technische universität
dortmund

Light-Activation of DNA-Methyltransferases

Dissertation

Submitted for the degree of

Doctor of Natural Sciences

(Dr. rer. nat.)

by

Jan Wolffgramm

TU Dortmund

Faculty of Chemistry and Chemical Biology

1st Referee: Prof. Dr. Daniel Summerer

2nd Referee: Prof. Dr. Hannes Mutschler

Dortmund, 2021

All truths are easy to understand once they are discovered - the point is to discover them.

Galileo Galilei

This work was supported by the TU Dortmund, the ERC Horizon 2020 program (ERC CoG EPICODE, Grant Nr. 723863 to Prof. Dr. Daniel Summerer), the Volkswagenstiftung (Lichtenberg program to Prof. Dr. Michal Ruth Schweiger), the DFG (KFO286) and the Center for Molecular Medicine Cologne (CMMC).

Experiments presented in this thesis were performed between July 2017 and February 2021 in the group of Prof. Dr. Daniel Summerer, TU Dortmund.

Pyrosequencing was performed in the group of Prof. Dr. Michal Ruth Schweiger, Center for Molecular Medicine Cologne.

Mass spectrometry of peptides was performed and analyzed in the project group of Dr. Petra Janning, Max Plank Institute of Molecular Physiology Dortmund.

Mainly, results of this work were published in *Angew. Chem. Int. Ed*. **2021**, 60, 24, 13507–13512 – an open access article enabled and organized by Projekt DEAL.

Acknowledgement

First and foremost, I thank Prof. Dr. Daniel Summerer for giving me the opportunity to work on this highly interesting and relevant topic in the field of epigenetics. Specifically, I am grateful for all discussions, help, new ideas and particular the relentless effort in promoting this work to be published.

Further, I thank Prof. Dr. Hannes Mutschler for being the second referee of this thesis.

I also thank the collaborator group of Prof. Dr. Michal Ruth Schweiger for fruitful discussions regarding SatIII targeted methylation. A special thanks goes to Julian Kanne and Felix Hess who performed the pyrosequencing of the SatIII locus. I also thank the project group of Dr. Petra Janning for measuring and analyzing the peptide masses. I thank the group of Prof. Dr. Daniel Rauh and the MPI of Molecular Physiology Dortmund for sharing their devices for measurements.

Especially important are all members of the Summerer group who helped if problems occurred, gave new thoughts and were always there for discussions and nice conversations. Namely I thank Dr. Álvaro Muñoz-López, Anne Jung, Brinja Kosel, Damian Schiller, Katharina Kuhr, Dr. Mario Gieß, Dr. Preeti Rathi, Dr. Sarah Flade, Dr. Shubhendu Palei, Simone Eppmann, and Sudakshina Banerjee. A special thanks goes to Dr. Benjamin Buchmuller for his massive help in analyzing bioinformatics data and Tzu-Chen Lin for proofreading this thesis. I also want to highlight Dr. Anna Witte, Nadine Schmidt and Dr. Sara Maurer who became friends during this time and who were always there, listened and gave good advice. A great thanks goes to Stefan Helmer for synthesizing the DMNB-Cys and cloning the synthetase/tRNA plasmid and to my former Bachelor student Alicia Hoffmann Benito for supporting me with her work effort during her bachelor thesis.

Not to forget and also especially important, I thank Maria Sergani, Martina Reibner, Petra Alhorn and Ulrich Schoppe from the administration for dealing with all kinds of general management and bureaucracy.

And last, but certainly not least, I thank my family and friends for the continuous support. I highly appreciate the encouragement I got from my parents and my brother. Also in more difficult times, they were there to strengthen the optimism which ultimately led to this thesis.

Table of Contents

List of Figures

List of Tables

Abbreviations

Abbreviation	Meaning
°C	Degrees Celsius
5-Aza	5-Aza-2'-deoxycytidine
5caC	5-Carboxylcytosine
5fC	5-Formylcytosine
5hmC	5-Hydroxymethylcytosine
5mC	5-Methylcytosine
A	Adenine
aaRS	Aminoacyl tRNA synthetase
AD	Activation domain
ADD	ATRX-DNMT3-DNMT3L
AF405	Alexa Fluor™ 405
AF488	Alexa Fluor™ 488
AML	Acute myeloid leukemia
APS	Ammonium peroxodisulfate
BAH	Bromo-adjacent homology
BER	Base excision repair
BRCA1	Breast cancer 1
BS	Bisulfite
BSA	Bovine serum albumin
C	Cytosine
CAM	Carbamidomethylation
Carb	Carbenicillin
CAT	Chloramphenicol acetyltransferase
CDKN2A	Cyclin-dependent kinase inhibitor 2A
CIB1	Cryptochrome-interacting basic-helix-loop-helix protein 1
CID	Chemical inducer of dimerization
CMV	Cytomegalovirus
CNB	α-Carboxy-2-nitrobenzyl
cPCR	Colony PCR
CRD	Central repeat domain
CRISPR-dCas9	Clustered regularly interspaced short palindromic repeats/deactivated CRISPR-associated protein 9
CRY2	Cryptochrome 2 protein
CTR	C-Terminal region
Da	Dalton

Abbreviation	Meaning
dATP	Deoxyadenosine triphosphate
DBD	DNA binding domain
dCTP	Deoxycytidine triphosphate
DEACM	Diethylaminocoumarinyl-4-methyl
dGTP	Deoxyguanosine triphosphate
DHFR	Dihydrofolate reductase
DKO	Double knock out
DMNB	4,5-Dimethoxy-2-nitrobenzyl
DMNPE	4,5-Dimethoxy-NPE
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DNMT3L	DNMT3-like
dNTPs	Deoxynucleoside triphosphates
DPBS	Dulbecco's phosphate buffered saline
DTT	Dithiothreitol
dTTP	Deoxythymidine triphosphate
E. coli	*Escherichia coli*
EcLeuRS	*E. coli* leucyl tRNA synthetase
EcTyrRS	*E. coli* tyrosyl tRNA synthetase
EDTA	Ethylenediaminetetraacetic acid
eRF	Eukaryotic release factors
ESI	Electrospray ionization
EtOH	Ethanol
FACS	Fluorescent activated cell sorting
FBS	Fetal bovine serum
FITC	Fluorescein
FLASH	Fast ligation-based automatable solid-phase high-throughput
fw	Forward
G	Guanine
GCE	Genetic code expansion
gDNA	Genomic DNA
GG1 or GG2	Golden gate reaction 1 or Golden gate reaction 2
gRNA	Guide RNA
h	Hour(s)
HAT	Histone acetyltransferase
HCD	High-energy collision-dissociation
HDAC	Histone deacetylase

Abbreviation	Meaning
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMT	Histone methyltransferase
Hrp	Hypersensitive response and pathogenicity
HS	Heat shock
HSF1	Heat shock factor 1
LCM-RRBS	Laser-capture microdissection RRBS
LINE	Long interspersed nuclear element
lnc	Long noncoding
LOV domain	Light-oxygen-voltage-sensing domain
LR	Last repeat
MBD	Methyl-binding domain
mESC	Murine embryonic stem cells
miRNA	MicroRNA
MNP	Methyl-*o*-nitropiperonyl
mRNA	Messenger RNA
NAD^+	Nicotinamide adenine dinucleotide
NB	*o*-Nitrobenzylic
ncAA	Non-canonical amino acid
ncRNA	Non-coding RNA
NGS	Next generation sequencing
NLS	Nuclear localization signals
NMR	Nuclear magnetic resonance
NP	*o*-Nitropiperonyl
NPE	*o*-Nitrophenylethyl
NPP	*o*-Nitrophenylpropyl
nSB	Nuclear stress body
nt	Nucleotides
NTR	N-Terminal region
PAM	Protospacer-adjacent motif
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEI	Polyethylenimine
PHD	Plant homeodomain
PLL	Poly-L-lysine
PMSF	Phenylmethylsulfonyl fluoride
PPG	Photoremovable protecting group
PPI	Protein-protein interactions
Pre-mRNA	Precursor mRNA

Abbreviation	Meaning
PTM	Posttranslational modifications
PVDF	Polyvinylidene difluoride
Pyl	Pyrrolysine
PYLIS	Pyrrolysine insertion sequence
RFD	Replication foci targeting domain
RHD	Rel-homology domain
RNA	Ribonucleic acid
rpm	Revolutions per minute
RRBS	Reduced representation bisulfite sequencing
rRNA	Ribosomal RNA
RT	Room temperature
rv	Reverse
RVD	Repeat variable diresidue
s	Second(s)
SAH	S-Adenosyl-L-homocysteine
SAM	S-Adenosyl methionine
SatII	Satellite II
SatIII	Satellite III
scFv	Single-chain variable fragment
scRRBS	Single cell RRBS
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SeC	Selenocysteine
SECIS	Selenocysteine insertion sequence
SINE	Short interspersed nuclear element
snRNA	Small nuclear RNA
SOC	Super optimal broth with catabolite repression
Spec	Spectinomycin
T	Thymine
T3S	Type III secretion system
TALE	Transcription activator like effector
TC-plate	Tissue culture plate
TDG	Thymine-DNA-glycosylase
TEMED	Tetramethylethylenediamine
TET	Ten-eleven translocation dioxygenases
THF	Tetrahydrofuran
TRD	Target recognition domain
TRF1	Telomere repeat binding factor 1

Abbreviation	Meaning
TRIS	Tris(hydroxymethyl)-aminomethan
tRNA	Transfer RNA
TSS	Transcription start site
U	Units
U	Uracil
UAA	Unnatural amino acid
UHRF1	Ubiquitin-like containing PHD and RING finger domains 1
w/o	Without
WGA	Whole genome amplifications
WT	Wild type
ZFP	Zinc finger proteins
α-KG	α-Ketoglutarate

Additionally, all common 1- and 3-letter codes for amino acids were used.

List of Publications

J. Wolffgramm, B. Buchmuller, S. Palei, Á. Muñoz-López, J. Kanne, P. Janning, M. R. Schweiger, D. Summerer. *Angew. Chem. Int. Ed*. **2021**, 60, 24, 13507–13512. doi: 10.1002/anie.202103945.

J. Kanne, M. Hussong, J. Isensee, Á. Muñoz-López, J. Wolffgramm, F. Heß, C. Grimm, S. Bessonov, L. Meder, J. Wang, H. C. Reinhardt, M. Odenthal, T. Hucho, R. Büttner, D. Summerer, M. R. Schweiger. *Cell Death Dis*. **2021**, 12, 6, 530. doi: 10.1038/s41419-021-03810-9.

Á. Muñoz-López, A. Jung, B. Buchmuller, J. Wolffgramm, S. Maurer, A. Witte, D. Summerer, *ChemBioChem* **2021**, 22, 4, 645–651. doi: 10.1002/cbic.202000563.

Á. Muñoz-López, B. Buchmuller, J. Wolffgramm, A. Jung, M. Hussong, J. Kanne, M. R. Schweiger, D. Summerer, *Angew. Chem. Int. Ed*. **2020**, 59, 23, 8927–8931. doi: 10.1002/anie.202001935.

S. Palei, B. Buchmuller, J. Wolffgramm, Á. Muñoz-Lopez, S. Jung, P. Czodrowski, D. Summerer. *J. Am. Chem. Soc*. **2020**, 142, 16, 7289–7294. doi: 10.1021/jacs.0c01193.

1. Abstract

The most prominent DNA modification in mammalian cells is the 5-methylation of cytosine (5mC) in a CpG context and several severe diseases like cancer are connected to aberrant cytosine methylation. 5mC is a dynamic key epigenetic modification regulating chromatin states and therefore gene transcription, and it is highly important to get new insights into the regulation and effects of such epigenetic mark. It is necessary to be able to write 5mC at user-defined genomic loci at specific time points to analyze locus specific and time-resolved downstream effects. 5mC is written by DNA methyltransferases (DNMTs) and several approaches were made to control 5mC levels either on a global or targeted gene level. However, these approaches had low spatial and/or temporal control and resulted in off-target effects. For now, it was not possible to control the catalytic activity of DNMTs itself. This work reports the direct light-control of in vivo DNMT activity to overcome drawbacks of previous approaches. A photocaged cysteine is genetically encoded on DNMT3a to replace the Cys710 which is essential for the catalytic activity. This leads to the expression of DNMT3a proteins in an inactive state and simultaneous activation of them is achieved by light irradiation. This results in the cleavage of the caging group and the remaining of an unmodified and active DNMT3a. Since upstream processes like transfection and protein expression are uncoupled from the actual catalytic activity after light-activation, the kinetics of 5mC writing alone can be monitored. This tool is used to study the effects of several DNMT3a mutations connected to acute myeloid leukemia on the catalytic activity, providing new in vivo observations to dissect the role of these mutations. In addition, it is a great advantage to be able to write 5mC at user-defined genomic loci to specifically alter the chromatin state or to monitor protein interactions associated with 5mC at such loci. Here, locus-specificity is reached by fusing a recombinant DNMT protein with a programmable transcription activator like effector protein. Thereby, inactive DNMTs are recruited to the target locus and activated at a given time point which prevents off-target methylation. It is noteworthy that the final 5mC level is tunable by adjusting the light irradiation time. Also, time-resolved effects of DNA methylation on the transcriptome are reported, providing data on the fast consequence of DNA methylation on gene expression. In conclusion, the ability to control the activity of DNMTs with least structural change by just incorporating a single non-canonical amino acid and the rapid activation of such by light gives previous unreachable spatio-temporal control over DNA methylation in living cells. With this tool, the kinetics of downstream effects of 5mC can be monitored and new insights into the 5mC-related epigenetic network can be obtained.

2. Zusammenfassung

Die wichtigste DNA-Modifikation in Säugetierzellen ist die 5-Methylierung von Cytosin (5mC) in einem CpG-Kontext, und mehrere schwere Krankheiten wie Krebs stehen im Zusammenhang mit einer anomalen Cytosin-Methylierung. 5mC ist eine dynamische epigenetische Schlüsselmodifikation, die den Chromatinzustand und damit die Gentranskription reguliert, und es ist wichtig, neue Erkenntnisse über die Regulierung und die Auswirkungen dieser epigenetischen Modifikation zu gewinnen. Es ist notwendig 5mC an benutzerdefinierten genomischen Loci zu bestimmten Zeitpunkten zu schreiben, um lokusspezifische und zeitaufgelöste Effekte analysieren zu können. 5mC wird durch DNA-Methyltransferasen (DNMTs) geschrieben, und es wurden mehrere Ansätze entwickelt, das 5mC Level entweder auf globaler oder gezielt auf Gen-Ebene zu kontrollieren. Diese Ansätze hatten jedoch nur eine geringe räumliche und/oder zeitliche Kontrolle und führten zu Off-Target-Effekten. Bislang war es nicht möglich, die katalytische Aktivität selbst von DNMTs zu kontrollieren. In dieser Arbeit wird über die direkte Lichtkontrolle der DNMT-Aktivität in vivo berichtet, um die Nachteile früherer Ansätze zu überwinden. Dazu wird das für die katalytische Aktivität essenzielle Cys710 mit einem Cystein ersetzt, welches zusätzlich eine lichtlabile Schutzgruppe besitzt. Dies führt zu Expression von transient inaktiven DNMT3a-Proteinen, wobei die gleichzeitige Aktivierung durch Lichtbestrahlung erreicht wird, die zur Abspaltung der lichtlabilen Gruppe und dem Verbleib von unmodifizierten und aktiven DNMT3a-Proteinen führt. Da vorgelagerte Prozesse wie Transfektion und Proteinexpression von der eigentlichen katalytischen Aktivität nach der Lichtaktivierung entkoppelt sind, kann allein die Kinetik des 5mC-Schreibens verfolgt werden. Mit dieser Methode werden die Auswirkungen verschiedener DNMT3a-Mutationen, die mit akuter myeloischer Leukämie in Verbindung gebracht werden, auf die katalytische Aktivität untersucht, was neue In-vivo-Beobachtungen zur Entschlüsselung der Rolle dieser Mutationen ermöglicht. Darüber hinaus ist es von großem Vorteil, 5mC an benutzerdefinierten genomischen Loci schreiben zu können, um den Chromatinzustand spezifisch zu verändern oder Proteininteraktionen im Zusammenhang mit 5mC an solchen Loci zu beobachten. Hier wird Locus-Spezifität erreicht, indem ein rekombinantes DNMT-Protein mit einem programmierbaren Transkriptionsaktivator-ähnlichen Effektorprotein fusioniert wird. Dadurch werden inaktive DNMTs an den Ziellocus rekrutiert und zu einem bestimmten Zeitpunkt aktiviert, was eine Off-Target-Methylierung verhindert. Dabei ist zu erwähnen, dass das finale 5mC Level durch Anpassung der Lichtbestrahlungszeit einstellbar ist. Außerdem werden zeitaufgelöste Auswirkungen der DNA-Methylierung auf das Transkriptom berichtet, wodurch Daten über die schnellen Auswirkungen der DNA-

Methylierung auf die Genexpression geliefert werden. Zusammenfassend lässt sich sagen, dass die Fähigkeit, die Aktivität von DNMTs mit minimalen strukturellen Veränderungen zu kontrollieren, indem nur eine einzige nicht-kanonische Aminosäure eingebaut wird, und die schnelle Aktivierung dieser Aminosäure durch Licht eine bisher unerreichte räumliche und zeitliche Kontrolle der DNA-Methylierung in lebenden Zellen ermöglicht. Mit dieser Methode kann die Kinetik von Effekten, welche durch 5mC ausgelöst werden, untersucht und neue Erkenntnisse über das 5mC-bezogene epigenetische Netzwerk gewonnen werden.

3. Introduction

3.1. DNA

All information to build up an organism from the molecular level over the cell level, up to the organ and whole organismal level must be stored in a heritable way. Nature fulfills this task by storing this information in the sequence of the linear polymer deoxyribonucleic acid (DNA) in each cell of a living organism.[1,2] The cells total DNA constitutes the so called 'genome' and can be replicated and inherited to new cells.[1,2] The DNA monomers are called nucleotides (Figure 1B) and consist of three parts: the nucleobase (purine based adenine (A) and guanine (G), or pyrimidine based thymine (T) and cytosine (C); Figure 1A), five-carbon sugar deoxyribose, and 5'-phosphate group. Deoxyribose with an attached nucleobase forms a nucleoside and after the addition of one to three phosphate groups it becomes a nucleotide.[1–3] Deoxynucleoside triphosphates (dNTPs) can be connected through a phosphodiester bond to the free 3'-hydroxy group after the elimination of pyrophosphate (Figure 1C).[1,2] Independent of the nucleobases, the deoxyriboses connected by phosphodiester bonds build the 'sugar phosphate backbone' of DNA with a free phosphate group at its 5'-end and a free hydroxy group at its 3'-end.

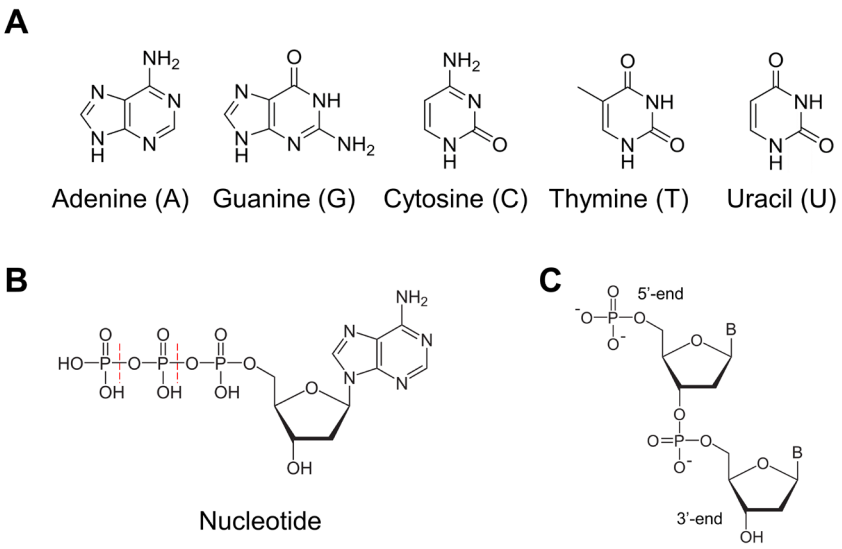


Figure 1: Chemical structures of DNA nucleobases and nucleotide. A: Chemical structures of the four DNA nucleobases adenine, guanine, cytosine, and thymine. Additionally, the structure of uracil is shown which is only found in RNA. **B:** Chemical structure of deoxyadenosine triphosphate with adenine as nucleobase. The nucleotide does also exist as di- or monophosphate (indicated by red dashed lines). **C:** Basic chemical structure of single stranded DNA. Single monophosphate nucleotides (see B) are connected via phosphodiester bonds building the sugar phosphate backbone with free 5' and 3' ends. The structure of nucleobases is not shown but shortened as 'B'.

Specific base pairing of adenine with thymine and guanine with cytosine known as Watson-Crick base pairing was discovered (Figure 2).[4] A·T pairing consists of two hydrogen bonds whereas the stronger G·C pairing forms three hydrogen bonds.[4] Due to this, two complementary DNA strands can hybridize in antiparallel orientation (5'-end of one strand is opposite to the 3'-end of the other strand) whereby the nucleobases are facing inward and the phosphate backbone faces outward of the formed double stranded DNA.[1,2,4] The DNA is not planar but forms a double helix which exists in three biologically active structures called A-DNA[5], B-DNA[4] and Z-DNA[6].

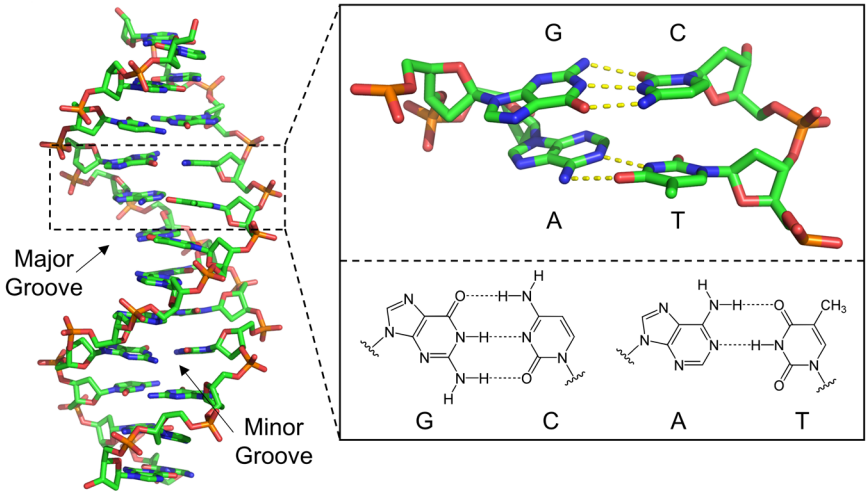


Figure 2: Structure of B-DNA. Crystal structure of B-DNA (PDB 1BNA[7]). Highlighted are the two Watson-Crick base pairs G·C and A·T with their three and two hydrogen bonds, respectively, shown in yellow.

The most common DNA structure in nature is the right-handed helical B-DNA (Figure 2).[1] As in all three mentioned DNA structures, the helix is mainly stabilized by Watson-Crick hydrogen bonds and π-π interactions of the aromatic rings on purines and pyrimidines (see close up of Figure 2).[8] In B-DNA, 10 – 10.5 bp are needed for a full helical turn which has a pitch (height of such helical turn) of ~34 Å with a diameter of 20 Å.[9] Since both strands are not exactly opposite to each other, the B-DNA helix shows a wider major groove and narrower minor groove (Figure 2).[10] While the minor groove is only 6 Å wide and 7.5 Å deep, the major groove is about 12 Å wide and 8.5 Å deep and represents a good protein interaction site since the diameter of protein α-helices with side chains is ~12 Å and can therefore fit into the major groove.[2,11] Although the Watson-Crick hydrogen bonds are no accessible information for other proteins in the minor or major groove, other parts of the nucleobases are and they can serve as important binding partners in crucial protein-DNA interactions.[1,2]

3.1.1. DNA Organization in Cells

The genome of prokaryotes is mainly organized in a single circular chromosome[12] with some exceptions[13], whereas eukaryotic DNA is linear, highly organized, and localized in the cell nucleus[14]. Histone proteins constitute the basic structure of such organization and represent half the mass of chromosomes in eukaryotic cells.[2] There are five histone proteins (H2A, H2B, H3, H4, and H1, with the exception in spermatozoa which use protamines[1,15]) of which all but H1 build a histone octamer (Figure 3A).[2] For this, histones H2A and H2B as well as H3 and H4 form heterodimers, while two H3-H4 dimers further combine to a tetramer. Then, two H2A-H2B dimers attach to the H3-H4 tetramer with their docking sequence to complete the histone octamer (Figure 36, page 112).[16,17] DNA fragments of 147 bp wrap around such histone octamers forming nucleosomes with ~100 Å diameter (Figure 3A and B) which are connected by accessible linker DNA of 157–240 bp length (Figure 3B, 'beads on a string').[2,16,18] The tails of histone proteins extend outwards from the core structure, they are enriched with lysine and arginine residues, and are common sites for protein modifications (see also 3.2.1).[17,19] Histone H1 binds to the exterior of nucleosomes where the linker DNA enters and leaves the nucleosome, and supports the formation of higher-order chromatin structures together with other proteins[20] (see 30-nm chromatin fiber in Figure 3B).[1,2,21] This DNA organization is extremely conserved between species with only two mutations in histone proteins between mammals and plants.[1] The entire nucleoprotein complex called chromatin (Figure 3B) can be divided into sparsely packed euchromatin and densely packed heterochromatin based on its accessibility.[17,19] Chromatin has its most condensed state during mitosis while forming chromosomes (Figure 3B), where it can be compacted by a factor of over 10000.[1] This high compaction allows the ~2 m DNA per human cell (all 46 chromosomes in a row) to fit into the nucleus.[1]

The protein-coding parts makes up almost the entire genome in most prokaryotic cells but only represents ~1.5 % of the human genome (Figure 3C).[22,23] Roughly more than a quarter of the genome are introns which are located within a gene and are transcribed but not translated into the final protein.[22] Other parts of the genome encode functional RNAs like structural RNAs (e.g. transfer RNA (tRNA), ribosomal RNA (rRNA) and small nuclear RNA (snRNA)) and regulatory RNAs (e.g. microRNA (miRNA) and long non-coding RNA (lncRNA)).[24] Interestingly, more than half of the human genome consists of repetitive elements.[22] The vast majority of the repetitive elements are transposable elements which can change their position in the genome and can interact with the genome in complex ways.[25] They include the short interspersed nuclear element (SINE) *Alu* which has

>1,000,000 copies and long interspersed nuclear element (LINE) *LINE1* which has >50,000 copies, which account for 11 % and 17 % of the human genome, respectively.[22,26]

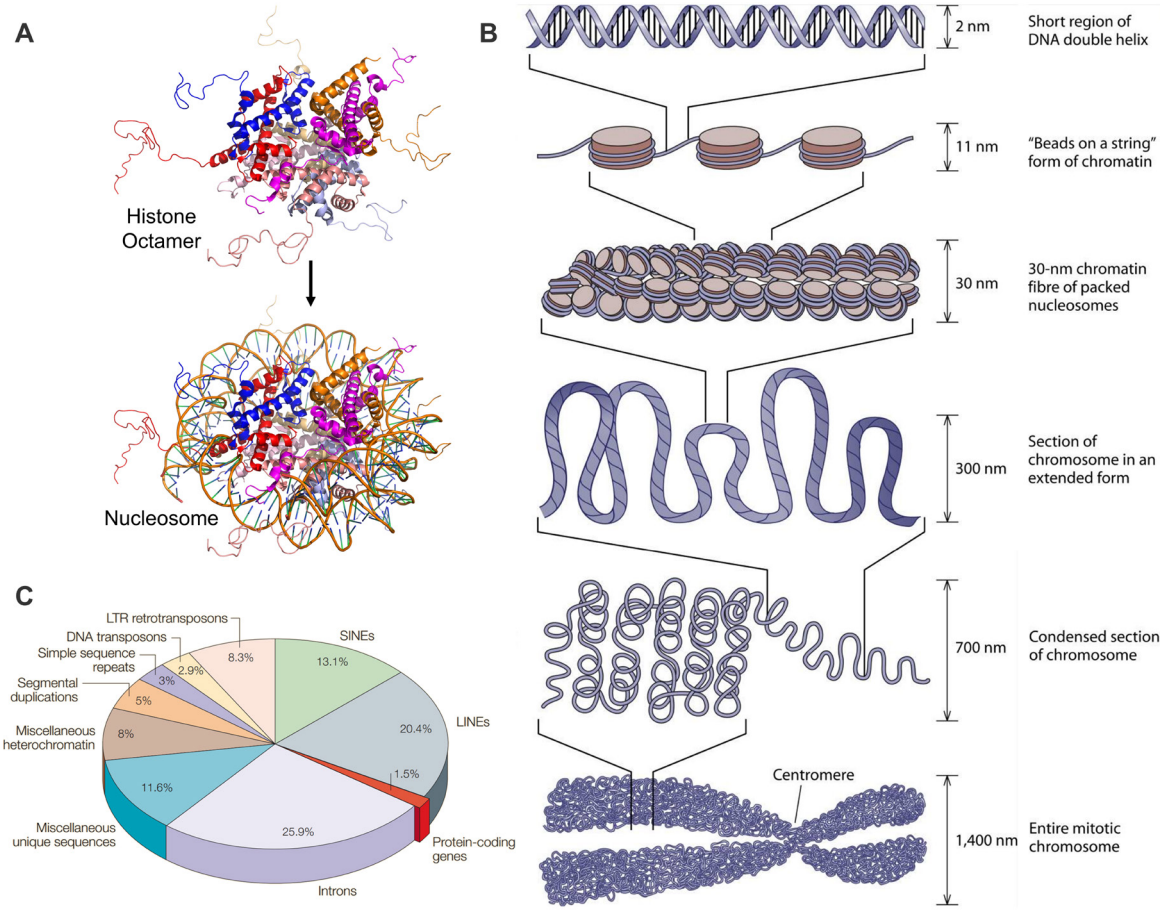


Figure 3: Composition and structure of genomic DNA. A: Crystal structure of eight histone proteins building a histone octamer or a nucleosome with additional DNA (histone color code: H3 (red and salmon), H4 (blue and light blue), H2A (magenta and light pink), H2B (orange and light orange). Adapted from PDB 1KX5[16,27]. **B:** Chromatin structure in eukaryotes. DNA is wrapped around histone octamers forming a nucleosome which further fold into chromatin fibers and other higher-order structures.[28] Figure was adapted from[28]. **C:** Components of the human genome.[22,26] Only about 1.5 % of the genome consists of protein-coding sequences. Figure was adapted from[22].

Another type of repetitive elements are tandem repeats which include microsatellites, minisatellites, and satellites.[29] Tandem repeats are short sequences which are highly repetitive in a head to tail fashion.[29] Satellite DNA repeats are found as heterochromatin at centromeres and telomeres.[30] They contribute to chromosome structure with several thousands of highly compact short repeats.[29,31] They can also bind chromatin regulatory proteins and can cause alterations to the regulatory process which leads to severe diseases.[32] At centromeres, satellite DNA plays an important role in sister chromatids cohesion and subsequent chromosome segregation during mitosis.[29,33] The alpha-satellite sequence with 170 bp units is the most abundant unit in human centromeric and pericentromeric heterochromatin but also the pentameric repeats satellite II (SatII) and satellite III (SatIII) occur frequently[29,34], which can even be transcribed as a part of the cells

stress response. SatIII transcription is a general stress response upon heat-shock, heavy metal treatment, UV-C irradiation, oxidative and hyper-osmotic stress.[35] After applying stress in form of heat to human cells, transcription of SatIII lncRNA is getting activated.[36] These RNAs are quite stable and are associated with the chromosome 9q12 region even during mitosis.[37] The transcription is dependent on heat shock factor 1 (HSF1) and a rapid relocalization of this protein to the SatIII regions was observed after heat-shock.[35,38,39] SatIII elements contain sequence motifs similar to the HSF1 binding site[40] and the binding of HSF1 together with the transcribed RNA forms so called nuclear stress bodies (nSBs).[40,41] nSBs are thought to play a role in rapid and transient global changes of gene expression and therefore aberrant control of SatIII transcription is connected to cancer.[38,42,43]

3.1.2. Transcription

Information stored in the DNA sequence of genes can be converted into functional biomolecules. DNA can be transcribed into RNA whereby these nucleic acids serve either as intermediate (e.g. messenger RNA (mRNA) for protein biosynthesis) or as the final biomolecule (e.g. rRNA and tRNA).[1] The main differences between DNA and RNA include the usage of ribose instead of deoxyribose as five-carbon sugar, the replacement of thymine by uracil (U) as nucleobase (Figure 1), and the fact that RNA is single stranded and much more unstable than DNA.[1,2,44] For protein biosynthesis, mRNA in eukaryotes is exported from the nucleus to the cytoplasm where the final biosynthesis takes place (Figure 4A).[1,44] mRNA also serves as an intermediate in prokaryotes despite that they do not have a nucleus.[45] In both, prokaryotes and eukaryotes, mRNA is degraded after protein biosynthesis.[46] Transcription of DNA is performed by RNA polymerases whereby only one type exists in prokaryotes and three main types in eukaryotes (RNA polymerase I, II and III).[47] The transcription process includes three steps (1st: initiation, 2nd: elongation, and 3rd: termination) and requires a promoter located upstream of the transcription start site (TSS) of a gene to serve as the binding site for transcription initiator proteins. In the initiation step, the RNA polymerase in bacteria recognizes and binds to specific DNA sequences in the promoter, while eukaryotes have additional transcription factors for the recognition of promoters leading to the binding of the RNA polymerase II needed for mRNA transcription.[1,48] The promoter sequence has a direct influence on the binding strength of the interacting proteins and therefore different sequences lead to different transcription levels.[49] In eukaryotes, the TATA-box is crucial and serves as the starting point of the assembly of different transcription factors.[50] The RNA polymerase transcribes the gene sequence from the TSS on with 5’ to 3’ direction.[1,2] The elongation in prokaryotes and eukaryotes is terminated in different ways. In bacteria, either specific G-C rich elements

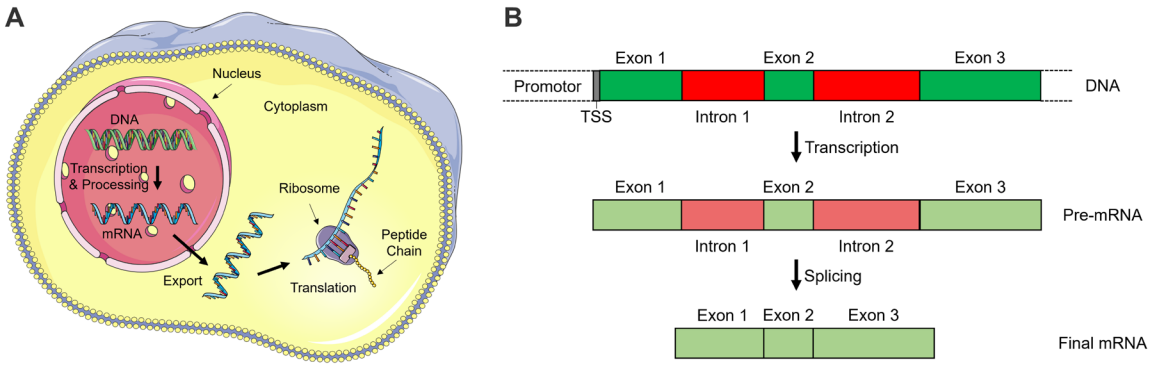


Figure 4: Schematic overview of transcription/translation and splicing process.
A: Transcription of DNA into mRNA takes place in the cell nucleus and the mRNA is subsequently exported into the cytoplasm to be translated into a protein ('peptide chain'). Single structures are not to scale. Figure is self-made with structures taken from[51]. **B:** Overview of the splicing process. The DNA sequence containing introns and exons is fully transcribed into pre-mRNA. Then, introns are spliced out to get the mature mRNA with exons only. mRNA modifications like 5'-cap or poly(A)-tail are not shown. Figure is based on[1].

with a poly-uracil sequence lead to a destabilization of the transcription complex or the termination factor Rho triggers the dissociation of the RNA polymerase.[52,53] However, the termination process in eukaryotes is less understood and RNA polymerase-dependent.[52] RNA polymerase II termination in eukaryotes is more complex and can happen through different pathways. One of the most-studied termination pathways is initiated by the mRNA cleavage at specific sites, followed by rapid 5'-3' degradation by exonucleases which finally clash with the polymerase resulting in its release.[52,54,55] Alternatively, a poly(A)-signal could lead to structural changes[52,56] or termination factors including RNA/DNA helicases are needed for the termination[52,55,57]. The mRNA after termination is directly ready for translation in prokaryotes, whereas it must be processed and modified in eukaryotes and is therefore called the precursor mRNA (pre-mRNA).[1] The 5'-end of the mRNA is modified with a 7-methylguanylate cap co-transcriptionally to protect it from degradation by exonucleases.[58] The 3'-end is cleaved off and a poly(A)-tail is attached which plays an important role for mRNA stability and protein translation.[59] In contrast to prokaryotes, the gene-coding sequence of eukaryotes consists of introns and exons whereby only exons are needed for translation instead of the full mRNA sequence. Therefore, introns must be spliced out of the pre-mRNA sequence (Figure 4B) and this process is done by the large spliceosome complex which consist of snRNA and various proteins.[60] These processing steps result in the mature mRNA which is exported from the nucleus into the cytoplasm where it can be translated into a protein (Figure 4A; for more information see 3.4.1).[1,2]

3.2. Dynamics of DNA Methylation

3.2.1. Writing and Erasing of DNA Methylation

As described before, every cell of an organism uses DNA with almost the same sequence as storage of genetic information and yet many different cell types exist within the same organism.[1,17] It is important that only specific genes defining the cells' identity are expressed.[17] The field of epigenetics contains mechanisms which can control gene expression without altering the DNA sequence itself.[17,19] Epigenetic modifications can change the chromatin state due to mainly DNA and histone modifications which lead to either loosely packed and active euchromatin or tightly packed and inactive heterochromatin.[17,19] Such epigenetic patterns can change over time and are also dependent on environmental changes like nutrition, chemical compounds, temperature, or stress.[61] The most abundant epigenetic DNA modification in mammalian cells is the 5-methylation of cytosine (Figure 5; further called 5mC or 'DNA methylation') which is mainly found in CpG dinucleotides.[62,63] 70 – 80 % of all CpGs in the genome are methylated whereby most unmethylated CpGs are found in GC-rich regions upstream of a gene (CpG islands).[64] Most genes in vertebrates are associated with CpG islands and they play specific roles in gene regulation.[65] 5mC is commonly known as a repressive mark and results in non-binding of transcription factors to the promoter region.[63] Other proteins however can specifically bind to methylated DNA and are part of a second mode of repression (more details later).[63,66] Due to its repressive nature, DNA methylation was associated with X-chromosome inactivation in female cells[67], suppression of transposable elements[68], and genomic imprinting[69]. Genomic imprinting describes the epigenetic mechanism that specific genes are only expressed from one of two parental alleles in cells and it is essential for normal growth.[69] In humans, more than 200 imprinted genes were discovered.[70] DNA methylation also plays a crucial role in chromatin stability and is altered in severe diseases like cancer.[71] While DNA is usually hypomethylated in cancer, single promoter regions of tumor-suppressor genes show hypermethylation.[72] 5-methylation of cytosines is catalyzed by DNA methyltransferases (DNMTs) which either maintain 5mC patterns after DNA replication or create them *de novo* (for detailed information see 3.2.2).[73] In the past, 5mC was believed to be a stable epigenetic DNA mark but this changed after the discovery of ten-eleven translocation dioxygenases (TET).[74] Then, passive dilution of 5mC by DNA replication in the absence of maintenance methylation was just one possible way of demethylation.[75] The other and more common one is based on an active demethylation pathway catalyzed by TET enzymes, Fe(II)/α-ketoglutarate (α-KG)-dependent dioxygenases (see Figure 5; Figure 37, page 113, shows the reaction

mechanism of TET enzymes).[75] Special about the DNA interaction is that the modification at the 5-position is not recognized by TET which allows the stepwise oxidation of 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) in the respective order.[76] 5fC and 5caC can be excised by thymine-DNA-glycosylase (TDG) leading to abasic sites which are restored to cytosines by base excision repair (BER).[77]

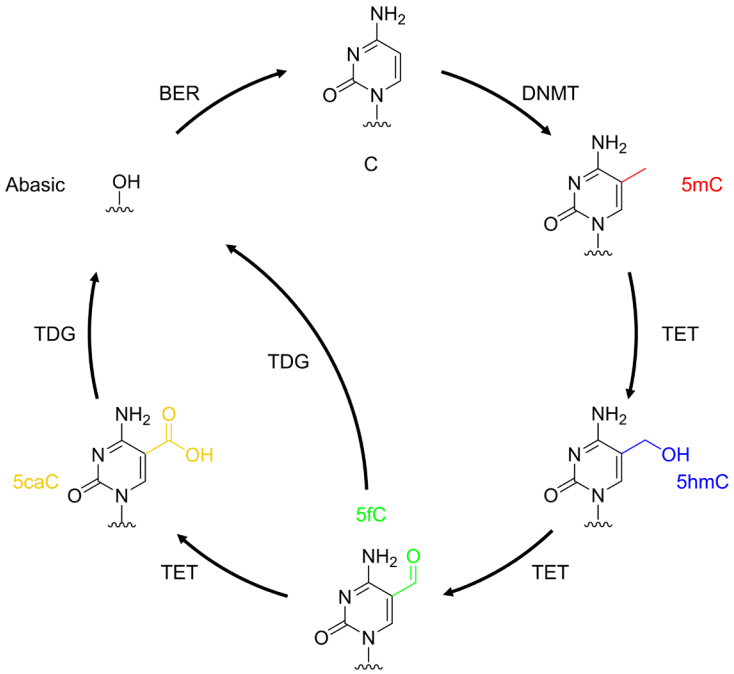


Figure 5: Cycle of active cytosine methylation and demethlytion. C: cytosine; 5mC: 5-methylcytosine; 5hmC: 5-hydroxymethylcytosine; 5fC: 5-formylcytosine; 5caC: 5-carboxylcytosine; DNMT: DNA methyltransferase; TET: ten-eleven translocation dioxygenase; TDG: thymine-DNA glycosylase; BER: base excision repair. Figure based on[78].

However, the different oxidized modifications are highly disproportional in the genome. 5hmC is the most abundant oxidized modification and some biological functions were also connected to it.[75] Its presence is cell-type dependent with highest amounts detectable in brain tissue.[79,80] For example, 0.3 – 0.7 % hydroxymethylated CpGs were detected in cells of the central nervous system, 0.15 – 0.17 % in kidney, bladder, heart and lung, and 0.03 – 0.06 % in liver and spleen.[79] High 5hmC levels are also found in embryonic tissues.[77] The higher oxidized modifications are even less abundant. In mouse embryonic stem cells only 1 – 10 % of total 5hmC is oxidized to 5fC and 5caC.[81] This distribution might be based on different reactivities of TET enzymes. Although they accept 5mC, 5hmC and 5fC as substrate, they are most active on 5mC.[82] Additionally, 5fC and 5caC are subject for TDG/BER-mediated active demethylation which leads to decreased levels.[83,84] Oxidized cytosine modifications were connected to different biological functions showing that they might not only be intermediates of active demethylation.[84,85] 5hmC can be recognized by other proteins and its presence in gene bodies showed a positive correlation to gene

expression.[84] The enrichment of 5hmC at exon-intron boundaries and enhancer also suggests a role in splicing and gene regulation.[84] Recruitment of specific proteins including transcription regulators and chromatin remodelers to 5fC containing sites was detected which suggest a direct role of 5fC in gene regulation.[86] 5fC and 5caC also have direct negative impact on the activity of RNA polymerase II.[87] It was also suggested that the transient pausing during transcription might lead to the recruitment of transcription factors or chromatin remodeling complexes.[87]

Chromatin structure has direct influence on gene expression and is not only controlled by DNA modifications but also histone modifications with both having an influence on each other.[88] Tails of histone proteins (see 3.1.1) are mainly subject for post-translational methylation and acetylation.[88,89] Histone methylation is introduced by histone methyltransferase (HMT) usually affecting arginine and lysine residues of histone H3 and H4.[88,90,91] They can be mono-, di-, and trimethylated and result in either active (H3R17, H3K4, H3K36, and H3K79) or repressive epigenetic marks (H3R8, H3K9, H3K27, and H4K20) whereby not all mechanisms are understood.[88,91,92] Like DNA methylation, also histone methylation is not irreversible and can be erased by histone demethylases.[88] These histone marks can recruit different reader proteins. For example, heterochromatin protein HP1 mediates gene silencing by binding trimethylated lysine 9 of histone 3 (H3K9me3) and subsequent recruitment of histone methylases SUV39H1 and SUV39H2.[88,93] It was shown that H3K36me2 can recruit DNMT3a to non-coding regions of euchromatin.[94] Histone acetylation of lysine residues reduces their positive charge and with this the interaction to negatively charged DNA resulting in more open and active DNA.[88,95] The repressive DNA mark 5mC is known to trigger histone deacetylation.[96] 5mC can be recognized by methyl-binding domains (MBDs) which are either part of histone deacetylases (HDAC) or proteins which can recruit HDACs.[96,97] Moreover, recruitment of HDACs through direct binding by DNMT1, DNMT3a and DNMT3b was demonstrated.[98] Histone acetyltransferases (HATs) can act on several histone lysine residues including H3K4, H3K9, H3K27, H3K36, H3K79, H4K5, H4K12, H4K20.[88] Active promoters and enhancers with H3K27ac usually also have the active mark H3K4me3 showing that different interplays between histone modifications exist.[99]

3.2.2. DNA Methyltransferases

DNA methylation in mammals is dependent on the three DNA methyltransferases DNMT1, DNMT3a and DNMT3b, whereby DNMT1 is important for maintenance methylation, and DNMT3a and DNMT3b for *de novo* methylation.[73] Although they are active in different contexts, they still use the same reaction mechanism to 5-methylate cytosine (Figure 6).[17,100] DNMTs use a base flipping mechanism to rotate the target cytosine out of the DNA

helix to bring it into close proximity to crucial amino acids and the cofactor in the catalytic pocket (Figure 15B, page 35).[101,102]

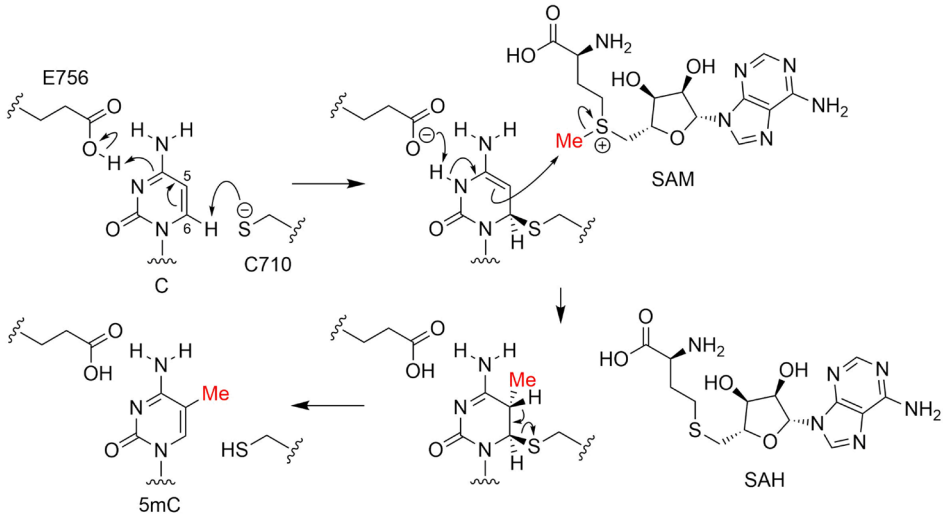


Figure 6: Chemical mechanism of cytosine 5-methylation by DNMT. The water molecule for the final β-elimination step is not shown. SAM: S-adenosyl methionine; SAH: S-adenosyl-L-homocysteine. Numbering of C710 and E756 according to human DNMT3a. Modified from[103], based on[100].

In a first step, the catalytically active cysteine (C710 in human DNMT3a) is deprotonated to a thiolate anion by a DNA phosphate group via a bridging water molecule and becomes a strong nucleophile.[100] It forms a covalent bond with the C6 atom of cytosine by a 1,4-addition reaction and the negative charge is simultaneously stabilized through the protonation of N3 by E756.[104] Next, N3 is deprotonated again which leaves more negative charge on the cytosine ring and due to electron delocalization it can make C5 of cytosine nucleophilic enough to target the methyl group (shown in red in Figure 6) of cofactor S-adenosyl methionine (SAM), leaving S-adenosyl-L-homocysteine (SAH).[100,105] A final β-elimination supported by a water molecule results in 5mC and the releasing of DNMT.[100,104] SAH is hydrolyzed to adenosine and homocysteine, whereby homocysteine is recycled by methylation to methionine and subsequent reaction with ATP to form SAM.[106] The shown methylation reaction mechanism is conserved throughout many DNA methyltransferases.[107] Ten motifs needed for this reaction are conserved between eukaryotic and procaryotic cytosine 5-methyl transferases.[108] Motifs I, IV, VI, VIII, IX, and X are highly conserved between mammalian DNMTs.[107] Motif I and X are important for binding the cofactor SAM.[109] The catalytically active cysteine with an upstream proline forms motif IV and with it the catalytic loop.[107,109] The glutamyl residue responsible for protonation of the N3 position is part of the conserved ENV sequence of motif VI and it also stabilizes the flipped base.[107,109,110] Motif VIII with two arginine residues seems to have a critical role in protonation of N3 and stabilization of the reactions intermediate.[107,110] The

variable region target recognition domain (TRD; located between motif VIII and IX) together with motif IX is responsible for target recognition and binding via specific contacts in the DNA major groove.[110,111]

Several other cytosine 5-methyl transferases were found in other organisms but some organisms like *Caenorhabditis elegans* and *Saccharomyces cerevisiae* lack these enzymes and 5mC completely.[112] Although DNMT2 shows sequence and structural similarities compared to DNMT1 and DNMT3, its activity has no impact on DNA methylation.[109,113] DNMT2 however showed methylation activity of small RNA molecules in different organisms.[113]

DNMT1 has a strong preference for hemimethylated DNA with a strict requirement of CpG sites and is therefore responsible for maintaining DNA methylation patterns after DNA replication when the newly synthesized DNA strand is not methylated.[109,114] This is highly important for inheritance of epigenetic DNA methylation and maintaining genomic integrity.[115] It is mainly found at replication foci during the S phase in proliferating cells.[107] With 1616 amino acid, DNMT1 is much larger than members from the DNMT3 family.[107] It is composed of two parts connected by a glycine-lysine linker – a longer regulatory N-terminal part with several functional modules, and a C-terminal part containing the catalytic domain (Figure 7).[107,109]

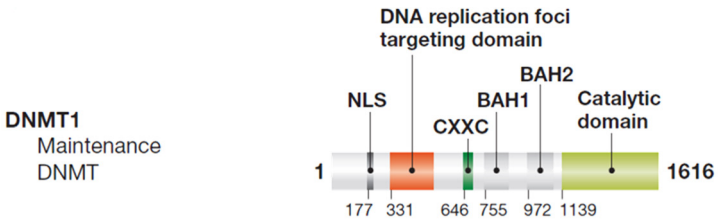


Figure 7: Schematic overview of DNMT1. Figure adapted from[116].

One of two modules responsible for cell localization is the nuclear localization signal (NLS) which is necessary for nuclear import of expressed DNMT1.[107] The other one is the replication foci targeting domain (RFD) to localize DNMT1 towards replication foci during S phase.[107] However, recruitment of DNMT1 to hemimethylated CpG sites is dependent on interactions with ubiquitin-like containing PHD and RING finger domains 1 (UHRF1) which can bind such sites.[117] On top, UHRF1 is responsible for flipping the cytosine of the targeted CpG out of the DNA helix.[118] Recent studies showed that the base flipping mechanism and subsequent DNMT1 mediated DNA methylation shows dependence on flanking sequences which can change the conformational dynamics of the active site and cofactor binding pocket.[101] To prevent *de novo* methylation of unmethylated CpGs after DNA replication, DNMT1 contains a cysteine-rich CXXC domain which binds unmethylated CpGs.[119] CXXC and BAH1 are separated by the CXXC-BAH1 autoinhibitory linker with a

highly acidic segment. After binding of unmethylated DNA by CXXC, the linker is relocated between DNA and the active site ensuring the exclusion of unmethylated CpGs by preventing DNA from entering the catalytic pocket.[119] A second autoinhibitory mechanism comes from the BAH2-TRD loop which positions TRD in a retracted way which blocks interaction between TRD and DNA.[119] Both bromo-adjacent homology (BAH) domains are connected via an α-helical linker and physically associated with the catalytic domain.[119] They present a large surface for interactions with other proteins connected to DNA methylation, replication and transcriptional regulation.[119,120]
The catalytic domain contains the previously mentioned ten conserved motifs and folds into two subdomains (catalytic core and the TRD).[107,119] Interestingly, the C-terminal part of DNMT1 is more related to bacterial restriction methyltransferases than to mammalian DNMT3a family.[109,121] The large cleft between catalytic core and TRD is filled with DNA.[119] The catalytic core structure exhibits the highly conserved motifs I, IV, VI, and VIII.[107] The TRD subdomain consists of motifs IX and X, and the region between motif VIII and IX. As described, TRD is responsible for targeting the DNA.[107]

The DNMT3 family has three members – DNMT3a, DNMT3b and DNMT3L.[107,122,123] While DNMT1 is responsible for maintenance of 5mC patterns, DNMT3a and DNMT3b are establishing them as *de novo* DNA methyltransferase.[102,107] The structures of DNMT3a and DNMT3b are quite similar (Figure 8) but both enzymes show differences in their expression pattern.[124] DNMT3a has two splicing isoforms of which DNMT3a1 is expressed at low levels in all tissues and the shorter DNMT3a2 lacking the N-terminal part mainly in embryonic tissue.[124,125] DNMT3b is highly expressed in embryonic stem cells and strongly downregulated after cell differentiation.[126] For DNMT3b, several isoforms were identified but most of them are catalytically inactive.[124] Although DNMT3b is mainly active in embryonic cells, inactive isoforms (especially DNMT3b3 and DNMT3b4) are expressed in somatic cells to regulate DNMT3a mediated methylation through direct interactions.[124,127] Like DNMT1, members of the DNMT3 family have a regulatory N-terminal region and catalytic C-terminal region whereby DNMT3L is lacking a catalytic domain.[107] However, the N-terminal part of DNMT3a and DNMT3b is not essential for activity of the catalytic domain alone which makes them valuable model systems in various studies.[128–130] The catalytic domains also have the typical ten conserved motifs and TRD mentioned before.[107,131] Some inactive DNMT3b isoforms may lack some of them due to alternative splicing.[107] The conserved motifs fulfill the previous described functions.[102,107] In more detail, binding of the cofactor SAM by DNMT3a and DNMT3b was described to be dependent on motifs I to III, as well as motifs V and X.[107,128]

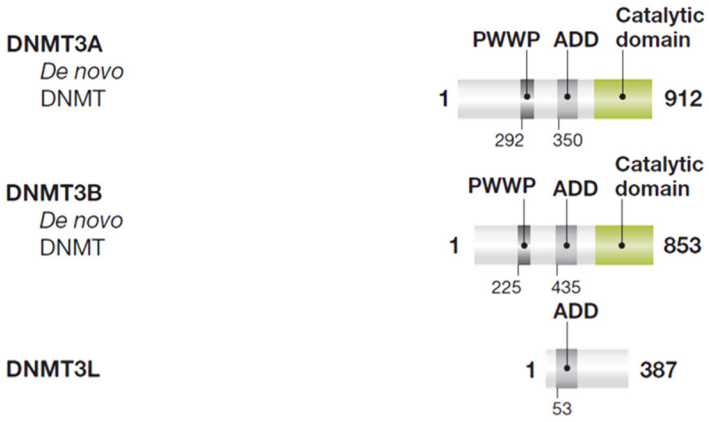


Figure 8: Schematic overview of DNMT3a, DNMT3b and DNMT3L. Figure adapted from[116].

Both active DNMT3s have similar affinity towards unmethylated and hemimethylated DNA and can uphold 5mC to a certain extend in DNMT1 lacking cells.[124,132] Though CpGs are the preferred substrate (DNMT3b has a lower CpG specificity than DNMT3a[131]), also non-CpG methylation was observed.[131,133,134] Mainly it is found in neurons, frontal cortex, and embryonic cells with up to 2 % of all cytosines showing non-CpG methylation.[133] These methylations occur in different sequence contexts with differentiated neurons mainly having CAC methylation and embryonic cells CAG methylation.[133] Since DNMT1 has a clear specificity for CpGs, non-CpG methylation must be generated *de novo* after DNA replication.[124] CAC and CAG preference can be also assigned to DNMT3a and DNMT3b, respectively.[135] This fits the observation that DNMT3a is highly expressed in brain tissue and that DNMT3b is important in early development.[124] CpA methylation was also proved to be important for expression regulation for MeCP2-controlled genes in brain tissue.[136] To a lesser extent, methylation in embryonic stem cells was also found in a CpT and CpC context.[137] Both DNMT3s also showed dependency on flanking sequences, and mutations in such were promoting cancer progression.[101,131] While DNMT1 can recognize its target due to the hemimethylated state, *de novo* methylation must be controlled by other mechanisms. It was shown that specific sequence motifs in the genome can regulate DNMT3 dependent DNA methylation.[138] The regulatory N-terminal part between DNMT3a and DNMT3b is variable but contains the conserved PWWP domain and ATRX-DNMT3-DNMT3L (ADD) domain (Figure 8).[107] The PWWP domain with its proline-tryptophan-tryptophan-proline motif is found in several chromatin-associated proteins and is responsible for binding of pericentric heterochromatin, and specific loci at metaphase.[107,139] Additional, specific interactions of the PWWP domain of DNMT3b with histone methylation (H3K36) in gene bodies was demonstrated.[140] A similar interaction was observed at active enhancer regions in human epidermal stem cells.[141] Yet, it resulted in different downstream effects since DNMT3a promoted TET2 dependent DNA hydroxymethylation, and DNMT3b

DNA methylation.[141] Mutations of the PWWP domain can result in aberrant genomic localization causing dwarfism phenotype.[142] The ADD domain is necessary for chromatin interaction and autoinhibition.[143,144] In the enzymes inactive conformation, the ADD domain showed interactions with the DNA-binding region of the catalytic domain which only changed after the ADD bound to the unmethylated histone H3 (H3K4me0) but not methylated one (H3K4me3).[144] The ADD domain contains a plant homeodomain (PHD)-like sequence which plays a role in protein-protein interactions (PPIs).[107,145] Interactions of ADD domain with transcriptional and epigenetic regulators including Myc, SUV39H1, and HP1 was shown.[107] This indicates that gene repression by DNMT3a and DNMT3b can be independent of DNA methylation.[98] Although UHRF1 does not have a similar regulation on DNMT3a and DNMT3b as it does on DNMT1, recent studies showed a negative dependency and it was also demonstrated that UHRF1 overexpression can lead to decreased *de novo* methylation by degradation of DNMT3a.[146]

The third member of the DNMT3 family, DNMT3L (DNMT3-like), shows clear homology to DNMT3a and DNMT3b but no intrinsic catalytic activity.[122,147] Key catalytic amino acids are mutated and DNMT3L cannot bind the cofactor.[122,147,148] However, it directly interacts with the catalytic domain of DNMT3a and DNMT3b (Figure 9) and stimulates their activity.[147–149] Therefore, it is a critical regulatory factor of DNA methylation, it is mainly expressed in embryonic and germ cells, and it is important for methylating imprinted genes.[107,147,148]

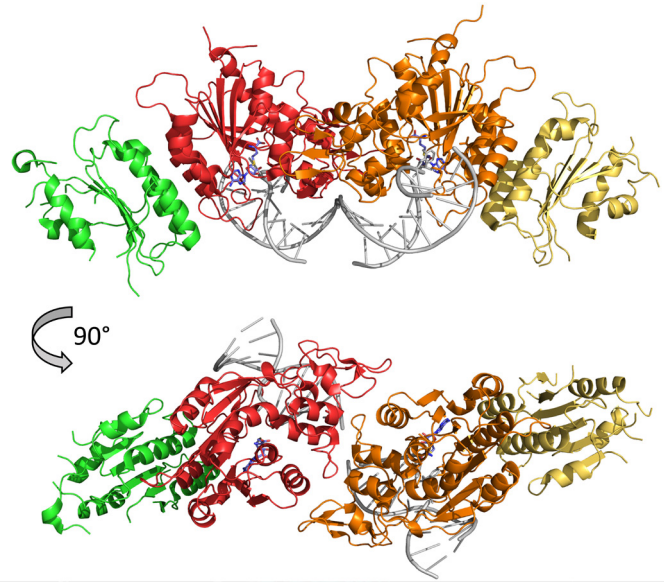


Figure 9: Crystal structure of DNMT3a3L with DNMT3a-DNMT3a homodimeric interface. Catalytic domain of DNMT3a is shown in red and orange; DNMT3L parts are shown in green and yellow. Structure is taken from PDB 6F57[102].

It seems that DNMT3a is dependent on DNMT3L since specific *de novo* methylation of DNMT3a conditional germline knockout animals did not differ from DNMT3L knockout ones.[148,150] In crystal structures with DNMT3a, it stabilizes the active site loop of DNMT3a

with its key catalytic cysteine.[102,148] Two monomers of DNMT3a and two monomers of DNMT3L build a heterotetramer with one DNMT3a-DNMT3a interface and two DNMT3a-DNMT3L interfaces (Figure 9).[102,148] Mutation analysis which led to disruption of the heterotetramer showed that both interfaces (3a-3a and 3a-3L) are important for catalysis.[148] However, it is not essential that two CpG sites are in close proximity as long as at least eight base pairs are on each side of the CpG so that both DNMT3a monomers have DNA contact.[148] Since this distance is required, CpGs with eight to ten base pairs in between show strong methylation correlation since they can be methylated by the same heterotetramer.[148] CpGs within these active sites though cannot be methylated resulting in absence of methylation in a specific pattern.[148]

3.3. Targeted Editing of DNA Methylation

3.3.1. Previous Approaches

Since DNA methylation has great impact on an organism and malfunctions can lead to severe diseases, several efforts were made to study DNA methylation by taking spatial control over it.[151] Such control allows the functional analysis of 5mC since specific genomic loci can be targeted and specifically altered. Mainly, effector proteins are recruited to user defined loci by fusing them to programmable DNA binding domains (DBDs).[151,152] Commonly used DBDs are zinc finger proteins (ZFPs)[153], transcription activator like effector proteins (TALEs; since TALEs were used in this thesis, detailed information can be found under 3.3.2)[154] and the clustered regularly interspaced short palindromic repeats/deactivated CRISPR-associated protein 9 (CRISPR-dCas9) system[155,156]. Protein sequences of ZFPs and TALEs are directly programmable whereas CRISPR-dCas9 needs a guide RNA (gRNA). gDNA recognizes the target sequence 5' to the important protospacer-adjacent motif (PAM) and intercalates into the DNA to mediate binding of dCas9.[156] ZFPs can recognize base triplets and are assembled according to the target sequence.[153] TALEs offer an easier way of assembly according to a one-to-one recognition code (see 3.3.2).[157] The first approaches of targeted DNA methylation was done by fusing ZFPs to the bacterial cytosine-5 methyltransferase M.SssI to methylate the $p21^{WAF1/CIP1}$ gene.[158] Others used the methyltransferases M.HhaI and M.HpaII but in all cases strong off-target effects were observed due to non-specific binding of ZFPs.[152,159] Still, over several years multiple studies using ZFP-based methylation tools were published including methylation with DNMT3a since no better alternative was available.[147,152] This changed with the usage of TALEs as DBDs and targeted DNA methylation was demonstrated using a DNMT3a-DNMT3L fusion protein.[160] The specific methylation of the CDKN2A locus with up to ~50 % 5mC level for single CpGs resulted in decreased gene

expression. However, CpGs had to be avoided in target sequences to guarantee binding since TALEs are sensitive to epigenetic DNA modifications (3.3.2). Binding of gRNA in CRISPR/dCas9-based tools is not sensitive to such modifications and the system was first used with DNMT3a to target the BACH2 and IL6ST genes.[161] The use of multiple gRNAs to target the CDKN2A gene resulted in up to 50 % methylation.[162] Previous mentioned studies used only the catalytic domain of DNTM3a, but in 2016 dCas9 was fused to full length DNMT3a.[163] As in the previous mentioned TALE-based study, other dCas9-based studies used a fusion construct from DNMT3a and DNMT3L since this improves the catalytic activity (see 3.2.2).[164] DNA was methylated ~400 bp around the binding site which is comparable to studies with TALEs.[160,164] This was improved to ~4.5 kb at HOXA locus using a SunTag fused dCas9 to recruit multiple DNMT3a proteins.[165,166] SunTag is a repeating polypeptide array which is recognized by single-chain variable fragments (scFv) of corresponding antibodies.[165,166] DNMT3a is fused to scFv and therefore recruited to the SunTag-dCas9.[166] Free DNMT3a however can result in high methylation background and this is why a DNMT3a mutant with lower DNA binding ability was used for the SunTag-based system.[167] A more rapid DNA methylation was achieved by fusing a modified version of M.SssI from *Mollicutes spiroplasma* ($MQ1^{Q147L}$) to dCas9.[168] A split version of M.SssI was used to reduce background methylation since the C-terminal part was fused to dCas9 which localizes functional enzymes to the target site.[169] Recruitment of DNMT3a to subtelomeric sequences was also done by merging it to telomere repeat binding factor 1 (TRF1) using the light-activatable heterodimerizing system of cryptochrome 2 protein (CRY2) and cryptochrome-interacting basic-helix-loop-helix protein 1 (CIB1).[170] All systems though use constitutively active enzymes which could lead to high methylation backgrounds, and it does not offer temporal control. DNA methylation can also be decreased on a global level by inhibition of DNMTs with the cytosine analog 5-aza-2'-deoxycytidine (5-Aza). 5-Aza is incorporated in place of cytosine during DNA replication. DNMTs (especially DNMT1 during DNA replication) can bind to such analog during the methylation process but since the carbon-5 atom is substituted with a nitrogen atom, the elimination step cannot happen and the DNMT is irreversibly bound to the DNA (see Figure 6, page 13). This leads to DNA damage and subsequent degradation of the bound DNMT, lowering the overall 5mC level by passive dilution.[171,172] However, 5-Aza does not offer any selectivity, and it is chemically instable and toxic.[173]

DNA methylation cannot only be studied by writing it with DNA methyltransferases, but also by erasing it. First efforts were made by recruiting the thymidine DNA glycosylase to specific loci using ZFPs or the DNA binding domain Rel-homology domain (RHD) of the transcription factor NF-κB.[174] Also, all three TET enzymes were fused to ZFPs for targeted demethylation.[175] Although TET2 showed the highest demethylation effect, other studies

using TALE- or CRISPR/dCas9-based methods (including SunTags) used TET1.[163,175,176,177]

Beside gene expression regulation, these tools can be used to examine the temporal stability and heritability of DNA methylation. While written 5mC patterns were found to be stable in mice, decreasing 5mC levels were reported in other studies.[152,178] Also, after the removement of 5mC with TET1, re-methylation was observed diminishing the editing effects.[176] However, all tools with only spatial control over DNA methylation do not allow precise kinetics measurements of DNA methylation itself or dynamic downstream processes since the start of activity can only be controlled poorly. To get temporal control, it is necessary to leave the effector domains inactive until precise and simultaneous activation at a user defined time point. Multiple approaches regarding temporal control of epigenetic effectors were made and are discussed in 3.4.

3.3.2. Transcription Activator Like Effectors

TALE proteins were found in the group of *Xanthomonas* bacterial plant pathogens. These proteins are translocated from the bacteria into cells of the infected plant via the hypersensitive response and pathogenicity (Hrp) type III secretion system (T3S) where they can bind specific promoter sequences resulting in upregulation of such genes promoting bacterial virulence.[179] The structure of TALEs can be divided into three main parts: The N-terminal region (NTR) with the secretion and translocation signal, the C-terminal region (CTR) with NLS needed to import the proteins into the plant cell nucleus and the activation domain (AD) necessary for upregulation of bound genes, and the central repeat domain (CRD) responsible for selective DNA binding.[179]

The CRD of most naturally occurring TALEs consists of nearly identical 34-amino acid (102 bp) long repeat elements (see sequence in Figure 10A) whereby number and order of such repeats are variable.[180] The last repeat with a length of 19 or 20 amino acids is shorter than the other repeats and is therefore described as the ‘last half repeat’.[179] TALEs found in nature have a length of 1.5 to 33.5 repeats with most showing a length of 15.5 to 19.5 repeats.[179] The short TALEs are probably nonfunctional since it was demonstrated that a minimum of 6.5 repeats is necessary for target gene expression.[154] Variations within the repeats mainly occur at amino acids four, twelve, thirteen and thirty-two (numbering of a 34 amino acid long repeat) of which the variations at the twelves and thirteens position are responsible for DNA base recognition and are called repeat variable diresidue (RVD).[154,181] Each repeat folds into two α-helices with different sizes connected by a loop containing the RVD facing the respective DNA base (Figure 10B) and by wrapping around the DNAs major groove, the TALE forms a right-handed superhelix as a strand-specific DNA-binder protein (Figure 10C).[182]

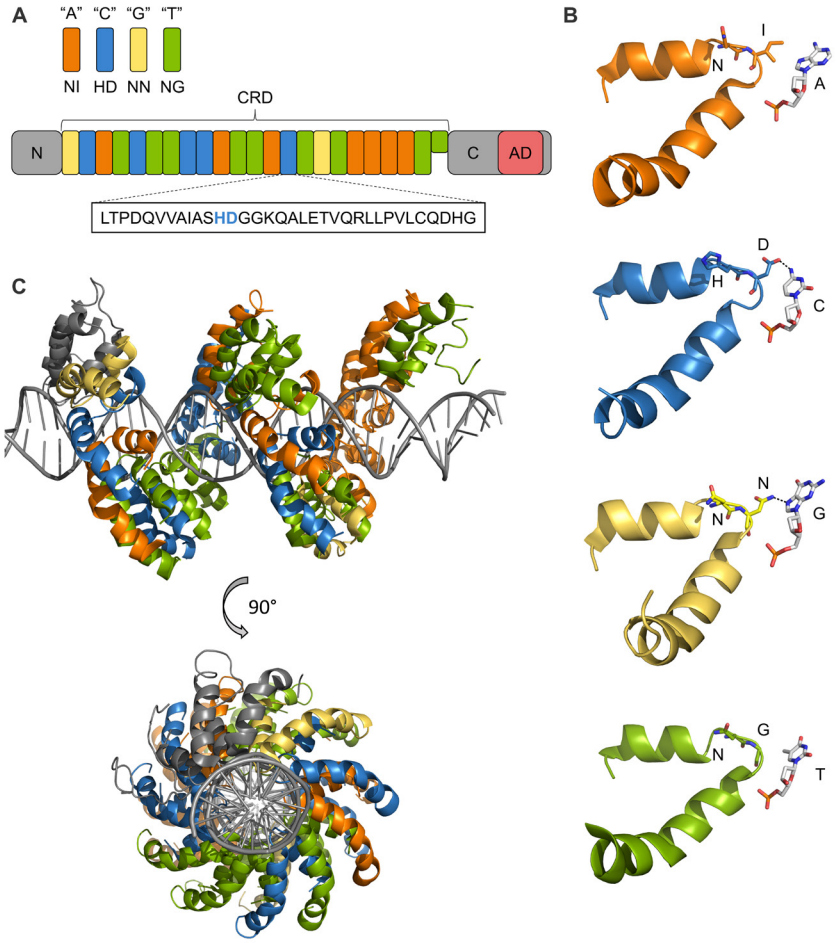


Figure 10: Structure of TALEs. A: Cartoon overview of TALEs consisting of the NTR ('N'), CTR ('C'), and the central repeat domain ('CRD') with its single repeat elements shown as colored rectangles of which only the 12th and 13th amino acid are variable to bind single DNA bases according to the code shown on top (here, the example HD for cytosine binding is shown). The last repeat is truncated, and the activation domain ('AD') is localized at the C-terminus. **B:** Crystal structures of single TALE repeats with RVDs NI, HD, NN and NG bound to the corresponding DNA base (white). Taken from PDB 3UGM.[182,183] **C:** Crystal structure of a TALE (PDB 3UGM[182,183]; color code of single repeats according to A bound to the major groove of the targeted DNA (gray).

