Imaging-Based Analysis of 5-Methylcytosine at Low-Repetitive Genomic Loci Using Transcription Activator-Like Effector Probes

Dissertation

Submitted for the degree of Doctor of Natural Sciences

(Dr. rer. nat.)

to the

Faculty of Chemistry and Chemical Biology

Technical University Dortmund

by

Anne Jung

born in Heidenheim an der Brenz

Dortmund 2023

“Science cannot solve the ultimate mystery of nature. And that is because, in the last analysis, we ourselves are a part of the mystery that we are trying to solve.”

– Max Planck

This work was prepared from August 2018 to July 2022 in the group of Prof. Dr. Daniel Summerer of the Faculty of Chemistry and Chemical Biology at the Technical University Dortmund.

The research project presented in this work was funded by the European Research Council (ERC) under the ERC consolidator grant EPICODE No. 723863 to Prof. Dr. Daniel Summerer.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to Prof. Dr. Daniel Summerer for giving me the opportunity to perform my PhD study in his research group. I am truly thankful for his trust and continuous support throughout my PhD journey. His excellent supervision and invaluable scientific guidance have always been a great source of motivation and enabled me to achieve my goals. I am grateful for the privilege of working with such an outstanding mentor.

My gratitude extends to Prof. Dr. Daniel Rauh for his time and dedication as the second examiner of my dissertation.

Furthermore, I would like to thank my students Corinna Tiggemann (M. Sc.), Cristina West (B. Sc.), Justin Kern (B. Sc.) and Pia Beer (B. Sc.) for their enthusiastic and fruitful work under my supervision. They have contributed significantly to the success of the project, and I am very grateful for their efforts and commitment.

Being part of a great team is not only a pleasure, but also an invaluable source of motivation and inspiration. I would like to thank all my friends and colleagues who made my time at the Summerer group such a memorable and enjoyable experience. My special thanks go to Dr. Álvaro Muñoz López, Dr. Anna Witte, Dr. Benjamin Buchmuller, Brinja Kosel, Carolin Gierse, Damian Schiller, Dominic Kamps, Dr. Jan Wolffgramm, Katrin Bigler, Lena Engelhard, Dr. Mario Gieß, Dr. Preeti Rathi, Ricarda Lüttig, Dr. Sara Maurer, Dr. Shubhendu Palei, Simone Eppmann, Suchet Nanda, Sudakshina Banerjee, and Dr. Tzu Chen Lin. I am truly grateful that I had the opportunity to work with such an amazing group of people and thank them all for the wonderful time we spent together in the lab and social activities.

Finally, I would like to express my deepest gratitude to my parents, brothers, and the rest of my family for the constant support and encouragement I have received from them throughout my studies. I would especially like to thank my boyfriend Álvaro, without whom I would not have been able to reach this far. I cannot thank him enough for the time and effort he spent helping me with my project. He has always been there for me, and I am forever grateful for his love and support.

Publications

Part of this work has been published in:

Jung A., Muñoz-López Á., Buchmuller B. C., Banerjee S., Summerer D. (2023). Imaging-Based In Situ Analysis of 5-Methylcytosine at Low Repetitive Single Gene Loci with Transcription-Activator-Like Effector Probes. *ACS Chemical Biology*, 18(2), 230-236.

Other publications:

Buchmuller B., Jung A., Muñoz-López Á., Summerer D. (2021). Programmable tools for targeted analysis of epigenetic DNA modifications. *Current Opinion in Chemical Biology*, 63, 1-10.

Muñoz-López Á., Jung A., Buchmuller B., Wolffgramm J., Maurer S., Witte A., Summerer D. (2020). Engineered TALE Repeats for Enhanced Imaging-based Analysis of Cellular 5-Methylcytosine. *ChemBioChem*, 22(4), 645-651.

Muñoz-López Á., Buchmuller B., Wolffgramm J., Jung A., Hussong M., Kanne J., Schweiger M. R. Summerer D. (2020). Designer Receptors for Nucleotide Resolution Analysis of Genomic 5-Methylcytosine by Cellular Imaging. *Angewandte Chemie International Edition*, 59(23), 8927-8931.

Table of Contents

List of Figures

List of Tables

Abbreviations

°C	Degree celsius
µL	Microliter
µm	Micrometer
5-aza-dC	5-aza-2′-deoxycytidine
5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
A	Adenine
Ab	Antibody
Ac	Acetylation
AD	Acidic activation domain
Ala	Alanine
AMOP	Nidogen (NIDO)-like, adhesion-associated domain
APOBEC	Apolipoprotein B mRNA editing catalytic polypeptide
Asn	Asparagine
au	Arbitrary units
BER	Base excision repair
BiFC	Bimolecular fluorescence complementation
bp	Base pair
BSA	Bovine serum albumin
BS-seq	Bisulfite sequencing
C	Cytosine
C4	Carbon-4 position of cytosine base
C5	Carbon-5 position of cytosine base
Carb	Carbenicillin
CARGO	Chimeric array of gRNA oligonucleotides
Cas9	CRISPR-associated protein 9
CpG	Cytosine-phosphate-guanine dinucleotide
CRD	Central repeat domain
Cre	P1 bacteriophage cyclization recombination recombinase
CRISPR	Clustered regularly interspaced short palindromic repeat
CT	C-terminal
CTR	C-terminal region
Cy2	Cyanine 2
Cy3	Cyanine 3
Cy5	Cyanine 5
Cy7	Cyanine 7
Cys	Cysteine
DAPI	4',6-diamidino-2-phenylindole
dCas9	Catalytically dead CRISPR-associated protein 9

DddA	Double-stranded DNA deaminase toxin A
DHU	Dihydrouracil
DKO	Double-knockouts
DMEM	Dulbecco's modified eagle medium
DMR	Differentially methylated region
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DNMT1	DNA methyltransferase 1
DNMT3A	DNA methyltransferase 3A
DNMT3a3L	Dnmt3a-Dnmt3L single-chain fusion protein protein
DNMT3B	DNA methyltransferase 3B
DNMT3L	DNA methyltransferase 3-like
dNTP	Deoxyribonucleoside triphosphate
DTT	Dithiothreitol
EBFP2	Enhanced blue fluorescent protein 2
EDTA	Ethylenediaminetetraacetic acid
eGFP	Enhanced green fluorescent protein
EM-seq	Enzymatic methyl sequencing
FBS	Fetal bovine serum
FISH	Fluorescence in situ hybridization
FITC	Fluorescein isothiocyanate
Flp	Flippase recombinase
FP	Fluorescent protein
FRACTAL	Fluorescent signal amplification via cyclic staining of target molecules
FS	Phenylalanine 12 + Serine 13 repeat variable diresidue
G	Guanine
g	Gram
G*	Glycine 12 + deletion at position 13 repeat variable diresidue
GAPDH	Glycerinaldehyd-3-phosphat-dehydrogenase
GCN4	General control noninducible 4
GFP	Green fluorescent protein
GG1/2	Golden Gate 1/2
GG	Glycine-Glycine diresidue
Gin	G-segment invertase
Gln	Glutamine
Gly	Glycine
gRNA	Guide RNA
h	Hour(s)
H1	Histone 1
H2A	Histone 2A
H2B	Histone 2B
H3	Histone 3
H4	Histone 4
HA	Histidine 12 + Alanine 13 repeat variable diresidue

