for m^6dA- and m^4dC-insensitive). For details about the measurement methods, see Table 1 and Supplementary Tables 3, 5 and 7.
▲ **CRITICAL STEP** Inject 3 µL less than the transferred volume, if the transferred volume is <43 µL.
? TROUBLESHOOTING

Evaluation ● Timing 20 min–1.5 h

49 Create a new folder called 'QuantResults' in the folder with the measured data.
50 Open the program QQQ Quantitative Analysis and click on 'File' > 'New Batch' and label the batch according to your experiment name/number.
51 Import all samples of interest.
52 Set up a method for analysis. Click on 'Method' > 'New' > 'New Method' under 'Acquired MRM Data' > 'Select a sample which was measured with this method' > 'Method Setup Tasks' > 'MRM Compound Setup'. Add UV, UV-dG and UV-dC as new compounds in the table (right-click on the table). Set the time segment 'TS' to '3' and the 'Type' to 'Target'. Enter the appropriate RT in the 'RT setup'. For UV, set it to 6 min with a 'Left RT Delta' of 6 min and a 'Right RT Delta' of 6 min. Go next to 'ISTD Setup'. In the 'ISTD conc.' column, add '1' for the ISTD concentration, make a tick in the 'ISTD Flag' column for each ISTD and select the appropriate ISTD for each target in the 'ISTD Compound Name' column. Finally, select 'Advanced Tasks' > '2D Compound Setup' and then select 'VWD' for the compounds UV, UV-dC and UV-dG. Click on 'Validate' to ensure that

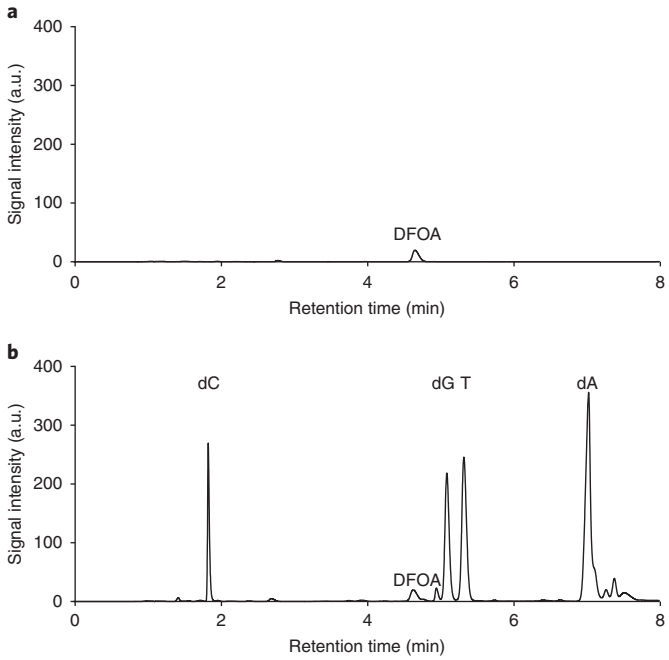


Fig. 4 | Chromatogram of different UV traces as represented in the QQQ quantitative analysis program. a,**b**, The chromatogram of a digestion blank sample (**a**) and the chromatogram of a gDNA sample from mESC line J1 (**b**).

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

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

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

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

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

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

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

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

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

? TROUBLESHOOTING

57 Use the provided Excel sheet (Supplementary Methods 1 or 2) for the evaluation of your data. For this, you need to copy the values of the peak areas (for A (MOI) and A^* (MOI*)) resulting from the

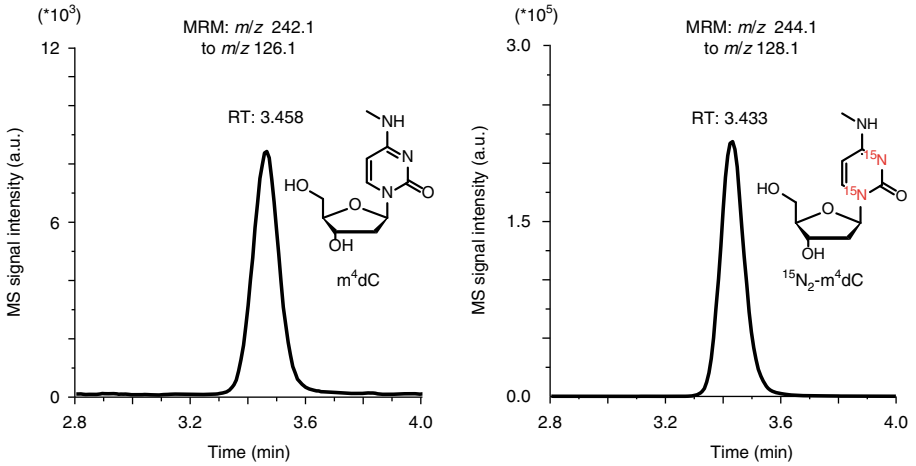


Fig. 5 | Chromatograms for m4dC and 15N2-m4dC. The left panel shows the nucleoside signal for the natural m^4dC and the right panel shows the nucleoside signal for the internal standard $^{15}N_2$-m^4dC.

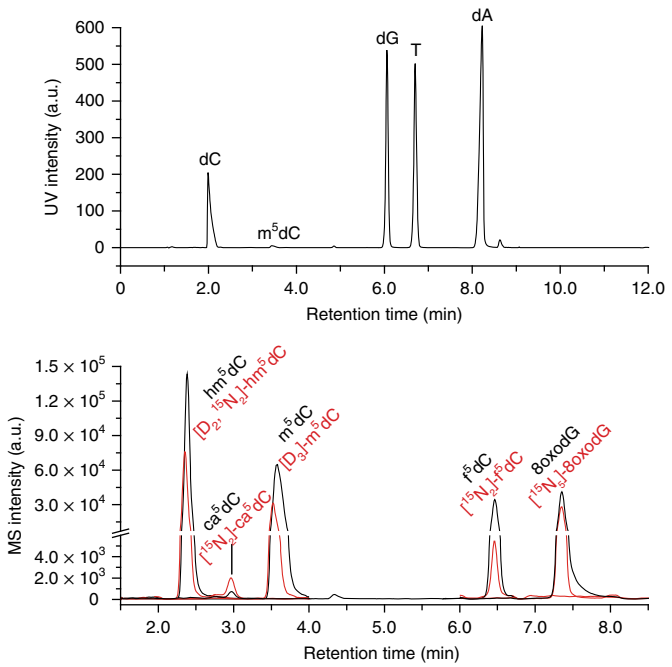


Fig. 6 | Chromatogram showing the analysis of hm^5dC, ca^5dC, m^5dC, f^5dC and 8oxodG with their corresponding isotopically labeled standards. The upper chromatogram shows the UV trace and the lower chromatogram shows the MS trace[11].

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

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

Troubleshooting

Troubleshooting advice can be found in Table 6.

Table 6 | Troubleshooting table

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

Timing

Steps 1–4, general procedure for dissolution and dilution of nucleosides: 30 min
Steps 5–11, preparation of ISTD mastermix and UHPLC–MS/MS system, and performance of sensitivity checks: 2–3 h
Steps 12–24, calculation of calibration curves: 1 d
Steps 25–36, isolation of gDNA: 1.5–3 h
Steps 37–42, DNA digestion: 7 h or overnight
Steps 43–48, sample preparation for UHPLC–MS/MS analysis: 1 h
Steps 49–57, evaluation: 20 min–1.5 h
Box 1, use of microscales for dilution of nucleosides: 15–30 min
Box 2, testing digestion conditions for an unknown sample: 8 h
Box 3, DNA digestion with Degradase Plus: 5 h

Anticipated results

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

The protocol described above details the steps needed to quantify DNA modifications and lesions such as m^5dC, hmdC or 8oxodG. The levels of 8oxodG in mammalian cells are probably between 0.07 and 0.9×10^{-6} 8oxodG/dN, based on the average of the results from several laboratories and various methods[68]. The background level of 8-oxodG has been estimated at 0.5 lesions per 10^{-6} dN[69,70]. Depending on the modification, there are big differences in the abundance of certain modifications between different cell lines or culture conditions. Whereas the abundance of m^5dC for most

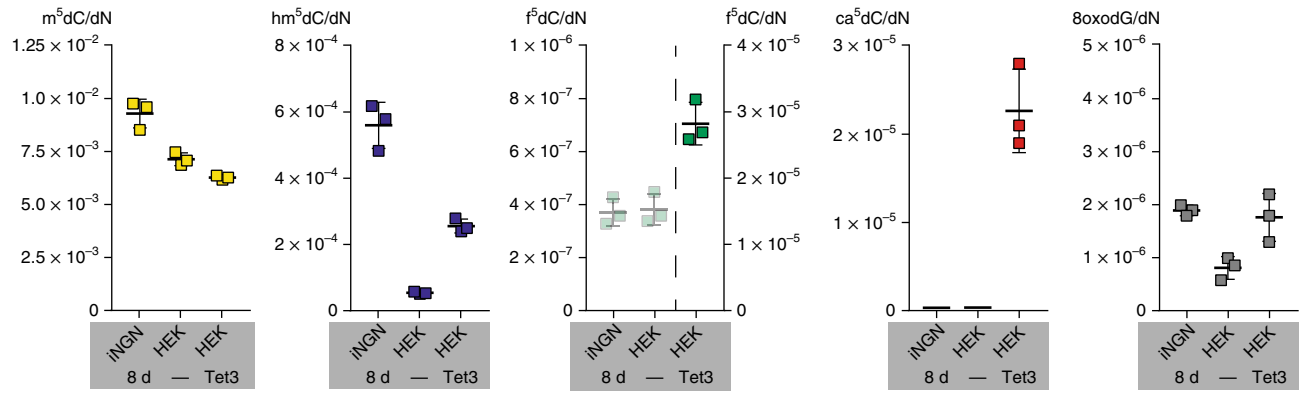


Fig. 7 | Quantities of m^5dC, hm^5dC, f^5dC, ca^5dC and 8oxodG in small molecule–inducible neurogenin human-induced pluripotent stem cells lines (iNGNs) (8 d after induction of differentiation), untransfected HEK293T cells and HEK293T cells that were transfected for 24 h with a plasmid coding for Tet3. The transparent data points at f^5dC for iNGNs and untransfected HEK293T cells indicate that f^5dC was above the LLOQ but did not meet the strict $LOD_{practical}$ criterion. Each data point represents the mean of one biological replicate that was measured in three technical replicates; the bar represents the mean of the biological replicates and error bars represent s.d.

Table 7 | Measurements for an exemplary calibration curve for m^4dC in technical triplicates

Calibration level	$A(m^4dC)$	$A^*(^{15}N_2\text{-}m^4dC)$	A/A^*	Replicate
L1	4,797	1,745,218	0.0027	1
L2	7,213	1,737,978	0.0042	
L3	10,664	1,692,721	0.0063	
L4	19,426	1,670,953	0.0116	
L5	34,395	1,677,652	0.0205	
L6	74,308	1,646,471	0.0451	
L7	164,901	1,566,133	0.1053	
L8	379,446	1,467,230	0.2586	
L9	869,825	1,340,740	0.6488	
L10	1,960,129	1,181,613	1.6589	
L11	4,164,191	971,267	4.2874	
L1	5,094	1,611,819	0.0032	2
L2	6,145	1,658,872	0.0037	
L3	10,207	1,658,337	0.0062	
L4	18,928	1,586,681	0.0119	
L5	34,678	1,570,039	0.0221	
L6	73,600	1,534,700	0.0480	
L7	163,872	1,558,445	0.1052	
L8	368,404	1,454,277	0.2533	
L9	875,082	1,321,236	0.6623	
L10	1,914,969	1,189,553	1.6098	
L11	4,264,798	885,407	4.8168	
L1	4,728	1,689,815	0.0028	3
L2	6,595	1,680,195	0.0039	
L3	10,841	1,677,259	0.0065	
L4	18,867	1,670,998	0.0113	
L5	35,859	1,696,961	0.0211	
L6	72,444	1,694,990	0.0427	
L7	169,019	1,543,636	0.1095	
L8	386,871	1,510,771	0.2561	
L9	897,343	1,315,416	0.6822	
L10	1,994,610	1,225,289	1.6279	
L11	4,063,788	944,247	4.3037	

mammalian cell lines or tissues is normally in a range between 0.6×10^{-2} and 1.0×10^{-2} m^5dC/dN, modifications that are less abundant can deviate by factors of hundreds or thousands, depending on the biological background of the sample (e.g., organism, culture conditions).

Figure 7 shows typical quantities of m^5dC, hm^5dC, f^5dC, ca^5dC and 8oxodG for iNGNs obtained 8 d after differentiation, for untransfected HEK293T cells and HEK293T cells that were transfected with a plasmid coding for Tet3 and incubated for an additional 24 h. For all the cell lines shown, 2 µg of gDNA was used per technical replicate for the digestion. As m^5dC is a very abundant DNA modification, only very little amounts of gDNA are needed for detection and quantification, and 2 µg is by far sufficient to exceed the LLOQ and the $LOD_{practical}$. Also, for hm^5dC, for which the typical amount is in the range of 10^{-4} hm^5dC/dN, 2 µg of gDNA is more than enough for the quantification, even in cell lines that have typically low quantities of this modification, such as HEK293T cells. However, for less abundant modifications, here f^5dC and ca^5dC, 2 µg is often not enough to meet the LLOQ or $LOD_{practical}$ criterion. Although the detected f^5dC levels in iNGNs and untransfected HEK293T cells were above the LLOQ, these results cannot be used for further interpretation, because in this case the background signal was too high and the samples therefore did not meet the $LOD_{practical}$ criterion. For ca^5dC, the abundance was above neither the $LOD_{practical}$ nor the LLOQ. One possible way to overcome this problem is to use more gDNA for the quantification. If you are interested only in a certain modification, you can also try to optimize the UHPLC–MS/MS system for

Table 8 | Calculations for an exemplary m^4dC calibration curve

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

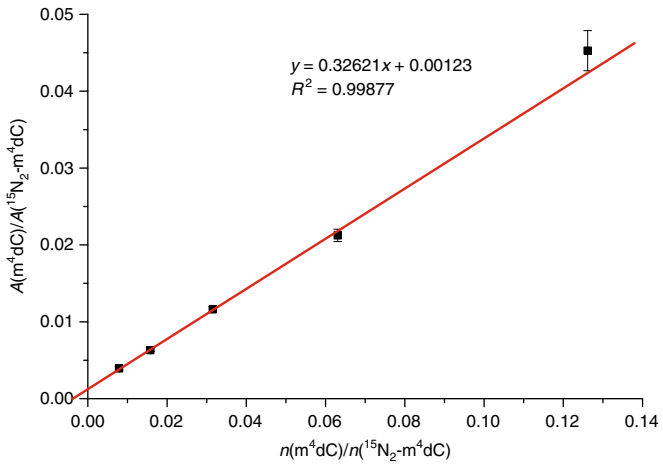


Fig. 8 | Exemplary calibration curve for m4dC. Three technical replicates were measured per data point. The mean is plotted and error bars represent s.d.