For each nucleobase a common respective RVD was found in nature: adenine is bound by NI (Asn12, Ile 13), cytosine by HD (His12, Asp13), guanine by NN (Asn12, Asn13) and thymine by NG (Asn12, Gly13), whereby NN is also recognizing adenine.[154,181] Of such four, RVDs NN and HD show a strong, and RVDs NI and NG a weaker DNA interaction.[184] Additionally, RVD NS was found to be a binder of all four DNA nucleobases and also other less common RVDs were found (HG against C or T and N* with a deletion at the 13th position mainly against C).[154] For all RVDs the amino acid at position 12 is only stabilizing the proteins backbone via contacts to the carbonyl oxygen of position 8 whereas the amino acid at position 13 is responsible for base-specific DNA interaction providing different mechanisms for the different RVDs (Figure 10B).[182] The aliphatic side chain of isoleucine of RVD NI builds non-polar Van der Waals contacts to C8 and N7 of the adenine purine ring.[182] The aspartate of HD not only builds Van der Waals contacts to cytosine but also an

additional hydrogen bond with the N4 atom.[182] A hydrogen bond is also found in the interaction of RVD NN and guanine. Asn13 builds such bond with the N7 of guanine and since also adenine has this N7 nitrogen in the purine ring, it explains the similar recognition by RVD NN.[182,185] Since the RVD NG is lacking any side chain at the 13th position (glycine), it gives enough space for the 5-methyl group of thymine on the one hand and also enables Van der Waals interactions between the glycine's Cα atom and the 5-methyl group.[179,182,186,187] 5mC does not differ from the DNA base thymine with respect to the positions involved in DNA-TALE binding and can be also bound by the RVD NG.[187] On top of the RVD-DNA interactions, amino acids Lys16 and Gln17 are contributing in all repeats to non-selective DNA binding by building hydrogen bonds to the DNAs phosphate backbone.

The N-terminus harbors cryptic repeats with similar secondary structures as the repeats in the CRD (repeats 0, -1, -2, and -3).[182,188] The indole ring of the conserved W232 of repeat -1 faces the methyl group of thymine forming Van der Waals forces. It was observed that target sites of TALEs found in nature start with a 5'-T and it was thought that this is necessary for binding.[154,182,189] Later studies showed similar affinities to T-rich and T-deficient dsDNA, promoting a more nonspecific DNA interaction of the NTR[188] whereas other studies showed severe decreased activity if the 5'-T was missing.[190] The NTR as a whole is very important for full TALE binding with more than 100 essential residues, especially positively charged ones, with respect to DNA binding of customized TALEs.[188,191] Affords were made to modify the NTR to bind all possible 5'-bases with the generation of a NTR harboring two amino acid substitutions (Q231S and W232R) to accept a preceding guanine instead of thymine.[190] It was also demonstrated that not only the NTR but also the overall length and sequence of the CRD and especially the RVD of repeat 1 has an influence on the binding capacity of the 5'-base.[192,193]

Since the design of the CRD follows the previously described nucleobase-RVD code, TALEs for almost every target sequence in the genome can be assembled and are less limited than ZFPs. For such assembly different methods are available ranging from fast ligation-based automatable solid-phase high-throughput (FLASH) assembly[194] over ligation-independent cloning[195] and isothermal assembly[196] to Golden Gate cloning[157]. Since only the CRD is changed between different TALEs, the entry vector used for assembly harboring the NTR and CTR can always stay the same which enables fast and flexible side by side assembly of various TALEs. The main drawback though is that the described nucleobase-RVD code has some limitations with respect to the affinity and accurate binding of target sequences. The TALE binding is influenced by the sequence context: position, neighboring repeats and overall number of single RVDs.[184,193,197,198] Nevertheless, TALEs are good DNA-binders and due to its easy and fast assembly, multiple TALEs can be tested

to find the best sequence to target a specific genomic locus. Additionally, off-target binding is also observed with other common programmable DNA-binding proteins like ZFPs or CRISPR/Cas9 based proteins.[155,199,200] A huge benefit of TALEs and the CRISPR/Cas9 system though is that tightly repressed genes can be targeted due to their chromatin remodeling abilities where TALEs showed the best performance.[200,201] To further improve the binding affinities of TALEs, different RVD libraries were tested resulting in various new RVDs to increase the specificity of TALEs, especially to overcome the double recognition of adenine and guanine by RVD NN.[198,202] Different to ZNF and CRISPR/Cas9 based proteins, the nucleobase information presented in the DNA major groove has an impact on the binding ability of TALEs due to its helical binding mechanism. As already observed with RVD NG, also cytosine modifications (in this case 5mC) can be bound by TALEs. Therefore, a broadened RVD spectrum including the specific recognition not only for A, C, G and T, but also for epigenetic cytosine modifications was developed. This sensitivity enables diagnostic approaches to analyze specific epigenetic marks of defined sequences based on the binding ability of TALEs. Also, it allows the specific binding of genes only if respective cytosine modifications occur. This discrimination might be useful to target genes only in cancer cells with a given aberrant epigenetic pattern whereas healthy cells with a normal pattern would not be targeted. First, it was demonstrated that the natural occurring RVD HD does not bind 5mC due to a steric clash between the C5-methyl group and the aspartic acid.[187,203,204] On the other hand, 5mC cannot only be bound by RVD NG due to the missing side chain but also by RVD N* with a deletion at the 13th position.[203] Such deletion leads to a smaller loop region since the conserved 15th amino acid of the repeat is not shifting completely into the empty space. Three RVDs, G*, S* and T*, showed strong binding towards all cytosine modifications (C, 5mC, 5hmC, 5fC and 5caC) and can serve as universal binder.[205,206] These RVDs guarantee TALE-binding if target sequences include a cytosine with unknown modifications. RVD P* on the other hand showed binding towards all modifications but 5caC and can be used as negative selective RVD to test for this modification.[205] Deletion of all amino acids from position 12 to 15 though revealed a weak but selective 5caC binding RVD.[207] R**** (amino acid 11 is mutated to R) leads to a size reduced loop region in which the side chain of arginine faces towards 5caC enabling the formation of hydrogen bonds.[207]

3.4. Concepts in Optobiology and Optochemical Biology

To investigate biological mechanisms within living cells it might be necessary to take control over the interaction or catalytic activity itself of specific proteins. This can be achieved for example with small molecules that can inhibit specific PPIs[208] or degrade proteins in a selective manner[209]. Drawbacks of this approach however are that small molecules must be cell permeable, must not interfere with other proteins, and that they only offer low spatio-temporal control within cells. The usage of light however offers a non-invasive method to trigger minimal chemical changes in cells with high spatio-temporal resolution. Therefore, many applications using light as stimulus were published including the control of protein activity, their nuclear localization, or gene expression.[210–214] Light can be also used to activate small molecules which at least improves the temporal control since these molecules can be given to cells before actual activation.[212] This was demonstrated with a photocaged version of the common chemical inducer of dimerization (CID) rapamycin[215], or with a photocaged trimethoprim group connected to a alkyl chloride to receive *E. coli* dihydrofolate reductase (DHFR) binding with the HaloTag enzyme[216]. Another strategy allowed the deactivation of PPIs by using a light cleavable CID.[217] DNA methylation was also controlled with photochemical methods. DNMT inhibitor 5-Aza was photocaged with diethylaminocoumarinyl-4-methyl (DEACM) which shows high absorption at 385 nm.[218] Upon light irradiation the inhibitor became active and 5mC levels decreased via passive dilution mainly due to DNMT1 inhibition.[218] However, this method is based on enzyme deactivation rather than activation, it has low temporal resolution since cell division is necessary for passive dilution, and the inhibitor is incorporated ubiquitously into the DNA and subsequent inhibits all DNMTs which does not allow gene specific editing, and it also leads to DNA damage.[172] In a different approach, the binding ability of small molecules towards the target protein was deactivated upon light irradiation.[219] This made it possible to control the concentration of active small molecules and enables multiple deactivation-activation rounds by adding new non-irradiated small molecules.

To get light control over single enzymes, multiple natural systems were made usable which become active during light irradiation due to conformational changes but inactive again without light irradiation. For example, expression of two different genes was regulated multichromatic by using green and red photoswitchable phytochromes.[220] Transcriptional activation can be also activated with blue light which leads to the binding of the photocontrolled transcription factor EL222 and subsequent recruitment of RNA polymerase.[221] Another natural light sensor activated with blue light are light-oxygen-voltage-sensing domains (LOV domains) interacting with a C-terminal α-helix (Jα) in the dark state leading to steric hindrance.[222] Blue light was also used to control T7 RNA

polymerases in *E. coli*.[223] For that, a split version of the polymerase was fused to either the nMag or pMag heterodimerizing protein domain which bind each other after blue-light induction.[223,224] Another heterodimerizing system is the CRY2-CIB1 dimerization upon blue light.[225] As already described in 3.3.1, it was for example used to recruit DNMT3a to subtelomeric regions after light irradiation by fusing CRY2 to DNMT3a, and CIB1 to TRF1.[170] These tools are switchable since they become inactive without light irradiation, but constant light irradiation can result in cell damage, and fusion proteins can lead to structural changes of the protein of interest and therefore possible alterations in protein behavior.[226] Therefore, better photoresponsive tools had to be designed to get precise temporal control over single enzymes with least structural changes. Genetic code expansion (GCE) describes the method to introduce photocaged non-canonical amino acids (ncAAs) site-specifically into proteins enabling the direct control of PPIs or the activity of the catalytic site itself.

3.4.1. Genetic Code Expansion

The translation of the mRNA sequence into a functional protein takes place in the cytoplasm of eukaryotic cells.[1,2] tRNAs can recognize so called base triplets or codons of the mRNA with their anticodon (shown in black in Figure 11A). With the four RNA bases, A, C, G, and U, there are 64 possible codons of which 61 have corresponding tRNAs decoding 20 proteinogenic amino acids (Figure 11B), which are the basic modules of proteins. The fact that only 20 amino acids are decoded (degenerated code) was commonly described as the ‘frozen accident’ because any additional change to this imperfect system was expected to be lethal.[227] Most amino acids are encoded by multiple codons – some with even six different ones (Leu, Ser, and Arg; see Figure 11B). The 3’-end of tRNAs with the sequence CCA is unpaired and represents the position of aminoacylation (Figure 11A), which describes the covalent addition of an amino acid to the respective tRNA faithfully by an aminoacyl tRNA synthetase (aaRS).[228] With this, tRNAs connect the mRNA codons with corresponding amino acids based on the genetic code.[229] The translation still requires the catalyzation by the ribosome consisting of the large and small subunit assembled by several ribosomal proteins and rRNA.[230] As described before, three of the 64 codons do not encode amino acid but serve as stop codons (also called nonsense codons; Figure 11B).[231] Instead of tRNAs, eukaryotic release factors (eRF) bind to the stop codon in eukaryotes which lead to the hydrolysis of the peptide chain and the disassembling of the ribosome.[1,232] Finally, the peptide chain folds into a specific three-dimensional structure. However, many proteins require posttranslational modifications (PTMs) like phosphorylation, acetylation, or methylation to gain their full and correct function.[233]

A

B

TTT → Phe (F) 0,46	TCT → Ser (S) 0,19	TAT → Tyr (Y) 0,44	TGT → Cys (C) 0,46
TTC → Phe (F) 0,54	TCC → Ser (S) 0,22	TAC → Tyr (Y) 0,56	TGC → Cys (C) 0,54
TTA → Leu (L) 0,08	TCA → Ser (S) 0,15	TAA → 0,30	TGA → 0,47
TTG → Leu (L) 0,13	TCG → Ser (S) 0,05	TAG → 0,24	TGG → Trp (W) 1,00
CTT → Leu (L) 0,13	CCT → Pro (P) 0,29	CAT → His (H) 0,42	CGT → Arg (R) 0,08
CTC → Leu (L) 0,20	CCC → Pro (P) 0,32	CAC → His (H) 0,58	CGC → Arg (R) 0,18
CTA → Leu (L) 0,07	CCA → Pro (P) 0,28	CAA → Gln (Q) 0,27	CGA → Arg (R) 0,11
CTG → Leu (L) 0,40	CCG → Pro (P) 0,11	CAG → Gln (Q) 0,73	CGG → Arg (R) 0,20
ATT → Ile (I) 0,36	ACT → Thr (T) 0,25	AAT → Asn (N) 0,47	AGT → Ser (S) 0,15
ATC → Ile (I) 0,47	ACC → Thr (T) 0,36	AAC → Asn (N) 0,53	AGC → Ser (S) 0,24
ATA → Ile (I) 0,17	ACA → Thr (T) 0,28	AAA → Lys (K) 0,43	AGA → Arg (R) 0,21
ATG → Met (M) 1,00	ACG → Thr (T) 0,11	AAG → Lys (K) 0,57	AGG → Arg (R) 0,21
GTT → Val (V) 0,18	GCT → Ala (A) 0,27	GAT → Asp (D) 0,46	GGT → Gly (G) 0,16
GTC → Val (V) 0,24	GCC → Ala (A) 0,40	GAC → Asp (D) 0,54	GGC → Gly (G) 0,34
GTA → Val (V) 0,12	GCA → Ala (A) 0,23	GAA → Glu (E) 0,42	GGA → Gly (G) 0,25
GTG → Val (V) 0,46	GCG → Ala (A) 0,11	GAG → Glu (E) 0,58	GGG → Gly (G) 0,25

Figure 11: Structure of tRNA and table of the genetic code. A: Crystal structure of yeast phenylalanine tRNA (PDB 1EHZ[234]). The three bases of the anticodon are shown as black sticks. Color code of other parts: red: CCA tail, green and pink: acceptor stem, orange: T arm, dark blue: variable loop, yellow: anticodon arm, light blue: D arm. **B:** Codon usage table with percentage distributions in *Homo sapiens* (preferred codon for each amino acid highlighted in green). The three stop codons TGA, TAA and TAG are shown in red. Figure taken from[235].

Though most organisms can only translate 61 codons into the previously described 20 amino acids, there are some exceptions in which organisms can reassign codons. Contrary to the 'frozen accident' hypothesis, the 'codon capture' hypothesis proposes that during evolution a codon and its respective tRNA can disappear just to reappear in a later stage with the tRNA now carrying either the former or a different amino acid.[236] Codon reassignments were early observed in mitochondria[237], *Mycoplasma capricolum*[238], and ciliated protozoa[239] in which mostly one or two of the three stop codons ochre (UAA), amber (UAG) or opal (UGA) got a respective tRNA. Later, additional codon reassignments were found with some tRNAs carrying different amino acids than the known 20 proteinogenic ones.[240] However, only two additional amino acids, selenocysteine (SeC) and pyrrolysine (Pyl), were found which shows how difficult it is to expand the natural 20-amino acid code.[240] SeC is also known as the 21st proteinogenic amino acid and was detected in all three domains of life and is translated by tRNAs recognizing the opal stop codon UGA.[241,242] It provides specific features for some enzymes but its incorporation is dependent on specific translation factors.[242–244] The mechanism was well studied in *E. coli* in which four genes (*selA*, *selB*, *selC*, and *selD*) as well as a cis-acting structure in the mRNA (selenocysteine insertion sequence (SECIS) element) located directly after the opal codon are needed to incorporate SeC.[242–244] The SECIS element is guiding the SeC-specific elongation factor SelB to the ribosome which catalyzes the insertion of SeC.[242–244] Compared with the standard factor EF-Tu, SelB has an extra C-terminal domain binding the SECIS element and it can recognize tRNAs charged with SeC.[242–244] Interestingly, there is no specific aaRS for SeC and tRNAs are initially aminoacylated with a seryl residue and subsequent converted to a selenocysteinyl moiety with different pathways in bacteria and

archaea.[242–246] The lysine derivative Pyl is the 22nd proteinogenic amino acid and is translated with a tRNA recognizing the amber stop codon UAG in a few methane-producing archaea, and bacteria.[246,247] Different to SeC incorporation there is a corresponding aaRS (PylRS) which can selectively aminoacylate the $tRNA_{Pyl}$ with Pyl.[248–250] Translation is performed by the standard mechanism whereby it was shown that a pyrrolysine insertion sequence (PYLIS) element is not essential but enhances the translation efficiency.[246,251] The fact that almost all natural occurring codon reassignment affects the stop codons shows that such changes have the least effects on the proteome. If no other tRNAs exist, the only mechanism to compete with is the termination process by release factors which could lead to truncation products. Additionally, the stop codons are the least occurring codons in a gene since they are only needed once to stop the translation. The mechanism of codon reassignment was also used to artificially expand the genetic code.[211] As said before, the natural chemical space of proteins is greatly expanded by PTMs and the study of such PTMs was complicated due to the poor opportunities to introduce them at specific amino acids of specific proteins in living cells. Such methods included solid-phase synthesis[252], expressed protein ligation[253] and chemical modification of amino acids (especially cysteines) in expressed proteins[254]. Drawbacks are size limitations of synthesized peptides, the need to import peptides and proteins into cells or low selectivity of chemical reactions. However, the ability to introduce ncAAs site-specifically into proteins by using the cells own translation machinery enables the easy analysis of protein functions, interactions, and structures altered by specific amino acid modifications.[255,256]

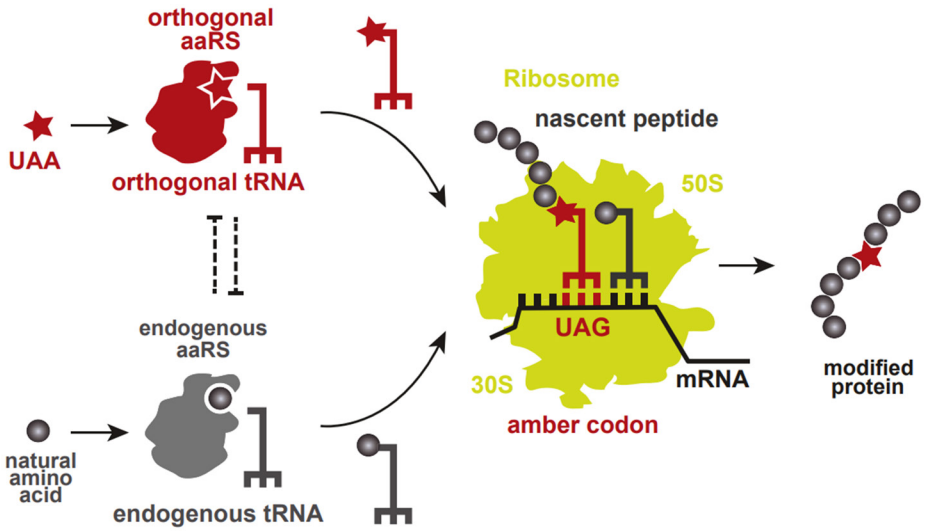


Figure 12: Principal overview of genetic code expansion. Unnatural amino acids (UAA, also known as ncAA) can be loaded onto orthogonal tRNAs which can recognize the UAG codon which naturally serves as a stop codon. During ribosomal translation (here shown with the prokaryotic 30S and 50S subunits) the UAA can be incorporated into the growing peptide chain. Figure was adapted from[211].

For this, usually the least occurring stop codon amber (TAG in Figure 11B) is used to replace the codon of a gene that encodes the amino acid which should be replaced with the modified ncAA.[256] Amber codons rarely terminate essential genes and they do not occur often among stop codons in the genome (9 % in *Escherichia coli*, 23% in *Saccharomyces cerevisiae* and 24 % in *Homo sapiens*).[235,256] Moreover, a tRNA recognizing the amber codon and a corresponding aaRS which can amino acetylate this tRNA selectively with the ncAA are needed. Both, the tRNA and the aaRS must be orthogonal which means that the aaRS can only recognize the orthogonal tRNA and ncAA, as well as that endogenous synthetases cannot recognize the orthogonal tRNA and ncAA (Figure 12).[256] The tRNA charged with ncAA also has to be functional with the cells ribosomal translation mechanism without any additional factors and the ncAA must be stable, cell permeable and non-toxic. Specific recognition of respective tRNAs by their cognate aaRS is based on positive (determinants) and negative (anti-determinants) tRNA identity elements.[257,258] These elements include structural motifs, base pairs in helices or just single nucleotides in single-stranded regions and they differ between the three life domains.[258,259] Because of this, some aaRS/tRNA pairs from one life domain were found to be orthogonal in another life domain.[256] After incorporation of ncAAs was demonstrated in vitro[260], first attempts of amber suppression in *E. coli* were done by mutating the anticodon of standard tRNAs to see which endogenous amino acid will be charged.[261] Artificial genetic code expansion with an ncAA in *E. coli* however was done by using the yeast phenylalanyl tRNA/synthetase pair to incorporate *p*-fluoro-phenylalanine with an efficiency of 64-75 %.[262] Much better amber suppression was achieved with a mutated tyrosyl tRNA/synthetase pair of the archaebacterium *Methanococcus jannaschii*.[263] The identity elements of this tRNA differ from the *E. coli* tyrosyl tRNA and the tyrosyl tRNA synthetase (TyrRS) has only a weak anticodon loop binding domain and no proofreading which allows the switch to a CUA anticodon recognizing the amber codon and the use of ncAAs.[263,264] Since *M. jannaschii* TyrRS recognizes tyrosine, additional mutations with subsequent selection rounds were necessary to get a fully orthogonal tRNA/aaRS pair which selectively incorporated the ncAA *O*-methyltyrosine at amber codons.[263] Although it was already shown that *M. jannaschii* $tRNA_{CUA}^{Tyr}$ is a poor substrate for the *E. coli* synthetases, eleven more nucleotides were randomly mutated to decrease the recognition of them.[263,265] Selection was done by rounds of negative and positive selection. To remove tRNAs recognized by *E. coli* synthetases, the amber codon was introduced in the lethal barnase gene which would be expressed if tRNAs are not orthogonal.[263] Positive selection was done by amber suppression of the β-lactamase gene which is essential for the cell to survive in the presence of ampicillin.[263] Positive-negative selection rounds were also done with the *M. jannaschii* TyrRS. For this, five amino acids in the active site in close proximity to the aryl ring of tyrosine were randomly

mutated.[263] Positive selection was done by amber suppression in the chloramphenicol acetyltransferase (CAT) gene which is crucial for survival in the presence of chloramphenicol.[263] To remove all clones which charged the $tRNA_{CUA}^{Tyr}$ with natural amino acids, the surviving clones were grown again in the presence of chloramphenicol but in the absence of *O*-methyltyrosine. Cells with orthogonal TyrRSs died but were recovered from a replica plate with the unnatural amino acid. After multiple selection rounds a fully orthogonal tRNA/aaRS pair was received which showed selective amber suppression only in the presence of *O*-methyltyrosine.[263] Later, the *M. jannaschii* $tRNA_{CUA}^{Tyr}$/TyrRS pair was used to incorporate L-3-(2-naphthyl)alanine[266] and *p*-benzoyl-L-phenylalanine[267]. Additional to the *M. jannaschii* pair which can be only used in *E. coli*, also other pairs were developed.[211] They include the *E. coli* tyrosyl tRNA synthetase (EcTyrRS)/$tRNA_{CUA}$ pair usable in yeast[268], the *Methanosarcina mazei* and *Methanosarcina barkeri* pyrrolysyl tRNA synthetase (PylRS)/$tRNA_{CUA}$ pair usable in prokaryotes and eukaryotes[248,249,269], and the *E. coli* leucyl tRNA synthetase (EcLeuRS)/$tRNA_{CUA}$ pair usable in eukaryotes[270,271]. The strategy to gain full orthogonality always includes mutagenesis of crucial amino acids of the aaRS responsible for tRNA and ncAA recognition or mutagenesis of the tRNA to decrease recognition by endogenous aaRSs with subsequent positive and negative selection rounds.[211,272] Recently, a computational approach was presented which allows the rapid screening of possible orthogonal tRNA/aaRS pairs in *E. coli*.[273] With all these orthogonal pairs, it is possible to incorporate a wide variety of ncAAs with different chemical properties in prokaryotes and eukaryotes.[210,211,274] Applications include protein cross-linking, altering redox potentials, addition of fluorophores and infrared probes, spin-label and NMR probes, post-translational modifications, site-specific protein conjugation, therapeutic applications, and the control of enzyme activities.[210]

Scientists were able to recode the genome of *E. coli* by deleting two sense codons (TCG and TCA encoding serine) and the amber stop codon TAG.[275] This demonstrated that organisms can indeed live with a reduced number of synonymous sense codons.[275] It also opens the possibility to reassign the non-used codons to expand the space of amino acids used in *E. coli*.[275,276] In further studies all endogenous amber stop codons were deleted in *E. coli* and genetic code expansion was done with the frameshift codon UAGA.[277] But even multicellular organisms were subject to genetic code expansion including *Caenorhabditis elegans*[278], *Drosophila melanogaster*[279], *Danio rerio*[280] and *Mus musculus*[281,282]. Other approaches to achieve genetic code expansion included the use of the opal stop codon TGA[283] or other quadruplet codons[284,285]. The combination of amber and frameshift codons made it possible to incorporate two different ncAAs into the same protein.[285] Later, the possibility to incorporate multiple ncAAs at the same time was even evolved.[286]

3.4.2. Photoresponsive Non-Canonical Amino Acids

With the strength of genetic code expansion it is possible to directly photocage proteins by site-specific incorporation of ncAAs carrying a light-labile chemical group, also called photoremovable protecting group (PPG; see Figure 13 for selected photocaged ncAAs).[213,287] This results in the expression of inactive proteins with least structural change and provides a high temporal control. After light-activation, the PPG is cleaved of and the wild type protein becomes active, enabling kinetics measurements of this protein in its natural context.

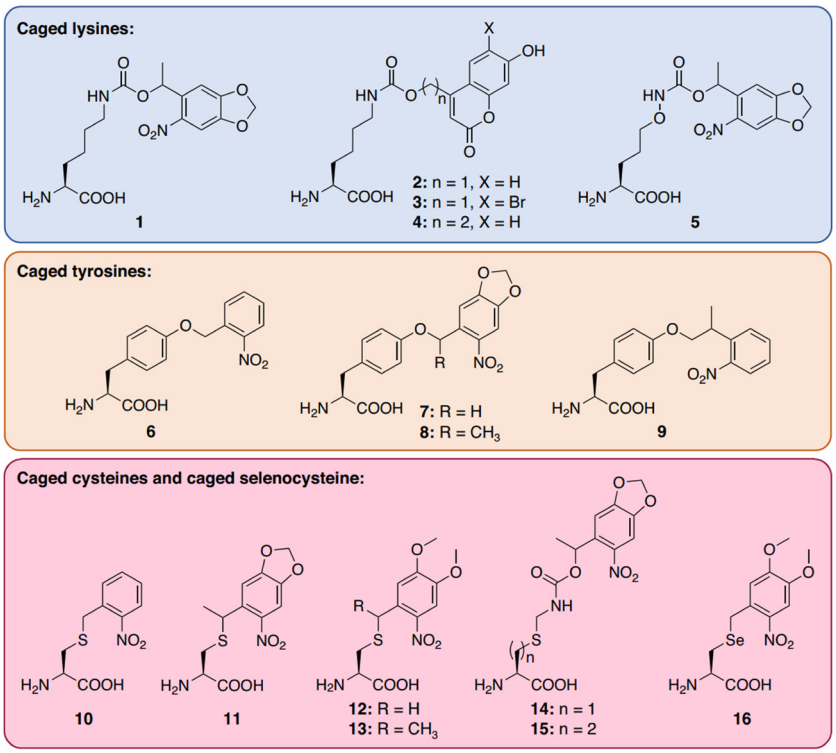


Figure 13: Structures of selected photoremovable protecting groups for lysine, tyrosine, and cysteine. Figure adapted from[213].

Widely used are *o*-nitrobenzylic (NB) derivatives which are cleaved of upon light irradiation.[270,287,288] NB itself is cleaved off with 365 nm light and was used for optical control of several enzyme classes (caged tyrosine (**6**) and caged cysteine (**10**) in Figure 13).[213,270,287,289] The strategy usually involves the replacement of a catalytically important amino acid with its photocaged analog.[213] Irradiation leads to the cleavage of the caging group leaving an unmodified and functional enzyme. If such important amino acid is not available, it is possible to use light-controlled inteins which are self-cleaved out of the protein after light irradiation.[290] To reduce the stress for living cells, other caging groups with a red-shifted absorption spectrum were developed to be able to use visible light (405 nm), or which give ketones instead of aldehydes as byproducts. These include *o*-nitropiperonyl (NP,

7[291] in Figure 13), methyl-*o*-nitropiperonyl (MNP, **8**[291] and **11**[292] in Figure 13), *o*-nitrophenylpropyl (NPP, **9**[291] in Figure 13), and 4,5-Dimethoxy-2-nitrobenzyl (DMNB, **12**[290] and **16**[293] in Figure 13). The absorption spectrum of DMNB allows uncaging with blue light (405 nm) due to their electron donating methoxy groups.[271] The uncaging mechanism of DMNB-caged cysteine which was used in this thesis is shown in Figure 14.

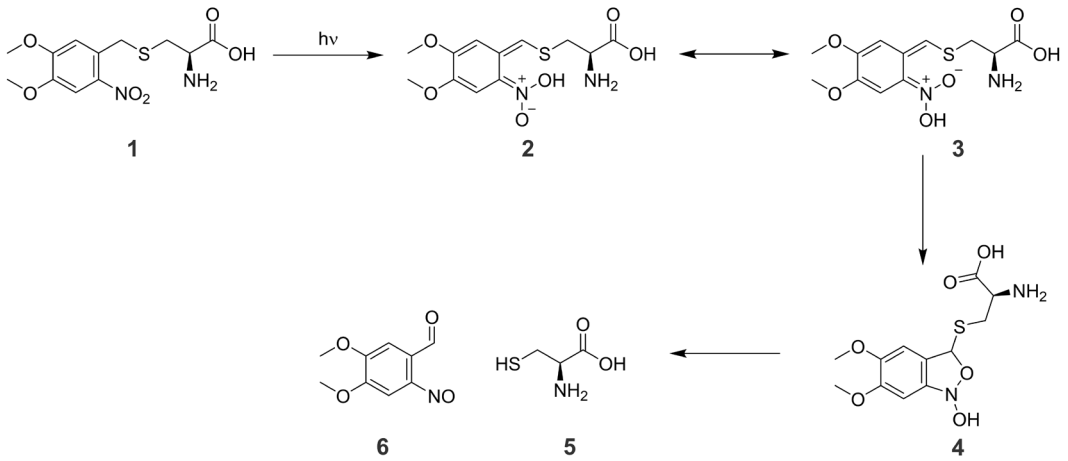


Figure 14: Decaging mechanism of DMNB-Cys. Photon absorption by **1** results in *Z*-nitronic acid (**2**) which is in equilibrium with *E*-nitronic acid (**3**). After cyclization to *N*-hydroxy-5,6-dimethoxy-2,1-benzisoxazoline (**4**), it decays to cysteine (**5**) and 4,5-dimethoxy-2-nitrosobenzaldehyde (**6**). Reaction mechanism similar as shown in[287,294,295].

Photon absorption leads to a transfer of the benzylic proton to the nitro group creating a *Z*-nitronic acid (**2** in Figure 14) which is in equilibrium with its isomeric *E*-nitronic acid (**3** in Figure 14). After cyclization to *N*-hydroxy-5,6-dimethoxy-2,1-benzisoxazoline (**4** in Figure 14) it immediately decays into the decaged cysteine (**5** in Figure 14) and 4,5-dimethoxy-2-nitrosobenzaldehyde (**6** in Figure 14).[287,294–296] Besides the previously mentioned light-controlled intein[290], also other biological functions were controlled with DMNB. The small molecule inhibitor anisomycin can inhibit protein biosynthesis by binding to the ribosome and was photocontrolled with DMNB.[297] Another possibility to inhibit the translation process was the photocaging of a crucial tyrosine with DMNB of 4E-BP proteins.[298] Initiation of cap-dependent translation is done by binding of 4E-BP proteins to the initiation factor eIF4E and caging of a conserved tyrosine residue leads to a steric clash.[298] Also, the specific role of central regulatory kinase mTORC1 was examined by indirect photocontrol in HeLa cells.[298] Instead of photocaging mTORC1, DMNB was attached to rapamycin which builds an inhibitor complex of mTORC1 with FKBP.[298] Introduction of DMNB-caged amino acids with genetic code expansion however was demonstrated in *Saccharomyces cerevisiae*, HEK293T cells, hippocampal neurons and the neocortex of mouse embryos.[271,281] The role of different serines in the transcription factor Pho4 was examined in *Saccharomyces cerevisiae*.[271] Phosphorylation of a specific serine was first blocked by the selective

incorporation of DMNB-caged serine and then triggered after light irradiation and cleavage of DMNB.[271] The application of DMNB-caged cysteine was proven in mammalian cell lines and even in mouse embryos. The bulky DMNB was used to photocage a cysteine in the pore of the inwardly rectifying potassium channel Kir2.1 which led to a physical blockage of it.[281,299] In all cases the mutated orthogonal leucyl *E. coli* amber suppressor $tRNA_{CUA}^{Leu5}$/$LeuRS^{BH5\ T252A}$ was used.[271,281]

Another widely used PPG is MeNPOC to photocage amongst others lysine (**1** and **5** in Figure 13)[300], cysteine (**14** in Figure 13)[301], or homocysteine (**15** in Figure 13)[301]. Lysines were also photocaged with coumarin derivatives allowing the photolysis at 405 nm (**2** in Figure 13), and 760 nm with two-photon irradiation (**3** in Figure 13).[302] Other PPGs are *o*-nitrophenylethyl (NPE)[303], its 4,5-dimethoxy analog (DMNPE, **13** in Figure 13)[290,304], and α-carboxy-2-nitrobenzyl (CNB)[305]. Not only the active site itself can be used for incorporation of photocaged amino acids but also interaction sites needed for PPIs. Using a photocaged lysine residue, the interaction between phosphatase MKP3 and ERK was controlled.[306]. Also the interaction of antibodies[307] and nanobodies[308] was subject for photocontrol.

With these tools it is possible to study the biology of specific enzymes using their wild type sequence with high temporal control. It could also improve the previously described systems for targeted editing of DNA methylation.

4. Aim of This Work

5mC as a dynamic epigenetic DNA modification plays a crucial role in transcription regulation and differentiation. However, investigations of effects by 5mC at specific loci are still restricted due to the inability to properly control DNMTs and with it 5mC with spatio-temporal control. The current tools for targeted editing of 5mC patterns all have severe drawbacks since most either offer no spatial control, or they use constitutively active DNMTs. This prohibits the analysis of kinetics of downstream processes like reader recruitment or gene silencing. The only approach to get temporal control over 5mC patterns required the fusion of split protein domains to constitutively active DNMTs. This approach does not allow the study of natural DNMTs and the usage of active DNMTs can lead to off-target effects. The aim of this work is to overcome the limitations of the current tools and to enable precise kinetics studies to dissect the role of DNA methylation in living cells. For this, the activity of DNMTs should be directly controlled by light irradiation since it offers a non-invasive and fast trigger. Using the advantages of genetic code expansion, a crucial cysteine in the active site of DNMTs should be replaced with a photocaged cysteine. This should lead to the expression of transiently inactive DNMTs and should allow the study of natural occurring DNMTs since only the unmodified cysteine is created after light-activation. Therefore, this approach offers much higher structural precision since the activity of DNMTs itself is controlled by replacement of a single amino acid. With this tool in hand, aberrant functions of DNMT mutations found in cancer, or transcriptome changes upon DNA methylation with defined starting points could be explored. To precisely edit 5mC patterns at user defined loci, light controlled DNMTs should be fused to TALEs. After ensuring the high fidelity of on-target methylation, the tunability of 5mC levels after targeted DNA methylation should be tested. This tool would allow the editing of 5mC patterns with control over the location, time, and level of 5mC and opens the possibility to explore the role of DNA methylation with previous unknown precision and controllability.

5. Results and Discussion

5.1. Characterization of DNMT3a3L and DMNB-Cysteine

For now, different tools using light were described to change the epigenetic environment either on a global or gene level. Yet, in these cases researchers used either photocaged 5-Aza[218] or the protein pair CRY2-CIB1 to recruit constitutively active DNMTs[170,309]. However, it is better to develop a tool to control the activity of DNMTs directly with the least structural interference as possible. In this study, a previously described fusion construct of DNMT3a and DNMT3L was used to 5-methylate cytosines.[160] As described in 3.2.2, DNMT3L serves thereby as an activator protein of DNMT3a by direct interaction of three α-helices[102,148,310] (Figure 15B) and therefore this fusion construct was chosen. The catalytic domain of human DNMT3a (amino acids 612 – 912) and amino acids 208 – 421 of mouse DNMT3L were fused with a 28-amino acid long linker, a N-terminal nuclear localization sequence and a C-terminal HA-tag (Figure 15A and Figure 62 (page 139), see 8.2 for full protein sequence). With genetic code expansion it is possible to incorporate ncAAs at user defined positions of proteins. As described in 3.4.1, the naturally occurring stop codon TAG is usually used as incorporation site using an orthogonal tRNA/aaRS pair recognizing the amber codon as the tRNAs anticodon as well as the ncAA. To get complete control over the DNMTs activity, it is necessary to exchange an amino acid crucial for the catalytic reaction. To choose the correct amino acid, the function and structure of DNA methyltransferases was examined. The crystal structure of the heterodimer of DNMT3L and DNMT3a (catalytic domain) is shown in Figure 15B. DNMT3a has direct contact with the DNA and the cytosine of the bound CpG is flipped out of the DNA helix in a first step of DNA methylation (see “Z” in close-up of Figure 15B). This repositions the C6 atom of cytosine to face the cysteine 710 residue (C710) and N3 atom to face the glutamic acid 756 (E756) residue so that the methylation reaction can occur as described in 3.2.2. C710 is thereby crucial for catalytic activity and the addition of a caging group to the sulfur atom should result in inhibition of DNMTs. Since the inhibition should be reversible and the activation should be triggered in a least invasive way, the caging group was chosen to be cleavable by light (3.4.2). This gives the opportunity to control the DNMTs activity by changing just a single amino acid with least structural change to the protein and should not interfere with protein folding. Subsequent light irradiation should cleave off the caging group and restore an unmodified cysteine, leading to activation of the actual DNA-methylation reaction. Previously, a photocaged cysteine with DMNB as caging group was already used in the context of controlling the activity of Kir2.1.[281,299] Also, the necessary tRNA/aaRS pair (*E. coli* leucyl $tRNA_{CUA}^{Leu5}$/$LeuRS^{BH5\ T252A}$ pair; Figure 63, page 139) needed for

incorporating DMNB-Cys was already evolved in yeast in the context of DMNB-Serine incorporation[271] and was also used for incorporating DMNB-Cys.

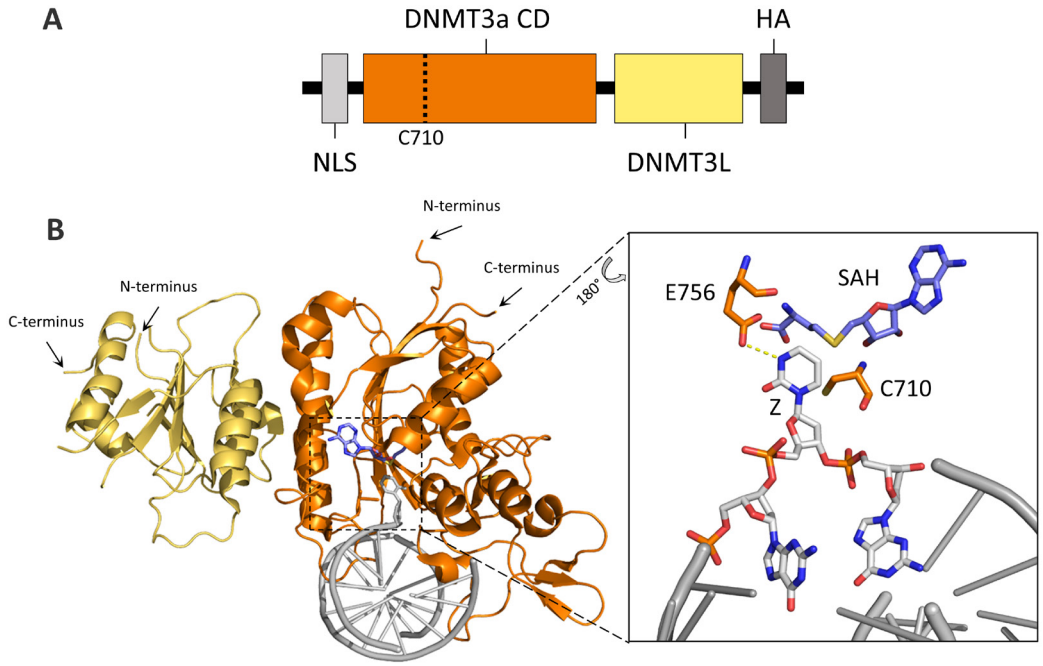


Figure 15: Crystal structure of DNMT3a3L with active site. A: Cartoon of DNMT3a3L construct. This design accounts to all used DNMT3a3L constructs in this thesis. For pcDNMT3a3L (explained later) the codon of C710 is mutated to 'TAG' for incorporation of DMNB-Cys via amber suppression. NLS: nuclear localization signal; HA: hemagglutinin tag. **B:** Crystal structure of catalytic domain of DNMT3a (orange, amino acids 628 – 912) and DNMT3L (yellow, amino acids 188 – 379) bound to DNA (grey), based on PDB 6F57[102]. Cytosine analog zebularine (Z)[311] and S-Adenosyl-L-homocysteine (SAH) were used to investigate the structure during the methylation reaction. Zebularine was used to trap C710 (bond is not shown in structure).

Therefore, this system should be used and tested in the context of controlling DNMT activity. After synthesizing DMNB-Cys (see Figure 69, page 156, and Figure 70, page 157 for NMR analysis), incorporation was tested in the easy-to-transfect HEK293T cell line. For this, an established mCherry-GFP fusion construct in which Tyr39 of GFP is mutated to the amber codon (Figure 16B) was used to confirm the fidelity of the tRNA/aaRS pair.[312] Recognition of other canonical amino acids by the synthetase would result in GFP expression in the absence of DMNB-Cys. Additionally, a flawed recognition of DMNB-Cys would result in an absence of GFP expression even in the presence of DMNB-Cys. The mCherry-GFP^{Y39amber} construct (further called 'transfection control') was expressed in HEK293T cells together with the $tRNA_{CUA}^{Leu5}$/LeuRS$^{BH5\ T252A}$ pair (throughout the works in this thesis, this pair was always co-transfected if photocaged DNMT constructs were used) in the presence or absence of 0.5 mM DMNB-Cys, then, cells were analyzed 24 h later by fluorescence microscopy (Figure 16C).

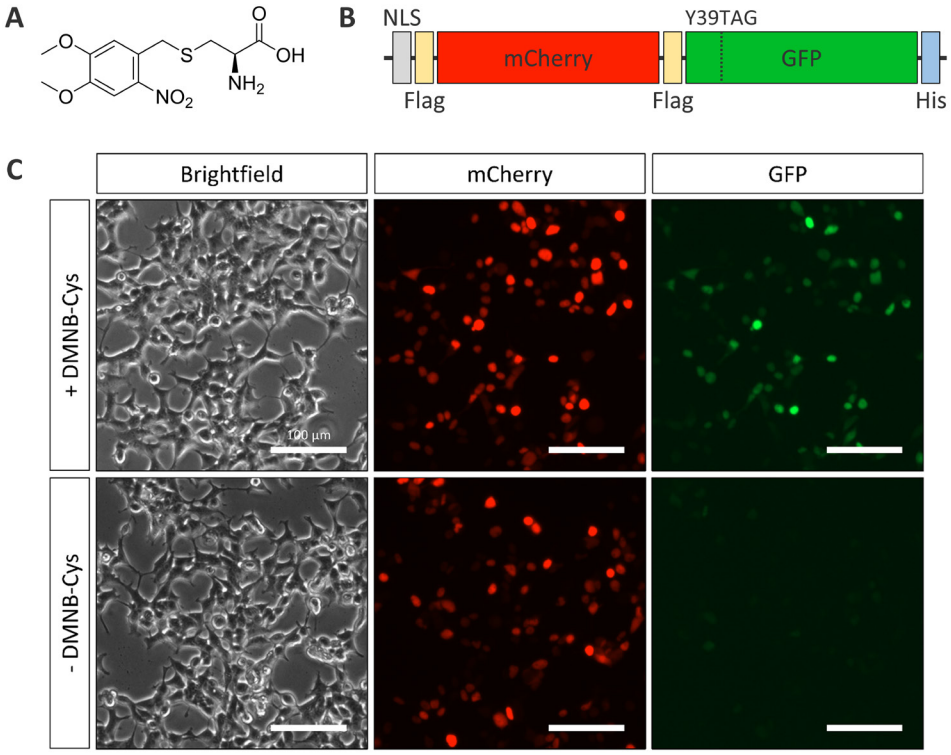
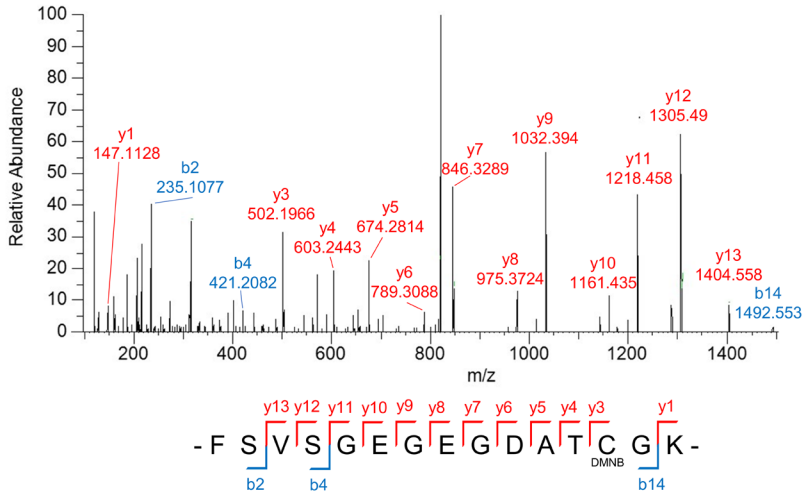


Figure 16: Selective expression of the transfection control in HEK293T cells with DMNB-Cys. **A:** Chemical structure of DMNB-Cys. **B:** Cartoon of the used transfection control mCherry-GFP^{Y39amber} with nuclear localization signal (NLS), Flag-Tags, and His-Tag. **C:** mCherry and GFP expression in p1660/p1169 transfected HEK293T cells in the presence or absence of 0.5 mM DMNB-Cys. Scale bar is 100 µm. Modified from Wolffgramm et al.[103]

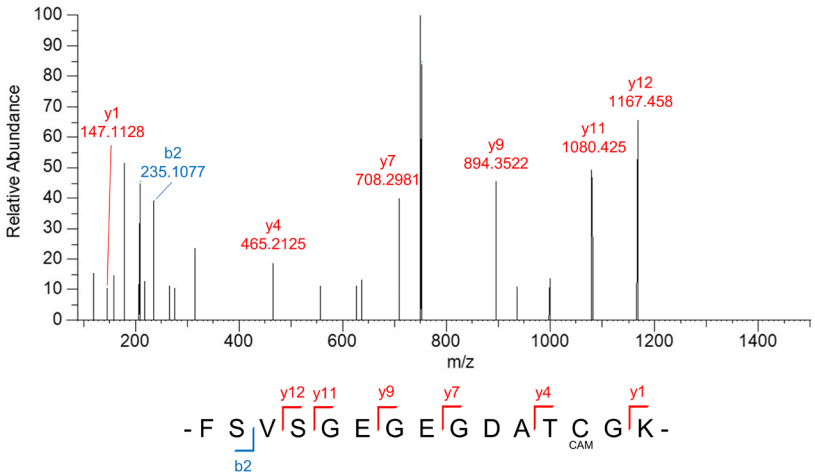
Very high transfection efficiencies and expression levels were observed for mCherry both with and w/o DMNB-Cys. Moreover, the expected high fidelity of the tRNA/aaRS pair was confirmed since high GFP expression was only observed in the presence of DMNB-Cys. Further analysis was conducted by flow cytometry measurement of GFP-positive cells (Figure 38, page 113). 13.96 % of single cells showed GFP-signal above a certain threshold in the presence of DMNB-Cys whereas only 0.97 % GFP-positive cells were measured in the absence of DMNB-Cys. After demonstrating the successful full-length expression of GFP, it was tested if the DMNB-caging group was indeed attached to the cysteine prior, and only cleaved off after light irradiation. For this, HEK293T cells were transfected again with the transfection control in the presence of DMNB-Cys and irradiated with 5 min of light (365 nm, see 7.2.1.21 and 7.2.2.3) 24 h later which is expected to lead to the cleavage of DMNB-Cys (Figure 19A, page 40). To verify whether light irradiation really restored the unmodified cysteine in living cells, peptide masses covering the DMNB-Cys were measured by electrospray ionization (ESI)-MS/MS. Whole cell lysate was applied to an SDS-PAGE and bands with the expected size of the transfection control were cut out. Proteins were extracted, its cysteines modified by carbamidomethylation (CAM), and finally digested by trypsin before analysis by ESI-MS/MS (7.2.1.21). After trypsin digestion of the transfection control, a 15-amino-acid long peptide (-FSVSGEGEGDATCGK-) was expected to form

which covers the initial Y39amber codon where DMNB-Cys was expected to be incorporated. The theoretical masses of peptides with caged cysteine (expected w/o light irradiation) and decaged cysteine (expected after light irradiation) were both detected (Figure 17C), proving the faithful incorporation of DMNB-Cys at the amber codon and decaging after light irradiation.

A



B



C

Sample	Cysteine modification	Theoretical mass	Measured mass
Caged w/o light	DMNB	1639.6660	1639.6637
Decaged with light	CAM	1501.6343	1501.6373

Figure 17: ESI-MS/MS spectra of GFP peptide. Detected fragments of the peptide -FSVSGEGEGDATCGK- from the transfection control mCherry-GFP^{Y39amber} covering the amber site used for incorporation of DMNB-Cys in HEK293T cells. **A:** Spectrum of the photocaged (DMNB) peptide w/o light. **B:** Spectrum of the decaged and carbamidomethylated cysteine (CAM) with light. **C:** Comparison of theoretical $[M+H]^{2+}$ and measured masses. Modified from Wolffgramm et al.[103]

Additionally, multiple y- and b-fragments of this peptide were detected with their theoretical masses (Figure 17A and B; Table 25, page 143, and Table 26, page 144). The signals for

the peptide with decaged cysteine after light irradiation though was much weaker than the signals for the peptide w/o light irradiation (Table 24, page 143). Also, even after light irradiation the peptide with caged DMNB-Cys was detected with better signals (Figure 39, page 114, and Table 27, page 144) showing that not all DMNB-Cys was decaged 5 min after light irradiation. Nevertheless, since the decaged cysteine was only detected after light irradiation (Table 24, page 143), it was proven that the caging group was stable, and it should guarantee the inhibition of DNMTs prior to light-activation. Also, the decaged fragment gave the first hint that the cleavage of DMNB with light was possible in living cells. After testing the incorporation and cleavage of DMNB-Cys with the transfection control, the same was repeated for the photocaged DNMT3a3L^{C710amber} construct (Figure 15A, further called 'pcDNMT3a3L'). HEK293T cells were transfected with pcDNMT3a3L in the presence of 0.05 mM DMNB-Cys for 24 h and then 5 min light was applied (no light for the control). Whole lysate was again applied to an SDS-PAGE and bands covering the expected DNMT3a3L size were cut out and analyzed. It was not possible to detect the peptide covering C710 and therefore no statement could be made regarding the incorporation and cleavage of DMNB-Cys. However, multiple peptides downstream of C710 were detected both with and w/o light (see 8.6.2), showing that the incorporation of DMNB-Cys was probably successful and resulted in full length expression. To verify that expression was indeed only in the presence of DMNB-Cys, a western blot analysis was done.

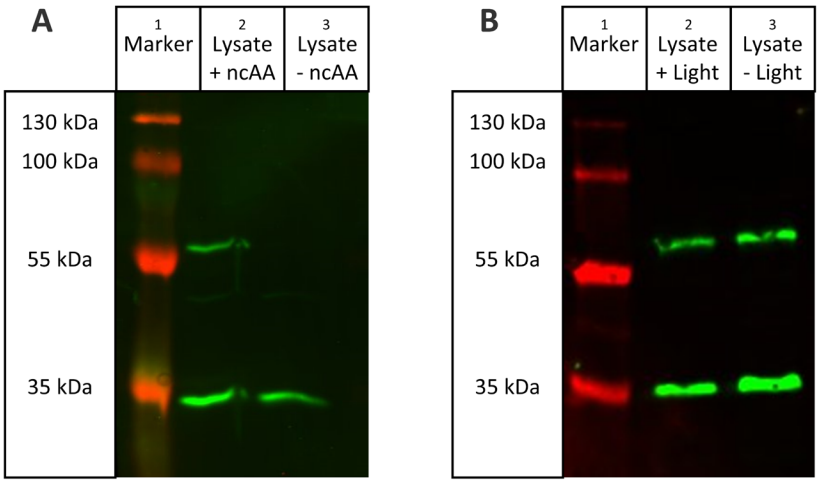


Figure 18: Western Blot of pcDNMT3a3L in HEK293T cells. A: Lysates from pcDNMT3a3L transfected HEK293T cells in the presence or absence of 0.5 mM DMNB-Cys were used for Western Blot analysis. **B:** Lysates from pcDNMT3a3L transfected and 5 min light irradiated HEK293T cells, and from the respective no light control were used for Western Blot analysis. 0.05 mM DMNB-Cys was added to the cell medium.
For both analyses, the membrane was stained using GAPDH-antibodies as internal control and HA-antibodies for pcDNMT3a3L detection, and subsequent DyLight800 labeled secondary antibodies. Expected sizes of 35.9 kDa for GAPDH and 64.9 kDa for pcDNMT3a3L were detected, proving full length expression of pcDNMT3a3L only in the presence of DMNB-Cys. As reference, the PageRuler™ Plus Prestained Protein Ladder was used. Modified from Wolffgramm et al.[103]

As before, whole cell lysate of HEK293T cells transfected with pcDNMT3a3L in the presence or absence of DMNB-Cys was applied to an SDS-PAGE and transferred to a PVDF membrane. Subsequent immunostaining of the C-terminal HA-tag of pcDNMT3a3L (Figure 15A) and endogenous housekeeping gene GAPDH as loading control was performed (Figure 18A; 7.2.1.19). Expression of full length pcDNMT3a3L with 64.9 kDa was only observed in the presence of DMNB-Cys (Figure 18A lane 2) whereas no band was detected in the absence of DMNB-Cys (Figure 18A lane 3). The GAPDH loading control though was positive in both cases. This again showed the high fidelity of the used tRNA/aaRS pair regarding the recognition of DMNB-Cys. It was also tested if light irradiation itself led to protein degradation or a change in expression levels of pcDNMT3a3L. Since for the ESI-MS/MS analysis only bands with an expected size of ~64.9 kDa were cut out and digested, no fragmentation would have been noticed. For western blot analysis though an extra band would appear after partial fragmentation since the C-terminal fragment with its HA tag would be also stained. As shown in Figure 18B, only the two expected bands for full length pcDNMT3a3L and GAPDH were detected for the light and no light samples. The same intensities of both pcDNMT3a3L bands showed that no degradation or change in expression level was observable.

After showing the correct incorporation of DMNB-Cys at the crucial C710 of pcDNMT3a3L, it was necessary to test the decaging kinetics of DMNB-Cys with the used light source to identify a suitable time window for tests in living cells. It was important to use the maximum time possible to achieve high decaging ratios with least cell stress. After light irradiation, the DMNB caging group is cleaved off leaving a nitrosoaldehyde and the unmodified cysteine (Figure 19A, and 3.4.2). This reaction can be monitored by measuring the absorption spectrum of the solution since the cleavage leads to a decreased absorbance at 365 nm and an increased absorbance at 410 nm.[313] For this, 1 mM DMNB-Cys was dissolved in DPBS and irradiated with light for up to 30 min. At different time points, small aliquots were taken, and the absorption spectrum was measured (Figure 19B). Clear decrease of absorption at 365 nm and increase at 410 nm was observed (Figure 40, page 114) up to 30 min. The decaging process though was slowing down after 5 min. As shown in Figure 19C, an almost linear decrease was observed for 365 nm absorption for the first 5 min after light irradiation followed by a slower decrease up to 10 min. For 410 nm though the increase of absorbance was almost linear up to 10 min after light. On the other hand, no change in absorption was detected for the control w/o light (Figure 19B-D). Although a large fraction seemed to be already decaged after 5 min, DMNB-Cys incorporated in cells was not fully decaged at this time as already observed in the ESI-MS/MS analysis (Figure 17) and longer light exposure times were necessary to achieve higher yields.

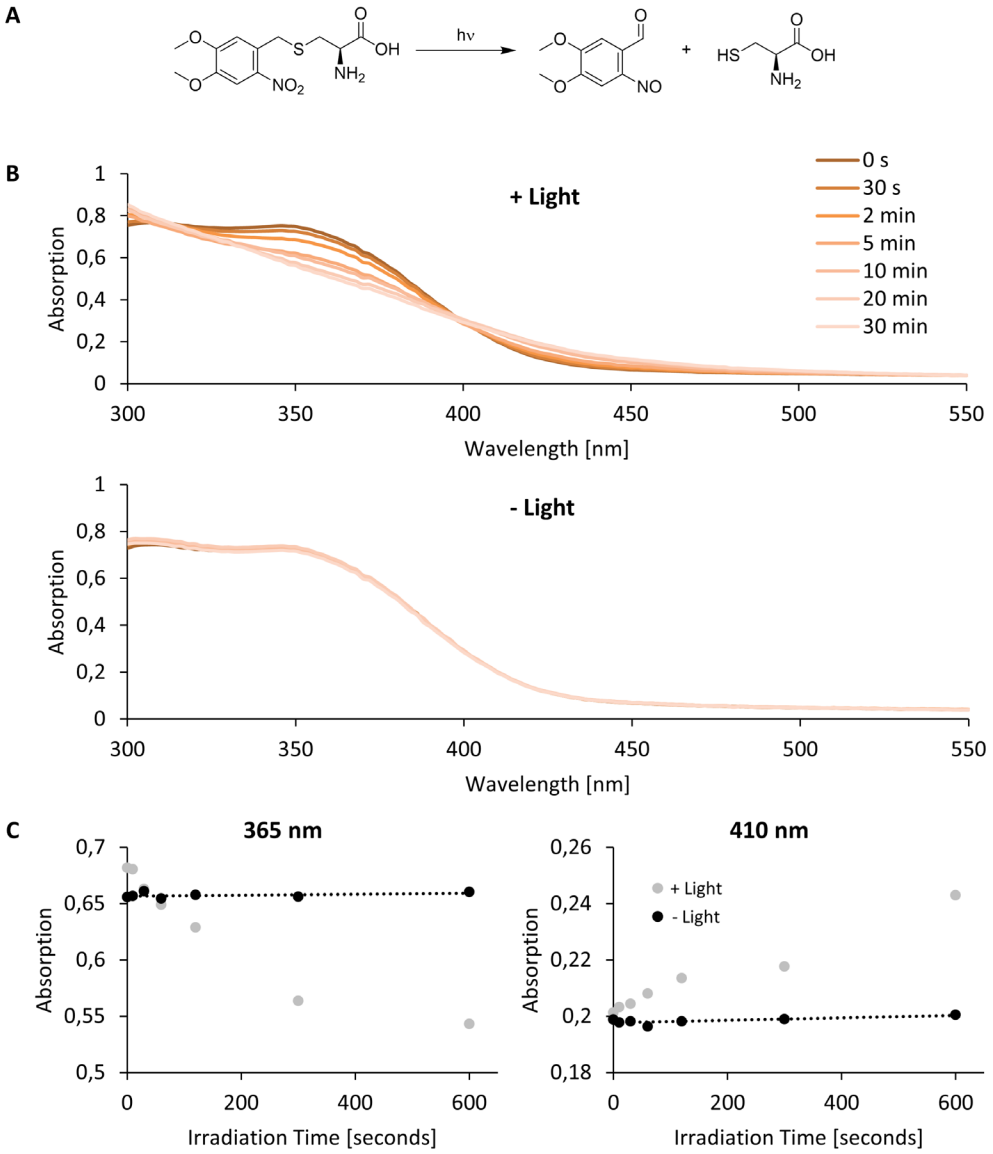


Figure 19: Absorption of DMNB-Cys from 300 nm to 550 nm. 1 mM DMNB-Cys in DPBS was irradiated differently long with 365 nm light. **A:** Decaging reaction of DMNB-Cys after light irradiation. **B:** Absorption spectra of samples taken after different times with or w/o light irradiation. **C:** Absorption at 365 nm and 410 nm after up to 10 min light irradiation and its respective controls w/o light. Experiment was done as a single replicate. Modified from Wolffgramm et al.[103]

To test if living cells can endure prolonged irradiation, HEK293T cells were irradiated with 0 min, 1 min, 3 min, 5 min and 10 min light and their cell morphology was checked by microscopy before light, 6 h after light, and 24 h after light (Figure 41, page 115). With up to 5 min light irradiation cells reached maximum confluency after 24 h whereas cells grew slower after 10 min light whereby the cell morphology stayed normal. Nevertheless, the increased cleavage from 5 min to 10 min (Figure 40, page 114) was not enough to justify a longer light irradiation in view of the increased stress for the cells. Therefore, 5 min light irradiation was used for cell experiments conducted for this thesis whereby longer exposure times with up to 10 min might be justifiable to use in future experiments if higher methylation effects should be tested.

5.2. Global Methylation

5.2.1. Proof of Concept

For now, it was decided to photocage the crucial C710 with a DMNB-caging group in the fusion construct DNMT3a3L. The precise incorporation of such amino acid was demonstrated in HEK293T cells, and an ideal light irradiation time was set to 5 min to get the highest decaging level with least cell stress. In the next step, a proper cell system must be established to validate DNA methylation after light-activation. Since HEK293T cells were already used in all previous experiments and offered good transfection efficiencies and incorporation of DMNB-Cys, these cells were used to test global DNA methylation. To reveal the maximum possible DNA methylation level, cells were transfected with wild type DNMT3a3L and global 5mC levels were analyzed by flow cytometry measurements of 5mC immunostaining 24 h later (Figure 20A). The overall signal however was not shifted to higher values for the cells transfected with DNMT3a3L compared to untransfected wild type (WT) HEK293T cells which could be explained with a high 5mC background in this cell line. To test this hypothesis, the endogenous DNMTs in HEK293T were inhibited by 5-Aza, which should result in lower 5mC levels (see 3.3.1). Since high 5-Aza concentrations are toxic to cells, different concentrations were tested to achieve the maximal possible effect. In a comparable study with 5-azacytidin in HEK293 cells, 2 µM final concentration was used for 3 days and low 5mC levels were observed on the fifth day.[314] However, here the growth of HEK293T cells was already diminished with 0.8 µM and 1 µM 5-Aza concentration after 5 days (Figure 43, page 117). Though, at a concentration of 0.6 µM, cell growth only showed a slight reduction for the first 5 days compared to untreated cells. Therefore, this concentration was used in the following studies. As expected, a slightly lower 5mC level was detected in cells treated with 5-Aza for 6 days compared to WT HEK293T cells (Figure 20B). To test the methylation with DNMT3a3L, 5-Aza was removed from the cell culture medium 24 h before transfection to minimize the inhibition effect on the transfected construct. Low 5-Aza concentrations in the DNA should not lead to inhibition of the overexpressed DNMT3a3L, but endogenous DNMT would be degraded which should affect DNA methylation. Cells were transfected with DNMT3a3L and global 5mC levels were analyzed 24 h later as before (Figure 20C). Compared to HEK293T cells w/o 5-Aza treatment, the effect was slightly higher but still trivial. The assay was repeated with pcDNMT3a3L and 5 min light irradiation 24 h after transfection. 24 h after light, the global 5mC level was minimally higher for the light sample compared to the control sample w/o light (Figure 20C). Limited methylation effects by pcDNMT3a3L were expected due to its lower expression levels compared to wild type DNMT3a3L which is caused by the

competition with release factors that also recognize the amber stop codon besides the $tRNA_{CUA}^{Leu5}$/$LeuRS^{BH5\ T252A}$ pair during translation. Nevertheless, the overall effect shown for pcDNMT3a3L was too low to judge if DNA was indeed methylated after light-activation.

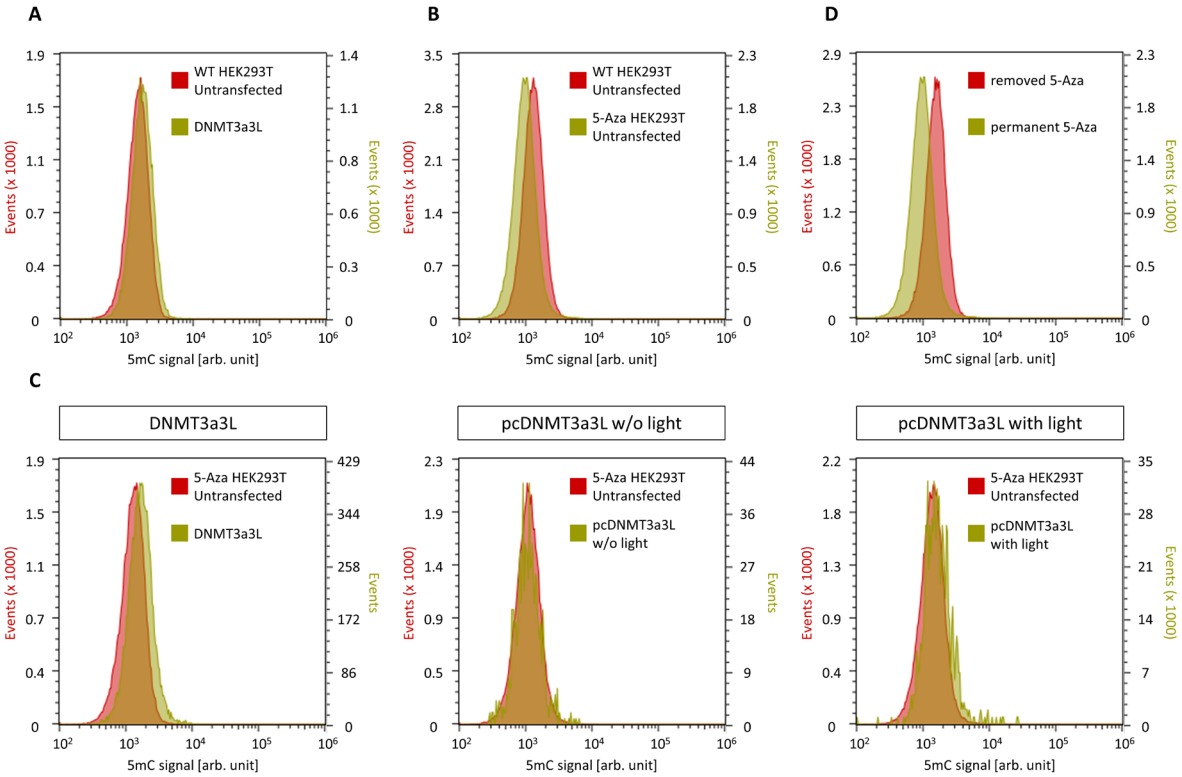


Figure 20: Histograms of global methylation in WT and 5-Aza HEK293T cells. A: WT HEK293T cells were transfected with DNMT3a3L (p1848) and analyzed 24 h after transfection by immunostaining. **B:** Comparison of basal 5mC levels by immunostaining in WT HEK293T cells and cells treated with 0.6 µM 5-Aza for 6 days. **C:** Global DNA-methylation by either DNMT3a3L (p1848), or pcDNMT3a3L (p1849) with or w/o light irradiation in 5-Aza HEK293T cells. Cells were cultivated 72 h with 0.6 µM 5-Aza and 24 h w/o 5-Aza prior to transfection. Analysis was done by immunostaining 24 h after transfection for DNMT3a3L. For pcDNMT3a3L, 5 min light irradiation was applied 24 h after transfection and cells were immunostained after additional 24 h. **D:** Basal 5mC levels in HEK293T cells treated with 0.6 µM 5-Aza for 6 days ("permanent 5-Aza") and cells grown in presence of 0.6 µM 5-Aza for 3 days and additional 3 days in absence of 5-Aza ("removed 5-Aza"). Immunostaining was done with HA- and 5mC-antibodies. Secondary antibodies were labeled with FITC (for HA) and AF405 (for 5mC). For all experiments, cells with a HA-signal above 20000 [arb. unit] were defined as DNMT3a3L- or pcDNMT3a3L-positive. Shown are the 5mC-signal histograms of the respective cell populations. For additional replicates see Figure 44, page 118.