HA	Influenza hemagglutinin
HCT116	Human colorectal carcinoma cell line
HD	Histidine 12 + Aspartic acid 13 repeat variable diresidue
HDAC	Histone deacetylase
His	Histidine
HR	Homologous recombination
HRP	Horseradish peroxidase
HSF1	Heat shock factor 1
ICON	Bipyridine-containing nucleic acid
k	Thousand
kb	Kilo–base-pair
kDa	Kilodaltons
KO	Knockout
KQ	Lysine-Glutamine diresidue
KRAB	Krüppel associated box
Lac	Lactose
LB	Lysogeny broth
L-Glu	L-Glutamine
log	Logarithmic
Lys	Lysine
MB	Molecular beacon
MBD	Methyl-CpG-binding domain
MBP	DNA methyl-binding protein
MCP	MS2 capsid protein
Me	Methylation
MeCP2	Methyl CpG binding protein 2
MeDIP	Methylated-DNA-immunoprecipitation
MeFISH	Methylation-specific fluorescence *in situ* hybridization
min	Minute(s)
mL	Milliliter
mM	Millimolar
MS2	*Emesvirus zinderi*
MUC4	Mucin 4, cell surface associated
MWCO	Molecular weight cutoff
Myc	Myelocytomatosis viral oncogene
N*	Asparagine 12 + deletion at position 13 repeat variable diresidue
NG	Asparagine 12 + Glycine 13 repeat variable diresidue
NHEJ	Non-homologous end joining
NI	Asparagine 12 + Isoleucine 13 repeat variable diresidue
NLS	Nuclear localization signal
nm	Nanometer
nM	Nanomolar
NN	Asparagine 12 + Asparagine 13 repeat variable diresidue
nt	Nucleotides
NT	N-terminal

NTR	N-terminal region
OD600	Optical density
P	Phosphorylation
P*	Proline 12 + deletion at position 13 repeat variable diresidue
p300	Histone acetyltransferase p300
PBS	PUF-binding site
PCP	PP7 coat protein
PCR	Polymerase chain reaction
Pen/Strep	Penicillin/Streptomycin
PGC	Primordial germ cell
Ph	Phosphorylation
PMSF	Phenylmethylsulfonyl fluoride
PP7	*Pseudomonas aeroginosa*
PUF	Pumilio/Fem3 RNA-binding factor
qPCR	Real-time quantitative PCR
R*	Arginine 12 + deletion at position 13 repeat variable diresidue
R****	Arginine 12 + deletion of four amino acids
RCA	Rolling circle amplification
rDNA	Ribosomal DNA
RFP	Red fluorescent protein
RG	Arginine 12 + Gylcine 13 repeat variable diresidue
RNA	Ribonucleic acid
rpm	Revolutions per minute
RPMI 1640	Roswell park memorial institute 1640 medium
RT	Room temperature
RVD	Repeat variable diresidue
S*	Serine 12 + deletion at position 13 repeat variable diresidue
SAH	*S*-adenosylhomocysteine
SAM	*S*-adenyl methionine
SatIII	Satellite 3
SBR	Signal-to-background ratio
scFv	Single chain variable fragment
SD	Standard deviation
sec	Second(s)
sfGFP	Split super-folder GFP
sgRNA	Single guide RNA
SID	mSin interaction domain
SMRT	Single molecule real-time
SNR	Signal-to-noise ratio
SunTag	Supernova tagging system
T	Thymine
T*	Threonine 12 + deletion at position 13 repeat variable diresidue
TALE	Transcription activator-like effector
TALEN	TALE-nuclease
TAPS	TET-assisted pyridine borane sequencing

TDG	Thymine DNA glycosylase
TET	Ten-eleven translocation dioxygenase
Tet operator	Tetracycline operator
TET1	Ten-eleven translocation dioxygenase 1
TM	Transmembrane region
Tris	Tris(hydroxymethyl)aminomethan
TSA	Tyramide signal amplification
TSH2B	Histone cluster 1 H2B family member
U	Uracil
U2OS	Human bone osteosarcoma epithelial
Ub	Ubiquitination
UHRF1	Ubiquitin like with PHD and ring finger domains 1
UTR	5'-untranslated region
V5	Simian virus 5-derived epitope
V_H	Heavy immunoglobulin chains
V_L	Light immunoglobulin chains
VP16	Herpes simplex viral protein 16
VP64	Tetrameric derivative of herpes simplex viral protein 16
vWD	Willebrand factor
w/o	Without
wt	Wild type
ZFP	Zinc finger protein

1. Abstract

5-Methylcytosine (5mC) is the main epigenetic modification of mammalian genomes. It plays significant roles during cell development and differentiation and is involved in the regulation of essential cellular processes such as the control of gene expression. Dysregulation of methylation can lead to aberrant epigenetic patterns associated with a variety of diseases. To analyze cellular 5mC *in situ*, fluorescently labeled transcription-activator-like effector (TALE) proteins can be used as 5mC-sensitive probes in imaging studies. TALEs are DNA-binding proteins that provide sequence and 5mC selectivity via a domain of modular repeats, each recognizing a specific nucleobase. This enables the design of TALE probes for sequence-specific analysis of 5mC in user-defined target sequences. In imaging studies, 5mC-sensitive and 5mC-insensitive TALE pairs are used in co-stainings to allow the analysis of 5mC independently of changes in target accessibility. However, until now this has been limited to highly repetitive genomic DNA sequences. To extend this approach for the analysis of 5mC in low-repetitive coding gene loci, this work develops a straightforward signal amplification strategy to increase the imaging sensitivity with TALEs. This is achieved by additional immunostaining of the employed TALE probes, enabling the imaging of only 32 theoretical repeat sequences in the low repetitive MUC4 gene locus. In co-staining experiments, this allows the detection of 5mC changes in this locus between cell types with different methylation levels, introduced by DNA methyltransferase knockouts or overexpression. The ability to detect 5mC differences in this small number of target sequences opens up new perspectives for the analysis of 5mC in non-repetitive genomic loci, providing new insights into the regulation of gene expression.

1. Zusammenfassung

5-Methylcytosin (5mC) ist die wichtigste epigenetische Modifikation des Säugetiergenoms. Es spielt eine bedeutende Rolle während der Zellentwicklung und -differenzierung und ist an der Regulation wesentlicher zellulärer Prozesse wie der Kontrolle der Genexpression beteiligt. Dysregulation von Methylierung kann zu abweichenden epigenetischen Mustern führen, die mit einer Vielzahl von Krankheiten in Verbindung gebracht werden. Um zelluläres 5mC *in situ* zu analysieren, können fluoreszenzmarkierte „Transcription Activator Like-Effector" (TALE) Proteine als 5mC-empfindliche Sonden in Bildgebungsstudien verwendet werden. TALEs sind DNA-bindende Proteine, die Sequenz- und 5mC-Selektivität über eine Domäne modularer Wiederholungen bereitstellen, die jeweils eine spezifische Nukleinbase erkennen. Dies ermöglicht das Design von TALE-Sonden für die sequenzspezifische Analyse von 5mC in benutzerdefinierten Zielsequenzen. In Bildgebungsstudien werden 5mC-empfindliche und 5mC unempfindliche TALE-Paare in Co-Färbungen verwendet, um die Analyse von 5mC unabhängig von Änderungen in der Zugänglichkeit zu ermöglichen. Bisher war dies jedoch auf stark repetitive genomische DNA-Sequenzen beschränkt. Um diesen Ansatz für die Analyse von 5mC in gering repetitiven kodierenden Genloci zu erweitern, entwickelt diese Arbeit eine einfache Signalamplifikationsstrategie, um die Bildgebungsempfindlichkeit mit TALEs zu erhöhen. Dies wird durch eine zusätzliche Immunfärbung der verwendeten TALE-Sonden erreicht, die die Detektion von nur 32 theoretischen Wiederholungssequenzen im gering repetitiven MUC4-Genlokus ermöglicht. In Co-Färbungsexperimenten ermöglicht dies die Detektion von 5mC-Veränderungen in diesem Lokus zwischen Zelltypen mit unterschiedlichen Methylierungsgraden, die entweder durch DNA-Methyltransferase-Knockouts oder Überexpression eingeführt wurden. Die Fähigkeit 5mC-Unterschiede in dieser geringen Anzahl an Zielsequenzen nachzuweisen zu können, eröffnet neue Perspektiven für die Analyse von 5mC in nicht repetitiven genomischen Loci, was neue Einblicke in die Regulation der Genexpression ermöglicht.