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

Evaluation of calibration curve

Table 7 shows an exemplary evaluation of the calibration curve of m^4dC as described in the Procedure. Consideration of all data points for linear regression would result in the following equation:

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

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

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

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

Ethics statement

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

Reporting Summary

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

References

1. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930 (2009).
2. Tahiliani, M. et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
3. Pfaffeneder, T. et al. The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Ed. Engl.* **50**, 7008–7012 (2011).
4. Ito, S. et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).
5. He, Y.-F. et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303-1307 (2011).
6. Wu, H. & Zhang, Y. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Gene. Dev.* **25**, 2436–2452 (2011).
7. Branco, M. R., Ficz, G. & Reik, W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat. Rev. Genet.* **13**, 7–13 (2011).
8. Langemeijer, S. M. C. et al. Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nat. Genet.* **41**, 838–842 (2009).
9. Ko, M. et al. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **468**, 839–843 (2010).
10. Jones, P. A. & Baylin, S. B. The epigenomics of cancer. *Cell* **128**, 683-692 (2007).
11. Pfaffeneder, T. et al. Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat. Chem. Biol.* **10**, 574–581 (2014).
12. Bachman, M. et al. 5-Formylcytosine can be a stable DNA modification in mammals. *Nat. Chem. Biol.* **11**, 555–557 (2015).
13. Iwan, K. et al. 5-Formylcytosine to cytosine conversion by C–C bond cleavage in vivo. *Nat. Chem. Biol.* **14**, 72–78 (2018).
14. Raiber, E.-A. et al. 5-Formylcytosine alters the structure of the DNA double helix. *Nat. Struct. Mol. Biol.* **22**, 44–49 (2014).
15. Song, C. X. et al. Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153**, 678–691 (2013).
16. Kellinger, M. W. et al. 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* **19**, 831–833 (2012).
17. Zhu, C. et al. Single-cell 5-formylcytosine landscapes of mammalian early embryos and ESCs at single-base resolution. *Cell Stem Cell* **20**, 720–731.e5 (2017).
18. Hill, P. W. S., Amouroux, R. & Hajkova, P. DNA demethylation, Tet proteins and 5-hydroxymethylcytosine in epigenetic reprogramming: an emerging complex story. *Genomics* **104**, 324–333 (2014).
19. Yu, M. et al. Base-resolution detection of N(4)-methylcytosine in genomic DNA using 4mC-Tet-assisted-bisulfite sequencing. *Nucleic Acids Res.* **43**, e148 (2015).
20. Ehrlich, M. et al. DNA methylation in thermophilic bacteria: N4-methylcytosine, 5-methylcytosine, and N6-methyladenine. *Nucleic Acids Res.* **13**, 1399–1412 (1985).
21. Arber, W. & Dussoix, D. Host specificity of DNA produced by *Escherichia coli*: I. Host controlled modification of bacteriophage λ. *J. Mol. Biol.* **5**, 18–36 (1962).
22. Wu, T. P. et al. DNA methylation on N6-adenine in mammalian embryonic stem cells. *Nature* **532**, 329–333 (2016).
23. Huang, W. et al. Determination of DNA adenine methylation in genomes of mammals and plants by liquid chromatography/mass spectrometry. *RSC Adv.* **5**, 64046–64054 (2015).
24. Koziol, M. J. et al. Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications. *Nat. Struct. Mol. Biol.* **23**, 24–30 (2015).
25. Liu, J. et al. Abundant DNA 6mA methylation during early embryogenesis of zebrafish and pig. *Nat. Commun.* **7**, 13052 (2016).
26. Schiffers, S. et al. Quantitative LC–MS provides no evidence for m^6dA or m^4dC in the genome of mouse embryonic stem cells and tissues. *Angew. Chem. Int. Ed. Engl.* **56**, 11268–11271 (2017).
27. Wang, D., Kreutzer, D. A. & Essigmann, J. M. Mutagenicity and repair of oxidative DNA damage: insights from studies using defined lesions. *Mutat. Res.* **400**, 99–115 (1998).

28. Cooke, M. S., Evans, M. D., Dizdaroglu, M. & Lunec, J. Oxidative DNA damage: mechanisms, mutation, and disease. *FASEB J.* **17**, 1195–1214 (2003).
29. Lunec, J. ESCODD: European standards committee on oxidative DNA damage. *Free Radical Res.* **29**, 601–608 (1998).
30. Murray, K. K. et al. Definitions of terms relating to mass spectrometry (IUPAC Recommendations 2013). *Pure Appl. Chem.* **85**, 1515–1609 (2013).
31. Oakeley, E. J. DNA methylation analysis: a review of current methodologies. *Pharmacol. Ther.* **84**, 389–400 (1999).
32. Taghizadeh, K. et al. Quantification of DNA damage products resulting from deamination, oxidation and reaction with products of lipid peroxidation by liquid chromatography isotope dilution tandem mass spectrometry. *Nat. Protoc.* **3**, 1287–1298 (2008).
33. Annesley, T. M. Ion suppression in mass spectrometry. *Clin. Chem.* **49**, 1041–1044 (2003).
34. Buhrman, D. L., Price, P. I. & Rudewiczcor, P. J . Quantitation of SR 27417 in human plasma using electrospray liquid chromatography–tandem mass spectrometry: a study of ion suppression.*J. Am. Soc. Mass Spectrom* **7**, 1099–1105 (1996).
35. Brückl, T., Globisch, D., Wagner, M., Müller, M. & Carell, T. Parallel isotope-based quantification of modified tRNA nucleosides. *Angew. Chem. Int. Ed. Engl.* **48**, 7932–7934 (2009).
36. Münzel, M. et al. Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew. Chem. Int. Ed. Engl.* **49**, 5375–5377 (2010).
37. Globisch, D. et al. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE* **5**, e15367 (2010).
38. Pearson, D. et al. LC-MS based quantification of 2′-ribosylated nucleosides Ar(p) and Gr(p) in tRNA. *Chem. Commun.* **47**, 5196–5198 (2011).
39. Globisch, D. et al. Systems-based analysis of modified tRNA bases. *Angew. Chem. Int. Ed. Engl.* **50**, 9739–9742 (2011).
40. Wetzel, C. & Limbach, P. A. Global identification of transfer RNAs by liquid chromatography–mass spectrometry (LC–MS). *J. Proteomics* **75**, 3450–3464 (2012).
41. Schiesser, S. et al. Deamination, oxidation, and C–C bond cleavage reactivity of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine. *J. Am. Chem. Soc.* **135**, 14593–14599 (2013).
42. Schröder, A. S. et al. Synthesis of a DNA promoter segment containing all four epigenetic nucleosides: 5-methyl-, 5-hydroxymethyl-, 5-formyl-, and 5-carboxy-2′-deoxycytidine. *Angew. Chem. Int. Ed. Engl.* **53**, 315–318 (2014).
43. Thienpont, B. et al. Tumour hypoxia causes DNA hypermethylation by reducing TET activity. *Nature* **537**, 63–68 (2016).
44. Rahimoff, R. et al. 5-Formyl- and 5-carboxydeoxycytidines do not cause accumulation of harmful repair intermediates in stem cells. *J. Am. Chem. Soc.* **139**, 10359–10364 (2017).
45. Turowski, M. et al. Deuterium isotope effects on hydrophobic interactions: the importance of dispersion interactions in the hydrophobic phase. *J. Am. Chem. Soc.* **125**, 13836–13849 (2003).
46. Kanao, E. et al. Isotope effects on hydrogen bonding and CH/CD−π interaction. *J. Phys. Chem.* **122**, 15026–15032 (2018).
47. Liu, S. & Wang, Y. Mass spectrometry for the assessment of the occurrence and biological consequences of DNA adducts. *Chem. Soc. Rev.* **44**, 7829–7854 (2015).
48. Lentini, A. et al. A reassessment of DNA-immunoprecipitation-based genomic profiling. *Nat. Methods* **15**, 499–504 (2018).
49. Shrivastava, A. & Gupta, V. B. HPLC: isocratic or gradient elution and assessment of linearity in analytical methods. *J. Adv. Sci. Res.* **3**, 12–20 (2012).
50. Schellinger, A. P. & Carr, P. W. Isocratic and gradient elution chromatography: a comparison in terms of speed, retention reproducibility and quantitation. *J. Chromatogr. A* **1109**, 253–266 (2006).
51. Zhang, J. J. et al. Analysis of global DNA methylation by hydrophilic interaction ultra high-pressure liquid chromatography tandem mass spectrometry. *Anal. Biochem.* **413**, 164–170 (2011).
52. Contreras-Sanz, A. et al. Simultaneous quantification of 12 different nucleotides and nucleosides released from renal epithelium and in human urine samples using ion-pair reversed-phase HPLC. *Purinergic Signal.* **8**, 741–751 (2012).
53. Yonekura, S., Iwasaki, M., Kai, M. & Ohkura, Y. High-performance liquid chromatographic determination of guanine and its nucleosides and nucleotides in biospecimens with precolumn fluorescence derivatization using phenylglyoxal. *Anal. Sci.* **10**, 247–251 (1994).
54. Giel-Pietraszuk, M. et al. Quantification of 5-methyl-2′-deoxycytidine in the DNA. *Acta Biochim. Pol.* **62**, 281–286 (2015).
55. Pitt, J. J. Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. *Clin. Biochem. Rev.* **30**, 19–34 (2009).
56. Niessen, W. M. A. State-of-the-art in liquid chromatography–mass spectrometry. *J. Chromatogr. A* **856**, 179–197 (1999).
57. Williamson, L. N. & Bartlett, M. G. Quantitative liquid chromatography/time-of-flight mass spectrometry. *Biomed. Chromatogr.* **21**, 567–576 (2007).
58. Payne, A. H. & Glish, G. L. Tandem mass spectrometry in quadrupole ion trap and ion cyclotron resonance mass spectrometers. *Methods Enzymol.* **402**, 109–148 (2005).

59. Tozuka, Z. et al. Strategy for structural elucidation of drugs and drug metabolites using (MS)*n* fragmentation in an electrospray ion trap. *J. Mass Spectrom.* **38**, 793–808 (2003).
60. Ens, W. & Standing, K. G. Hybrid quadrupole/time-of-flight mass spectrometers for analysis of biomolecules. *Methods Enzymol.* **402**, 49–78 (2005).
61. Commission Decision 2002/657/EC. *Implementation of Council Directive 96/23/EC Concerning the Performance of Analytical Methods and the Interpretation of Results* (Brussels, Official Journal of the European Communities, 2002).
62. Zhao, Z. & Zhang, F. Sequence context analysis in the mouse genome: single nucleotide polymorphisms and CpG island sequences. *Genomics* **87**, 68–74 (2006).
63. Kellner, S. et al. Oxidation of phosphorothioate DNA modifications leads to lethal genomic instability. *Nat. Chem. Biol.* **13**, 888–894 (2017).
64. Kellner, S. et al. Absolute and relative quantification of RNA modifications via biosynthetic isotopomers. *Nucleic Acids Res.* **42**, e142 (2014).
65. Amouroux, R. et al. De novo DNA methylation drives 5hmC accumulation in mouse zygotes. *Nat. Cell Biol.* **18**, 225–233 (2016).
66. Busskamp, V. et al. Rapid neurogenesis through transcriptional activation in human stem cells. *Mol. Syst. Biol.* **10**, 760 (2014).
67. Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
68. Collins, A. R., Cadet, J., Mőller, L., Poulsen, H. E. & Viña, J. Are we sure we know how to measure 8-oxo-7,8-dihydroguanine in DNA from human cells? *Arch. Biochem. Biophys.* **423**, 57–65 (2004).
69. ESCODD. Comparative analysis of baseline 8-oxo-7,8-dihydroguanine in mammalian cell DNA, by different methods in different laboratories: an approach to consensus. *Carcinogenesis* **23**, 2129–2133 (2002).
70. Ravanat, J.-L. et al. Cellular background level of 8-oxo-7,8-dihydro-2′-deoxyguanosine: an isotope based method to evaluate artefactual oxidation of DNA during its extraction and subsequent work-up. *Carcinogenesis* **23**, 1911–1918 (2002).

Acknowledgements

We thank the Deutsche Forschungsgemeinschaft for funding through SFB749, SFB1032, SPP1784 and CA275-11/1. Further support is acknowledged from the Excellence Cluster CiPSM (Center for Integrated Protein Science). F.R.T. thanks the Boehringer Ingelheim Fonds for her PhD fellowship. S.K. thanks the Fonds der Chemischen Industrie for the Liebig Fellowship. We thank T. Pfaffeneder for early method development and helpful input. We thank S. Michalakis (Department of Pharmacy, Ludwig-Maximilians–Universität München) for providing mouse tissues.

Author contributions

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

Competing interests

The authors declare no competing interests.

Additional information

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

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

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

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

Published online: 17 December 2018

Related links

Key references using this protocol

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

Additional Published Work

I also contributed to the following published work.

"2'(R)-Fluorinated mC, hmC, fC and caC triphosphates are substrates for DNA polymerases and TET-enzymes" by *Schröder et al.*[404] describes the synthesis and chemistry of 2'-(R)-fluorinated derivates of the 5mdC, 5hmdC, 5fdC and 5cadC nucleotides. These triphosphates could be incorporated by DNA polymerases and TET enzymes are performing the oxidation reaction on F-5mdC, F-5hmdC and F-5fdC, while TDG does not remove F-5fdC and F-5cadC. These features make them ideal tool substances to investigate *in vitro* and *in vivo* potential C-C cleavage reactions to remove the methyl group of 5mdC.

I contributed to this work by expressing GFP-Tet1cd in HEK293T cells and subsequently purified the enzyme using GFP nanobeads. GFP-Tet1cd was then used in the *in vitro* assays to monitor whether GFP-Tet1cd was able to oxidize the fluorinated derivates.

License

"Influencing Epigenetic Information with a Hydrolytically Stable Carbocyclic 5-Aza-2'-deoxycytidine" by *Wildenhof et al.*[405] describes the synthesis and chemistry of carbocyclic 5-aza-2'-deoxyxcytidine. This carbocyclic version does not decompose to create abasic sites in the genome, but still inhibits DNMTs. Therefore, it can be used as a tool molecule to study the effects of (impaired) DNA-methylation on cells, without perturbating the results by massive DNA lesion formation.

I contributed to this work by feeding the compound to mESCs and analyzed the outcome by UHPLC-QQQ-MS. Both the tasks were done in cooperation with *Dr. Sarah Schiffers*. The figures were created by and the manuscript was written by *Prof. Dr. Thomas Carell*, *Dr. Sarah Schiffers* and me.

License

Closing Remarks

The Point of View “The chemistries and consequences of DNA and RNA methylation and demethylation” by *Traube et al.*[8] was jointly written by *Prof. Dr. Thomas Carell* and me.

License

I contributed to the published protocol about isotope-dilution MS “Isotope-dilution mass spectrometry for exact quantification of noncanonical DNA nucleosides” by *Traube et al.* [376] design and development of the protocol in cooperation with *Dr. Sarah Schiffers* and *Dr. Katharina Iwan* and the manuscript was written by *Dr. Sarah Schiffers*, *Dr. Katharina Iwan*, *Dr. Stefanie Kellner*, *Dr. Fabio Spada*, *Dr. Markus Müller*, *Prof. Dr. Thomas Carell* and me.

License

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.

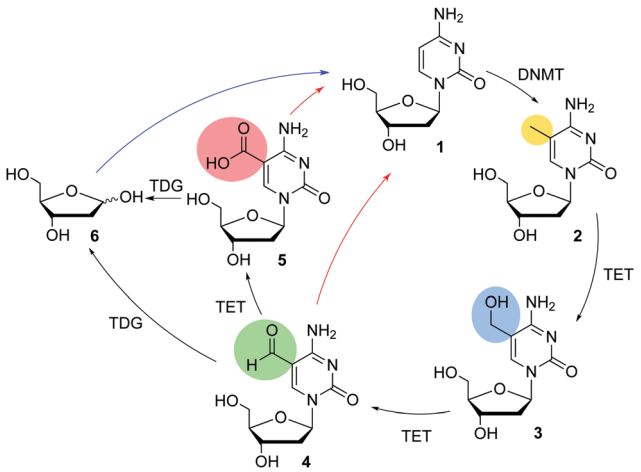


Fig. 1 Putative active demethylation pathways. DNMT: DNA methyltransferase, TET: ten-eleven-translocation methylcytosine dioxygenase, TDG: thymine DNA glycosylase. Red arrows: deformylation/decarboxylation, blue arrow: BER based active demethylation.

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.

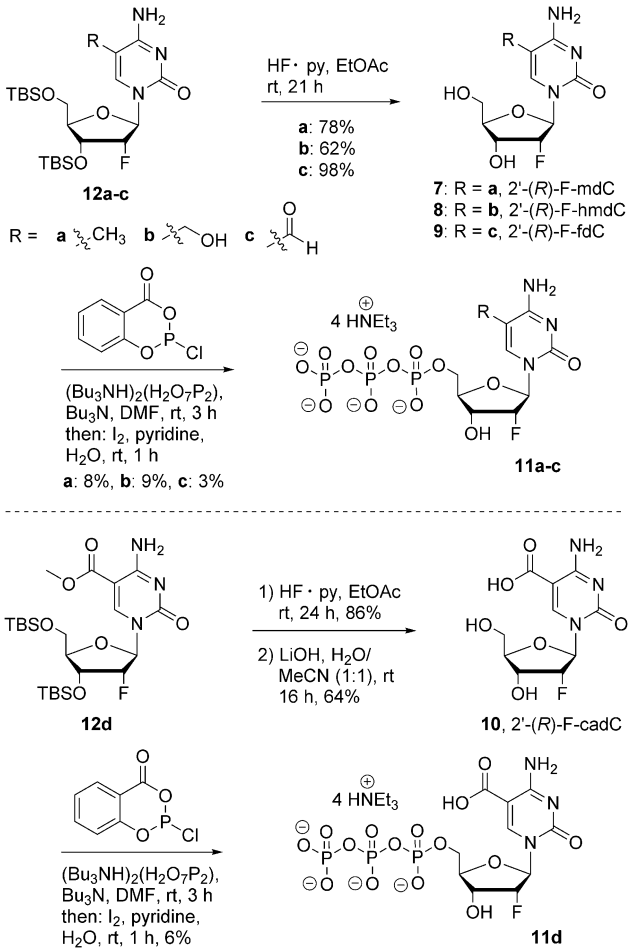
TET enzymes and this together makes them ideal tool substances to study active demethylation *via* the putative C–C bond cleavage.