The assay was performed based on the assumption that background DNA methylation is not restored by endogenous DNMTs during the time of the experiment. To test this assumption, 5mC levels of HEK293T cells treated with 5-Aza for 3 days and cultured w/o 5-Aza for another 3 days were compared with cells treated continuously with 5-Aza for 6 days (Figure 20D). Contrary to the expectations, DNA methylation fully recovered during the time w/o 5-Aza treatment, explaining the trivial difference between untransfected and transfected cells in Figure 20C. Overall, the usage of 5-Aza revealed multiple drawbacks, including the fast recovery of 5mC background, small methylation effects by DNMT3a3L in

the presence of 5-Aza, poor cell survival, and reduced transfection efficiency (data not shown). Since low 5mC background is necessary for the assay and since it was not possible to reduce such background in HEK293T cells, the cell line was switched.

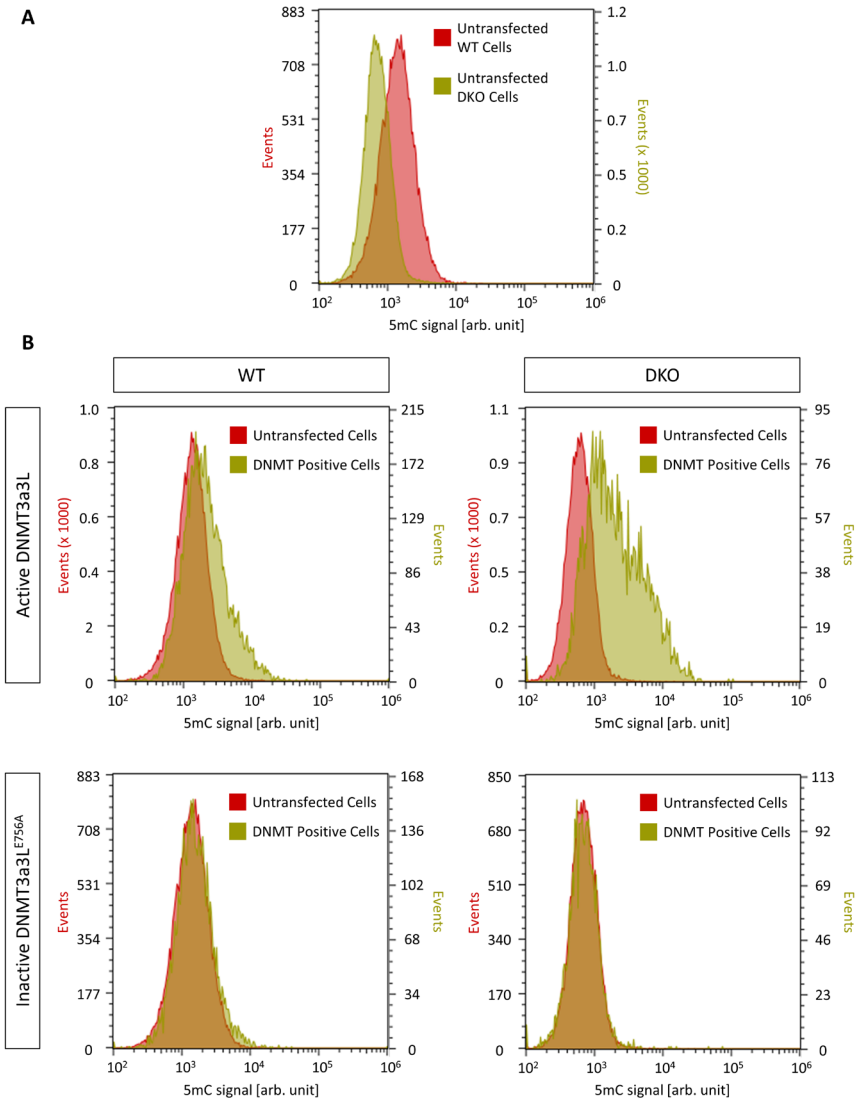


Figure 21: Global methylation in WT and DKO HCT116 cells. A: Histograms of 5mC immunostainings of untransfected WT and DKO HCT116 cells showing the basal 5mC levels. **B:** Global methylation with DNMT3a3L and inactive DNMT3a3L^{E756A} in WT and DKO HCT116 cells 48 h after transfection. Cells were immunostained with HA- and 5mC-antibodies. Cells with a minimum HA-signal of 20000 [arb. unit] were defined as DNMT3a3L-positive. For additional replicates see Figure 45, page 119.

HCT116 cells, a human colorectal carcinoma cell line, were available as WT and double knock out (DKO) cells in which DNMT1 (Δexons3-5/Δexons3-5) and DNMT3B (-/-) were knocked out. This double knock out results in 95 % reduction of 5mC level compared to the WT cells[315] and was confirmed by immunostaining and flow cytometry analysis (Figure 21A). Also, no cell morphology changes were observed after 5 min light irradiation (Figure 46, page 120). With such cell lines in hand, global methylation by DNMT3a3L was tested again in both WT and DKO HCT116 cells (Figure 21B). Additionally, an inactive

DNMT3a3L^{E756A} construct was used as the control in which the mutation of the key glutamic acid needed for stabilization (see 3.2.2) to alanine results in the loss of activity.[128] A clear increase in 5mC immunostaining levels 48 h after transfection was observed in the WT HCT116 cells. This effect was even stronger in the DKO HCT116 cells whereas 5mC levels did not change for the inactive control. Therefore, DKO HCT116 cells offered a good methylation range and could be used to study the light-activation of pcDNMT3a3L. A drawback though was the much lower transfection efficiencies observed with various transfection reagents (Figure 47 - Figure 50, pages 121 - 124). Only ~1.5 % of DKO HCT116 cells showed pcDNMT3a3L expression 48 h after transfection when analyzed by immunostaining of the C-terminal HA-tag, which is roughly 50 % of the expression level observed with DNMT3a3L (Figure 22A). Nonetheless, only negligible HA levels were observed in the absence of DMNB-Cys showing that this cell line also has high fidelities of ncAA recognition and amber suppression. It was even possible to stop the expression of pcDNMT3a3L by removing DMNB-Cys from the medium 24 h after transfection (Figure 22B). First, it shows clearly that 48 h expression time is needed to reach proper pcDNMT3a3L levels, and second, it proves that DMNB-Cys is removed efficiently from cells before light irradiation. This reduces the potential toxicity from nitrosoaldehyde in the medium and gives the opportunity to work with lower expression levels when desired. The long expression time needed however highlights the benefits of light-activatable DNMT constructs. Since pcDNMT3a3L stays inactive before light irradiation, the prolonged expression time does not lead to off-target activation, which would be an obstacle when using constitutively active DNMTs. This also makes the pcDNMT3a3L system independent of cell lines with good transfection and expression efficiencies. To test the light-activation of pcDNMT3a3L, DKO HCT116 cells were irradiated with light for 5 min 48 h after transfection and analyzed by 5mC immunostaining 24 h later (Figure 22C, populations were gated according to Figure 51, page 125). As desired, an increase of global 5mC level only after light irradiation proved the activation of pcDNMT3a3L with light, providing for the first time the precise and direct control of DNMT3a activity in living cells. Since no 5mC increase was observed for the inactive mutant after light irradiation, it was proven that light itself does not influence DNA methylation and that the observed effect was indeed based on activated pcDNMT3a3L. Moreover, it was possible to analyze the effect from different expression levels of pcDNMT3a3L. For this, cell populations were gated into low and high expression based on their HA-signal, and the 5mC levels were analyzed (Figure 52, page 126). Thereby, higher expression led to higher DNA methylation which allows expression level-based studies.

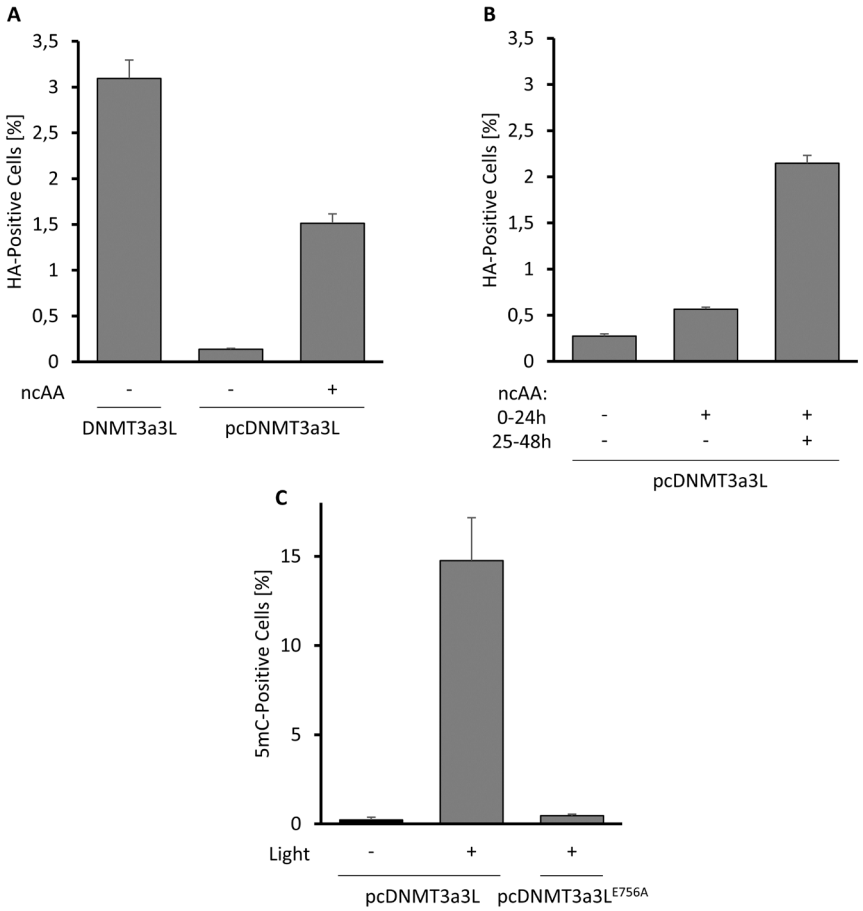


Figure 22: Expression of DNMT3a3L and pcDNMT3a3L and light-activation in DKO HCT116 cells. A: Cells were transfected with either p1848 (DNMT3a3L), or p1849 (pcDNMT3a3L)/p1169 with or w/o 0.05 mM DMNB-Cys ("ncAA") and analyzed 48 h later by immunostaining. **B:** DKO HCT116 cells were transfected with p1849/p1169 with or w/o 0.05 mM DMNB-Cys ("ncAA") for 48 h, or 24 h in the presence of ncAA + 24 h w/o ncAA. Immunostaining for **A** and **B** was done with HA-antibodies and FITC-labeled secondary antibodies. Cells with a FITC-signal above 20000 [arb. unit] were considered as HA-positive and are shown as share of all detected single cells. Error bars show standard deviations from three independent biological replicates. Since the experiments and immunostainings in **A** and **B** were done independently, the absolute numbers differ slightly. **C:** Global methylation with pcDNMT3a3L 24 h after 5 min light-activation. Before, cells were transfected with pcDNMT3a3L or inactive pcDNMT3a3L^{E756A} in presence of 0.05 mM DMNB-Cys for 48 h in DKO HCT116 cells. Cells were immunostained with HA- and 5mC- antibodies and gates were set according Figure 51.
Modified from Wolffgramm et al.[103]

5.2.2. Studies of Cancer-Related Mutations

DNMT3a is very frequently mutated in hematological malignancies, specifically acute myeloid leukemia (AML), but it remains a challenge to study the effects of these mutations on the catalytic activity alone.[316] Since the mutation can also alter upstream processes like protein expression and folding, it is very difficult to uncouple the catalytic activity from these processes when mutants are expressed in their constitutively active form. With the here described light-activatable pcDNMT3a3L, these challenges can be overcome by expression of the transiently inactive form and simultaneous activation when full and comparable

expression levels are reached. Ten different mutations (Table 1, amino acids are shown in bold in Figure 23) were chosen, which are either highly abundant in AML, have reported in vitro or in vivo kinetics data, or are mutated at interesting positions within the crystal structure where interactions to the DNA, cofactor, or DNMT3L are shown (Figure 23).

Table 1: COSMIC database data of analyzed DNMT3a mutations.
Data is retrieved from "COSMIC: the Catalogue Of Somatic Mutations In Cancer"[317] and is as of 31st December 2020.

Mutation	COSMIC Database Count	Genomic Mutation ID
S714C	28	COSV53039274
R736H	21	COSV53036512
R771Q	8	COSV53037113
R771G	1	COSV53073691
T835M	5	COSV53041391
R836W	1	COSV53059547
R836A	0	-
R882C	397	COSV53036332
R882H	905	COSV53036153
W893S	13	COSV53041651

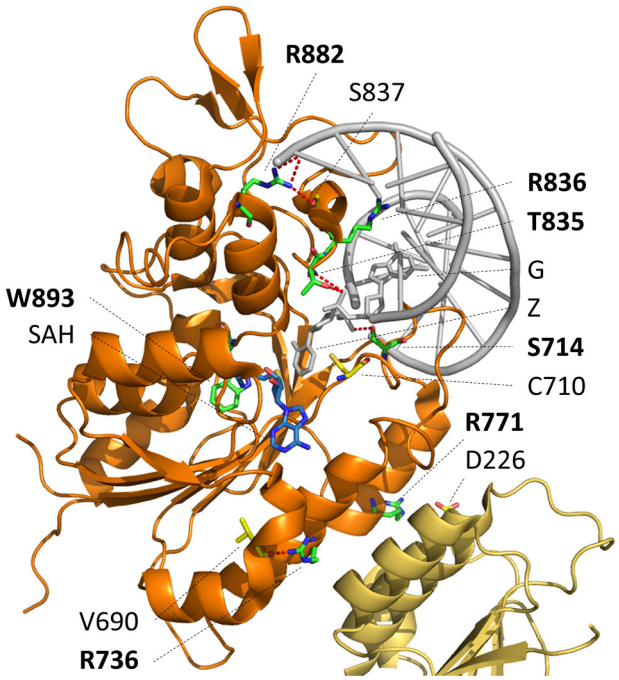


Figure 23: Location of examined mutated amino acids in DNMT3aCD. Crystal structure (PDB 6F57[102]) of DNMT3aCD (orange), part of DNMT3L (yellow) and DNA (grey). Amino acids chosen for mutation studies are shown in bold. Z: cytosine analog zebularine[311], SAH: S-Adenosyl-L-homocysteine. Modified from Wolffgramm et al.[103]

Four amino acids (S714, T835, R836, and R882) show either direct or water mediated hydrogen bonds to the DNA, W893 interacts with the cofactor SAH, and two arginine residues, R736 and R771, are localized at the interface to DNMT3L. With this, all relevant parts of the DNMT3a3L construct were covered and most of them are sitting near the active site (see C710 and Z in Figure 23). First, all ten mutants were cloned into the wild type

DNMT3a3L construct w/o the C710amber mutation, and their overall activities in DKO HCT116 cells were examined. Global 5mC levels were analyzed by immunostaining and flow cytometry measurement (Figure 51, page 125) 48 h after transfection as described before and normalized to WT DNMT3a3L (Figure 24).

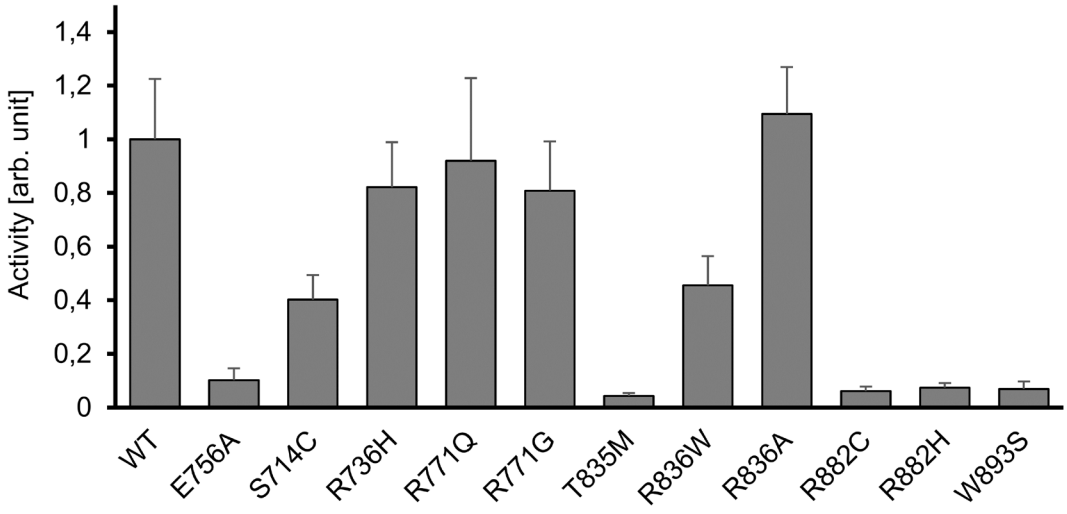


Figure 24: Activity of DNMT3a3L-mutants in DKO HCT116 cells. Cells were analyzed 48 h after transfection with wild type DNMT3a3L (WT) or its mutants by immunostaining of HA-tagged DNMT3a3L and 5mC. Data was analyzed according to Figure 51, page 125. Since the collection of the data of WT DNMT3a3L and DNMT3a3L^{S714C} failed in the first experiment, the shown data was collected in an independent experiment together with DNMT3a3L^{R736H} (see Figure 53, page 126). The data of other mutants was normalized to the activity of DNMT3a3L^{R736H} to get comparable 5mC data from both experiments. Error bars show standard deviations from three independent biological replicates.

Surprisingly, multiple mutants showed almost no DNA methylation which was even lower than the previous used 'inactive' mutant DNMT3a3L^{E756A}. Among them were R882C and R882H – by far the most frequent mutations in AML (Table 1).[318] R882 is located near the interface of the DNMT3a-DNMT3a homodimer (not shown in Figure 23) building hydrogen bonds to the DNA phosphate backbone and amino acid S837 in the TRD loop.[102] However, conflicting results regarding the effect of the R882H mutations have been reported. While it was initially assumed to have a dominant-negative effect due to the formation of functionally deficient tetramers[319,320], other reports questioned such hypothesis and referred to other mechanisms in cells[321]. Additionally, different activities of these R882 mutants have been reported, ranging from WT-comparable activity in HEK293T cells[322], a slight reduction in vitro[323], to almost inactive in murine embryonic stem cells[319]. Moreover, a restoration of activity for DNMT3a^{R882H} by DNMT3L has been reported.[324] Latest structural analysis revealed that the orientation of the DNMT3a-DNMT3a interface and the TRD loop is changed by the R882H mutation, resulting in an altered recognition of CpG nucleotide context.[325] Since no activity was observed in the DKO HCT116 cells, these R882 mutations could not be used for further characterization. Another inactive mutant was T835M which

has been reported to be active in TF-1 cells[102]. T835 is located in the TRD loop and builds a hydrogen bond to the DNA backbone (Figure 23). W893 is located near the cofactor SAH and stabilizes the crucial E756 (not shown in Figure 23) via a hydrogen bond. In DKO HCT116 cells, the W893S substitution was also inactive, though different in vitro activities have been reported.[326] Two mutations, S714C and R836W, showed a reduction in activities to 40.3 % and 45.5 % compared to WT, respectively. S714 is a conserved amino acid of the catalytic loop and stabilizes the DNA via a hydrogen bond to the DNA backbone (Figure 23)[102,327], while R836 is located in the TRD loop and builds a water-mediated hydrogen bond to the DNA phosphate of the targeted CpG (not shown in Figure 23).[102] In both cases, mutation seems to weaken the enzyme-DNA interaction, which is in line with previous publications for R836W.[102] For S714C, a complete loss of function in mESC[326] has been reported whereas in vitro studies[327] showed a similar decreasing effect as observed here. Since both mutations seem to have similar effects in DKO HCT116 cells and since the second examined R836 mutation (R836A) showed high activity, only S714C was taken for further studies. The other three mutants (R736H, R771Q, and R771G) showed comparable activities to WT in which R736H and R771G were the least active mutants in this subset. It is noteworthy that no mutant showed much higher 5mC levels than WT, leaving the possibility that a plateau of DNA methylation was already reached 48 h after transfection. Since it was shown in Figure 22B that the expression 24 h after transfection was very low, it was not possible to measure the 5mC levels at this time point. This again shows the need to express transiently inactive enzymes which can be simultaneously activated. Together with S714C, the actual activities of these mutants were examined by light-activation. This enables new insights into the mutant's impact on DNMT3a catalytic activity alone, especially since the published in vitro data are not comparable to the in vivo activity observed here for mutants R882C, R882H, T835M and W893S. DKO HCT116 cells were transfected with the respective pcDNMT3a3L mutants containing the additional C710amber mutation in the presence of 0.05 mM DMNB-Cys for 48 h to achieve saturated protein expression. Then, enzymes were activated by 5 min light irradiation and global 5mC levels were analyzed by immunostaining and flow cytometry measurement as described before. Expression levels were additional monitored by immunostaining of the C-terminal HA tag of pcDNMT3a3L. As shown in Figure 25, all expression levels were comparable at the time of analysis. As mentioned before, no further expression was seen after removing DMNB-Cys from the cell culture medium (Figure 22B) so that the expression level at the time of light irradiation also had to be comparable. Additionally, since the controls w/o light irradiation showed similarly low 5mC levels for all mutants, it can be stated that the 5mC level at the time of light-activation were also comparable in the light-activated samples. Therefore, the observed differences in activity (Figure 25) were faithfully reflecting the influence of the mutation itself

and was not altered by other upstream processes or different expression and 5mC levels at the time of light-activation.

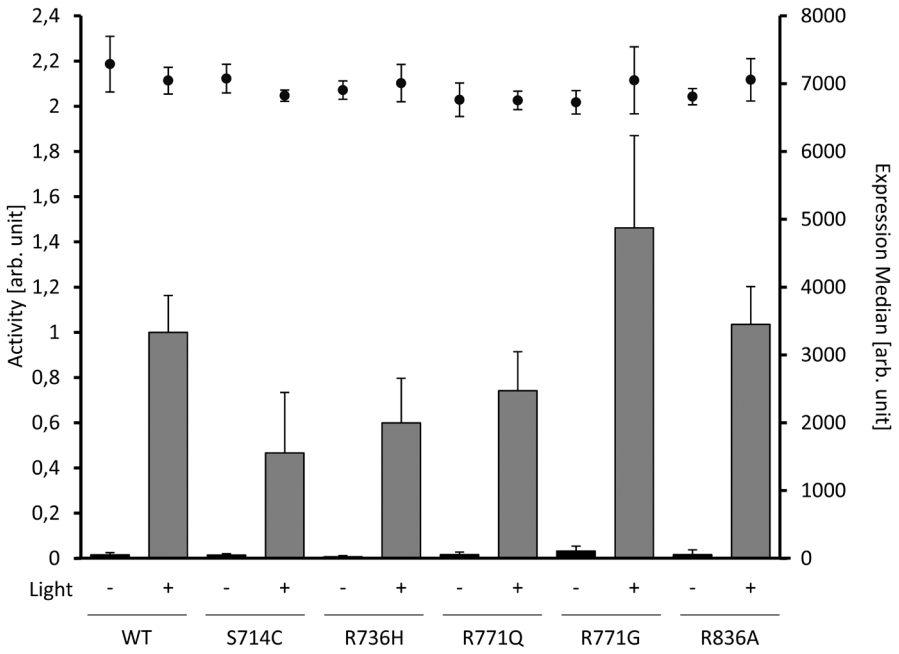


Figure 25: Comparison of active mutants of pcDNMT3a3L in DKO HCT116 cells. Cells were irradiated with 5 min light 48 h after transfection with pcDNMT3a3L (WT) or its mutants (no light for "- Light" controls). After another 24 h incubation, cells were analyzed by immunostaining of HA-tagged pcDNMT3a3L and 5mC. Data was analyzed according to Figure 51, page 125. Additionally, the median value of expression (staining of HA-tag with AF405) is shown. Error bars show standard deviations from three independent biological replicates. Modified from Wolffgramm et al.[103]

Activity of S714C was again reduced to ~46 % of WT which follows the observed effect in Figure 24. The next mutated amino acid R736 is located at the DNMT3a-DNMT3L interface with a stabilizing intra-helix hydrogen bond to V690 (Figure 23). Disruptions of the DNMT3a-DNMT3L interface can lead to wrong 5mC patterns.[328] R736H mutation in DNMT3a has different reported in vitro activities ranging from 2-fold decrease to 4-fold increase, whereas no significant change has been observed in a DNMT3a-DNMT3L tetramer.[326] It also showed much higher DNA methylation in a non-CpG context.[326] Here, the activity was reduced to 60 % of WT which was slightly less than observed in Figure 24 with 82.1 %. Another amino acid in the DNMT3a-DNMT3L interface was R771 which builds a salt bridge with D226 of DNMT3L (Figure 23). Interestingly, R771Q and R771G mutations have shown largely increased in vitro activity of 6-fold and 2-fold, respectively, given that R771G also methylates non-CpG contexts.[326] Surprisingly, R771Q was less active and R771G more active than WT after light-activation (Figure 25), while in Figure 24 R771G was slightly less active than R771Q. This severe change in activity might indicate a strong effect of the mutation on upstream processes which play a role in Figure 24. The upregulation by R771G was in accordance with the published in vitro data. However, additional studies must be

made to understand the role of R771Q mutation in DKO HCT116 cells. Since R771G was the only mutant that showed high 5mC level, it was proven that 5mC did not saturate 24 h after light irradiation for all other tested DNMT3a3L constructs. It remains unclear though if R771G reached such saturation and if the activity differences compared to WT might be even higher. The last studied mutation was R836A for which in vitro data of a similar DNMT3a3L fusion construct has been published.[102] The reported activity was slightly increased compared to WT and showed significant DNA methylation in a CpA context. Here, the activity after light-activation was similar with the activity measured in Figure 24 and in accordance with published data. This shows that the mutation apparently has no impact on upstream processes and in this case the in vitro observation was comparable to the in vivo observation. Methylated cytosines in a CpA context should be stained by the 5mC-antibody similarly as methylated cytosines in a CpG context. Since the overall 5mC level for this mutant did not change compared to WT and since the previously published data has shown increased CpA methylation, the CpG methylation is expected to be decreased to keep the overall 5mC level stable. With specific binders recognizing CpG or CpA methylation separately, the cytosine context dependency of the mutation could be studied. Furthermore, the light-activation of such mutant could be used to examine the overall effect of increased CpA methylation in cells.

Summarized, most studied mutations showed similar activities after light-activation compared to constitutively active mutants in DKO HCT116 cells (Figure 54, page 127), although it is difficult to compare between the two assays due to the different active times (48 h after transfection vs. 24 h after light irradiation). Only R771G showed a large change in activity and further studies are needed to unveil its impact. Though the complex roles of the studied mutations cannot be completely solved by the studies described here, they can provide additional insights into the in vivo activities that are uncoupled from upstream processes in DKO HCT116 cells. This may reveal a missing piece in solving the impact of these mutations on DNA methylation. Moreover, it showed that for R836A the published in vitro data and the here presented activity after light-activation was comparable using similar DNMT3a3L constructs.

5.3. Targeted Methylation

After deploying photocaged DNMT on a global level in DKO HCT116 cells, the tool should be applied to single genes. Epigenome editing of user-defined single loci gives great advantages in studying direct downstream effects of DNA methylation. Knowing the exact location and time of DNA methylation in living cells opens many possibilities, for example monitoring the kinetics of specific 5mC reader recruitment or gene silencing. For this, the pcDNMT3a3L construct was fused C-terminally to a TALE protein (Figure 26A). TALEs are programmable DNA-binding proteins which can be used to recruit effector domains to loci of interest (see 3.3.2). By directing DNMTs to a specific target gene, methylation of such gene is not only ensured but also background methylation of other genes is lowered. Since the construct is expressed and recruited to the target gene in an inactive form before light-activation, the targeting effect should be even more prominent compared to other published approaches with constitutively active DNMTs (see 3.3.1).[170,309]

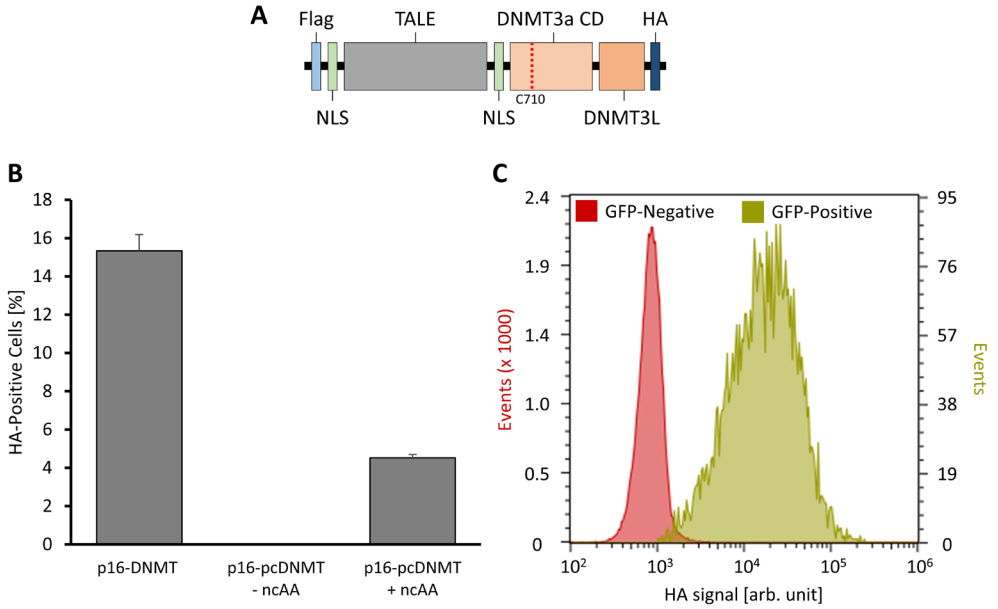


Figure 26: Expression of p16-DNMT and p16-pcDNMT in HEK293T cells.
A: Cartoon of TALE-DNMT fusion constructs. This design accounts to all used TALE constructs in this thesis. For TALE-pcDNMT the codon of C710 is mutated to 'TAG' for incorporation of DMNB-Cys via amber suppression. **B:** Expression levels of p16-DNMT, and p16-pcDNMT in the absence or presence of 1 mM DMNB-Cys analyzed by immunostaining 24 h after transfection. p16-pcDNMT is selectively expressed only in the presence of DMNB-Cys. **C:** Histograms of p16-pcDNMT expressing cells (staining of HA-tag with AF405) in GFP-negative or -positive cells. Both expressions are correlated to each other.
Cells were stained with anti-HA primary and AF405 labeled secondary antibodies. Cells with a HA- or GFP-signal over 20000 [arb. unit] were considered HA- or GFP-positive, respectively. Error bars show standard deviations from two independent biological replicates. Modified from Wolffgramm et al.[103]

First, the CDKN2A gene encoding p16, a cell cycle regulator associated with cellular senescence and age-related diseases[329], was targeted. This gene was already subject of

targeted methylation with constitutively active DNMT3a3L fused to a TALE protein in HeLa cells.[160] Starting with a known system gave the opportunity to establish this approach in HEK293T cells. This easy-to-transfect cell line was used for targeted methylation since it is only important that the target gene has low 5mC levels rather than on a global level. The TALE sequence (Table 17) was assembled into plasmids encoding DNMT3a3L (Figure 60, page 138) or DNMT3a3L^{C710amber} leading to fusion constructs named p16-DNMT (Figure 61, page 138) or p16-pcDNMT, respectively. Again, expression of the photocaged construct p16-pcDNMT was confirmed in the presence of DMNB-Cys by staining of the C-terminal HA-tag and subsequent flow cytometry analysis (Figure 26B). In accordance with the results in Figure 22A, no expression in the absence of DMNB-Cys but sufficient expression levels of p16-DNMT for proper analysis in the presence of DMNB-Cys were observed. The plasmids also encoded the previous used mCherry-GFP^{Y39amber} gene (Figure 61, page 138) for transfection and amber suppression control to circumvent the need to stain the HA-tag in fixed cells. The correlation of GFP and p16-pcDNMT expression was validated and shown in Figure 26C, where 98.95 ± 0.73 % of GFP-positive cells (above a certain GFP-threshold) also showed expression of p16-pcDNMT, while GFP-negative cells did not show any p16-pcDNMT expression. Due to this clear correlation, high GFP-signals can serve as an indicator for the expression of the TALE-pcDNMT construct and enabled sorting of GFP-positive cells without prior fixation. Additionally, it was also confirmed that light irradiation did not change the mCherry or GFP expression (Figure 42, page 116). To examine the maximal possible methylation, HEK293T cells were transfected with p16-DNMT or inactive p16-DNMT^{E756A} and the mCherry-positive cells were sorted 48 h later. Then, total gDNA was isolated, bisulfite converted, and the p16 locus was amplified by PCR (7.2.3.1). PCR amplicons were barcoded and sequenced by deep amplicon sequencing (Illumina platform). In total, 5mC levels of 16 CpGs with a distance of 18 to 227 nucleotides downstream to the TALE binding site (counting starts with 5-T of TALE binding site) were analyzed, while 5mC levels of untransfected HEK293T cells were used as control. A strong methylation effect by p16-DNMT was observed, especially for CpGs 42, 129, and 217 for which up to 52.42 % 5mC (average level) was reached (Figure 27). It is noticeable that small local maxima for these three CpGs were also observed in untransfected HEK293T cells, indicating a possible preference for such methylation pattern. Overall, the CDKN2A locus was a good target since only very low DNA methylation in untransfected cells were measured with CpG 217 showing the highest level of 5.93 % on average. The exact same methylation pattern was observed in cells treated with inactive p16-DNMT^{E756A} except for CpG 42 with a difference of 9.31 % compared to untransfected cells. Still, all monitored CpGs were well below the values reached by p16-DNMT, showing the great possibility to edit the epigenome landscape of single loci using TALE-fused

DNMTs. In addition, it was proven that the TALE-DNMT system also worked in HEK293T cells, and it gave the reliance that CpGs in a range of more than 200 nucleotides can be targeted.

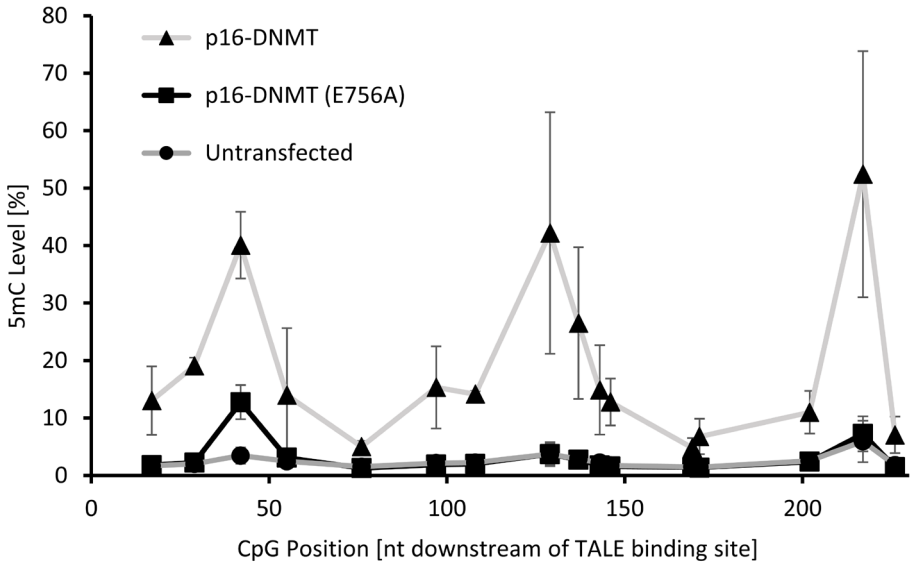


Figure 27: 5mC levels of p16 gene in p16-DNMT or p16-DNMT^{E756A} transfected HEK293T cells. Methylation levels of CpGs 18 to 227 nt downstream of the p16-DNMT binding site were analyzed by Illumina sequencing of bisulfite converted and amplified gDNA. Numbering starts with 5-T of TALE target site. Cells were analyzed 48 h after transfection with p16-DNMT or inactive p16-DNMT^{E756A}. Error bars show standard deviations from two independent biological replicates. Modified from Wolffgramm et al.[103]

In the same analysis run, the methylation by p16-pcDNMT 24 h after 5 min light irradiation was tested. Though the effect was much weaker, a similar methylation pattern with regard of the three highly methylated CpGs was observed (Figure 55, page 127, one replicate). The total methylation levels were difficult to compare since cells expressing p16-DNMT were analyzed 48 h after transfection and cells expressing p16-pcDNMT were analyzed 24 h after light irradiation. Longer incubation after light irradiation was not done since cell confluency would be too high and extra passaging might influence overall cell growth and activity, thus preventing any comparability. Though only one replicate of p16-pcDNMT transfected cells was analyzed, it gave the indication that targeted methylation with the light-activatable construct might be possible. Nevertheless, further studies were necessary and for these the target locus was switched to repetitive pericentromeric SatIII (see 3.1.1).[330] This locus also shows aberrant methylation in cancer[42,43] and due to its repetitive nature, multiple TALE-fusion proteins can bind simultaneously which might lead to stronger methylation effects. A TALE with a 20-base long recognition site (Table 17) was fused to pcDNMT3a3L ("SatIII-pcDNMT", based on Figure 61, page 138) similar to p16-pcDNMT. Also for this construct, proper expression in the presence of DMNB-Cys (Figure 56A, page 128) and a correlation between SatIII-pcDNMT and GFP expression (Figure 56B, page 128)

was observed. Additionally, it should be ensured that the expression level plateau was reached to maximize the methylation effect. HEK293T cells were transfected with SatIII-pcDNMT in the presence of DMNB-Cys with concentrations between 0.05 mM and 0.5 mM, and the expression level was analyzed by immunostaining of the C-terminal HA-tag 24 h later (Figure 28). Indeed, the maximal expression level was already reached in the presence of 0.05 mM DMNB-Cys with almost no visible difference for total transfected cells across different concentrations (Figure 28A). This also accounted for the average expression levels in these transfected cells (Figure 28B) showing that higher concentrations did not lead to higher numbers of transfected plasmids per cell and with it to higher protein expression levels.

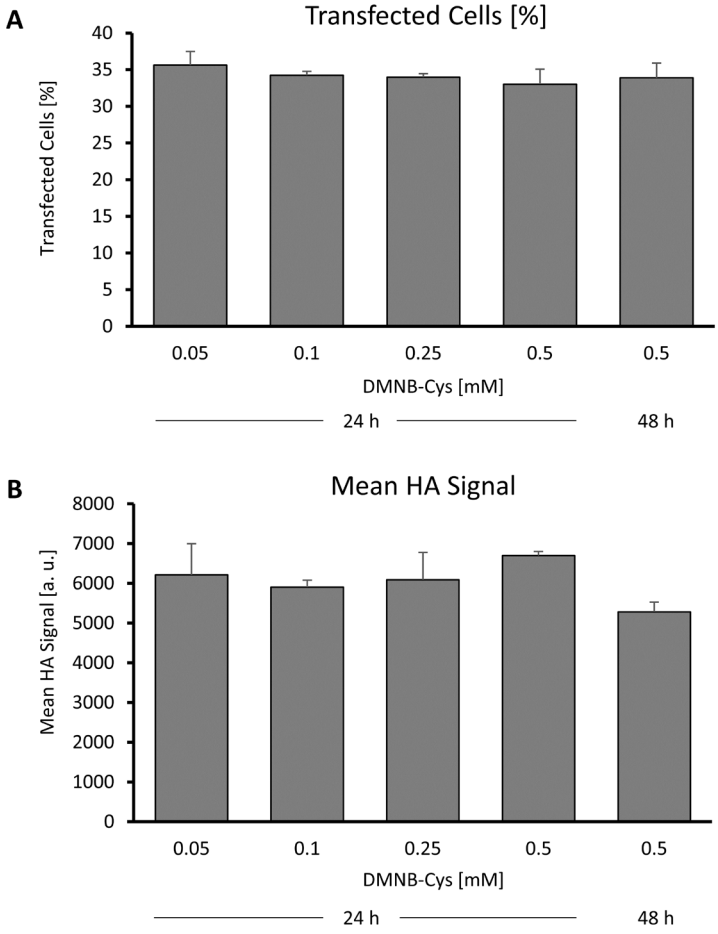


Figure 28: Expression levels of SatIII-pcDNMT in HEK293T cells with different DMNB-Cys concentrations. Cells were transfected with p1660/p1169 and analyzed by HA-immunostaining 24 h or 48 h after transfection. **A:** Share of HA-positive cells from all measured single cells. **B:** Mean HA-signal of all HA-positive cells (AF405 used for staining).
Modified from Wolffgramm et al.[103]

It was also tested if the 24 h time window after transfection was sufficient to reach saturation in expression. After incubating the cells for 48 h in the presence of 0.5 mM DMNB-Cys, no increase in expression was observed compared to 24 h incubation (Figure 28A and B),

proving that the expression has saturated at 24 h after expression in the presence of 0.05 mM DMNB-Cys. Since the SatIII locus is highly repetitive, it was expected that the bound SatIII-pcDNMTs can be visualized by HA-immunostaining and with that the correct and selective localization of the construct can be ensured. The SatIII locus plays a crucial role in the heat shock response in human cells (see 3.1.1).[37,40,331] After treating cells with a heat shock, the transcription factor HSF1 is recruited to SatIII regions and subsequent nuclear stress bodies are formed.[40,41]

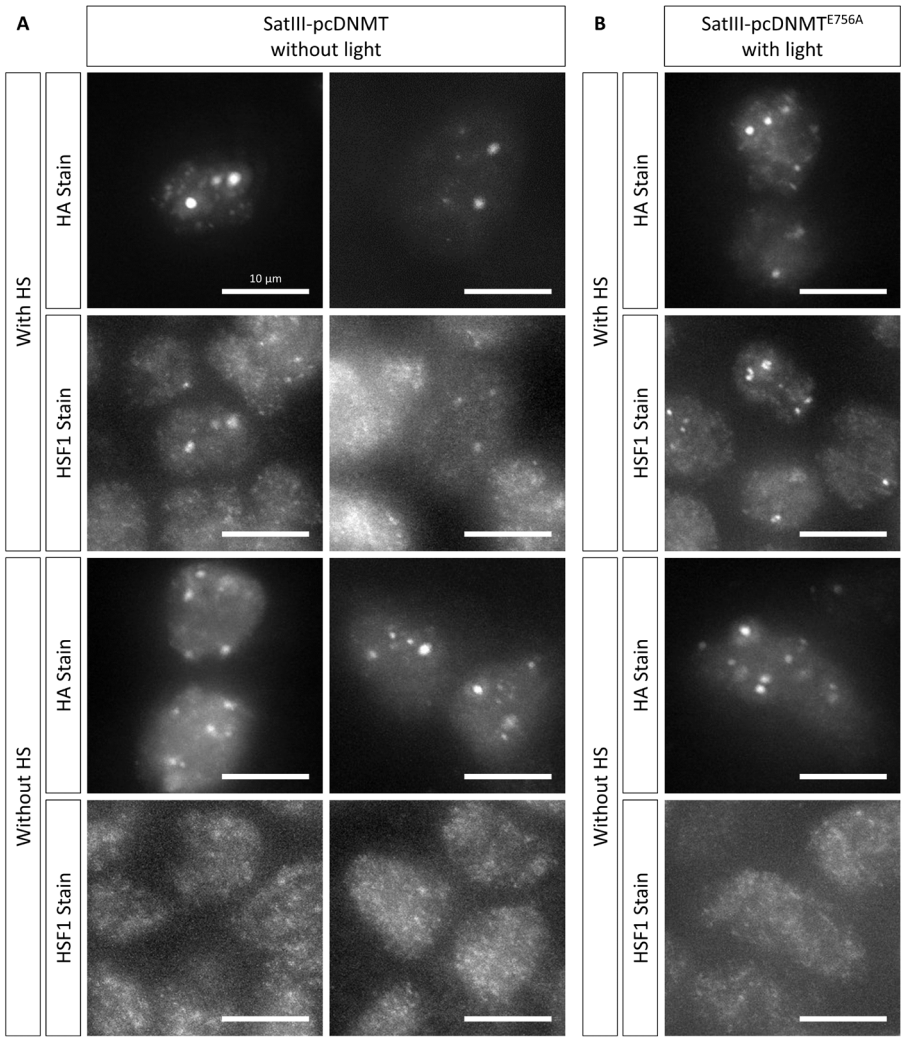


Figure 29: Colocalization of SatIII-pcDNMT and HSF1 in heat stressed HEK293T cells. A: Cells were transfected with SatIII-pcDNMT and analyzed 48 h after transfection. **B:** Cells with inactive SatIII-pcDNMT^{E756A} were light irradiated for 5 min 24 h after transfection and analyzed after another 24 h of incubation.
For all conditions, 0.05 mM DMNB-Cys was used, and cells were incubated at 44 °C (with HS) or 37 °C (w/o HS) for 1 h before fixation. Further analysis was done by immunostaining with HA- and HSF1-antibodies. Secondary antibodies were labeled with AF488 or Cy5, respectively. Scale bar is 10 µm. Modified from Wolffgramm et al.[103]

Hence, the co-localization of HSF1 and SatIII-pcDNMT can serve as an indication for the fidelity of pcDNMT-recruitment by SatIII-TALE.[332] To avoid bleed-through in microscopy, the transfection and suppression control mCherry-GFP^{Y39amber} was deleted from the

plasmid (Figure 67, page 141). HEK293T cells were transfected with such plasmid in the presence of 0.05 mM DMNB-Cys for 48 h, and a heat shock for 1 h at 44 °C was carried out to trigger a heat shock response. Subsequent cell fixation and immunostaining against the HA-tag of SatIII-pcDNMT and endogenous HSF1 was done (7.2.2.5) and cells were analyzed by fluorescence microscopy (Figure 29A). Clearly, HSF1 foci were only detected after heat shock and they also colocalized with almost all detected foci after HA-staining, verifying selective binding of the SatIII-pcDNMT construct. It was also tested if the light irradiation itself led to different localization of the construct. For this, cells were transfected with inactive SatIII-pcDNMT^{E756A} for 24 h followed by 5 min light. After another 24 h of incubation, the heat shock and immunostaining were carried out as described before (Figure 29B). Again, HSF1 accumulation was only seen after heat shock and it colocalized with the DNMT construct, proving that light alone did not influence the correct binding of the designed TALE. After ensuring the expression and the correct nuclear localization of the construct, the performance of it was tested. First, the maximum methylation achievable at this locus in HEK293T cells should be examined. For this, wild type SatIII-DNMT was transfected, mCherry-positive cells were sorted, and the SatIII locus from bisulfite converted gDNA was amplified to perform pyrosequencing analysis (7.2.2.7). Methylation levels of a single CpG in the SatIII sequence which is localized 22 bp downstream of the SatIII-TALE binding site (see “first CpG” in 8.3) was analyzed. After 24 h, a strong methylation of ~55 % 5mC by SatIII-DNMT and only ~9 % by inactive SatIII-DNMT^{E756A} was quantified (Figure 30A). Additionally, 5mC levels were measured after 48 h to see if a plateau was already reached after 24 h, but an increase to ~68 % after 48 h was observed. Nevertheless, it was chosen to test the methylation by SatIII-pcDNMT 24 h after light irradiation, since 48 h of incubation after light would result in the previously described challenges including high confluency and the need to passage cells, while the observed increase was not substantial enough in comparison to these challenges. 5mC levels after 24 h were high enough to serve as the optimal time window for testing light irradiated samples. For this, HEK293T cells were transfected with SatIII-pcDNMT in the presence of 0.05 mM DMNB-Cys, irradiated with 5 min light 24 h later, and 5mC levels were measured as described before after another 24 h of incubation. For controls, 5mC levels of inactive SatIII-pcDNMT^{E756A} as well as non-irradiated and irradiated untransfected HEK293T cells were measured. A clear methylation effect for SatIII-pcDNMT after light irradiation reaching almost 20 % was observed (Figure 30B). As expected, no increase for the sample w/o light was seen, staying at ~9 % as observed for untransfected HEK293T cells. The same was true for the non-irradiated inactive SatIII-pcDNMT^{E756A} whereas a slight increase was observed after light irradiation of inactive SatIII-pcDNMT^{E756A}. This increase might came from a weak activity of the DNMT^{E756A} mutation as also observed in Figure 30A or previously in Figure 24,

page 47, although reported differently[128,130]. The increase should not be light induced since no increase for light irradiated untransfected cells was observed. Overall, light-activatable DNA methylation of a targeted locus by a TALE-pcDNMT fusion construct in living cells was demonstrated for the first time.

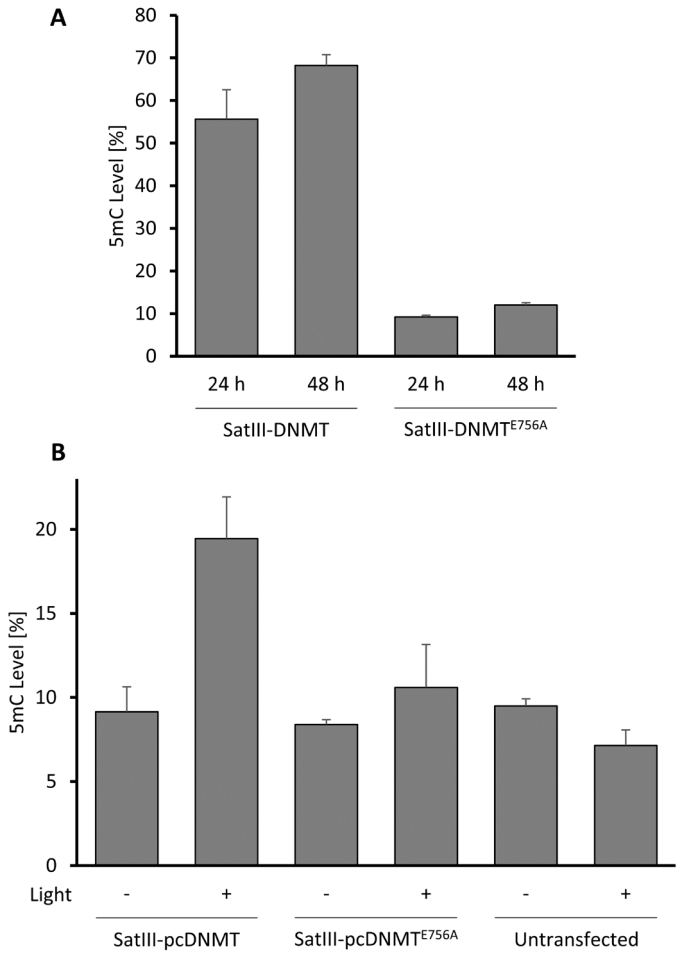


Figure 30: SatIII methylation with SatIII-DNMT and SatIII-pcDNMT in HEK293T cells. A: SatIII methylation levels were analyzed by pyrosequencing of bisulfite converted and amplified gDNA 24 h or 48 h after transfection with SatIII-DNMT or inactive SatIII-DNMT^{E756A}. 48 h data was also published in Muñoz-López et al.[332] **B:** Cells were irradiated with 5 min light 24 h after transfection (plus no light controls and untransfected cells) and incubated another 24 h later. Before analysis of SatIII methylation levels, both untransfected samples were incubated at 44 °C for 1 h. The heat shock itself did not change 5mC levels as the comparisons of heat shock vs. no heat shock samples showed in Figure 58, page 129.
Error bars show standard deviations from two independent biological replicates. Modified from Wolffgramm et al.[103]

To make sure that only on-target methylation was introduced by SatIII-pcDNMT, 5mC levels of two off-target loci (BRCA1 and p16) were evaluated by Sanger sequencing. Single loci were amplified from bisulfite converted gDNA from cells sorted 24 h after light irradiation. After bisulfite conversion, signals of thymine (red) are expected in Sanger sequencing for non-methylated cytosines and signals of cytosine (blue) for methylated cytosines (see 7.2.1.15).

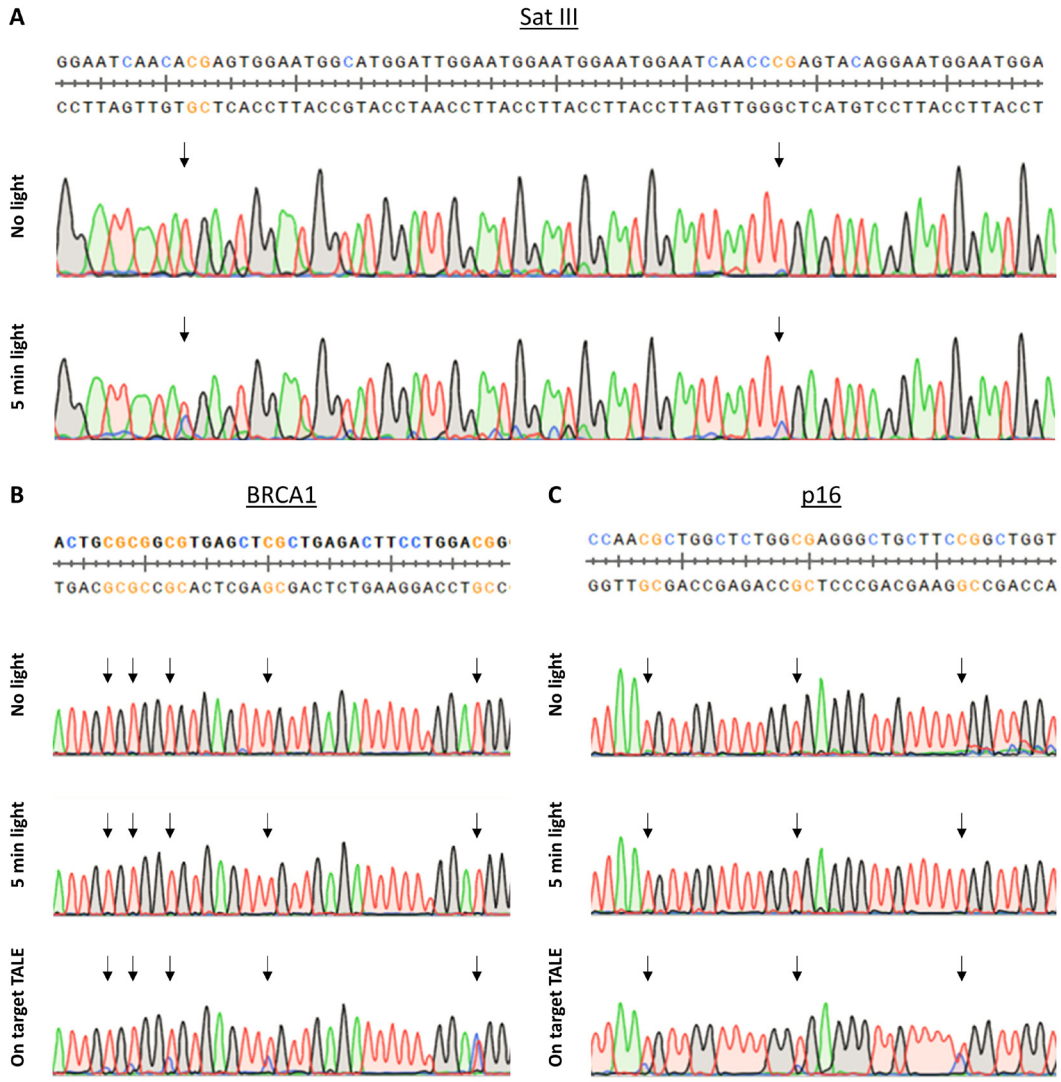


Figure 31: Bisulfite-Sanger sequencing of SatIII, BRCA1 and p16. HEK293T cells were transfected with SATIII-pcDNMT and gDNA was isolated 24 h after 5 min light irradiation (plus the control w/o light). After conversion with bisulfite, SatIII (**A**), BRCA1 (**B**) and p16 (**C**) genes were amplified and sequenced. Additionally, BRCA1 (in **B**) and p16 (in **C**) was analyzed in HEK293T cells 24 h after transfection with BRCA1-DNMT or p16-DNMT ("On target TALE"), respectively. Modified from Wolffgramm et al.[103]

No methylation was seen at the five CpGs analyzed for BRCA1 and three CpGs analyzed for p16 after light irradiation (indicated by arrows in Figure 31B and C, respectively). Figure 31A shows that cytosine signals for CpGs with ~20 % 5mC level (see first CpG (left arrow) which was analyzed in Figure 30B) can be detected, clearly indicating that BRCA1 and p16 are indeed unmethylated. To test if the absence of methylation at BRCA1 and p16 loci was only caused by condensed chromatin states which would prohibit the binding of DNMT3a3L, TALE-DNMT constructs targeting these loci (see 7.1.13, page 89 for TALE-sequences) were applied in HEK293T cells. mCherry-positive cells were sorted 24 h after transfection, then the respective locus was amplified from bisulfite converted gDNA and analyzed by Sanger-sequencing. Indeed, 5mC signals for all analyzed CpGs (see arrows for "On target TALE" in Figure 31B and C) indicated that the chromatin was accessible and therefore

methylation was possible. For p16, the three CpGs are corresponding to the first three CpGs analyzed in Figure 27. Assuming that the reached 5mC levels are comparable with previous experiments, the 5mC-signal in Sanger-sequencing can account for 13% 5mC level (comparing the first CpG in Figure 27 and Figure 31C). This proves that the absence of 5mC-signal at the BRCA1 and p16 loci with SatIII-pcDNMT indeed indicates that the actual 5mC levels are close to 0 %. It proves additionally that the absence of off-target methylation was due to the high fidelity of SatIII-pcDNMT.

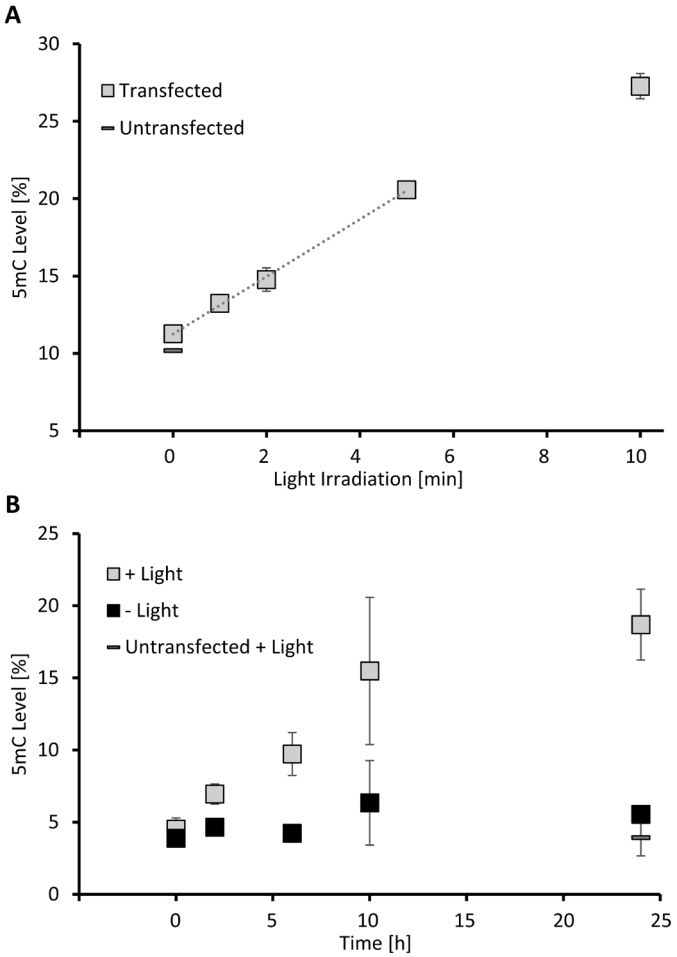


Figure 32: Light titration of SatIII-pcDNMT and kinetics of SatIII methylation. A: HEK293T cells were transfected with SatIII-pcDNMT and irradiated with 0 to 10 min light 24 h later. After another 24 h incubation SatIII methylation level was analyzed. Coefficient of determination of trend line: $R^2 = 0.9988$. **B:** Kinetics of SatIII methylation by SatIII-pcDNMT up to 24 h after 5 min light irradiation in HEK293T cells. As control, SatIII-pcDNMT w/o light for all time points and untransfected cells 24 h after light irradiation are shown.
Error bars show standard deviations from three independent biological replicates. p1660/p1169 or p1666/p1169 with 0.05 µM DMNB-Cys was used and 5mC levels were measured by pyrosequencing (**A**) or Illumina sequencing (**B**) of bisulfite converted and amplified gDNA. Modified from Wolffgramm et al.[103]

Next, it was tested if the 5mC level of SatIII is precisely titratable by different light irradiation times. For this, HEK293T cells were transfected with SatIII-pcDNMT and different light irradiation times ranging from 0 min to 10 min were applied. 5mC levels were measured by

pyrosequencing as described before (Figure 32). Additionally, the 5mC level of untransfected HEK293T cells was examined, showing again that the SatIII-pcDNMT construct stayed completely inactive w/o light. Further, a linear dependency of 5mC levels with respect to the applied light dosage for the first 5 min was visible. Such linear response is in line with the linear absorbance change of DMNB-Cys at 365 nm for 5 min shown in Figure 19, page 40. This observation extends the usefulness of the tool since it can be used to titrate specific 5mC levels needed with very high accuracy to control the strength of downstream effects. Higher 5mC levels can be seen after 10 min light though it does not follow a linear trend anymore. Nevertheless, it shows that complete decaging was not reached after 5 min which was already measured in Figure 17, page 37. Since higher light irradiation times led to more stress and morphological changes in HEK293T cells (Figure 42, page 116), 5 min remained the optimal time with the used light source.

For now, an effect of targeted methylation of the SatIII locus 24 h after light irradiation was proven. Next, time-resolved quantification of SatIII methylation should be investigated to get a higher time resolution. HEK293T cells were transfected with SatIII-pcDNMT, 5 min light was applied 24 h after transfection, and cells were sorted for GFP-signal either immediately (zero hour sample), two, six, ten or twenty-four hours later. Additionally, no light controls for all time points were done and untransfected HEK293T cells 24 h after light irradiation were measured to prove again that light itself did not change 5mC levels. After amplifying the SatIII locus from the BS-converted gDNA, amplicon deep sequencing was conducted (see 7.2.3.1). A fast and almost linear increase of 5mC levels during the first ten hours for the light samples (Figure 32B) was observed. Then, the curve began to flatten reaching around 20 % 5mC 24 h after light. As expected, there was no increase for the samples w/o light proving again that pcDNMT stays in its inactive form and with this the high fidelity of the method. The control samples also did not reach much higher 5mC levels than untransfected cells with light showing that transfection and light irradiation on its own do not interfere with SatIII methylation. Since amplicon deep sequencing was applied, methylation levels of a second CpG four bases upstream of the TALE binding site (see “second CpG” in 8.3) was additionally measured and a similar trend over 24 h was observed (Figure 57, page 128). However, the maximum level of ~13 % reached at the second CpG was lower than the maximum of the first CpG in Figure 32. This could be explained by a better accessibility of the first CpG by DNMT3a3L since it is C-terminally fused to the SatIII-TALE and thereby directed to the first CpG (8.2 and 8.3). Nevertheless, it proved again that the methylation of multiple CpGs of a defined region is possible. Since pcDNMT-dependent methylation was already seen two hours after light irradiation, the tool can be used to examine rapid cell responses to DNA methylation within a short time window.

5.4. Studying Downstream Effects

After demonstrating proper DNA methylation after light irradiation on a global and targeted gene level, downstream effects of introduced 5mC patterns should be studied. For this, the DNMT construct was changed by deleting the DNMT3L part, leaving the catalytic domain of DNMT3a (further “DNMT3aCD”, see Figure 68, page 142). The activity of DNMT3aCD was compared to DNMT3a3L and found to be even more active on a global level in both HEK293T cells and DKO HCT116 cells (Figure 59, page 129). Against the assumption that DNMT3a3L is more active because of the known positive effect of DNMT3L on DNMT3a activity (see 3.2.2), an actual ~1.5-fold increase of activity for DNMT3aCD in DKO HCT116 cells and even a ~1.7-fold increase of activity in HEK293T cells was observed. Since 5mC is known to be a repressive mark in gene regulation[63,68], rapid transcriptomal changes in living cells based on DNA-methylation should be observed. First, SatIII expression in HEK293T cells was tested after methylation by SatIII-pcDNMT but resulted in inconsistent data (not shown), which led to the assumption that SatIII expression is not solely regulated by DNA methylation of the targeted CpGs. Since specific CpGs responsible for gene silencing must be methylated, it was chosen to target the genome with DNMT3aCD on a global level with subsequent analysis of the whole transcriptome instead of targeting a single locus. Although DKO HCT116 cells have very low 5mC background that would serve as an ideal cell line to study changes in gene expression levels, the transfection efficiency in this cell line was too low for proper analysis. Therefore, HEK293T cells were used despite their high 5mC background (Figure 20, page 42), which should still have unmethylated genes serving as the target of DNMT3aCD. HEK293T cells were transfected with photocaged DNMT3aCD (“pcDNMT3aCD”) in the presence of 0.05 mM DMNB-Cys and irradiated with light for 5 min 24 h after transfection. GFP-positive cells were sorted either immediately (0 h sample), 4 h, or 8 h later, then total mRNA was extracted and sent for RNA-sequencing (7.2.3.2). Indeed, significant changes in mRNA levels were observed between the light irradiated group and the control group w/o light-activation. The genes were assigned to four main groups based on their transcriptional behavior (Figure 33, for additional data see 8.8). The first group consisted of 9 genes which showed slightly higher gene expression 8 h after light irradiation and slight downregulation in the control w/o light. This behavior was against the assumption that 5mC leads to gene silencing and further studies are needed for these genes to test if the effect was due to light irradiation or DNA methylation. All other main groups though showed the expected downregulation of mRNA levels at different time points. The third and fourth group were also small with only eleven and four genes, respectively, and showed lower mRNA levels already at 4 h after light irradiation. Genes in the third group though showed again an upregulation to the initial levels

after 8 h. Overall, the effect of these two groups were weaker than the effects observed in the large second group.

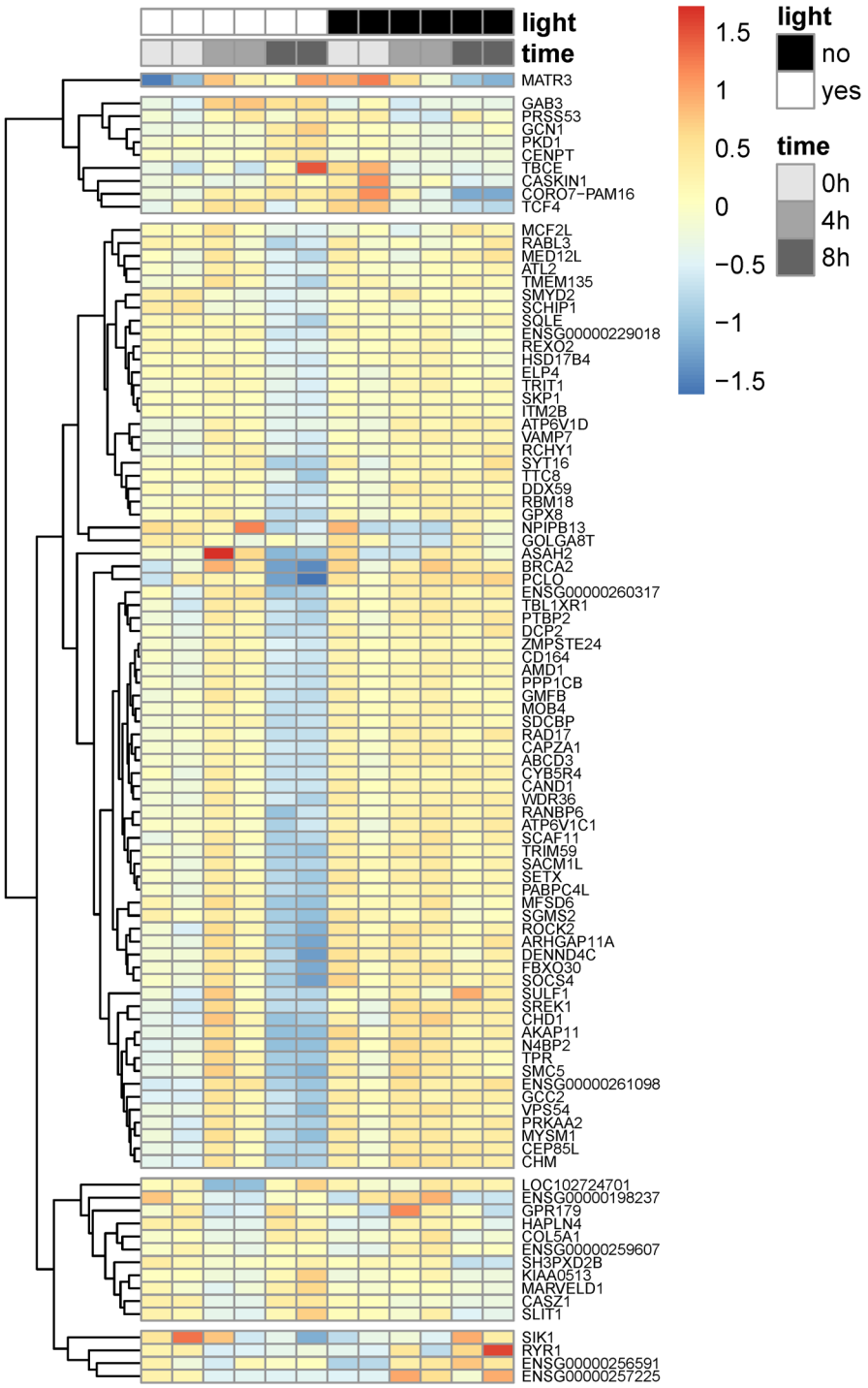


Figure 33: Transcriptome analysis of pcDNMT3a transfected HEK293T cells. Total RNA was isolated from pcDNMT3a transfected HEK293T cells 0 h, 4 h and 8 h after decaging of DMNB-Cys by 5 min light irradiation. The same time points were used for the respective control w/o light. Shown is the clustering of differentially expressed genes with a p-value of 0.05. The color code shows the difference of regularized log2-normalized counts of the mean expression of all samples. Modified from Wolffgramm et al.[103]

Strikingly, several of the genes in the third group play roles in cell structure and cell cycle regulation (HAPLN4, COL5A1, SH3PXD2B, MARVELD1 and SIK1), which could be a hint that this slight downregulation is due to fast structural change upon light but not DNA methylation since most genes do also show normal expression levels after 8 h. However, morphology changes of HEK293T cells were not observed 6 h after 5 min light irradiation (Figure 41, page 115). For some genes (SH3PXD2B, KIAA0513, MARVELD1, CASZ1 and SLIT1) the mRNA levels of the control w/o light also showed slight downregulation but after 8 h instead of 4 h. Therefore, a more specific testing for these genes is necessary and light effects should be tested with the inactive pcDNMT3aCD^{E756A} control. Nevertheless, for several mentioned genes a connection between gene expression and DNA methylation has been already reported (HAPLN4[333], SH3PXD2B[334], MARVELD1[335], SLIT1[336], RYR1[337]). However, for some genes the here observed downregulation effect was in contradiction to previous publications where no connection of gene expression to 5mC levels was observed (COL5A1[338], CASZ1[339], SIK1[340]). CASZ1 though plays a role in transcription activation and a fast downregulation due to DNA methylation might be possible.