2. Introduction

2.1 The Structure and Function of DNA

Life as we know it only exists because of the capability to store and process genetic information. The complete set of genetic information, the so-called genome, provides all the information required by an organism to function[1]. It encodes for all proteins within each cell of an organism, defines their shape and is responsible for the regulation of cellular processes. In all living organisms, genetic information is encoded by deoxyribonucleic acid (DNA) molecules and is passed from generation to generation to enable development, maintenance and reproduction[1,2]. DNA molecules are made up of long polynucleotide strands composed of four different nucleotide monomers whose specific order along the DNA strand encodes the genetic information[3]. Each nucleotide monomer contains the five-carbon sugar deoxyribose, attached to a phosphate group and a nitrogen-containing nucleobase, which can be either adenine (A), guanine (G), cytosine (C) or thymine (T) (Figure 1a)[4]. The nucleotides within a DNA strand are covalently linked by phosphodiester bonds between the 5'-phosphate and 3'-OH groups of adjacent nucleotides, forming a sugar-phosphate backbone from which the nucleobases extend (Figure 1b). Due to the asymmetry of the sugar-phosphate backbone, each DNA strand has a 3'-hydroxyl and a 5'-phosphate end, giving the strand a chemical polarity[2].

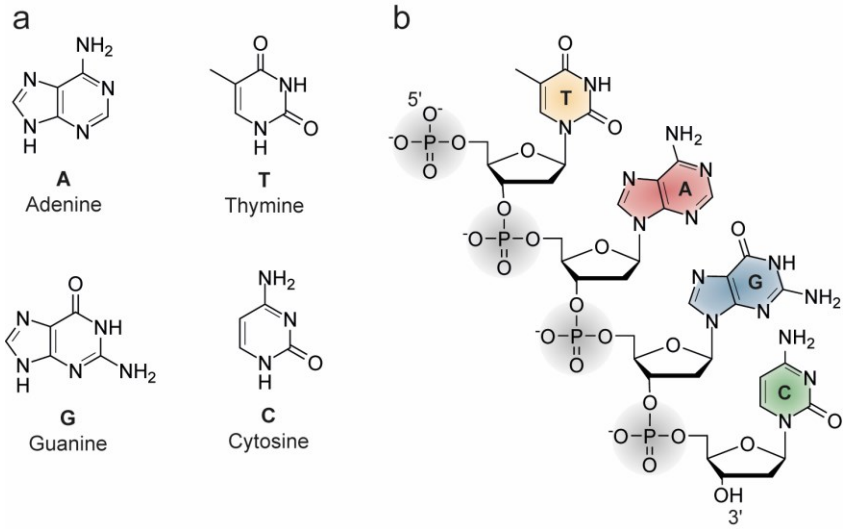


Figure 1. Chemical structures of nucleobases and DNA strand. a) Purine nucleobases adenine (A) and guanine (G) and pyrimidine nucleobases thymine (T) and cytosine (C). b) Nucleotides are covalently linked to each other by phosphodiester bonds.

In the naturally occurring double-stranded form of DNA, two anti-parallel DNA strands are paired such that the 5'-end of one strand is paired with the 3'-end of its complementary strand (Figure 2). The two polynucleotide strands wind around each other to form a double-helix in which the bases point towards the inside of the double-helix and the sugar-phosphate backbones are on the outside[5]. The helical structure of DNA is variable and depends on the sequence. Depending on the number of bases per turn and the twist angle, three main helical DNA structures can be distinguished, A-DNA, B-DNA and Z-DNA. A- and B-DNA are right-handed helices, while Z-DNA has a left-handed orientation of the helix[6]. In the most common B-form of DNA, the bases are regularly spaced at 0.34 nm apart from each other along the axis of the helix[1]. A complete helical turn includes 10.4 base pairs (bp) and is made every 3.4 nm[2]. Due to the coiling of the DNA strands around each other, two unequally sized grooves are formed which run along the entire length of the DNA double-helix. In the smaller minor groove, the space between the two antiparallel DNA strands is 1.2 nm wide, whereas it is 2.2 nm wide in the major groove[7].

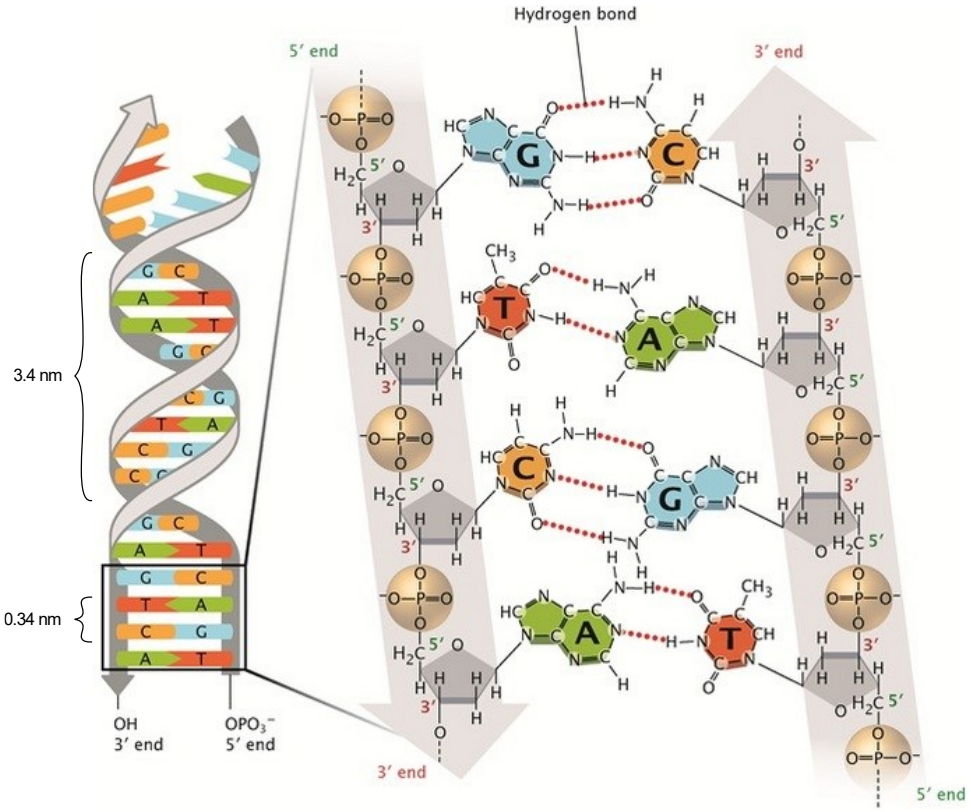


Figure 2. Double-helical structure and base pairing of DNA. Complementary bases A and T are connected by two, and complementary bases C and G by three hydrogen bonds. Figure adapted from (Pray, 2008) with permission (8.3).

The two nucleotide strands of the DNA double-helix are connected via hydrogen bonds between the purine nucleobases (A and G) and pyrimidine nucleobases (T and C)[2,8]. According to the Watson-Crick base-pairing, two hydrogen bonds are formed between A and T and three hydrogen bonds between G and C[9]. As a consequence of this base-pairing, the sequence of each strand of a DNA molecule is exactly complementary to the nucleotide sequence of its partner strand. This specific pairing provides the basis for the ability of cells to replicate its DNA during cellular reproduction, as each strand can serve as a template for the synthesis of a new complementary strand[2].

2.1.1 DNA Organization within the Nucleus

Each human somatic cell is diploid and contains two sets of homologous chromosomes, one set inherited from the mother and one from the father. Human somatic cells have 23 pairs of chromosomes which makes a total of 46 chromosomes[10,11]. Twenty-two of these pairs, called autosomes carry the genetic information for everything except the sex determination, which is determined by the 23rd chromosome pair and differs between males and females. Females have two copies of the X chromosome while males have one X and one Y chromosome[12].