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]

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,



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).

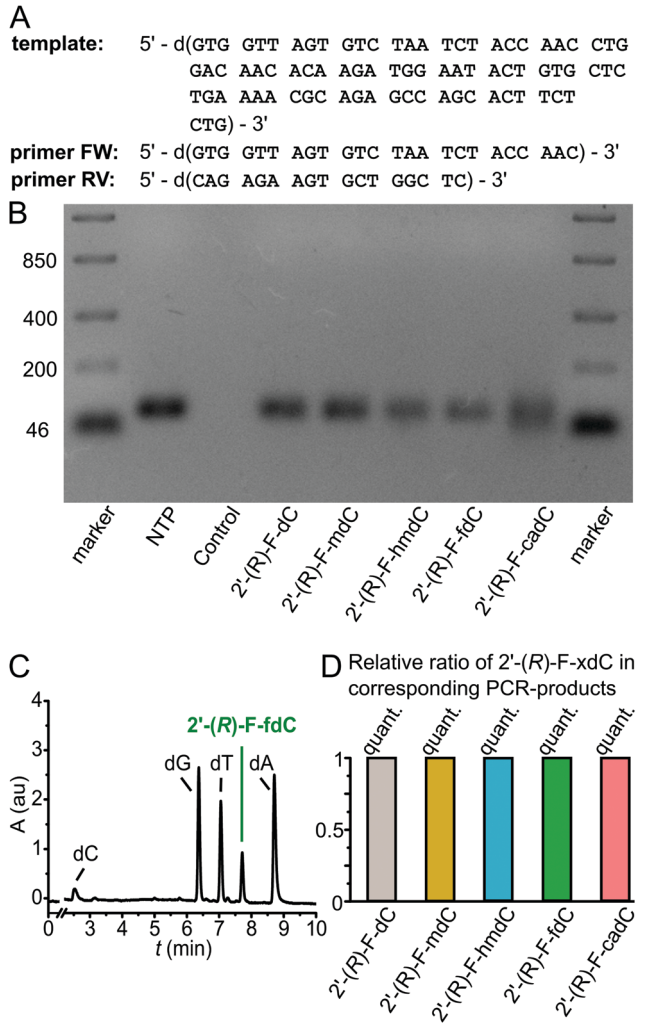


Fig. 2 Synthesis of 2′-(*R*)-F-xdC containing OCT4 promoter fragments by PCR. (A) Sequence of the template and primers used for the PCR (FW: forward, RV: reverse). (B) Result of the PCR analyzed by gel electrophoresis. The retention of the band corresponds to the expected PCR products. (Control: PCR 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

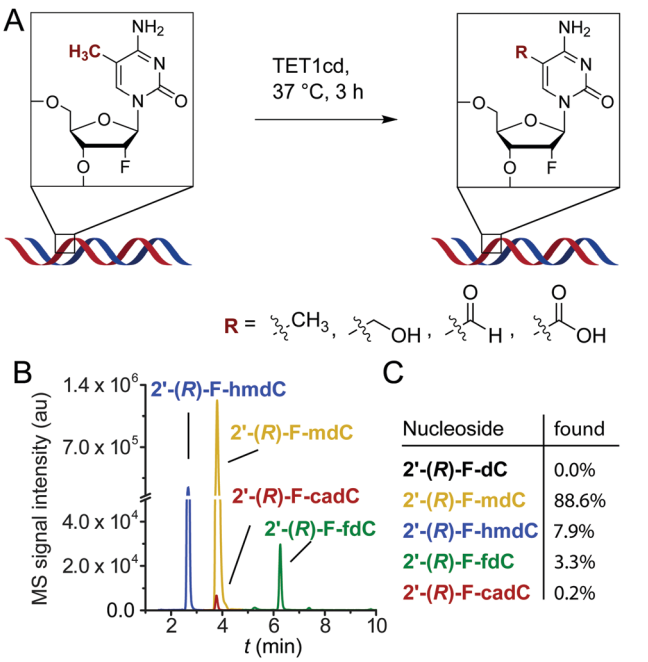


Fig. 3 TET1cd oxidation assay. (A) Scheme of the TET1cd dependent oxidation reaction. (B) UHPL-MS/MS-trace of 2′-(*R*)-F-xdC nucleosides after enzymatic digestion of the DNA fragment. (C) Quantification data of the product after complete enzymatic digestion and UHPLC-MS/MS (QQQ) analysis.

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.

Epigenetics Hot Paper

International Edition: DOI: 10.1002/anie.201904794
German Edition: DOI: 10.1002/ange.201904794

Influencing Epigenetic Information with a Hydrolytically Stable Carbocyclic 5-Aza-2′-deoxycytidine

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

Dedicated to Professor Julius Rebek, Jr. on the occasion of his 75th birthday

Abstract: *5-Aza-2′-deoxycytidine (AzadC) is an antimetabolite in clinical use, which reduces the level of the epigenetic modification 5-methyl-2′-deoxycytidine (mdC). AzadC is incorporated into the genome of proliferating cells, where it inhibits DNA methyltransferases (DNMTs), leading to a reduction of mdC. The loss of mdC, which is a transcriptional silencer in the promoter region found upstream of genes, leads to the reactivation of the corresponding gene, including tumor-suppressor genes, which elicits a beneficial effect. The problem associated with AzadC is that the compound is hydrolytically unstable. It decomposes during treatment to a variety of poorly characterized hydrolysis products. After its incorporation into the genome, this hydrolytic instability generates abasic sites. It is consequently difficult to dissect whether the activity of the compound is caused by DNMT inhibition or more generally by DNA lesion formation. We now discovered that a disarmed version of AzadC, in which the ribose oxygen was replaced by a CH_2 group, is surprisingly stable under a variety of pH values while keeping activity against the DNMTs.*

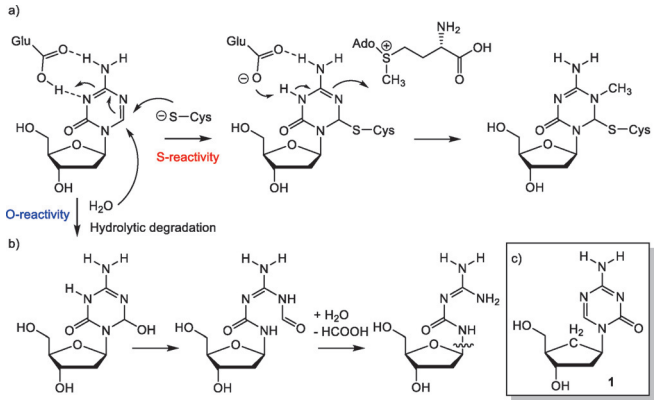


Figure 1. Structure of 5-aza-2′-deoxycytidine (decitabine, AzadC) together with its mode of action. a) Active-site thiol reacts with the C6-position of AzadC. b) Hydrolytic degradation pathway via reaction of a water molecule with the C6-position (O-reactivity) of AzadC. This leads to a final base loss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

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

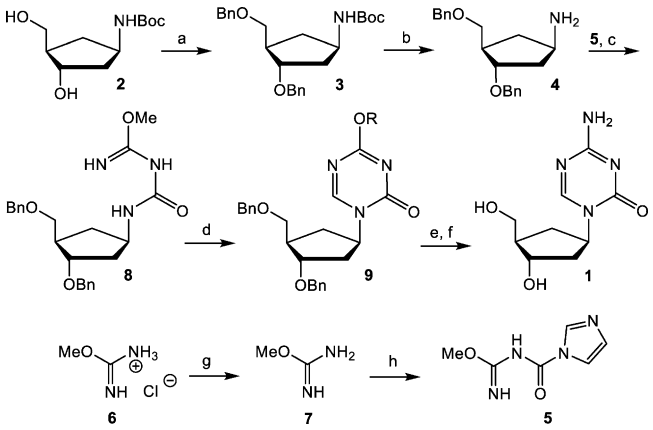
The problem associated with AzadC is that the compound is hydrolysed in aqueous solution following the path depicted in Figure 1 b. This hydrolysis compromises the activity of AzadC, particularly over long treatment times. In order to circumvent this problem, it is necessary to generate an AzadC compound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

Here, we report that replacing the oxygen of the ribose with a CH_2 group has a surprisingly large remote effect on the reactivity of the heterocycle. The created carbocyclic version of AzadC (cAzadC, **1**) still inhibits DNMTs but is hydrolytically stable (Figure 1 c).

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

Supporting information (including all synthetic, cell culture, and analytical procedures) and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/anie.201904794.

The synthesis of cAzadC (**1**) is depicted in Scheme 1. It starts with the Boc-protected aminocyclopentane derivative **2** that we used previously to synthesize DNA lesion analogues.[12–15] Compound **2** was first benzyl-protected to **3**, Boc-deprotected to **4**, and then reacted with carbimidazole **5**,



Scheme 1. Synthesis of the carbocyclic 5-aza-2′-deoxycytidine (cAzadC, **1**). a) NaH, BnBr, DMF, 0 °C, 1.5 h and stirred for additional 2 h at RT; b) TFA (30%), CH_2Cl_2 then Na_2CO_3, 10 min RT; c) CH_3CN, reflux, 2 h; d) $HC(OEt)_3$, TFA cat., reflux, 3 h; e) NH_3 (7 N, MeOH), 3 h, RT, then H_2O; f) CH_2Cl_2, −78 °C, BCl_3, 1 h, then →RT, 2 h, MeOH, 20 min. g) KOH, Et_2O/H_2O (39:1), −15 °C, 30 min, h) carbonyldiimidazole, THF, RT., 3 h; R = Me or Et. Bn = benzyl, DMF = dimethylformamide, TFA = trifluoroacetic acid, THF = tetrahydrofuran.

which was prepared in two steps from isomethylurea **6** after generation of the free base **7** with potassium hydroxide and reaction of **7** with carbonyldiimidazole. This provides the carbamoylurea-cyclopentane nucleoside analogue **8**. Cyclization to the triazine base **9** was subsequently performed with triethylorthoformate. Reaction of **9** with NH_3 in methanol and deprotection of the benzyl groups with BCl_3 in dichloromethane furnished the final compound cAzadC (**1**) as the free nucleoside.

Recrystallization of cAzadC from hot methanol gave colourless needles, which allowed us to solve the crystal structure that is depicted in Figure 2. Interesting is the observation that cAzadC exists with two different cyclopentane conformations in the crystal (Figure 2; Figure S1 in the Supporting Information). One conformer adopts a C6′-*endo* (P = 88.2°, ν_{max} = 47.8°) conformation (Figure 2 a), while the second exists as the C2′-*endo*-C3′-*exo* (South, P = 150.8°, ν_{max} = 45.4°) conformer (Figure 2 b). The latter conformation is typical for 2′-deoxynucleosides in DNA. This shows that the cAzadC nucleoside can adopt the correct DNA-type conformation, thus fuelling hope that the analogue has the potential to be phosphorylated and integrated into the genome.

We next investigated the stability of cAzadC in direct comparison to the pharmaceutical AzadC (Figure 3). Since one treatment cycle goes over four weeks, we decided to measure the stability at a time point related to a half cycle (14 d). We dissolved AzadC and cAzadC at a concentration of 100 mM in a phosphate buffer (100 mM) at three different pH

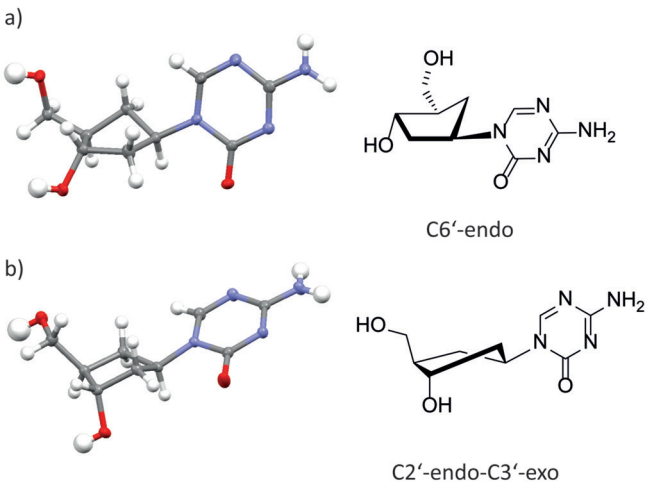


Figure 2. Crystal structure of carbocyclic 5-aza-2′-deoxycytidine (cAzadC), showing the molecule in the observed C6′-*endo* conformation (a) and the C2′-*endo*-C3′-*exo* conformation ($^2T_1$) (b).

values (7.4, 5.5 and 8.5) and measured NMR spectra after keeping the solutions at RT. Since tumour cells often provide an acidic microenvironment,[16] the stability under slightly acidic pH is particularly informative. As evident from the data shown in Figure 3, the pharmaceutical AzadC strongly degraded within these 14 d. Importantly, at pH 5.5 and at pH 8.5, intact AzadC was hardly detectable anymore. At physiological pH (7.4), AzadC was still present after 14 d but the level of degradation is dramatic. In contrast to these results, for cAzadC we did not observe degradation at any of the tested pH values, including pH 5.5. This result led to the surprising discovery that the simple O→CH_2 exchange causes a strong remote disarming effect that seems to change the properties of the triazine ring so that reaction with water is stopped.

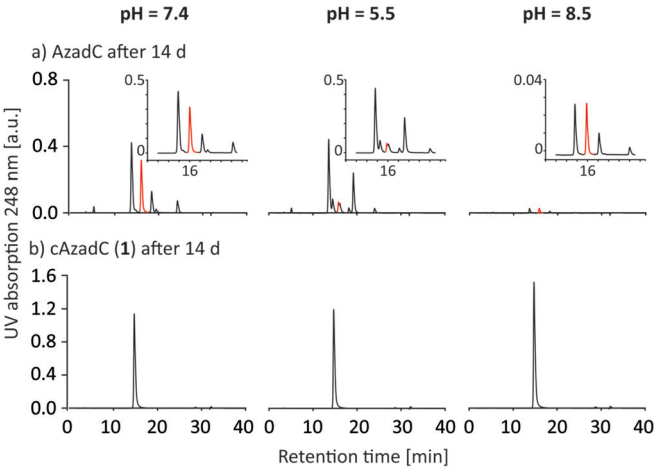


Figure 3. HPLC-based stability measurements. a) Severe hydrolytic decomposition of 5-aza-2′-deoxycytidine (AzadC) solutions at different pH values. b) The carbocyclic compound cAzadC (**1**) was stable at all three pH values. The inset table in (a) shows the chromatogram between t_1 = 10 min and t_2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

We next investigated whether this disarming effect would influence the biological functions. For this purpose, we used mouse embryonic stem cells (mESCs) that were primed in serum/LIF as a model system, since mdC levels increase from the naïve to primed state.[17] We added cAzadC in two different concentrations (1 μM and 5 μM) to mESCs that have been primed for 48 h and allowed the cells to further proliferate under priming conditions in the presence of cAzadC for additional 72 h. After the 72 h, we harvested the cells, isolated the DNA and digested the DNA down to the nucleoside level using our described protocol.[18] The levels of mdC were finally precisely quantified using isotope dilution UHPLC-MS^2. To this end, isotopically labelled standards of the nucleosides were spiked in for exact quantification.[18,19] In addition to mdC, we quantified the levels of 5-hydroxymethyl-2′-deoxycytidine (hmdC), which is formed from mdC by the action of TET enzymes.[20,21] The absolute levels of hmdC are in mESC more than ten times lower than the mdC levels.[20,22] The consequence is that even after a substantial reduction of mdC, there should be sufficient mdC to keep the hmdC levels constant. The question of if and by how much the hmdC level is affected can therefore inform us about how epigenetic reprogramming is organized. Parallel to the quantification of mdC and hmdC, we also quantified to which extent cAzadC itself was incorporated into the genome of the mESCs. Detection of AzadC in the genome of treated cells is only possible after treatment of the DNA with $NaBH_4$. Application of $NaBH_4$ reduces the C(5)=C(6) double bond, which stabilizes the compound so that its quantification becomes possible.[19,23] To our delight, we noted that the stability of cAzadC allowed its quantification without this pre-treatment. We also noted that the applied enzymatic digestion protocol allowed digestion of genomic DNA (gDNA) even in the presence of large amounts of cAzadC. Taken together, quantification of cAzadC by UHPLC-MS^2 using an external calibration curve (Figure S2) was possible in parallel with quantification of canonical and epigenetic bases.