The large second group contained 73 genes, and some showed strong downregulation of mRNA levels 8 h after light irradiation whereas no changes were observed for the control w/o light. It is noticeable that 34 genes of this group can be assigned to functions regarding RNA processing and transcription regulation, transport within the cell, cell structure, and cell cycle regulation. Especially RNA processing and transcription regulation might require quick response to DNA methylation. 14 genes were assigned to such group: MED12L, ELP4, TRIT1, DDX59, RBM18, TBL1XR1, PTBP2, DCP2, ZMPSTE24, WDR36, SCAF11, SETX, PABPC4L, SREK1. Precisely, MED12L (mediator of RNA Pol II)[341], ELP4 (part of the RNA Pol II elongator complex)[342], and SETX (modulates RNA Pol II binding)[343] are associated to RNA Pol II activity. General RNA binders (DDX59, RBM18, PTBP2, and PABPC4L) and processors of different RNAs (TRIT1, DCP2, WDR36, SCAF11, and SREK1) were also found. This shows that many downregulated genes were associated with gene transcription. Another 11 genes (ATP6V1D, VAMP7, SYT16, SDCBP, ABCD3, RANBP6, DENND4C, TPR, GCC2, VPS54, CHM) showed functions regarding transport, however, no specific type of transportation was found and rather a broad range of genes with such affiliation was downregulated. Again, multiple genes were found to play roles in the cell structure and cell cycle regulation: MCF2L (guanine nucleotide exchange factor for RhoA and CDC42)[344], RABL3 (KRAS signaling)[345], SKP1 (part of the SCF complex which ubiquitinates proteins involved in the cell cycle)[346], RCHY1 (ubiquitin ligase and HDAC2 regulator)[347], AMD1, CAPZA1 (capping protein of actin filaments)[348], ROCK2 (regulator of actin)[349], ARHGAP11A (regulator of Rho GTPases)[350], and SOCS4. Besides SKP1, other two genes (CAND1 and FBXO30) linked to the SCF complex were also downregulated. Especially

genes involved in the actin cytoskeleton and protein ubiquitination were found showing that DNA methylation has a fast impact on these cell functions. Three genes connected to histone modifications were also downregulated (SMYD2, MYSM1, and CHD1). SMYD2 introduces trimethylation on histone H3 (H3K4me3), which is found at active genes[351], while CHD1 selectively binds to such histone mark[352] and is involved in the maintenance of open chromatin[353]. MYSM1 is a deubiquitinase of monoubiquitinated histone H2A, a repressive mark regarding gene expression.[354] The downregulation of these three genes fits to the assumption that 5mC leads to gene silencing. Furthermore, six genes (GPX8, BRCA2, RAD17, CYB5R4, N4BP2, SMC5) which play roles in stress response and DNA repair were downregulated, which is not expected if light irradiation has caused DNA damage. Since none of the slightly upregulated genes (first group) showed connection to stress response, it could be that the six genes are indeed downregulated due to DNA methylation. For some of the mentioned genes a connection between high or low DNA methylation and decreased or increased gene expression, respectively, has been reported: MCF2L[355], SQLE[356], AMD1[357], ABCD3[358], ARHGAP11A[359], FBXO30[360], SOCS4[361]. This supports the assumption that downregulation was caused by DNA methylation, while for other genes no 5mC related expression change has been reported in previous publications (SMYD2[362], HSD17B4[363], PPP1CB[364], SULF1[365]). Since these studies were conducted in different cell lines under different conditions, it is difficult to compare with the observed effects in HEK293T cells. Nevertheless, it is possible to conclude that higher 5mC levels were indeed connected to the downregulation of mRNA levels within a short time window of up to 8 h, despite some specific genes that only showed weak effects, or also slight effects for the no light controls. This shows that the tool can be used to monitor rapid downstream effects in cells, which is a challenge with constitutively active DNMTs. With this, downstream effects in living cells after induced global DNA methylation by light-activating pcDNMT3aCD was demonstrated for the first time, and genes that respond quickly to DNA methylation were revealed which provides the first insight about the regulatory mechanisms of the transcriptome. Furthermore, it is possible to target specific genes with the respective TALE-pcDNMT constructs to unveil the role of DNA methylation in these genes.

6. Summary and Outlook

This work reports for the first time the direct activation of DNMTs by light in eukaryotic cells. This novel tool provides unprecedented spatio-temporal control over DNA methylation in living cells and allows the dissection of the role of 5mC. Using genetic code expansion, the essential catalytic Cys710 residue of DNMT3a was replaced with a photolabile DMNB-Cys, resulting in the expression of transiently inactive DNMT3aCD-DNMT3L fusion enzymes. Upon light irradiation at a defined time point, a rapid cleavage of the DMNB group restored the unmodified cysteine and thus the enzymatic activity. The successful incorporation of DMNB-Cys was characterized, and the efficacy of global DNA methylation after light-activation was demonstrated. In contrast to previous approaches, the ability to define the starting point of enzymatic activity of unmodified DNMT3aCD allowed kinetics studies of cancer-associated mutants. It was possible to monitor the direct impact of chosen mutations on the catalytic activity itself in living cells since the catalytic activity was uncoupled from upstream processes like transfection and protein expression. Targeted methylation of user-defined loci was enabled by fusing TALEs to photocaged DNMTs. High fidelity of on-target methylation was demonstrated at the SatIII locus. Furthermore, 5mC levels could be adjusted to specific levels by applying different light irradiation times with a linear correlation for up to 5 min. This allows studies of effects of tunable 5mC levels at specific genes. The tool was further used to explore fast effects of 5mC on the transcriptome. Multiple genes that are involved in cell functions that might require fast adaption to DNA methylation, including gene expression, cell cycle regulation, and transport, showed lower expression levels after DNA methylation. This highlights the usefulness of light-controlled DNMTs which can be applied to unveil kinetics of downstream processes triggered by 5mC in living cells either on a global level or user-defined gene loci.

In future studies, the strategy of photocaging the catalytic cysteine could be transferred to other DNA methyltransferases since the methylation reaction is highly conserved throughout this group of enzymes (see 3.2.2). Also, full-length DNMT3a can be studied to ensure that protein interactions and regulations of all DNMT3a domains come to effect. Moreover, it is interesting to examine if *de novo* methylation by DNMT3a in DNMT-deficient and non-methylated genomes (e.g. DKO HCT116 cells) will result in similar 5mC patterns found in naturally methylated genomes (e.g. WT HCT116 cells) and if specific sequences are preferred for early methylation[138]. This might give novel insights into the mechanisms of regulation directed by DNMTs. Furthermore, it should be tested how long globally or locus-specifically introduced 5mC patterns can be maintained and whether targeted epigenetic imprinting can be achieved. Such study could be done in embryonic stem cells

targeting crucial developmental genes. It is also interesting to study the downstream effects of DNA methylation in a different pattern or context (CpA methylation) created by specific DNMT3a mutants like R836A. Such studies are only feasible due to the ability to write DNA methylation with high spatio-temporal control. The developed tool can also be applied on time-resolved monitoring of changes in the chromatin state caused by reader/eraser recruitment or changed histone modifications triggered by DNA methylation. This will unveil kinetics of 5mC reader and eraser protein recruitment in relevant cellular contexts. It might also provide insights into the control of the epigenome in living cells during cell development or disease progression.

7. Materials and Methods

7.1. Materials

7.1.1. Laboratory Equipment

Table 2: Used Laboratory Equipment.

Type	Model	Supplier
Agarose Chamber	Midicell Primo EC330	ThermoEC
Autoclave	Varioklav	H+P
Autoclave	VX-150	Systec
Beaker	Different volumes	Simax and VWR
Bunsen Burner	1010	Usbeck
Calculator	Fx-82 solar	Casio
Camera	PowerShot G10	Canon
Camera Chamber	BDAdigital	Biometra
Cell Sorter	SH800S	SONY
Centrifuge	5415R; 5424; 5702R; 5810R	Eppendorf
Centrifuge	Micro Centrifuge	Carl Roth
Centrifuge	Micro centrifuge Model I R 220VAC	Carl Roth
Centrifuge	Sorvall Evolution RC with Sorvall SLA-3000	Thermo Scientific
Centrifuge	Sorvall Lynx 6000 with Fiberlite F14-6x250y	Thermo Scientific
Concentrator	Concentrator Plus	Eppendorf
Dewar	Dewar	Isotherm
Erlenmeyer Flask	Different volumes	Simax and VWR
Flask	Different volumes	VWR
Freezer	-152 °C	Sanyo
Freezer	-20 °C Profi Line	Liebherr
Freezer	-80 °C	New Brunswick Scientific
Fridge	4 °C Profi Line	Liebherr

Type	Model	Supplier
Fridge	C10GH 2D NordCap Val Mod C10	NordCap
Glass Beads	Glass Beads	VWR
Heating Block	AccuBlock	Labnet
Heating Block	CHB-202	Bioer
Heating Block	ThermoStat Plus	Eppendorf
Ice Machine	AF20	Scotsman
Incubator	100-800	Memmert
Incubator – Shaker	I26	New Brunswick Scientific
Magnetic Stirrer	MR Hei-Mix S; MR Hei-Standard	Heidolph
Measuring Cylinder	Different volumes	Brand and Hirschmann
Microwave	ED 8525.3S	Exquisit
Nano-HPLC	UltimateTM 3000 RSLCnano	Thermo Scientific
PCR Cycler	MyCycler™ Thermal Cycler System	Bio-Rad
PCR Cycler	SimpliAmp	Life Technologies
PCR Cycler	T Personal Thermocycler	Biometra
pH-Meter	Five easy	Mettler-Toledo
Pico Tip emitter	ID 20, Tip-ID 10	New Objective
Pipette	0.1-2.5 µL Research Plus	Eppendorf
Pipette	0.5-10 µL Research Plus	Eppendorf
Pipette	0.5-10 µL Xplorer (12 channel)	Eppendorf
Pipette	100-1000 µL Research Plus	Eppendorf
Pipette	10-100 µL Research Plus	Eppendorf
Pipette	10-100 µL Xplorer (12 channel)	Eppendorf
Pipette	2-20 µL Research Plus	Eppendorf
Pipette	50-1200 µL Picus 1200 (12 channel)	Sartorius
Plate Reader	Infinite M1000 plate reader	Tecan
Power Supply	EV233	Consort
Pyrosequencer	PSQ HS 96ATwo	Qiagen
Rack	0.2 mL tubes	VWR

Type	Model	Supplier
Rack	1.5 mL tubes	GLW and VWR
Rack	15 mL tubes	GLW and neoLab
Rack	50 mL tubes	GLW and neoLab
Rack	MagRack 6	GE Healthcare Life Sciences
Rotator	88881002	Thermo Scientific
Scale	3500-2NM	Kern PLJ
Scale	AX224	Sartorius
Scale	PM400	Mettler
Scalpel	Cutfix	Braun
Scanner	CanonScan 9000F	Canon
Scanner	Typhoon FLA 9500	GE Healthcare Life Sciences
SDS PAGE Chamber	Mini-PROTEAN Tetra System with PowerPac Basic	Bio-Rad
Timer	3 Channel Display Clock Timer	Jumbo
Ultrasonic Bath	Ultrasonic Cleaner	VWR
UV Protection	EN 166-3F	Bollé
UV Table	UVstar 312 nm	Biometra
UV-Vis Spectral Photometer	NanoDrop2000	Thermo Scientific
Vacuum Manifold	QIAvac 24 Plus	Qiagen
Vacuum Pump	VNC2	Vacuubrand
Vortex Mixer	RS-VA 10	Phoenix Instrument
Water	A10 TOC	Millipore
Water	Purelab flex 2	ELGA
Waterbath	JB Aqua 12 Plus	Grant
Western Blot	Pierce™ Power Blotter	Thermo Scientific
Western Blot Scanner	Odyssey® CLx	LI-COR

7.1.2. Cell Culture and Microscopy Equipment

Table 3: Used Cell Culture and Microscopy Equipment.

Type	Model	Supplier
Biological Safety Cabinet	NU S440 500E	Nuaire
Biological Safety Cabinet	Cellguard ES Class II	Nuaire
Calculator	ELSI MATE EL-2435	Sharp

Type	Model	Supplier
Centrifuge	5810R	Eppendorf
Centrifuge	Sprout	Heathrow Scientific
Cryo Container	Mr. Frosty (5100-0001)	Nalgene
Hemocytometer	718605 (Neubauer)	Brand
Incubator	HeraCell 150	Thermo Electron Corp.
Incubator	CellXpert C170i	Eppendorf
Microscope1	IX81	Olympus
Microscope1 Camera System	ORCA-R^2 – C10600	Hamamatsu
Microscope1 Controller	STC	Olympus
Microscope1 Lamp	IX2-UCB	Olympus
Microscope1 Objective	UPlanSApo 10x	Olympus
Microscope1 Objective	UPlanSApo 20x	Olympus
Microscope1 Objective	UPlanSApo 40x	Olympus
Microscope2	Eclipse TE200	Nikon
Microscope2 Objective	10x	Nikon
Microscope2 Objective	20x	Nikon
Microscope3	EVOS XL Imaging System	Life Technologies
Microscope3 Objective	2x Amep 4631	Life Technologies
Microscope3 Objective	4x Amep 4632	Life Technologies
Microscope3 Objective	10x Amep 4633	Life Technologies
Microscope3 Objective	20x Amep 4634	Life Technologies
Pipet Filler	Accu-jet	Brand
Pipet Filler	Pipetus	Hirschmann
Pipette	0.1-2.5 µL Research Plus	Eppendorf
Pipette	0.5-10 µL Research Plus	Eppendorf
Pipette	10-100 µL Research Plus	Eppendorf
Pipette	100-1000 µL Research Plus	Eppendorf
Pipette	0.1-2.5 µL Research	Eppendorf
Pipette	0.5-10 µL Research	Eppendorf
Pipette	10-100 µL Research	Eppendorf
Pipette	100-1000 µL Research	Eppendorf
Tally Counter	H20	Link
Thermomixer	Comfort Compact	Eppendorf
Timer	C5079	Carl Roth

Type	Model	Supplier
UV Table	WUV-L10	Witeg
Vacuum Pump	Vacusafe comfort	IBS Integra Biosciences
Vacuum Pump	uniVACUUSYS1	LLG Labware
Vortex Mixer	Vortex-Genie 2	Scientific Industries
Waterbath	3047	Köttermann
Waterbath-Supplement	Waroklar	Waldeck

7.1.3. Consumables

Table 4: Used Consumables.

Type	Model	Supplier
Autoclave Bags	V200300	Diagonal
Bottle Top Filter	83.1823.101	Sarstedt
C18 PepMap 100 column	5 µm, 100 Å, 300 µm ID * 5 mm 3 µm, 100 Å, 75 µm ID * 50 cm	ThermoScientific
Cannulas	Sterican 0.80 x 120 mm; Sterican 0.90 x 70 mm	Braun
Cover Film	Self-adhesive, soft PVC	Ratiolab
Cryotube	1.8 mL (72.379)	Sarstedt
FACS Tube	5 mL Round Bottom (352058)	Falcon
FACS Tube	5 mL Round Bottom with Cell Strainer (352235)	Falcon
Filter	Filtropur S 0.2 (83.1826.001)	Sarsted
Fixation/Permeabilization	Fix + Perm Cell Permeabilization Kit	Life Technologies
Gas	C206 GLS Super	Campingaz
Gloves	Nitrile – L	Semperguard and VWR
Laboratory Film	PM-996 – Parafilm	Bemis
Petri Dish	82.1473.001	Sarstedt
Pipette	3.5 mL (86.1171)	Sarstedt
Pipette Tip	10 µL (70.1130); 200 µL (70.760.002); 1000 µL (70.3050.020)	Sarstedt

Type	Model	Supplier
Pipette Tip	10 µL (0030 072.006)	Eppendorf
Pipette Tip	200 µL Biosphere (70.760.202); 1000 µL Biosphere (70.762.200)	Sarstedt
Pipette Tip	1200 µL Optifit Tip	Sartorius
Pipette Tip with Filter	MultiGuard3 200 µL (35240)	Sorenson BioScience
Pipette Tip with Filter	100 µL low-binding (732-2385)	VWR
Pipette Tip with Filter	20 µL (732222)	Brand
Reaction Plate	96 well plates black	Greiner Bio-One
Reaction Plate	96 well PCR plate (82006-636)	VWR
Reaction Tube	0.2 mL (72.737.002); 0.5 mL (72.704.200); 1.5 mL (72.706); 1.5 mL brown (72.690.004); 1.5 mL low binding (72.706.700) 2 mL (72.695.500)	Sarstedt
Reaction Tube	1.5 mL low binding (0030 108.116)	Eppendorf
Serological Pipette	10 mL (86.1254.001); 25 mL (86.1685.001)	Sarstedt
TC-Plate	100 mm (83.3902) 35 mm (83.3900) 6-Well (83.3920)	Sarstedt
TC-Plate	µ-Dish 35 mm (81158)	Ibidi
Tissues	Facial Tissues	PrimeSource
Tissues	Kimtech Science 7552	Kimberly-Clark
Transfer Membrane	Immobilon-FL	Merck
Tube	15 mL (62.554.502) 50 mL (62.547.254) 50 mL brown (62.547.354)	Sarstedt
ZipTip	Pierce C18 Tips 100 µL (87784)	Thermo Scientific

7.1.4. Kits

Table 5: Used Kits.

Type	Name	Supplier
Bisulfite Conversion	EpiTect Bisulfite Kit	Qiagen
Bisulfite PCR	EpiTect MSP Kit	Qiagen
gDNA Isolation	QIAamp DNA Mini Kit	Qiagen
Gel Extraction and PCR Purification	GeneJET Gel Extraction Kit	Thermo Fisher Scientific
Gel Extraction and PCR Purification	NucleoSpin Gel and PCR Clean-up	Macherey-Nagel
PCR Purification	Monarch® PCR & DNA Cleanup Kit	NEB
Plasmid Isolation	GeneJET Plasmid MiniPrep	Thermo Fisher Scientific
Plasmid Isolation	NucleoSpin Plasmid EasyPure; NucleoBond Xtra Maxi Plus Kit	Macherey-Nagel
Purification	QIAquick Nucleotide Removal Kit	Qiagen
RNA Isolation	RNeasy Mini Kit	Qiagen

7.1.5. Biological Reagents and Chemicals

Table 6: Used Biological Reagents and Chemicals.

Name	Supplier
2-Log DNA Ladder	NEB
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES; ≥99.5 %)	Carl Roth
4,5-Dimethoxy-2-Nitrobenzyl Bromide	Sigma Aldrich
5-aza-2-deoxycytidine (Decitabine)	Selleckchem
Acetic Acid	Carl Roth
Acetonitrile	Sigma Aldrich
Ammonium bicarbonate	Merck
Ammonium peroxodisulfate	Carl Roth
Boric acid (≥99.8 %)	Carl Roth
Bovine serum albumin (BSA)	Cell Signaling
Bromophenol blue	Sigma Aldrich

Name	Supplier
Calcium chloride	Fisher Scientific
Carbenicillin · 2 Na	Carl Roth
Chloroform	Sigma Aldrich
Coomassie Brilliant Blue G250	Carl Roth
D-(+)-Glucose	Sigma-Aldrich
Deoxyadenosine triphosphate (dATP)	NEB
Deoxycytidine triphosphate (dCTP)	NEB
Deoxyguanosine triphosphate (dGTP)	NEB
Deoxythymidine triphosphate (dTTP)	NEB
Diethyl ether	Sigma Aldrich and Carl Roth
Dimethyl sulfoxide (DMSO; ≥99.7 %)	Merck
Disodium phosphate ($Na_2HPO_4 \cdot 2\ H_2O$)	Sigma-Aldrich
Dithiothreitol (DTT; ≥99 %)	Carl Roth
DNA-ExitusPlus	AppliChem Panreac
dNTPs (10 mM)	NEB
Dulbecco's Modified Eagle Medium (DMEM; w/o L-Glutamine; with D-Glucose; with Pyruvate)	PAN Biotech
Dulbecco's phosphate-buffered saline (DPBS)	PAN Biotech
Ethanol (EtOH; ≥99.8 %)	VWR Chemicals and Fisher Chemical
Ethidium bromide	Carl Roth
Ethylenediaminetetraacetic acid (EDTA; ≥99 %)	Carl Roth
Fetal bovine serum (FBS; Premium South America)	PAN Biotech
Formaldehyde (37 %)	Merck
Formic Acid	Carl Roth
FuGENE® 6 Transfection Reagent	Promega
Glycerole	Carl Roth
Glycine	Carl Roth
H_2O Nuclease-free	Qiagen
H_2O Nuclease-free	Promega
Hydrochloric acid (HCl; 37 %)	Merck and Carl Roth
Imidazole	Carl Roth
Kanamycinsulfate	Carl Roth
LB-Agar (Lennox)	Carl Roth
LB-Medium (Lennox)	Carl Roth

Name	Supplier
L-Cysteine Hydrochloride	Carl Roth
LE Agarose	Biozym
L-Glutamine	PAN Biotech
Lipofectamine 2000 Reagent	Invitrogen
Magnesium chloride ($MgCl_2$)	NEB
Magnesium chloride hexahydrate ($MgCl_2 \cdot 6\ H_2O$; 99%)	Acros Organics
Magnesium sulfate ($MgSO_4 \cdot x\ H_2O$)	Sigma-Aldrich
Methanol	Alfa Aesar
Monopotassium phosphate (KH_2PO_4)	Sigma-Aldrich
Nicotinamide adenine dinucleotide (NAD^+)	NEB
Odyssey Blocking Buffer	LI-COR
Opti-MEM® I	Gibco
PageRuler™ Plus Prestained Protein Ladder	Thermo Fisher Scientific
PEG-8000 (Molecular Biology Grade)	Promega
Penicillin-Streptomycin (10,000 U/mL Penicillin; 10 mg/mL Streptomycin)	PAN Biotech
Phenol-Chloroform-Isoamyl alcohol mixture (PCI)	Sigma Aldrich
Phenylmethylsulfonyl fluoride	Carl Roth
Pierce™ Iodoacetamide, Single Use	Thermo Fisher Scientific
Polyethylenimine (PEI)	Alfa Aesar
Poly-L-lysine hydrobromide (PLL)	Sigma-Aldrich
Potassium chloride (KCl; ≥99.5 %)	Carl Roth
Propan-2-ol	Fisher Chemical
RIPA Buffer	Thermo Scientific
Rotiphorese Gel 40	Carl Roth
RPMI 1640 (with L-Glutamine; with 2.0 g/L NaHCO3)	PAN Biotech
Sodium acetate	Merck
Sodium azide	Merck
Sodium chloride (NaCl; ≥99.5%)	Carl Roth
Sodium dodecyl sulfate (SDS)	Carl Roth
Sodium hydroxide (NaOH)	Fisher Scientific
Spectinomycin	Alfa Aesar
Tetracycline	Sigma-Aldrich

Name	Supplier
Tetracycline hydrochloride	Carl Roth
Tetrahydrofuran (THF)	Merck
Tetramethylethylenediamine (TEMED)	Carl Roth
Trifluoroacetic acid (TFA)	Sigma Aldrich
Tris(hydroxymethyl)-aminomethan (TRIS; 99.9 %)	Carl Roth
Triton X-100	Fluka Chemika
TRIzol™ Reagent	Life Technologies
Trypsin/EDTA (w/o Ca^{2+}, Mg^{2+})	PAN Biotech
Tryptone	Carl Roth
Tween20	Fisher Bioreagents
X-Gal	Thermo Fisher Scientific
X-tremeGENE9 DNA Transfection Reagent	Roche
Xylene cyanol	Carl Roth
Yeast Extract	Carl Roth

7.1.6. Buffers and Media

Table 7: Used Buffers and Media.

Name	Formulation
0.5x TBE	• 89 mM Tris • 89 mM Boric acid • 2 mM EDTA (diluted from 0.5 M solution, pH = 8)
1.33x Gibson Mastermix	• 1.33x Isothermal Reaction Buffer • 1 U/µL T5 exonuclease • 2 U/µL Phusion DNA polymerase • 40 U/µL Taq DNA ligase
4x PBS	• 548 mM NaCl • 43 mM KCl • 69 mM $Na_2HPO_4 \cdot 2H_2O$ • 3.2 g KH_2PO_4 • Adjust pH to 8

Name	Formulation
5x Isothermal Reaction Buffer	• 25 % PEG-8000 • 500 mM Tris-HCl pH 7.5 • 50 mM $MgCl_2$ • 50 mM DTT • 1 mM dATP • 1 mM dTTP • 1 mM dCTP • 1 mM dGTP • 5 mM NAD^+
CutSmart Buffer (Supplied by NEB)	• 50 mM KOAc • 20 mM Tris-acetate • 10 mM $Mg(OAc)_2$ • 100 µg/mL BSA • pH 7.9
DNA Storage Buffer	• Elution Buffer (GeneJET Plasmid MiniPrep; Thermo Fisher Scientific) • 1 mM EDTA
E. coli Storage Buffer	• 50 mM $CaCl_2$ • 15 % (v/v) Glycerol
FACS Blocking Buffer	• DPBS • 1 % BSA • 0.05 % Tween20
FACS Sorting Buffer	• DPBS (Ca^{2+}/Mg^{2+} free) • 5 mM EDTA • 25 mM HEPES • 1 % BSA • Freshly add 10 U/mL DNase I
FACS Washing Buffer	• DPBS • 0.1 % NaN_3 • 5 % FBS
WT and DKO HCT116 Growth Medium	• RPMI (with L-Glutamine; with 2 g/L $NaHCO_3$) • 10 % FBS Premium • 1 % Penicillin/Streptomycin

Name	Formulation
HEK293T Growth Medium	• DMEM (w/o L-Glutamine; with D-Glucose; with Pyruvate) • 10 % FBS Premium • 1 % L-Glutamine • 1 % Penicillin/Streptomycin
NEB Buffer 3.1 (Supplied by NEB)	• 100 mM NaCl • 50 mM Tris-HCl • 10 mM $MgCl_2$ • 100 µg/mL BSA • pH 7.9
PBS + 0.25 % Triton X-100	• 1x PBS • 0.25 % Triton® X-100
PBS + 4 % Formaldehyde	• 1x PBS • 4 % Formaldehyde
PBS-T	• DPBS • 0.05 % Tween20
Pfu Reaction Buffer (10x)	• 100 mM Tris-HCl pH 8.8 • 500 mM KCl • 20 mM $MgCl_2$ • 1 % Triton® X-100
PLL Coating	• 0.01 % PLL in DPBS
Resolving SDS Gel (10 %)	• 3535 µL H_2O • 1365 µL Rotiphorese Gel 40 • 420 µL Tris (1.5 M, pH 8.8) • 66.7 µL SDS (10 %) • 66.7 µL APS (10 %) • 6.67 µL TEMED
SDS Gel Comassie Staining	• Ethanol 50 % (v/v) • Acetic Acid 10 % (v/v) • H_2O 40 % (v/v) • Comassie Brilliant Blue G250 0.1 % (w/v)
SDS Gel Destaining	• Ethanol 20 % (v/v) • Acetic Acid 10 % (v/v) • H_2O 70 % (v/v)

Name	Formulation
SDS loading buffer (4 x) pH 6.8	• Tris-HCl 200 mM • Glycerol 40 % (v/v) • 2-Mercaptoethanol 1 % (v/v) • SDS 8 % (w/v) • Bromphenol blue 0.008 % (w/v)
SDS Running Buffer (10x)	• Tris 30.3 g/L • Glycine 144.1 g/L • SDS 10 g/L
SOC	• 0.58 g/L NaCl • 2.03 g/L $MgCl_2 \cdot 6H_2O$ • 2.46 g/L $MgSO_4 \cdot 7H_2O$ • 20 g/L Tryptone • 5 g/L Yeast Extract • Water • Adjust pH to 7-7.5 • Add 20 mL of 1 M sterile glucose solution after autoclaving
Stacking SDS Gel (5 %)	• 1000 µL H_2O • 208 µL Rotiphorese Gel 40 • 420 µL Tris (0.5 M, pH 6.8) • 16.7 µL SDS (10 %) • 16.7 µL APS (10 %) • 1.67 µL TEMED
T4 DNA ligase buffer	• 50 mM Tris-HCl • 10 mM MgCl2 • 1 mM ATP • 10 mM DTT • pH = 7.5
TBS-T	• 50 mM Tris-HCl, pH 7.5 • 150 mM NaCl • 0.05 % (w/v) Tween20
ThermoPol® Reaction Buffer (Supplied by NEB)	• 20 mM Tris-HCl • 10 mM $(NH_4)_2SO_4$ • 10 mM KCl • 2 mM $MgSO_4 \cdot 7H_2O$ • 0.1 % Triton® X-100 • pH 8.8

7.1.7. Antibiotics

Table 8: Used Antibiotics.

Name	Solvent	Concentration [mg/mL]
Carbenicillin · 2Na	EtOH/H_2O	50 mg/mL
Kanamycin	H_2O	50 mg/mL
Penicillin-Streptomycin (Supplied by PAN Biotech)		
Spectinomycin	H_2O	100 mg/mL
Tetracycline	EtOH	12.5 mg/mL

7.1.8. Enzymes

Table 9: Used Enzymes.

Application	Name	Supplier
Digestion for MS Analysis	Trypsin-ultra MS grade	NEB
FACS Buffer	DNase I	Thermo Scientific
GG-assembly	BsaI	NEB
GG-assembly	BsmBI (Esp3I)	NEB
GG-assembly	Plasmid-Safe DNase	Epicentre
Gibson-Assembly	T5-Exonuclease	NEB
Gibson-Assembly	Taq-DNA Ligase	NEB
Ligation / GG-assembly	T4-DNA Ligase	NEB
PCR	Pfu	AG Summerer
PCR	Q5	NEB
PCR	Taq	AG Summerer
PCR and Gibson-Assembly	Phusion	NEB
Restriction	AscI	NEB
Restriction	DpnI	NEB
Restriction	PacI	NEB
RNA Isolation	DNase I	NEB

7.1.9. Strains

7.1.9.1. Bacteria Strains

Table 10: Used Bacterial Strains.

E. coli Strain	Lab Code	Supplier
DH5 alpha	sBeB378	Michael Knop, Heidelberg University
GH371	sDaS58	iGEM

7.1.9.2. Mammalian Strains

Table 11: Used Mammalian Strains.

Strain	Lab Code	Supplier
DKO HCT116	sJaW387	Horizon Discovery Ltd (HD R02-022)
WT HCT116	sJaW386	Horizon Discovery Ltd (HD PAR-033)
HEK293T	sAnW376	MPI Dortmund

7.1.10. Antibodies

Table 12: Used Primary Antibodies.

Epitope	Description	Company (Cat. #)
5mC	Anti-5-methylcytosine Clone 33D3 (mouse monoclonal)	Merck (MABE146)
GAPDH	GAPDH (14C10) Rabbit mAb	Cell Signaling Technology (2118S)
HA	HA-tag (C29F4) Rabbit mAb	Cell Signaling Technologies (3724S)
HA	anti-HA.11 Epitope tag (Mouse)	BioLegend (901502)
HSF1	HSF1-antibody (Rabbit)	Cell Signaling Technology (4356S)

Table 13: Used Secondary Antibodies.

Description	Company (Cat. #)
Alexa Fluor 405 goat anti-mouse IgG (H+L)	Invitrogen (A31553)
Alexa Fluor 405 goat anti-rabbit IgG (H+L)	Invitrogen (A31556)
Alexa Fluor 488 goat anti-mouse IgG (H+L)	Invitrogen (A11029)
Fluorescein goat anti-rabbit IgG (H+L)	Invitrogen (F2765)
Cyanine5 goat anti-rabbit IgG (H+L)	Invitrogen (A10523)
Anti-rabbit IgG (H+L) DyLight™ 800 4X PEG Conjugate	Cell Signaling (5151S)

7.1.11. Oligonucleotides

All oligonucleotides were synthesized by Sigma-Aldrich.

Table 14: Used Oligonucleotides for Cloning

Name	Sequence (5' → 3')
o734	TTGATGCCTGGCAGTTCCCT
o735	CGAACCGAACAGGCTTATGT
o2038	CCCTTCTTCTGGCTCTTTGCCAATGTGGTGGCCATGGGCG
o2039	CCATGGCCACCACATTGGCAAAGAGCCAGAAGAAGGGGCG
o2071	AAACGCAAAGTTGGGCGCGCCAACCATGACCAGGAATTTGACC
o2072	GAACGTCGTACGGGTAGTTAATAAGAGGAAGTGAGTTTTGAG
o2175	GATTGGAGGCAGTCCCTAGAATGACCTCTCCATTGTCAACC
o2176	GACAATGGAGAGGTCATTCTAGGGACTGCCTCCAATCACC
o2253	GGCTAGCGCCATGGCTAGCCCGAAAAAGAAACGCAAAGTTG
o2254	CTTTTTCGGGCTAGCCATGGCGCTAGCCAGCTTGGGTCTCC
o2330	ACCGATTCCACCATCCGGGCGGTGTTTCGTCCTTTCCAC
o2331	GGGCGGTATTGCTCTTCCATGGTGGCAAGCTTCCGTGCAG
o2332	CTGCACGGAAGCTTGCCACCATGGAAGAGCAATACCGC
o2333	GTCGAGGCTGATCAGCGGGTTTAGCCAACGACCAGATTG
o2334	CAATCTGGTCGTTGGCTAAACCCGCTGATCAGCCTCGAC
o2335	CAAGTCCCGCTCCGGGTACCATTTTTCGGGGAAATGTGCGC
o2336	GCCCGGATGGTGGAATCGGTAGACACAAGGGATTCTAAATCCCTCGGCGTTCGCGCT GTGCGGGTTCAAGTCCCGCTCCGGGTACCA
o2337	TGGTACCCGGAGCGGGACTTGAACCCGCACAGCGCGAACGCCGAGGGATTTAGAAT CCCTTGTGTCTACCGATTCCACCATCCGGGC
o2433	ACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGC
o2434	CCAATCCTCCCCCTTGCTGTCCTGCCCCACCCCACCCCCCAGAATAGAAT
o2435	ATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTT
o2436	AGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCC
o2475	CTACCCGCCCGGACGCCTTTATGGGTTGTAC
o2476	ACAACCCATAAAGGCGTCCGGGCGGGTAGTG
o3352	CTGACGTCGACGGATCGGGAGCGGCCGCTTCGAGCAGACA
o3355	TGTCTGCTCGAAGCGGCCGCTCCCGATCCGTCGACGTCAG
o3569	GGAATATTTTGCTTGTGTGATTAACTACCCGTACGACGTTC
o3570	GAACGTCGTACGGGTAGTTAATCACACAAGCAAAATATTCC
o3574	ATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGG
o3575	ACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTG

Name	Sequence (5' → 3')
o3670	CAGTCCCTGCAATGACCTCTgCATTGTCAACCCTGCCCGC
o3671	GCGGGCAGGGTTGACAATGcAGAGGTCATTGCAGGGACTG
o3672	GCAGTCCCTagAATGACCTCTgCATTGTCAACCCTGCCCG
o3673	CGGGCAGGGTTGACAATGcAGAGGTCATTctAGGGACTGC
o3674	GCCTCTTCTTTGAGTTCTACCaCCTCCTGCATGATGCGCG
o3675	CGCGCATCATGCAGGAGGtGGTAGAACTCAAAGAAGAGGC
o3676	GTGACAAGAGGGACATCTCGCaATTTCTTGAGTCTAACCCC
o3677	GGGGTTAGACTCAAGAAATtGCGAGATGTCCCTCTTGTCAC
o3678	CAAGAGGGACATCTCGgGATTTCTTGAGTCTAACCCCGTG
o3679	CACGGGGTTAGACTCAAGAAATCcCGAGATGTCCCTCTTG
o3680	GCAAAGTGAGGACCATTACCAtgAGGTCAAACTCTATAAAG
o3681	CTTTATAGAGTTTGACCTcaTGGTAATGGTCCTCACTTTGC
o3684	GTGAGGACCATTACCACCgccTCAAACTCTATAAAGCAG
o3685	CTGCTTTATAGAGTTTGAggcGGTGGTAATGGTCCTCAC
o3686	GACGTCTCCAACATGAGCtGCTTGGCGAGGCAGAGACTGCT
o3687	AGCAGTCTCTGCCTCGCCAAGCaGCTCATGTTGGAGACGTC
o3688	GACGTCTCCAACATGAGCCaCTTGGCGAGGCAGAGACTGCT
o3689	AGCAGTCTCTGCCTCGCCAAGtGGCTCATGTTGGAGACGTC
o3690	CTGCTGGGCCGATCGTcGAGCGTGCCGGTCATCCGCCACC
o3691	GGTGGCGGATGACCGGCACGCTCgACGATCGGCCCAGCAG
o3772	CCGTCCACTACACAGACGTgTCCAACATGAGCCGCTTGG
o3773	CCAAGCGGCTCATGTTGGAcACGTCTGTGTAGTGGACGG
o3774	GTCACAAATGTCGTGAGGAGgGACGTGGAGAAATGGGG
o3775	CCCCATTTCTCCACGTCcCTCCTCACGACATTTGTGAC

Table 15: Used Oligonucleotides for Bisulfite PCR.

Name	Sequence (5' → 3')
o3266_fw SatIII Pyroseq.	GGAATGGATTTAATTTGAATG
o3439_rv SatIII Pyroseq.	[Btn]TTCCATTCCATTCCTATACT
o3268_fw SatIII Sanger	CTTCCTGGCACGAGGGAATGGATTTAATTTGAATG
o3269_rv SatIII Sanger	GAAACAGCTATGACTTCCATTCCATTCCTATACT
o2348_fw BRCA1 Sanger	CTTCCTGGCACGAGATTGGGTGGTTAATTTAGAG

Name	Sequence (5' → 3')
o2350_fw BRCA1 Sanger	CTTCCTGGCACGAGGGGAATTATAGATAAATTAA
o2220_rv BRCA1 Sanger	CCAATACCCCAAAACATCAC
o2339_fw p16_Part 1	CTTCCTGGCACGAGGGGTAGGTGGGGAGGAGTTTAG
o2344_fw p16_Part 1	CTTCCTGGCACGAGGGTGGGGAGGAGTTTAGT
o2199_rv p16_ Part 1	ATATCTTTCCAAACAAAAAAAC
o2340_fw p16_ Part 2	CTTCCTGGCACGAGGGGGGAGATTTAATTTGGGG
o2201_rv p16_ Part 2	AACCCCTCCTCTTTCTTCCTCC
o2344_fw p16 Sanger	CTTCCTGGCACGAGGGTGGGGAGGAGTTTAGT
o2800_rv p16 Sanger	CTATCCCTCAAATCCTC
o2363_fw Sanger 2nd PCR/Barcode	ATCACGCTTCCTGGCACGAG
o2370_fw Sanger 2nd PCR/Barcode	ACTTGACTTCCTGGCACGAG
o2366_fw Barcode	TGACCACTTCCTGGCACGAG
o2406_fw Barcode	TATAATCTTCCTGGCACGAG
o2407_fw Barcode	TCATTCCTTCCTGGCACGAG
o2410_fw Barcode	TCGGCACTTCCTGGCACGAG
o2364_fw Barcode	CGATGTCTTCCTGGCACGAG
o2365_fw Barcode	TTAGGCCTTCCTGGCACGAG
o2367_fw Barcode	ACAGTGCTTCCTGGCACGAG
o2368_fw Barcode	GCCAATCTTCCTGGCACGAG
o2369_fw Barcode	CAGATCCTTCCTGGCACGAG
o2371_fw Barcode	GATCAGCTTCCTGGCACGAG
o2372_fw Barcode	TAGCTTCTTCCTGGCACGAG
o2373_fw Barcode	GGCTACCTTCCTGGCACGAG
o2374_fw Barcode	CTTGTACTTCCTGGCACGAG
o2375_fw Barcode	AGTCAACTTCCTGGCACGAG
o2376_fw Barcode	AGTTCCCTTCCTGGCACGAG
o2377_fw Barcode	ATGTCACTTCCTGGCACGAG
o2378_fw Barcode	CCGTCCCTTCCTGGCACGAG
o2379_fw Barcode	GTAGAGCTTCCTGGCACGAG
o2380_fw Barcode	GTCCGCCTTCCTGGCACGAG
o2381_fw Barcode	GTGAAACTTCCTGGCACGAG
o2382_fw Barcode	GTGGCCCTTCCTGGCACGAG
o2383_fw Barcode	GTTTCGCTTCCTGGCACGAG

Name	Sequence (5' → 3')
o2384_fw Barcode	CGTACGCTTCCTGGCACGAG
o2385_fw Barcode	GAGTGGCTTCCTGGCACGAG
o2386_fw Barcode	GGTAGCCTTCCTGGCACGAG
o2387_fw Barcode	ACTGATCTTCCTGGCACGAG
o2388_fw Barcode	ATGAGCCTTCCTGGCACGAG
o2389_fw Barcode	ATTCCTCTTCCTGGCACGAG
o2390_fw Barcode	CAAAAGCTTCCTGGCACGAG
o2391_fw Barcode	CAACTACTTCCTGGCACGAG
o2392_fw Barcode	CACCGGCTTCCTGGCACGAG
o2399_fw Barcode	CGGAATCTTCCTGGCACGAG
o2400_fw Barcode	CTAGCTCTTCCTGGCACGAG
o2401_fw Barcode	CTATACCTTCCTGGCACGAG
o3033_rv Barcode	ATCACGGAAACAGCTATGAC
o3034_rv Barcode	CGATGTGAAACAGCTATGAC
o3035_rv Barcode	TTAGGCGAAACAGCTATGAC
o3036_rv Barcode	TGACCAGAAACAGCTATGAC
o3037_rv Barcode	ACAGTGGAAACAGCTATGAC
o3038_rv Barcode	GCCAATGAAACAGCTATGAC
o3039_rv Barcode	CAGATCGAAACAGCTATGAC
o3040_rv Barcode	ACTTGAGAAACAGCTATGAC
o3041_rv Barcode	GATCAGGAAACAGCTATGAC
o3042_rv Barcode	TAGCTTGAAACAGCTATGAC
o3043_rv Barcode	GGCTACGAAACAGCTATGAC
o3044_rv Barcode	CTTGTAGAAACAGCTATGAC
o3045_rv Barcode	AGTCAAGAAACAGCTATGAC
o3046_rv Barcode	AGTTCCGAAACAGCTATGAC
o3047_rv Barcode	ATGTCAGAAACAGCTATGAC
o3048_rv Barcode	CCGTCCGAAACAGCTATGAC
o3049_rv Barcode	GTAGAGGAAACAGCTATGAC
o3050_rv Barcode	GTCCGCGAAACAGCTATGAC
o3051_rv Barcode	GTGAAAGAAACAGCTATGAC
o3052_rv Barcode	GTGGCCGAAACAGCTATGAC
o3053_rv Barcode	GTTTCGGAAACAGCTATGAC
o3054_rv Barcode	CGTACGGAAACAGCTATGAC
o3055_rv Barcode	GAGTGGGAAACAGCTATGAC
o3056_rv Barcode	GGTAGCGAAACAGCTATGAC
o3057_rv Barcode	ACTGATGAAACAGCTATGAC

Name	Sequence (5’ → 3’)
o3058_rv Barcode	ATGAGCGAAACAGCTATGAC
o3059_rv Barcode	ATTCCTGAAACAGCTATGAC
o3060_rv Barcode	CAAAAGGAAACAGCTATGAC
o3061_rv Barcode	CAACTAGAAACAGCTATGAC
o3062_rv Barcode	CACCGGGAAACAGCTATGAC
o3069_rv Barcode	CGGAATGAAACAGCTATGAC
o3070_rv Barcode	CTAGCTGAAACAGCTATGAC
o3071_rv Barcode	CTATACGAAACAGCTATGAC

7.1.12. Plasmids

All used plasmids have an encoded ampicillin resistance. Only exception is pJaW824 which has an encoded kanamycin resistance. Plasmids pDaS49, pDaS273, pJaW824, pJaW891, pJaW911, pJaW915, pJaW1038, pJaW1131, pStH1147 and pJaW2557 were only used for cloning. All other plasmids were used in experiment involving mammalian cell lines.

Table 16: Used Plasmids.

Lab Code	Description and Important Genes	Similarity
pDaS49	LeuRS-BH5 and $tRNA^{Leu}$(CUA) genes	---
pDaS273	PylRS-AF and $tRNA^{Pyl}$(CUA) (source: Edward Lemke)	pStH1169
pJaW824	DNMT3a3L gene (source: Addgene #71827)	---
pAnW891	GG2-entry vector with transfection control for TALE-VP64	pJaW911
pJaW911	GG2-entry vector with transfection control for TALE-DNMT3a3L	pJaW911
pJaW915	GG2-entry vector with transfection control for TALE-inactiveDNMT3a3L	pJaW911
pJaW926	pJaW911 with p16-DNMT	pJaW926
pJaW935	pJaW915 with inactive p16-$DNMT^{E756A}$	pJaW926
pJaW1002	pJaW911 with BRCA1-DNMT	pJaW926
pJaW1038	pJaW911 with pcDNMT3a3L ($DNMT3a3L^{C710amber}$)	pJaW911
pJaW1041	pJaW915 with inactive pcDNMT3a3L ($DNMT3a3L^{C710amber, E756A}$)	pJaW911
pJaW1125	pJaW1038 with p16-pcDNMT ($DNMT3a3L^{C710amber}$)	pJaW926
pJaW1131	DNMT3a3L with transfection control	pJaW1131
pStH1147	LeuRS-BH5 and $tRNA^{Leu}$(CUA)	pStH1169
pStH1169	pStH1147 with LeuRS-$BH5^{T252A}$	pStH1169
pJaW1559	pJaW1131 with pcDNMT3a3L ($DNMT3a3L^{C710amber}$)	pJaW1131
pJaW1648	pJaW911 with SatIII-DNMT	pJaW926
pJaW1654	pJaW911 with inactive SatIII-$DNMT^{E756A}$	pJaW926
pJaW1660	pJaW911 with SatIII-pcDNMT ($DNMT3a3L^{C710amber}$)	pJaW926
pJaW1666	pJaW911 with inactive SatIII-pcDNMT^{E756A} ($DNMT3a3L^{C710amber, E756A}$)	pJaW926
pJaW1848	pJaW1131 w/o transfection control	pJaW1848
pJaW1849	pJaW1559 w/o transfection control	pJaW1848
pJaW1850	pJaW1848 with inactive $DNMT3a3L^{E756A}$	pJaW1848
pJaW1851	pJaW1849 with inactive $DNMT3a3L^{E756A}$	pJaW1848
pAlH1895	pJaW1848 with DNMT3aCD	pAlH1895
pAlH1898	pJaW1848 with DNMT3a3L S714C	pJaW1848
pAlH1899	pJaW1848 with DNMT3a3L R736H	pJaW1848

Lab Code	Description and Important Genes	Similarity
pAlH1900	pJaW1848 with DNMT3a3L R771Q	pJaW1848
pAlH1901	pJaW1848 with DNMT3a3L R771G	pJaW1848
pAlH1902	pJaW1848 with DNMT3a3L T835M	pJaW1848
pAlH1903	pJaW1848 with DNMT3a3L R836W	pJaW1848
pAlH1904	pJaW1848 with DNMT3a3L R836A	pJaW1848
pAlH1905	pJaW1848 with DNMT3a3L R882C	pJaW1848
pAlH1906	pJaW1848 with DNMT3a3L R882H	pJaW1848
pAlH1907	pJaW1848 with DNMT3a3L W893S	pJaW1848
pJaW1910	pJaW1038 w/o transfection control	pJaW1910
pJaW1928	pJaW1910 with inactive DNMT3a3L^{E756A}	pJaW1910
pJaW1976	pJaW1910 w/o BsmBI sites in DNMT3a3L	pJaW1910
pJaW1977	pJaW1928 w/o BsmBI sites in DNMT3a3L	pJaW1910
pJaW1978	pJaW1849 with pcDNMT3a3L S714C	pJaW1848
pJaW1979	pJaW1849 with pcDNMT3a3L R736H	pJaW1848
pJaW1980	pJaW1849 with pcDNMT3a3L R771Q	pJaW1848
pJaW1981	pJaW1849 with pcDNMT3a3L R771G	pJaW1848
pJaW1984	pJaW1849 with pcDNMT3a3L R836A	pJaW1848
pJaW1998	pJaW1976 with SatIII-pcDNMT (DNMT3a3L^{C710amber})	pJaW1998
pJaW1999	pJaW1977 with inactive SatIII-pcDNMT^{E756A} (DNMT3a3L$^{C710amber, E756A}$)	pJaW1998
pJaW2557	pJaW1895 with transfection control	pJaW2557
pJaW2559	pJaW2557 with pcDNMT3aCD (DNMT3a3L^{C710amber})	pJaW2557

7.1.13. TALEs

DNA sequence and its length includes 5'-T. Shown RVDs don't include this thymine.

Table 17: Sequence of used TALEs.

Gene	Length	DNA Sequence	RVDs for Assembly
p16	25	TCCTCCTTCCTTGCCAACGCTGGCT[160]	HD HD NG HD HD NG NG HD HD NG NG NN HD HD NI NI HD NN HD NG NN NN HD NG
BRCA1	26	TCAGGAGGCCTTCACCCTCTGCTCTG	HD NI NN NN NI NN NN HD HD NG NG HD NI HD HD HD NG HD NG NN HD NG HD NG NN
SatIII	20	TGATTCCATTCCATTCCATT[332]	NN NI NG NG HD HD NI NG NG HD HD NI NG NG HD HD NI NG NG

7.1.14. Software

Table 18: Used Software.

Program	Company
BioDoc Analyze	Analytik Jena AG
Cell Sorter Software	Sony Biotechnology
CFX Manager	Bio-Rad
ChemBioDraw	Cambridgesoft
Citavi	Swiss Academic Software GmbH
ImageJ	National Institutes of Health
MaxQuant	[366]
MestReNova	Mestrelab Research
Office 365 ProPlus	Microsoft
Photoshop	Adobe
PSQ HS 96A	Qiagen
Pymol	Schrödinger LLC
RStudio	RStudio
SnapGene	GSL Biotech

7.2. Methods

Some of the described methods were already published in Muñoz-López et al.[332], Palei et al.[367], or Wolffgramm et al.[103]

7.2.1. Biomolecular Methods

7.2.1.1. Preparation of chemically competent *E. coli* cells

20 mL LB-medium was inoculated with a single colony of GH371 or DH5α *E. coli* cells and incubated overnight at 37 °C and 180 rpm. 8 mL of this culture were transferred into 800 mL LB-medium and incubated at 37 °C and 180 rpm until an OD_{600} of 0.4 was reached. The culture was stored on ice for 30 min and then aliquoted in 50 mL tubes and pelleted at 3500 xg at 4 °C for 10 min. Pellets were resuspended in 10 mL ice-cold 100 mM $MgCl_2$ buffer and centrifuged again as before. Pellets were resuspended in 10 mL ice-cold 50 mM $CaCl_2$ buffer and stored on ice for 30 min. After another centrifugation step, cells were resuspended in 4 mL *E. coli* storage buffer (50 mM $CaCl_2$ + 15 % (v/v) glycerol), aliquoted into 25 µL or 50 µL, shock frozen in liquid nitrogen and stored at -80 °C for long term storage.

7.2.1.2. Transformation of chemically competent *E. coli* cells

To amplify plasmid DNA, $CaCl_2$-treated chemically competent GH371 or DH5α *E. coli* cells were used for transformation. Aliquots of 25 µL or 50 µL were thawed slowly on ice and up to 5 µL DNA was added and mixed gently by pipetting. The amount of DNA was dependent on the application with low volumes for retransformations of pure plasmid DNA and high volumes for cloning approaches, especially Site-Directed Mutagenesis (7.2.1.6). Cells were incubated on ice for 30 min and then a heat shock at 42 °C for 45 s was performed. After another 2 min on ice, cell mixture was transferred into 500 µL or 1 mL (depending on the usage of 25 µL or 50 µL cell suspension, respectively) 37 °C-prewarmed SOC medium and incubated for 1 h at 37 °C and 600 rpm. In the meantime, LB-agar plates with the respective antibiotic were prewarmed to 37 °C. Depending on the application, different volumes of the transformation mixture were added onto the LB-agar plates and distributed equally using glass beads. As before, a retransformation required less and cloning approaches higher volumes with up to 1 mL. In the latter case, cells were centrifuged 3 min at 400 xg, 800 µL supernatant was discarded, and cells were resuspended in the remaining SOC medium. Finally, LB-agar plates were incubated overnight at 37 °C.

7.2.1.3. Overnight Culture and Isolation of Plasmid DNA

6 µL of the respective 1000x antibiotic stock solution was added to 6 mL LB-medium in 15 mL tubes and mixed thoroughly. If many samples were prepared, a master mix was done in beakers and distributed equally to the 15 mL tubes and inoculated with a single bacterial colony from LB-agar plates. After incubation at 37 °C and 180-200 rpm overnight, cells were pelleted at 4000 rpm for 10 min. For cell lysis and plasmid DNA purification, the "GeneJET Plasmid MiniPrep" kit or "NucleoSpin Plasmid EasyPure" kit was used according to the manufactures protocol. Elution from columns was performed with 50 µL H_2O and the plasmid DNA concentration was measured photometrically (7.2.1.4). For long term storage, plasmid DNA was diluted 1:10 in DNA storage buffer and stored at -20 °C.

If larger amounts of plasmid DNA were needed, a MaxiPrep was performed. For this, 250 mL LB-medium was mixed with the respective antibiotic and inoculated with a single bacterial colony from LB-agar plates. After overnight incubation at 37 °C and 180 rpm, cells were pelleted at 12000 xg at 4 °C for 20 min. Isolation of plasmid DNA was performed according to the manufactures protocol of "NucleoBond Xtra Maxi Plus Kit". Final elution was done with 1 mL TRIS buffer of the kit.

7.2.1.4. Concentration Measurement of DNA and RNA

Concentrations of isolated DNA (gDNA or plasmid DNA) or RNA were measured photometrically using the NanoDrop2000. First, a blanking step was performed with the solvent. Then, 1 µL solution was placed on the middle of the measuring module and the concentration was determined by the NanoDrop2000 software. Since the absorption was measured from 220 nm to 350 nm, the ratios of absorption A260/A280 and A260/A230 gave additional information about the purity.[368]

7.2.1.5. Isolation of Plasmid DNA from Addgene Stocks

To isolate plasmid DNA from Addgene bacterial agar stocks, cells were streaked on a LB-agar plate with the respective antibiotic using an inoculation loop leading to the growth of isolated colonies. After incubation at 37 °C overnight, a single colony was used to inoculate an overnight culture for subsequent plasmid DNA-isolation (7.2.1.3).

7.2.1.6. Site-Directed Mutagenesis (Quickchange)

Site-directed mutagenesis was performed by whole plasmid PCR based on the "QuickChange" protocol from Agilent Technologies (see instruction manual of product #200518). For this, 35 to 45 bases long primers covering the site to be mutated contained the desired mutation in the center of their sequence. Forward and reverse primers were

either fully complementary to each other or shifted by 4 bases. Using Pfu polymerase, the following protocol was performed:

Table 19: Reaction mixture for site-directed mutagenesis.

Component	Volume
Pfu Reaction Buffer (10x)	5 µL
Template	0.5 µL
Forward Primer (10 µM)	2.5 µL
Reverse Primer (10 µM)	2.5 µL
dNTPs (10 mM)	1.5 µL
Pfu Polymerase	1 µL
H_2O	37 µL

Table 20: Thermocycler protocol for site-directed mutagenesis.

Step	Temperature	Time	Repeats
Denaturation	95 °C	30 s	1x
Denaturation	95 °C	30 s	12 – 18x
Annealing	55 °C	1 min	
Elongation	68 °C	2 min/1000 bases	
Hold	4 °C	∞	

The cycle repeats were dependent on the number of mutated bases. Specifically, 12 repeats were used for point mutations, 16 repeats for single amino acid changes and 18 repeats for multiple amino acid changes. Since the PCR mixture still contained template plasmid DNA without mutations, a DpnI-digestion was performed by adding 5 µL NEB Cutsmart buffer, 1 µL DpnI and 4 µL H_2O to 40 µL PCR mixture. Digestion was done for at least 3 h at 37 °C. Then, 5 µL were used for transformation in GH371 or DH5α cells (7.2.1.2) and 1 mL was plated onto LB-agar plates with the respective antibiotic. Single colonies were picked for overnight cultures (7.2.1.3) and cloning success was checked by DNA sequencing (7.2.1.12).

7.2.1.7. Gibson Assembly

Gibson assembly was based on a previous published protocol[369]. Linear DNA fragments with compatible ends were generated by either PCR amplification using high-fidelity polymerases like Q5 or Phusion, or by digestion with restriction enzymes. A total volume of 5 µL of these fragments were mixed with 15 µL 1.33x Gibson master mix and incubated for 1 h at 50 °C. The molar ratio of fragments was based on their length with higher molarity for smaller fragments. 5 µL reaction mixture was used for transformation in GH371 or DH5α

cells (7.2.1.2) and 1 mL was plated onto LB-agar plates with the respective antibiotic. Single colonies were picked for overnight cultures (7.2.1.3) and cloning success was checked by DNA sequencing (7.2.1.12).

If a short fragment (e.g. tRNA) was needed for cloning, two artificial oligonucleotides were hybridized. For this, 40 µL of each 10 mM oligonucleotide were mixed with 10 µL NEB Cutsmart buffer and 10 µL H_2O and incubated at 95 °C for 5 min. The mixture cooled down slowly by placing the reaction tube at RT and it was subsequently purified (7.2.1.10).

7.2.1.8. TALE Assembly

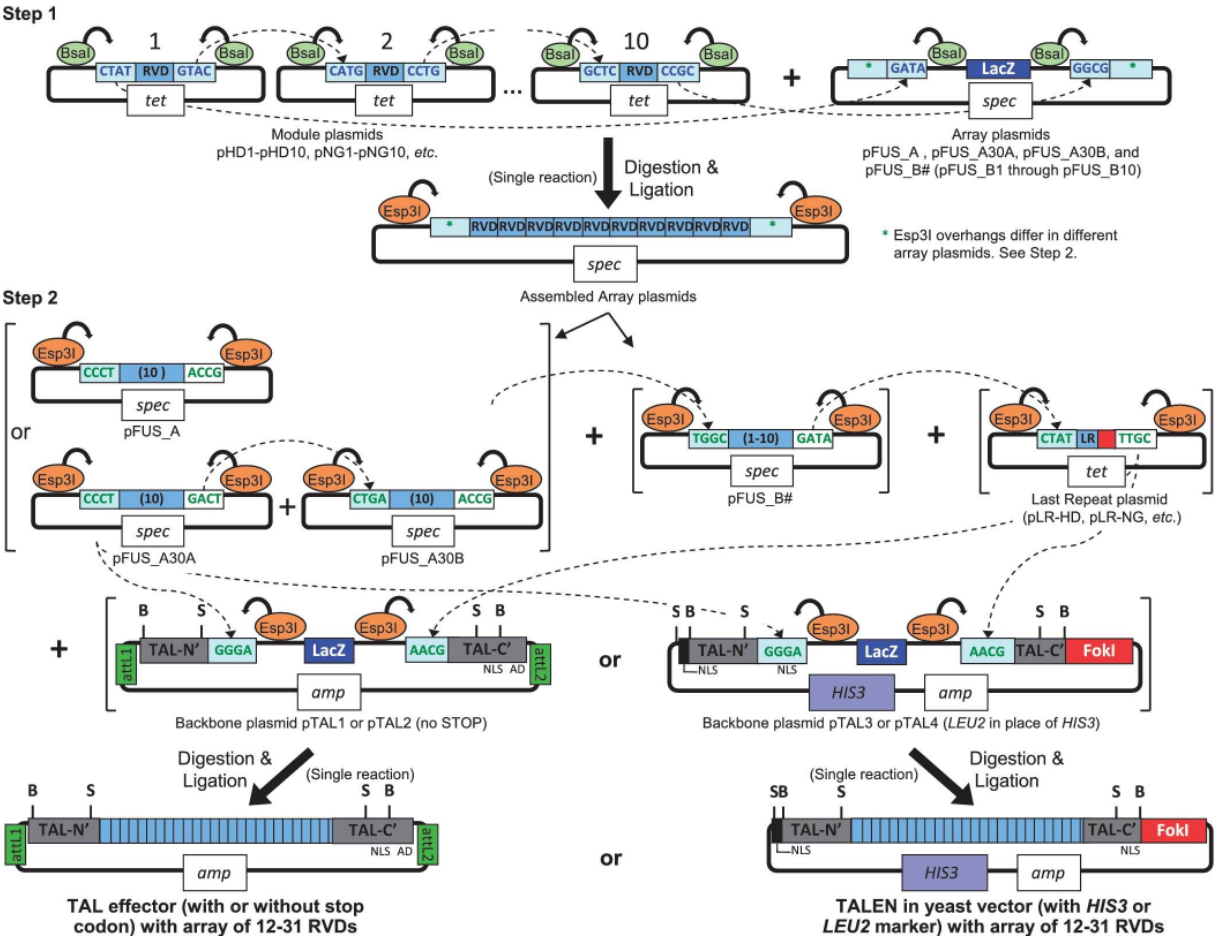


Figure 34: Schematic overview of TALE assembly. In this thesis, different backbone plasmids were used. In those, FokI was replaced by DNMT3a3L or pcDNMT3a3L, and they had additional NLSs and tags (see plasmid maps (page 138) and protein sequences (page 130)). Golden Gate 1-reaction is described in step 1 followed by the Golden Gate 2-reaction in step 2. This figure is adapted from [157].

TALE assembly was performed based on a previously published protocol by Cermak et al.[157] which is schematically shown in Figure 34. There are two independent steps: First, the array plasmids are assembled in the Golden Gate 1 (GG1)-reaction which are then used for incorporation into the backbone plasmid in the Golden Gate 2 (GG2)-reaction. In GG1, up to 10 repeats were cloned from the module plasmids into the respective

array plasmid. For the p16- and BRCA1-TALE in this thesis the first 10 repeats were cloned into the pFUS_A30A vector, the next 10 repeats into the pFUS_A30B vector, and the rest except the last repeat into the corresponding pFUS_B vector. For the SatIII-TALE, the first 10 repeats were cloned into the pFUS_A vector and the rest except the last repeat into the corresponding pFUS_B vector (GG1 cloning of SatIII-TALE was done by Dr. Álvaro Muñoz-López). For GG1-reactions, a protocol of 10 cycles was performed with 5 min 37 °C and 10 min 16 °C plus final incubation steps at 50 °C and heat inactivation at 80 °C for 5 min each, using 5 U BsaI and 200 U T4 DNA ligase in a 10 µL total volume mixture. To digest linear DNA, 0.5 µL 10 mM ATP and 0.5 µL Plasmid-Safe DNase were added to the reaction mixture and incubated at 37 °C for 60 min followed by a heat inactivation at 70 °C for 20 min. 2.5 µL of this reaction was then used for transformation in GH371 or DH5α cells using Spec/X-Gal plates (7.2.1.2). Cloning success was examined by cPCR of single white colonies using oligonucleotides o734 and o735 as PCR-primers (7.2.1.11). Colonies showing bands with the expected size on an agarose gel (7.2.1.9) were used for overnight cultures (7.2.1.3). Subsequent DNA sequencing confirmed the correct order of RVDs in the GG1-plasmids (7.2.1.12). In the GG2 step, the previous assembled GG1 parts were cloned together with the last repeat (LR) onto the backbone plasmid. A similar protocol as in GG1-reactions with 10 cycles of 10 min 37 °C and 15 min 16 °C, followed by final incubation at 37 °C for 15 min and 80 °C for 5 min was performed, using 5 U BsmBI and 200 U T4 DNA ligase. The backbone plasmids were designed in a way that DNMT3a3L was expressed C-terminally as a fusion protein with the assembled TALE (see plasmid maps on page 138). 2.5 µL were used for transformation in GH371 or DH5α cells using Carb/X-Gal plates. White colonies were used for overnight cultures and subsequent sequencing without additional cPCR.

7.2.1.9. Agarose Gel Analysis and Purification

In general, 1 % agarose gels were used except for analysis of bisulfite-PCRs in which 2 % agarose gels were used. Agarose powder was molten in 0.5x TBE buffer in a microwave and cooled down slightly to prevent deformation of the gel casting chamber. In such, liquefied agarose gel was poured, the respective comb was inserted, and the gel cooled down to RT. The gel was then placed into the running chamber and covered fully with 0.5x TBE buffer. 1 µL loading buffer (1:20 dilution of 3 % glycerol mixed with bromophenol blue and xylene cyanol in H_2O) was mixed with up to 10 µL sample and loaded onto the gel. 2-log ladder was used as reference. For 1 % gels, a voltage of 80 V for 50 min, and for 2 % gels, 90 V for 60 min was used. Afterwards, staining was done with 0.1 % (w/v) ethidium bromide and visualized by UV light.

For purification of PCR products, up to 50 µL sample was loaded into one large well. Staining was done as described before but only weak UV light was used for visualization to reduce DNA damage. All other steps were performed as described before.

7.2.1.10. Purification of PCR Products

If there were multiple PCR products in a PCR reaction, the product of interest had to be isolated by agarose gel purification (7.2.1.9). The band was cut out of the gel using gloves and a UV protection shield for the eyes. DNA was isolated following the manufactures protocol of “GeneJET Gel Extraction Kit” or “NucleoSpin Gel and PCR Clean-up“ kit. Final elution was done with 20 – 50 µL H_2O.
For hybridized oligonucleotides or if the PCR product was without byproducts, DNA purification was directly performed using the solution and the previous mentioned kits. Final elution was done with 20 – 50 µL H_2O.

7.2.1.11. Colony PCR

Colony PCR (cPCR) was performed with transformed GH371 or DH5α cells after GG1 reactions or cloning approaches to check if the plasmid DNA contained the gene of interest. Single colonies from a LB-agar plate were picked and resuspended in 10 µL H_2O each. To save the single clones, 1 µL of each cell suspension was transferred to a specific labeled spot on a new LB-agar plate with the respective antibiotic and incubated at 37 °C for several hours. 4 µL of the same solution was used as template in the PCR reaction with Taq polymerase, dNTPs and specific forward and reverse primers covering the gene of interest. PCR was analyzed on a 1 % agarose gel (7.2.1.9). Clones showing the expected bands were used for overnight cultures taking the leftovers of the cell suspensions or the safeplate for inoculation (7.2.1.3). DNA sequencing was performed to validate the successful cloning (7.2.1.12).

7.2.1.12. Sanger Sequencing

Sequencing of plasmid DNA was either performed using the “LIGHTRUN night express” service by GATC or “Economy Run” service by Microsynth Seqlab. For GATC, up to 500 ng plasmid DNA was mixed with 2.5 µL 10 µM sequencing primer and if necessary, filled up with water to a final volume of 10 µL. For Microsynth Seqlab, up to 1200 ng plasmid DNA was mixed with 3 µL 10 µM sequencing primer and if necessary, filled up with water to a final volume of 15 µL.
To sequence plasmid DNA directly from transformed GH371 or DH5α cells without plasmid DNA isolation, the “Ecoli NightSeq” service by Microsynth Seqlab was used.
Purified bisulfite-PCR products were sequenced with the “Economy Run” service by

Microsynth Seqlab with a final concentration of 4 ng/µL in a 15 µL mixture including 3 µL 10 µM sequencing primer. For further information see 7.2.1.16.

7.2.1.13. Vector Construction

All enzymes were purchased from New England Biolabs. Plasmid maps can be found in the appendix 8.5, page 138.

Golden Gate 2 backbone plasmid pJaW911 with DNMT3a3L was cloned by Gibson assembly (7.2.1.7). As backbone for this cloning, plasmid pAnW891 (cloned by Dr. Anna Witte based on Addgene plasmid #47389) was used which encoded the transfection control mCherry-GFP^{Y39amber} and the transfection activator VP64 with an HA-tag C-terminal to the TALE assembly site. Such VP64 gene was cut out with PacI and AscI and the linearized backbone was mixed with the insert DNMT3a3L which was amplified from plasmid pJaW824 (Addgene #71827) using primers o2071/o2072. Mutations were introduced by Quickchange site directed mutagenesis (7.2.1.6). For the catalytically inactive DNMT3a3L^{E756A} and pcDNMT3a3L (DNMT3a3L^{C710amber}) primer pairs o2038/o2039 or o2175/o2176 were used, respectively, resulting in plasmids pJaW915 (DNMT3a3L^{E756A}), pJaW1038 (pcDNMT3a3L) and pJaW1041 (pcDNMT3a3L^{E756A}).
TALE assembly into the backbone plasmids was done as described before (7.2.1.8) and RVD sequences can be found in section 7.1.13. For SATIII-TALE constructs, all four previous described backbone plasmids were used, resulting in pJaW1648 (SatIII-DNMT), pJaW1654 (inactive SatIII-DNMT^{E756A}), pJaW1660 (SatIII-pcDNMT) and pJaW1666 (inactive SatIII-DNMT^{E756A}). TALEs targeting BRCA1 and p16 genes were assembled into pJaW911 resulting in pJaW926 and pJaW1002, respectively, and for p16 also in pJaW915 (inactive DNMT3a3L^{E756A}) resulting in pJaW935 and pJaW1038 (pcDNMT3a3L) resulting in pJaW1125.
To receive the final *E. coli* leucyl-tRNA-synthetase and suppressor tRNA$_{CUA}$, plasmid pStH1147 was first cloned by Stefan Helmer. For this, the synthetase ecLRS-BH5 with five previously reported mutations M40G, L41Q, Y499L, Y527G, H537F[271] was amplified with primer pair o2332/o2333 from plasmid pescLRSNVOCBH5-4 (pDas49) and the backbone from plasmid pCMV_PylRS_AF (pDaS273) with primer pairs o2330/o2331 and o2334/2335. Further, the tRNA (Leu5$_{CUA}$[270]) was obtained by hybridizing oligonucleotides o2336/o2337 (7.2.1.7), and all four fragments were assembled by Gibson protocol. Further, the mutation T252A[271] was introduced into the synthetase ecLRS-BH5 by Quickchange site directed mutagenesis using primer pair o2475/o2476, resulting in pStH1169 used for all genetic code expansion experiments in this thesis.
For DNMT3a3L without TALE (pJaW1131) the Golden Gate 2 entry site was removed from pJaW911 by a single fragment Gibson assembly. For this, the backbone of pJaW911 was

amplified with primers o2253/o2254. With the same primers as before, amber mutation was introduced resulting in pJaW1559 (pcDNMT3a3L).