On average, a human somatic cell contains about 6 billion nucleotide pairs with a total length of about two meters of DNA [13]. Most of a cell's DNA is located in the nucleus, a membrane-bound organelle with an average diameter of 10 μm. In order to fit this large amount of DNA into the small nucleus, the DNA must be tightly organized and condensed. This is done by a process of winding and folding during which the DNA is packaged into physically separate chromosomes made up of single, very long, linear DNA molecules. To establish DNA compaction, DNA-histone protein complexes are formed, known as chromatin[14]. The basic repeating unit of chromatin is the nucleosome, which consists of 146 bp of DNA wrapped approximately 1.65 times around a histone protein octamer[15]. The histone octamer is composed of two copies of the histone proteins H2A, H2B, H3 and H4. The central domains of these histone proteins are structured and form the globular part of the nucleosome, while the N-terminal domains of all four histone proteins and C-terminal domains of H2A and H2B, the so-called histone tails are poorly structured and protrude from the nucleosome[16]. To establish a tight interaction between the histone core and the DNA molecule, positively charged

lysine and arginine residues of histone proteins interact with the negatively charged phosphate backbone of the DNA[17]. Each nucleosome core is connected to the adjacent one through a segment of internucleosomal DNA, which is often stabilized by the linker histone proteins H1 or H5. Nucleosomes can self-assemble to form 30 nm chromatin fibers which then form loops of about 300 nm length. These loops are compressed and folded to a 250 nm wide fiber, which is tightly coiled to form the chromatid of a chromosome (Figure 3)[18].

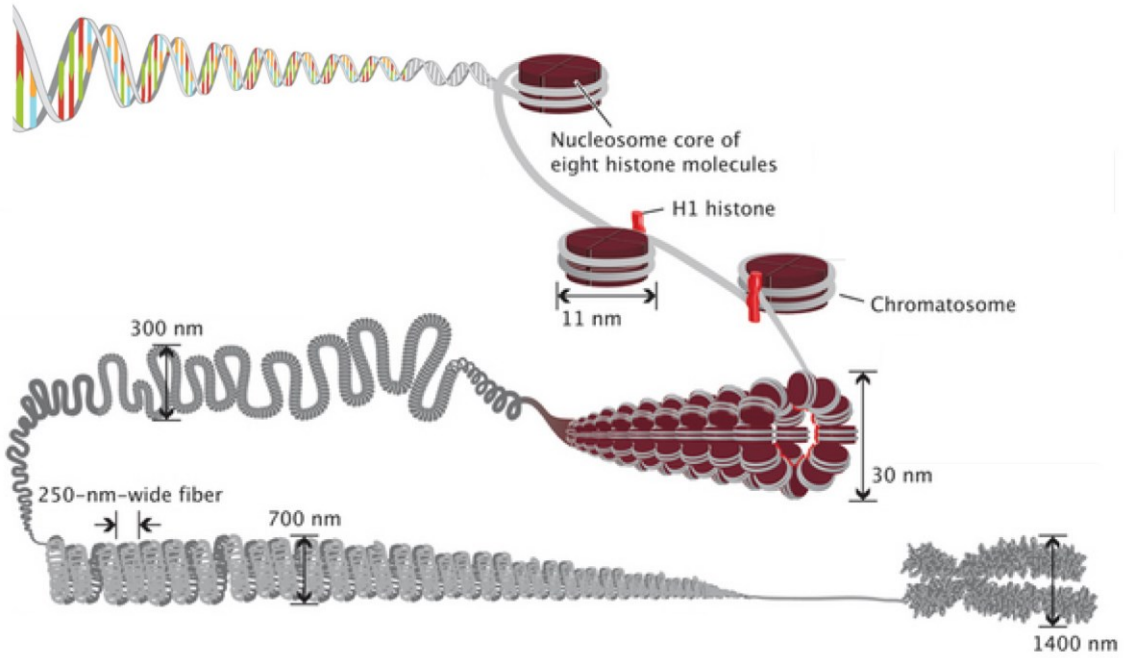


Figure 3. Compaction of DNA. DNA is wrapped around histone proteins which are tightly folded to form a chromosome. Modified from (Pierce, 2013) (8.3).

The compaction of DNA by nucleosomes has the advantage that large parts of the DNA can be present in a highly compacted form to preserve its integrity, while other segments of the genome can be easily accessed for essential functions. Heterochromatin, the highly condensed form of chromatin is typically gene-poor and transcriptionally repressed, whereas euchromatin is less condensed, gene-rich and more accessible to transcription[19]. Pericentromeric and telomeric regions that contain a high density of repetitive DNA elements, such as clusters of satellite sequences and transposons, are mainly present as heterochromatin[20]. By contrast, specific genomic segments related with active gene expression are often found as euchromatin.

2.1.2 Epigenetic Modifications Regulate Chromatin Structure and Gene Expression

Heterochromatin and euchromatin can typically be distinguished by certain histone and DNA modifications. Common histone modifications include acetylation, methylation, phosphorylation and ubiquitination[21]. Most histone modifications occur at the N-terminal tails of histones, with some exceptions including ubiquitination of the C-terminal tails of H2A and H2B and acetylation and methylation of the globular domain of H3 at K56 and K79, respectively[22] (Figure 4). Acetylation mainly occurs at lysine residues located in the amino-terminal domains of histones H3 and H4 and results in neutralization of the positively charged lysine residues. This weakens the histone:DNA interactions and leads to a more open chromatin environment to facilitate transcription factor binding and gene expression[23]. By contrast, histone methylation does not alter the histone charge or influence histone:DNA interactions but it can have different impacts on transcription depending on the modified amino acids. Methyl groups can be added to lysine, arginine and histidine residues of histones. Lysines can be mono-, di-, or tri-methylated on their ε amine group, whereas arginines can be mono-, symmetrically di- or asymmetrically di-methylated on their guanidinyl group. While arginine methylation typically promotes transcriptional activation, lysine methylation can influence both transcriptional activation and repression, depending on the methylation site. By contrast, histidine methylation appears to be rare and monomethylated[24]. Histone phosphorylation occurs on serine, threonine and tyrosine residues. Various effector proteins contain phospho-binding domains that recognize phosphorylated histones and lead to various downstream effects which influence transcriptional regulation, chromosome condensation, DNA damage repair and apoptosis[25]. Histone ubiquitination plays a central role in DNA damage response and predominantly occurs as mono-ubiquitination on lysine residues of H2A and H2B in the vicinity of DNA double-strand breaks. The less common poly-ubiquitination is usually induced by DNA damage but also targets histone proteins for proteasome mediated degradation[26].

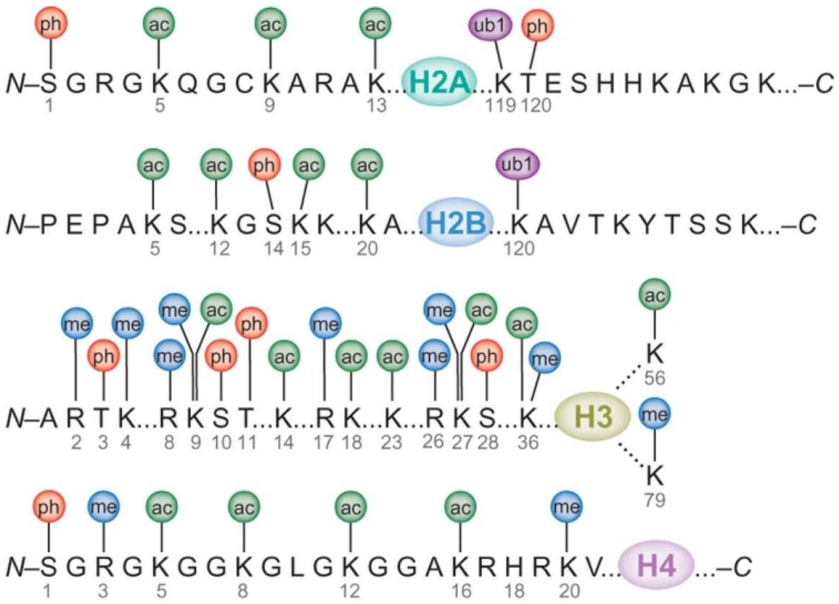


Figure 4. Post-translational modifications of human nucleosomal histones. Acetylation (ac), methylation (me), phosphorylation (ph) and ubiquitination (ub1) at the N- and C-terminal tails of histone proteins H2A, H2B, H3 and H4. Adapted from (Bhaumik et al, 2007) with permission (8.3).