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

Importantly, after exposing the mESC for 72 h at 1 μM cAzadC in the medium, we detected a reduction of the mdC values by almost 30% (Figure 4b). At 5 μM concentration in the medium, the mdC levels dropped even to about 50% of the original value. A decrease to 50% is observed for AzadC as well. Here, however, the 50% reduction is reached faster (24 h) and with lower AzadC concentration (1 μM).[19] The data show that the carbocyclic version, cAzadC, simply needs more time to affect the mdC levels by the same amount. We

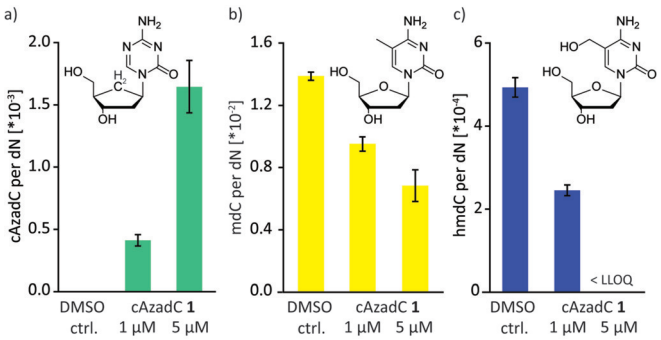


Figure 4. Quantification data for DNA modifications in cAzadC-treated mouse embryonic stem cells (mESCs) obtained by UHPLC-MS^2. For each condition, three biological replicates were measured in technical triplicates. For each technical replicate, 0.5 μg of DNA were digested. Bars represent the mean, error bars show the standard deviation. LLOQ indicates the lower limit of quantification.

believe that this effect is caused by a potentially slower conversion of cAzadC into the triphosphate. The slower kinetics of cAzadC, however, is not necessarily a disadvantage given the long treatment times that are applied in the clinic.

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

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

Acknowledgements

We thank the Deutsche Forschungsgemeinschaft (DFG) for financial support via the programs SFB1309 (TP-A4), SFB1361 (TP-02), SPP1784, and GRK2338/1 (P12). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020

research and innovation programme (grant agreement n° EPiR 741912). F.R.T. thanks the Boehringer Ingelheim Fonds for a PhD fellowship.

Conflict of interest

The authors declare no conflict of interest.

Keywords: 5-methyl-2′-deoxycytidine · DNA modification · epigenetics · methylation · nucleoside analogues

How to cite: *Angew. Chem. Int. Ed.* **2019**, *58*, 12984–12987
Angew. Chem. **2019**, *131*, 13118–13121

[1] M. Daskalakis, T. T. Nguyen, C. Nguyen, P. Guldberg, G. Kohler, P. Wijermans, P. A. Jones, M. Lubbert, *Blood* **2002**, *100*, 2957–2964.
[2] H. Kantarjian, J. P. Issa, C. S. Rosenfeld, J. M. Bennett, M. Albitar, J. DiPersio, V. Klimek, J. Slack, C. de Castro, F. Ravandi, R. Helmer III, L. Shen, S. D. Nimer, R. Leavitt, A. Raza, H. Saba, *Cancer* **2006**, *106*, 1794–1803.
[3] C. Stresemann, F. Lyko, *Int. J. Cancer* **2008**, *123*, 8–13.
[4] Y. Koh, Y. A. Kim, K. Kim, J.-A. Sim, S.-S. Yoon, S. M. Park, Y. H. Yun, *Blood* **2016**, *128*, 2381–2381.
[5] M. Nieto, P. Demolis, E. Béhanzin, A. Moreau, I. Hudson, B. Flores, H. Stemplewski, T. Salmonson, C. Gisselbrecht, D. Bowen, F. Pignatti, *Oncologist* **2016**, *21*, 692–700.
[6] S. S. Smith, B. E. Kaplan, L. C. Sowers, E. M. Newman, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 4744–4748.
[7] G. G. Wilson, R. J. Roberts, S. Kumar, J. Posfai, M. Sha, S. Klimasauskas, X. Cheng, *Nucleic Acids Res.* **1994**, *22*, 1–10.
[8] A. Bird, *Genes Dev.* **2002**, *16*, 6–21.
[9] R. J. Klose, A. P. Bird, *Trends Biochem. Sci.* **2006**, *31*, 89–97.
[10] R. Jüttermann, E. Li, R. Jaenisch, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 11797–11801.
[11] S. Gabbara, A. S. Bhagwat, *Biochem. J.* **1995**, *307*, 87–92.
[12] M. Ober, H. Müller, C. Pieck, J. Gierlich, T. Carell, *J. Am. Chem. Soc.* **2005**, *127*, 18143–18149.
[13] H. Müller, T. Carell, *Eur. J. Org. Chem.* **2007**, 1438–1445.
[14] F. Büsch, J. C. Pieck, M. Ober, J. Gierlich, G. W. Hsu, L. S. Beese, T. Carell, *Chem. Eur. J.* **2008**, *14*, 2125–2132.
[15] T. H. Gehrke, U. Lischke, K. L. Gasteiger, S. Schneider, S. Arnold, H. C. Müller, D. S. Stephenson, H. Zipse, T. Carell, *Nat. Chem. Biol.* **2013**, *9*, 455.
[16] Y. Kato, S. Ozawa, C. Miyamoto, Y. Maehata, A. Suzuki, T. Maeda, Y. Baba, *Cancer Cell Int.* **2013**, *13*, 89.
[17] S. Takahashi, S. Kobayashi, I. Hiratani, *Cell. Mol. Life Sci.* **2018**, *75*, 1191–1203.
[18] F. R. Traube, S. Schiffers, K. Iwan, S. Kellner, F. Spada, M. Müller, T. Carell, *Nat. Protoc.* **2019**, *14*, 283–312.
[19] S. Schiffers, T. M. Wildenhof, K. Iwan, M. Stadlmeier, M. Müller, T. Carell, *Helv. Chim. Acta* **2019**, *102*, e1800229.
[20] M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930–935.
[21] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300–1303.
[22] M. Münzel, D. Globisch, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 6460–6468; *Angew. Chem.* **2011**, *123*, 6588–6596.
[23] A. Unnikrishnan, A. N. Q. Vo, R. Pickford, M. J. Raftery, A. C. Nunez, A. Verma, L. B. Hesson, J. E. Pimanda, *Leukemia* **2018**, *32*, 900–910.

Manuscript received: April 17, 2019
Revised manuscript received: May 31, 2019
Accepted manuscript online: June 17, 2019
Version of record online: July 25, 2019

References

1. Zeng, H. & Sanes, J.R. Neuronal cell-type classification: challenges, opportunities and the path forward. *Nat. Rev. Neurosci.* **18**, 530-546 (2017).
2. Jäkel, S. & Dimou, L. Glial Cells and Their Function in the Adult Brain: A Journey through the History of Their Ablation. *Front. Cell. Neurosci.* **11**, 24 (2017).
3. Guan, J.S., Xie, H. & Ding, X. The role of epigenetic regulation in learning and memory. *Exp. Neurol.* **268**, 30-36 (2015).
4. Riccio, A. Dynamic epigenetic regulation in neurons: enzymes, stimuli and signaling pathways. *Nat. Neurosci.* **13**, 1330 (2010).
5. Murao, N., Noguchi, H. & Nakashima, K. Epigenetic regulation of neural stem cell property from embryo to adult. *Neuroepigenetics* **5**, 1-10 (2016).
6. Borrelli, E., Nestler, E.J., Allis, C.D. & Sassone-Corsi, P. Decoding the Epigenetic Language of Neuronal Plasticity. *Neuron* **60**, 961-974 (2008).
7. Probst, A.V., Dunleavy, E. & Almouzni, G. Epigenetic inheritance during the cell cycle. *Nat. Rev. Mol. Cell Biol.* **10**, 192-206 (2009).
8. Traube, F.R. & Carell, T. The chemistries and consequences of DNA and RNA methylation and demethylation. *RNA Biol.* **14**, 1099-1107 (2017).
9. Morris, M.J. & Monteggia, L.M. Role of DNA methylation and the DNA methyltransferases in learning and memory. *Dialogues Clin. Neurosci.* **16**, 359-371 (2014).
10. Santiago, M., Antunes, C., Guedes, M., Sousa, N. & Marques, C.J. TET enzymes and DNA hydroxymethylation in neural development and function - how critical are they? *Genomics* **104**, 334-340 (2014).
11. Khare, T. *et al.* 5-hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary. *Nat. Struct. Mol. Biol.* **19**, 1037-1043 (2012).
12. Mellén, M., Ayata, P. & Heintz, N. 5-hydroxymethylcytosine accumulation in postmitotic neurons results in functional demethylation of expressed genes. *Proc. Natl. Acad. Sci. USA* **114**, E7812-E7821 (2017).
13. Tahiliani, M. *et al.* Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science* **324**, 930-935 (2009).
14. Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129 - 1133 (2010).
15. Beck, D.B. *et al.* Delineation of a Human Mendelian Disorder of the DNA Demethylation Machinery: TET3 Deficiency. *Am. J. Hum. Genet.* **106**, 234-245 (2020).
16. Yu, H. *et al.* Tet3 regulates synaptic transmission and homeostatic plasticity via DNA oxidation and repair. *Nat. Neurosci.* **18**, 836-843 (2015).
17. Amir, R.E. *et al.* Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.* **23**, 185 (1999).
18. Guy, J., Hendrich, B., Holmes, M., Martin, J.E. & Bird, A. A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. *Nat. Genet.* **27**, 322 - 326 (2001).
19. Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S. & Heintz, N. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* **151**, 1417-1430 (2012).
20. Globisch, D. *et al.* Tissue Distribution of 5-Hydroxymethylcytosine and Search for Active Demethylation Intermediates. *PLoS One* **5**, e15367 (2010).
21. Szwagierczak, A., Bultmann, S., Schmidt, C.S., Spada, F. & Leonhardt, H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.* **38**, e181 (2010).
22. Sun, X. & Lin, Y. Npas4: Linking Neuronal Activity to Memory. *Trends Neurosci.* **39**, 264-275 (2016).
23. Kowiański, P. *et al.* BDNF: A Key Factor with Multipotent Impact on Brain Signaling and Synaptic Plasticity. *Cell. Mol. Neurobiol.* **38**, 579-593 (2018).
24. Carey, B.W., Finley, L.W., Cross, J.R., Allis, C.D. & Thompson, C.B. Intracellular alpha-ketoglutarate maintains the pluripotency of embryonic stem cells. *Nature* **518**, 413-416 (2015).

25. TeSlaa, T. *et al.* alpha-Ketoglutarate Accelerates the Initial Differentiation of Primed Human Pluripotent Stem Cells. *Cell Metab.* **24**, 485-493 (2016).
26. Nikiforov, A., Kulikova, V. & Ziegler, M. The human NAD metabolome: Functions, metabolism and compartmentalization. *Crit. Rev. Biochem. Mol. Biol.* **50**, 284-297 (2015).
27. Busskamp, V. *et al.* Rapid neurogenesis through transcriptional activation in human stem cells. *Mol. Syst. Biol.* **10**, 760 - 781 (2014).
28. Pray, L. Discovery of DNA structure and function: Watson and Crick. *Nature Education* **1(1):100** (2008).
29. Biesecker, L.G. & Spinner, N.B. A genomic view of mosaicism and human disease. *Nat. Rev. Genet.* **14**, 307 (2013).
30. Takahashi, K. Cellular reprogramming. *Cold Spring Harb. Perspect. Biol.* **6**, a018606 (2014).
31. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676 (2006).
32. Luger, K., Dechassa, M.L. & Tremethick, D.J. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat. Rev. Mol. Cell Biol.* **13**, 436-447 (2012).
33. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693-705 (2007).
34. Spruijt, C.G. *et al.* Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**, 1146-1159 (2013).
35. Peterson, C.L. & Laniel, M.A. Histones and histone modifications. *Curr. Biol.* **14**, R546-551 (2004).
36. Doerks, T., Copley, R.R., Schultz, J., Ponting, C.P. & Bork, P. Systematic identification of novel protein domain families associated with nuclear functions. *Genome Res.* **12**, 47-56 (2002).
37. Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F. & Richmond, T.J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251 - 260 (1997).
38. Gates, L.A., Foulds, C.E. & O'Malley, B.W. Histone Marks in the Driver's Seat: Functional Roles in Steering the Transcription Cycle. *Trends Biochem. Sci.* **42**, 977-989 (2017).
39. *The Merck Index - An Encyclopedia of Chemicals, Drugs, and Biologicals*. (Whitehouse Station, NJ: Merck and Co., Inc., 2006).
40. Eberharter, A. & Becker, P.B. Histone acetylation: a switch between repressive and permissive chromatin. *EMBO Rep.* **3**, 224-229 (2002).
41. Rossetto, D., Avvakumov, N. & Côté, J. Histone phosphorylation: a chromatin modification involved in diverse nuclear events. *Epigenetics* **7**, 1098-1108 (2012).
42. Allshire, R.C. & Madhani, H.D. Ten principles of heterochromatin formation and function. *Nat. Rev. Mol. Cell Biol.* **19**, 229-244 (2018).
43. Hergeth, S.P. & Schneider, R. The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. *EMBO Rep.* **16**, 1439-1453 (2015).
44. Schaukowitch, K. & Kim, T.K. Emerging epigenetic mechanisms of long non-coding RNAs. *Neuroscience* **264**, 25-38 (2014).
45. Volpe, T.A. *et al.* Regulation of Heterochromatic Silencing and Histone H3 Lysine-9 Methylation by RNAi. *Science* **297**, 1833-1837 (2002).
46. Costa, F.F. Non-coding RNAs, epigenetics and complexity. *Gene* **410**, 9-17 (2008).
47. Luo, M. *et al.* Long Non-Coding RNAs Control Hematopoietic Stem Cell Function. *Cell Stem Cell* **16**, 426-438 (2015).
48. Zlotorynski, E. X-chromosome inactivation unravelled. *Nat. Rev. Mol. Cell Biol.* **16**, 325 (2015).
49. Hermann, A., Goyal, R. & Jeltsch, A. The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *J. Biol. Chem.* **279**, 48350-48359 (2004).
50. Di Ruscio, A. *et al.* DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* **503**, 371 - 376 (2013).
51. Jeltsch, A. & Jurkowska, R.Z. Allosteric control of mammalian DNA methyltransferases - a new regulatory paradigm. *Nucleic Acids Res.* **44**, 8556-8575 (2016).