To avoid bleed-through in microscopy and flow cytometry the transfection control (mCherry-GFP^{Y39amber}) was removed from plasmid pJaW1038 by a two fragments Gibson assembly (the backbone was amplified with primer pairs o3352/o2434 and o2433/o3355) resulting in plasmid pJaW1910. Inactive E756A mutation was introduced as described before, resulting in plasmid pJaW1928. Additionally, the BsmBI sites in the DNMT3a3L gene were removed to make the GG2 assembly easier. For this, two Quickchange site directed mutagenesis were performed using primer pairs o3772/o3773 and o3774/o3775 resulting in pJaW1976 (as pJaW1910) and pJaW1977 (as pJaW1928). In both, SatIII-TALE was assembled resulting in plasmids pJaW1998 (SatIII-pcDNMT) and pJaW1999 (inactive SatIII-pcDNMT^{E756A}). The transfection control was also removed from the two global DNMT plasmids pJaW1131 and pJaW1559 by Gibson assembly (the backbone was amplified with primer pairs o3352/o2435 and o2436/o3355). On both, the E756A mutation was introduced as described before, resulting in plasmids pJaW1848 (DNMT3a3L), pJaW1850 (inactive DNMT3a3L^{E756A}), pJaW1849 (pcDNMT3a3L) and pJaW1851 (inactive pcDNMT3a3L^{E756A}). For the analysis of DNMT3a mutations, following mutations were introduced on plasmids pJaW1848 or pJaW1849 by Quickchange site directed mutagenesis (cloning was done together with Alicia Hoffmann Benito):

Table 21: Mutagenesis of DNMT3a mutations

Mutation	Primer pair	Resulting plasmid (DNMT3a3L)	Resulting plasmid (pcDNMT3a3L)
S714C	o3670/o3671 (for DNMT3a3L o3672/o3673 (for pcDNMT3a3L)	pAlH1898	pJaW1978
R736H	o3674/o3675	pAlH1899	pJaW1979
R771Q	o3676/o3677	pAlH1900	pJaW1980
R771G	o3678/o3679	pAlH1901	pJaW1981
T835M	o3680/o3681	pAlH1902	---
R836W	o3682/o3683	pAlH1903	---
R836A	o3684/o3685	pAlH1904	pJaW1984
R882C	o3686/o3687	pAlH1905	---
R882H	o3688/o3689	pAlH1906	---
W893S	o3690/o3691	pAlH1907	---

For whole transcriptome analysis, the 3L part from plasmids pJaW1131 and pJaW1848 were deleted by a two fragments Gibson assembly (each plasmid was amplified with primers o3569/o3574 and o3575/o3570) resulting in plasmids pJaW2557 (with transfection control) and pAlH1895 (w/o transfection control). C710amber mutation was introduced in pJaW2557, resulting in pJaW2559 (pcDNMT3aCD).

7.2.1.14. Isolation of Genomic DNA

Genomic DNA (gDNA) was isolated from either untransfected HEK293T, WT HCT116 and DKO HCT116, or sorted HEK293T cells transfected with TALE-DNMT constructs or DNMT3aCD. For untransfected cells, cells of a confluent TC-plate were trypsinized (7.2.2.1.1 and 7.2.2.1.2), counted, and no more than $5 \cdot 10^6$ cells were transferred into a 15 mL tube containing 10 mL DPBS. After centrifugation at 500 xg for 4 min, the supernatant was discarded, and cells were resuspended in 200 µL DPBS, and 20 µL proteinase K from the “QIAamp DNA Mini Kit” was added. Further steps were executed as described from step 3 on in the “Protocol: DNA Purification from Blood or Fluids (Spin Protocol)” of the kit. Final elution was done with 200 µL H_2O with a previous 5 min incubation for 5 min at RT. Sorted cells (7.2.2.6.1) were kept in sheath fluid (PBS) on ice and then centrifuged at 500 xg for 4 min. The supernatant was removed to a residual volume of 200 µL and isolation of gDNA was done as described before. If the volume of sorted cells was less than 200 µL, DPBS was added to a final volume of 200 µL and the protocol was continued from the Proteinase K step on. Final elution was also done in 200 µL H_2O. If only a small number of cells was sorted, concentration was increased by decreasing the volume in an Eppendorf Concentrator to ~20 µL. Isolated gDNA was stored at -20 °C.

7.2.1.15. Bisulfite Conversion and Bisulfite PCR

Isolated gDNA was bisulfite-converted using the “EpiTect Bisulfite Kit” following the manufactures protocol. Before elution with 20 µL EB-buffer, the column was incubated for 1 min at RT. If high concentrations were expected, another elution step with 20 µL EB-buffer was performed into the same 1.5 mL collection tube. DNA concentrations were measured, and the converted DNA was stored at -20 °C.
For bisulfite PCR primers see 7.1.11 and for Sanger sequencing with bisulfite PCR products see 7.2.1.16.
The “EpiTect MSP Kit” was used for PCR with bisulfite-converted DNA. Template was mixed with primers and RNase-free water. Then, MSP master mix was added, and the cycling protocol was executed.

Table 22: Reaction mixture for bisulfite PCR.

Component	Volume
MSP Master Mix	12.5 µL or 25 µL
Template	0.5 µL or 1 µL
Forward Primer [10 µM]	0.75 µL or 1.5 µL
Reverse Primer [10 µM]	0.75 µL or 1.5 µL
RNase-free H_2O	10.5 µL or 21 µL

Table 23: Thermocycler protocol for bisulfite PCR.

Step	Temperature	Time	Repeats
Denaturation	95 °C	10 min	1x
Denaturation	94 °C	15 s	40x
Annealing	55 °C	30 s	
Elongation	72 °C	30 s	
Elongation	72 °C	10 min	1x
Hold	4 °C	∞	

Of this reaction 1 µL unpurified sample was used as template for either nested PCR reactions (with EpiTect MSP Kit) or a PCR with Taq polymerase to add the universal barcodes at the 5'-end of the amplicons for NGS or to get higher amounts of amplicons for Sanger sequencing. The barcode PCR was performed in a 100 µL volume and subsequently purified and eluted in 50 µL H_2O (7.2.1.10).

Since bisulfite conversion was used as the main method to detect 5mC in this thesis, it is described in more detail in the following. 5mC can be analyzed strand specifically by bisulfite-PCR (BS-PCR) due to a different reactivity of C and 5mC with sodium bisulfite.[370,371,372] Cytosines in single stranded DNA react with sodium bisulfite and form a 5,6-dihydrocytosine-6-sulfonate intermediate (Figure 35).[371] Following a deamination at acid pH, the bisulfite of the formed uracil sulphonate is eliminated at alkaline pH (Figure 35).[371,372]

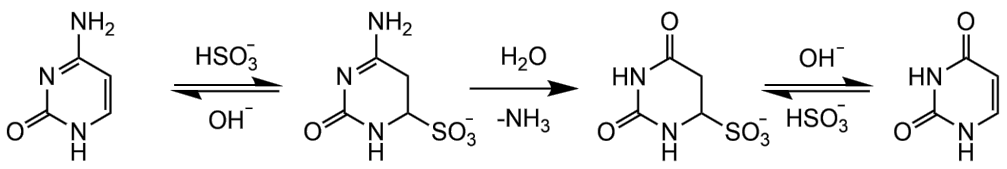


Figure 35: Reaction of cytosine with sodium bisulfite. Reaction of cytosine with sodium bisulfite leads to a cytosine sulfonate intermediate which is getting deaminated in the following step. In alkaline pH, the bisulfite is eliminated resulting in unmodified uracil.

The formed uracil can be amplified by DNA polymerases that tolerate uracil as residues (e.g. Taq) in PCR as thymine.[373] 5mC in single stranded DNA however only reacts poorly with sodium bisulfite and therefore will be amplified as C in PCR.[372] Sequencing of specific amplicons or whole genome amplifications (WGA) produced with bisulfite converted DNA gives information about the original cytosine modification status since original C will be sequenced as T and original 5mC as C.[374] After bisulfite treatment the two DNA strands are no longer complementary and each strand must be amplified separately which allows single-base and strand-specific analysis.[375] For applications like whole genome sequencing an enrichment for CpG rich regions might be beneficial. Reduced representation bisulfite sequencing (RRBS) uses methylation-insensitive restriction enzymes such as MspI and ApeKI to digest gDNA next to CpG sites.[376,377] Small fragments are isolated, bisulfite converted, amplified and sequenced (usually NGS).[376,377] Advanced protocols reduced the required amount of input DNA (laser-capture microdissection RRBS (LCM-RRBS)[378], and single cell RRBS (scRRBS)[379]).

7.2.1.16. Sanger Sequencing with Bisulfite PCR

Analysis of on- and off-target effects by SatIII-pcDNMT and on-target effects by BRCA1-DNMT and p16-DNMT was done with amplified loci from bisulfite-converted gDNA (7.2.1.14 and 7.2.1.15). For amplification of SatIII, BRCA1, and p16 gene, the Qiagen MSP Kit and primer pairs o3268/o3269, o2348/o2220 and o2344/o2800 were used, respectively. In a second, consecutive PCR, the product was used as template and amplified using Taq polymerase, o2363 as forward primer and the previously used reverse primers. After purification (7.2.1.10), amplicons were sequenced (7.2.1.12) using o2363 for SatIII, o2220 for BRCA1 or o2199 for p16. A slightly different protocol was used for amplification of BRCA1 locus methylated with BRCA1-DNMT. The first PCR with primers o2348/o2220 was used as template in a second PCR with Qiagen MSP Kit and primers o2350/o2220. In a third PCR using Taq polymerase and primers o2370/o2220, the second PCR was used as template. Purification and sequencing were performed as described before (7.2.1.10 and 7.2.1.12).

7.2.1.17. Absorption Measurement of DMNB-Cysteine

Decaging of DMNB-Cys was monitored photometrically. 2 mL of 1 mM DMNB-Cys in DPBS in a 3.5 cm TC-plate was placed in the middle of the UV-table (6x 15 W, 365 nm). After different time points (0 s up to 30 min), light was turned off and 80 µL solution was transferred to a 96-well plate (flat bottom, black from Greiner Bio-One). For controls, the same samples were taken from a 3.5 cm TC-plate which was not irradiated with light. Samples with 80 µL DPBS were also added to the plate for background subtraction.

Absorption between 300 nm and 550 nm of each well was measured with the Infinite M1000 plate reader (Tecan). Absorption values were normalized to 550 nm of the “0 s” sample.

7.2.1.18. SDS-PAGE

For Western Blot analysis and mass spectrometry of proteins with prior in-gel digestion, a SDS-PAGE was performed. In a gel casting chamber of the Bio-Rad Mini-PROTEAN Tetra System, a 10 % resolving gel (3535 µL H_2O, 1365 µL Rotiphorese Gel 40, 420 µL Tris (1.5 M, pH 8.8), 66.7 µL SDS (10 %), 66.7 µL APS (10 %) and 6.67 µL TEMED) was poured and covered with isopropanol. After polymerization, the casting chamber was filled up with a 5 % stacking gel (1000 µL H_2O, 208 µL Rotiphorese Gel 40, 420 µL Tris (0.5 M, pH 6.8), 16.7 µL SDS (10 %), 16.7 µL APS (10 %) and 1.67 µL TEMED) and a comb was inserted. Samples were mixed with 4x SDS loading buffer and incubated at 95 °C for 5 min. 10 µL to 25 µL sample were loaded into each well and gels run at 80 V for 15 min and 130 V for 50 min. Gels were transferred into SDS gel staining buffer and heated up briefly in a microwave. After 15 min shaking at RT, buffer was exchanged with SDS gel destaining buffer and heated up again briefly. Further destaining was done in H_2O overnight and subsequent documentation was done with the CanonScan 9000F.

7.2.1.19. Western Blot

HEK293T cells on a 10 cm TC-plate were transfected with pcDNMT3a3L (pJaW1559) + pStH1169 in the presence or absence of 0.5 mM DMNB-Cys for Figure 18A or p1559/p1169 in the presence of 0.05 mM DMNB-Cys and light irradiated for 5 min 24 h later (no light for control, 7.2.2.3) for Figure 18B. For Figure 18A, cells were trypsinized and lysed in 4 mL RIPA buffer (Thermo Scientific™ #89900) + 11.5 µL 0.1 M PMSF 24 h after transfection. The mixture was incubated on ice for 15 min at 400 rpm and then centrifuged at 3200 xg, 4 °C for 15 min. For Figure 18B, 24h after light irradiation, cells were sorted for GFP-signal (7.2.2.6.1), pelleted and lysed with 0.75 mL RIPA buffer + 2.15 µL 0.1 M PMSF. The mixture was incubated on ice for 15 min at 400 rpm and then centrifuged at 14000 xg, 4 °C for 15 min. The supernatant was applied to a 10 % SDS-PAGE gel (7.2.1.18) and proteins were transferred to a polyvinylidene difluoride (PVDF) membrane using Thermo Scientific Pierce™ Power Blotter. Membrane was washed in H_2O for 5 min (shacking at RT) and blocked in LI-COR Odyssey Blocking Buffer for 60 min. Then, membrane was washed three times with TBS-T buffer and incubated with HA-antibody (Cell Signaling #3724) and GAPDH-antibody (Cell Signaling #2118) at 4 °C overnight. After another three washing steps with TBS-T, the membrane was stained with anti-rabbit secondary antibody conjugated with DyLight™ 800 (Cell Signaling #5151) for 60 min at RT. After another three

washing steps with TBS-T, imaging of the membrane was done with LI-COR Odyssey® CLx.

7.2.1.20. RNA Isolation

RNA from HEK293T cells was either isolated using TRIzol (Life Technologies) or the RNeasy Mini Kit (Qiagen) according to the manufacture's protocols.
For TRIzol extraction, sorted cells were pelleted and resuspended in 800 µL TRIzol and homogenized by pipetting. 160 µL chloroform was added and mixed by flipping for 30 s and incubated at RT for 3 min. Centrifugation at 12000 xg, 4 °C for 15 min led to a three-phase separation. The upper aqueous phase was transferred to a new 1.5 mL tube without touching the interphase. 400 µL isopropanol was added, mixed by flipping, and incubated for 10 min at RT. RNA was pelleted at 12000 xg, 4 °C for 10 min and resuspended in 1 mL 75 % ethanol by short vortexing. After another centrifugation step at 12000 xg, 4 °C for 5 min, supernatant was discarded, and the pellet was air-dried. As soon as the pellet was dry, 45 µL nuclease free H_2O with 0.1 mM EDTA was added and incubated at 55 °C for 10 min. The concentration was measured (7.2.1.4) and 1 U DNase I per 1 µg RNA with 5 µL 10x DNase I buffer was added and incubated at 37 °C for 40 min. 150 µL nuclease free H_2O and 200 µL Phenol-Chloroform-Isoamyl alcohol mixture were added and incubated at 37 °C, 700 rpm for 2 min followed by 2 min at RT without shaking. After centrifugation at 12000 xg, RT for 2 min, the upper H_2O phase was transferred into a new 1.5 mL tube and 200 µL chloroform was added and inverted five times. After 2 min incubation at RT, samples were centrifuged at 12000 xg, RT for 2 min. Supernatant was transferred into a new 1.5 mL tube and 20 µL 3 M NaOAc (pH 5.2) was added, vortexed and centrifuged briefly. Precipitation at -20 °C overnight was done after the addition of 475 µL 100 % ethanol. Samples were centrifuged at 16000 xg, 4 °C for 15 min, resuspended in 1 mL 75 % ethanol and centrifuged again at 16000 xg, 4 °C for 5 min. Pellet was air-dried, resuspended in 20 µL nuclease free H_2O and the concentration was measured (7.2.1.4).

7.2.1.21. Mass Spectrometry of Proteins

All used solutions were sterile filtered (Filtropur S 0.2 with 0.2 µm pores), and work was done under a clean bench to avoid contamination with keratin. Whole cell lysate produced as described in 7.2.1.19 (for Figure 18B) was applied to a 10 % SDS-PAGE gel (7.2.1.18) and stained. The part covering the transfection control and DNMT3a3L band was cut out in very small pieces, transferred to a 1.5 mL low-binding tube, and washed in 200 µL 25 mM NH_4HCO_3 3:1 in Acetonitrile for 30 min, 37 °C at 600 rpm. Then, sample was washed in 200 µL 25 mM NH_4HCO_3 1:1 in Acetonitrile for 15 min, 37 °C at 600 rpm. Both washing steps were repeated once. Reduction was done with 50 mM DTT in 25 mM NH_4HCO_3 for

45 min, 37 °C at 600 rpm. The peptides were alkylated in the dark with 55 mM iodacetamide in 25 mM NH_4HCO_3 for 1 h at 25 °C and washed twice with 25 mM NH_4HCO_3 1:1 in Acetonitrile. The gel was dehydrated with 100 µL acetonitrile and fully air-dried afterwards. 125 µL digest solution (0.1 µg/µL Trypsin MS-grade (in 10 mM HCl) diluted 1:10 in 25 mM NH_4HCO_3) was added and after 15 min at 25 °C another 125 µL 25 mM NH_4HCO_3 was added and shaken at 350 rpm 30 °C overnight. Addition of 16.6 µL 10 % TFA stopped the digestion and samples were sonified on ice for 30 min. The supernatant was transferred into a new 1.5 mL low binding tube and 100 µL acetonitrile was added to the remaining gel and incubated for 15 min, 25 °C at 350 rpm. Supernatant was combined with the previous collected supernatant and the step was repeated once. The solvent of combined supernatants was fully removed in a concentrator and stored at -20 °C until measurement. Protein fragments were analyzed by the group of Dr. Petra Janning by nano-HPLC-MS/MS using an UltimateTM 3000 RSLC nano-HPLC system and a Q ExactiveTM Plus Hybrid Quadrupole-Orbitrap equipped with a nano-spray source (all from Thermo Fisher Scientific). The tryptic peptides were suspended in 20 µL 0.1% TFA and 1 µL of the samples were injected onto and enriched on a C18 PepMap 100 column (5 µm, 100 Å, 300 µm ID * 5 mm, Thermo Scientific) using 0.1 % TFA, at a flow rate of 30 µL/min, for 5 min. Subsequently, the peptides were separated on a C18 PepMap 100 column (3 µm, 100 Å, 75 µm ID * 50 cm) using a linear gradient, starting with 95 % solvent A/5 % solvent B and increasing to 30 % solvent B in 90 min with a flow rate of 300 nL/min (solvent A: water containing 0.1 % formic acid; solvent B: acetonitrile containing 0.1 % formic acid). The nano-HPLC apparatus was coupled online with the mass spectrometer using a standard coated Pico Tip emitter (ID 20, Tip-ID 10, New Objective, Woburn, MA, USA). Signals in the mass range of m/z 300 to 1650 were acquired at a resolution of 70,000 followed by up to ten high-energy collision-dissociation (HCD) MS/MS scans of the most intense at least doubly charged ions at a resolution of 17,500. Identification of the artificial modification of the protein of interest was performed by using MaxQuant[366] v.1.6.14.0, including the Andromeda search algorithm and searching a database containing the sequence of the proteins of interest and known contaminants in parallel. Briefly, an MS/MS ion search was performed for enzymatic trypsin cleavage, allowing two missed cleavages. Carbamidomethylation of cysteine, DMNB-modification of cysteine (formula change $C_9NO_4H_9$, Δm = 195.05316 u), acetylation of protein N-termini, and oxidation of methionine were set as variable modifications. The mass accuracy was set to 20 ppm for the first search, and to 4.5 ppm for the second search. The false discovery rates for peptide and protein identification were set to 0.01.

7.2.2. Cell-Biological Methods

7.2.2.1. Cultivation of Mammalian Cells

Cells were constantly treated under sterile conditions and the incubator had always a temperature of 37 °C, a CO_2-level of 5 % and a humidity of ≥ 95 %.
All growth media were sterile filtered, and cells were negatively tested for mycoplasma infections.

7.2.2.1.1. Culturing of HEK293T Cells

HEK293T cells were cultured on 10 cm TC-plates with 10 mL growth medium (DMEM (w/o L-Glutamine; with D-Glucose; with Pyruvate), 10 % FBS Premium, 1 % L-Glutamine, 1 % Penicillin/Streptomycin). At a confluency of 80 – 90 %, cells were passaged. For that, medium was discarded, cells were washed very gently with 1 mL DPBS and 1 mL 37 °C-prewarmed Trypsin (0.05% with 0.02% EDTA (PAN Biotech)) was added. After 5 min incubation at 37 °C, cells were completely detached and homogenized by gently pipetting with 4 mL 37 °C-prewarmed growth medium. For a two-day split, 800 µL, for a three-day split, 350 µL, and for a four-day split, 170 µL cell solution was transferred to 10 mL prewarmed growth medium on a new TC-plate and homogenized by gentle shacking. Cells were incubated at 37 °C for 2 – 4 days.

7.2.2.1.2. Culturing of WT HCT116 and DKO HCT116 Cells

WT HCT116 and DKO HCT116 cells (DNMT1 (Δexons3-5/Δexons3-5), DNMT3B (-/-), Horizon Discovery Ltd., #HD R02-022) were cultured on 10 cm TC-plates with 10 mL growth medium (RPMI (with L-Glutamine; with 2 g/L NaHCO3), 10 % FBS Premium, 1 % Penicillin/Streptomycin). At a confluency of 90 – 95 %, cells were trypsinized as described before (7.2.2.1.1). For a two-day split, 1600 µL and for a three-day split, 1000 µL cell solution was transferred to 10 mL prewarmed growth medium on a new TC-plate and homogenized by gentle shacking. Cells were incubated at 37 °C for 2 – 3 days.

7.2.2.1.3. Cryoconservation

Cryoconservation was used for long term storage of mammalian cell lines and stored at -152 °C. Confluent TC-plates were trypsinized (7.2.2.1.1 and 7.2.2.1.2) and total cell number was determined using a hemocytometer. Cells were transferred to a 15 mL tube with 10 mL DPBS and pelleted at 500 xg for 4 min. The cell pellet was gently resuspended in cold freezing medium (growth medium with 5 % DMSO) to a final cell concentration of $1 \cdot 10^6$ cells per mL for HEK293T cells and up to $3 \cdot 10^6$ cells per mL for WT HCT116 and DKO cells. 1 mL aliquots in cryotubes were placed into an isopropanol bath and stored at -80 °C for 24 h to decrease the temperature very slowly. Then, the tubes were transferred

to -152 °C for long term storage. For the usage, the cryoculture was incubated in a 37 °C water bath until it was fully thawed. Cells were transferred to 4 mL fresh medium, centrifuged at 500 xg for 4 min and resuspended in 10 mL 37 °C prewarmed growth medium. The following cultivation was done as described before (7.2.2.1.1 and 7.2.2.1.2).

7.2.2.1.4. 5-Aza Treatment

Describes experiment of Figure 43: On day 0, 500k HEK293T cells were seeded in each well of a 6-well TC-plate in the presence of different concentrations of (5-Aza). Before passaging on days 2 and 5, pictures were taken with EVOS FL microscope. On day 2, cells were splitted 1:15 and on day 5, cells were splitted 1:5 for all 5-Aza concentrations, regardless of the confluency at the time of splitting. Thus, the toxicity was kept up so it could be judged by the final confluency of the TC-plates. The concentration of 5-Aza remained constant for the respective concentrations during passaging.
HEK293T cells treated with 5-Aza which were used for methylation by DNMT3a3L or pcDNMT3a3L were cultured in the presence of 0.6 µM 5-Aza for several days. Before splitting for the actual experiment, cells were trypsinized, washed in DPBS and centrifuged at 500 xg for 4 min before counting to remove all remaining 5-Aza which would also inhibit the transfected DNMT3a3L.

7.2.2.2. Transient Transfections

7.2.2.2.1. With HEK293T Cells

600.000 HEK293T cells were seeded on a 3.5 cm TC-plate or on a µ-Dish 35 mm plate (Ibidi) for microscopy in 2 mL growth medium. If a heat shock had to be done, Ibidi plates and TC-plates were additionally coated with 0.01 % poly L-lysine in DPBS for 1 h at 37°C with a subsequent washing step with DPBS to prevent cell detaching. On the next day at a confluency of ~70 %, 3 µL FuGENE 6 or PEI (1 mg/mL stock solution) per 1000 ng plasmid DNA were mixed with 100 µL OptiMEM and incubated for 5 min at RT. Then, 2 µg TALE-DNMT/DNMT3a3L plasmid or 2 µg TALE-pcDNMT/pcDNMT3a3L/pcDNMT3aCD + 1 µg p1169 was added, mixed by pipetting, and incubated for 15 min at RT. For photocaged DNMT constructs, 0.05 mM DMNB-Cys (if not stated otherwise) was additionally added to the transfection mix and then added dropwise to the cells. Although DMNB-Cys was dissolved in 0.1 mM NaOH, it was not further neutralized before addition to the cells. Cell survival of HEK293T cells was tested with up to 1.5 mM NaOH (data not shown). Photocaged DNMT constructs were light irradiated 24 h later. DNMT constructs w/o C710amber mutation were analyzed 24 h or 48 h after transfection.
For mass spectrometry of the transfection control, 10 cm TC-plates were used. $6 \cdot 10^6$ cells were seeded and 7.5 µg pcDNMT3a3L + 5 µg p1169 were used.

7.2.2.2.2. With WT and DKO HCT116 Cells

400.000 WT or DKO HCT116 cells were seeded on a 3.5 cm TC-plate in 2 mL growth medium. On the next day, 3 µL X-tremeGENE9 per 1000 ng plasmid DNA was mixed with 150 µL OptiMEM. Also, plasmid DNA (2 µg TALE-DNMT/DNMT3a3L plasmid or 2 µg TALE-pcDNMT/pcDNMT3a3L/pcDNMT3aCD + 1 µg p1169) was mixed in additional 150 µL OptiMEM and both mixtures were combined after 5 min incubation at RT and incubated for additional 20 min. If needed, DMNB-Cys was added to the transfection mix with a final concentration of 0.05 mM (if not stated otherwise), and the transfection mixture was added dropwise to the cells. Photocaged DNMT constructs were light irradiated 48 h later. DNMT constructs w/o C710amber mutation were analyzed 48 h after transfection.
During this thesis, also FuGENE 6, Lipofectamin2000, and PEI was tested with 2 µg pcDNMT3a3L + 1 µg p1169 in the presence of 0.05 mM DMNB-Cys. As described before, 9 µL transfection reagent was incubated in 150 µL OptiMEM for 5 min. At the same time, plasmid DNA was incubated in 150 µL OptiMEM for 5 min. Both solutions were mixed and incubated 20 min at RT. DMNB-Cys was added, and the solution was applied dropwise to the cells. Analysis by microscopy was done 24 h and 48 h later.

7.2.2.3. Light Irradiation

HEK293T cells were irradiated with light 24 h and DKO HCT116 cells 48 h after transfection. Medium was discarded and 1 mL 37 °C prewarmed DPBS (for 3.5 cm TC-plates) was added gently. Samples were placed in the middle of a 365 nm UV-table (WUV-L10, 6x 15 W, Witeg) and irradiated for 5 min if not indicated differently. For controls w/o light, medium was also changed to DPBS, and samples were placed at RT w/o light irradiation for 5 min. Then, DPBS was exchanged with 2 mL growth medium (w/o DMNB-Cys) and cells were incubated for another 24 h before analysis (both cell lines).

7.2.2.4. Heat Shock Treatment

HEK293T cells plated on PLL coated TC-plates were incubated at 44 °C, 5 % CO_2 for 1 h. Subsequent analysis by cell sorting or microscopy was done immediately. Heat shock was performed 24 h after light irradiation for pcDNMT3a3L samples.

7.2.2.5. Microscopy

Some microscopy pictures of cells which did not require immunostaining were taken with the EVOS FL microscope (Life Technologies) using 2x, 4x, 10x and 20x objectives. For experiments with immunostaining or if a 60x objective was needed, microscopes 1 or 2 were used (see 7.1.1).

For immunostaining, cells were either fixed by adding 54 µL 37 % formaldehyde directly to the 2 mL growth medium (final concentration 1 % formaldehyde) or by exchanging the growth medium with 1 mL DPBS + 4 % formaldehyde. After incubation at 37 °C for 5 min, cells were washed once with 1 mL DPBS. Permeabilization was done with 1 mL DPBS + 0.25 % Triton-X 100 (15 min at RT). Cells were blocked in 1 mL blocking buffer (DPBS, 1 % BSA, 0.05 % Tween20) for 30 min at RT. Primary antibodies (versus HA-tag and HSF1) were added as 1:800 dilutions in 500 µL blocking buffer and incubated for 2 h at RT. After one washing step with PBS-T (PBS + 0.05 % Tween20), secondary antibodies (conjugated with AF488 or Cy5) were added as 1:800 dilutions in 500 µL blocking buffer and incubated for 1 h at RT in the dark. Cells were washed twice with PBS-T and stored in 2 mL DPBS for microscopy. For SatIII-pcDNMT and HSF1 localization experiments, images were taken as z-stacks of FretGFP and Cy5 channels using a 60x oil objective with an Olympus IX81 microscope coupled with Hamamatsu model C10600-10B-H camera with an exposure time of 10 ms each. The subcellular localization of foci was analyzed from maximal intensity z-projections of image stacks using the Fiji distribution of ImageJ.[380]

7.2.2.6. Flow Cytometry

7.2.2.6.1. Cell Sorting

HEK293T cells were washed with 1 mL DPBS and trypsinized (500 µL for 3.5 cm TC-plates, 1 mL for 10 cm TC-plates, 4 min 37 °C, see 7.2.2.1.1 and 7.2.2.1.2). Growth medium was added (1 mL for 3.5 cm TC-plates, 4 mL for 10 cm TC-plates) and the cell suspension was transferred into 15 mL tubes with DPBS. After centrifugation (5 min, 500 xg), the pellet was resuspended either in 300 µL 4 °C cold DPBS or in FACS sorting buffer, filtered into 5 mL tubes (round bottom with cell strainer, Falcon #352235) and kept on ice until sorting. Sorting for high mCherry or high GFP expressed cells (GFP for TALE-pcDNMT, pcDNMT3a3L or pcDNMT3aCD constructs) was done with a SONY SH800S cell sorter with prior compensation of bleed-through fluorescence (488 nm and 561 nm lasers with manufactures optical filter pattern 2). 2 mL or 15 mL collection tubes were coated with 10 % FBS in DPBS (rotation for 1 h at 4 °C) to avoid cell sticking to the tube. Sorted cells were centrifuged at 500 xg for 5 min. For experiments where gDNA had to be isolated, the supernatant was removed to a final residual volume of ~200 µL and Proteinase K was added (for more see 7.2.1.14). For RNA isolation, the supernatant was discarded completely, and cells were resuspended either in TRIzol or according to the RNA isolation kits protocol (7.2.1.20).

7.2.2.6.2. Fixation, Permeabilization and Staining for Analysis

The following protocol was used for experiments in which cells were not sorted but only analyzed by immunostaining and subsequent measurement with SONY SH800S (405 nm, 488 nm, 561 nm, and 638 nm lasers with manufactures optical filter pattern 2). For all experiments, cells were cultured on 3.5 cm TC-plates and compensation of bleed-through fluorescence was done with samples immunostained with only one dye.

Cells were trypsinized (500 µL, 4 min, see 7.2.2.1.1 and 7.2.2.1.2), homogenized with 1 mL growth medium and transferred to 5 mL tubes (round bottom, Falcon #352058) with 3 mL DPBS. After centrifugation at 500 xg for 5 min, cells were resuspended in 100 µL medium A of Fix + Perm Cell Permeabilization Kit (Life Technologies) and incubated for 10-15 min at RT. After washing with 2 mL FACS washing buffer (DPBS + 5% FBS + 0.1% NaN_3), cells were permeabilized with 100 µL medium B for 20 min and then washed again. If 5mC should be stained, chromosomal DNA was denatured by incubation with 2 N HCl for 20 min, centrifugation, and resuspension in 1 mL DPBS with 15 min incubation. Cells were blocked with 1 mL FACS blocking buffer (DPBS + 1% BSA + 0.05% Tween20) overnight at 4 °C and 500 rpm. Cells were centrifuged and 100 µL solution with antibodies against 5mC and/or DNMT3a3Ls C-terminal HA-tag were added as 1:400 dilutions in blocking buffer. After 1 h incubation at RT and 500 rpm, cells were washed twice with PBS-T. Then, cells were incubated with secondary antibodies (diluted 1:400 in 100 µL blocking buffer, see 7.1.10 for list of used antibodies. Dyes were AF405, AF488 and fluorescein) for 1 h at RT in the dark and 500 rpm. Cells were washed again twice with PBS-T and once with DPBS. After filtering the cells (round bottom with cell strainer, Falcon #352235), they were measured by flow cytometry.

7.2.2.6.3. Analysis of DNMT3a3L Activity

For activity analysis of DNMT3a3L and its mutants, flow cytometry data of single cells were exported with Cell Sorter Software (Sony Biotechnology) to a comma-separated values file and analyzed with R (v. 3.6.2)[381] using the data.table package[382] and visualized using ggplot2[383]. Untransfected cells were selected as the most abundant unstained population from a Gaussian mixture model[384] and transfected cells were defined with respect to a threshold of 8.4 times above the median of untransfected cells in the HA stain. Of such cells, 5mC-positive cells were defined with a threshold of 15.85 times the median of the untransfected cells under the same conditions in the 5mC stain (see Figure 51, page 125).

7.2.2.7. Pyrosequencing

Pyrosequencing analysis was done as previously reported.[332]
In short, bisulfite-converted gDNA was used to amplify the SatIII-locus with primers o3266/o3439 (7.2.1.14 and 7.2.1.15). The PCR product was prepared for sequencing and mixed with sequencing primers as reported. The pyrosequencing reaction was performed in a PSQ HS 96ATwo Pyrosequencer and analyzed with the PSQ HS 96A software. Shown 5mC levels were absolute numbers received in the sequencing run and were not normalized to standard samples.

7.2.3. Next Generation Sequencing

7.2.3.1. Illumina DNA Sequencing and Analysis

Methylation of the p16 gene by p16-DNMT, p16-DNMT^{E756A} and p16-pcDNMT was analyzed by next generation Illumina sequencing. HEK293T cells were either transfected with pJaW926 or pJaW935 and sorted for mCherry 48h later, or with pJaW1125 + pStH1169 + 0.5 mM DMNB-Cys and irradiated with light 24 h later. The light samples were sorted for GFP-signal after another 24 h incubation. gDNA was isolated and bisulfite converted as described before (7.2.1.14 and 7.2.1.15). Then, two parts of the p16 locus downstream of the TALE binding site was amplified in multiple PCRs (7.2.1.15). In a first PCR, primer pairs o2344/o2210 and o2340/2201 were used. The first mentioned product was then used in another PCR using primers o2339/o2199. A DNA-barcode was added to all amplicons by using the same reverse primer in a second PCR using Taq polymerase and o2363 or o2406 (for p16-DNMT samples); o2366 or o2407 (for p16-DNMT samples) or o2410 (for p16-pcDNMT) as forward primers. All amplicons were purified (7.2.1.10) and pooled equimolar in a 1.5 mL tube. This mixture was further purified with the “QIAquick Nucleotide Removal Kit”. Library preparation and Illumina sequencing was done by GATC (2x 150 bp paired end), and analysis was conducted using Qiagen Genomic Workbench.

Monitoring of SatIII methylation in HEK293T cells with SatIII-pcDNMT up to 24 h after light irradiation (Figure 32) was also analyzed by Illumina DNA sequencing. Cells were seeded on 3.5 cm TC-plates and transfected with p1660/p1169 and 0.05 mM DMNB-Cys as described before. 24 h later, cells were 5 min light irradiated and ‘0 h’ samples were immediately sorted for GFP-signal. Other samples were sorted at the respective time points. After gDNA isolation and bisulfite conversion, SatIII locus was amplified using the Qiagen MSP Kit and primer pair o3268/o3269. As before, a DNA-barcode was added by PCR using Taq polymerase and corresponding barcode primer pairs from o2363/o3033 up to o2401/o3071 (see Table 28). Single PCR reactions were purified with NEB Monarch® PCR & DNA Cleanup Kit (#T1030L) and pooled equimolar according to Table 28 in 1.5 mL tubes.

These pools were purified again with Macherey Nagel PCR Clean up kit and eluted in 20 µL NE-buffer. 25 µL of each pool with a concentration of 20 ng/µL were sent to GENEWIZ using their Amplicon-EZ service (Illumina Miseq platform). For analysis (executed by Dr. Benjamin Buchmuller), 250-mer pairs were trimmed and then merged with Pandaseq v. 2.11[385] using a k-mer length of 3. Mapping was done against the *in silico* bisulfite-converted sequence of the sense strand (see SatIII Sequence, page 134), flanked on both ends with the appropriate (non-converted) primer binding sites and sample barcodes using BBMap v. 38.87[386]. Sample barcodes (Table 28) and both CpGs were degenerated (NNNNNN or NN, respectively). The BAM files were sorted and indexed and taken for further analysis in R v. 4.0.1[381] (for read counts see Table 29). The aligned subsequences at the four degenerated positions were extracted from the BAM files using the GenomicAlignments package of the Bioconductor suite[387] and the barcodes demultiplexed using DNABarcodes package[388] using the sequence-Levenstein model for error correction, i.e., allowing insertions and deletions. Alignments had to fulfill requirements to be kept guaranteeing that the respective sample is identified correctly: When both barcodes corresponded without error to the designed barcodes; if one of them was missing, but the other corresponded without error; when both matched the same designed (and unambiguous) barcode within less than four operations. Additionally, alignments were discarded if one or both CpGs were not read as “TG” or “CG”. Finally, the fraction of “CG” alignments over all alignments for a CpG was determined for each sample (see Table 30).

7.2.3.2. RNA Sequencing and Transcriptome Analysis

HEK293T cells on a 3.5 cm TC-plate were transfected with pcDNMT3aCD (pJaW2559) and pStH1169 in the presence of 0.05 mM DMNB-Cys. 5 min light irradiation was done 24 h later as described before (7.2.2.3) and GFP-positive cells were sorted either immediately (‘0 h’ sample), 4 h or 8 h later. From these cells RNA was extracted with TRIzol (7.2.1.20). RNA-sequencing was performed by BGI Genomics (DNBseq RNA-Seq). Subsequent analysis was already described by Palei et al.[367] and executed by Dr. Benjamin Buchmuller. Cleaned, single-end RNA-sequencing reads were aligned to the GENCODE reference annotation[389] of the human transcriptome (v32) which was expanded to include transcripts from the transfected plasmids pcDNMT3aCD and the synthetase/tRNA pair and using the GRCh38 primary assembly (v32) as a decoy. Mapping was performed with Salmon v. 1.3.0[390] adopting a k-mer length of 27 and the --validate Mappings (‘selective alignment’) option[391]. Transcript abundance was quantitated with the same tool considering a fragment length of 160 to 250 bp (205 ± 45 bp) and correcting for random hexamer priming bias[392] Count data was analyzed with R v. 4.0.1[381] after projecting transcript counts onto genes[393]

using DESeq2[394] and applying adaptive shrinkage[395,396] to control the false-discovery rate in differential gene expression.

7.2.4. Synthesis of DMNB-Cys

The photocaged amino acid DMNB-Cys was synthesized by Stefan Helmer according to a procedure adapted from Pedersen et al.[397]

12.75 mmol L-Cysteine (2.01 g) was resolved in 16 mL 1 M NaOH and 12.54 mmol 4,5-dimethoxy-2-nitrobenzyl bromide (3.5 g) was resolved in 8 mL THF. The bromide was added to the stirred L-Cysteine solution over 1 h. After the addition of 12 mL THF and 12 mL 1 M NaOH, the mixture stirred overnight. The suspension was filtered, and the brown solid was washed with THF, ethyl acetate and Et_2O and dried over filter paper. The crude product was resolved in 20 mL H_2O and lyophilized over 3 days. The product was obtained as a light-yellow solid (2,5 g, 7,9 mmol, 61,9 %). For cell experiment, DMNB-Cys was dissolved as 100 mM stock solutions in 0.1 mM NaOH.

^{1}H NMR (400 MHz, DMSO) δ = 7.67 (s, 1H), 7.32 (s, 1H), 4.08 (dd, *J*=30.1, 13.5, 2H), 3.95 (s, 3H), 3.87 – 3.82 (m, 3H), 2.90 (dd, *J*=14.6, 3.9, 1H), 2.75 (dd, *J*=14.6, 7.5, 1H).

^{13}C NMR (126 MHz, DMSO) δ 169.0, 152.7, 147.4, 139.6, 129.1, 114.9, 108.9, 56.5, 56.1, 53. 3, 40.0, 39.8, 39.7, 39.5, 39.3, 39.2, 39.0, 32.3, 31.8.

For NMR spectra, see Figure 69 and Figure 70 on pages 156 and 157.

8. Appendix

8.1. Supporting Figures

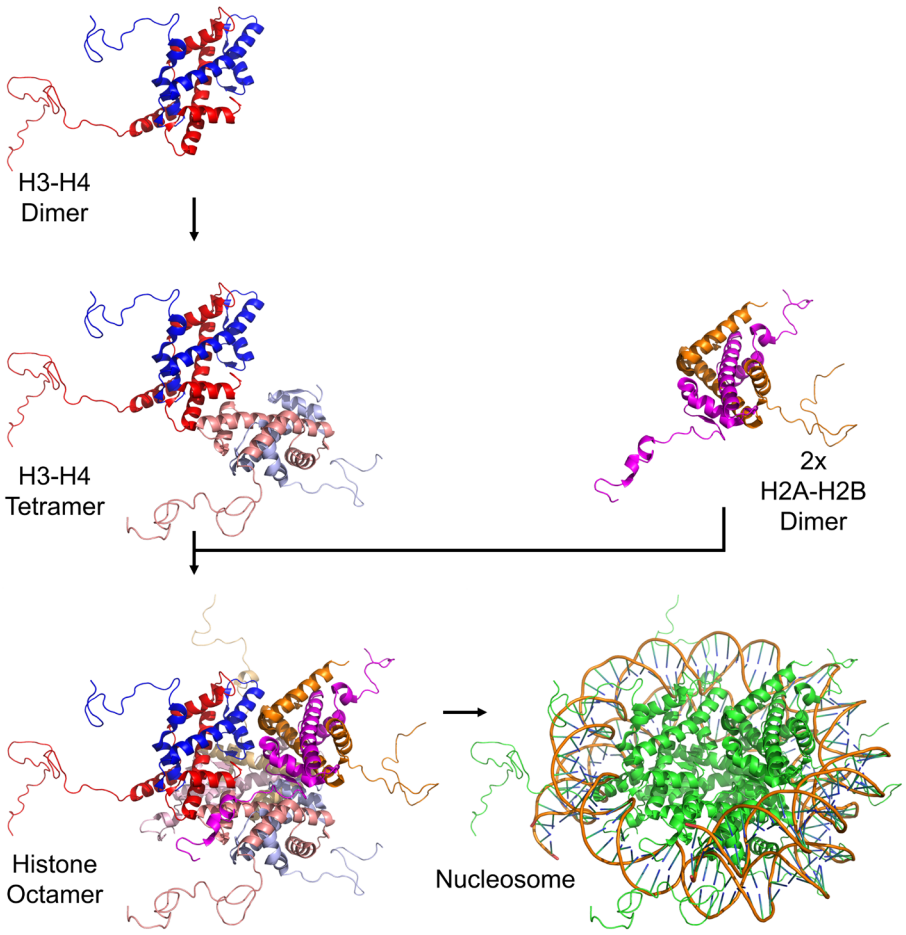


Figure 36: Assembly of nucleosomes. Nucleosomes consist of eight histone proteins forming an octamer and the DNA which is wrapped around it. First, histone H3 (red) forms a dimer with histone H4 (blue) as well as histone H2A (magenta) with histone H2B (orange). Then, two H3-H4 dimers congregate, resulting in a tetramer (additional histones: H3 (salmon) and H4 (light blue)). This tetramer and two H2A-H2B dimers build the histone octamer (additional histones: H2A (light pink) and H2B (light orange)) which becomes the final nucleosome after the DNA is wrapped around it (histone octamer is shown in green for better visualization). Structures adopted from PDB 1KX5[16,27].

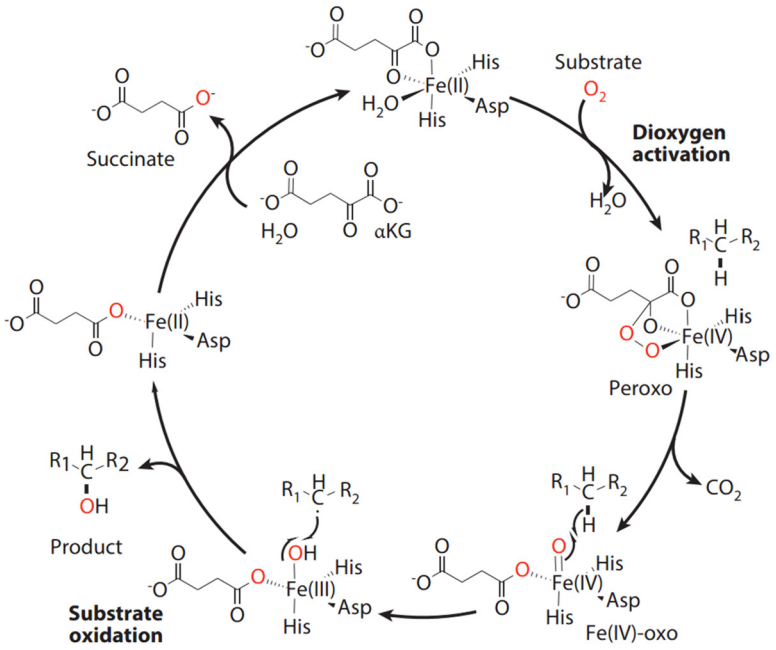


Figure 37: Stepwise oxidation reaction by TET enzymes. The reaction takes place in two stages starting with the dioxygen activation.[398] Fe(II) as well as α-KG contribute two electrons for the activation of dioxygen via a bridged peroxo intermediate.[398,399] The highly active Fe(IV)-oxo species is used in the following substrate oxidation stage to oxidize the inert C–H bond of the substrate with one oxygen atom of the consumed dioxygen.[398] To complete the catalytic cycle, the oxidized cofactor succinate with the other oxygen atom is substituted with a new α-KG.[398] Figure adapted from[398].

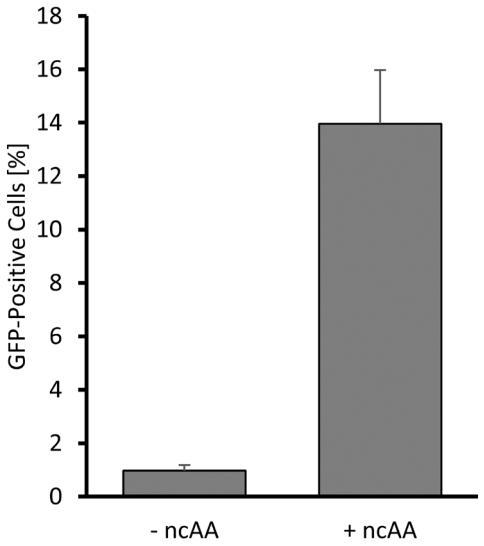


Figure 38: GFP^{Y39amber} expression w/o and with DMNB-Cys. HEK293T cells were transfected with p1660/p1169 in the presence or absence of 0.5 mM DMNB-Cys and analyzed by flow cytometry 24 h after transfection. Cells with a GFP value above 100-times the median GFP value of untransfected cells were considered GFP-positive. Error bars show standard deviations from two independent biological replicates. Modified from Wolffgramm et al.[103]

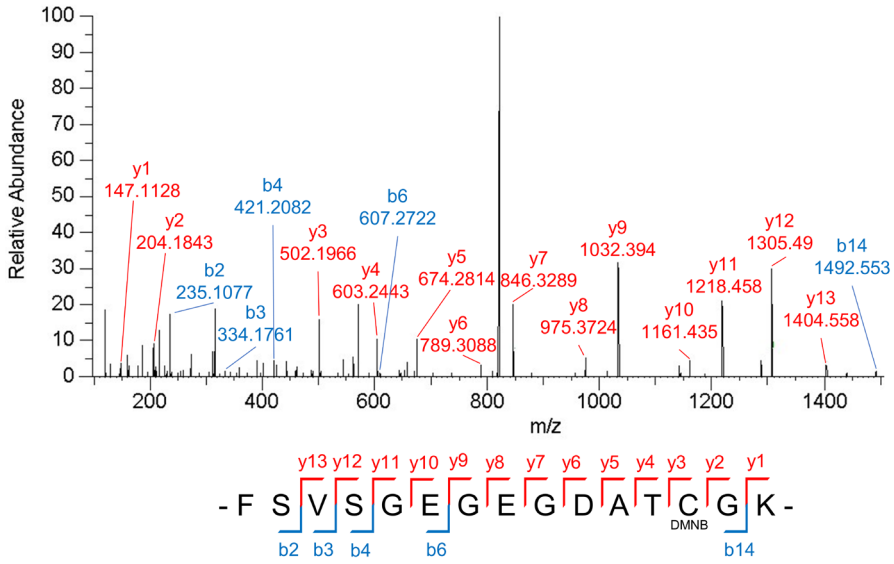


Figure 39: Additional ESI-MS/MS spectra of GFP peptide. Spectrum of the photocaged (DMNB) peptide with light. The spectrum of the decaged and carbamidomethylated cysteine with light is shown in Figure 17B. Comparison of theoretical and measured masses are found in Table 27, page 144.

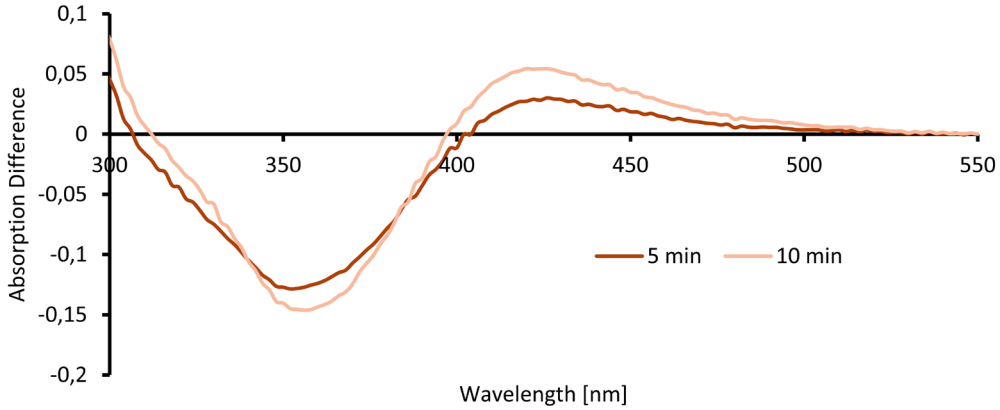


Figure 40: Absorption difference of DMNB-Cys for 5 min and 10 min light samples. Differences in absorption between 5 min or 10 min light samples and 0 s sample in Figure 19B “+ Light”. 1 mM DMNB-Cys in DPBS was irradiated with 365 nm light.

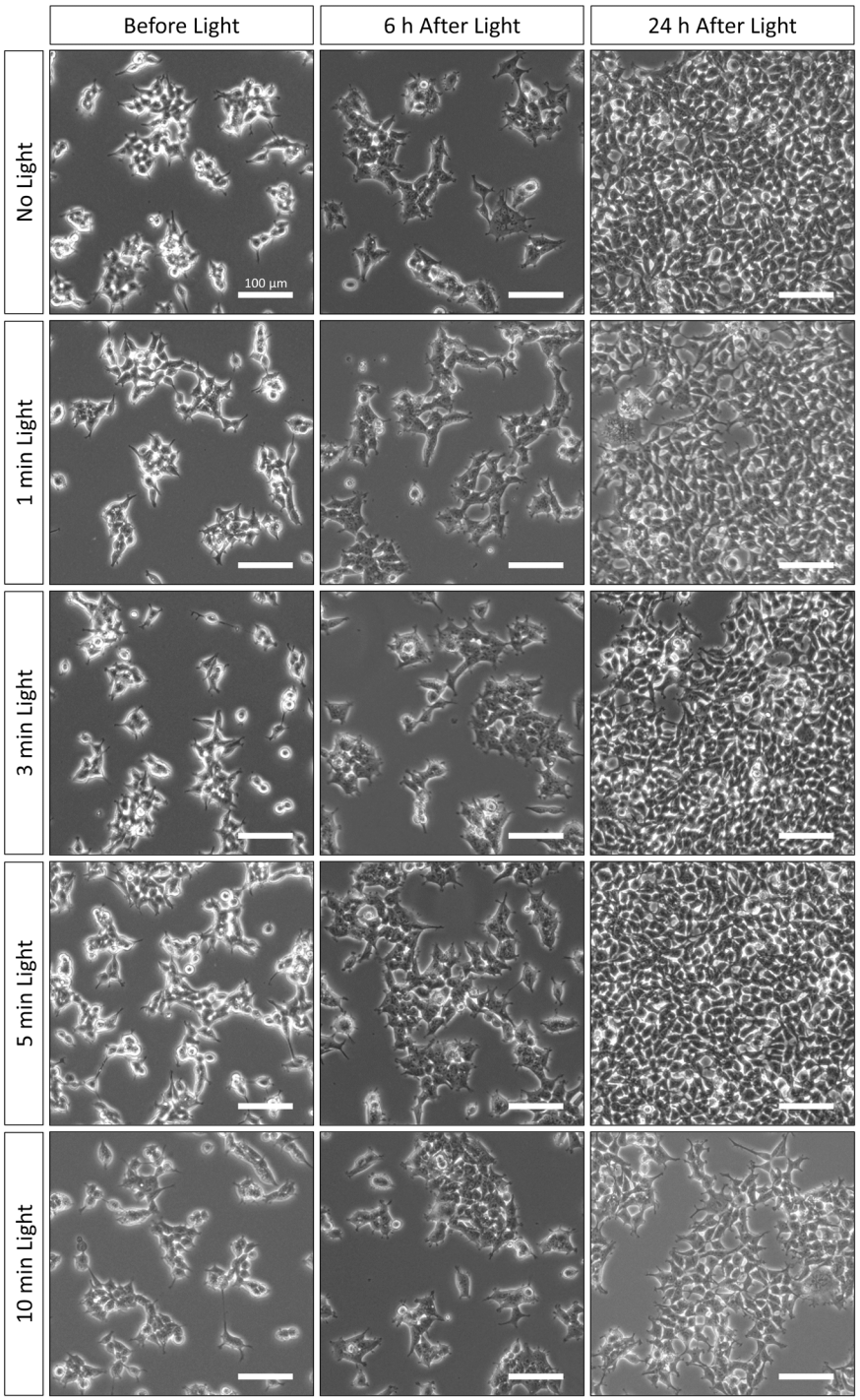


Figure 41: HEK293T morphology after light irradiation. Cell morphology was monitored directly before light irradiation (in DPBS) and 6 h and 24 h after (in medium). Scale bar is 100 µm.

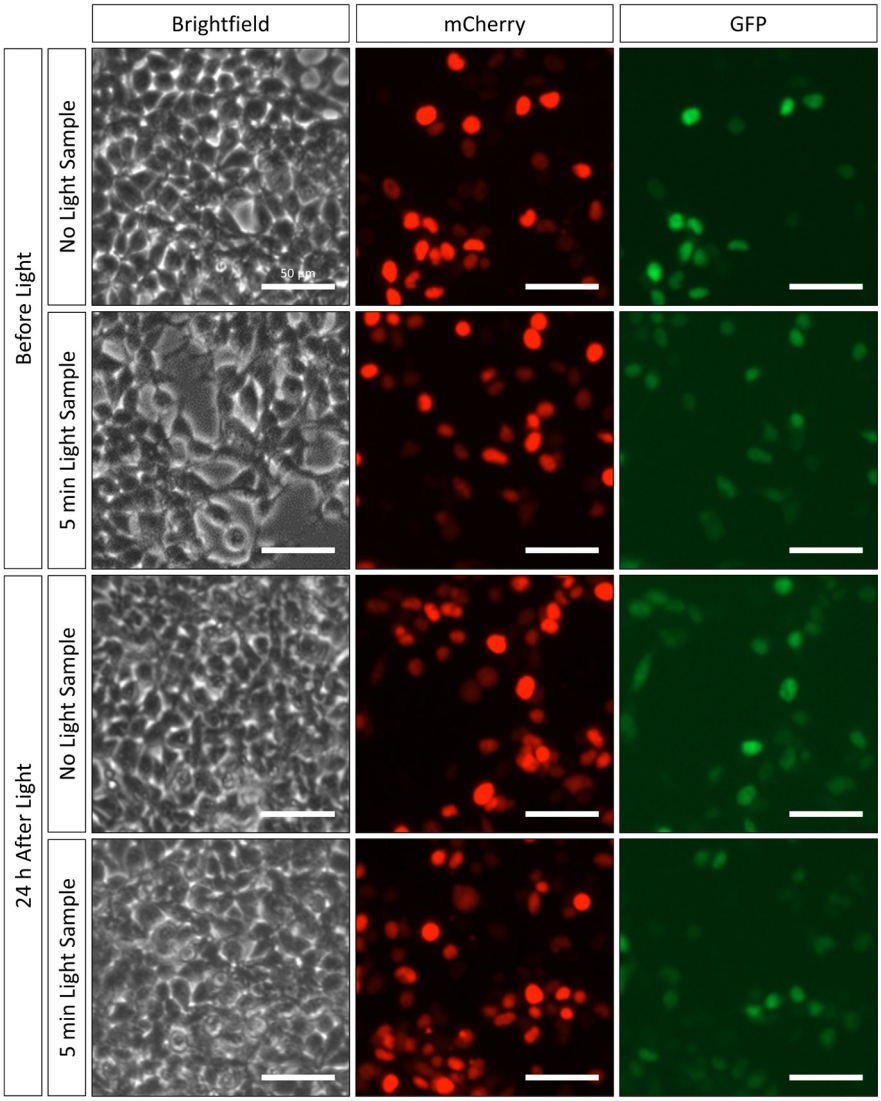


Figure 42: Cell morphology of transfected HEK293T cells after light irradiation. Cells were irradiated with 5 min light 24 h after transfection with SatIII-pcDNMT and 0.05 mM DMNB-Cys. Scale bar is 50 µm. Modified from Wolffgramm et al.[103]

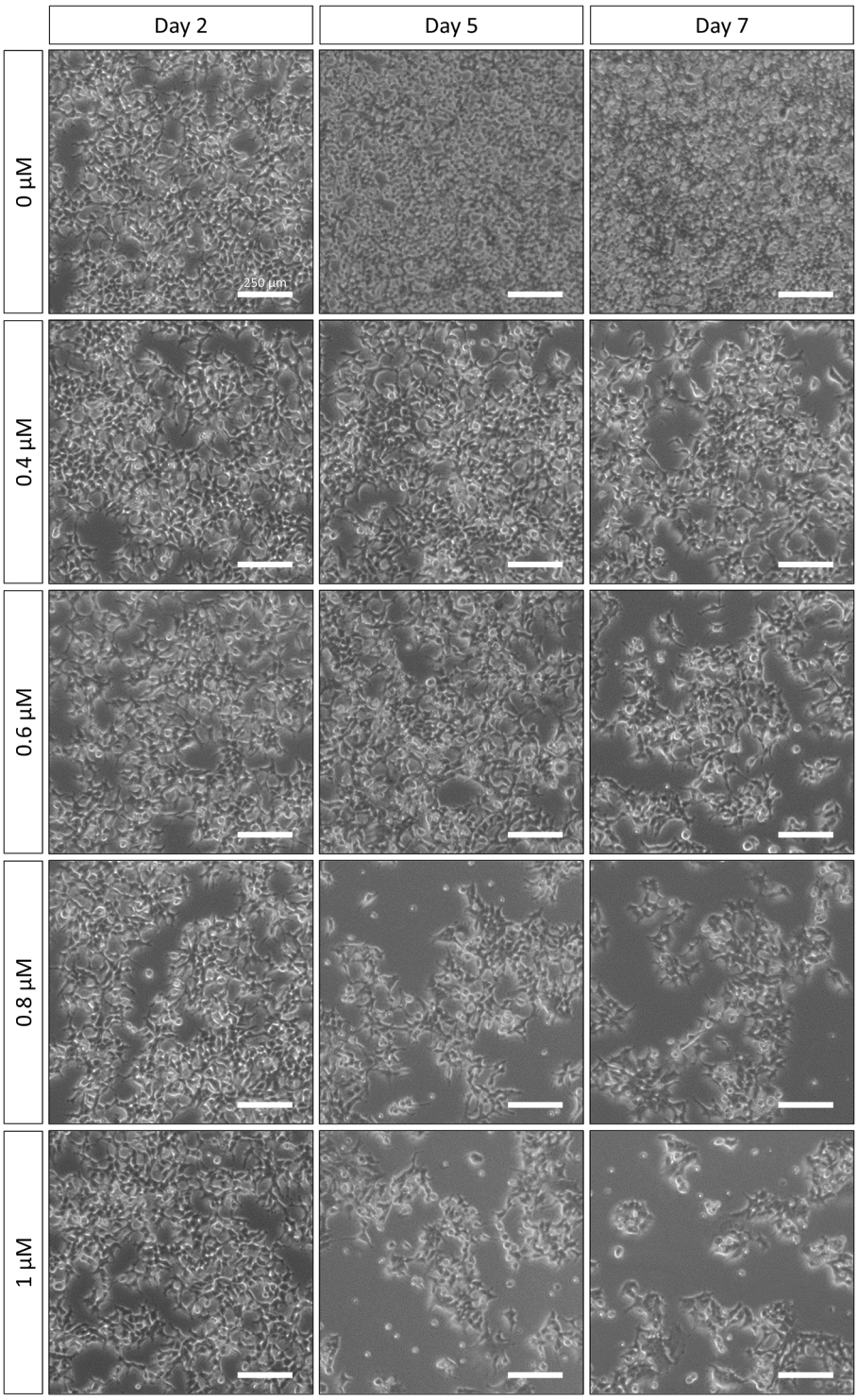


Figure 43: Survival of HEK293T cells at different 5-Aza concentrations. Cells were passaged on day 0, day 2 and day 5. Pictures were taken before passaging on day 2 and day 5 and additional on day 7. See methods 7.2.2.1.4.

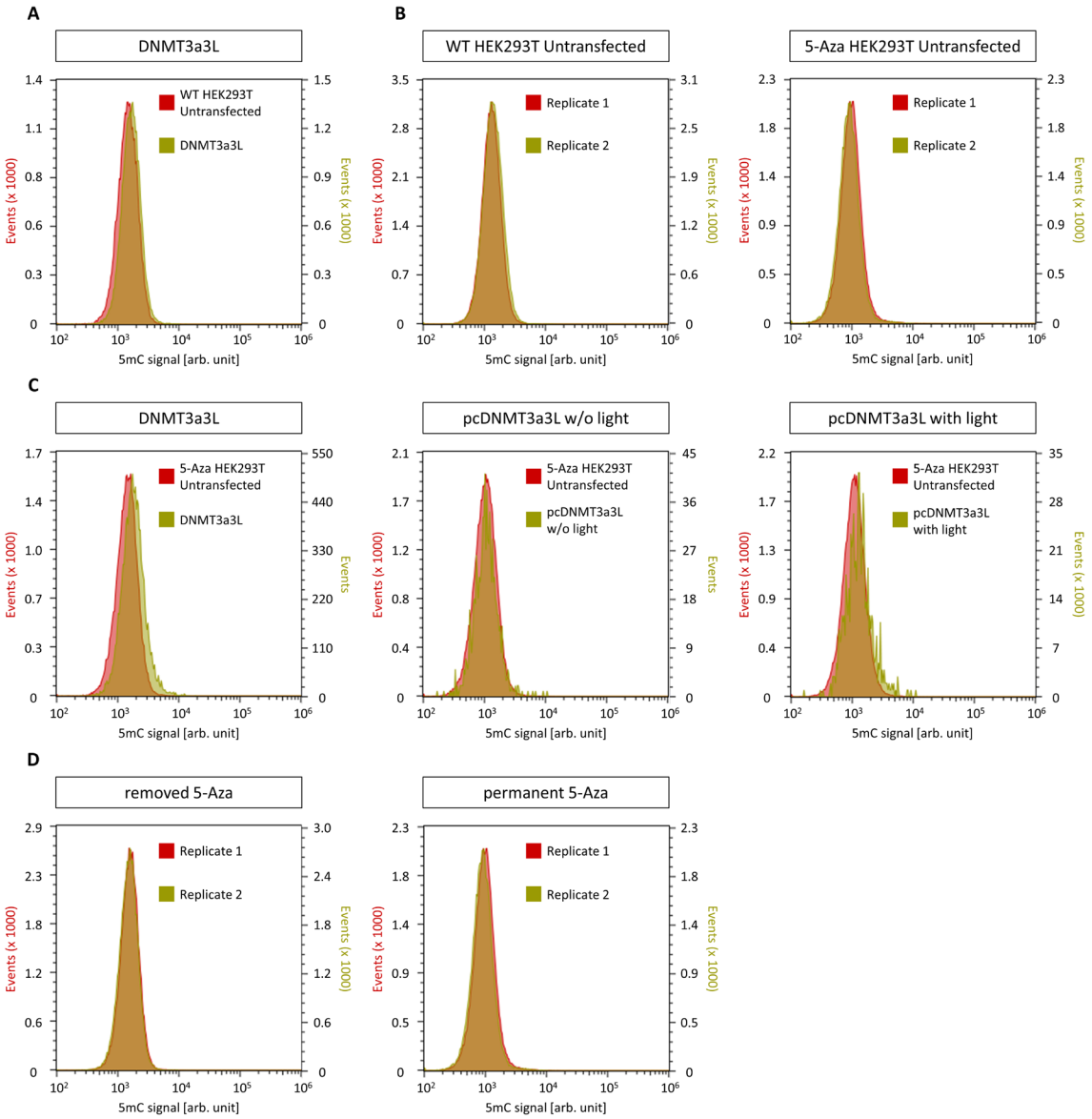


Figure 44: Additional replicates of Figure 20 showing global methylation in WT and 5-Aza HEK293T cells. A: Second replicate showing WT HEK293T cells which were transfected with DNMT3a3L (p1848) and analyzed 24 h after transfection by immunostaining. **B:** Both replicates as overlay for WT HEK293T cells and cells treated with 0.6 µM 5-Aza for 6 days. Both replicates '1' were used as overlay in Figure 20B. **C:** Second replicates for each condition in Figure 20C. Global DNA-methylation by either DNMT3a3L (p1848) or pcDNMT3a3L (p1849) with or w/o light irradiation in 5-Aza HEK293T cells. Cells were cultivated 72 h with 0.6 µM 5-Aza and 24 h w/o 5-Aza prior to transfection. Analysis was done by immunostaining 24 h after transfection for DNMT3a3L. For pcDNMT3a3L, 5 min light irradiation was done 24 h after transfection and cells were immunostained after additional 24 h. **D:** Both replicates as overlay showing the basal 5mC levels in HEK293T cells treated with 0.6 µM 5-Aza for 6 days ("permanent 5-Aza") and cells grown in presence of 0.6 µM 5-Aza for 3 days and additional 3 days in absence of 5-Aza ("removed 5-Aza"). Both replicates '1' were used as overlay in Figure 20D.
Immunostaining was done with HA- and 5mC-antibodies. Secondary antibodies were labeled with FITC (for HA) and AF405 (for 5mC). For all experiments, cells with a HA-signal above 20000 [arb. unit] were defined as DNMT3a3L- or pcDNMT3a3L-positive. Shown are the 5mC-signal histograms of the respective cell populations.