Histone modifications are carefully controlled and are recognized by enzymes which are highly specific for particular amino acid positions. This results in different biological responses depending on the type and location of the modification[27]. Additionally, histone modifications can occur as combinations at one or several histone tails and constitute a so-called histone code, which can be read by reader proteins that contain binding motifs specific for each modification[28,29]. Many chromatin-regulatory proteins and complexes contain multiple histone-binding domains and have in some cases been shown to respond to specific histone modification combinations[30].

Besides histone modifications, modifications of DNA itself can alter the accessibility and transcriptional activity of specific DNA segments. The most common DNA modification is the methylation of cytosine residues in cytosine guanine dinucleotides (CpG)[31]. CpG methylation is typically associated with transcriptional repression and is essential for the regulation of gene expression, silencing of specific genomic elements, genomic imprinting and X chromosome inactivation[32]. Importantly, depending on the underlying genetic sequence, DNA methylation may exert different influences on gene activities[32]. For example, methylation sites in promoter or other regulatory regions mainly mediate transcriptional repression[33]. This is either caused by direct interference of the methyl groups with the binding of transcription factors to their target sites, or alternatively by

the recruitment of methyl-binding factors that suppress gene expression[32,34]. By contrast, heavy gene-body methylation can be found in actively transcribed genes[35]. DNA methylation and histone modifications are highly interrelated and mechanistically rely on each other[36]. For example, DNA methyl-binding proteins (MBPs) such as the methyl CpG binding protein 2 (MeCP2) or other family members bind methylated DNA and repress transcription by recruitment of histone-modifying proteins like histone deacetylases (HDAC). Subsequent histone deacetylation promotes chromatin condensation, resulting in further repression of transcription[37]. Due to the complex interplay between DNA methylation and histone modifications, the structure and dynamics of chromatin and thereby gene expression and silencing are highly regulated[38,39].

2.2 DNA Methylation

The epigenetic mechanism of DNA methylation involves the covalent transfer of a methyl group to the cytosine nucleobase to form 5-methylcytosine (5mC) (Figure 5). The reaction is catalyzed by a family of DNA methyltransferase enzymes (DNMTs) that transfer a methyl group from the cofactor *S*-adenyl methionine (SAM) to the C-5 carbon of cytosine[40].

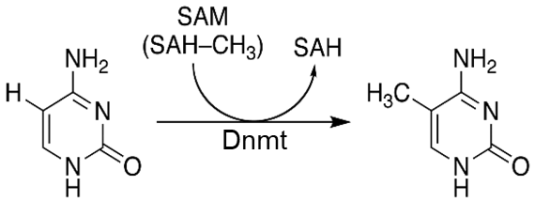


Figure 5. Cytosine methylation catalyzed by DNA methyltransferase (DNMT). DNMT proteins transfer a methyl group from the cofactor *S*-adenosylmethionine (SAM) to the fifth carbon of the cytosine base. SAH: *S*-adenosylhomocysteine.

In mammals, cytosine residues are methylated at levels between 3.5 and 4.5% in differentiated tissues, which accounts for only ~1% of all nucleic acids[41]. However, DNA methylation mainly occurs within the context of palindromic cytosine guanine dinucleotides (CpG). In contrast to other combinations of dinucleotides, CpG sites are depleted in mammalian genomes which can be explained by the mutagenic potential of 5mC to deaminate to thymine[42,43]. CpG sites are distributed throughout the genome

where they show high methylation levels of 70~80% of all CpGs being methylated[44], with the exception of CpG islands. CpG islands are short, interspersed CpG-rich DNA sequences which are predominantly nonmethylated[45]. The majority of annotated gene promoters are associated with CpG islands, making this the most common promoter type for the initiation of transcription[46]. In particular, promoters for housekeeping genes are often imbedded in CpG islands and it appears that CpG islands, associated with promoters are highly conserved between mice and humans, as well as evolutionary[47,48]. CpG islands contain less nucleosomes compared to other stretches of DNA, which are additionally often labelled with histone modifications involved in activation of gene expression[49]. Together, this enhances the accessibility of DNA and promotes transcription factor binding. By contrast, the methylation of CpG islands can result in reduced transcription factor binding, the recruitment of repressive methyl-binding proteins and stable silencing of gene expression[32]. Although methylation predominantly occurs at CpG dinucleotides, about 15% of the total cytosine methylation accounts for non-CpG methylation[50]. The abundancy of non-CpG methylation varies strongly among different cell and tissue types. While it is highly abundant in neurons and oocytes and moderately distributed in brain and embryonic stem cells, it is almost negligible in other cell types[50–52]. Non-CpG methylation exhibits cell-type-specific enrichment patterns and is potentially linked to cell-type-specific gene regulation, but its functional role is still unclear[50,53].

DNA methylation is crucial for normal mammalian development and is involved in a variety of functions in the mammalian genome. While DNA methylation can be a stable and heritable epigenetic mark, methylation and demethylation can also be highly dynamic[54]. Particularly during embryogenesis, mammalian cells undergo a process called epigenetic reprogramming that involves dramatic changes of DNA methylation patterns[55]. Epigenetic reprogramming consists of two main waves that occur initially during germ cell development and later during preimplantation development (Figure 6)[56].

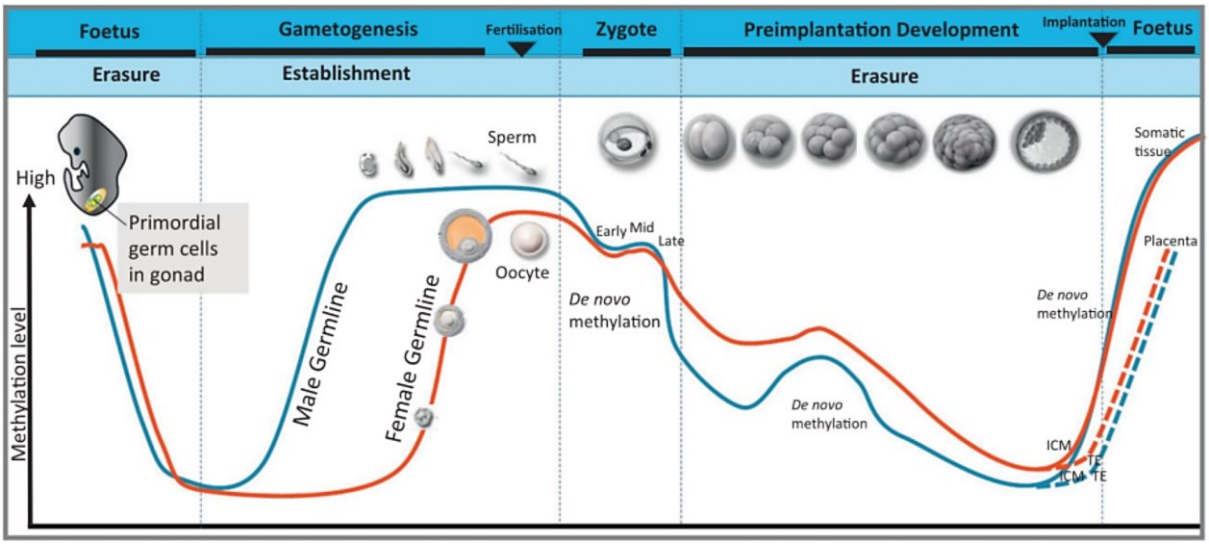


Figure 6. Epigenetic reprogramming during gametogenesis and preimplantation development. Global DNA methylation levels of the male and female germline/genome are represented by a blue line or red line, respectively. Adapted from (Huntriss, 2021) with permission (8.3).