52. Dean, W. The elusive Dnmt1 and its role during early development. *Epigenetics* **3**, 175-178 (2014).
53. Budhavarapu, V.N., Chavez, M. & Tyler, J.K. How is epigenetic information maintained through DNA replication? *Epigenetics Chromatin* **6**, 32-32 (2013).
54. Kaelin, W.G., Jr. & McKnight, S.L. Influence of metabolism on epigenetics and disease. *Cell* **153**, 56-69 (2013).
55. Rolfe, D.F. & Brown, G.C. Cellular energy utilization and molecular origin of standard metabolic rate in mammals. *Physiol. Rev.* **77**, 731-758 (1997).
56. Locasale, J.W. & Cantley, L.C. Metabolic flux and the regulation of mammalian cell growth. *Cell Metab.* **14**, 443-451 (2011).
57. Belanger, M., Allaman, I. & Magistretti, P.J. Brain energy metabolism: focus on astrocyte-neuron metabolic cooperation. *Cell Metab.* **14**, 724-738 (2011).
58. Díaz-García, C.M. *et al.* Neuronal Stimulation Triggers Neuronal Glycolysis and Not Lactate Uptake. *Cell Metab.* **26**, 361-374 (2017).
59. Vander Heiden, M.G., Cantley, L.C. & Thompson, C.B. Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. *Science* **324**, 1029-1033 (2009).
60. Berg, J.M., Tymoczko, J.L. & Stryer, L. Glycogen Metabolism, in *Biochemistry, 5th edition* (W H Freeman, New York; 2002).
61. Berg, J.M., Tymoczko, J.L. & Stryer, L. The Pentose Phosphate Pathway Generates NADPH and Synthesizes Five-Carbon Sugars, in *Biochemistry, 5th edition* (W H Freeman, New York; 2002).
62. Berg, J.M., Tymoczko, J.L. & Stryer, L. Glycolysis Is an Energy-Conversion Pathyway in Many Organisms, in *Biochemistry, 5th edition* (W H Freeman, New York; 2002).
63. Berg, J.M., Tymoczko, J.L. & Stryer, L. The Citric Acid Cycle, in *Biochemistry, 5th edition* (W H Freeman, New York; 2002).
64. Berg, J.M., Tymoczko, J.L. & Stryer, L. Oxidative Phosphorylation, in *Biochemistry, 5th edition* (W H Freeman, New York; 2002).
65. Kim, J.W., Tchernyshyov, I., Semenza, G.L. & Dang, C.V. HIF-1-mediated expression of pyruvate dehydrogenase kinase: a metabolic switch required for cellular adaptation to hypoxia. *Cell Metab.* **3**, 177-185 (2006).
66. Patel, M.S. & Roche, T.E. Molecular biology and biochemistry of pyruvate dehydrogenase complexes. *FASEB J.* **4**, 3224-3233 (1990).
67. Zhang, J. *et al.* UCP2 regulates energy metabolism and differentiation potential of human pluripotent stem cells. *EMBO J.* **30**, 4860-4873 (2011).
68. Mathieu, J. *et al.* Hypoxia-inducible factors have distinct and stage-specific roles during reprogramming of human cells to pluripotency. *Cell Stem Cell* **14**, 592-605 (2014).
69. Yanes, O. *et al.* Metabolic oxidation regulates embryonic stem cell differentiation. *Nat. Chem. Biol.* **6**, 411-417 (2010).
70. Tischler, J. *et al.* Metabolic regulation of pluripotency and germ cell fate through α-ketoglutarate. *EMBO J.* **38**, e99518 (2019).
71. Battle, S.L. *et al.* Enhancer Chromatin and 3D Genome Architecture Changes from Naive to Primed Human Embryonic Stem Cell States. *Stem Cell Rep.* **12**, 1129-1144 (2019).
72. Moussaieff, A. *et al.* Glycolysis-mediated changes in acetyl-CoA and histone acetylation control the early differentiation of embryonic stem cells. *Cell Metab.* **21**, 392-402 (2015).
73. Bannister, A.J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **21**, 381-395 (2011).
74. Sperber, H. *et al.* The metabolome regulates the epigenetic landscape during naive-to-primed human embryonic stem cell transition. *Nat. Cell Biol.* **17**, 1523-1535 (2015).
75. Zhou, W. *et al.* HIF1alpha induced switch from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/hESC transition. *EMBO J.* **31**, 2103-2116 (2012).
76. Klingenberg, M. & Echtay, K.S. Uncoupling proteins: the issues from a biochemist point of view. *Biochim. Biophys. Acta Bioenerg.* **1504**, 128-143 (2001).

77. Sone, M. *et al.* Hybrid Cellular Metabolism Coordinated by Zic3 and Esrrb Synergistically Enhances Induction of Naive Pluripotency. *Cell Metab.* **25**, 1103-1117 (2017).
78. Cliff, T.S. *et al.* MYC Controls Human Pluripotent Stem Cell Fate Decisions through Regulation of Metabolic Flux. *Cell Stem Cell* **21**, 502-516 (2017).
79. Sim, Y.-J. *et al.* 2i Maintains a Naive Ground State in ESCs through Two Distinct Epigenetic Mechanisms. *Stem Cell Rep.* **8**, 1312-1328 (2017).
80. Silva, J. *et al.* Promotion of Reprogramming to Ground State Pluripotency by Signal Inhibition. *PLoS Biol.* **6**, e253 (2008).
81. Tamm, C., Pijuan Galitó, S. & Annerén, C. A Comparative Study of Protocols for Mouse Embryonic Stem Cell Culturing. *PLoS One* **8**, e81156 (2013).
82. Hwang, I.Y. *et al.* Psat1-Dependent Fluctuations in alpha-Ketoglutarate Affect the Timing of ESC Differentiation. *Cell Metab.* **24**, 494-501 (2016).
83. Loenarz, C. & Schofield, C.J. Expanding chemical biology of 2-oxoglutarate oxygenases. *Nat. Chem. Biol.* **4**, 152-156 (2008).
84. Lin, A.P. *et al.* D2HGDH regulates alpha-ketoglutarate levels and dioxygenase function by modulating IDH2. *Nat. Commun.* **6**, 7768 (2015).
85. Park, J.O. *et al.* Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. *Nat. Chem. Biol.* **12**, 482-489 (2016).
86. Tarhonskaya, H. *et al.* Studies on the Interaction of the Histone Demethylase KDM5B with Tricarboxylic Acid Cycle Intermediates. *J. Mol. Biol.* **429**, 2895-2906 (2017).
87. Laukka, T. *et al.* Fumarate and Succinate Regulate Expression of Hypoxia-inducible Genes via TET Enzymes. *J. Biol. Chem.* **291**, 4256-4265 (2016).
88. Wu, N. *et al.* Alpha-Ketoglutarate: Physiological Functions and Applications. *Biomol. Ther. (Seoul)* **24**, 1-8 (2016).
89. Yang, Q. *et al.* AMPK/α-Ketoglutarate Axis Dynamically Mediates DNA Demethylation in the Prdm16 Promoter and Brown Adipogenesis. *Cell Metab.* **24**, 542-554 (2016).
90. Wellen, K.E. *et al.* The hexosamine biosynthetic pathway couples growth factor-induced glutamine uptake to glucose metabolism. *Genes Dev.* **24**, 2784-2799 (2010).
91. Garten, A., Petzold, S., Körner, A., Imai, S.-i. & Kiess, W. Nampt: Linking NAD biology, metabolism, and cancer. *Trends Endocrinol. Metab.* **20**, 130-138 (2009).
92. Contreras, L. & Satrustegui, J. Calcium signaling in brain mitochondria: interplay of malate aspartate NADH shuttle and calcium uniporter/mitochondrial dehydrogenase pathways. *J. Biol. Chem.* **284**, 7091-7099 (2009).
93. McKenna, M.C., Waagepetersen, H.S., Schousboe, A. & Sonnewald, U. Neuronal and astrocytic shuttle mechanisms for cytosolic-mitochondrial transfer of reducing equivalents: Current evidence and pharmacological tools. *Biochem. Pharmacol.* **71**, 399-407 (2006).
94. Chen, W.W., Freinkman, E., Wang, T., Birsoy, K. & Sabatini, D.M. Absolute Quantification of Matrix Metabolites Reveals the Dynamics of Mitochondrial Metabolism. *Cell* **166**, 1324-1337 (2016).
95. Wang, R. & Brattain, M.G. The maximal size of protein to diffuse through the nuclear pore is larger than 60kDa. *FEBS Lett.* **581**, 3164-3170 (2007).
96. Borle, A.B. Kinetic analyses of calcium movements in HeLa cell cultures. I. Calcium influx. *J. Gen. Physiol.* **53**, 43-56 (1969).
97. García-Pérez, A.I. *et al.* Molecular Crowding and Viscosity as Determinants of Translational Diffusion of Metabolites in Subcellular Organelles. *Arch. Biochem. Biophys.* **362**, 329-338 (1999).
98. Webster, M., Witkin, K.L. & Cohen-Fix, O. Sizing up the nucleus: nuclear shape, size and nuclear-envelope assembly. *J. Cell Sci.* **122**, 1477-1486 (2009).
99. Chinwalla, A.T. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520-562 (2002).
100. Zhao, Z. & Zhang, F. Sequence context analysis in the mouse genome: Single nucleotide polymorphisms and CpG island sequences. *Genomics* **87**, 68-74 (2006).

101. Pritchard, J.B. Intracellular alpha-ketoglutarate controls the efficacy of renal organic anion transport. *J. Pharmacol. Exp. Ther.* **274**, 1278-1284 (1995).
102. 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).
103. Blaschke, K. *et al.* Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* **500**, 222-226 (2013).
104. Linster, C.L., Van Schaftingen, E. & Hanson, A.D. Metabolite damage and its repair or pre-emption. *Nat. Chem. Biol.* **9**, 72 - 80 (2013).
105. Xu, W. *et al.* Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. *Cancer Cell* **19**, 17-30 (2011).
106. Dang, L. *et al.* Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739 - 744 (2009).
107. Intlekofer, Andrew M. *et al.* Hypoxia Induces Production of L-2-Hydroxyglutarate. *Cell Metab.* **22**, 304-311 (2015).
108. Keith, B. & Simon, M.C. Hypoxia-inducible factors, stem cells, and cancer. *Cell* **129**, 465-472 (2007).
109. Xie, Y. *et al.* Defining the role of oxygen tension in human neural progenitor fate. *Stem Cell Rep.* **3**, 743-757 (2014).
110. Dunwoodie, S.L. The Role of Hypoxia in Development of the Mammalian Embryo. *Dev. Cell* **17**, 755-773 (2009).
111. Danet, G.H., Pan, Y., Luongo, J.L., Bonnet, D.A. & Simon, M.C. Expansion of human SCID-repopulating cells under hypoxic conditions. *J. Clin. Invest.* **112**, 126-135 (2003).
112. Forristal, C.E., Wright, K.L., Hanley, N.A., Oreffo, R.O.C. & Houghton, F.D. Hypoxia inducible factors regulate pluripotency and proliferation in human embryonic stem cells cultured at reduced oxygen tensions. *Reproduction* **139**, 85-97 (2010).
113. Ezashi, T., Das, P. & Roberts, R.M. Low O2 tensions and the prevention of differentiation of hES cells. *Proc. Natl. Acad. Sci. USA* **102**, 4783-4788 (2005).
114. Schöler, H.R., Ruppert, S., Suzuki, N., Chowdhury, K. & Gruss, P. New type of POU domain in germ line-specific protein Oct-4. *Nature* **344**, 435-439 (1990).
115. Thienpont, B. *et al.* Tumour hypoxia causes DNA hypermethylation by reducing TET activity. *Nature* **537**, 63 - 68 (2016).
116. Boukouris, A.E., Zervopoulos, S.D. & Michelakis, E.D. Metabolic Enzymes Moonlighting in the Nucleus: Metabolic Regulation of Gene Transcription. *Trends Biochem. Sci.* **41**, 712-730 (2016).
117. Zhong, X.H. & Howard, B.D. Phosphotyrosine-containing lactate dehydrogenase is restricted to the nuclei of PC12 pheochromocytoma cells. *Mol. Cell. Biol.* **10**, 770-776 (1990).
118. Popanda, O., Fox, G. & Thielmann, H.W. Modulation of DNA polymerases α, δ and ε by lactate dehydrogenase and 3-phosphoglycerate kinase1. *Biochim. Biophys. Acta Gene Struct. Expr.* **1397**, 102-117 (1998).
119. Grosse, F., Nasheuer, H.-P., Scholtissek, S. & Schomburg, U. Lactate dehydrogenase and glyceraldehyde-phosphate dehydrogenase are single-stranded DNA-binding proteins that affect the DNA-polymerase-α–primase complex. *Eur. J. Biochem.* **160**, 459-467 (1986).
120. Castonguay, Z., Auger, C., Thomas, S.C., Chahma, M.h. & Appanna, V.D. Nuclear lactate dehydrogenase modulates histone modification in human hepatocytes. *Biochem. Biophys. Res. Commun.* **454**, 172-177 (2014).
121. Yogev, O. *et al.* Fumarase: A Mitochondrial Metabolic Enzyme and a Cytosolic/Nuclear Component of the DNA Damage Response. *PLoS Biol.* **8**, e1000328 (2010).
122. De, P. & Chatterjee, R. Evidence of nucleolar succinic dehydrogenase activity. *Exp. Cell Res.* **27**, 172-173 (1962).
123. Jung, S.-J., Seo, Y., Lee, K.-C., Lee, D. & Roe, J.-H. Essential function of Aco2, a fusion protein of aconitase and mitochondrial ribosomal protein bL21, in mitochondrial translation in fission yeast. *FEBS Lett.* **589**, 822-828 (2015).

124. Haraguchi, C.M., Mabuchi, T. & Yokota, S. Localization of a Mitochondrial Type of NADP-dependent Isocitrate Dehydrogenase in Kidney and Heart of Rat: An Immunocytochemical and Biochemical Study. *J. Histochem. Cytochem.* **51**, 215-226 (2003).
125. McEwen, B.S., Allfrey, V.G. & Mirsky, A.E. Studies on Energy-yielding Reactions in Thymus Nuclei. *J. Biol. Chem.* **238**, 2571-2578 (1963).
126. Nagaraj, R. *et al.* Nuclear Localization of Mitochondrial TCA Cycle Enzymes as a Critical Step in Mammalian Zygotic Genome Activation. *Cell* **168**, 210-223 e211 (2017).
127. Reed, L.J. & Cox, D.J. Macromolecular Organization of Enzyme Systems. *Annu. Rev. Biochem.* **35**, 57-84 (1966).
128. Sutendra, G. *et al.* A nuclear pyruvate dehydrogenase complex is important for the generation of acetyl-CoA and histone acetylation. *Cell* **158**, 84-97 (2014).
129. Douglas, R. & Martin, K. Neocortex, in *The synaptic organization of the brain, 4th ed.* 459-509 (Oxford University Press, New York; 1998).
130. Rockel, A.J., Hiorns, R.W. & Powell, T.P.S. The Basic Uniformity in Structure of the Neocortex. *Brain* **103**, 221-244 (1980).
131. Allen, N.J. & Eroglu, C. Cell Biology of Astrocyte-Synapse Interactions. *Neuron* **96**, 697-708 (2017).
132. Sofroniew, M.V. & Vinters, H.V. Astrocytes: biology and pathology. *Acta Neuropathol.* **119**, 7-35 (2010).
133. Nave, K.-A. Myelination and the trophic support of long axons. *Nat. Rev. Neurosci.* **11**, 275 - 283 (2010).
134. Dimou, L., Simon, C., Kirchhoff, F., Takebayashi, H. & Götz, M. Progeny of Olig2-Expressing Progenitors in the Gray and White Matter of the Adult Mouse Cerebral Cortex. *J. Neurosci.* **28**, 10434-10442 (2008).
135. Sun, W. & Dietrich, D. Synaptic integration by NG2 cells. *Front. Cell. Neurosci.* **7**, 1-13 (2013).
136. Carter, A.M. & Enders, A.C. Placentation in mammals: Definitive placenta, yolk sac, and paraplacenta. *Theriogenology* **86**, 278-287 (2016).
137. Ramon y Cajal, S. *Histologie du Système Nerveux de l'Homme et des Vertébrés*. (Oxford University Press, New York; 1899).
138. Bernard, A., Sorensen, S.A. & Lein, E.S. Shifting the paradigm: new approaches for characterizing and classifying neurons. *Curr. Opin. Neurobiol.* **19**, 530-536 (2009).
139. Fishell, G. & Heintz, N. The Neuron Identity Problem: Form Meets Function. *Neuron* **80**, 602-612 (2013).
140. Gusel'nikova, V.V. & Korzhevskiy, D.E. NeuN As a Neuronal Nuclear Antigen and Neuron Differentiation Marker. *Acta Naturae* **7**, 42-47 (2015).
141. Dehmelt, L. & Halpain, S. The MAP2/Tau family of microtubule-associated proteins. *Genome Biol.* **6**, 204-204 (2005).
142. Conde, C. & Cáceres, A. Microtubule assembly, organization and dynamics in axons and dendrites. *Nat. Rev. Neurosci.* **10**, 319-332 (2009).
143. Missler, M., Südhof, T.C. & Biederer, T. Synaptic cell adhesion. *Cold Spring Harb. Perspect. Biol.* **4**, a005694 (2012).
144. O'Rourke, N.A., Weiler, N.C., Micheva, K.D. & Smith, S.J. Deep molecular diversity of mammalian synapses: why it matters and how to measure it. *Nat. Rev. Neurosci.* **13**, 365-379 (2012).
145. Connors, B.W. & Long, M.A. Electrical synapses in the mammalian brain. *Annu. Rev. Neurosci.* **27**, 393-418 (2004).
146. Nagy, J.I., Dudek, F.E. & Rash, J.E. Update on connexins and gap junctions in neurons and glia in the mammalian nervous system. *Brain Res. Rev.* **47**, 191-215 (2004).
147. Sohl, G., Maxeiner, S. & Willecke, K. Expression and functions of neuronal gap junctions. *Nat. Rev. Neurosci.* **6**, 191-200 (2005).
148. Evans, W.H. & Martin, P.E. Gap junctions: structure and function (Review). *Mol. Membr. Biol.* **19**, 121-136 (2002).