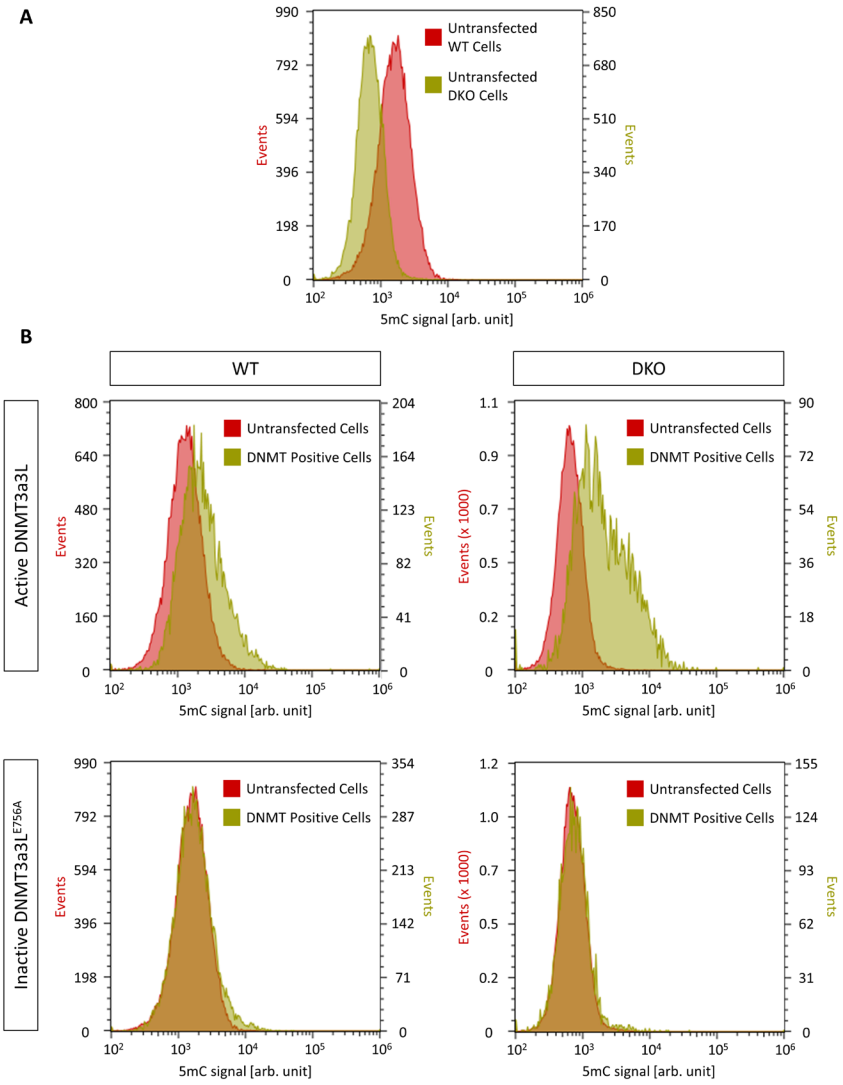


Figure 45: Additional replicates of Figure 21 showing global methylation in WT and DKO HCT116 cells. A: Histograms of 5mC immunostainings of untransfected WT and DKO HCT116 cells showing the basal 5mC levels. **B:** Global methylation with DNMT3a3L and inactive DNMT3a3L^{E756A} in WT and DKO HCT116 cells 48 h after transfection. Cells were immunostained with HA- and 5mC-antibodies. Cells with a minimum HA-signal of 20000 [arb. unit] were defined as DNMT3a3L-positive.

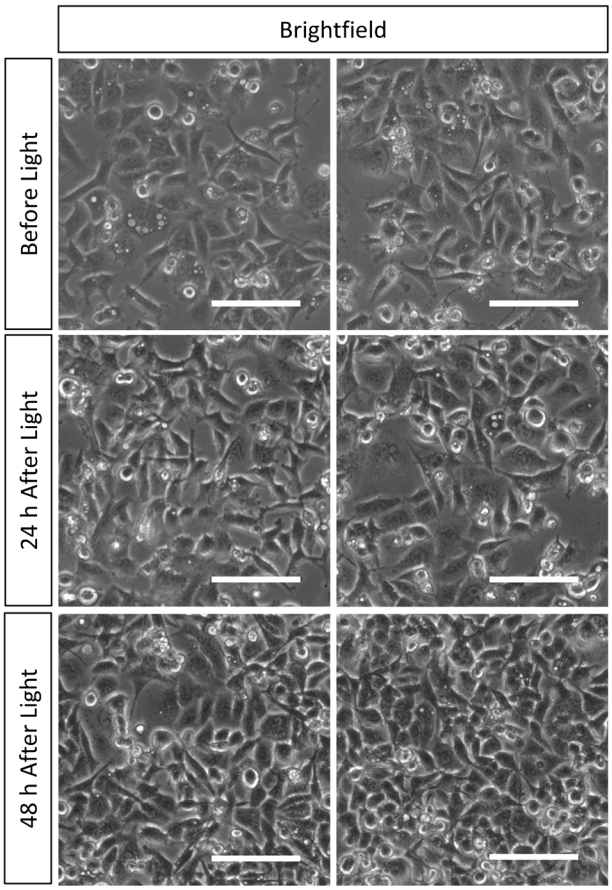


Figure 46: Cell morphology of DKO HCT116 cells before and after light irradiation. Cell morphology was monitored directly before light irradiation (in DPBS) and 24 h and 48 h after (in medium). Scale bar is 100 µm.

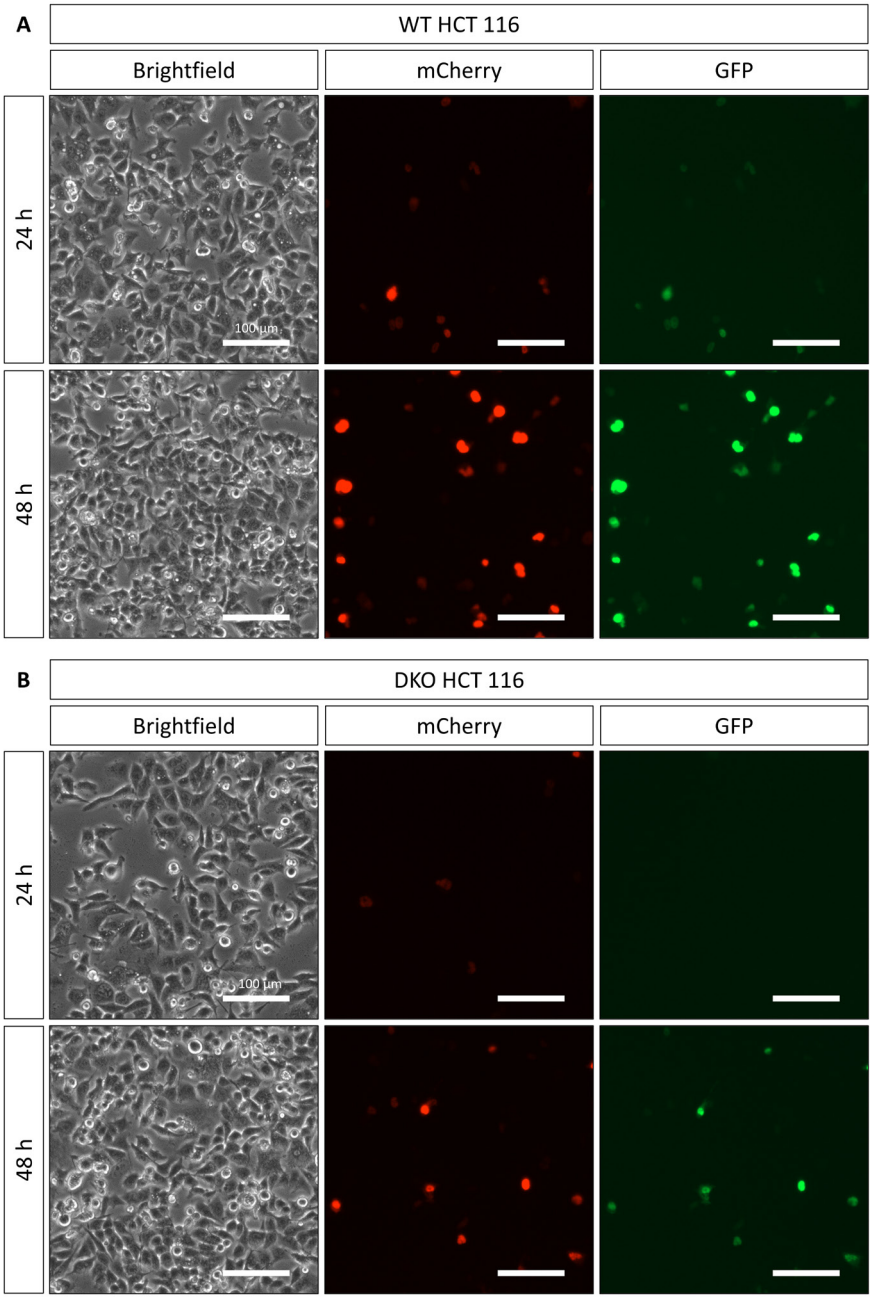


Figure 47: Transfection efficiency of XtremeGENE 9 in WT and DKO HCT116 cells. Pictures of WT **(A)** and DKO **(B)** HCT116 cells were taken 24 h and 48 h after transfection with 2000 ng p1559, 1000 ng p1169 and 9 µL transfection reagent. Scale bar is 100 µm.

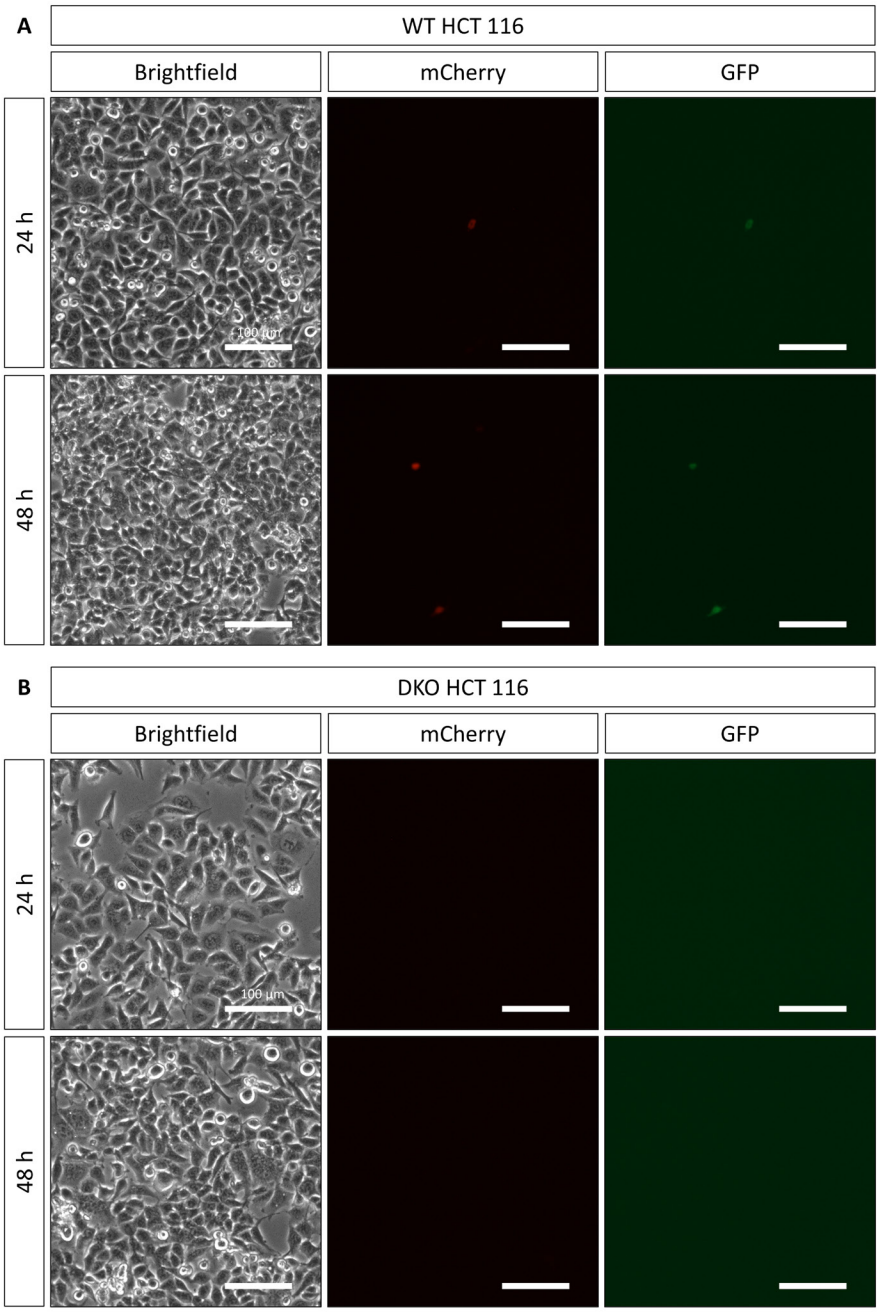


Figure 48: Transfection efficiency of FuGENE 6 in WT and DKO HCT116 cells. Pictures of WT **(A)** and DKO **(B)** HCT116 cells were taken 24 h and 48 h after transfection with 2000 ng p1559, 1000 ng p1169 and 9 µL transfection reagent. Scale bar is 100 µm.

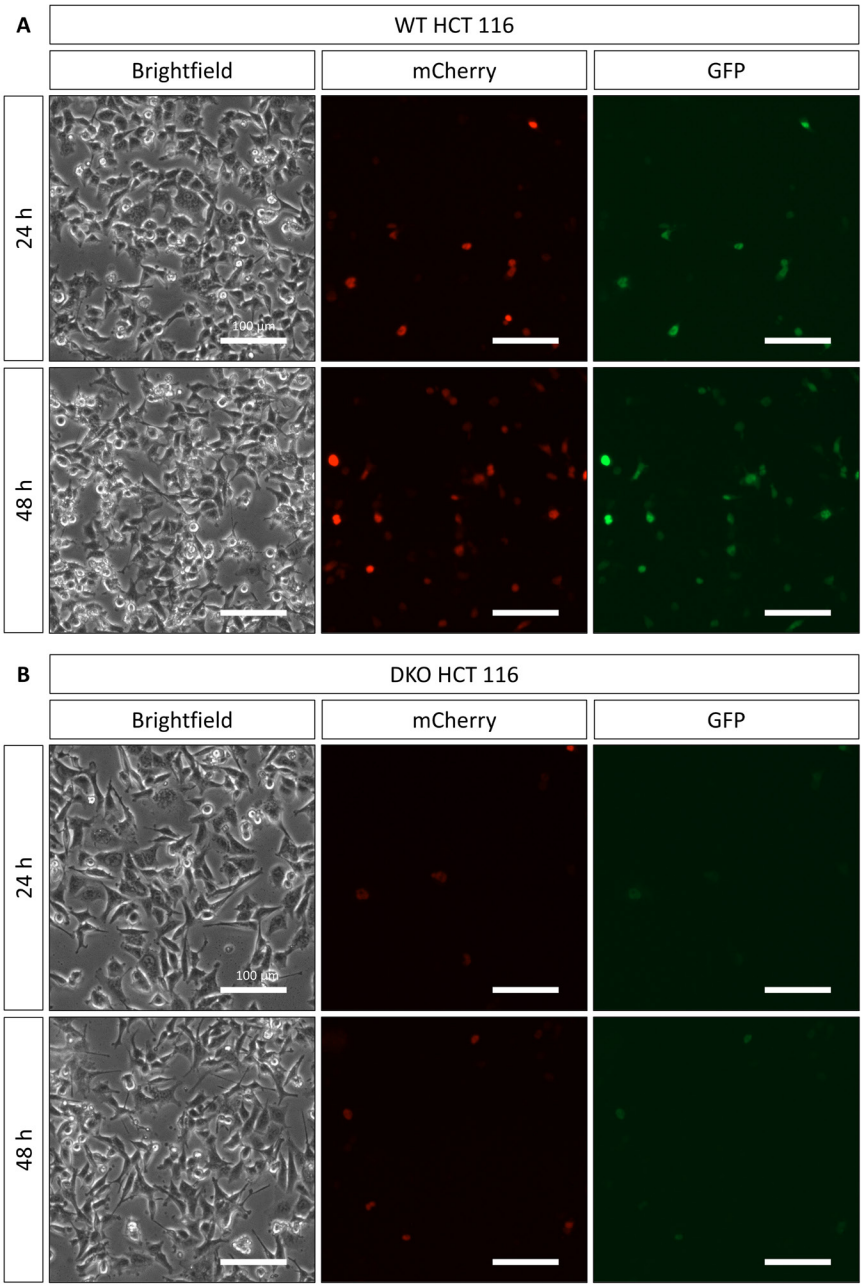


Figure 49: Transfection efficiency of PEI in WT and DKO HCT116 cells. Pictures of WT **(A)** and DKO **(B)** HCT116 cells were taken 24 h and 48 h after transfection with 2000 ng p1559, 1000 ng p1169 and 9 µL transfection reagent. Scale bar is 100 µm.

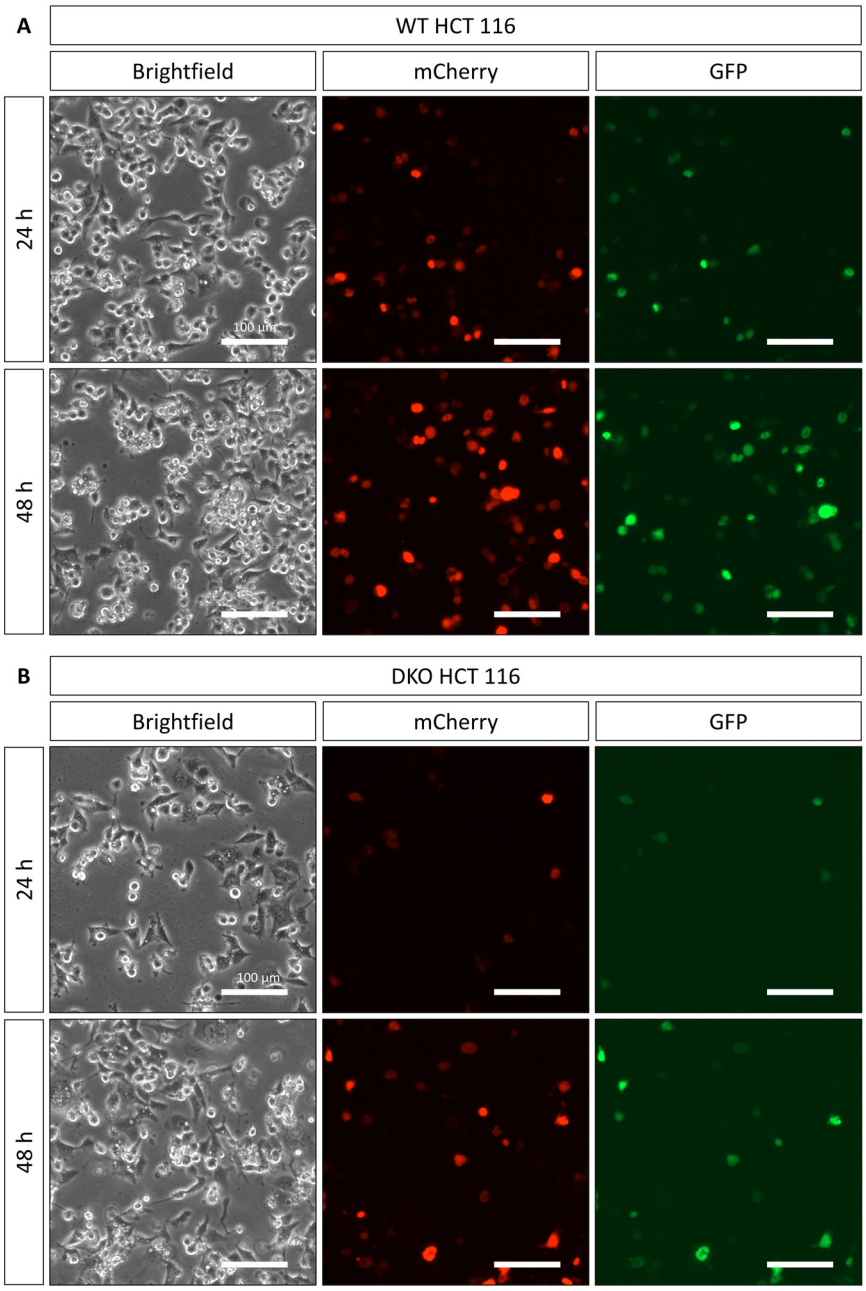


Figure 50: Transfection efficiency of Lipofectamine 2000 in WT and DKO HCT116 cells. Pictures of WT **(A)** and DKO **(B)** HCT116 cells were taken 24 h and 48 h after transfection with 2000 ng p1559, 1000 ng p1169 and 9 µL transfection reagent. Scale bar is 100 µm.

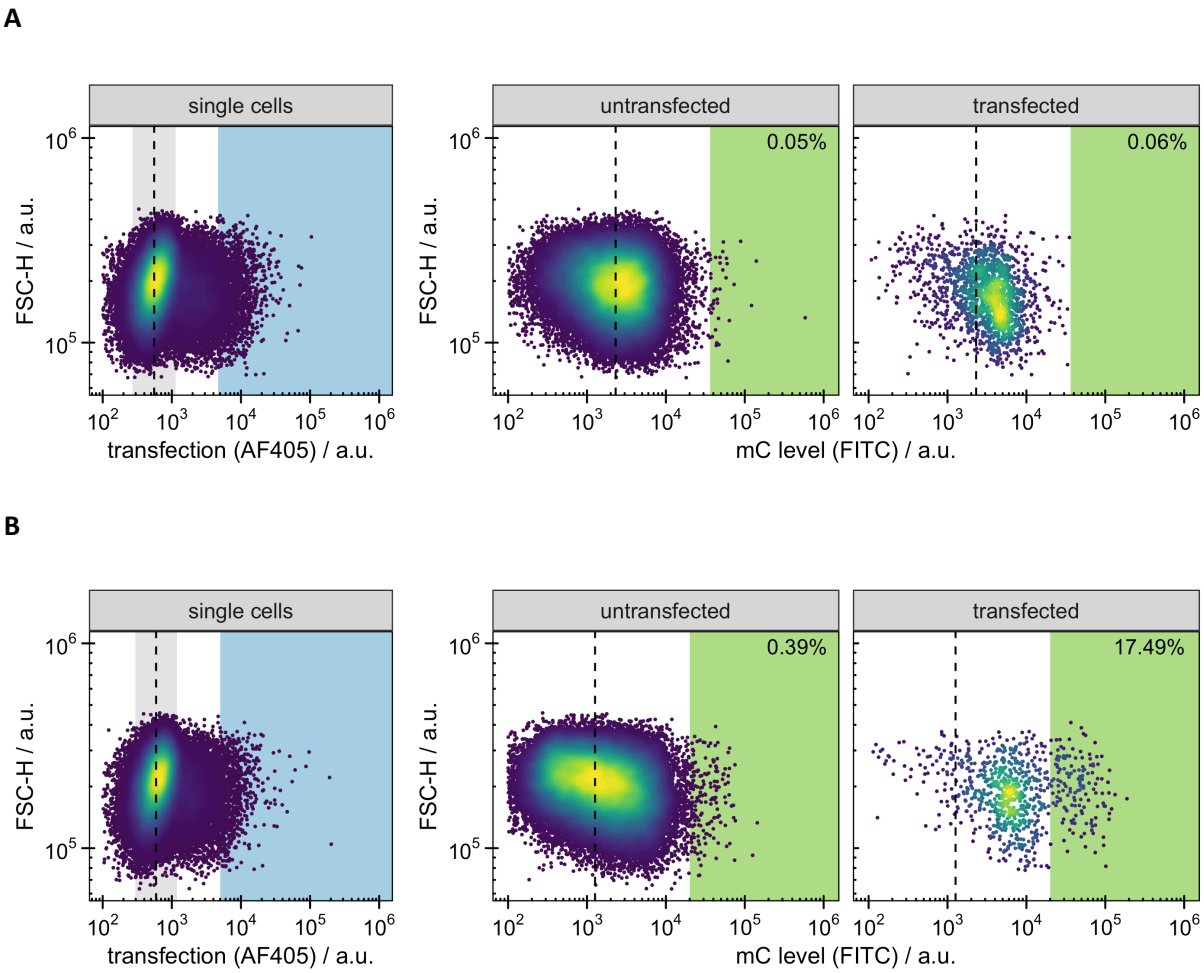


Figure 51: Analysis of flow cytometry data of pcDNMT3a3L activity in DKO HCT116 cells. Representative images of flow cytometry data analyzed with R. **A:** One replicate of wild type pcDNMT3a3L w/o light irradiation. **B:** One replicate of wild type pcDNMT3a3L with 5 min light irradiation.
The blue gate describes the transfected cells (immunostaining of HA-tag with AF405-labeled secondary antibodies). The threshold is defined as 8.4 times above the median of untransfected cells (dashed line). The green gate in the second and third plot describes the 5mC-positive cells (based on immunostaining of 5mC with FITC) and its threshold is defined as 15.4 times above the median of untransfected cells (dashed line). Modified from Wolffgramm et al.[103]

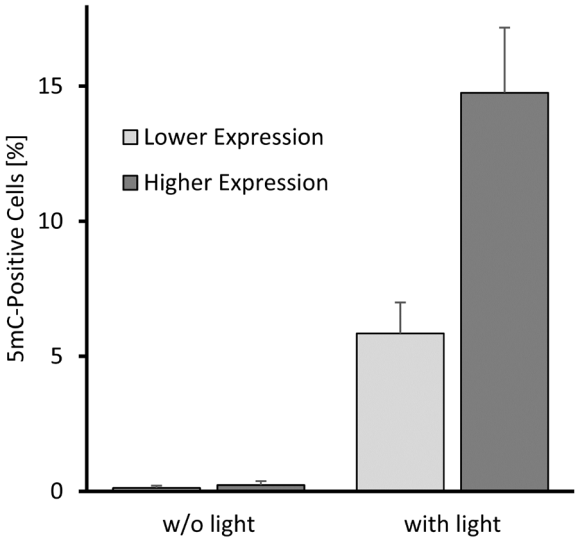


Figure 52: Activity of pcDNMT3a3L in DKO HCT116 cells with different expression levels. Cells were transfected with pcDNMT3a3L and irradiated with light for 5 min 48 h later. After another 24 h, cells were grouped into “lower expression”, or “higher expression” based on their HA immunostaining of pcDNMT3a3L. Lower expression was defined as 6 to 8 times the median and higher expression with a threshold of 8.4 times above the median of HA immunostaining of untransfected cells. Modified from Wolffgramm et al.[103]

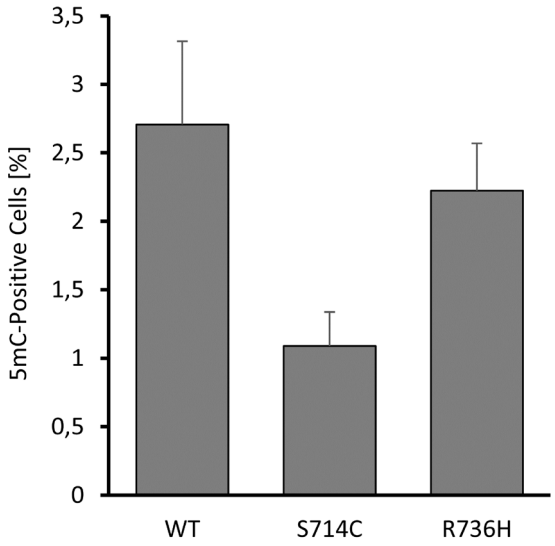


Figure 53: Activity of DNMT3a3L WT, S714C and R736H in DKO HCT116 cells. Cells were analyzed 48 h after transfection with WT DNMT3a3L, DNMT3a3L^{S714C} or DNMT3a3L^{R736H} by immunostaining of HA-tagged DNMT3a3L and 5mC. Data was analyzed according to Figure 51 and used for Figure 24. For this, data of DNMT3a3L and DNMT3a3L^{S714C} was normalized based on the mean activity of DNMT3a3L^{R736H}. Error bars show standard deviations from three independent biological replicates.

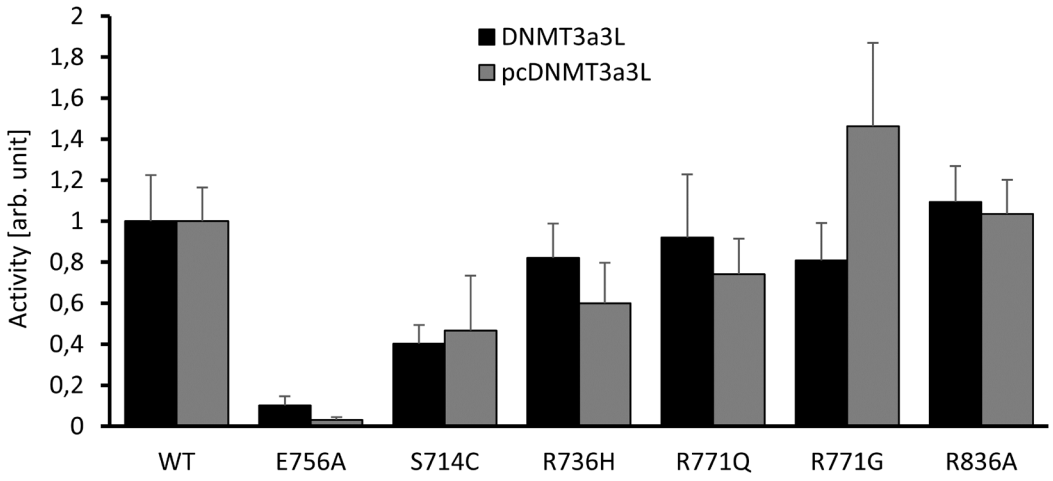


Figure 54: Comparison of DNMT3a3L- and pcDNMT3a3L-mutants in DKO HCT116 cells. Data shown in Figure 24 and Figure 25 were each normalized to WT DNMT3a3L. For DNMT3a3L w/o C710amber mutation, cells were analyzed 48 h after transfection. For DNMT3a3L with C710amber mutation, cells were 5 min light irradiated 48 h after transfection and analyzed after another 24 h incubation. Error bars show standard deviations from three independent biological replicates.

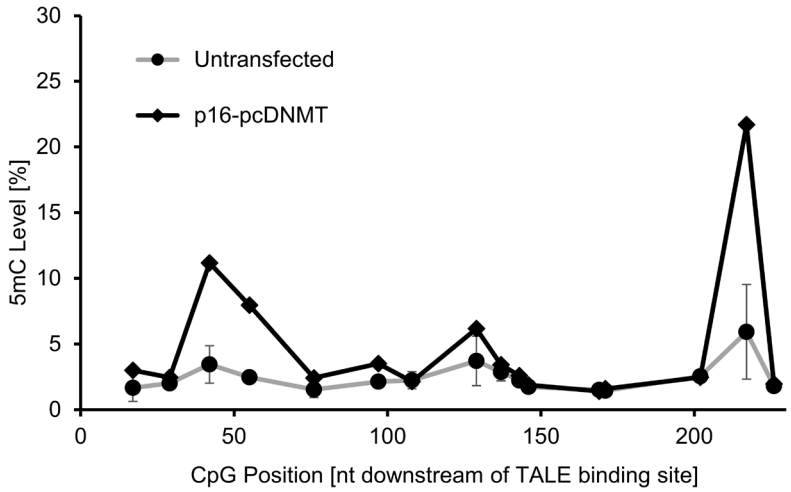


Figure 55: Methylation of p16 gene in p16-pcDNMT transfected HEK293T cells. Cells were transfected with p16-pcDNMT and irradiated with 5 min light 24 h later. After another 24 h incubation, 5mC levels of the CpGs 18 to 227 nt downstream of the p16-pcDNMT binding site were analyzed by Illumina sequencing. Numbering starting with 5'-T of TALE target site. Experiment was done as single replicate for p16-pcDNMT. For untransfected cells, error bars show standard deviations from two independent biological replicates.

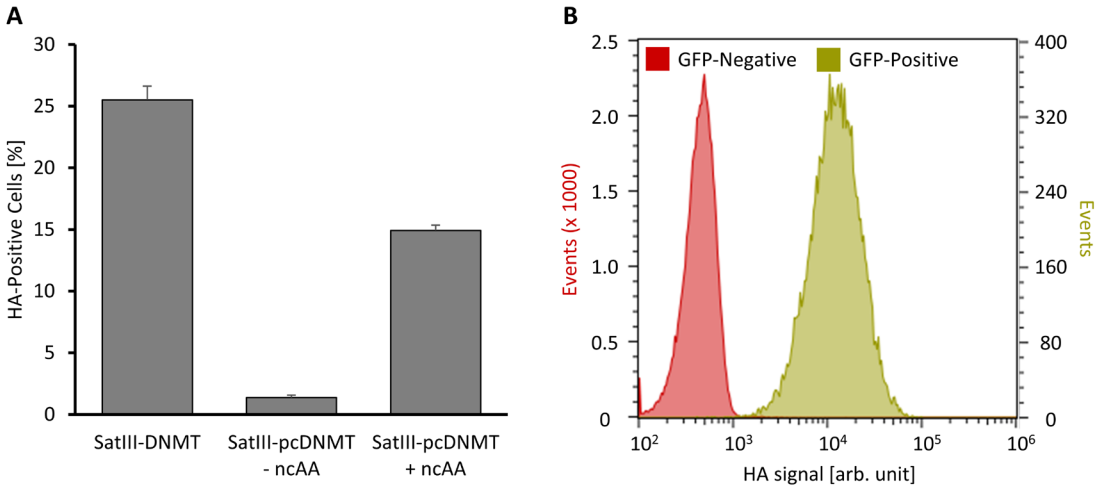


Figure 56: Expression of SatIII-DNMT and SatIII-pcDNMT in HEK293T cells. A: Expression levels of SatIII-DNMT and SatIII-pcDNMT in the absence or presence of 0.05 mM DMNB-Cys analyzed by immunostaining 24 h after transfection. Cells were transfected with either p1996 or p1998/p1169 and stained with anti-HA primary and FITC labeled secondary antibodies. SatIII-pcDNMT is selectively expressed only in the presence of DMNB-Cys. **B:** Histograms of SatIII-pcDNMT expressing cells (staining of HA-tag with AF405) in GFP-negative or -positive HEK293T cells 24 h after transfection with p1660/p1169 + 0.05 mM DMNB-Cys. Both expressions are correlated to each other.
Cells with a HA- or GFP-signal 100-times the median of untransfected cells or above were considered HA- or GFP-positive, respectively. Error bars show standard deviations from three independent biological replicates. Modified from Wolffgramm et al.[103]

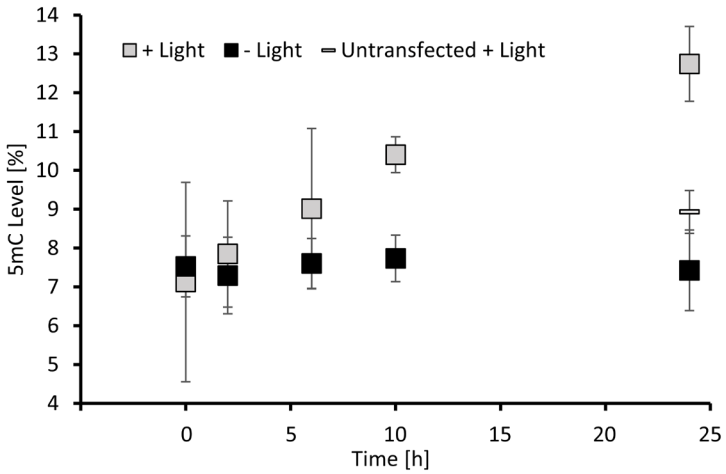


Figure 57: Kinetics of SatIII methylation - Second CpG. Kinetics of second CpG of SatIII methylation by SatIII-pcDNMT up to 24 h after 5 min light irradiation in HEK293T cells. As control, SatIII-pcDNMT w/o light for all time points and untransfected cells 24 h after light irradiation are shown. For SatIII sequence see 8.3.
Error bars show standard deviations from three independent biological replicates. p1660/p1169 or p1666/p1169 with 0.05 µM DMNB-Cys was used and 5mC levels were measured by Illumina sequencing of bisulfite converted and amplified gDNA. Modified from Wolffgramm et al.[103]

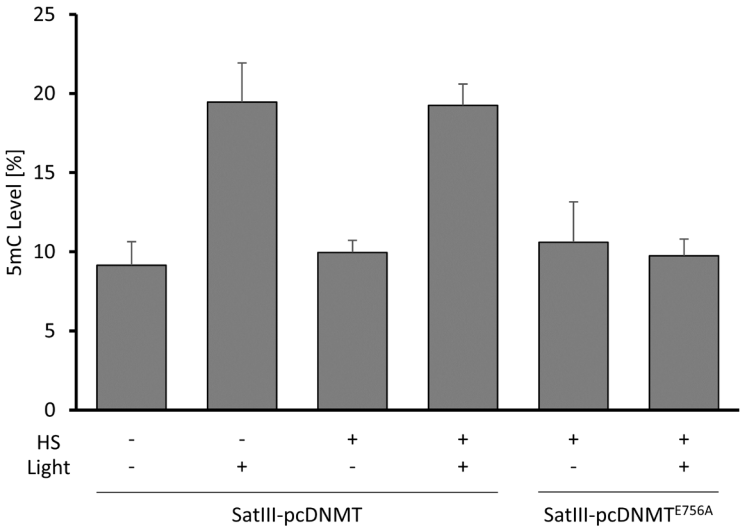


Figure 58: SatIII methylation with SatIII-pcDNMT in heat stressed HEK293T cells. Cells were transfected either with SatIII-pcDNMT or inactive SatIII-pcDNMT^{E756A}. 24 h later, cells were irradiated with light for 5 min (no light for controls) and incubated another 24 h before the HS of 1 h at 44 °C. Then, gDNA was isolated, bisulfite converted and the SatIII methylation level was analyzed by pyrosequencing. Error bars show standard deviations from two independent biological replicates. Modified from Wolffgramm et al.[103]

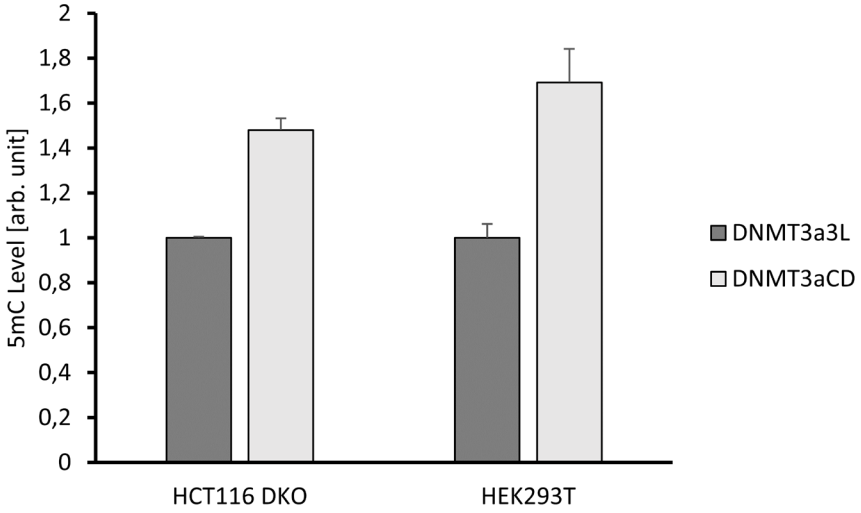


Figure 59: Activity of DNMT3a3L and DNMT3aCD in DKO HCT116 and HEK293T cells. Cells were transfected with either DNMT3a3L or DNMT3aCD, and immunostained with HA- and 5mC-antibodies 48 h (for DKO HCT116 cells) or 24 h (for HEK293T) after transfection. Shown is the respective median 5mC value of HA-positive cells normalized to DNMT3a3L. Cells with a HA-value between 8.4 times to 18.65 times of the mode HA-value of untransfected cells were defined as HA-positive. Error bars show standard deviations from three independent biological replicates.

8.2. Protein Sequences

DNMT3a3L

MASPKKKRKVGRANHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVD
RYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSP**C**NDLSIVNP
ARKGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNP
VMIDAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRS
NSIKQGKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSV
PVIRHLFAPLKEYFACV*SSGNSNANSRGPSFSSGLVPLSLRGSHM*GPMEIYKTVSAWKR
QPVRVLSLFRNIDKVLKSLGFLESGSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGS
TQPLGSSCDRCPGWYMFQFHRILQYALPRQESQRPFFWIFMDNLLLTEDDQETTTRFLQ
TEAVTLQDVRGRDYQNAMRVWSNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDL
LVKNCLLPLREYFKYFSQNSLPLINYPYDVPDYAS*

SV40 NLS1

human DNMT3A part (amino acids 612-912, with catalytic C710 in red)

S714, R736, E756, R771, T835, R836, R882 and W893 are highlighted in grey

Linker

mouse DNMT3L part (amino acids 208-421)

HA Tag

Properties: 568 amino acids; 64.94 kDa

DNMT3aCD

MASPKKKRKVGRANHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVD
RYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSP**C**NDLSIVNP
ARKGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNP
VMIDAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRS
NSIKQGKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSV
PVIRHLFAPLKEYFACVINYPYDVPDYAS*

SV40 NLS1

human DNMT3A part (amino acids 612-912, with catalytic C710 in red)

S714, R736, E756, R771, T835, R836, R882 and W893 are highlighted in grey

HA Tag

Properties: 326 amino acids; 37.37 kDa

p16-DNMT

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRGSVDLRTLGYSQQQQEKIKPKVRS
TVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVTYQHIITALPEATHEDIVGVGKQWSG
ARALEALLTDAGELRGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNALTGAPLNLTPDQV
VAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQ
DHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETV
QRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNG
GGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPD
QVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVL
CQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALE
TVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASH
DGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTP
DQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLC
QDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALET
VQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASN
GGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTP
DQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALESIVAQLSRPDPALAALTNDHLVALACLGGRPAM
DAVKKGLPHAPELIRRVNRRIGERTSHRVAGSKASPKKKRKVGRANHDQEFDPPKVYPP
VPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVGD
VRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPKEG
DDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRARYFWGNLPGMNRP
LASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDILWCTEM
ERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACV*SSGNSNANSRG*
*PSFSSGLVPLSLRGSHM*GPMEIYKTVSAWKRQPVRVLSLFRNIDKVLKSLGFLESGSGS
GGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQFHRILQYAL
PRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQNAMRVWSNIPGL
KSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYFKYFSQNSLPLINYPY
DVPDYAS*

Flag Tag

SV40 NLS1

p16-TALE (with **RVDs**)

human DNMT3A part (amino acids 612-912, with catalytic C710 in red and E756)

Linker

mouse DNMT3L part (amino acids 208-421)

HA Tag

Properties: 1607 amino acids; 174.42 kDa

BRCA1-DNMT

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRGSVDLRTLGYSQQQQEKIKPKVRS
TVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVTYQHIITALPEATHEDIVGVGKQWSG
ARALEALLTDAGELRGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNALTGAPLNLTPDQV
VAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NI**GGKQALETVQRLLPVLCQD
HGLTPDQVVAIAS**NN**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGKQALETVQ
RLLPVLCQDHGLTPDQVVAIAS**NI**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GG
KQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGKQALETVQRLLPVLCQDHGLTPDQV
VAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQ
DHGLTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NG**GGKQALETV
QRLLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NI**G
GKQALETVQRLLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQ
VVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLC
QDHGLTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**HD**GGKQALET
VQRLLPVLCQDHGLTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**N**
NGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTP
DQVVAIAS**NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPV
LCQDHGLTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGKQAL
ESIVAQLSRPDPALAALTNDHLVALACLGGRPAMDAVKKGLPHAPELIRRVNRRIGERTS
HRVAGSKASPKKKRKVGRANHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKD
LGIQVDRYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSP**C**ND
LSIVNPARKGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRF
LESNPVMIDAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVR
TITTRSNSIKQGKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLG
RSWSVPVIRHLFAPLKEYFACV*SSGNSNANSRGPSFSSGLVPLSLRGSHM*GPMEIYKTV
SAWKRQPVRVLSLFRNIDKVLKSLGFLESGSGSGGGTLKYVEDVTNVVRRDVEKWGPF
DLVYGSTQPLGSSCDRCPGWYMFQFHRILQYALPRQESQRPFFWIFMDNLLLTEDDQET
TTRFLQTEAVTLQDVRGRDYQNAMRVWSNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLD
APKVDLLVKNCLLPLREYFKYFSQNSLPLINYPYDVPDYAS*

Flag Tag

SV40 NLS1

BRCA1-TALE (with **RVDs**)

human DNMT3A part (amino acids 612-912, with catalytic C710 in red and E756)

Linker

mouse DNMT3L part (amino acids 208-421)

HA Tag

Properties: 1641 amino acids; 178.02 kDa

SatIII-DNMT

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRGSVDLRTLGYSQQQQEKIKPKVRS
TVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVTYQHIITALPEATHEDIVGVGKQWSG
ARALEALLTDAGELRGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNALTGAPLNLTPDQV
VAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQD
HGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQ
RLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDG
GKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQV
VAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQ
DHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETV
QRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGG
GKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQ
VVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLC
QDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETV
QRLLPVLCQDHGLTPDQVVAIASNGGGKQALESIVAQLSRPDPALAALTNDHLVALACLG
GRPAMDAVKKGLPHAPELIRRVNRRIGERTSHRVAGSKASPKKKRKVGRANHDQEFDP
PKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKI
MYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHD
ARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRARYFWGNL
PGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDIL
WCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACV*SSGNS*
*NANSRGPSFSSGLVPLSLRGSHM*GPMEIYKTVSAWKRQPVRVLSLFRNIDKVLKSLGFLE
SGSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQFHR
ILQYALPRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQNAMRVW
SNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYFKYFSQNSLPL
INYPYDVPDYAS*

Flag Tag

SV40 NLS1

SatIII-TALE (with **RVDs**)

human DNMT3A part (amino acids 612-912, with catalytic C710 in red and E756)

Linker

mouse DNMT3L part (amino acids 208-421)

HA Tag

Properties: 1437 amino acids; 156.40 kDa

mCherry-GFP transfection control

MGRLESTPPKKKRKVEDSASDYKDDDDKVAYAVSKGEEDNMAIIKEFMRFKVHMEGSV
NGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPD
YLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKK
TMGWEASSERMYPEDGALKGEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIK
LDITSHNEDYTIVEQYERAEGRHSTGGMDELYKTLQEFPPPPASDYKDDDDKVSKGEEL
FTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYG
VQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELK
GIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKANFKIRHNIEDGSVQLADHYQQNT
PIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKHHHHHH*

SV40 NLS1

Flag Tag

mCherry

GFP (Y39 which was mutated to amber stop codon TAG is highlighted in grey)

His Tag

Properties: 530 amino acids; 59.96 kDa

8.3. SatIII Sequence

TALE binding site is highlighted and assembled according to the antisense sequence. Methylation was always measured from CpGs on the sense strand. Methylation level of first CpG was analyzed in all experiments using pyrosequencing. Methylation level of second CpG was only analyzed by Illumina DNA sequencing.

Sense:

AATCAACCCGAGTGCAATCGAATGGAATCGAATGAATGGAATGCAATGGAATGGATT
CAACTTGAATGGAATGGAAAGAATGGAATCAACACGAGTGGAATGGCATGGATTGGA
ATGGAATGGAATGGAATCAACCCGAGTACAGGAATGGAATGGAA

Antisense:

TTCCATTCCATTCCTGTACTCGGGTTGATTCCATTCCATTCCATTCCAATCCATGCCAT
TCCACTCGTGTTGATTCCATTCTTTCCATTCCATTCAAGTTGAATCCATTCCATTGCAT
TCCATTCATTCGATTCCATTCGATTGCACTCGGGTTGATT

8.4. Transfection Protocols

Provided DMNB-Cys concentration was used only for samples with C710amber mutation.

Experiment	Cells	DMNB-Cys [mM]	Plasmids
Figure 16	HEK293T	0.5	pJaW1660 pStH1169
Figure 17	HEK293T	0.05	pJaW1559 pStH1169
Figure 18	HEK293T	0.05 and 0.5	pJaW1559 pStH1169
Figure 19	---	1	---
Figure 20	HEK293T and 5-Aza HEK293T	0.05	pJaW1848 pJaW1849 pStH1169
Figure 21	WT and DKO HCT116	---	pJaW1848 pJaW1850
Figure 22	DKO HCT116	0.05	pJaW1848 pJaW1849 pJaW1851 pStH1169
Figure 24	DKO HCT116	---	pJaW1848 pJaW1850 pAlH1898 pAlH1899 pAlH1900 pAlH1901 pAlH1902 pAlH1903 pAlH1904 pAlH1905 pAlH1906 pAlH1907
Figure 25	DKO HCT116	0.05	pJaW1849 pJaW1978 pJaW1979 pJaW1980 pJaW1981 pJaW1984 pStH1169
Figure 26	HEK293T	1	pJaW926 pJaW1125 pStH1169
Figure 27	HEK293T	---	pJaW926 pJaW935
Figure 28	HEK293T	0.05 - 0.5	pJaW1660 pStH1169
Figure 29	HEK293T	0.05	pJaW1998 pJaW1999 pStH1169

Experiment	Cells	DMNB-Cys [mM]	Plasmids
Figure 30	HEK293T	0.05	pJaW1648 pJaW1654 pJaW1660 pJaW1666 pStH1169
Figure 31	HEK293T	0.05	pJaW926 pJaW1002 pJaW1660 pStH1169
Figure 32	HEK293T	0.05	pJaW1660 pStH1169
Figure 33	HEK293T	0.05	pJaW2559 pStH1169
Figure 38	HEK293T	0.5	pJaW1660 pStH1169
Figure 39	HEK293T	0.05	pJaW1559 pStH1169
Figure 40	---	1	---
Figure 41	HEK293T	---	---
Figure 42	HEK293T	0.05	pJaW1660 pStH1169
Figure 43	HEK293T	---	---
Figure 44	HEK293T and 5-Aza	0.05	pJaW1848 pJaW1849 pStH1169
Figure 45	WT and DKO HCT116	---	pJaW1848 pJaW1850
Figure 46	DKO HCT116	---	---
Figure 47	WT and DKO HCT116	0.05	pJaW1559 pStH1169
Figure 48	WT and DKO HCT116	0.05	pJaW1559 pStH1169
Figure 49	WT and DKO HCT116	0.05	pJaW1559 pStH1169
Figure 50	WT and DKO HCT116	0.05	pJaW1559 pStH1169
Figure 51	DKO HCT116	0.05	pJaW1849 pStH1169
Figure 52	DKO HCT116	0.05	pJaW1849 pStH1169
Figure 53	DKO HCT116	---	pJaW1848 pAlH1898 pAlH1899

Experiment	Cells	DMNB-Cys [mM]	Plasmids
Figure 54	DKO HCT116	0.05	pJaW1848 pJaW1850 pAlH1898 pAlH1899 pAlH1900 pAlH1901 pAlH1904 pJaW1849 pJaW1851 pJaW1978 pJaW1979 pJaW1980 pJaW1981 pJaW1984 pStH1169
Figure 55	HEK293T	0.05	pJaW1125 pStH1169
Figure 56	HEK293T	0.05	pJaW1648 pJaW1660 pStH1169
Figure 57	HEK293T	0.05	pJaW1660 pJaW1666 pStH1169
Figure 58	HEK293T	0.05	pJaW1660 pJaW1666 pStH1169
Figure 59	HEK293T and DKO HCT116	---	pJaW1848 pAlH1895

8.5. Plasmid Maps

All plasmid maps were created with SnapGene (from Insightful Science; available at snapgene.com).

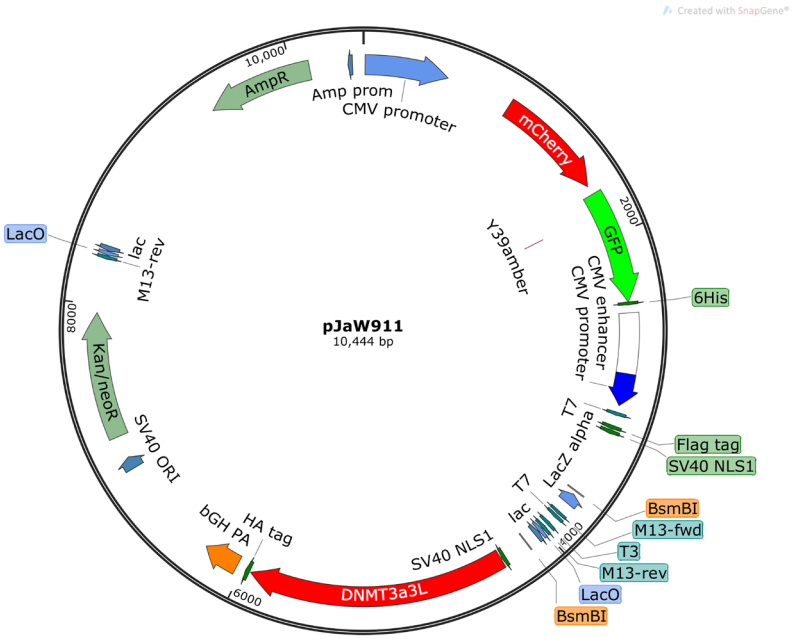


Figure 60: Plasmid map of pJaW911.
Golden Gate 2 entry plasmid with DNMT3a3L and the transfection control mCherry-GFP^{Y39amber}.

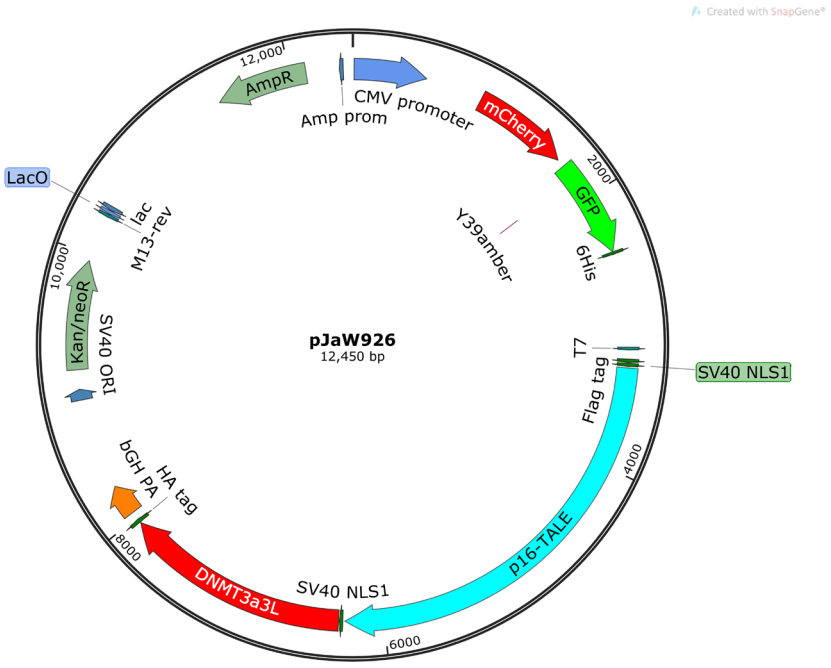


Figure 61: Plasmid map of pJaW926.
p16-DNMT construct and the transfection control mCherry-GFP^{Y39amber}.

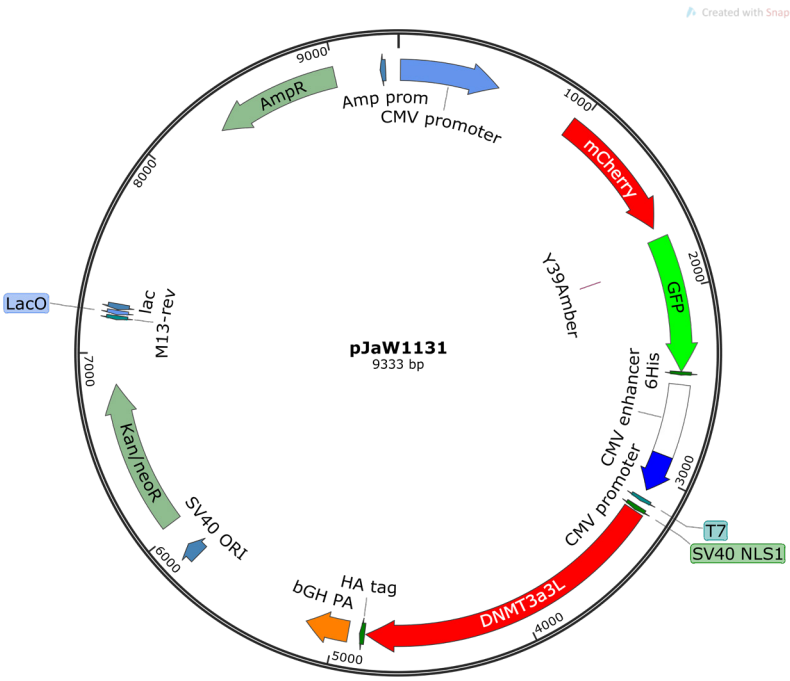


Figure 62: Plasmid map of pJaW1131.
Global DNMT3a3L and the transfection control mCherry-GFP^{Y39amber}.

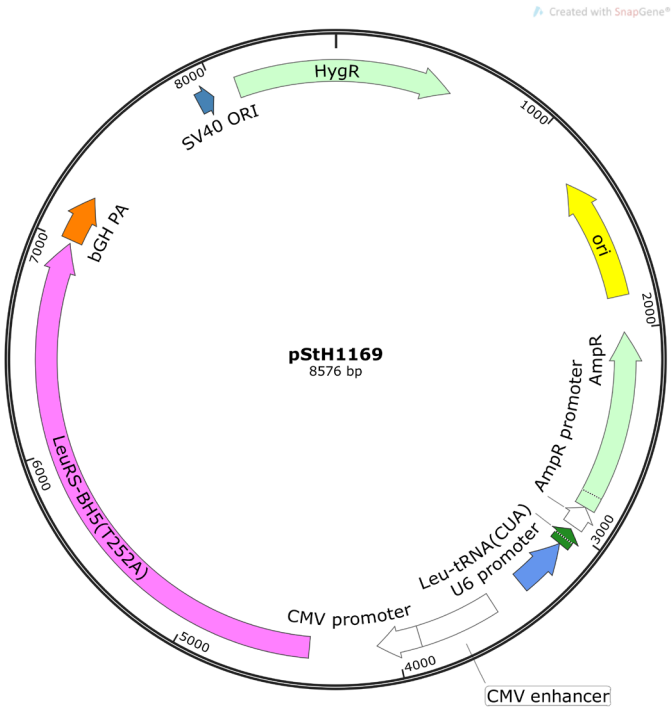


Figure 63: Plasmid map of pStH1169.
Encodes Leu-tRNA$_{CUA}$ and synthetase pLRS_BH5^{T252A}.

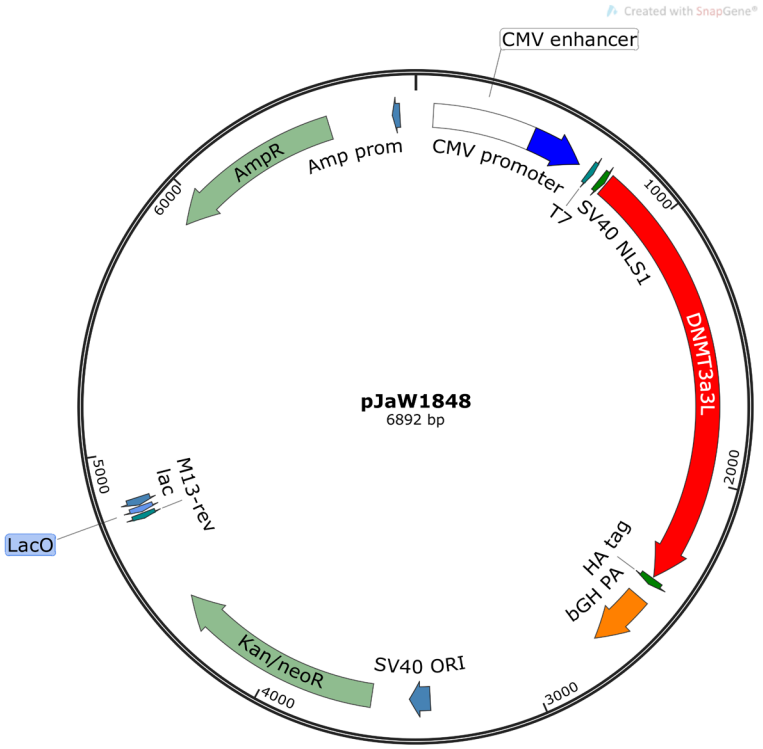


Figure 64: Plasmid map of pJaW1848.
Global DNMT3a3L construct w/o transfection control.

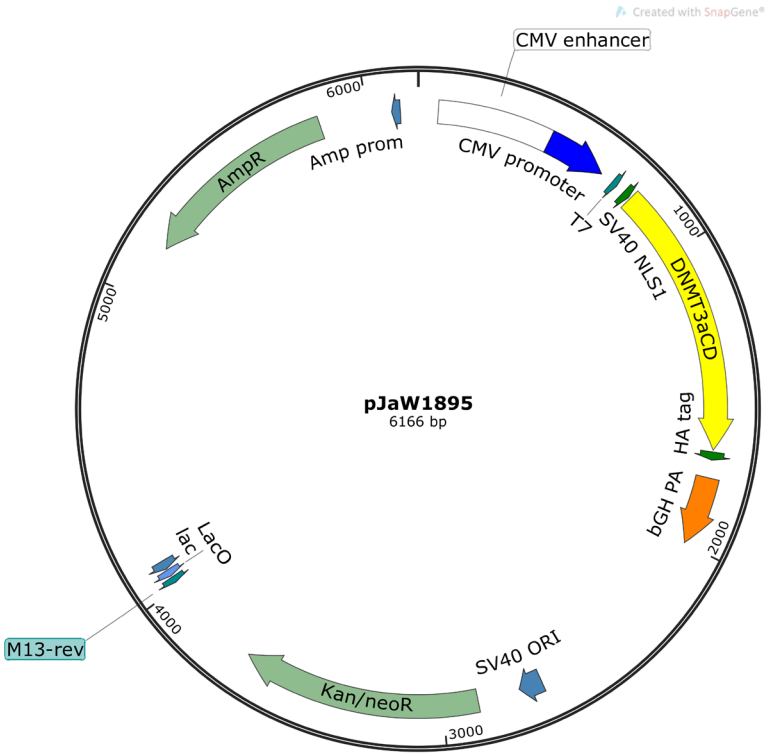


Figure 65: Plasmid map of pJaW1895.
Encodes global DNMT3aCD w/o transfection control.

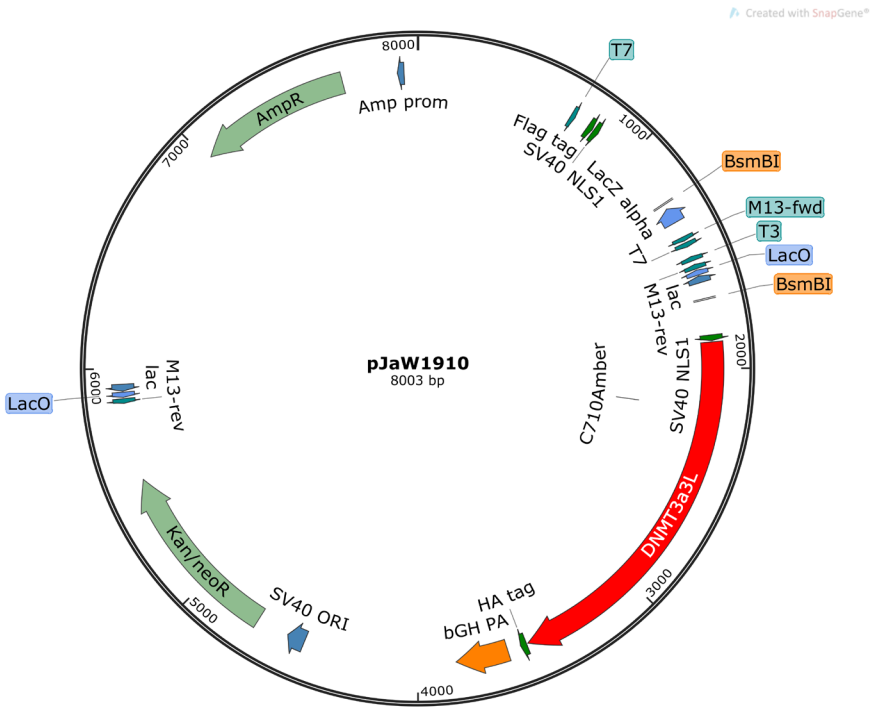


Figure 66: Plasmid map of pJaW1910.
Golden Gate 2 entry plasmid with pcDNMT3a3L (DNMT3a3L^{C710amber}) w/o transfection control.

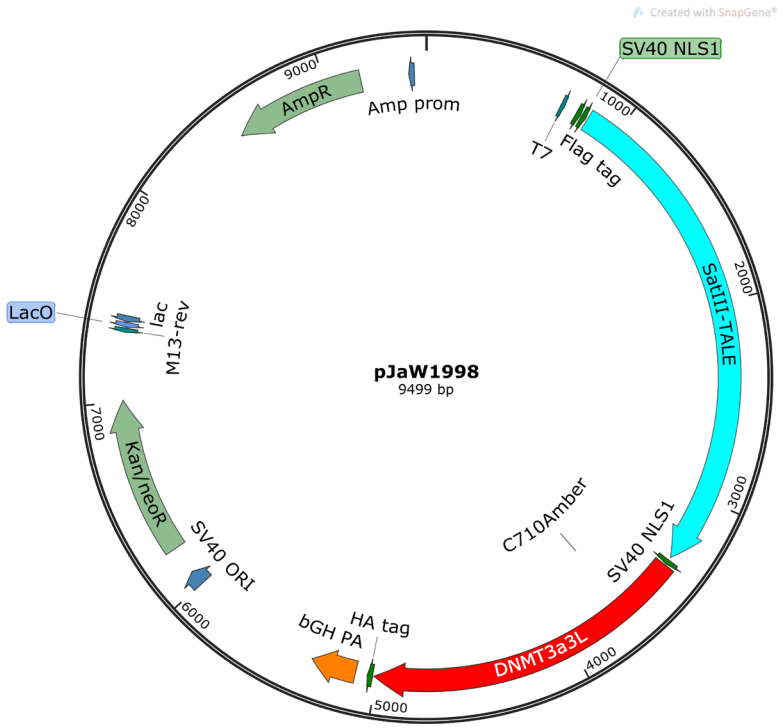


Figure 67: Plasmid map of pJaW1998.
Encodes SatIII-pcDNMT (DNMT3a3L^{C710amber}) construct w/o transfection control.

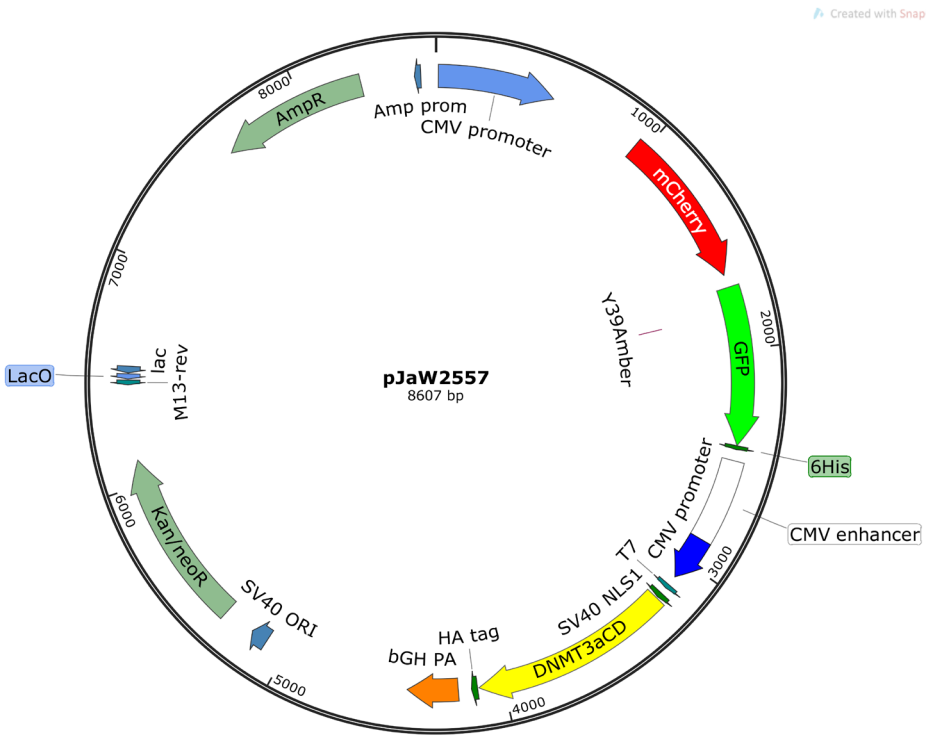


Figure 68: Plasmid map of pJaW2557.
Encodes global DNMT3aCD and the transfection control mCherry-GFP^{Y39amber}.

8.6. Additional Data of Mass Spectrometry

8.6.1. mCherry-GFP^{Y39amber}

Table 24: PEP-number and score of the detected peptide of mCherry-GFP^{Y39amber}. CAM: Carbamidomethyl; PEP: posterior error probability; Score: Andromeda score of the best identified modified peptide containing this site.

Peptide	Light	MS/MS m/z	Charge	MS/MS Count	PEP	Score	Intensity
FSVSGEGEG DATC(CAM)GK	yes	750.8187	2	1	0.00013612	55.426	$5.6*10^{5}$
FSVSGEGEG DATC(DMNB)GK	yes	819.8307	2	2	$3.86*10^{-138}$	196.16	$2.0*10^{7}$
FSVSGEGEG DATC(CAM)GK	no	Not detected					
FSVSGEGEG DATC(DMNB)GK	no	819.8318	2	2	$1.40*10^{-158}$	208.17	$3.6*10^{7}$

Table 25: Exact theoretical and measured masses of peptide fragments shown in Figure 17A.

Peptide Fragment	Theoretical Mass [u]	Measured Mass [u]	ΔMass
y1	147.1128	147.1128	0.0000
y3	502.1966	502.1966	0.0000
y4	603.2443	603.2443	0.0000
y5	674.2814	674.2814	0.0000
y6	789.3084	789.3088	-0.0004
y7	846.3298	846.3289	0.0009
y8	975.3724	975.3724	0.0000
y9	1032.3939	1032.394	-0.0001
y10	1161.4365	1161.435	0.0015
y11	1218.4579	1218.458	-0.0001
y12	1305.4900	1305.49	0.0000
y13	1404.5584	1404.558	0.0004
b2	235.1077	235.1077	0.0000
b4	421.2081	421.2082	-0.0001
b14	1492.5532	1492.553	0.0002

Table 26: Exact theoretical and measured masses of peptide fragments shown in Figure 17B.

Peptide Fragment	Theoretical Mass [u]	Measured Mass [u]	ΔMass
y1	147.1128	147.1128	0.0000
y4	465.2126	465.2125	0.0001
y7	708.2981	708.2981	0.0000
y9	894.3622	894.3522	0.0100
y11	1080.4262	1080.425	0.0012
y12	1167.4583	1167.458	0.0003
b2	235.1077	235.1077	0.0000

Table 27: Exact theoretical and measured masses of peptide fragments shown in Figure 39.