In the first wave during gametogenesis, highly methylated primordial germ cells (PGCs) undergo rapid genome-wide demethylation. This results in erasure and resetting of parent-of-origin specific marks that include methylation of imprinted differentially methylated regions (DMRs) associated with allele-specific gene expression[57]. Subsequent genome-wide *de novo* DNA methylation is first initiated in males and shortly after in females, such that the mature gametes will become highly methylated[56]. The second phase of epigenetic reprogramming occurs between fertilization and formation of the blastocyst (Figure 6). Immediately after fertilization, within 6-8 hours, a rapid paternal-specific active demethylation is observed before onset of DNA replication. By contrast, the maternal genome is passively demethylated during several cleavage divisions[58]. In this process, the methylation level declines in a stepwise fashion due to the absence of the primary DNA methyltransferase DNMT1 during replication[59,60]. The initiation of *de novo* methylation occurs after the fifth cell cycle. This coincides with the first differentiation into two cell lineages that become asymmetrically methylated. The inner cell mass, which gives rise to all tissues of the adult, becomes hypermethylated, while the trophectoderm, that forms most of the structure of the placenta, becomes hypomethylated[56]. Further *de novo* DNA methylation creates new epigenetic information, characteristic for each next phase of development[61]. Ultimately, this leads to fully differentiated cells with a stable and unique DNA methylation pattern that regulates tissue-specific gene transcription[32].

New methylation patterns that are established during *de novo* methylation are carried out by the *de novo* DNA methyltransferases DNMT3A and DNMT3B, which are extremely similar in structure and function[62]. Both methylate DNA with no preference for unmethylated or hemimethylated DNA[63], however DNMT3A prefers CpG and non-CpG sites followed by a 3'-pyrimidine, whereas DNMT3B favors a 3'-purine[64]. Both *de novo* methyltransferases show different gene expression patterns. While DNMT3A is expressed relatively ubiquitously, DNMT3B is expressed at very low levels in most differentiated tissues[65]. Based on DNMT3 knockout experiments in mice, DNMT3B has been shown to be required during early embryonic development, whereas DNMT3A is required for normal cellular differentiation[63]. An additional member of the DNMT family, involved in these processes is the catalytically inactive DNA methyltransferase 3-like (DNMT3L), which lacks the catalytical domain to transfer methyl groups. However, it associates with DNMT3A and DNMT3B and stimulates their methyltransferase activity[66]. Moreover, it recognizes nucleosomes with an unmethylated histone H3 lysine 4 (H_3K_4) and recruits DNMT3A and DNMT3B to their targets[67]. It has been shown that DNMT3L is essential for DNMT3A-mediated *de novo* methylation in germ cells but is dispensable for *de novo* methylation during embryogenesis, which is mainly mediated by DNMT3B[68].

In addition to *de novo* methylation, the maintenance of already existing methylation patterns after DNA replication is performed by the maintenance methyltransferase DNMT1. DNMT1 interacts with the multidomain protein UHRF1 (Ubiquitin Like With PHD And Ring Finger Domains 1) and preferentially methylates hemimethylated DNA by copying existing methylation patterns onto newly replicated DNA strands[69,70]. DNMT1 is highly expressed in mammalian tissues and localized to the replication fork where newly synthesized hemimethylated DNA strands are created[71]. The base pairing of CpG allows the reciprocal maintenance of methylation during consecutive replication cycles, due to which methylation can be considered a long-term, relatively stable epigenetic mark[32]. Additionally, DNMT1 is involved in DNA damage repair in a DNA methylation-independent manner to protect cells from mutagenic events[72,73].

DNA demethylation can occur passively or actively. Since DNMT1 actively maintains DNA methylation during cell replication, its absence, inhibition or dysfunction results in a reduction of methylation following each cell division. This decrease in DNA methylation is known as passive DNA demethylation and occurs in dividing cells[74]. By contrast, active DNA demethylation can occur in both, dividing and nondividing cells but

requires enzymatic reactions for the demethylation process[75]. It includes a sequential oxidation of the methyl group by ten-eleven translocation (TET) dioxygenases that can oxidize 5mC to the three additional derivatives 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Figure 7). Further excision of 5fC and 5caC by thymine DNA glycosylase (TDG) generates an abasic site that is subsequently repaired by the base excision repair (BER) pathway, which leads to the restoration of an unmodified cytosine[75,76]. Additionally, oxidized bases are poorly recognized by the DNMT1-UHRF1 complex, resulting in replication dependent passive demethylation mediated by TET[77,78].

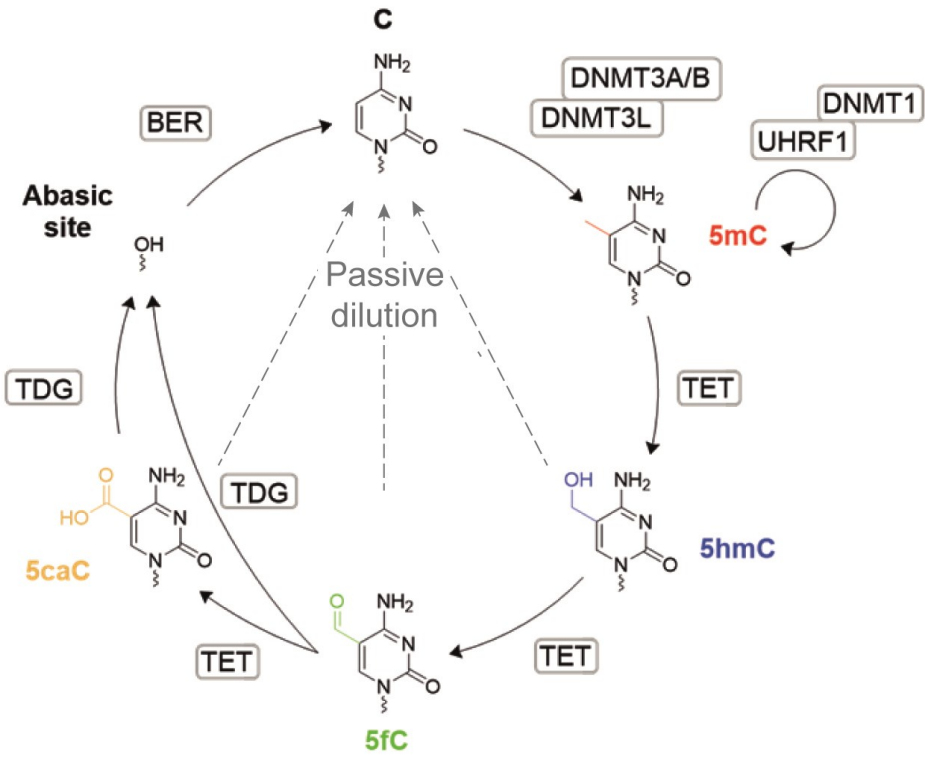


Figure 7. Cycle of cytosine methylation and demethylation. *De novo* DNA methyltransferases (DNMT3A/B) and associated DNA methyltransferase 3 like (DNMT3L) introduce a methyl group at the fifth carbon of cytosine (C), resulting in 5-methylcytosine (5mC). Maintenance DNA methyltransferase (DNMT1) in complex with UHRF1 (Ubiquitin like with PHD and Ring Finger domains 1) maintain already existing 5mC after DNA replication. Sequential oxidation of 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) is catalyzed by ten-eleven translocation (TET) dioxygenases. All modified cytosine bases can be passively removed by replication-dependent dilution. 5fC and 5caC can be cleaved from the ribose moiety by the thymine DNA glycosylase (TDG), leaving an abasic site that is subsequently repaired by the base excision repair (BER) pathway with an unmodified cytosine. Created based on (Muñoz-López et al., 2017) with permission (8.3).

2.2.1 Detection and Quantification of 5-Methylcytosine

Because if its biological significance and potential use as biomarker in disease diagnosis, several strategies for the detection and quantification of 5mC have been developed. Amongst them, single-base resolution methods have been shown to be highly sensitive and provide promising opportunities to comprehensively characterize DNA modifications. Bisulfite sequencing (BS-seq) is the most widely used method for profiling 5mC methylation because of its capacity to generate qualitative and quantitative single-nucleotide resolution information[79]. In this method, the treatment of single-stranded DNA with sodium bisulfite results in deamination of C residues to uracil (U), while 5mC remains nonreactive. After subsequent polymerase chain reaction (PCR) amplification, the converted Cs are read as T, whereas 5mC is still read as C during sequencing. Consequently, the comparison of bisulfite-treated and untreated DNA enables the quantitative discrimination between C and 5mC[80]. Many different BS-seq based methods to detect 5mC and further oxidized bases have been developed, which have different advantages and disadvantages. However, all of them require harsh treatment with bisulfite, which can cause severe degradation of the template DNA[81]. Additionally, the sequence complexity of the DNA sample is reduced by converting unmodified Cs to T, which lowers the sequencing quality and mapping rate[82].