149. Marder, E. Electrical Synapses: Rectification Demystified. *Curr. Biol.* **19**, R34-R35 (2009).
150. Alcami, P. & Pereda, A.E. Beyond plasticity: the dynamic impact of electrical synapses on neural circuits. *Nat. Rev. Neurosci.* **20**, 253-271 (2019).
151. Savtchenko, L.P. & Rusakov, D.A. The optimal height of the synaptic cleft. *Proc. Natl. Acad. Sci. USA* **104**, 1823-1828 (2007).
152. Allen, N.J. Astrocyte regulation of synaptic behavior. *Annu. Rev. Cell. Dev. Biol.* **30**, 439-463 (2014).
153. Eliasson, L. The exocytotic machinery. *Acta Physiol.* **210**, 455-457 (2014).
154. Grienberger, C. & Konnerth, A. Imaging Calcium in Neurons. *Neuron* **73**, 862-885 (2012).
155. Lu, B. *et al.* Extracellular Calcium Controls Background Current and Neuronal Excitability via an UNC79-UNC80-NALCN Cation Channel Complex. *Neuron* **68**, 488-499 (2010).
156. Schneggenburger, R. & Neher, E. Intracellular calcium dependence of transmitter release rates at a fast central synapse. *Nature* **406**, 889-893 (2000).
157. Smart, T.G. & Paoletti, P. Synaptic neurotransmitter-gated receptors. *Cold Spring Harb. Perspect. Biol.* **4**, a009662 (2012).
158. Neves, G., Cooke, S.F. & Bliss, T.V.P. Synaptic plasticity, memory and the hippocampus: a neural network approach to causality. *Nat. Rev. Neurosci.* **9**, 65-75 (2008).
159. Pereda, A.E. Electrical synapses and their functional interactions with chemical synapses. *Nat. Rev. Neurosci.* **15**, 250-263 (2014).
160. Martin, S.J., Grimwood, P.D. & Morris, R.G.M. Synaptic Plasticity and Memory: An Evaluation of the Hypothesis. *Annu. Rev. Neurosci.* **23**, 649-711 (2000).
161. Massey, P.V. & Bashir, Z.I. Long-term depression: multiple forms and implications for brain function. *Trends Neurosci.* **30**, 176-184 (2007).
162. Oberheim, N.A. *et al.* Uniquely Hominid Features of Adult Human Astrocytes. *J. Neurosci.* **29**, 3276-3287 (2009).
163. Khaitovich, P. *et al.* Regional patterns of gene expression in human and chimpanzee brains. *Genome Res.* **14**, 1462-1473 (2004).
164. Kanton, S. *et al.* Organoid single-cell genomic atlas uncovers human-specific features of brain development. *Nature* **574**, 418-422 (2019).
165. Kornack, D.R. & Rakic, P. Continuation of neurogenesis in the hippocampus of the adult macaque monkey. *Proc. Natl. Acad. Sci. USA* **96**, 5768-5773 (1999).
166. Seri, B., García-Verdugo, J.M., McEwen, B.S. & Alvarez-Buylla, A. Astrocytes Give Rise to New Neurons in the Adult Mammalian Hippocampus. *J. Neurosci.* **21**, 7153-7160 (2001).
167. van Praag, H. *et al.* Functional neurogenesis in the adult hippocampus. *Nature* **415**, 1030-1034 (2002).
168. Kempermann, G., Kuhn, H.G. & Gage, F.H. More hippocampal neurons in adult mice living in an enriched environment. *Nature* **386**, 493-495 (1997).
169. van Praag, H., Kempermann, G. & Gage, F.H. Running increases cell proliferation and neurogenesis in the adult mouse dentate gyrus. *Nat. Neurosci.* **2**, 266-270 (1999).
170. Eriksson, P.S. *et al.* Neurogenesis in the adult human hippocampus. *Nat. Med.* **4**, 1313-1317 (1998).
171. Spalding, Kirsty L. *et al.* Dynamics of Hippocampal Neurogenesis in Adult Humans. *Cell* **153**, 1219-1227 (2013).
172. Sorrells, S.F. *et al.* Human hippocampal neurogenesis drops sharply in children to undetectable levels in adults. *Nature* **555**, 377 - 381 (2018).
173. Lister, R. *et al.* Global epigenomic reconfiguration during mammalian brain development. *Science* **341**, 1237905 (2013).
174. Wagner, M. *et al.* Age-Dependent Levels of 5-Methyl-, 5-Hydroxymethyl-, and 5-Formylcytosine in Human and Mouse Brain Tissues. *Angew. Chem. Int. Ed. Engl.* **54**, 12511-12514 (2015).
175. Wen, L. *et al.* Whole-genome analysis of 5-hydroxymethylcytosine and 5-methylcytosine at base resolution in the human brain. *Genome Biol.* **15**, R49 (2014).

176. Bird, C.M. & Burgess, N. The hippocampus and memory: insights from spatial processing. *Nat. Rev. Neurosci.* **9**, 182 - 194 (2008).
177. Smith, D.M. & Bulkin, D.A. The Form and Function of Hippocampal Context Representations. *Neurosci. Biobehav. Rev.* **40**, 52-61 (2014).
178. Morris, R.G.M., Garrud, P., Rawlins, J.N.P. & O'Keefe, J. Place navigation impaired in rats with hippocampal lesions. *Nature* **297**, 681 - 683 (1982).
179. Andersen, P., Morris, R., Amaral, D., Bliss, T. & O'Keefe, J. *The Hippocampus Book*. (Oxford University Press, New York; 2007).
180. Lorente de Nó, R. Studies on the structure of the cerebral cortex. II. Continuation of the study of the ammonic system. *J. Neurol. Psychol.* **46**, 113 - 177 (1934).
181. van Strien, N.M., Cappaert, N.L.M. & Witter, M.P. The anatomy of memory: an interactive overview of the parahippocampal–hippocampal network. *Nat. Rev. Neurosci.* **10**, 272 - 282 (2009).
182. Schultz, C. & Engelhardt, M. Anatomy of the hippocampal formation. *Front. Neurol. Neurosci.* **34**, 6-17 (2014).
183. Freund, T.F. & Buzsáki, G. Interneurons of the hippocampus. *Hippocampus* **6**, 347-470 (1996).
184. Group, P.I.N. Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nat. Rev. Neurosci.* **9**, 557 - 568 (2008).
185. Kepecs, A. & Fishell, G. Interneuron cell types are fit to function. *Nature* **505**, 318-326 (2014).
186. Ishizuka, N., Weber, J. & Amaral, D.G. Organization of intrahippocampal projections originating from CA3 pyramidal cells in the rat. *J. Comp. Neurol.* **295**, 580-623 (1990).
187. Amaral, D.G., Dolorfo, C. & Alvarez-Royo, P. Organization of CA1 projections to the subiculum: A PHA-L analysis in the rat. *Hippocampus* **1**, 415-435 (1991).
188. Finch, D.M., Nowlin, N.L. & Babb, T.L. Demonstration of axonal projections of neurons in the rat hippocampus and subiculum by intracellular injection of HRP. *Brain Res.* **271**, 201-216 (1983).
189. Henze, D.A., Urban, N.N. & Barrionuevo, G. The multifarious hippocampal mossy fiber pathway: a review. *Neuroscience* **98**, 407-427 (2000).
190. Lee, K.J. *et al.* Mossy fiber-CA3 synapses mediate homeostatic plasticity in mature hippocampal neurons. *Neuron* **77**, 99-114 (2013).
191. Buckmaster, P.S., Strowbridge, B.W. & Schwartzkroin, P.A. A comparison of rat hippocampal mossy cells and CA3c pyramidal cells. *J. Neurophysiol.* **70**, 1281-1299 (1993).
192. Li, X.G., Somogyi, P., Ylinen, A. & Buzsáki, G. The hippocampal CA3 network: An in vivo intracellular labeling study. *J. Comp. Neurol.* **339**, 181-208 (1994).
193. Swanson, L., Sawchenko, P. & Cowan, W. Evidence for collateral projections by neurons in Ammon's horn, the dentate gyrus, and the subiculum: a multiple retrograde labeling study in the rat. *J. Neurosci.* **1**, 548-559 (1981).
194. Christer, K. Intrinsic projections of the retrohippocampal region in the rat brain. I. The subicular complex. *J. Comp. Neurol.* **236**, 504-522 (1985).
195. Dudek, S.M., Alexander, G.M. & Farris, S. Rediscovering area CA2: unique properties and functions. *Nat. Rev. Neurosci.* **17**, 89-102 (2016).
196. Gaarskjaer, F. The organization and development of the hippocampal mossy fiber system. *Brain Res. Rev.* **11**, 335-357 (1986).
197. Woodhams, P.L., Celio, M.R., Ulfig, N. & Witter, M.P. Morphological and functional correlates of borders in the entorhinal cortex and hippocampus. *Hippocampus* **3**, 303-311 (1993).
198. Mercer, A., Trigg, H.L. & Thomson, A.M. Characterization of Neurons in the CA2 Subfield of the Adult Rat Hippocampus. *J. Neurosci.* **27**, 7329-7338 (2007).
199. Lein, E.S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168 - 176 (2006).
200. Llorens-Martín, M., Jurado-Arjona, J., Avila, J. & Hernández, F. Novel connection between newborn granule neurons and the hippocampal CA2 field. *Exp. Neurol.* **263**, 285-292 (2015).

201. Karbowski, J. Global and regional brain metabolic scaling and its functional consequences. *BMC Biol.* **5**, 18 (2007).
202. Zielke, H.R., Zielke, C.L. & Baab, P.J. Direct measurement of oxidative metabolism in the living brain by microdialysis - Review. *J. Neurochem.* **109**, 24-29 (2009).
203. Obel, L.F. *et al.* Brain glycogen-new perspectives on its metabolic function and regulation at the subcellular level. *Front. Neuroenergetics* **4**, 3 (2012).
204. Bak, L.K., Walls, A.B., Schousboe, A. & Waagepetersen, H.S. Astrocytic glycogen metabolism in the healthy and diseased brain. *J. Biol. Chem.* **293**, 7108-7116 (2018).
205. Agostini, M. *et al.* Metabolic reprogramming during neuronal differentiation. *Cell Death Differ.* **23**, 1502-1514 (2016).
206. Zheng, X. *et al.* Metabolic reprogramming during neuronal differentiation from aerobic glycolysis to neuronal oxidative phosphorylation. *eLife* **5**, e13374 (2016).
207. Mangia, S. *et al.* Response to 'comment on recent modeling studies of astrocyte-neuron metabolic interactions': much ado about nothing. *J. Cereb. Blood Flow Metab.* **31**, 1346-1353 (2011).
208. Magistretti, P.J. & Allaman, I. A cellular perspective on brain energy metabolism and functional imaging. *Neuron* **86**, 883-901 (2015).
209. Aubert, A. & Costalat, R. Interaction between Astrocytes and Neurons Studied using a Mathematical Model of Compartmentalized Energy Metabolism. *J. Cereb. Blood Flow Metab.* **25**, 1476-1490 (2005).
210. Pellerin, L. *et al.* Activity-dependent regulation of energy metabolism by astrocytes: An update. *Glia* **55**, 1251-1262 (2007).
211. DiNuzzo, M., Mangia, S., Maraviglia, B. & Giove, F. Changes in glucose uptake rather than lactate shuttle take center stage in subserving neuroenergetics: evidence from mathematical modeling. *J. Cereb. Blood Flow Metab.* **30**, 586-602 (2010).
212. Simpson, I.A., Carruthers, A. & Vannucci, S.J. Supply and Demand in Cerebral Energy Metabolism: The Role of Nutrient Transporters. *J. Cereb. Blood Flow Metab.* **27**, 1766-1791 (2007).
213. Hall, C.N., Klein-Flügge, M.C., Howarth, C. & Attwell, D. Oxidative Phosphorylation, Not Glycolysis, Powers Presynaptic and Postsynaptic Mechanisms Underlying Brain Information Processing. *J. Neurosci.* **32**, 8940-8951 (2012).
214. Mongeon, R., Venkatachalam, V. & Yellen, G. Cytosolic NADH-NAD+ Redox Visualized in Brain Slices by Two-Photon Fluorescence Lifetime Biosensor Imaging. *Antioxi. Redox Signal.* **25**, 553-563 (2016).
215. Ivanov, A.I. *et al.* Glycolysis and Oxidative Phosphorylation in Neurons and Astrocytes during Network Activity in Hippocampal Slices. *J. Cereb. Blood Flow Metab.* **34**, 397-407 (2013).
216. Zhou, Y. & Danbolt, N.C. Glutamate as a neurotransmitter in the healthy brain. *J. Neural. Transm.* **121**, 799-817 (2014).
217. Hassel, B. & Dingledine, R. Chapter 17 - Glutamate and Glutamate Receptors, in *Basic Neurochemistry, 8th edition* 342-366 (Academic Press, New York; 2012).
218. Choi, D.W. Chapter 6 - Glutamate receptors and the induction of excitotoxic neuronal death, in *Progress in Brain Research* 47-51 (Elsevier, Amsterdam; 1994).
219. Sinor, J.D. *et al.* NMDA and Glutamate Evoke Excitotoxicity at Distinct Cellular Locations in Rat Cortical Neurons In Vitro. *J. Neurosci.* **20**, 8831-8837 (2000).
220. Hynd, M.R., Scott, H.L. & Dodd, P.R. Glutamate-mediated excitotoxicity and neurodegeneration in Alzheimer's disease. *Neurochem. Int.* **45**, 583-595 (2004).
221. Gross, O.P. & von Gersdorff, H. Recycling at synapses. *eLife* **5**, e17692 (2016).
222. Gandhi, S.P. & Stevens, C.F. Three modes of synaptic vesicular recycling revealed by single-vesicle imaging. *Nature* **423**, 607 - 613 (2003).
223. Siddoway, B., Hou, H. & Xia, H. Glutamatergic Synapses: Molecular Organisation, in *Encyclopedia of Life Sciences* (John Wiley & Sons, New Jersey; 2011).