Peptide Fragment	Theoretical Mass [u]	Measured Mass [u]	ΔMass
y1	147.1128	147.1128	0.0000
y2	204.1342	204.1343	-0.0001
y3	502.1966	502.1966	0.0000
y4	603.2443	603.2443	0.0000
y5	674.2814	674.2814	0.0000
y6	789.3084	789.3083	0.0001
y7	846.3298	846.3298	0.0000
y8	975.3724	975.3724	0.0000
y9	1032.3939	1032.394	-0.0001
y10	1161.4365	1161.436	0.0005
y11	1218.4579	1218.458	-0.0001
y12	1305.4900	1305.49	0.0000
y13	1404.5584	1404.558	0.0004
b2	235.1077	235.1077	0.0000
b3	334.1761	334.1761	0.0000
b4	421.2081	421.2082	-0.0001
b6	607.2722	607.2722	0.0000
b14	1492.5532	1492.553	0.0002

8.6.2. DNMT3a3L

Detected fragments of DNMT3a3L:

C710 which is used for genetic code expansion and incorporation of DMNB-Cys is shown in red.

Detected fragments are shown in green.

With light

MASPKKKRKVGRANHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLK**DLGIQVD**
RYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSP**C**NDLSIVNP
ARKGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISR**FLESNP**
VMIDAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRS
NSIKQGKDQHFPVFMNEKEDILWCTEMER**VFGFPVHYTDVSNMSR**LARQRLLGR**SWSV**
PVIRHLFAPLKEYFACVSSGNSNANSR**GPSFSSGLVPLSLR**GSHMGPMEIYKTVSAWKR
QPVRVLSLFRNIDKVLK**SLGFLESGSGSGGGTLKYVEDVTNVVR**RDVEKWGPFDLVYGS
TQPLGSSCDRCPGWYMFQFHR**ILQYALPR**QESQRPFFWIFMDNLLLTEDDQETTTR**FL**
QTEAVTLQDVRGRDYQNAMR**VWSNIPGLK**SK**HAPLTPKEEEYLQAQVR**SRSKLDAPKV
DLLVKNCLLPLREYFKYFSQNSLPLINYPYDVPDYAS*

Without light

MASPKKKRKVGRANHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLK**DLGIQVD**
RYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSP**C**NDLSIVNP
AR**KGLYEGTGR**LFFEFYR**LLHDARPK**EGDDRPFFWLFENVVAMGVSDKRDISR**FLESN**
PVMIDAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTR
SNSIKQGKDQHFPVFMNEKEDILWCTEMER**VFGFPVHYTDVSNMSR**LARQRLLGRSWS
VPVIR**HLFAPLK**EYFACVSSGNSNANSR**GPSFSSGLVPLSLR**GSHMGPMEIYKTVSAWK
RQPVRVLSLFRNIDKVLK**SLGFLESGSGSGGGTLKYVEDVTNVVR**RDVEKWGPFDLVY
GSTQPLGSSCDRCPGWYMFQFHR**ILQYALPR**QESQRPFFWIFMDNLLLTEDDQETTTR
FLQTEAVTLQDVRGRDYQNAMR**VWSNIPGLK**SKHAPLTPK**EEEYLQAQVR**SRSKLDAP
KVDLLVKNCLLPLREYFKYFSQNSLPLINYPYDVPDYAS*

8.7. Additional Data of SatIII Time Course Sequencing

Table 28: Information about SatIII bisulfite-converted amplicons. Data adapted from Wolffgramm et al.[103]

Run ID	Treatment	Time Point	Replicate	Primer fw	Primer rv	Barcode fw
Sample1	with_light	0	Rep1	o2363	o3033	ATCACG
Sample1	with_light	0	Rep2	o2364	o3034	CGATGT
Sample1	with_light	0	Rep3	o2365	o3035	TTAGGC
Sample1	no_light	0	Rep1	o2366	o3036	TGACCA
Sample1	no_light	0	Rep2	o2367	o3037	ACAGTG
Sample1	no_light	0	Rep3	o2368	o3038	GCCAAT
Sample1	with_light	2	Rep1	o2369	o3039	CAGATC
Sample1	with_light	2	Rep2	o2370	o3040	ACTTGA
Sample1	with_light	2	Rep3	o2371	o3041	GATCAG
Sample1	no_light	2	Rep1	o2372	o3042	TAGCTT
Sample1	no_light	2	Rep2	o2373	o3043	GGCTAC
Sample1	no_light	2	Rep3	o2374	o3044	CTTGTA
Sample1	with_light	6	Rep1	o2375	o3045	AGTCAA
Sample1	with_light	6	Rep2	o2376	o3046	AGTTCC
Sample1	with_light	6	Rep3	o2377	o3047	ATGTCA
Sample1	no_light	6	Rep1	o2378	o3048	CCGTCC
Sample1	no_light	6	Rep2	o2379	o3049	GTAGAG
Sample1	no_light	6	Rep3	o2380	o3050	GTCCGC
Sample1	with_light	10	Rep1	o2381	o3051	GTGAAA
Sample2	with_light	10	Rep2	o2382	o3052	GTGGCC
Sample2	with_light	10	Rep3	o2383	o3053	GTTTCG
Sample2	no_light	10	Rep1	o2384	o3054	CGTACG
Sample2	no_light	10	Rep2	o2385	o3055	GAGTGG
Sample2	no_light	10	Rep3	o2386	o3056	GGTAGC
Sample2	with_light	24	Rep1	o2387	o3057	ACTGAT
Sample2	with_light	24	Rep2	o2388	o3058	ATGAGC
Sample2	with_light	24	Rep3	o2389	o3059	ATTCCT
Sample2	no_light	24	Rep1	o2390	o3060	CAAAAG
Sample2	no_light	24	Rep2	o2391	o3061	CAACTA
Sample2	no_light	24	Rep3	o2392	o3062	CACCGG
Sample2	untransfected	24	Rep1	o2399	o3069	CGGAAT
Sample2	untransfected	24	Rep2	o2400	o3070	CTAGCT
Sample2	untransfected	24	Rep3	o2401	o3071	CTATAC

Table 29: Read counts of sequenced amplicons. Data adapted from Wolffgramm et al.[103]

Run ID	Raw Reads	Merged Reads	Mapped Reads	Assigned after Demultiplexing	CpG1 AND CpG2
Sample1	230,677	227,214 (99%)	219,207 (96%)	134,685 (61%)	114,296 (85%)
Sample2	190,166	185,797 (97%)	175,217 (95%)	120,999 (69%)	99,530 (82%)

Table 30: Methylation levels of individual samples for both analyzed SatIII CpGs. Data adapted from Wolffgramm et al.[103]

Time Point	Treatment	Replicate	Name	CG	TG	5mC [%]	Reads
0	Without Light	Rep1	First CpG	137	3898	3,40	4035
0	Without Light	Rep1	Second CpG	338	3697	8,38	4035
0	Without Light	Rep2	First CpG	288	6785	4,07	7073
0	Without Light	Rep2	Second CpG	483	6590	6,83	7073
0	Without Light	Rep3	First CpG	314	7236	4,16	7550
0	Without Light	Rep3	Second CpG	557	6993	7,38	7550
0	With Light	Rep1	First CpG	353	8555	3,96	8908
0	With Light	Rep1	Second CpG	802	8106	9,00	8908
0	With Light	Rep2	First CpG	172	3930	4,19	4102
0	With Light	Rep2	Second CpG	335	3767	8,17	4102
0	With Light	Rep3	First CpG	175	3060	5,41	3235
0	With Light	Rep3	Second CpG	136	3099	4,20	3235
2	Without Light	Rep1	First CpG	158	3096	4,86	3254
2	Without Light	Rep1	Second CpG	205	3049	6,30	3254
2	Without Light	Rep2	First CpG	159	3068	4,93	3227
2	Without Light	Rep2	Second CpG	267	2960	8,27	3227
2	Without Light	Rep3	First CpG	229	5275	4,16	5504
2	Without Light	Rep3	Second CpG	402	5102	7,30	5504
2	With Light	Rep1	First CpG	348	5195	6,28	5543
2	With Light	Rep1	Second CpG	384	5159	6,93	5543
2	With Light	Rep2	First CpG	341	4607	6,89	4948
2	With Light	Rep2	Second CpG	466	4482	9,42	4948
2	With Light	Rep3	First CpG	493	5927	7,68	6420
2	With Light	Rep3	Second CpG	462	5958	7,20	6420
6	Without Light	Rep1	First CpG	123	2650	4,44	2773
6	Without Light	Rep1	Second CpG	224	2549	8,08	2773
6	Without Light	Rep2	First CpG	371	8023	4,42	8394

Time Point	Treatment	Replicate	Name	CG	TG	5mC [%]	Reads
6	Without Light	Rep2	Second CpG	577	7817	6,87	8394
6	Without Light	Rep3	First CpG	291	7371	3,80	7662
6	Without Light	Rep3	Second CpG	602	7060	7,86	7662
6	With Light	Rep1	First CpG	831	7764	9,67	8595
6	With Light	Rep1	Second CpG	653	7942	7,60	8595
6	With Light	Rep2	First CpG	564	6262	8,26	6826
6	With Light	Rep2	Second CpG	550	6276	8,06	6826
6	With Light	Rep3	First CpG	880	6948	11,24	7828
6	With Light	Rep3	Second CpG	891	6937	11,38	7828
10	Without Light	Rep1	First CpG	213	1983	9,70	2196
10	Without Light	Rep1	Second CpG	184	2012	8,38	2196
10	Without Light	Rep2	First CpG	237	4612	4,89	4849
10	Without Light	Rep2	Second CpG	349	4500	7,20	4849
10	Without Light	Rep3	First CpG	192	4154	4,42	4346
10	Without Light	Rep3	Second CpG	331	4015	7,62	4346
10	With Light	Rep1	First CpG	1087	7331	12,91	8418
10	With Light	Rep1	Second CpG	834	7584	9,91	8418
10	With Light	Rep2	First CpG	688	4968	12,16	5656
10	With Light	Rep2	Second CpG	612	5044	10,82	5656
10	With Light	Rep3	First CpG	1627	5993	21,35	7620
10	With Light	Rep3	Second CpG	798	6822	10,47	7620
24	Without Light	Rep1	First CpG	207	3796	5,17	4003
24	Without Light	Rep1	Second CpG	332	3671	8,29	4003
24	Without Light	Rep2	First CpG	207	3507	5,57	3714
24	Without Light	Rep2	Second CpG	233	3481	6,27	3714
24	Without Light	Rep3	First CpG	163	2613	5,87	2776
24	Without Light	Rep3	Second CpG	214	2562	7,71	2776
24	Untransfected + Light	Rep1	First CpG	141	4664	2,93	4805
24	Untransfected + Light	Rep1	Second CpG	455	4350	9,47	4805
24	Untransfected + Light	Rep2	First CpG	318	5645	5,33	5963
24	Untransfected + Light	Rep2	Second CpG	499	5464	8,37	5963
24	Untransfected + Light	Rep3	First CpG	262	7271	3,48	7533
24	Untransfected + Light	Rep3	Second CpG	674	6859	8,95	7533
24	With Light	Rep1	First CpG	1232	5850	17,40	7082
24	With Light	Rep1	Second CpG	844	6238	11,92	7082

Time Point	Treatment	Replicate	Name	CG	TG	5mC [%]	Reads
24	With Light	Rep2	First CpG	1446	6987	17,15	8433
24	With Light	Rep2	Second CpG	1054	7379	12,50	8433
24	With Light	Rep3	First CpG	1275	4653	21,51	5928
24	With Light	Rep3	Second CpG	818	5110	13,80	5928

8.8. Additional Data of Transcriptome Analysis

Table 31: Description and statistics of RNA sequencing samples. Alignment was done with Salmon v. 1.3.0 against an expanded version of the human transcriptome including transcripts originating from the transfected plasmids. Data adapted from Wolffgramm et al.[103]

Run ID	Light	Time	Clean Reads	Mapping Rate	Mapped Counts [1e6]
V300055926_L2_HK500HUMhliEAAARAAPEI-509_1	yes	0h	36281971	90.37%	32.8
V300055926_L2_HK500HUMhliEAABRAAPEI-510_1	yes	0h	35360906	90.72%	32.1
V300055926_L2_HK500HUMhliEAACRAAPEI-511_1	yes	4h	35204944	89.92%	31.7
V300055926_L2_HK500HUMhliEAADRAAPEI-512_1	yes	4h	35348904	90.47%	32.0
V300055926_L2_HK500HUMhliEAAERAAPEI-513_1	yes	8h	43547913	89.78%	39.1
V300055926_L2_HK500HUMhliEAAFRAAPEI-514_1	yes	8h	41683354	90.93%	37.9
V300055926_L2_HK500HUMhliEAAIRAAPEI-517_1	no	0h	37410535	89.77%	33.6
V300055926_L2_HK500HUMhliEAAJRAAPEI-518_1	no	0h	36268131	89.88%	32.6
V300055926_L2_HK500HUMhliEAAKRAAPEI-519_1	no	4h	34718412	90.32%	31.4
V300055926_L2_HK500HUMhliEAALRAAPEI-520_1	no	4h	35230629	90.70%	32.0
V300055912_L3_HK500HUMhliEAAMRAAPEI-521_1	no	8h	34629709	89.84%	31.1
V300055912_L3_HK500HUMhliEAANRAAPEI-522_1	no	8h	36161999	89.21%	32.3

Table 32: Number of up- and down-regulated genes 4h and 8h after light irradiation. Absolute log2-fold change (LFC) and significance level of 0.05. Percentage reported with respect to 39,520 genes that had non-zero counts. Data adapted from Wolffgramm et al.[103]

Time	LFC	Up-Regulated	Down-Regulated
4h	>0	68 (0.17%)	30 (0.08%)
8h	>0	19 (0.05%)	79 (0.20%)

Table 33: Log2-fold changes of genes after 4h. Data shown in Figure 33. Data adapted from Wolffgramm et al.[103]

ENSEMBL	Symbol	BaseMean	Log2FoldChange	lfcSE	Stat	pvalue	padj
ENSG00000257225.1	–	14.13	-14.33	2.61	38.02	5.55E-09	1.99E-05
ENSG00000187664.9	HAPLN4	10.79	-13.40	2.73	24.03	6.04E-06	4.50E-03
ENSG00000259607.1	–	8.87	-8.78	2.18	16.97	2.06E-04	4.61E-02
ENSG00000275496.4	LOC102724701	48.52	-8.16	1.65	42.37	6.31E-10	3.77E-06
ENSG00000198237.8	–	21.42	-5.75	2.02	18.78	8.38E-05	2.54E-02
ENSG00000277399.4	GPR179	20.53	-5.70	1.62	16.69	2.37E-04	4.76E-02
ENSG00000196218.12	RYR1	44.14	-3.09	1.14	17.00	2.03E-04	4.61E-02
ENSG00000256591.5	–	73.85	-2.81	0.76	17.69	1.44E-04	3.58E-02
ENSG00000142178.9	SIK1	78.82	-1.47	1.20	19.10	7.11E-05	2.32E-02
ENSG00000130635.15	COL5A1	290.37	-1.22	0.35	27.92	8.64E-07	1.19E-03
ENSG00000143499.14	SMYD2	292.08	-1.18	0.35	17.05	1.98E-04	4.61E-02
ENSG00000187122.17	SLIT1	210.61	-1.10	0.48	24.54	4.69E-06	4.00E-03
ENSG00000135709.12	KIAA0513	269.09	-0.96	0.41	18.27	1.08E-04	3.02E-02
ENSG00000283154.2	SCHIP1	201.62	-0.77	0.34	19.94	4.67E-05	1.86E-02
ENSG00000155254.13	MARVELD1	500.88	-0.77	0.36	19.07	7.24E-05	2.32E-02
ENSG00000130940.15	CASZ1	581.09	-0.66	0.23	36.03	1.50E-08	4.48E-05
ENSG00000174705.13	SH3PXD2B	1138.35	-0.62	0.27	40.15	1.91E-09	8.55E-06
ENSG00000119446.14	RBM18	796.72	-0.50	0.27	34.68	2.95E-08	7.56E-05
ENSG00000076043.10	REXO2	428.79	-0.49	0.25	16.57	2.52E-04	4.80E-02
ENSG00000118197.14	DDX59	146.59	-0.44	0.41	16.48	2.63E-04	4.92E-02
ENSG00000109911.19	ELP4	554.11	-0.41	0.28	16.64	2.43E-04	4.76E-02
ENSG00000165533.18	TTC8	412.62	-0.28	0.42	16.48	2.64E-04	4.92E-02
ENSG00000155097.12	ATP6V1C1	475.48	-0.24	0.41	17.48	1.60E-04	3.92E-02
ENSG00000136156.15	ITM2B	1743.03	-0.24	0.22	20.03	4.47E-05	1.86E-02
ENSG00000104549.12	SQLE	741.09	-0.23	0.31	20.82	3.01E-05	1.50E-02
ENSG00000043514.17	TRIT1	1188.44	-0.14	0.24	16.86	2.19E-04	4.61E-02
ENSG00000133835.16	HSD17B4	970.84	-0.13	0.20	18.95	7.68E-05	2.37E-02
ENSG00000113558.18	SKP1	777.43	-0.05	0.23	21.21	2.48E-05	1.39E-02
ENSG00000152942.19	RAD17	448.81	-0.03	0.34	19.49	5.87E-05	2.19E-02
ENSG00000186472.20	PCLO	219.03	-0.01	0.70	29.55	3.84E-07	7.64E-04
ENSG00000100554.12	ATP6V1D	456.45	0.02	0.26	18.74	8.53E-05	2.55E-02
ENSG00000164294.14	GPX8	281.17	0.03	0.39	24.19	5.58E-06	4.50E-03
ENSG00000117528.14	ABCD3	727.67	0.04	0.31	21.07	2.66E-05	1.40E-02
ENSG00000139973.16	SYT16	34.02	0.05	0.91	24.97	3.78E-06	3.98E-03
ENSG00000116489.13	CAPZA1	2492.82	0.13	0.25	17.73	1.42E-04	3.57E-02
ENSG00000137040.10	RANBP6	208.85	0.15	0.44	16.63	2.45E-04	4.76E-02
ENSG00000229018.5	–	18.01	0.17	0.96	17.90	1.30E-04	3.40E-02
ENSG00000144840.9	RABL3	210.33	0.19	0.51	16.98	2.05E-04	4.61E-02
ENSG00000102901.13	CENPT	1916.58	0.20	0.22	16.92	2.12E-04	4.61E-02
ENSG00000213186.8	TRIM59	259.19	0.23	0.42	17.72	1.42E-04	3.57E-02
ENSG00000124333.16	VAMP7	1223.52	0.24	0.25	18.22	1.11E-04	3.02E-02
ENSG00000115540.15	MOB4	559.47	0.30	0.27	24.60	4.56E-06	4.00E-03
ENSG00000008710.19	PKD1	1529.40	0.31	0.23	17.28	1.77E-04	4.28E-02
ENSG00000151690.15	MFSD6	58.86	0.33	0.64	19.98	4.59E-05	1.86E-02
ENSG00000117569.18	PTBP2	1576.88	0.33	0.32	24.68	4.37E-06	4.00E-03
ENSG00000137575.12	SDCBP	616.76	0.35	0.26	23.94	6.33E-06	4.50E-03
ENSG00000065615.14	CYB5R4	205.46	0.36	0.43	16.44	2.69E-04	4.92E-02
ENSG00000164023.14	SGMS2	201.14	0.37	0.41	21.36	2.30E-05	1.37E-02
ENSG00000153914.16	SREK1	694.96	0.39	0.40	16.91	2.13E-04	4.61E-02
ENSG00000139218.18	SCAF11	604.61	0.41	0.40	16.44	2.69E-04	4.92E-02
ENSG00000213639.10	PPP1CB	1483.96	0.42	0.27	21.07	2.66E-05	1.40E-02
ENSG00000107290.14	SETX	611.62	0.44	0.34	20.52	3.49E-05	1.62E-02
ENSG00000135535.17	CD164	1101.65	0.45	0.26	23.69	7.18E-06	4.77E-03
ENSG00000211456.12	SACM1L	347.14	0.45	0.36	19.35	6.28E-05	2.30E-02
ENSG00000084073.9	ZMPSTE24	1386.77	0.46	0.25	21.20	2.49E-05	1.39E-02

ENSEMBL	Symbol	BaseMean	Log2FoldChange	lfcSE	Stat	pvalue	padj
ENSG00000143952.20	VPS54	374.66	0.49	0.37	18.65	8.90E-05	2.61E-02
ENSG00000111530.13	CAND1	1560.57	0.50	0.29	16.79	2.26E-04	4.70E-02
ENSG00000172795.16	DCP2	508.89	0.50	0.38	19.19	6.82E-05	2.32E-02
ENSG00000123505.17	AMD1	2613.63	0.50	0.24	28.87	5.39E-07	8.78E-04
ENSG00000111860.14	CEP85L	359.31	0.51	0.41	16.70	2.36E-04	4.76E-02
ENSG00000118496.5	FBXO30	279.56	0.52	0.49	16.98	2.06E-04	4.61E-02
ENSG00000047410.14	TPR	1007.40	0.53	0.39	20.22	4.06E-05	1.78E-02
ENSG00000163743.13	RCHY1	503.70	0.56	0.29	20.45	3.63E-05	1.62E-02
ENSG00000134987.11	WDR36	707.44	0.56	0.38	16.85	2.19E-04	4.61E-02
ENSG00000260317.1	–	56.91	0.56	0.67	25.32	3.18E-06	3.56E-03
ENSG00000254535.4	PABPC4L	262.88	0.59	0.38	18.28	1.07E-04	3.02E-02
ENSG00000197045.13	GMFB	465.59	0.59	0.32	20.48	3.56E-05	1.62E-02
ENSG00000153922.10	CHD1	649.21	0.59	0.47	16.66	2.41E-04	4.76E-02
ENSG00000089154.11	GCN1	5266.75	0.65	0.29	26.36	1.89E-06	2.26E-03
ENSG00000198887.9	SMC5	554.70	0.69	0.45	19.78	5.06E-05	1.97E-02
ENSG00000126217.21	MCF2L	128.46	0.70	0.53	18.22	1.10E-04	3.02E-02
ENSG00000188419.14	CHM	506.00	0.70	0.37	20.57	3.41E-05	1.62E-02
ENSG00000180008.9	SOCS4	390.43	0.70	0.50	19.26	6.56E-05	2.30E-02
ENSG00000119787.14	ATL2	636.20	0.74	0.32	19.00	7.47E-05	2.35E-02
ENSG00000162601.11	MYSM1	1501.36	0.81	0.42	19.07	7.21E-05	2.32E-02
ENSG00000162409.11	PRKAA2	1053.00	0.81	0.39	19.09	7.14E-05	2.32E-02
ENSG00000177565.16	TBL1XR1	623.67	0.84	0.39	16.88	2.16E-04	4.61E-02
ENSG00000137145.20	DENND4C	393.43	0.89	0.47	17.93	1.28E-04	3.40E-02
ENSG00000198826.11	ARHGAP11A	301.12	0.91	0.47	29.32	4.30E-07	7.70E-04
ENSG00000134318.14	ROCK2	241.38	0.93	0.49	18.21	1.11E-04	3.02E-02
ENSG00000167971.16	CASKIN1	636.11	0.99	0.44	27.43	1.11E-06	1.42E-03
ENSG00000144893.12	MED12L	301.03	0.99	0.51	16.76	2.29E-04	4.72E-02
ENSG00000078177.14	N4BP2	156.39	1.03	0.54	16.96	2.08E-04	4.61E-02
ENSG00000166575.17	TMEM135	270.80	1.05	0.41	22.32	1.42E-05	8.77E-03
ENSG00000023516.9	AKAP11	511.75	1.06	0.45	20.88	2.92E-05	1.49E-02
ENSG00000139618.15	BRCA2	389.39	1.34	0.60	28.14	7.75E-07	1.16E-03
ENSG00000135968.20	GCC2	189.05	1.45	0.44	22.39	1.37E-05	8.77E-03
ENSG00000137573.14	SULF1	41.31	1.50	0.94	19.49	5.87E-05	2.19E-02
ENSG00000261098.1	–	72.00	1.51	0.57	29.81	3.36E-07	7.52E-04
ENSG00000285053.1	TBCE	155.22	2.21	0.83	24.59	4.58E-06	4.00E-03
ENSG00000196628.18	TCF4	240.24	2.52	0.55	24.03	6.04E-06	4.50E-03
ENSG00000188611.15	ASAH2	125.17	2.59	1.01	17.88	1.31E-04	3.40E-02
ENSG00000103426.12	CORO7-PAM16	65.29	2.63	0.75	79.71	4.90E-18	8.78E-14
ENSG00000160219.12	GAB3	64.96	2.86	0.67	20.10	4.32E-05	1.84E-02
ENSG00000151006.7	PRSS53	68.44	2.87	0.65	19.28	6.50E-05	2.30E-02
ENSG00000198064.13	NPIPB13	21.84	5.27	1.82	23.88	6.53E-06	4.50E-03
ENSG00000280987.4	MATR3	124.60	5.66	0.92	72.68	1.65E-16	1.48E-12
ENSG00000261247.1	GOLGA8T	20.36	6.24	1.85	16.61	2.47E-04	4.76E-02

Table 34: Log2-fold changes of genes after 8h. Data shown in Figure 33. Data adapted from Wolffgramm et al.[103]

ENSEMBL	Symbol	BaseMean	Log2FoldChange	lfcSE	Stat	pvalue	padj
ENSG00000257225.1	–	14.13	-13.87	2.61	38.02	5.55E-09	1.99E-05
ENSG00000229018.5	–	18.01	-6.12	1.73	17.90	1.30E-04	3.40E-02
ENSG00000259607.1	–	8.87	-6.05	2.13	16.97	2.06E-04	4.61E-02
ENSG00000198064.13	NPIPB13	21.84	-5.93	2.08	23.88	6.53E-06	4.50E-03
ENSG00000187664.9	HAPLN4	10.79	-5.37	2.35	24.03	6.04E-06	4.50E-03
ENSG00000142178.9	SIK1	78.82	-5.37	1.19	19.10	7.11E-05	2.32E-02
ENSG00000139973.16	SYT16	34.02	-4.94	1.16	24.97	3.78E-06	3.98E-03
ENSG00000196218.12	RYR1	44.14	-4.80	1.13	17.00	2.03E-04	4.61E-02
ENSG00000186472.20	PCLO	219.03	-3.46	0.71	29.55	3.84E-07	7.64E-04
ENSG00000260317.1	–	56.91	-3.09	0.78	25.32	3.18E-06	3.56E-03
ENSG00000137573.14	SULF1	41.31	-2.86	0.98	19.49	5.87E-05	2.19E-02
ENSG00000256591.5	–	73.85	-2.85	0.76	17.69	1.44E-04	3.58E-02
ENSG00000151690.15	MFSD6	58.86	-2.60	0.70	19.98	4.59E-05	1.86E-02
ENSG00000139618.15	BRCA2	389.39	-1.93	0.61	28.14	7.75E-07	1.16E-03
ENSG00000188611.15	ASAH2	125.17	-1.86	1.02	17.88	1.31E-04	3.40E-02
ENSG00000261098.1	–	72.00	-1.83	0.64	29.81	3.36E-07	7.52E-04
ENSG00000144840.9	RABL3	210.33	-1.81	0.53	16.98	2.05E-04	4.61E-02
ENSG00000164294.14	GPX8	281.17	-1.69	0.39	24.19	5.58E-06	4.50E-03
ENSG00000118197.14	DDX59	146.59	-1.68	0.42	16.48	2.63E-04	4.92E-02
ENSG00000198826.11	ARHGAP11A	301.12	-1.67	0.48	29.32	4.30E-07	7.70E-04
ENSG00000165533.18	TTC8	412.62	-1.63	0.42	16.48	2.64E-04	4.92E-02
ENSG00000119446.14	RBM18	796.72	-1.61	0.28	34.68	2.95E-08	7.56E-05
ENSG00000155097.12	ATP6V1C1	475.48	-1.61	0.41	17.48	1.60E-04	3.92E-02
ENSG00000126217.21	MCF2L	128.46	-1.56	0.53	18.22	1.10E-04	3.02E-02
ENSG00000137040.10	RANBP6	208.85	-1.53	0.45	16.63	2.45E-04	4.76E-02
ENSG00000283154.2	SCHIP1	201.62	-1.53	0.34	19.94	4.67E-05	1.86E-02
ENSG00000180008.9	SOCS4	390.43	-1.53	0.51	19.26	6.56E-05	2.30E-02
ENSG00000164023.14	SGMS2	201.14	-1.52	0.43	21.36	2.30E-05	1.37E-02
ENSG00000118496.5	FBXO30	279.56	-1.50	0.50	16.98	2.06E-04	4.61E-02
ENSG00000213186.8	TRIM59	259.19	-1.49	0.44	17.72	1.42E-04	3.57E-02
ENSG00000065615.14	CYB5R4	205.46	-1.36	0.44	16.44	2.69E-04	4.92E-02
ENSG00000104549.12	SQLE	741.09	-1.35	0.31	20.82	3.01E-05	1.50E-02
ENSG00000152942.19	RAD17	448.81	-1.34	0.34	19.49	5.87E-05	2.19E-02
ENSG00000143499.14	SMYD2	292.08	-1.33	0.35	17.05	1.98E-04	4.61E-02
ENSG00000153922.10	CHD1	649.21	-1.32	0.48	16.66	2.41E-04	4.76E-02
ENSG00000198887.9	SMC5	554.70	-1.31	0.45	19.78	5.06E-05	1.97E-02
ENSG00000078177.14	N4BP2	156.39	-1.31	0.57	16.96	2.08E-04	4.61E-02
ENSG00000117528.14	ABCD3	727.67	-1.23	0.31	21.07	2.66E-05	1.40E-02
ENSG00000153914.16	SREK1	694.96	-1.22	0.41	16.91	2.13E-04	4.61E-02
ENSG00000134318.14	ROCK2	241.38	-1.21	0.50	18.21	1.11E-04	3.02E-02
ENSG00000047410.14	TPR	1007.40	-1.19	0.39	20.22	4.06E-05	1.78E-02
ENSG00000117569.18	PTBP2	1576.88	-1.19	0.32	24.68	4.37E-06	4.00E-03
ENSG00000111860.14	CEP85L	359.31	-1.18	0.42	16.70	2.36E-04	4.76E-02
ENSG00000139218.18	SCAF11	604.61	-1.17	0.40	16.44	2.69E-04	4.92E-02
ENSG00000172795.16	DCP2	508.89	-1.16	0.38	19.19	6.82E-05	2.32E-02
ENSG00000137145.20	DENND4C	393.43	-1.15	0.48	17.93	1.28E-04	3.40E-02
ENSG00000211456.12	SACM1L	347.14	-1.15	0.37	19.35	6.28E-05	2.30E-02
ENSG00000109911.19	ELP4	554.11	-1.14	0.28	16.64	2.43E-04	4.76E-02
ENSG00000144893.12	MED12L	301.03	-1.14	0.52	16.76	2.29E-04	4.72E-02
ENSG00000143952.20	VPS54	374.66	-1.10	0.38	18.65	8.90E-05	2.61E-02
ENSG00000254535.4	PABPC4L	262.88	-1.08	0.39	18.28	1.07E-04	3.02E-02
ENSG00000107290.14	SETX	611.62	-1.07	0.34	20.52	3.49E-05	1.62E-02
ENSG00000162601.11	MYSM1	1501.36	-1.05	0.42	19.07	7.21E-05	2.32E-02
ENSG00000023516.9	AKAP11	511.75	-1.04	0.46	20.88	2.92E-05	1.49E-02
ENSG00000076043.10	REXO2	428.79	-1.03	0.25	16.57	2.52E-04	4.80E-02

ENSEMBL	Symbol	BaseMean	Log2FoldChange	lfcSE	Stat	pvalue	padj
ENSG00000134987.11	WDR36	707.44	-1.01	0.39	16.85	2.19E-04	4.61E-02
ENSG00000115540.15	MOB4	559.47	-1.01	0.27	24.60	4.56E-06	4.00E-03
ENSG00000100554.12	ATP6V1D	456.45	-0.99	0.27	18.74	8.53E-05	2.55E-02
ENSG00000188419.14	CHM	506.00	-0.99	0.37	20.57	3.41E-05	1.62E-02
ENSG00000113558.18	SKP1	777.43	-0.97	0.24	21.21	2.48E-05	1.39E-02
ENSG00000136156.15	ITM2B	1743.03	-0.96	0.22	20.03	4.47E-05	1.86E-02
ENSG00000162409.11	PRKAA2	1053.00	-0.93	0.40	19.09	7.14E-05	2.32E-02
ENSG00000043514.17	TRIT1	1188.44	-0.92	0.24	16.86	2.19E-04	4.61E-02
ENSG00000137575.12	SDCBP	616.76	-0.91	0.26	23.94	6.33E-06	4.50E-03
ENSG00000166575.17	TMEM135	270.80	-0.90	0.41	22.32	1.42E-05	8.77E-03
ENSG00000197045.13	GMFB	465.59	-0.89	0.33	20.48	3.56E-05	1.62E-02
ENSG00000116489.13	CAPZA1	2492.82	-0.85	0.25	17.73	1.42E-04	3.57E-02
ENSG00000133835.16	HSD17B4	970.84	-0.85	0.21	18.95	7.68E-05	2.37E-02
ENSG00000135535.17	CD164	1101.65	-0.81	0.26	23.69	7.18E-06	4.77E-03
ENSG00000123505.17	AMD1	2613.63	-0.81	0.24	28.87	5.39E-07	8.78E-04
ENSG00000213639.10	PPP1CB	1483.96	-0.80	0.27	21.07	2.66E-05	1.40E-02
ENSG00000124333.16	VAMP7	1223.52	-0.80	0.25	18.22	1.11E-04	3.02E-02
ENSG00000177565.16	TBL1XR1	623.67	-0.79	0.40	16.88	2.16E-04	4.61E-02
ENSG00000163743.13	RCHY1	503.70	-0.79	0.30	20.45	3.63E-05	1.62E-02
ENSG00000111530.13	CAND1	1560.57	-0.71	0.30	16.79	2.26E-04	4.70E-02
ENSG00000084073.9	ZMPSTE24	1386.77	-0.68	0.25	21.20	2.49E-05	1.39E-02
ENSG00000261247.1	GOLGA8T	20.36	-0.67	1.18	16.61	2.47E-04	4.76E-02
ENSG00000119787.14	ATL2	636.20	-0.66	0.32	19.00	7.47E-05	2.35E-02
ENSG00000135968.20	GCC2	189.05	-0.61	0.46	22.39	1.37E-05	8.77E-03
ENSG00000275496.4	LOC102724701	48.52	0.09	0.82	42.37	6.31E-10	3.77E-06
ENSG00000130635.15	COL5A1	290.37	0.63	0.35	27.92	8.64E-07	1.19E-03
ENSG00000130940.15	CASZ1	581.09	0.74	0.23	36.03	1.50E-08	4.48E-05
ENSG00000155254.13	MARVELD1	500.88	0.80	0.36	19.07	7.24E-05	2.32E-02
ENSG00000135709.12	KIAA0513	269.09	0.82	0.41	18.27	1.08E-04	3.02E-02
ENSG00000102901.13	CENPT	1916.58	0.85	0.22	16.92	2.12E-04	4.61E-02
ENSG00000008710.19	PKD1	1529.40	0.95	0.23	17.28	1.77E-04	4.28E-02
ENSG00000151006.7	PRSS53	68.44	1.00	0.65	19.28	6.50E-05	2.30E-02
ENSG00000174705.13	SH3PXD2B	1138.35	1.08	0.27	40.15	1.91E-09	8.55E-06
ENSG00000277399.4	GPR179	20.53	1.16	1.65	16.69	2.37E-04	4.76E-02
ENSG00000187122.17	SLIT1	210.61	1.29	0.48	24.54	4.69E-06	4.00E-03
ENSG00000089154.11	GCN1	5266.75	1.47	0.29	26.36	1.89E-06	2.26E-03
ENSG00000196628.18	TCF4	240.24	2.21	0.56	24.03	6.04E-06	4.50E-03
ENSG00000167971.16	CASKIN1	636.11	2.33	0.44	27.43	1.11E-06	1.42E-03
ENSG00000160219.12	GAB3	64.96	2.35	0.67	20.10	4.32E-05	1.84E-02
ENSG00000285053.1	TBCE	155.22	4.23	0.82	24.59	4.58E-06	4.00E-03
ENSG00000198237.8	–	21.42	5.33	2.39	18.78	8.38E-05	2.54E-02
ENSG00000280987.4	MATR3	124.60	8.62	0.95	72.68	1.65E-16	1.48E-12
ENSG00000103426.12	CORO7-PAM16	65.29	10.84	1.62	79.71	4.90E-18	8.78E-14

Table 35: Overall differential gene expression after light irradiation showing light vs non-light log2-fold changes shrunken using ashr[395]. Figure 33 displays values of Table 33 and Table 34 to show time-resolution. Data adapted from Wolffgramm et al.[103]

ENSEMBL	Symbol	BaseMean	Log2FoldChange	lfcSE	pvalue	padj
ENSG00000280987.4	MATR3	124.60	-5.03	0.68	1.65E-16	1.48E-12
ENSG00000167971.16	CASKIN1	636.11	-0.15	0.44	1.11E-06	1.42E-03
ENSG00000135968.20	GCC2	189.05	-0.12	0.40	1.37E-05	8.77E-03
ENSG00000103426.12	CORO7-PAM16	65.29	-0.03	0.27	4.90E-18	8.78E-14
ENSG00000285053.1	TBCE	155.22	-0.01	0.17	4.58E-06	4.00E-03
ENSG00000188419.14	CHM	506.00	-0.01	0.09	3.41E-05	1.62E-02
ENSG00000023516.9	AKAP11	511.75	-0.01	0.09	2.92E-05	1.49E-02

ENSEMBL	Symbol	BaseMean	Log2FoldChange	lfcSE	pvalue	padj
ENSG00000261098.1	–	72.00	-0.00	0.08	3.36E-07	7.52E-04
ENSG00000078177.14	N4BP2	156.39	-0.00	0.05	2.08E-04	4.61E-02
ENSG00000123505.17	AMD1	2613.63	-0.00	0.02	5.39E-07	8.78E-04
ENSG00000162409.11	PRKAA2	1053.00	-0.00	0.03	7.14E-05	2.32E-02
ENSG00000196628.18	TCF4	240.24	-0.00	0.04	6.04E-06	4.50E-03
ENSG00000111530.13	CAND1	1560.57	-0.00	0.01	2.26E-04	4.70E-02
ENSG00000084073.9	ZMPSTE24	1386.77	-0.00	0.01	2.49E-05	1.39E-02
ENSG00000089154.11	GCN1	5266.75	-0.00	0.01	1.89E-06	2.26E-03
ENSG00000137575.12	SDCBP	616.76	-0.00	0.01	6.33E-06	4.50E-03
ENSG00000162601.11	MYSM1	1501.36	-0.00	0.02	7.21E-05	2.32E-02
ENSG00000177565.16	TBL1XR1	623.67	-0.00	0.02	2.16E-04	4.61E-02
ENSG00000213639.10	PPP1CB	1483.96	-0.00	0.01	2.66E-05	1.40E-02
ENSG00000143952.20	VPS54	374.66	-0.00	0.02	8.90E-05	2.61E-02
ENSG00000135535.17	CD164	1101.65	-0.00	0.01	7.18E-06	4.77E-03
ENSG00000116489.13	CAPZA1	2492.82	-0.00	0.01	1.42E-04	3.57E-02
ENSG00000134318.14	ROCK2	241.38	-0.00	0.02	1.11E-04	3.02E-02
ENSG00000115540.15	MOB4	559.47	-0.00	0.01	4.56E-06	4.00E-03
ENSG00000107290.14	SETX	611.62	-0.00	0.01	3.49E-05	1.62E-02
ENSG00000111860.14	CEP85L	359.31	-0.00	0.02	2.36E-04	4.76E-02
ENSG00000163743.13	RCHY1	503.70	-0.00	0.01	3.63E-05	1.62E-02
ENSG00000117569.18	PTBP2	1576.88	-0.00	0.01	4.37E-06	4.00E-03
ENSG00000124333.16	VAMP7	1223.52	-0.00	0.01	1.11E-04	3.02E-02
ENSG00000197045.13	GMFB	465.59	-0.00	0.01	3.56E-05	1.62E-02
ENSG00000198826.11	ARHGAP11A	301.12	-0.00	0.02	4.30E-07	7.70E-04
ENSG00000211456.12	SACM1L	347.14	-0.00	0.01	6.28E-05	2.30E-02
ENSG00000139618.15	BRCA2	389.39	-0.00	0.02	7.75E-07	1.16E-03
ENSG00000254535.4	PABPC4L	262.88	-0.00	0.01	1.07E-04	3.02E-02
ENSG00000137145.20	DENND4C	393.43	-0.00	0.02	1.28E-04	3.40E-02
ENSG00000134987.11	WDR36	707.44	-0.00	0.01	2.19E-04	4.61E-02
ENSG00000047410.14	TPR	1007.40	-0.00	0.01	4.06E-05	1.78E-02
ENSG00000198887.9	SMC5	554.70	-0.00	0.01	5.06E-05	1.97E-02
ENSG00000172795.16	DCP2	508.89	-0.00	0.01	6.82E-05	2.32E-02
ENSG00000139218.18	SCAF11	604.61	-0.00	0.01	2.69E-04	4.92E-02
ENSG00000153922.10	CHD1	649.21	-0.00	0.01	2.41E-04	4.76E-02
ENSG00000119787.14	ATL2	636.20	-0.00	0.01	7.47E-05	2.35E-02
ENSG00000153914.16	SREK1	694.96	-0.00	0.01	2.13E-04	4.61E-02
ENSG00000180008.9	SOCS4	390.43	-0.00	0.01	6.56E-05	2.30E-02
ENSG00000118496.5	FBXO30	279.56	-0.00	0.01	2.06E-04	4.61E-02
ENSG00000151006.7	PRSS53	68.44	-0.00	0.02	6.50E-05	2.30E-02
ENSG00000213186.8	TRIM59	259.19	-0.00	0.01	1.42E-04	3.57E-02
ENSG00000166575.17	TMEM135	270.80	-0.00	0.01	1.42E-05	8.77E-03
ENSG00000152942.19	RAD17	448.81	-0.00	0.01	5.87E-05	2.19E-02
ENSG00000117528.14	ABCD3	727.67	-0.00	0.01	2.66E-05	1.40E-02
ENSG00000137040.10	RANBP6	208.85	-0.00	0.01	2.45E-04	4.76E-02
ENSG00000144893.12	MED12L	301.03	-0.00	0.01	2.29E-04	4.72E-02
ENSG00000164023.14	SGMS2	201.14	-0.00	0.01	2.30E-05	1.37E-02
ENSG00000065615.14	CYB5R4	205.46	-0.00	0.01	2.69E-04	4.92E-02
ENSG00000160219.12	GAB3	64.96	-0.00	0.01	4.32E-05	1.84E-02
ENSG00000008710.19	PKD1	1529.40	-0.00	0.01	1.77E-04	4.28E-02
ENSG00000100554.12	ATP6V1D	456.45	-0.00	0.01	8.53E-05	2.55E-02
ENSG00000137573.14	SULF1	41.31	-0.00	0.01	5.87E-05	2.19E-02
ENSG00000186472.20	PCLO	219.03	-0.00	0.01	3.84E-07	7.64E-04
ENSG00000260317.1	–	56.91	-0.00	0.01	3.18E-06	3.56E-03
ENSG00000155097.12	ATP6V1C1	475.48	-0.00	0.01	1.60E-04	3.92E-02
ENSG00000188611.15	ASAH2	125.17	-0.00	0.01	1.31E-04	3.40E-02
ENSG00000151690.15	MFSD6	58.86	-0.00	0.01	4.59E-05	1.86E-02
ENSG00000229018.5	–	18.01	-0.00	0.01	1.30E-04	3.40E-02
ENSG00000043514.17	TRIT1	1188.44	-0.00	0.01	2.19E-04	4.61E-02

ENSEMBL	Symbol	BaseMean	Log2FoldChange	lfcSE	pvalue	padj
ENSG00000164294.14	GPX8	281.17	-0.00	0.01	5.58E-06	4.50E-03
ENSG00000261247.1	GOLGA8T	20.36	0.00	0.01	2.47E-04	4.76E-02
ENSG00000102901.13	CENPT	1916.58	0.00	0.01	2.12E-04	4.61E-02
ENSG00000139973.16	SYT16	34.02	0.00	0.01	3.78E-06	3.98E-03
ENSG00000198064.13	NPIPB13	21.84	0.00	0.01	6.53E-06	4.50E-03
ENSG00000275496.4	LOC102724701	48.52	0.00	0.01	6.31E-10	3.77E-06
ENSG00000144840.9	RABL3	210.33	0.00	0.01	2.05E-04	4.61E-02
ENSG00000198237.8	–	21.42	0.00	0.01	8.38E-05	2.54E-02
ENSG00000277399.4	GPR179	20.53	0.00	0.01	2.37E-04	4.76E-02
ENSG00000113558.18	SKP1	777.43	0.00	0.01	2.48E-05	1.39E-02
ENSG00000187122.17	SLIT1	210.61	0.00	0.01	4.69E-06	4.00E-03
ENSG00000133835.16	HSD17B4	970.84	0.00	0.01	7.68E-05	2.37E-02
ENSG00000118197.14	DDX59	146.59	0.00	0.01	2.63E-04	4.92E-02
ENSG00000165533.18	TTC8	412.62	0.00	0.01	2.64E-04	4.92E-02
ENSG00000126217.21	MCF2L	128.46	0.00	0.01	1.10E-04	3.02E-02
ENSG00000130635.15	COL5A1	290.37	0.00	0.01	8.64E-07	1.19E-03
ENSG00000155254.13	MARVELD1	500.88	0.00	0.01	7.24E-05	2.32E-02
ENSG00000104549.12	SQLE	741.09	0.00	0.01	3.01E-05	1.50E-02
ENSG00000135709.12	KIAA0513	269.09	0.00	0.01	1.08E-04	3.02E-02
ENSG00000136156.15	ITM2B	1743.03	0.00	0.01	4.47E-05	1.86E-02
ENSG00000119446.14	RBM18	796.72	0.00	0.01	2.95E-08	7.56E-05
ENSG00000196218.12	RYR1	44.14	0.00	0.02	2.03E-04	4.61E-02
ENSG00000174705.13	SH3PXD2B	1138.35	0.00	0.01	1.91E-09	8.55E-06
ENSG00000076043.10	REXO2	428.79	0.00	0.01	2.52E-04	4.80E-02
ENSG00000143499.14	SMYD2	292.08	0.00	0.01	1.98E-04	4.61E-02
ENSG00000283154.2	SCHIP1	201.62	0.00	0.01	4.67E-05	1.86E-02
ENSG00000109911.19	ELP4	554.11	0.00	0.01	2.43E-04	4.76E-02
ENSG00000130940.15	CASZ1	581.09	0.00	0.01	1.50E-08	4.48E-05
ENSG00000142178.9	SIK1	78.82	0.01	0.13	7.11E-05	2.32E-02
ENSG00000256591.5	–	73.85	0.01	0.11	1.44E-04	3.58E-02
ENSG00000259607.1	–	8.87	0.02	0.34	2.06E-04	4.61E-02
ENSG00000187664.9	HAPLN4	10.79	0.04	0.49	6.04E-06	4.50E-03
ENSG00000257225.1	–	14.13	0.04	0.49	5.55E-09	1.99E-05

8.9. NMR Data

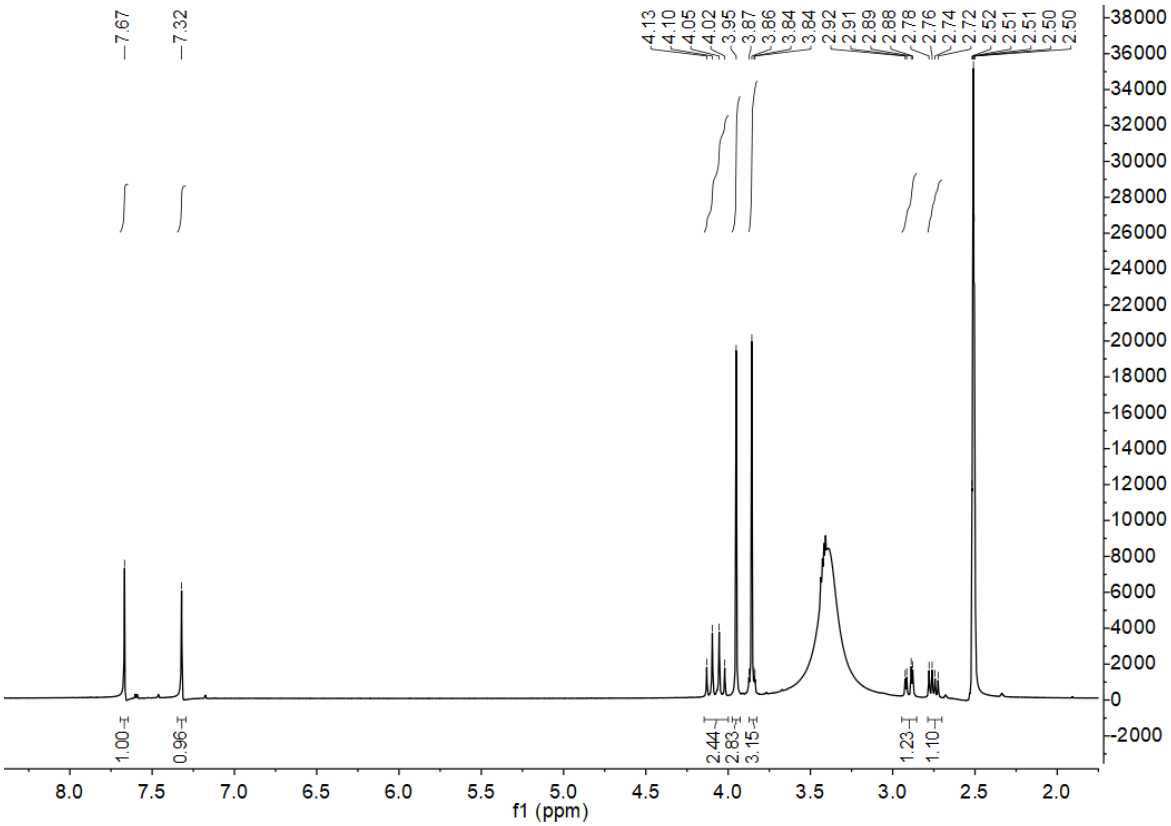


Figure 69: 1H spectrum of DMNB-Cys.
^{1}H NMR (400 MHz, DMSO) δ = 7.67 (s, 1H), 7.32 (s, 1H), 4.08 (dd, *J*=30.1, 13.5, 2H), 3.95 (s, 3H), 3.87 – 3.82 (m, 3H), 2.90 (dd, *J*=14.6, 3.9, 1H), 2.75 (dd, *J*=14.6, 7.5, 1H).
NMR was done by Stefan Helmer.

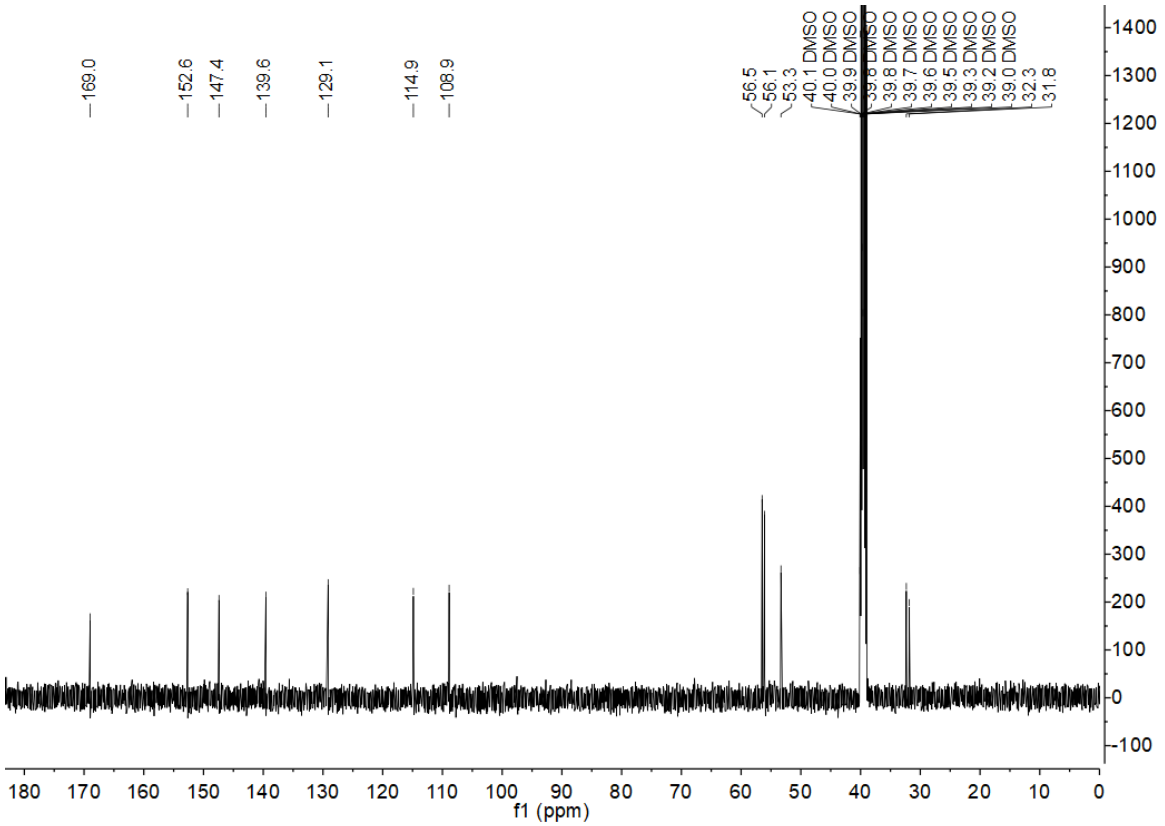


Figure 70: 13C spectrum of DMNB-Cys.
^{1}H NMR (400 MHz, DMSO) δ = 7.67 (s, 1H), 7.32 (s, 1H), 4.08 (dd, *J*=30.1, 13.5, 2H), 3.95 (s, 3H), 3.87 – 3.82 (m, 3H), 2.90 (dd, *J*=14.6, 3.9, 1H), 2.75 (dd, *J*=14.6, 7.5, 1H).
NMR was done by Stefan Helmer.

9. References

[1] W. Müller-Esterl, U. Brandt, O. Anderka, S. Kerscher, *Biochemie*. *Eine Einführung für Mediziner und Naturwissenschaftler,* 2. Aufl. ed., Spektrum, Heidelberg, **2011**.

[2] J. M. Berg, J. L. Tymoczko, G. J. Gatto jr., L. Stryer, A. Held, G. Maxam, L. Seidler, B. Häcker, B. Jarosch, *Biochemistry,* Eighth edition ed., Springer Spektrum, Berlin, Heidelberg, **2017**.

[3] S. Müller, *Nucleic acids from A to Z*. *A concise encyclopedia / edited by Sabine Müller*, Wiley-VCH; Chichester : John Wiley [distributor], Weinheim, **2008**.

[4] J. D. WATSON, F. H. CRICK, *Nature* **1953**, *171*, 737–738.

[5] R. E. Franklin, R. G. Gosling, *Acta Cryst* **1953**, *6*, 673–677.

[6] Y. Mitsui, R. Langridge, B. E. Shortle, C. R. Cantor, R. C. Grant, M. Kodama, R. D. Wells, *Nature* **1970**, *228*, 1166–1169.

[7] a) H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, R. E. Dickerson, *Proceedings of the National Academy of Sciences* **1981**, *78*, 2179–2183; b) H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, R. E. Dickerson, *STRUCTURE OF A B-DNA DODECAMER. CONFORMATION AND DYNAMICS*, **1981**;

[8] P. Yakovchuk, E. Protozanova, M. D. Frank-Kamenetskii, *Nucleic acids research* **2006**, *34*, 564–574.

[9] a) A. Rich, A. Nordheim, A. H. Wang, *Annu. Rev. Biochem.* **1984**, *53*, 791–846; b) R. E. Dickerson, H. R. Drew, B. N. Conner, R. M. Wing, A. V. Fratini, M. L. Kopka, *Science* **1982**, *216*, 475–485;

[10] R. Wing, H. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura, R. E. Dickerson, *Nature* **1980**, *287*, 755–758.

[11] C. O. Pabo, R. T. Sauer, *Annu. Rev. Biochem.* **1984**, *53*, 293–321.

[12] a) R. Y. Samson, S. D. Bell, *Journal of molecular microbiology and biotechnology* **2014**, *24*, 420–427; b) M. S. Ferdows, A. G. Barbour, *Proceedings of the National Academy of Sciences* **1989**, *86*, 5969–5973;

[13] G. Chaconas, C. W. Chen in *The Bacterial Chromosome* (Ed.: N. P. Higgins), ASM Press, Washington, DC, USA, **2004**.

[14] a) K. van Bortle, V. G. Corces, *Annual review of cell and developmental biology* **2012**, *28*, 163–187; b) D. U. Gorkin, D. Leung, B. Ren, *Cell stem cell* **2014**, *14*, 762–775; c) J. Dekker, T. Misteli, *Cold Spring Harbor perspectives in biology* **2015**, *7*, a019356; d) B. Bonev, G. Cavalli, *Nature reviews. Genetics* **2016**, *17*, 661–678; e) R. A. Beagrie, A. Scialdone, M. Schueler, D. C. A. Kraemer, M. Chotalia, S. Q. Xie, M. Barbieri, I. de Santiago, L.-M. Lavitas, M. R. Branco et al., *Nature* **2017**, *543*, 519–524; f) C. B. Hug, J. M. Vaquerizas, *Trends in genetics : TIG* **2018**, *34*, 903–914; g) H. Zheng, W. Xie, *Nat Rev Mol Cell Biol* **2019**, *20*, 535–550;

[15] R. Balhorn, *Genome biology* **2007**, *8*, 227.

[16] C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder, T. J. Richmond, *Journal of molecular biology* **2002**, *319*, 1097–1113.

[17] L. Armstrong, *Epigenetics*, Garland Science, New York NY, **2014**.

[18] R. D. Kornberg, *Science* **1974**, *184*, 868–871.

[19] C. D. Allis, T. Jenuwein, M.-L. Caparros, D. Reinberg, M. Lachner, *Epigenetics,* Second edition / associate editor, Monika Lachner, Max Planck Institute of Immunobiology and Epigenetics, Freiburg ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, **2015**.

[20] a) D. Canzio, E. Y. Chang, S. Shankar, K. M. Kuchenbecker, M. D. Simon, H. D. Madhani, G. J. Narlikar, B. Al-Sady, *Molecular cell* **2011**, *41*, 67–81; b) N. J. Francis, R. E. Kingston, C. L. Woodcock, *Science (New York, N.Y.)* **2004**, *306*, 1574–1577;

[21] a) F. Thoma, T. Koller, A. Klug, *The Journal of cell biology* **1979**, *83*, 403–427; b) P. J. J. Robinson, L. Fairall, A. T. van Huynh, D. Rhodes, *Proceedings of the National Academy of Sciences* **2006**, *103*, 6506–6511;

[22] T. R. Gregory, *Nature reviews. Genetics* **2005**, *6*, 699–708.

[23] E. V. Koonin, Y. I. Wolf, *Nature reviews. Genetics* **2010**, *11*, 487–498.

[24] a) E. Birney, J. A. Stamatoyannopoulos, A. Dutta, R. Guigó, T. R. Gingeras, E. H. Margulies, Z. Weng, M. Snyder, E. T. Dermitzakis, R. E. Thurman et al., *Nature* **2007**, *447*, 799–816; b) *Nature cell biology* **2019**, *21*, 535; c) S. Washietl, J. S. Pedersen, J. O. Korbel, C. Stocsits, A. R. Gruber, J. Hackermüller, J. Hertel, M. Lindemeyer, K. Reiche, A. Tanzer et al., *Genome research* **2007**, *17*, 852–864;

[25] a) M. G. Kidwell, D. R. Lisch, *Trends in Ecology & Evolution* **2000**, *15*, 95–99; b) M. G. Kidwell, D. R. Lisch, *Evolution; international journal of organic evolution* **2001**, *55*, 1–24;

[26] E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh et al., *Nature* **2001**, *409*, 860–921.

[27] C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder, T. J. Richmond, *X-Ray Structure of the Nucleosome Core Particle, NCP147, at 1.9 A Resolution*, **2002**.

[28] an Jansen, K. J. Verstrepen, *Microbiology and molecular biology reviews : MMBR* **2011**, *75*, 301–320.

[29] J. Padeken, P. Zeller, S. M. Gasser, *Current opinion in genetics & development* **2015**, *31*, 12–19.

[30] a) G.-F. Richard, A. Kerrest, B. Dujon, *Microbiology and molecular biology reviews : MMBR* **2008**, *72*, 686–727; b) H. S. Malik, S. Henikoff, *Cell* **2009**, *138*, 1067–1082;

[31] K. H. Miga, *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* **2015**, *23*, 421–426.

[32] L. L. Hall, M. Byron, D. M. Carone, T. W. Whitfield, G. P. Pouliot, A. Fischer, P. Jones, J. B. Lawrence, *Cell reports* **2017**, *18*, 2943–2956.

[33] P. Bernard, J. F. Maure, J. F. Partridge, S. Genier, J. P. Javerzat, R. C. Allshire, *Science* **2001**, *294*, 2539–2542.

[34] a) C. Tyler-Smith, W. R. Brown, *Journal of molecular biology* **1987**, *195*, 457–470; b) K. H. Miga, *Genes* **2019**, *10*;

[35] R. Valgardsdottir, I. Chiodi, M. Giordano, A. Rossi, S. Bazzini, C. Ghigna, S. Riva, G. Biamonti, *Nucleic acids research* **2008**, *36*, 423–434.

[36] E. Col, N. Hoghoughi, S. Dufour, J. Penin, S. Koskas, V. Faure, M. Ouzounova, H. Hernandez-Vargash, N. Reynoird, S. Daujat et al., *Scientific reports* **2017**, *7*, 5418.

[37] C. Jolly, A. Metz, J. Govin, M. Vigneron, B. M. Turner, S. Khochbin, C. Vourc'h, *The Journal of cell biology* **2004**, *164*, 25–33.

[38] G. Biamonti, C. Vourc'h, *Cold Spring Harbor perspectives in biology* **2010**, *2*, a000695.

[39] a) P. Mähl, Y. Lutz, E. Puvion, J. P. Fuchs, *The Journal of cell biology* **1989**, *109*, 1921–1935; b) K. D. Sarge, S. P. Murphy, R. I. Morimoto, *Molecular and cellular biology* **1993**, *13*, 1392–1407; c) M. Denegri, D. Moralli, M. Rocchi, M. Biggiogera, E. Raimondi, F. Cobianchi, L. de Carli, S. Riva, G. Biamonti, *Molecular Biology of the Cell* **2002**, *13*, 2069–2079;

[40] N. Rizzi, M. Denegri, I. Chiodi, M. Corioni, R. Valgardsdottir, F. Cobianchi, S. Riva, G. Biamonti, *Molecular Biology of the Cell* **2004**, *15*, 543–551.

[41] M. Akerfelt, R. I. Morimoto, L. Sistonen, *Nat Rev Mol Cell Biol* **2010**, *11*, 545–555.

[42] Y. Saito, Y. Kanai, M. Sakamoto, H. Saito, H. Ishii, S. Hirohashi, *Hepatology (Baltimore, Md.)* **2001**, *33*, 561–568.

[43] N. I. Enukashvily, R. Donev, I. S.-R. Waisertreiger, O. I. Podgornaya, *Cytogenetic and genome research* **2007**, *118*, 42–54.

[44] S. BRENNER, F. JACOB, M. MESELSON, *Nature* **1961**, *190*, 576–581.

[45] F. GROS, H. HIATT, W. GILBERT, C. G. KURLAND, R. W. RISEBROUGH, J. D. WATSON, *Nature* **1961**, *190*, 581–585.

[46] A. A. Bicknell, E. P. Ricci, *Biochemical Society transactions* **2017**, *45*, 339–351.

[47] P. Cramer, *Current Opinion in Structural Biology* **2002**, *12*, 89–97.

[48] R. G. Roeder, *Trends in Biochemical Sciences* **1991**, *16*, 402–408.

[49] a) K. Hammer, I. Mijakovic, P. R. Jensen, *Trends in biotechnology* **2006**, *24*, 53–55; b) Q. Liu, M. Liu, W. Wu, *Journal of computational biology : a journal of computational molecular cell biology* **2018**, *25*, 1152–1160;

[50] O. Parra-Marín, K. López-Pacheco, R. Hernández, I. López-Villaseñor, *Molecular and biochemical parasitology* **2020**, *239*, 111312.

[51] S. A. LES LABORATOIRES SERVIER, to be found under smart.servier.com.

[52] J. Griesenbeck, H. Tschochner, D. Grohmann, *Sub-cellular biochemistry* **2017**, *83*, 225–270.

[53] a) E. Skordalakes, J. M. Berger, *Cell* **2003**, *114*, 135–146; b) M. H. Larson, W. J. Greenleaf, R. Landick, S. M. Block, *Cell* **2008**, *132*, 971–982;

[54] D. Tollervey, *Nature* **2004**, *432*, 456–457.

[55] J. N. Kuehner, E. L. Pearson, C. Moore, *Nat Rev Mol Cell Biol* **2011**, *12*, 283–294.

[56] a) A. Nag, K. Narsinh, A. Kazerouninia, H. G. Martinson, *RNA (New York, N.Y.)* **2006**, *12*, 1534–1544; b) I. J. Orozco, S. J. Kim, H. G. Martinson, *The Journal of biological chemistry* **2002**, *277*, 42899–42911;

[57] a) H. E. Mischo, N. J. Proudfoot, *Biochimica et biophysica acta* **2013**, *1829*, 174–185; b) O. Porrua, D. Libri, *Nat Rev Mol Cell Biol* **2015**, *16*, 190–202;

[58] P. Sarnow, *Proceedings of the National Academy of Sciences* **1989**, *86*, 5795–5799.

[59] a) J. Guhaniyogi, G. Brewer, *Gene* **2001**, *265*, 11–23; b) P. Bernstein, J. Ross, *Trends in Biochemical Sciences* **1989**, *14*, 373–377; c) A. Sachs, *Current Opinion in Cell Biology* **1990**, *2*, 1092–1098;

[60] C. L. Will, R. Lührmann, *Cold Spring Harbor perspectives in biology* **2011**, *3*.

[61] R. Feil, M. F. Fraga, *Nature reviews. Genetics* **2012**, *13*, 97–109.

[62] a) R. Holliday, J. E. Pugh, *Science* **1975**, *187*, 226–232; b) A. D. Riggs, *Cytogenetics and cell genetics* **1975**, *14*, 9–25; c) R. Lister, M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q.-M. Ngo et al., *Nature* **2009**, *462*, 315–322;

[63] A. Bird, *Genes & development* **2002**, *16*, 6–21.

[64] a) M. Ehrlich, M. A. Gama-Sosa, L. H. Huang, R. M. Midgett, K. C. Kuo, R. A. McCune, C. Gehrke, *Nucleic acids research* **1982**, *10*, 2709–2721; b) A. P. Bird, *Trends in genetics : TIG* **1987**, *3*, 342–347;

[65] a) A. L. Hughes, J. R. Kelley, R. J. Klose, *Biochimica et biophysica acta. Gene regulatory mechanisms* **2020**, *1863*, 194567; b) P. Saravanaraman, M. Selvam, C. Ashok, L. Srijyothi, S. Baluchamy, *Biochimie* **2020**, *176*, 85–102;

[66] a) S. H. Cross, R. R. Meehan, X. Nan, A. Bird, *Nat Genet* **1997**, *16*, 256–259; b) B. Hendrich, A. Bird, *Molecular and cellular biology* **1998**, *18*, 6538–6547; c) X. Nan, F. Campoy, A. Bird, *Cell* **1997**, *88*, 471–481;

[67] E. P. Nora, E. Heard, *Cold Spring Harbor symposia on quantitative biology* **2010**, *75*, 333–344.

[68] Z. D. Smith, A. Meissner, *Nature reviews. Genetics* **2013**, *14*, 204–220.

[69] A. C. Ferguson-Smith, *Nature reviews. Genetics* **2011**, *12*, 565–575.

[70] V. Tucci, A. R. Isles, G. Kelsey, A. C. Ferguson-Smith, *Cell* **2019**, *176*, 952–965.

[71] a) D. Subramaniam, R. Thombre, A. Dhar, S. Anant, *Frontiers in oncology* **2014**, *4*, 80; b) Y. C.-Y. Yu, T. Z. Hui, T.-H. Kao, H.-F. Liao, C.-Y. Yang, C.-C. Hou, H.-T. Hsieh, J.-Y. Chang, Y.-T. Tsai, M. Pinskaya et al., *Frontiers in cell and developmental biology* **2020**, *8*, 103;

[72] M. Esteller, *The New England journal of medicine* **2008**, *358*, 1148–1159.

[73] S. Tajima, I. Suetake, K. Takeshita, A. Nakagawa, H. Kimura, *Advances in experimental medicine and biology* **2016**, *945*, 63–86.

[74] M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind et al., *Science (New York, N.Y.)* **2009**, *324*, 930–935.

[75] W. A. Pastor, L. Aravind, A. Rao, *Nat Rev Mol Cell Biol* **2013**, *14*, 341–356.

[76] L. Hu, Z. Li, J. Cheng, Q. Rao, W. Gong, M. Liu, Y. G. Shi, J. Zhu, P. Wang, Y. Xu, *Cell* **2013**, *155*, 1545–1555.

[77] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science (New York, N.Y.)* **2011**, *333*, 1300–1303.

[78] F. Wang, O. K. Zahid, U. Ghanty, R. M. Kohli, A. R. Hall, *Scientific reports* **2020**, *10*, 20253.

[79] D. Globisch, M. Münzel, M. Müller, S. Michalakis, M. Wagner, S. Koch, T. Brückl, M. Biel, T. Carell, *PLoS ONE* **2010**, *5*, e15367.

[80] A. Szwagierczak, S. Bultmann, C. S. Schmidt, F. Spada, H. Leonhardt, *Nucleic acids research* **2010**, *38*, e181.

[81] J. An, A. Rao, M. Ko, *Experimental & molecular medicine* **2017**, *49*, e323.

[82] L. Hu, J. Lu, J. Cheng, Q. Rao, Z. Li, H. Hou, Z. Lou, L. Zhang, W. Li, W. Gong et al., *Nature* **2015**, *527*, 118–122.

[83] Y.-F. He, B.-Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li et al., *Science (New York, N.Y.)* **2011**, *333*, 1303–1307.