To overcome these problems, bisulfite-free enzyme-based sequencing methods have been developed. One approach, called TET-assisted pyridine borane sequencing (TAPS) is based on a less destructive pyridine borane reductive decarboxylation and deamination chemistry[83]. It converts methylated C to T and therefore preserves genomic complexity. To achieve this, TET enzymes are used to oxidize 5mC and 5hmC to 5caC. 5caC is then converted to dihydrouracil (DHU) via borane reduction. Subsequent PCR reaction leads to the conversion of DHU to T. To sequence 5mC alone, β-glucosyltransferase can be used to protect 5hmC from TET oxidation and borane reduction so that only 5mC is converted to T[84]. Another approach called enzymatic methyl sequencing (EM-seq) detects 5mC and 5hmC, using two sets of enzymatic reactions[85]. In the first one, TET2 and T4-BGT are used to oxidize 5mC and 5hmC to protect them from deamination by APOBEC (apolipoprotein B mRNA editing catalytic polypeptide). In the second reaction, unmodified Cs are deaminated by APOBEC to Us. This leads to the same read-out as with bisulfite-converted DNA and enables the identification of 5mC and 5hmC. In general, bisulfite-free enzyme-based sequencing

approaches use milder reaction conditions, which preserves genomic complexity and allows longer reads with less starting material. However, they involve multiple steps of enzymatic and chemical reactions.

To directly assess cytosine modifications without the requirement for prior chemical or enzymatic conversions and PCR amplification, single molecule real-time (SMRT) or nanopore sequencing can be directly applied to native DNA[86,87]. These third-generation sequencing technologies enable single molecule sequencing in real time and offer the opportunity to directly detect epigenetic base modifications[88]. SMRT sequencing relies on the direct observation of single DNA polymerases while incorporating fluorescently labelled bases during the synthesis of new DNA molecules. Kinetic characteristics of the DNA polymerization such as the time duration between two successive base incorporations are altered by the presence of a modified base and allow its detection[86]. In nanopore sequencing, single-stranded DNA is translocated through a nanopore under the use of a DNA polymerase for kinetic control. The measurement of specific electric signal characteristics enables the characterization of the different canonical and epigenetic nucleobases, while the DNA molecule passes through the nanopore[87]. Even though these methods are straightforward and can map 5mC in the genome at base resolution, they are not applicable to studies at a cellular level because genomic DNA extraction is required.

At a cellular level, global DNA methylation patterns can be analyzed with imaging-based strategies by using either anti-5mC antibodies[59] or methylated DNA binding domain fusion proteins[89,90]. These strategies enable the visualization of global methylation patterns and provide insights into the functional consequences of DNA methylation on gene expression and chromatin structure. However, they do not provide sequence information.

One method to enable *in situ* analysis of 5mC in a sequence-specific context, called methylation-specific fluorescence *in situ* hybridization (MeFISH), has been developed for microscopic visualization of 5mC in individual cells[91]. This method is based on a differential reactivity of 5mC and C in target DNA to crosslink with osmium and bipyridine-containing nucleic acid (ICON) probes. In particular, heat-denatured cellular DNA is hybridized with fluorescently labeled ICON probes which contain a bipyridine-attached adenine derivative opposite to the cytosine of interest. Osmium treatment

results in a 5mC-dependent crosslinking, whereas probes opposite to a nonmodified cytosine do not form a stable osmium complex and are removed in a subsequent denaturation step. The remaining MeFISH signals can be observed by fluorescence microscopy and allow the sequence-specific detection of 5mC in individual cells[91]. Furthermore, this method is compatible with RNA FISH and protein immunostaining and therefore enables spatiotemporal co-analysis of modified cytosines with other regulatory elements or chromatin-associated processes[92,93].

In alternative strategies for sequence-specific detection of 5mC in living cells, reader proteins of methylated cytosines are combined with programmable DNA-binding domains in fluorescence complementation designs. In these approaches, the 5mC-reader and the programmable DNA-binding domain are each fused to non-fluorescent fragments of a fluorescent protein (FP). The presence of 5mC at the target sequence causes a close proximity of both fragments and leads to reconstitution of a functional fluorophore[94]. As 5mC-reader, a methyl-CpG-binding domain (MBD) protein is commonly used. It can be combined with various programmable DNA-binding domains like Cys_2His_2 zinc finger proteins (ZFPs), transcription activator-like effector (TALE) proteins and clustered regularly interspaced short palindromic repeat (CRISPR)-catalytically dead CRISPR-associated protein 9 CRISPR-(dCas9)[94]. The combination of these programmable DNA-binding domains with selective 5mC-readers in fluorescence complementation designs allows the dynamic and locus-specific monitoring of epigenetic modifications at genomic loci in living cells[94]. However, it does not provide nucleotide and strand resolution because of the requirement for two different receptor molecules.

To detect 5mC sequence-specifically with nucleotide and strand resolution, TALE proteins have been engineered to obtain additional selectivity for modified DNA nucleobases[95–98]. This allowed the generation of receptor molecules that provide sequence specificity and 5mC selectivity in a single scaffold. By exploiting the different 5mC sensitivities of natural and engineered TALE repeats, an imaging-based approach has been developed to analyze 5mC levels in single cells[99,100].

2.3 TALEs Transcription Activator-Like Effectors

2.3.1 The Natural Role of TALEs and their Structure

Transcription activator-like effector (TALE) proteins are sequence-specific DNA binding proteins. They are naturally secreted by plant pathogenic bacteria of the genus *Xanthomonas* that can cause severe diseases on many crop plants[101]. TALEs enter plant cells via a bacterial type III secretion system[102] and are imported into the nucleus where they function as virulence factors by binding defined promoter sequences to activate the transcription of genes that support bacterial infection[103,104]. TALEs recognize DNA through a modular central repeat domain (CRD) composed of a variable number of tandem 34 amino acid repeats (Figure 8). In naturally occurring TALEs, the number of repeats has been found to range from 1.5 to 33.5 with an average of 17.5 repeats[101]. The last repeat of the CRD (most C-terminal repeat) is truncated to 20 amino acids, but still functional and therefore referred to as half repeat[105]. The sequence of each tandem repeat is highly conserved but with variations occurring at position 12 and 13. These two positions are termed repeat variable diresidues (RVDs) and are responsible for the specific recognition of a unique nucleotide[106]. The four most abundant RVDs from naturally occurring TALEs are HD, NG, NI and NN which respectively recognize the nucleotides C, T, A and G or A[105,107,108]. Thus, DNA targeting is enabled by a basic code in with each RVD corresponds to a specific nucleotide.

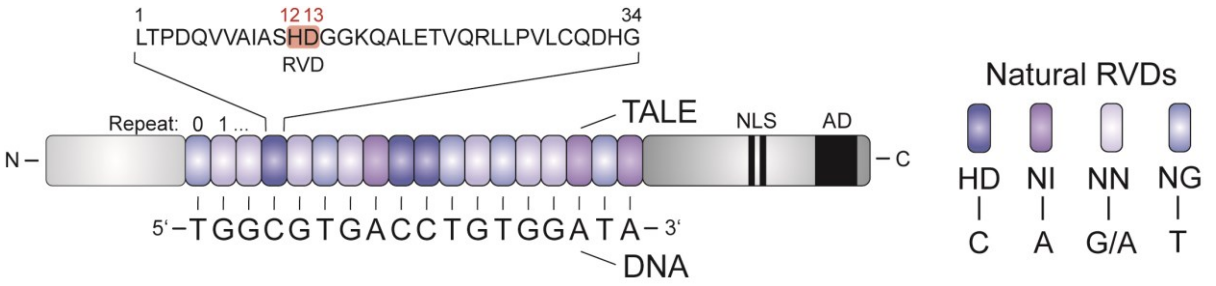


Figure 8. TALE domain organization. TALEs contain a central repeat domain (CRD) with a variable number of tandem repeats that bind sequence-specifically to a target nucleotide sequence. Sequence of one repeat is shown with RVD highlighted in red. Nucleotide selectivities are shown on the right. The N-terminal domain contributes to TALE specificity and binding affinity. The C-terminus contains nuclear localization signals (NLS) and an acidic activation domain (AD) for transcriptional activation. Created based on (Muñoz-López et al., 2020) with permission (8.3).