224. Barco, A., Bailey, C.H. & Kandel, E.R. Common molecular mechanisms in explicit and implicit memory. *J. Neurochem.* **97**, 1520-1533 (2006).
225. Verkhratsky, A. & Kirchhoff, F. NMDA Receptors in Glia. *Neuroscientist* **13**, 28-37 (2007).
226. Swanson, C.J. *et al.* Metabotropic glutamate receptors as novel targets for anxiety and stress disorders. *Nat. Rev. Drug Discov.* **4**, 131-144 (2005).
227. Rodríguez, A. & Ortega, A. Glutamine/Glutamate Transporters in Glial Cells: Much More Than Participants of a Metabolic Shuttle, in *Glial Amino Acid Transporters* 169-183 (Springer International Publishing, Cham; 2017).
228. Schousboe, A., Scafidi, S., Bak, L.K., Waagepetersen, H.S. & McKenna, M.C. Glutamate metabolism in the brain focusing on astrocytes. *Adv. Neurobiol.* **11**, 13-30 (2014).
229. Olsen, R.W. & Li, G.-D. Chapter 18 - GABA, in *Basic Neurochemistry, 8th edition* 367-376 (Academic Press, New York; 2012).
230. Bao, X. *et al.* Transgenic expression of Glud1 (glutamate dehydrogenase 1) in neurons: in vivo model of enhanced glutamate release, altered synaptic plasticity, and selective neuronal vulnerability. *J. Neurosci.* **29**, 13929-13944 (2009).
231. Frigerio, F. *et al.* Deletion of glutamate dehydrogenase 1 (Glud1) in the central nervous system affects glutamate handling without altering synaptic transmission. *J. Neurochem.* **123**, 342-348 (2012).
232. Raizen, D.M. *et al.* Central nervous system hyperexcitability associated with glutamate dehydrogenase gain of function mutations. *J. Pediatr.* **146**, 388-394 (2005).
233. Whitelaw, B.S. & Robinson, M.B. Inhibitors of glutamate dehydrogenase block sodium-dependent glutamate uptake in rat brain membranes. *Front. Endocrinol.* **4**, 123 (2013).
234. Szulwach, K.E. *et al.* 5-hmC–mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat. Neurosci.* **14**, 1607 - 1616 (2011).
235. Moore, L.D., Le, T. & Fan, G. DNA Methylation and Its Basic Function. *Neuropsychopharmacology* **38**, 23 - 28 (2012).
236. Levenson, J.M. *et al.* Evidence that DNA (cytosine-5) methyltransferase regulates synaptic plasticity in the hippocampus. *J. Biol. Chem.* **281**, 15763-15773 (2006).
237. Miller, C.A. & Sweatt, J.D. Covalent modification of DNA regulates memory formation. *Neuron* **53**, 857-869 (2007).
238. Goto, K. *et al.* Expression of DNA methyltransferase gene in mature and immature neurons as well as proliferating cells in mice. *Differentiation* **56**, 39-44 (1994).
239. Jian, F., Hua, C., En, L. & Guoping, F. Dynamic expression of de novo DNA methyltransferases Dnmt3a and Dnmt3b in the central nervous system. *J. Neurosci. Res.* **79**, 734-746 (2005).
240. Fan, G. *et al.* DNA Hypomethylation Perturbs the Function and Survival of CNS Neurons in Postnatal Animals. *J. Neurosci.* **21**, 788-797 (2001).
241. Hutnick, L.K. *et al.* DNA hypomethylation restricted to the murine forebrain induces cortical degeneration and impairs postnatal neuronal maturation. *Hum. Mol. Genet.* **18**, 2875-2888 (2009).
242. Wu, H. *et al.* Dnmt3a-Dependent Nonpromoter DNA Methylation Facilitates Transcription of Neurogenic Genes. *Science* **329**, 444-448 (2010).
243. Feng, J. *et al.* Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. *Nat. Neurosci.* **13**, 423-430 (2010).
244. Xie, W. *et al.* Base-resolution Analyses of Sequence and Parent-of-Origin Dependent DNA Methylation in the Mouse Genome. *Cell* **148**, 816-831 (2012).
245. Varley, K.E. *et al.* Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res.* **23**, 555-567 (2013).
246. De Felipe, J., Marco, P., Fairen, A. & G Jones, E. Inhibitory synaptogenesis in mouse somatosensory cortex. *Cereb. Cortex* **7**, 619-634 (1996).
247. Hahn, M.A. *et al.* Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis. *Cell Rep.* **3**, 291-300 (2013).

248. Song, C.X. *et al.* Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat. Biotechnol.* **29**, 68-72 (2011).
249. Wheldon, L.M. *et al.* Transient accumulation of 5-carboxylcytosine indicates involvement of active demethylation in lineage specification of neural stem cells. *Cell Rep.* **7**, 1353-1361 (2014).
250. Liu, Y. *et al.* Oligodendrocyte and astrocyte development in rodents: an in situ and immunohistological analysis during embryonic development. *Glia* **40**, 25-43 (2002).
251. Delaunay, D. *et al.* Early neuronal and glial fate restriction of embryonic neural stem cells. *J. Neurosci.* **28**, 2551-2562 (2008).
252. Colquitt, B.M., Allen, W.E., Barnea, G. & Lomvardas, S. Alteration of genic 5-hydroxymethylcytosine patterning in olfactory neurons correlates with changes in gene expression and cell identity. *Proc. Natl. Acad. Sci. USA* **110**, 14682-14687 (2013).
253. Münzel, M. *et al.* Quantification of the Sixth DNA Base Hydroxymethylcytosine in the Brain. *Angew. Chem. Int. Ed. Engl.* **49**, 5375-5377 (2010).
254. Rose, N.R. & Klose, R.J. Understanding the relationship between DNA methylation and histone lysine methylation. *Biochim. Biophys. Acta Gene Regul. Mech.* **1839**, 1362-1372 (2014).
255. Morey, L. & Helin, K. Polycomb group protein-mediated repression of transcription. *Trends Biochem. Sci.* **35**, 323-332 (2010).
256. Nakayama, J.-i., Rice, J.C., Strahl, B.D., Allis, C.D. & Grewal, S.I.S. Role of Histone H3 Lysine 9 Methylation in Epigenetic Control of Heterochromatin Assembly. *Science* **292**, 110-113 (2001).
257. Creyghton, M.P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. USA* **107**, 21931-21936 (2010).
258. Jin, S.G., Wu, X., Li, A.X. & Pfeifer, G.P. Genomic mapping of 5-hydroxymethylcytosine in the human brain. *Nucleic Acids Res.* **39**, 5015-5024 (2011).
259. Fasolino, M., Welsh, S.A. & Zhou, Z. TET and 5hmC in Neurodevelopment and the Adult Brain, in *DNA Modifications in the Brain*. (ed. T.W. Bredy) 61-79 (2017).
260. Yao, B. *et al.* DNA N6-methyladenine is dynamically regulated in the mouse brain following environmental stress. *Nat. Commun.* **8**, 1122 (2017).
261. Li, X. *et al.* The DNA modification N6-methyl-2'-deoxyadenosine (m6dA) drives activity-induced gene expression and is required for fear extinction. *Nat. Neurosci.* **22**, 534-544 (2019).
262. Schiffers, S. *et al.* Quantitative LC–MS Provides No Evidence for m6dA or m4dC in the Genome of Mouse Embryonic Stem Cells and Tissues. *Angew. Chem. Int. Ed. Engl.* **56**, 11268-11271 (2017).
263. Ma, D.K. *et al.* Neuronal Activity–Induced Gadd45b Promotes Epigenetic DNA Demethylation and Adult Neurogenesis. *Science* **323**, 1074-1077 (2009).
264. Guo, J.U. *et al.* Neuronal activity modifies DNA methylation landscape in the adult brain. *Nat. Neurosci.* **14**, 1345-1351 (2011).
265. Martinowich, K. *et al.* DNA Methylation-Related Chromatin Remodeling in Activity-Dependent Bdnf Gene Regulation. *Science* **302**, 890-893 (2003).
266. Huang, E.J. & Reichardt, L.F. Neurotrophins: Roles in Neuronal Development and Function. *Annu. Rev. Neurosci.* **24**, 677-736 (2001).
267. Bekinschtein, P. *et al.* BDNF is essential to promote persistence of long-term memory storage. *Proc. Natl. Acad. Sci. USA* **105**, 2711-2716 (2008).
268. Jin, S.G. *et al.* Tet3 Reads 5-Carboxylcytosine through Its CXXC Domain and Is a Potential Guardian against Neurodegeneration. *Cell Rep.* **14**, 493-505 (2016).
269. Dawlaty, M.M. *et al.* Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell* **9**, 166-175 (2011).
270. Rudenko, A. *et al.* Tet1 is critical for neuronal activity-regulated gene expression and memory extinction. *Neuron* **79**, 1109-1122 (2013).
271. Zhang, R.R. *et al.* Tet1 regulates adult hippocampal neurogenesis and cognition. *Cell Stem Cell* **13**, 237-245 (2013).

272. Coutellier, L., Beraki, S., Ardestani, P.M., Saw, N.L. & Shamloo, M. Npas4: A Neuronal Transcription Factor with a Key Role in Social and Cognitive Functions Relevant to Developmental Disorders. *PLoS One* **7**, e46604 (2012).
273. Kaas, G.A. *et al.* TET1 controls CNS 5-methylcytosine hydroxylation, active DNA demethylation, gene transcription, and memory formation. *Neuron* **79**, 1086-1093 (2013).
274. Guo, J.U., Su, Y., Zhong, C., Ming, G.L. & Song, H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145**, 423-434 (2011).
275. Jessop, P. & Toledo-Rodriguez, M. Hippocampal TET1 and TET2 Expression and DNA Hydroxymethylation Are Affected by Physical Exercise in Aged Mice. *Front. Cell Dev. Biol.* **6**, 45 (2018).
276. Dawlaty, M.M. *et al.* Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. *Dev. Cell* **24**, 310-323 (2013).
277. Tan, L. & Shi, Y.G. Tet family proteins and 5-hydroxymethylcytosine in development and disease. *Development* **139**, 1895-1902 (2012).
278. Li, X. *et al.* Neocortical Tet3-mediated accumulation of 5-hydroxymethylcytosine promotes rapid behavioral adaptation. *Proc. Natl. Acad. Sci. USA* **111**, 7120-7125 (2014).
279. Quirk, G.J. & Mueller, D. Neural Mechanisms of Extinction Learning and Retrieval. *Neuropsychopharmacology* **33**, 56-72 (2008).
280. Südhof, T.C. & Malenka, R.C. Understanding synapses: past, present, and future. *Neuron* **60**, 469-476 (2008).
281. Turrigiano, G. Homeostatic synaptic plasticity: local and global mechanisms for stabilizing neuronal function. *Cold Spring Harb. Perspect. Biol.* **4**, a005736 (2012).
282. Lee, H.-K. Ca-permeable AMPA receptors in homeostatic synaptic plasticity. *Front. Mol. Neurosci.* **5**, 17 (2012).
283. Fatt, P. & Katz, B. Spontaneous subthreshold activity at motor nerve endings. *J. Physiol.* **117**, 109-128 (1952).
284. Sibarov, D.A. & Antonov, S.M. Characteristics of Postsynaptic Currents in Primary Cultures of Rat Cerebral Cortical Neurons. *Neurosci. Behav. Physiol.* **45**, 431-439 (2015).
285. Poo, M.-M. Neurotrophins as synaptic modulators. *Nat. Rev. Neurosci.* **2**, 24-32 (2001).
286. Rutherford, L.C., Nelson, S.B. & Turrigiano, G.G. BDNF Has Opposite Effects on the Quantal Amplitude of Pyramidal Neuron and Interneuron Excitatory Synapses. *Neuron* **21**, 521-530 (1998).
287. Bray, N. Upscale, downscale. *Nat. Rev. Neurosci.* **19**, 184 (2018).
288. Iwan, K. *et al.* 5-Formylcytosine to cytosine conversion by C–C bond cleavage in vivo. *Nat. Chem. Biol.* **14**, 72 - 78 (2017).
289. Guan, J.-S. *et al.* HDAC2 negatively regulates memory formation and synaptic plasticity. *Nature* **459**, 55 - 60 (2009).
290. Fischer, A., Sananbenesi, F., Wang, X., Dobbin, M. & Tsai, L.H. Recovery of learning and memory is associated with chromatin remodelling. *Nature* **447**, 178-182 (2007).
291. Broide, R.S. *et al.* Distribution of histone deacetylases 1–11 in the rat brain. *J. Mol. Neurosci.* **31**, 47-58 (2007).
292. Davie, J.R. Inhibition of Histone Deacetylase Activity by Butyrate. *J. Nutr.* **133**, 2485 -2493 (2003).
293. Kim, M.-S. *et al.* An Essential Role for Histone Deacetylase 4 in Synaptic Plasticity and Memory Formation. *J. Neurosci.* **32**, 10879-10886 (2012).
294. Wittmann, M. *et al.* Synaptic activity induces dramatic changes in the geometry of the cell nucleus: interplay between nuclear structure, histone H3 phosphorylation, and nuclear calcium signaling. *J. Neurosci.* **29**, 14687-14700 (2009).
295. Rossetto, D., Avvakumov, N. & Cote, J. Histone phosphorylation: a chromatin modification involved in diverse nuclear events. *Epigenetics* **7**, 1098-1108 (2012).
296. Qureshi, I.A. & Mehler, M.F. Emerging roles of non-coding RNAs in brain evolution, development, plasticity and disease. *Nat. Rev. Neurosci.* **13**, 528 - 541 (2012).

297. Klein, M.E. *et al.* Homeostatic regulation of MeCP2 expression by a CREB-induced microRNA. *Nat. Neurosci.* **10**, 1513-1514 (2007).
298. Alvarez-Saavedra, M. *et al.* miRNA-132 orchestrates chromatin remodeling and translational control of the circadian clock. *Hum. Mol. Genet.* **20**, 731-751 (2011).
299. Briggs, J.A., Wolvetang, E.J., Mattick, J.S., Rinn, J.L. & Barry, G. Mechanisms of Long Non-coding RNAs in Mammalian Nervous System Development, Plasticity, Disease, and Evolution. *Neuron* **88**, 861-877 (2015).
300. Mercer, T.R., Dinger, M.E., Sunkin, S.M., Mehler, M.F. & Mattick, J.S. Specific expression of long noncoding RNAs in the mouse brain. *Proc. Natl. Acad. Sci. USA* **105**, 716-721 (2008).
301. Feng, J. *et al.* The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. *Genes Dev.* **20**, 1470-1484 (2006).
302. Bernard, D. *et al.* A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. *EMBO J.* **29**, 3082-3093 (2010).
303. Mercer, T.R. *et al.* Long noncoding RNAs in neuronal-glial fate specification and oligodendrocyte lineage maturation. *BMC Neurosci.* **11**, 14 (2010).
304. Ng, S.Y., Johnson, R. & Stanton, L.W. Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *EMBO J.* **31**, 522 - 533 (2012).
305. Khalil, A.M. *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. USA* **106**, 11667-11672 (2009).
306. Bienvenu, T. & Chelly, J. Molecular genetics of Rett syndrome: when DNA methylation goes unrecognized. *Nat. Rev. Genet.* **7**, 415-426 (2006).
307. Kishi, N. & Macklis, J.D. MECP2 is progressively expressed in post-migratory neurons and is involved in neuronal maturation rather than cell fate decisions. *Mol. Cell. Nerosci.* **27**, 306-321 (2004).
308. Miltenberger-Miltenyi, G. & Laccone, F. Mutations and polymorphisms in the human methyl CpG-binding protein MECP2. *Hum. Mutat.* **22**, 107-115 (2003).
309. Neul, J.L. *et al.* Rett syndrome: revised diagnostic criteria and nomenclature. *Ann. Neurol.* **68**, 944-950 (2010).
310. Bodda, C. *et al.* Mild Overexpression of Mecp2 in Mice Causes a Higher Susceptibility toward Seizures. *Am J. Pathol.* **183**, 195-210 (2013).
311. Chen, R.Z., Akbarian, S., Tudor, M. & Jaenisch, R. Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. *Nat. Genet.* **27**, 327 - 331 (2001).
312. Guy, J., Gan, J., Selfridge, J., Cobb, S. & Bird, A. Reversal of Neurological Defects in a Mouse Model of Rett Syndrome. *Science* **315**, 1143-1147 (2007).
313. Luikenhuis, S., Giacometti, E., Beard, C.F. & Jaenisch, R. Expression of MeCP2 in postmitotic neurons rescues Rett syndrome in mice. *Proc. Natl. Acad. Sci. USA* **101**, 6033-6038 (2004).
314. Reichwald, K. *et al.* Comparative sequence analysis of the MECP2-locus in human and mouse reveals new transcribed regions. *Mamm. Genome* **11**, 182 - 190 (2000).
315. Shahbazian, M.D. *et al.* Mice with Truncated MeCP2 Recapitulate Many Rett Syndrome Features and Display Hyperacetylation of Histone H3. *Neuron* **35**, 243-254 (2002).
316. Szczesna, K. *et al.* Improvement of the Rett syndrome phenotype in a MeCP2 mouse model upon treatment with levodopa and a dopa-decarboxylase inhibitor. *Neuropsychopharmacology* **39**, 2846-2856 (2014).
317. Tudor, M., Akbarian, S., Chen, R.Z. & Jaenisch, R. Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional changes in the brain. *Proc. Natl. Acad. Sci. USA* **99**, 15536-15541 (2002).
318. Vogel Ciernia, A. *et al.* Early motor phenotype detection in a female mouse model of Rett syndrome is improved by cross-fostering. *Hum. Mol. Genet.* **26**, 1839-1854 (2017).
319. McGraw, C.M., Samaco, R.C. & Zoghbi, H.Y. Adult neural function requires MeCP2. *Science* **333**, 186-186 (2011).