[84] X. Wu, Y. Zhang, *Nature reviews. Genetics* **2017**, *18*, 517–534.

[85] a) C.-X. Song, K. E. Szulwach, Q. Dai, Y. Fu, S.-Q. Mao, L. Lin, C. Street, Y. Li, M. Poidevin, H. Wu et al., *Cell* **2013**, *153*, 678–691; b) L. Shen, H. Wu, D. Diep, S. Yamaguchi, A. C. D'Alessio, H.-L. Fung, K. Zhang, Y. Zhang, *Cell* **2013**, *153*, 692–706;

[86] a) E.-A. Raiber, P. Murat, D. Y. Chirgadze, D. Beraldi, B. F. Luisi, S. Balasubramanian, *Nature structural & molecular biology* **2015**, *22*, 44–49; b) M. Iurlaro, G. Ficz, D. Oxley, E.-A. Raiber, M. Bachman, M. J. Booth, S. Andrews, S. Balasubramanian, W. Reik, *Genome biology* **2013**, *14*, R119; c) C. G. Spruijt, F. Gnerlich, A. H. Smits, T. Pfaffeneder, P. W. T. C. Jansen, C. Bauer, M. Münzel, M. Wagner, M. Müller, F. Khan et al., *Cell* **2013**, *152*, 1146–1159;

[87] L. Wang, Y. Zhou, L. Xu, R. Xiao, X. Lu, L. Chen, J. Chong, H. Li, C. He, X.-D. Fu et al., *Nature* **2015**, *523*, 621–625.

[88] Y. Zhang, Z. Sun, J. Jia, T. Du, N. Zhang, Y. Tang, Y. Fang, D. Fang, *Advances in experimental medicine and biology* **2021**, *1283*, 1–16.

[89] B. D. Strahl, C. D. Allis, *Nature* **2000**, *403*, 41–45.

[90] E. L. Greer, Y. Shi, *Nature reviews. Genetics* **2012**, *13*, 343–357.

[91] J. Zhang, L. Jing, M. Li, L. He, Z. Guo, *Molecular medicine reports* **2019**, *19*, 3963–3971.

[92] J. C. Black, C. van Rechem, J. R. Whetstine, *Molecular cell* **2012**, *48*, 491–507.

[93] J. C. Eissenberg, S. C. R. Elgin, *Current opinion in genetics & development* **2000**, *10*, 204–210.

[94] D. N. Weinberg, S. Papillon-Cavanagh, H. Chen, Y. Yue, X. Chen, K. N. Rajagopalan, C. Horth, J. T. McGuire, X. Xu, H. Nikbakht et al., *Nature* **2019**, *573*, 281–286.

[95] V. G. ALLFREY, R. FAULKNER, A. E. MIRSKY, *Proceedings of the National Academy of Sciences* **1964**, *51*, 786–794.

[96] A. P. Bird, A. P. Wolffe, *Cell* **1999**, *99*, 451–454.

[97] H. H. Ng, Y. Zhang, B. Hendrich, C. A. Johnson, B. M. Turner, H. Erdjument-Bromage, P. Tempst, D. Reinberg, A. Bird, *Nat Genet* **1999**, *23*, 58–61.

[98] W. A. Burgers, F. Fuks, T. Kouzarides, *Trends in genetics : TIG* **2002**, *18*, 275–277.

[99] M. P. Creyghton, A. W. Cheng, G. G. Welstead, T. Kooistra, B. W. Carey, E. J. Steine, J. Hanna, M. A. Lodato, G. M. Frampton, P. A. Sharp et al., *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107*, 21931–21936.

[100] R. Zangi, A. Arrieta, F. P. Cossío, *Journal of molecular biology* **2010**, *400*, 632–644.

[101] S. Adam, H. Anteneh, M. Hornisch, V. Wagner, J. Lu, N. E. Radde, P. Bashtrykov, J. Song, A. Jeltsch, *Nat Commun* **2020**, *11*, 3723.

[102] Z.-M. Zhang, R. Lu, P. Wang, Y. Yu, D. Chen, L. Gao, S. Liu, D. Ji, S. B. Rothbart, Y. Wang et al., *Nature* **2018**, *554*, 387–391.

[103] J. Wolffgramm, B. Buchmuller, S. Palei, Á. Muñoz-López, J. Kanne, P. Janning, M. R. Schweiger, D. Summerer, *Angewandte Chemie International Edition* **2021**, *60*, 13507–13512.

[104] X. Cheng, *Annual review of biophysics and biomolecular structure* **1995**, *24*, 293–318.

[105] M. Peräkylä, *J. Am. Chem. Soc.* **1998**, *120*, 12895–12902.

[106] T. Shlomi, J. D. Rabinowitz, *Nature chemical biology* **2013**, *9*, 293–294.

[107] F. Xu, C. Mao, Y. Ding, C. Rui, L. Wu, A. Shi, H. Zhang, L. Zhang, Z. Xu, *Current medicinal chemistry* **2010**, *17*, 4052–4071.

[108] a) S. Kumar, X. Cheng, S. Klimasauskas, S. Mi, J. Posfai, R. J. Roberts, G. G. Wilson, *Nucleic acids research* **1994**, *22*, 1–10; b) J. Pósfai, A. S. Bhagwat, G. Pósfai, R. J. Roberts, *Nucleic acids research* **1989**, *17*, 2421–2435; c) R. Lauster, T. A. Trautner, M. Noyer-Weidner, *Journal of molecular biology* **1989**, *206*, 305–312;

[109] T. H. Bestor, *Human molecular genetics* **2000**, *9*, 2395–2402.

[110] O. V. Lukashevich, N. A. Cherepanova, R. Z. Jurkovska, A. Jeltsch, E. S. Gromova, *BMC biochemistry* **2016**, *17*, 7.

[111] T. H. Bestor, G. L. Verdine, *Current Opinion in Cell Biology* **1994**, *6*, 380–389.

[112] M. G. Goll, T. H. Bestor, *Annual review of biochemistry* **2005**, *74*, 481–514.

[113] M. G. Goll, F. Kirpekar, K. A. Maggert, J. A. Yoder, C.-L. Hsieh, X. Zhang, K. G. Golic, S. E. Jacobsen, T. H. Bestor, *Science (New York, N.Y.)* **2006**, *311*, 395–398.

[114] J. A. Yoder, N. S. Soman, G. L. Verdine, T. H. Bestor, *Journal of molecular biology* **1997**, *270*, 385–395.

[115] A. Hermann, H. Gowher, A. Jeltsch, *Cellular and molecular life sciences : CMLS* **2004**, *61*, 2571–2587.

[116] H. Denis, '. N. Ndlovu, F. Fuks, *EMBO Reports* **2011**, *12*, 647–656.

[117] a) M. Bostick, J. K. Kim, P.-O. Estève, A. Clark, S. Pradhan, S. E. Jacobsen, *Science (New York, N.Y.)* **2007**, *317*, 1760–1764; b) J. Sharif, M. Muto, S. Takebayashi, I. Suetake, A. Iwamatsu, T. A. Endo, J. Shinga, Y. Mizutani-Koseki, T. Toyoda, K. Okamura et al., *Nature* **2007**, *450*, 908–912; c) T. Li, L. Wang, Y. Du, S. Xie, X. Yang, F. Lian, Z. Zhou, C. Qian, *Nucleic acids research* **2018**, *46*, 3218–3231;

[118] a) H. Hashimoto, J. R. Horton, X. Zhang, M. Bostick, S. E. Jacobsen, X. Cheng, *Nature* **2008**, *455*, 826–829; b) K. Arita, M. Ariyoshi, H. Tochio, Y. Nakamura, M. Shirakawa, *Nature* **2008**, *455*, 818–821;

[119] J. Song, O. Rechkoblit, T. H. Bestor, D. J. Patel, *Science (New York, N.Y.)* **2011**, *331*, 1036–1040.

[120] I. Callebaut, J.-C. Courvalin, J.-P. Mornon, *FEBS Letters* **1999**, *446*, 189–193.

[121] T. Bestor, A. Laudano, R. Mattaliano, V. Ingram, *Journal of molecular biology* **1988**, *203*, 971–983.

[122] U. Aapola, K. Kawasaki, H. S. Scott, J. Ollila, M. Vihinen, M. Heino, A. Shintani, S. Minoshima, K. Krohn, S. E. Antonarakis et al., *Genomics* **2000**, *65*, 293–298.

[123] M. Okano, S. Xie, E. Li, *Nat Genet* **1998**, *19*, 219–220.

[124] H. Gowher, A. Jeltsch, *Biochemical Society transactions* **2018**, *46*, 1191–1202.

[125] T. Chen, Y. Ueda, S. Xie, E. Li, *The Journal of biological chemistry* **2002**, *277*, 38746–38754.

[126] J. Liao, R. Karnik, H. Gu, M. J. Ziller, K. Clement, A. M. Tsankov, V. Akopian, C. A. Gifford, J. Donaghey, C. Galonska et al., *Nat Genet* **2015**, *47*, 469–478.

[127] C. A. Gordon, S. R. Hartono, F. Chédin, *PLoS ONE* **2013**, *8*, e69486.

[128] H. Gowher, P. Loutchanwoot, O. Vorobjeva, V. Handa, R. Z. Jurkowska, T. P. Jurkowski, A. Jeltsch, *Journal of molecular biology* **2006**, *357*, 928–941.

[129] H. Gowher, A. Jeltsch, *The Journal of biological chemistry* **2002**, *277*, 20409–20414.

[130] S. Reither, F. Li, H. Gowher, A. Jeltsch, *Journal of molecular biology* **2003**, *329*, 675–684.

[131] L. Gao, M. Emperle, Y. Guo, S. A. Grimm, W. Ren, S. Adam, H. Uryu, Z.-M. Zhang, D. Chen, J. Yin et al., *Nature communications* **2020**, *11*, 3355.

[132] I. Rhee, K. W. Jair, R. W. Yen, C. Lengauer, J. G. Herman, K. W. Kinzler, B. Vogelstein, S. B. Baylin, K. E. Schuebel, *Nature* **2000**, *404*, 1003–1007.

[133] Y. He, J. R. Ecker, *Annual review of genomics and human genetics* **2015**, *16*, 55–77.

[134] J. Arand, D. Spieler, T. Karius, M. R. Branco, D. Meilinger, A. Meissner, T. Jenuwein, G. Xu, H. Leonhardt, V. Wolf et al., *PLoS genetics* **2012**, *8*, e1002750.

[135] J.-H. Lee, S.-J. Park, K. Nakai, *Scientific reports* **2017**, *7*, 11295.

[136] a) L. Chen, K. Chen, L. A. Lavery, S. A. Baker, C. A. Shaw, W. Li, H. Y. Zoghbi, *Proceedings of the National Academy of Sciences of the United States of America* **2015**, *112*, 5509–5514; b) B. Kinde, D. Y. Wu, M. E. Greenberg, H. W. Gabel, *Proceedings of the National Academy of Sciences of the United States of America* **2016**, *113*, 15114–15119; c) B. Kinde, H. W. Gabel, C. S. Gilbert, E. C. Griffith, M. E. Greenberg, *Proceedings of the National Academy of Sciences of the United States of America* **2015**, *112*, 6800–6806;

[137] a) B. H. Ramsahoye, D. Biniszkiewicz, F. Lyko, V. Clark, A. P. Bird, R. Jaenisch, *Proceedings of the National Academy of Sciences* **2000**, *97*, 5237–5242; b) L. Laurent, E. Wong, G. Li, T. Huynh, A. Tsirigos, C. T. Ong, H. M. Low, K. W. Kin Sung, I. Rigoutsos, J. Loring et al., *Genome research* **2010**, *20*, 320–331;

[138] M. Wang, K. Zhang, V. Ngo, C. Liu, S. Fan, J. W. Whitaker, Y. Chen, R. Ai, Z. Chen, J. Wang et al., *Nucleic acids research* **2019**, *47*, 6753–6768.

[139] a) T. Chen, N. Tsujimoto, E. Li, *Molecular and cellular biology* **2004**, *24*, 9048–9058; b) C. Qiu, K. Sawada, X. Zhang, X. Cheng, *Nature structural biology* **2002**, *9*, 217–224; c) I. Stec, S. B. Nagl, G.-J. B. van Ommen, J. T. den Dunnen, *FEBS Letters* **2000**, *473*, 1–5; d) Y.-Z. Ge, M.-T. Pu, H. Gowher, H.-P. Wu, J.-P. Ding, A. Jeltsch, G.-L. Xu, *The Journal of biological chemistry* **2004**, *279*, 25447–25454;

[140] a) T. Baubec, D. F. Colombo, C. Wirbelauer, J. Schmidt, L. Burger, A. R. Krebs, A. Akalin, D. Schübeler, *Nature* **2015**, *520*, 243–247; b) F. Neri, S. Rapelli, A. Krepelova, D. Incarnato, C. Parlato, G. Basile, M. Maldotti, F. Anselmi, S. Oliviero, *Nature* **2017**, *543*, 72–77; c) G. Rondelet, T. Dal Maso, L. Willems, J. Wouters, *Journal of structural biology* **2016**, *194*, 357–367; d) E. J. Wagner, P. B. Carpenter, *Nat Rev Mol Cell Biol* **2012**, *13*, 115–126;

[141] L. Rinaldi, D. Datta, J. Serrat, L. Morey, G. Solanas, A. Avgustinova, E. Blanco, J. I. Pons, D. Matallanas, A. von Kriegsheim et al., *Cell stem cell* **2016**, *19*, 491–501.

[142] a) D. N. Weinberg, P. Rosenbaum, X. Chen, D. Barrows, C. Horth, M. R. Marunde, I. K. Popova, Z. B. Gillespie, M.-C. Keogh, C. Lu et al., *Nat Genet* **2021**, *53*, 794–800; b) K. Kibe, K. Shirane, H. Ohishi, S. Uemura, H. Toh, H. Sasaki, *PLoS genetics* **2021**, *17*, e1009570;

[143] A. Argentaro, J.-C. Yang, L. Chapman, M. S. Kowalczyk, R. J. Gibbons, D. R. Higgs, D. Neuhaus, D. Rhodes, *Proceedings of the National Academy of Sciences* **2007**, *104*, 11939–11944.

[144] X. Guo, L. Wang, J. Li, Z. Ding, J. Xiao, X. Yin, S. He, P. Shi, L. Dong, G. Li et al., *Nature* **2015**, *517*, 640–644.

[145] a) F. Fuks, W. A. Burgers, N. Godin, M. Kasai, T. Kouzarides, *The EMBO Journal* **2001**, *20*, 2536–2544; b) K. E. Bachman, M. R. Rountree, S. B. Baylin, *The Journal of biological chemistry* **2001**, *276*, 32282–32287;

[146] a) S. Maenohara, M. Unoki, H. Toh, H. Ohishi, J. Sharif, H. Koseki, H. Sasaki, *PLoS genetics* **2017**, *13*, e1007042; b) Y. Jia, P. Li, L. Fang, H. Zhu, L. Xu, H. Cheng, J. Zhang, F. Li, Y. Feng, Y. Li et al., *Cell discovery* **2016**, *2*, 16007;

[147] A. N. Siddique, S. Nunna, A. Rajavelu, Y. Zhang, R. Z. Jurkowska, R. Reinhardt, M. G. Rots, S. Ragozin, T. P. Jurkowski, A. Jeltsch, *Journal of molecular biology* **2013**, *425*, 479–491.

[148] D. Jia, R. Z. Jurkowska, X. Zhang, A. Jeltsch, X. Cheng, *Nature* **2007**, *449*, 248–251.

[149] K. Hata, M. Okano, H. Lei, E. Li, *Development* **2002**, *129*, 1983–1993.

[150] a) D. Bourc'his, G. L. Xu, C. S. Lin, B. Bollman, T. H. Bestor, *Science* **2001**, *294*, 2536–2539; b) D. Bourc'his, T. H. Bestor, *Nature* **2004**, *431*, 96–99; c) M. Kaneda, M. Okano, K. Hata, T. Sado, N. Tsujimoto, E. Li, H. Sasaki, *Nature* **2004**, *429*, 900–903;

[151] J. Gallego-Bartolomé, *The New phytologist* **2020**, *227*, 38–44.

[152] Y. Lei, Y.-H. Huang, M. A. Goodell, *Genome biology* **2018**, *19*, 187.

[153] A. Klug, *Annual review of biochemistry* **2010**, *79*, 213–231.

[154] J. Boch, H. Scholze, S. Schornack, A. Landgraf, S. Hahn, S. Kay, T. Lahaye, A. Nickstadt, U. Bonas, *Science (New York, N.Y.)* **2009**, *326*, 1509–1512.

[155] P. D. Hsu, D. A. Scott, J. A. Weinstein, F. A. Ran, S. Konermann, V. Agarwala, Y. Li, E. J. Fine, X. Wu, O. Shalem et al., *Nature biotechnology* **2013**, *31*, 827–832.

[156] P. D. Hsu, E. S. Lander, F. Zhang, *Cell* **2014**, *157*, 1262–1278.

[157] T. Cermak, E. L. Doyle, M. Christian, L. Wang, Y. Zhang, C. Schmidt, J. A. Baller, N. V. Somia, A. J. Bogdanove, D. F. Voytas, *Nucleic acids research* **2011**, *39*, e82.

[158] G. L. Xu, T. H. Bestor, *Nat Genet* **1997**, *17*, 376–378.

[159] a) A. R. McNamara, P. J. Hurd, A. E. F. Smith, K. G. Ford, *Nucleic acids research* **2002**, *30*, 3818–3830; b) A. E. Smith, K. G. Ford, *Nucleic acids research* **2007**, *35*, 740–754;

[160] D. L. Bernstein, J. E. Le Lay, E. G. Ruano, K. H. Kaestner, *The Journal of clinical investigation* **2015**, *125*, 1998–2006.

[161] A. Vojta, P. Dobrinić, V. Tadić, L. Bočkor, P. Korać, B. Julg, M. Klasić, V. Zoldoš, *Nucleic acids research* **2016**, *44*, 5615–5628.

[162] J. I. McDonald, H. Celik, L. E. Rois, G. Fishberger, T. Fowler, R. Rees, A. Kramer, A. Martens, J. R. Edwards, G. A. Challen, *Biology open* **2016**, *5*, 866–874.

[163] X. S. Liu, H. Wu, X. Ji, Y. Stelzer, X. Wu, S. Czauderna, J. Shu, D. Dadon, R. A. Young, R. Jaenisch, *Cell* **2016**, *167*, 233-247.e17.

[164] P. Stepper, G. Kungulovski, R. Z. Jurkowska, T. Chandra, F. Krueger, R. Reinhardt, W. Reik, A. Jeltsch, T. P. Jurkowski, *Nucleic acids research* **2017**, *45*, 1703–1713.

[165] M. E. Tanenbaum, L. A. Gilbert, L. S. Qi, J. S. Weissman, R. D. Vale, *Cell* **2014**, *159*, 635–646.

[166] Y.-H. Huang, J. Su, Y. Lei, L. Brunetti, M. C. Gundry, X. Zhang, M. Jeong, W. Li, M. A. Goodell, *Genome biology* **2017**, *18*, 176.

[167] D. Hofacker, J. Broche, L. Laistner, S. Adam, P. Bashtrykov, A. Jeltsch, *International journal of molecular sciences* **2020**, *21*.

[168] Y. Lei, X. Zhang, J. Su, M. Jeong, M. C. Gundry, Y.-H. Huang, Y. Zhou, W. Li, M. A. Goodell, *Nature communications* **2017**, *8*, 16026.

[169] T. Xiong, G. E. Meister, R. E. Workman, N. C. Kato, M. J. Spellberg, F. Turker, W. Timp, M. Ostermeier, C. D. Novina, *Scientific reports* **2017**, *7*, 6732.

[170] S. R. Choudhury, Y. Cui, A. Narayanan, D. P. Gilley, N. Huda, C.-L. Lo, F. C. Zhou, D. Yernool, J. Irudayaraj, *Oncotarget* **2016**, *7*, 50380–50391.

[171] a) P. A. Jones, S. M. Taylor, *Cell* **1980**, *20*, 85–93; b) C. Stresemann, F. Lyko, *International journal of cancer* **2008**, *123*, 8–13;

[172] D. V. Santi, A. Norment, C. E. Garrett, *Proceedings of the National Academy of Sciences* **1984**, *81*, 6993–6997.

[173] C. Gros, J. Fahy, L. Halby, I. Dufau, A. Erdmann, J.-M. Gregoire, F. Ausseil, S. Vispé, P. B. Arimondo, *Biochimie* **2012**, *94*, 2280–2296.

[174] a) D. J. Gregory, L. Mikhaylova, A. V. Fedulov, *Epigenetics* **2012**, *7*, 344–349; b) D. J. Gregory, Y. Zhang, L. Kobzik, A. V. Fedulov, *Epigenetics* **2013**, *8*, 1205–1212;

[175] H. Chen, H. G. Kazemier, M. L. de Groote, M. H. J. Ruiters, G.-L. Xu, M. G. Rots, *Nucleic acids research* **2014**, *42*, 1563–1574.

[176] M. L. Maeder, J. F. Angstman, M. E. Richardson, S. J. Linder, V. M. Cascio, S. Q. Tsai, Q. H. Ho, J. D. Sander, D. Reyon, B. E. Bernstein et al., *Nat Biotech* **2013**, *31*, 1137–1142.

[177] a) S. Morita, H. Noguchi, T. Horii, K. Nakabayashi, M. Kimura, K. Okamura, A. Sakai, H. Nakashima, K. Hata, K. Nakashima et al., *Nature biotechnology* **2016**, *34*, 1060–1065; b) S. R. Choudhury, Y. Cui, K. Lubecka, B. Stefanska, J. Irudayaraj, *Oncotarget* **2016**, *7*, 46545–46556; c) X. S. Liu, H. Wu, M. Krzisch, X. Wu, J. Graef, J. Muffat, D. Hnisz, C. H. Li, B. Yuan, C. Xu et al., *Cell* **2018**, *172*, 979-992.e6;

[178] S. Stolzenburg, A. S. Beltran, T. Swift-Scanlan, A. G. Rivenbark, R. Rashwan, P. Blancafort, *Oncogene* **2015**, *34*, 5427–5435.

[179] J. Boch, U. Bonas, *Annual review of phytopathology* **2010**, *48*, 419–436.

[180] U. Bonas, R. E. Stall, B. Staskawicz, *Molecular & general genetics : MGG* **1989**, *218*, 127–136.

[181] M. J. Moscou, A. J. Bogdanove, *Science (New York, N.Y.)* **2009**, *326*, 1501.

[182] A. N.-S. Mak, P. Bradley, R. A. Cernadas, A. J. Bogdanove, B. L. Stoddard, *Science* **2012**, *335*, 716–719.

[183] A. Mak, P. Bradley, R. A. Cernadas, A. J. Bogdanove, B. L. Stoddard, *Structure of TAL effector PthXo1 bound to its DNA target*, **2012**.

[184] J. M. Rogers, L. A. Barrera, D. Reyon, J. D. Sander, M. Kellis, J. K. Joung, M. L. Bulyk, *Nat Commun* **2015**, *6*, 7440.

[185] R. Rohs, X. Jin, S. M. West, R. Joshi, B. Honig, R. S. Mann, *Annual review of biochemistry* **2010**, *79*, 233–269.

[186] D. Deng, C. Yan, X. Pan, M. Mahfouz, J. Wang, J.-K. Zhu, Y. Shi, N. Yan, *Science (New York, N.Y.)* **2012**, *335*, 720–723.

[187] D. Deng, P. Yin, C. Yan, X. Pan, X. Gong, S. Qi, T. Xie, M. Mahfouz, J.-K. Zhu, N. Yan et al., *Cell research* **2012**, *22*, 1502–1504.

[188] H. Gao, X. Wu, J. Chai, Z. Han, *Cell research* **2012**, *22*, 1716–1720.

[189] A. J. Bogdanove, S. Schornack, T. Lahaye, *Current opinion in plant biology* **2010**, *13*, 394–401.

[190] B. M. Lamb, A. C. Mercer, C. F. Barbas, *Nucleic acids research* **2013**, *41*, 9779–9785.

[191] a) J. C. Miller, S. Tan, G. Qiao, K. A. Barlow, J. Wang, D. F. Xia, X. Meng, D. E. Paschon, E. Leung, S. J. Hinkley et al., *Nature biotechnology* **2011**, *29*, 143–148; b) C. Mussolino, R. Morbitzer, F. Lütge, N. Dannemann, T. Lahaye, T. Cathomen, *Nucleic acids research* **2011**, *39*, 9283–9293; c) N. Sun, J. Liang, Z. Abil, H. Zhao, *Molecular bioSystems* **2012**, *8*, 1255–1263; d) F. Zhang, Le Cong, S. Lodato, S. Kosuri, G. M. Church, P. Arlotta, *Nature biotechnology* **2011**, *29*, 149–153;

[192] B. P. Hubbard, A. H. Badran, J. A. Zuris, J. P. Guilinger, K. M. Davis, L. Chen, S. Q. Tsai, J. D. Sander, J. K. Joung, D. R. Liu, *Nat Meth* **2015**, *12*, 939–942.

[193] T. Schreiber, U. Bonas, *Nucleic acids research* **2014**, *42*, 7160–7169.

[194] D. Reyon, S. Q. Tsai, C. Khayter, J. A. Foden, J. D. Sander, J. K. Joung, *Nature biotechnology* **2012**, *30*, 460–465.

[195] J. L. Schmid-Burgk, T. Schmidt, V. Kaiser, K. Höning, V. Hornung, *Nature biotechnology* **2013**, *31*, 76–81.

[196] S. Gogolok, C. Garcia-Diaz, S. M. Pollard, *Scientific reports* **2016**, *6*, 33209.

[197] a) J. F. Meckler, M. S. Bhakta, M.-S. Kim, R. Ovadia, C. H. Habrian, A. Zykovich, A. Yu, S. H. Lockwood, R. Morbitzer, J. Elsäesser et al., *Nucleic acids research* **2013**, *41*, 4118–4128; b) A. Garg, J. J. Lohmueller, P. A. Silver, T. Z. Armel, *Nucleic acids research* **2012**, *40*, 7584–7595;

[198] J. C. Miller, L. Zhang, D. F. Xia, J. J. Campo, I. V. Ankoudinova, D. Y. Guschin, J. E. Babiarz, X. Meng, S. J. Hinkley, S. C. Lam et al., *Nat Meth* **2015**, *12*, 465–471.

[199] a) T. J. Cradick, E. J. Fine, C. J. Antico, G. Bao, *Nucleic acids research* **2013**, *41*, 9584–9592; b) Y. Fu, J. A. Foden, C. Khayter, M. L. Maeder, D. Reyon, J. K. Joung, J. D. Sander, *Nature biotechnology* **2013**, *31*, 822–826; c) X. R. Bao, Y. Pan, C. M. Lee, T. H. Davis, G. Bao, *Nature protocols* **2021**, *16*, 10–26; d) E. J. Fine, T. J. Cradick, C. L. Zhao, Y. Lin, G. Bao, *Nucleic acids research* **2014**, *42*, e42;

[200] L. R. Polstein, P. Perez-Pinera, D. D. Kocak, C. M. Vockley, P. Bledsoe, L. Song, A. Safi, G. E. Crawford, T. E. Reddy, C. A. Gersbach, *Genome research* **2015**, *25*, 1158–1169.

[201] S. Jain, S. Shukla, C. Yang, M. Zhang, Z. Fatma, M. Lingamaneni, S. Abesteh, S. T. Lane, X. Xiong, Y. Wang et al., *Nature communications* **2021**, *12*, 606.

[202] a) A. Juillerat, C. Pessereau, G. Dubois, V. Guyot, A. Maréchal, J. Valton, F. Daboussi, L. Poirot, A. Duclert, P. Duchateau, *Scientific reports* **2015**, *5*, 8150; b) J. Yang, Y. Zhang, P. Yuan, Y. Zhou, C. Cai, Q. Ren, D. Wen, C. Chu, H. Qi, W. Wei, *Cell research* **2014**, *24*, 628–631;

[203] J. Valton, A. Dupuy, F. Daboussi, S. Thomas, A. Maréchal, R. Macmaster, K. Melliand, A. Juillerat, P. Duchateau, *The Journal of biological chemistry* **2012**, *287*, 38427–38432.

[204] a) B. I. M. Wicky, M. Stenta, M. Dal Peraro, *PLoS ONE* **2013**, *8*, e80261; b) S. Bultmann, R. Morbitzer, C. S. Schmidt, K. Thanisch, F. Spada, J. Elsaesser, T. Lahaye, H. Leonhardt, *Nucleic acids research* **2012**, *40*, 5368–5377;

[205] S. Maurer, M. Giess, O. Koch, D. Summerer, *ACS Chemical Biology* **2016**, *11*, 3294–3299.

[206] P. Rathi, S. Maurer, G. Kubik, D. Summerer, *Journal of the American Chemical Society* **2016**, *138*, 9910–9918.

[207] S. Maurer, B. Buchmuller, C. Ehrt, J. Jasper, O. Koch, D. Summerer, *Chemical science* **2018**, *9*, 7247–7252.

[208] S. Choi, K.-Y. Choi, *Expert opinion on drug discovery* **2017**, *12*, 293–303.

[209] D. P. Bondeson, C. M. Crews, *Annual review of pharmacology and toxicology* **2017**, *57*, 107–123.

[210] D. D. Young, P. G. Schultz, *ACS Chemical Biology* **2018**, *13*, 854–870.

[211] K. Krauskopf, K. Lang, *Current opinion in chemical biology* **2020**, *58*, 112–120.

[212] N. Ankenbruck, T. Courtney, Y. Naro, A. Deiters, *Angewandte Chemie International Edition* **2018**, *57*, 2768–2798.

[213] T. Courtney, A. Deiters, *Current opinion in chemical biology* **2018**, *46*, 99–107.

[214] R. M. Hughes, *Critical reviews in biochemistry and molecular biology* **2018**, *53*, 453–474.

[215] a) A. V. Karginov, Y. Zou, D. Shirvanyants, P. Kota, N. V. Dokholyan, D. D. Young, K. M. Hahn, A. Deiters, *J. Am. Chem. Soc.* **2011**, *133*, 420–423; b) N. Umeda, T. Ueno, C. Pohlmeyer, T. Nagano, T. Inoue, *J. Am. Chem. Soc.* **2011**, *133*, 12–14; c) K. A. Brown, Y. Zou, D. Shirvanyants, J. Zhang, S.

Samanta, P. K. Mantravadi, N. V. Dokholyan, A. Deiters, *Chemical communications (Cambridge, England)* **2015**, *51*, 5702–5705;

[216] a) E. R. Ballister, C. Aonbangkhen, A. M. Mayo, M. A. Lampson, D. M. Chenoweth, *Nat Commun* **2014**, *5*, 5475; b) E. R. Ballister, S. Ayloo, D. M. Chenoweth, M. A. Lampson, E. L. F. Holzbaur, *Current biology : CB* **2015**, *25*, R407-R408;

[217] M. Zimmermann, R. Cal, E. Janett, V. Hoffmann, C. G. Bochet, E. Constable, F. Beaufils, M. P. Wymann, *Angewandte Chemie International Edition* **2014**, *53*, 4717–4720.

[218] H. P. Nguyen, S. Stewart, M. N. Kukwikila, S. F. Jones, D. Offenbartl-Stiegert, S. Mao, S. Balasubramanian, S. Beck, S. Howorka, *Angewandte Chemie International Edition* **2019**, *58*, 6620–6624.

[219] A. Baumschlager, M. Rullan, M. Khammash, *Nat Commun* **2020**, *11*, 3834.

[220] a) N. C. Rockwell, Y.-S. Su, J. C. Lagarias, *Annual review of plant biology* **2006**, *57*, 837–858; b) Y. Hirose, T. Shimada, R. Narikawa, M. Katayama, M. Ikeuchi, *Proceedings of the National Academy of Sciences of the United States of America* **2008**, *105*, 9528–9533; c) J. J. Tabor, A. Levskaya, C. A. Voigt, *Journal of molecular biology* **2011**, *405*, 315–324;

[221] P. Jayaraman, K. Devarajan, T. K. Chua, H. Zhang, E. Gunawan, C. L. Poh, *Nucleic acids research* **2016**, *44*, 6994–7005.

[222] E. Peter, B. Dick, S. A. Baeurle, *Nat Commun* **2010**, *1*, 122.

[223] A. Baumschlager, S. K. Aoki, M. Khammash, *ACS synthetic biology* **2017**, *6*, 2157–2167.

[224] F. Kawano, H. Suzuki, A. Furuya, M. Sato, *Nat Commun* **2015**, *6*, 6256.

[225] M. J. Kennedy, R. M. Hughes, L. A. Peteya, J. W. Schwartz, M. D. Ehlers, C. L. Tucker, *Nature methods* **2010**, *7*, 973–975.

[226] J. Icha, M. Weber, J. C. Waters, C. Norden, *BioEssays : news and reviews in molecular, cellular and developmental biology* **2017**, *39*.

[227] F. Crick, *Journal of molecular biology* **1968**, *38*, 367–379.

[228] P. Lengyel, D. Söll, *Bacteriological reviews* **1969**, *33*, 264–301.

[229] F. H. CRICK, L. BARNETT, S. BRENNER, R. J. WATTS-TOBIN, *Nature* **1961**, *192*, 1227–1232.

[230] J. Baßler, E. Hurt, *Annual review of biochemistry* **2019**, *88*, 281–306.

[231] C. T. Caskey, R. Tompkins, E. Scolnick, T. Caryk, M. Nirenberg, *Science* **1968**, *162*, 135–138.

[232] M. R. Capecchi, *Proceedings of the National Academy of Sciences* **1967**, *58*, 1144–1151.

[233] a) G. A. Khoury, R. C. Baliban, C. A. Floudas, *Scientific reports* **2011**, *1*; b) A. G. Beck-Sickinger, K. Mörl, *Angewandte Chemie International Edition* **2006**, *45*, 1020;

[234] a) H. Shi, P. B. Moore, *RNA (New York, N.Y.)* **2000**, *6*, 1091–1105; b) H. Shi, P. B. Moore, *The crystal structure of yeast phenylalanine tRNA at 1.93 A resolution*, **2000**;

[235] GSL Biotech, SnapGene® software.

[236] S. Osawa, T. H. Jukes, *Journal of molecular evolution* **1989**, *28*, 271–278.

[237] B. G. Barrell, A. T. Bankier, J. Drouin, *Nature* **1979**, *282*, 189–194.

[238] F. Yamao, A. Muto, Y. Kawauchi, M. Iwami, S. Iwagami, Y. Azumi, S. Osawa, *Proceedings of the National Academy of Sciences* **1985**, *82*, 2306–2309.

[239] a) F. Caron, E. Meyer, *Nature* **1985**, *314*, 185–188; b) N. Hanyu, Y. Kuchino, S. Nishimura, H. Beier, *The EMBO Journal* **1986**, *5*, 1307–1311; c) E. Helftenbein, *Nucleic acids research* **1985**, *13*, 415–433; d) S. Horowitz, M. A. Gorovsky, *Proceedings of the National Academy of Sciences* **1985**, *82*, 2452–2455; e) J. R. Preer, L. B. Preer, B. M. Rudman, A. J. Barnett, *Nature* **1985**, *314*, 188–190;

[240] A. Ambrogelly, S. Palioura, D. Soll, *Nat Chem Biol* **2007**, *3*, 29–35.

[241] D. Söll, *Nature* **1988**, *331*, 662–663.

[242] L. Johansson, G. Gafvelin, E. S. J. Arnér, *Biochimica et biophysica acta* **2005**, *1726*, 1–13.

[243] A. Böck, K. Forchhammer, J. Heider, W. Leinfelder, G. Sawers, B. Veprek, F. Zinoni, *Molecular microbiology* **1991**, *5*, 515–520.

[244] T. C. Stadtman, *Annu. Rev. Biochem.* **1996**, *65*, 83–100.

[245] X.-M. Xu, B. A. Carlson, H. Mix, Y. Zhang, K. Saira, R. S. Glass, M. J. Berry, V. N. Gladyshev, D. L. Hatfield, *PLoS biology* **2007**, *5*, e4.

[246] M. Rother, J. A. Krzycki, *Archaea (Vancouver, B.C.)* **2010**, *2010*.

[247] G. Srinivasan, C. M. James, J. A. Krzycki, *Science (New York, N.Y.)* **2002**, *296*, 1459–1462.

[248] C. Polycarpo, A. Ambrogelly, A. Bérubé, S. M. Winbush, J. A. McCloskey, P. F. Crain, J. L. Wood, D. Söll, *Proceedings of the National Academy of Sciences* **2004**, *101*, 12450–12454.

[249] S. K. Blight, R. C. Larue, A. Mahapatra, D. G. Longstaff, E. Chang, G. Zhao, P. T. Kang, K. B. Green-Church, M. K. Chan, J. A. Krzycki, *Nature* **2004**, *431*, 333–335.

[250] B. Hao, G. Zhao, P. T. Kang, J. A. Soares, T. K. Ferguson, J. Gallucci, J. A. Krzycki, M. K. Chan, *Chemistry & biology* **2004**, *11*, 1317–1324.

[251] D. G. Longstaff, S. K. Blight, L. Zhang, K. B. Green-Church, J. A. Krzycki, *Molecular microbiology* **2007**, *63*, 229–241.

[252] a) R. B. Merrifield, *Advances in enzymology and related areas of molecular biology* **1969**, *32*, 221–296; b) J. M. Palomo, *RSC Adv* **2014**, *4*, 32658–32672;

[253] a) T. W. Muir, *Annual review of biochemistry* **2003**, *72*, 249–289; b) D. Schwarzer, P. A. Cole, *Current opinion in chemical biology* **2005**, *9*, 561–569;

[254] a) M. Ghosh, I. Ichetovkin, X. Song, J. S. Condeelis, D. S. Lawrence, *Journal of the American Chemical Society* **2002**, *124*, 2440–2441; b) S. B. Gunnoo, A. Madder, *ChemBioChem* **2016**, *17*, 529–553;

[255] C. C. Liu, P. G. Schultz, *Annual review of biochemistry* **2010**, *79*, 413–444.

[256] J. Xie, P. G. Schultz, *Current opinion in chemical biology* **2005**, *9*, 548–554.

[257] a) C. A. Aldinger, A.-K. Leisinger, G. L. Igloi, *The FEBS journal* **2012**, *279*, 3622–3638; b) D. C. Larkin, A. M. Williams, S. A. Martinis, G. E. Fox, *Nucleic acids research* **2002**, *30*, 2103–2113; c) C. L. Quinn, N. Tao, P. Schimmel, *Biochemistry* **1995**, *34*, 12489–12495;

[258] R. Giegé, M. Sissler, C. Florentz, *Nucleic acids research* **1998**, *26*, 5017–5035.

[259] a) C. R. Woese, G. J. Olsen, M. Ibba, D. Söll, *Microbiology and molecular biology reviews : MMBR* **2000**, *64*, 202–236; b) G. S. Zamudio, M. Palacios-Pérez, M. V. José, *Theory in biosciences = Theorie in den Biowissenschaften* **2020**, *139*, 77–85;

[260] a) T. G. Heckler, Y. Zama, T. Naka, S. M. Hecht, *The Journal of biological chemistry* **1983**, *258*, 4492–4495; b) C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, P. G. Schultz, *Science* **1989**, *244*, 182–188;

[261] a) J. Normanly, L. G. Kleina, J.-M. Masson, J. Abelson, J. H. Miller, *Journal of molecular biology* **1990**, *213*, 719–726; b) L. G. Kleina, J.-M. Masson, J. Normanly, J. Abelson, J. H. Miller, *Journal of molecular biology* **1990**, *213*, 705–717;

[262] R. Furter, *Protein science : a publication of the Protein Society* **1998**, *7*, 419–426.

[263] L. Wang, A. Brock, B. Herberich, P. G. Schultz, *Science* **2001**, *292*, 498–500.

[264] a) B. A. Steer, P. Schimmel, *The Journal of biological chemistry* **1999**, *274*, 35601–35606; b) H. Jakubowski, E. Goldman, *Microbiological reviews* **1992**, *56*, 412–429;

[265] L. Wang, T. J. Magliery, D. R. Liu, P. G. Schultz, *J. Am. Chem. Soc.* **2000**, *122*, 5010–5011.

[266] L. Wang, A. Brock, P. G. Schultz, *J. Am. Chem. Soc.* **2002**, *124*, 1836–1837.

[267] J. W. Chin, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *Proceedings of the National Academy of Sciences* **2002**, *99*, 11020–11024.

[268] J. W. Chin, T. A. Cropp, J. C. Anderson, M. Mukherji, Z. Zhang, P. G. Schultz, *Science (New York, N.Y.)* **2003**, *301*, 964–967.

[269] a) S. M. Hancock, R. Uprety, A. Deiters, J. W. Chin, *J. Am. Chem. Soc.* **2010**, *132*, 14819–14824; b) H. Neumann, S. Y. Peak-Chew, J. W. Chin, *Nature chemical biology* **2008**, *4*, 232–234;

[270] N. Wu, A. Deiters, T. A. Cropp, D. King, P. G. Schultz, *Journal of the American Chemical Society* **2004**, *126*, 14306–14307.

[271] E. A. Lemke, D. Summerer, B. H. Geierstanger, S. M. Brittain, P. G. Schultz, *Nature chemical biology* **2007**, *3*, 769–772.

[272] a) L. Wang, P. G. Schultz, *Chemistry & biology* **2001**, *8*, 883–890; b) D. Summerer, S. Chen, N. Wu, A. Deiters, J. W. Chin, P. G. Schultz, *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 9785–9789;

[273] D. Cervettini, S. Tang, S. D. Fried, J. C. W. Willis, L. F. H. Funke, L. J. Colwell, J. W. Chin, *Nature biotechnology* **2020**, *38*, 989–999.

[274] a) W. Brown, J. Liu, A. Deiters, *ACS Chemical Biology* **2018**, *13*, 2375–2386; b) D. de La Torre, J. W. Chin, *Nature reviews. Genetics* **2021**, *22*, 169–184;

[275] J. Fredens, K. Wang, D. de La Torre, L. F. H. Funke, W. E. Robertson, Y. Christova, T. Chia, W. H. Schmied, D. L. Dunkelmann, V. Beránek et al., *Nature* **2019**, *569*, 514–518.

[276] J. W. Chin, *Nature* **2017**, *550*, 53–60.

[277] A. Chatterjee, M. J. Lajoie, H. Xiao, G. M. Church, P. G. Schultz, *ChemBioChem* **2014**, *15*, 1782–1786.

[278] S. Greiss, J. W. Chin, *J. Am. Chem. Soc.* **2011**, *133*, 14196–14199.

[279] A. Bianco, F. M. Townsley, S. Greiss, K. Lang, J. W. Chin, *Nature chemical biology* **2012**, *8*, 748–750.

[280] a) Y. Chen, J. Ma, W. Lu, M. Tian, M. Thauvin, C. Yuan, M. Volovitch, Q. Wang, J. Holst, M. Liu et al., *Cell research* **2017**, *27*, 294–297; b) J. Liu, J. Hemphill, S. Samanta, M. Tsang, A. Deiters, *J. Am. Chem. Soc.* **2017**, *139*, 9100–9103;

[281] J.-Y. Kang, D. Kawaguchi, I. Coin, Z. Xiang, D. D. M. O'Leary, P. A. Slesinger, L. Wang, *Neuron* **2013**, *80*, 358–370.

[282] R. J. Ernst, T. P. Krogager, E. S. Maywood, R. Zanchi, V. Beránek, T. S. Elliott, N. P. Barry, M. H. Hastings, J. W. Chin, *Nature chemical biology* **2016**, *12*, 776–778.

[283] S. T. Cload, D. R. Liu, W. A. Froland, P. G. Schultz, *Chemistry & biology* **1996**, *3*, 1033–1038.

[284] a) T. J. Magliery, J. C. Anderson, P. G. Schultz, *Journal of molecular biology* **2001**, *307*, 755–769; b) B. Moore, B. C. Persson, C. C. Nelson, R. F. Gesteland, J. F. Atkins, *Journal of molecular biology* **2000**, *298*, 195–209;

[285] J. C. Anderson, N. Wu, S. W. Santoro, V. Lakshman, D. S. King, P. G. Schultz, *Proceedings of the National Academy of Sciences* **2004**, *101*, 7566–7571.

[286] a) H. Neumann, K. Wang, L. Davis, M. Garcia-Alai, J. W. Chin, *Nature* **2010**, *464*, 441–444; b) G. Hayashi, Y. Goto, H. Suga, *Chemistry & biology* **2010**, *17*, 320–321;

[287] P. Klán, T. Šolomek, C. G. Bochet, A. Blanc, R. Givens, M. Rubina, V. Popik, A. Kostikov, J. Wirz, *Chem. Rev.* **2013**, *113*, 119–191.

[288] a) A. Patchornik, B. Amit, R. B. Woodward, *J. Am. Chem. Soc.* **1970**, *92*, 6333–6335; b) J. M. Amatrudo, J. P. Olson, H. K. Agarwal, G. C. R. Ellis-Davies, *The European journal of neuroscience* **2015**, *41*, 5–16;

[289] A. Deiters, D. Groff, Y. Ryu, J. Xie, P. G. Schultz, *Angewandte Chemie (International ed. in English)* **2006**, *45*, 2728–2731.

[290] W. Ren, A. Ji, H. Ai, *Journal of the American Chemical Society* **2015**, *137*, 2155–2158.

[291] J. Luo, J. Torres-Kolbus, J. Liu, A. Deiters, *ChemBioChem* **2017**, *18*, 1442–1447.

[292] D. P. Nguyen, M. Mahesh, S. J. Elsässer, S. M. Hancock, C. Uttamapinant, J. W. Chin, *Journal of the American Chemical Society* **2014**, *136*, 2240–2243.

[293] R. Rakauskaitė, G. Urbanavičiūtė, A. Rukšėnaitė, Z. Liutkevičiūtė, R. Juškėnas, V. Masevičius, S. Klimašauskas, *Chemical communications (Cambridge, England)* **2015**, *51*, 8245–8248.

[294] J. E. Corrie, T. Furuta, R. Givens, A. L. Yousef, M. Goeldner in *Dynamic Studies in Biology* (Eds.: M. Goeldner, R. S. Givens), Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, FRG, **2005**.

[295] J. W. Walker, G. P. Reid, J. A. McCray, D. R. Trentham, *J. Am. Chem. Soc.* **1988**, *110*, 7170–7177.

[296] Y. V. Il'ichev, J. Wirz, *J. Phys. Chem. A* **2000**, *104*, 7856–7870.

[297] M. Goard, G. Aakalu, O. D. Fedoryak, C. Quinonez, J. St Julien, S. J. Poteet, E. M. Schuman, T. M. Dore, *Chemistry & biology* **2005**, *12*, 685–693.

[298] O. Sadovski, A. S. I. Jaikaran, S. Samanta, M. R. Fabian, R. J. O. Dowling, N. Sonenberg, G. A. Woolley, *Bioorganic & medicinal chemistry* **2010**, *18*, 7746–7752.

[299] J.-Y. Kang, D. Kawaguchi, L. Wang, *Journal of visualized experiments : JoVE* **2016**, e53818.

[300] a) A. Gautier, A. Deiters, J. W. Chin, *J. Am. Chem. Soc.* **2011**, *133*, 2124–2127; b) J. Hemphill, E. K. Borchardt, K. Brown, A. Asokan, A. Deiters, *J. Am. Chem. Soc.* **2015**, *137*, 5642–5645; c) J. Hemphill, C. Chou, J. W. Chin, A. Deiters, *J. Am. Chem. Soc.* **2013**, *135*, 13433–13439; d) A. Liaunardy-Jopeace, B. L. Murton, M. Mahesh, J. W. Chin, J. R. James, *Nature structural & molecular biology* **2017**, *24*, 1155–1163; e) J. Luo, E. Arbely, J. Zhang, C. Chou, R. Uprety, J. W. Chin, A. Deiters, *Chemical communications (Cambridge, England)* **2016**, *52*, 8529–8532; f) O. S. Walker, S. J. Elsässer, M. Mahesh, M. Bachman, S. Balasubramanian, J. W. Chin, *J. Am. Chem. Soc.* **2016**, *138*, 718–721;

[301] R. Uprety, J. Luo, J. Liu, Y. Naro, S. Samanta, A. Deiters, *ChemBioChem* **2014**, *15*, 1793–1799.

[302] J. Luo, R. Uprety, Y. Naro, C. Chou, D. P. Nguyen, J. W. Chin, A. Deiters, *J. Am. Chem. Soc.* **2014**, *136*, 15551–15558.

[303] J. E. T. Corrie, Y. Katayama, G. P. Reid, M. Anson, D. R. Trentham, R. M. Sweet, K. Moffat, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **1992**, *340*, 233–244.

[304] R. A. Blidner, K. R. Svoboda, R. P. Hammer, W. T. Monroe, *Molecular bioSystems* **2008**, *4*, 431–440.

[305] J. E. T. Corrie, V. R. N. Munasinghe, D. R. Trentham, A. Barth, *Photochemical & photobiological sciences : Official journal of the European Photochemistry Association and the European Society for Photobiology* **2008**, *7*, 84–97.

[306] T. M. Courtney, A. Deiters, *Nature communications* **2019**, *10*, 4384.

[307] T. Bridge, S. A. Shaikh, P. Thomas, J. Botta, P. J. McCormick, A. Sachdeva, *Angewandte Chemie International Edition* **2019**, *58*, 17986–17993.

[308] B. Jedlitzke, Z. Yilmaz, W. Dörner, H. D. Mootz, *Angewandte Chemie International Edition* **2020**, *59*, 1506–1510.

[309] C.-L. Lo, S. R. Choudhury, J. Irudayaraj, F. C. Zhou, *Scientific reports* **2017**, *7*, 42047.

[310] a) H. Gowher, K. Liebert, A. Hermann, G. Xu, A. Jeltsch, *Journal of Biological Chemistry* **2005**, *280*, 13341–13348; b) X. Cheng, R. M. Blumenthal, *Structure (London, England : 1993)* **2008**, *16*, 341–350;

[311] L. Zhou, X. Cheng, B. A. Connolly, M. J. Dickman, P. J. Hurd, D. P. Hornby, *Journal of molecular biology* **2002**, *321*, 591–599.

[312] A. Chatterjee, H. Xiao, M. Bollong, H. Ai, P. G. Schultz, *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110*, 11803–11808.

[313] F. Buhr, J. Kohl-Landgraf, S. tom Dieck, C. Hanus, D. Chatterjee, A. Hegelein, E. M. Schuman, J. Wachtveitl, H. Schwalbe, *Angewandte Chemie (International ed. in English)* **2015**, *54*, 3717–3721.

[314] Y. Zhang, C. Rohde, S. Tierling, T. P. Jurkowski, C. Bock, D. Santacruz, S. Ragozin, R. Reinhardt, M. Groth, J. Walter et al., *PLoS genetics* **2009**, *5*, e1000438.

[315] I. Rhee, K. E. Bachman, B. H. Park, K.-W. Jair, R.-W. C. Yen, K. E. Schuebel, H. Cui, A. P. Feinberg, C. Lengauer, K. W. Kinzler et al., *Nature* **2002**, *416*, 552–556.

[316] L. Yang, R. Rau, M. A. Goodell, *Nature reviews. Cancer* **2015**, *15*, 152–165.

[317] J. G. Tate, S. Bamford, H. C. Jubb, Z. Sondka, D. M. Beare, N. Bindal, H. Boutselakis, C. G. Cole, C. Creatore, E. Dawson et al., *Nucleic acids research* **2019**, *47*, D941-D947.

[318] T. J. Ley, L. Ding, M. J. Walter, M. D. McLellan, T. Lamprecht, D. E. Larson, C. Kandoth, J. E. Payton, J. Baty, J. Welch et al., *The New England journal of medicine* **2010**, *363*, 2424–2433.

[319] S. J. Kim, H. Zhao, S. Hardikar, A. K. Singh, M. A. Goodell, T. Chen, *Blood* **2013**, *122*, 4086–4089.

[320] D. A. Russler-Germain, D. H. Spencer, M. A. Young, T. L. Lamprecht, C. A. Miller, R. Fulton, M. R. Meyer, P. Erdmann-Gilmore, R. R. Townsend, R. K. Wilson et al., *Cancer cell* **2014**, *25*, 442–454.

[321] M. Emperle, M. Dukatz, S. Kunert, K. Holzer, A. Rajavelu, R. Z. Jurkowska, A. Jeltsch, *Scientific reports* **2018**, *8*, 13242.

[322] J. Koya, K. Kataoka, T. Sato, M. Bando, Y. Kato, T. Tsuruta-Kishino, H. Kobayashi, K. Narukawa, H. Miyoshi, K. Shirahige et al., *Nature communications* **2016**, *7*, 10924.

[323] X.-J. Yan, J. Xu, Z.-H. Gu, C.-M. Pan, G. Lu, Y. Shen, J.-Y. Shi, Y.-M. Zhu, L. Tang, X.-W. Zhang et al., *Nat Genet* **2011**, *43*, 309–315.

[324] C. Holz-Schietinger, D. M. Matje, N. O. Reich, *The Journal of biological chemistry* **2012**, *287*, 30941–30951.

[325] H. Anteneh, J. Fang, J. Song, *Nature communications* **2020**, *11*, 2294.

[326] J. E. Sandoval, Y.-H. Huang, A. Muise, M. A. Goodell, N. O. Reich, *The Journal of biological chemistry* **2019**, *294*, 4898–4910.

[327] D. A. Khrabrova, A. G. Loiko, A. A. Tolkacheva, N. A. Cherepanova, M. I. Zvereva, O. V. Kirsanova, E. S. Gromova, *Biomolecules* **2019**, *10*.

[328] C. Holz-Schietinger, D. M. Matje, M. F. Harrison, N. O. Reich, *The Journal of biological chemistry* **2011**, *286*, 41479–41488.

[329] a) W. R. Jeck, A. P. Siebold, N. E. Sharpless, *Aging cell* **2012**, *11*, 727–731; b) N. Martin, D. Beach, J. Gil, *Trends in molecular medicine* **2014**, *20*, 667–674;

[330] R. Valgardsdottir, I. Chiodi, M. Giordano, F. Cobianchi, S. Riva, G. Biamonti, *Molecular Biology of the Cell* **2005**, *16*, 2597–2604.

[331] C. Jolly, L. Konecny, D. L. Grady, Y. A. Kutskova, J. J. Cotto, R. I. Morimoto, C. Vourc'h, *The Journal of cell biology* **2002**, *156*, 775–781.

[332] Á. Muñoz-López, B. Buchmuller, J. Wolffgramm, A. Jung, M. Hussong, J. Kanne, M. R. Schweiger, D. Summerer, *Angewandte Chemie (International ed. in English)* **2020**, *59*, 8927–8931.

[333] A. Brulport, D. Vaiman, M.-C. Chagnon, L. Le Corre, *Chemosphere* **2020**, *241*, 125092.

[334] Y. Wang, C. Ma, Y. Sun, Y. Li, L. Kang, Y. Jiang, *BMC genomics* **2017**, *18*, 780.

[335] M. Shi, S. Wang, Y. Yao, Y. Li, H. Zhang, F. Han, H. Nie, J. Su, Z. Wang, L. Yue et al., *Scientific reports* **2014**, *4*, 7545.

[336] D. Kandilya, S. Shyamasundar, D. K. Singh, A. Banik, M. P. Hande, W. Stünkel, Y. S. Chong, S. T. Dheen, *Scientific reports* **2020**, *10*, 15676.

[337] D. A. van der Plaat, K. de Jong, M. de Vries, C. C. van Diemen, I. Nedeljković, N. Amin, H. Kromhout, R. Vermeulen, D. S. Postma, C. M. van Duijn et al., *Occupational and environmental medicine* **2018**, *75*, 427–435.

[338] C. Li, Y. Li, G. Zhou, Y. Gao, S. Ma, Y. Chen, J. Song, X. Wang, *BMC genomics* **2018**, *19*, 638.

[339] J. W. Abdulrahim, L. C. Kwee, E. Grass, I. C. Siegler, R. Williams, R. Karra, W. E. Kraus, S. G. Gregory, S. H. Shah, *Journal of the American Heart Association* **2019**, *8*, e013228.

[340] S. Fan, J. Tang, N. Li, Y. Zhao, R. Ai, K. Zhang, M. Wang, W. Du, W. Wang, *NPJ genomic medicine* **2019**, *4*, 2.

[341] C. Jeronimo, F. Robert, *Trends in cell biology* **2017**, *27*, 765–783.

[342] G. S. Winkler, T. G. Petrakis, S. Ethelberg, M. Tokunaga, H. Erdjument-Bromage, P. Tempst, J. Q. Svejstrup, *The Journal of biological chemistry* **2001**, *276*, 32743–32749.

[343] A. Suraweera, Y. Lim, R. Woods, G. W. Birrell, T. Nasim, O. J. Becherel, M. F. Lavin, *Human molecular genetics* **2009**, *18*, 3384–3396.

[344] M. B. Klinger, B. Guilbault, R. J. Kay, *European journal of immunology* **2004**, *34*, 806–816.

[345] S. Nissim, I. Leshchiner, J. D. Mancias, M. B. Greenblatt, O. Maertens, C. A. Cassa, J. A. Rosenfeld, A. G. Cox, J. Hedgepeth, J. I. Wucherpfennig et al., *Nat Genet* **2019**, *51*, 1308–1314.

[346] K. M. Sakamoto, K. B. Kim, A. Kumagai, F. Mercurio, C. M. Crews, R. J. Deshaies, *Proceedings of the National Academy of Sciences* **2001**, *98*, 8554–8559.

[347] M. Choi, Y. M. Choi, I.-S. An, S. Bae, J. H. Jung, S. An, *Biochemical and biophysical research communications* **2020**, *521*, 37–41.

[348] a) D. Huang, L. Cao, S. Zheng, *Journal of experimental & clinical cancer research : CR* **2017**, *36*, 13; b) M. W. Kilimann, G. Isenberg, *The EMBO Journal* **1982**, *1*, 889–894;

[349] L. Truebestein, D. J. Elsner, E. Fuchs, T. A. Leonard, *Nature communications* **2015**, *6*, 10029.

[350] M. Florio, M. Albert, E. Taverna, T. Namba, H. Brandl, E. Lewitus, C. Haffner, A. Sykes, F. K. Wong, J. Peters et al., *Science (New York, N.Y.)* **2015**, *347*, 1465–1470.

[351] H. Santos-Rosa, R. Schneider, A. J. Bannister, J. Sherriff, B. E. Bernstein, N. C. T. Emre, S. L. Schreiber, J. Mellor, T. Kouzarides, *Nature* **2002**, *419*, 407–411.

[352] J. F. Flanagan, L.-Z. Mi, M. Chruszcz, M. Cymborowski, K. L. Clines, Y. Kim, W. Minor, F. Rastinejad, S. Khorasanizadeh, *Nature* **2005**, *438*, 1181–1185.

[353] A. Gaspar-Maia, A. Alajem, F. Polesso, R. Sridharan, M. J. Mason, A. Heidersbach, J. Ramalho-Santos, M. T. McManus, K. Plath, E. Meshorer et al., *Nature* **2009**, *460*, 863–868.

[354] P. Zhu, W. Zhou, J. Wang, J. Puc, K. A. Ohgi, H. Erdjument-Bromage, P. Tempst, C. K. Glass, M. G. Rosenfeld, *Molecular cell* **2007**, *27*, 609–621.

[355] M. Hong, J.-T. Hwang, E. J. Shin, H. J. Hur, K. Kang, H.-K. Choi, M.-Y. Chung, S. Chung, M. J. Sung, J.-H. Park, *PLoS ONE* **2019**, *14*.

[356] T. Z. Parris, A. Kovács, S. Hajizadeh, S. Nemes, M. Semaan, M. Levin, P. Karlsson, K. Helou, *Oncogenesis* **2014**, *3*, e95-.

[357] J. A. Gross, L. M. Fiori, B. Labonté, J. P. Lopez, G. Turecki, *Journal of psychiatric research* **2013**, *47*, 513–519.

[358] X. Tan, R. Liu, S. Xing, Y. Zhang, Q. Li, M. Zheng, G. Zhao, J. Wen, *International journal of molecular sciences* **2020**, *21*.

[359] A. Fu, D. I. Jacobs, A. E. Hoffman, T. Zheng, Y. Zhu, *Carcinogenesis* **2015**, *36*, 1094–1102.

[360] M. T. Bjerre, S. H. Strand, M. Nørgaard, H. Kristensen, A. K. I. Rasmussen, M. M. Mortensen, J. Fredsøe, P. Mouritzen, B. Ulhøi, T. Ørntoft et al., *International journal of molecular sciences* **2019**, *20*.

[361] D. F. Calvisi, S. Ladu, A. Gorden, M. Farina, J.-S. Lee, E. A. Conner, I. Schroeder, V. M. Factor, S. S. Thorgeirsson, *The Journal of clinical investigation* **2007**, *117*, 2713–2722.

[362] B. J. Toghill, A. Saratzis, P. J. Freeman, N. Sylvius, M. J. Bown, *Clinical epigenetics* **2018**, *10*, 29.

[363] K. Zglejc-Waszak, E. M. Waszkiewicz, A. Franczak, *Theriogenology* **2019**, *123*, 185–193.

[364] F. He, D. S. Lupu, M. D. Niculescu, *International journal of developmental neuroscience : the official journal of the International Society for Developmental Neuroscience* **2014**, *36*, 38–44.

[365] A. Shire, G. Lomberk, J.-P. Lai, H. Zou, N. Tsuchiya, I. Aderca, C. D. Moser, K. H. Gulaid, A. Oseini, C. Hu et al., *Medical epigenetics* **2015**, *3*, 1–18.

[366] J. Cox, M. Mann, *Nature biotechnology* **2008**, *26*, 1367–1372.

[367] S. Palei, B. Buchmuller, J. Wolffgramm, Á. Muñoz-Lopez, S. Jung, P. Czodrowski, D. Summerer, *Journal of the American Chemical Society* **2020**, *142*, 7289–7294.

[368] A. Hauk, *Biologie in unserer Zeit* **2013**, *43*, 278.

[369] D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison, H. O. Smith, *Nature methods* **2009**, *6*, 343–345.

[370] a) H. Hayatsu, Y. Wataya, K. Kai, S. Iida, *Biochemistry* **1970**, *9*, 2858–2865; b) A. Chatterjee, P. A. Stockwell, E. J. Rodger, I. M. Morison, *Nucleic acids research* **2012**, *40*, e79;

[371] R. Shapiro, V. DiFate, M. Welcher, *J. Am. Chem. Soc.* **1974**, *96*, 906–912.

[372] R. Y. Wang, C. W. Gehrke, M. Ehrlich, *Nucleic acids research* **1980**, *8*, 4777–4790.

[373] a) D. Millar, Y. Christova, P. Holliger, *Nucleic acids research* **2015**, *43*, e155; b) M. Frommer, L. E. McDonald, D. S. Millar, C. M. Collis, F. Watt, G. W. Grigg, P. L. Molloy, C. L. Paul, *Proceedings of the National Academy of Sciences* **1992**, *89*, 1827–1831;

[374] a) S. Lee, J. Kim, *MethodsX* **2016**, *3*, 1–7; b) C. Liu, X. Cui, B. S. Zhao, P. Narkhede, Y. Gao, J. Liu, X. Dou, Q. Dai, L.-S. Zhang, C. He, *J. Am. Chem. Soc.* **2020**, *142*, 4539–4543;

[375] M. F. Fraga, M. Esteller, *BioTechniques* **2002**, *33*, 632, 634, 636-49.

[376] P. Boyle, K. Clement, H. Gu, Z. D. Smith, M. Ziller, J. L. Fostel, L. Holmes, J. Meldrim, F. Kelley, A. Gnirke et al., *Genome biology* **2012**, *13*, R92.

[377] J. Wang, Y. Xia, L. Li, D. Gong, Y. Yao, H. Luo, H. Lu, N. Yi, H. Wu, X. Zhang et al., *BMC genomics* **2013**, *14*, 11.

[378] M. Schillebeeckx, A. Schrade, A.-K. Löbs, M. Pihlajoki, D. B. Wilson, R. D. Mitra, *Nucleic acids research* **2013**, *41*, e116.

[379] H. Guo, P. Zhu, X. Wu, X. Li, L. Wen, F. Tang, *Genome research* **2013**, *23*, 2126–2135.

[380] a) J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid et al., *Nature methods* **2012**, *9*, 676–682; b) C. A. Schneider, W. S. Rasband, K. W. Eliceiri, *Nature methods* **2012**, *9*, 671–675;

[381] R Core Team, *R: A Language and Environment for Statistical Computing*, **2019**, Vienna, Austria, to be found under http://www.R-project.org/.

[382] A. S. Matt Dowle, *data.table: Extension of data.frame. R package version 1.12.8*, **2019,** to be found under https://CRAN.R-project.org/package=data.table.

[383] a) H. Wickham, *ggplot2*, Springer International Publishing, Cham, **2016**; b) Baptiste Auguie, *gridExtra: Miscellaneous Functions for "Grid" Graphics*, **2017,** to be found under https://CRAN.R-project.org/package=gridExtra;

[384] Luca Scrucca, Michael Fop, Thomas Brendan Murphy, Adrian E. Raftery, *The R Journal* **2016**, *8*, 205–233.

[385] A. P. Masella, A. K. Bartram, J. M. Truszkowski, D. G. Brown, J. D. Neufeld, *BMC bioinformatics* **2012**, *13*, 31.

[386] B. Bushnell, "BBMap: A Fast, Accurate, Splice-Aware Aligner", to be found under https://www.sourceforge.net/projects/bbmap/.

[387] M. Lawrence, W. Huber, H. Pagès, P. Aboyoun, M. Carlson, R. Gentleman, M. T. Morgan, V. J. Carey, *PLoS computational biology* **2013**, *9*, e1003118.

[388] T. Buschmann, L. V. Bystrykh, *BMC bioinformatics* **2013**, *14*, 272.

[389] A. Frankish, M. Diekhans, A.-M. Ferreira, R. Johnson, I. Jungreis, J. Loveland, J. M. Mudge, C. Sisu, J. Wright, J. Armstrong et al., *Nucleic acids research* **2019**, *47*, D766-D773.

[390] R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, C. Kingsford, *Nature methods* **2017**, *14*, 417–419.

[391] A. Srivastava, L. Malik, H. Sarkar, M. Zakeri, F. Almodaresi, C. Soneson, M. I. Love, C. Kingsford, R. Patro, *Alignment and mapping methodology influence transcript abundance estimation*, **2019**.

[392] A. Roberts, C. Trapnell, J. Donaghey, J. L. Rinn, L. Pachter, *Genome biology* **2011**, *12*, R22.

[393] a) M. I. Love, C. Soneson, P. F. Hickey, L. K. Johnson, N. T. Pierce, L. Shepherd, M. Morgan, R. Patro, *Tximeta: reference sequence checksums for provenance identification in RNA-seq*, **2019**; b) Hervé Pagès, Marc Carlson, Seth Falcon, Nianhua Li, AnnotationDbi: Manipulation of SQLite-based annotations in Bioconductor. R package version 1.48.0, Bioconductor, **2019**;

[394] M. I. Love, W. Huber, S. Anders, *Genome biology* **2014**, *15*, 550.

[395] M. Stephens, *Biostatistics (Oxford, England)* **2017**, *18*, 275–294.

[396] Matthew Stephens, Peter Carbonetto, David Gerard, Mengyin Lu, Lei Sun, Jason Willwerscheid, Nan Xiao, *ashr: Methods for Adaptive Shrinkage, using Empirical Bayes. R package version 2.2-47.*, **2020,** to be found under https://cran.r-project.org/package=ashr.

[397] H. Pedersen, S. Hölder, D. P. Sutherlin, U. Schwitter, D. S. King, P. G. Schultz, *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95*, 10523–10528.

[398] L. Shen, C.-X. Song, C. He, Y. Zhang, *Annu. Rev. Biochem.* **2014**, *83*, 585–614.

[399] C. Krebs, D. Galonić Fujimori, C. T. Walsh, J. M. Bollinger, *Accounts of Chemical Research* **2007**, *40*, 484–492.

Eidesstattliche Versicherung (Affidavit)

Wolffgramm, Jan
Name, Vorname
(Surname, first name)

Matrikel-Nr.
(Enrolment number)

Belehrung:

Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50.000,00 € geahndet werden. Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden, § 63 Abs. 5 Hochschulgesetz NRW.

Die Abgabe einer falschen Versicherung an Eides statt ist strafbar.

Wer vorsätzlich eine falsche Versicherung an Eides statt abgibt, kann mit einer Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft werden, § 156 StGB. Die fahrlässige Abgabe einer falschen Versicherung an Eides statt kann mit einer Freiheitsstrafe bis zu einem Jahr oder Geldstrafe bestraft werden, § 161 StGB.

Die oben stehende Belehrung habe ich zur Kenntnis genommen:

Official notification:

Any person who intentionally breaches any regulation of university examination regulations relating to deception in examination performance is acting improperly. This offence can be punished with a fine of up to EUR 50,000.00. The competent administrative authority for the pursuit and prosecution of offences of this type is the chancellor of the TU Dortmund University. In the case of multiple or other serious attempts at deception, the candidate can also be unenrolled, Section 63, paragraph 5 of the Universities Act of North Rhine-Westphalia.

The submission of a false affidavit is punishable.

Any person who intentionally submits a false affidavit can be punished with a prison sentence of up to three years or a fine, Section 156 of the Criminal Code. The negligent submission of a false affidavit can be punished with a prison sentence of up to one year or a fine, Section 161 of the Criminal Code.

I have taken note of the above official notification.

Dortmund,
Ort, Datum
(Place, date)

Unterschrift
(Signature)

Titel der Dissertation:
(Title of the thesis):

Light-Activation of DNA-Methyltransferases

Ich versichere hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel selbstständig und ohne unzulässige fremde Hilfe angefertigt habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht.
Die Arbeit hat in gegenwärtiger oder in einer anderen Fassung weder der TU Dortmund noch einer anderen Hochschule im Zusammenhang mit einer staatlichen oder akademischen Prüfung vorgelegen.

I hereby swear that I have completed the present dissertation independently and without inadmissible external support. I have not used any sources or tools other than those indicated and have identified literal and analogous quotations.

The thesis in its current version or another version has not been presented to the TU Dortmund University or another university in connection with a state or academic examination.*

*Please be aware that solely the German version of the affidavit ("Eidesstattliche Versicherung") for the PhD thesis is the official and legally binding version.

Dortmund,
Ort, Datum
(Place, date)

Unterschrift
(Signature)