320. Nguyen, M.V.C. *et al.* MeCP2 Is Critical for Maintaining Mature Neuronal Networks and Global Brain Anatomy during Late Stages of Postnatal Brain Development and in the Mature Adult Brain. *J. Neurosci.* **32**, 10021-10034 (2012).
321. Klose, R.J. *et al.* DNA Binding Selectivity of MeCP2 Due to a Requirement for A/T Sequences Adjacent to Methyl-CpG. *Mol. Cell* **19**, 667-678 (2005).
322. Nan, X., Campoy, F.J. & Bird, A. MeCP2 Is a Transcriptional Repressor with Abundant Binding Sites in Genomic Chromatin. *Cell* **88**, 471-481 (1997).
323. Bird, A. & Kriaucionis, S. The major form of MeCP2 has a novel N-terminus generated by alternative splicing. *Nucleic Acids Res.* **32**, 1818-1823 (2004).
324. Mnatzakanian, G.N. *et al.* A previously unidentified MECP2 open reading frame defines a new protein isoform relevant to Rett syndrome. *Nat. Genet.* **36**, 339 - 341 (2004).
325. Dannenberg, J.-H. *et al.* mSin3A corepressor regulates diverse transcriptional networks governing normal and neoplastic growth and survival. *Genes Dev.* **19**, 1581-1595 (2005).
326. Lyst, M.J. *et al.* Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nat. Neurosci.* **16**, 898-902 (2013).
327. Jones, P.L. *et al.* Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet.* **19**, 187-191 (1998).
328. Nan, X. *et al.* Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**, 386-389 (1998).
329. Della Sala, G. & Pizzorusso, T. Synaptic plasticity and signaling in Rett syndrome. *Dev. Neurobiol.* **74**, 178-196 (2014).
330. Charman, T. *et al.* Dimensional phenotypic analysis and functional categorisation of mutations reveal novel genotype–phenotype associations in Rett syndrome. *Eur. J. Hum. Genet.* **13**, 1121 - 1130 (2005).
331. Sharma, K., Singh, J., Frost, E.E. & Pillai, P.P. MeCP2 in central nervous system glial cells: current updates. *Acta Neurobiol. Exp.* **78**, 30-40 (2018).
332. Khrapunov, S. *et al.* Unusual characteristics of the DNA binding domain of epigenetic regulatory protein MeCP2 determine its binding specificity. *Biochemistry* **53**, 3379-3391 (2014).
333. Jordan, C., Li, H.H., Kwan, H.C. & Francke, U. Cerebellar gene expression profiles of mouse models for Rett syndrome reveal novel MeCP2 targets. *BMC Med. Genet.* **8**, 36 (2007).
334. Colantuoni, C. *et al.* Gene Expression Profiling in Postmortem Rett Syndrome Brain: Differential Gene Expression and Patient Classification. *Neurobiol. Dis.* **8**, 847-865 (2001).
335. Shovlin, S. & Tropea, D. Transcriptome level analysis in Rett syndrome using human samples from different tissues. *Orphanet J. Rare Dis.* **13**, 113 (2018).
336. Chahrour, M. *et al.* MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science* **320**, 1224-1229 (2008).
337. Young, J.I. *et al.* Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. *Proc. Natl. Acad. Sci. USA* **102**, 17551-17558 (2005).
338. Chen, W.G. *et al.* Derepression of BDNF Transcription Involves Calcium-Dependent Phosphorylation of MeCP2. *Science* **302**, 885-889 (2003).
339. Cholewa-Waclaw, J. *et al.* Quantitative modelling predicts the impact of DNA methylation on RNA polymerase II traffic. *Proc. Natl. Acad. Sci. USA* **116**, 14995-15000 (2019).
340. Maunakea, A.K., Chepelev, I., Cui, K. & Zhao, K. Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res.* **23**, 1256 - 1269 (2013).
341. Monteggia, L.M. & Kavalali, E.T. Rett Syndrome and the Impact of MeCP2 Associated Transcriptional Mechanisms on Neurotransmission. *Biol. Psychiatry* **65**, 204-210 (2009).
342. Nimchinsky, E.A., Sabatini, B.L. & Svoboda, K. Structure and function of dendritic spines. *Annu. Rev. Physiol.* **64**, 313-353 (2002).
343. Tau, G.Z. & Peterson, B.S. Normal development of brain circuits. *Neuropsychopharmacology* **35**, 147-168 (2010).

344. Zhou, S. & Yu, Y. Synaptic E-I Balance Underlies Efficient Neural Coding. *Front. Neurosci.* **12**, 46 (2018).
345. Dani, V.S. *et al.* Reduced cortical activity due to a shift in the balance between excitation and inhibition in a mouse model of Rett syndrome. *Proc. Natl. Acad. Sci. USA* **102**, 12560-12565 (2005).
346. Nelson, E.D., Kavalali, E.T. & Monteggia, L.M. MeCP2-dependent transcriptional repression regulates excitatory neurotransmission. *Curr. Biol.* **16**, 710-716 (2006).
347. Asaka, Y., Jugloff, D.G.M., Zhang, L., Eubanks, J.H. & Fitzsimonds, R.M. Hippocampal synaptic plasticity is impaired in the Mecp2-null mouse model of Rett syndrome. *Neurobiol. Dis.* **21**, 217-227 (2006).
348. Moretti, P. *et al.* Learning and memory and synaptic plasticity are impaired in a mouse model of Rett syndrome. *J. Neurosci.* **26**, 319-327 (2006).
349. Chao, H.-T., Zoghbi, H.Y. & Rosenmund, C. MeCP2 controls excitatory synaptic strength by regulating glutamatergic synapse number. *Neuron* **56**, 58-65 (2007).
350. Chao, H.-T. *et al.* Dysfunction in GABA signalling mediates autism-like stereotypies and Rett syndrome phenotypes. *Nature* **468**, 263-269 (2010).
351. Maezawa, I., Swanberg, S., Harvey, D., LaSalle, J.M. & Jin, L.W. Rett syndrome astrocytes are abnormal and spread MeCP2 deficiency through gap junctions. *J. Neurosci.* **29**, 5051-5061 (2009).
352. Kumar, N.M. & Gilula, N.B. The Gap Junction Communication Channel. *Cell* **84**, 381-388 (1996).
353. Maezawa, I. & Jin, L.-W. Rett syndrome microglia damage dendrites and synapses by the elevated release of glutamate. *J. Neurosci.* **30**, 5346-5356 (2010).
354. Rakela, B., Brehm, P. & Mandel, G. Astrocytic modulation of excitatory synaptic signaling in a mouse model of Rett syndrome. *eLife* **7**, e31629 (2018).
355. Lehtonen, H.J. *et al.* Increased risk of cancer in patients with fumarate hydratase germline mutation. *J. Med. Genet.* **43**, 523-526 (2006).
356. King, A., Selak, M.A. & Gottlieb, E. Succinate dehydrogenase and fumarate hydratase: linking mitochondrial dysfunction and cancer. *Oncogene* **25**, 4675 - 4682 (2006).
357. Lloyd, S.J., Lauble, H., Prasad, G.S. & Stout, C.D. The mechanism of aconitase: 1.8 A resolution crystal structure of the S642a:citrate complex. *Protein Sci.* **8**, 2655-2662 (1999).
358. Beeckmanns, S. & Kanarek, L. Demonstration of Physical Interactions between Consecutive Enzymes of the Citric Acid Cycle and of the Aspartate-Malate Shuttle. *Eur. J. Biochem.* **117**, 527-535 (1981).
359. Hohnholt, M.C. *et al.* Glutamate dehydrogenase is essential to sustain neuronal oxidative energy metabolism during stimulation. *J. Cereb. Blood Flow Metab.* **38**, 1754-1768 (2018).
360. Lee, J.-H. *et al.* A Robust Approach to Identifying Tissue-Specific Gene Expression Regulatory Variants Using Personalized Human Induced Pluripotent Stem Cells. *PLoS Genet.* **5**, e1000718 (2009).
361. Sun, Y. *et al.* Neurogenin Promotes Neurogenesis and Inhibits Glial Differentiation by Independent Mechanisms. *Cell* **104**, 365-376 (2001).
362. Thoma, E.C. *et al.* Ectopic Expression of Neurogenin 2 Alone is Sufficient to Induce Differentiation of Embryonic Stem Cells into Mature Neurons. *PLoS One* **7**, e38651 (2012).
363. Raghavendra Rao, V.L. & Murthy, C.R.K.J.N.R. Uptake and metabolism of glutamate and aspartate by astroglial and neuronal preparations of rat cerebellum. *Neurochem. Res.* **18**, 647-654 (1993).
364. Divito, C.B. & Underhill, S.M. Excitatory amino acid transporters: roles in glutamatergic neurotransmission. *Neurochem. Int.* **73**, 172-180 (2014).
365. Lam, R.S., Topfer, F.M., Wood, P.G., Busskamp, V. & Bamberg, E. Functional Maturation of Human Stem Cell-Derived Neurons in Long-Term Cultures. *PLoS One* **12**, e0169506 (2017).
366. Wahlin, K.J. *et al.* Epigenetics and cell death: DNA hypermethylation in programmed retinal cell death. *PLoS One* **8**, e79140 (2013).

367. Naniche, N., Sau, D. & Pasinelli, P. In vivo and in vitro determination of cell death markers in neurons. *Methods Mol. Biol.* **793**, 9-21 (2011).
368. Cai, J., Yang, J. & Jones, D. Mitochondrial control of apoptosis: the role of cytochrome c. *Biochim. Biophys. Acta Bioenerg.* **1366**, 139-149 (1998).
369. Tarasov, V. *et al.* Differential regulation of microRNAs by p53 revealed by massively parallel sequencing: miR-34a is a p53 target that induces apoptosis and G1-arrest. *Cell Cycle* **6**, 1586-1593 (2007).
370. Martinet, W. *et al.* Gene Expression Profiling of Apoptosis-Related Genes in Human Atherosclerosis. *Arter. Thromb. Vasc. Biol.* **22**, 2023-2029 (2002).
371. Levy-Strumpf, N. & Kimchi, A. Death associated proteins (DAPs): from gene identification to the analysis of their apoptotic and tumor suppressive functions. *Oncogene* **17**, 3331 - 3340 (1998).
372. Perera, A. *et al.* TET3 Is Recruited by REST for Context-Specific Hydroxymethylation and Induction of Gene Expression. *Cell Rep.* **11**, 283-294 (2015).
373. Tyssowski, K.M. *et al.* Different Neuronal Activity Patterns Induce Different Gene Expression Programs. *Neuron* **98**, 530 - 546 (2018).
374. Choi, D.W., Viseskul, V., Amirthanayagam, M. & Monyer, H. Aspartate neurotoxicity on cultured cortical neurons. *J. Neurosci. Res.* **23**, 116-121 (1989).
375. Thornburg, J.M. *et al.* Targeting aspartate aminotransferase in breast cancer. *Breast Cancer Res.* **10**, R84 (2008).
376. Traube, F.R. *et al.* Isotope-dilution mass spectrometry for exact quantification of noncanonical DNA nucleosides. *Nat. Protoc.* **14**, 283-312 (2019).
377. Guillemin, I., Becker, M., Ociepka, K., Friauf, E. & Nothwang, H.G. A subcellular prefractionation protocol for minute amounts of mammalian cell cultures and tissue. *Proteomics* **5**, 35-45 (2005).
378. Kulak, N.A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **11**, 319 - 324 (2014).
379. Vizcaíno, J.A. *et al.* 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **44**, D447-D456 (2016).
380. Chen, Y. *et al.* NS21: re-defined and modified supplement B27 for neuronal cultures. *J. Neurosci. Methods* **171**, 239-247 (2008).
381. Pfaffl, M.W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, e45 (2001).
382. Pfaffl, M.W., Horgan, G.W. & Dempfle, L. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res.* **30**, e36 (2002).
383. Maunakea, A.K., Chepelev, I., Cui, K. & Zhao, K. Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res.* **23**, 1256 - 1269 (2013).
384. Maxwell, S.S., Pelka, G.J., Tam, P.P.L. & El-Osta, A. Chromatin context and ncRNA highlight targets of MeCP2 in brain. *RNA Biol.* **10**, 1741-1757 (2013).
385. Gabel, H.W. *et al.* Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature* **522**, 89 - 93 (2015).
386. Nott, A. *et al.* Histone deacetylase 3 associates with MeCP2 to regulate FOXO and social behavior. *Nat. Neurosci.* **19**, 1497 - 1505 (2016).
387. Kinde, B., Wu, D.Y., Greenberg, M.E. & Gabel, H.W. DNA methylation in the gene body influences MeCP2-mediated gene repression. *Proc. Natl. Acad. Sci. USA* **113**, 15114-15119 (2016).
388. Kaddoum, L. *et al.* Isoform-specific anti-MeCP2 antibodies confirm that expression of the e1 isoform strongly predominates in the brain. *F1000Res* **2**, 204 (2013).

389. Bueno, C., Tabares-Seisdedos, R., Moraleda, J.M. & Martinez, S. Rett Syndrome Mutant Neural Cells Lacks MeCP2 Immunoreactive Bands. *PLoS One* **11**, e0153262 (2016).
390. Huang, X. & Miller, W. A time-efficient, linear-space local similarity algorithm. *Adv. Appl. Math.* **12**, 337-357 (1991).
391. Grabrucker, A., Vaida, B., Bockmann, J. & Boeckers, T.M. Synaptogenesis of hippocampal neurons in primary cell culture. *Cell Tissue Res.* **338**, 333 - 341 (2009).
392. Lam, R.S., Töpfer, F.M., Wood, P.G., Busskamp, V. & Bamberg, E. Functional Maturation of Human Stem Cell-Derived Neurons in Long-Term Cultures. *PLoS One* **12**, e0169506 (2017).
393. Cesca, F., Baldelli, P., Valtorta, F. & Benfenati, F. The synapsins: Key actors of synapse function and plasticity. *Prog. Neurobiol.* **91**, 313-348 (2010).
394. Costa-Silva, J., Domingues, D. & Lopes, F.M. RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS One* **12**, e0190152 (2017).
395. Cho, W.C.S. Proteomics Technologies and Challenges. *Genom. Proteom. Bioinf.* **5**, 77-85 (2007).
396. Liu, Y., Beyer, A. & Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165**, 535-550 (2016).
397. Wang, D. *et al.* A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol. Syst. Biol.* **15**, e8503 (2019).
398. Gaffuri, A.L., Ladarre, D. & Lenkei, Z. Type-1 Cannabinoid Receptor Signaling in Neuronal Development. *Pharmacology* **90**, 19-39 (2012).
399. Gundelfinger, E.D., Reissner, C. & Garner, C.C. Role of Bassoon and Piccolo in Assembly and Molecular Organization of the Active Zone. *Front. Synaptic. Neurosci.* **7**, 19 (2016).
400. Dignam, J.D., Lebovitz, R.M. & Roeder, R.G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11**, 1475-1489 (1983).
401. Eden, E., Lipson, D., Yogev, S. & Yakhini, Z. Discovering Motifs in Ranked Lists of DNA sequences. *PLoS Comput. Biol.* **3**, e39 (2007).
402. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48 (2009).
403. Chevallet, M., Luche, S. & Rabilloud, T. Silver staining of proteins in polyacrylamide gels. *Nat. Protoc.* **1**, 1852-1858 (2006).
404. Schröder, A.S. *et al.* 2'-(R)-Fluorinated mC, hmC, fC and caC triphosphates are substrates for DNA polymerases and TET-enzymes. *Chem. Comm.* **52**, 14361-14364 (2016).
405. Wildenhof, T.M., Schiffers, S., Traube, F.R., Mayer, P. & Carell, T. Influencing Epigenetic Information with a Hydrolytically Stable Carbocyclic 5-Aza-2'-deoxycytidine. *Angew. Chem. Int. Ed. Engl.* **58**, 12984-12987 (2019).