Each TALE repeat forms a left-handed, two-α-helix bundle in which the RVD is located within a well-ordered interhelical loop that connects the two helices[109]. The first amino acid (His or Asn) in each RVD, at position 12, orients away from the DNA but engages a hydrogen bond with the carbonyl oxygen of a conserved Ala8 to stabilize the interhelical loop. The second amino acid at position 13 projects deep into the major groove of the DNA and interacts sequence-specifically with a single nucleotide base of one strand of the double-stranded DNA[108] (Figure 9 and 10). These interactions are dependent on the individual RVD and the corresponding nucleotide. Directional hydrogen-bonds are observed for RVD HD in contact with C, or RVD NN in contact with purines. Highly complementary steric packing in the absence of hydrogen bonds can be observed between RVD NG and T, and interactions that achieve reduced specificity through steric exclusion of alternate bases seem to appear with RVD NI[108]. Due to this, each RVD contributes differently to the overall binding affinity and specificity. Stronger interactions like the hydrogen bonds, formed by RVD HD and NN contribute more, in contrast to the weaker van der Waals interactions formed by NG and NI[110,111]. However, it has been shown that non-specific interactions between the DNA backbone of the sense strand and residues Gly14, Lys16 and Gln17 in each repeat confer most of the binding energy and that the RVDs contribute only relatively little to the overall binding energy[112].

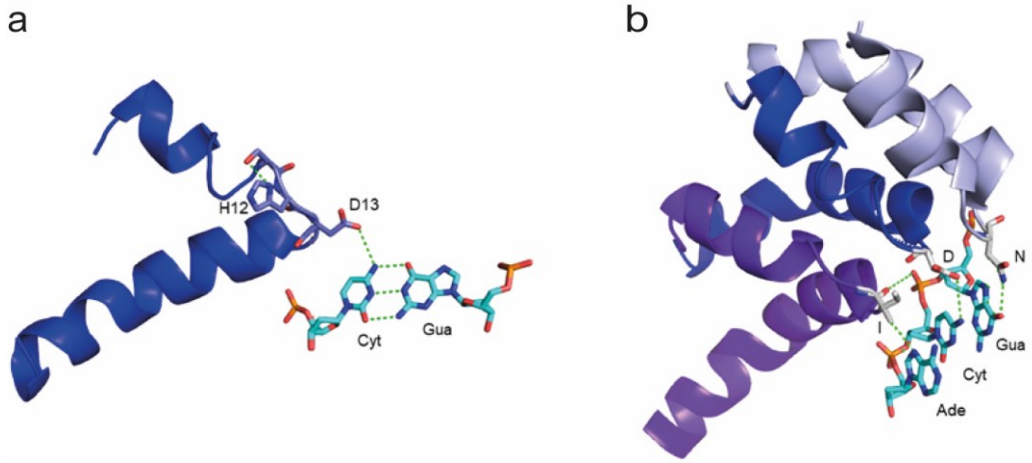


Figure 9. Crystal structure of TALE repeats and interactions with nucleotides. a) Left-handed, two-helix bundle of TALE-repeat with RVD HD in interhelical loop. Amino acid histidine at position 12 interacts with protein backbone. Aspartic acid at position 13 forms hydrogen bond with cytosine (Cyt). b) Interactions of RVD NN, HD and NI with guanine (Gua), cytosine and adenine (Ade). Hydrogen bonds are represented as dotted green lines. C = grey/cyan, N = blue, O = red, P = orange. (PDB ID: 3UGM[113])

The individual repeats of the CRD carry a relatively neutral overall charge and are able to self-assemble. By this, a right-handed superhelix is formed that wraps along the DNA major groove and makes sequence-specific interactions with the sense strand of the DNA duplex[114] (Figure 10). Thus, the TALE recognizes the target DNA sequence from the 5' to the 3' nucleotide, in an N- to C-terminal orientation[115].

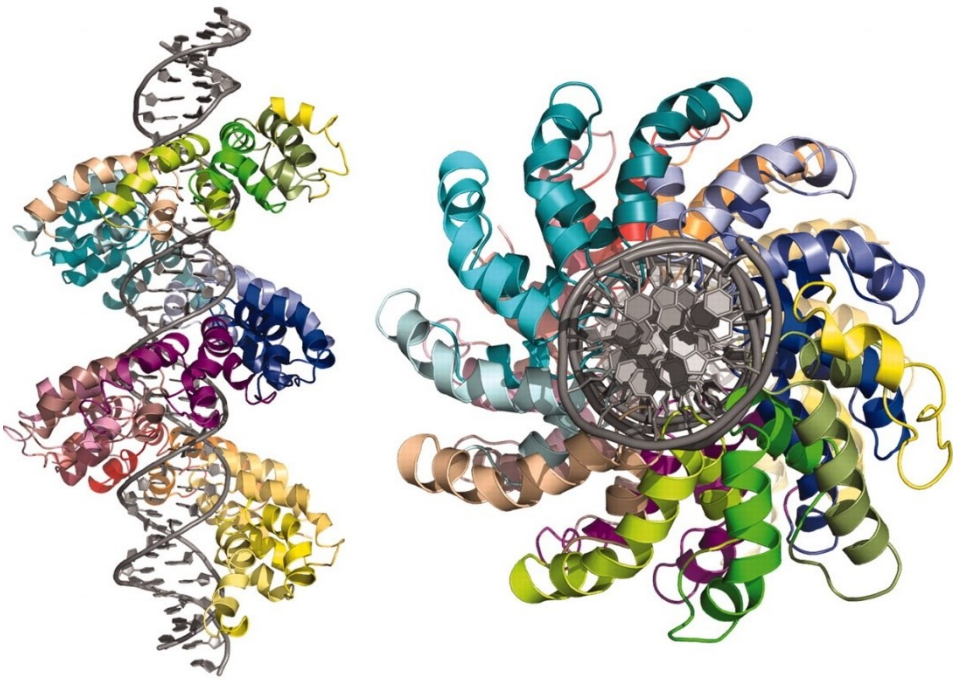


Figure 10. Crystal structure of TALE bound to target DNA in the major groove. DNA binding region from TALE PthXo1 in complex with its DNA target site. Single repeat units are shown in different colors. Adapted from (L. DeFrancesco, 2012) with permission (8.3).

In addition to the CRD, the non-repetitive N-terminal region (NTR) contributes to the TALE specificity and binding mechanism and provides most of the binding energy required for the high affinity binding[108,116]. Very similar to the CRD, it harbors a right-handed superhelical structure with four repeat-like helical bundles (cryptic repeats) that bind to DNA in a sequence-independent manner[117]. In naturally occurring TALEs, binding sites always begin with a T, which is specified by a tryptophan within one of the cryptic repeats of the N-terminus. This tryptophan is strictly conserved among natural occurring TALEs and interacts with the T base prior the RVD encoded bases[105,107]. Additionally, the type III translocation system required for secretion, is located in the NTR. The C-terminus (CTR) of TALEs is not required for DNA-binding and typically contains nuclear localization signals (NLS) that direct the TALE to the nucleus. It also

contains functional domains required for transcriptional activation, such as an acidic activation domain (AD)[118] (Figure 8).

To find their unique target site within millions of non-specific DNA sequences, TALEs need to accommodate a rapid search dynamic and a stable binding to their target site. To accomplish this, they exhibit distinct modes of action: a search, a recognition, and a bound mode[117,119] (Figure 11). These modes are facilitated by the NTR and CRD of the TALE. The search process is initiated when the NTR binds non-specifically to DNA and promotes the wrapping of the CRD around the DNA in a loose confirmation. This allows subsequent non-rotational sliding and micro-dissociation/reassociation of the TALE along the DNA to perform a rapid, one-dimensional and unbiased search. The process is known as sliding and hopping mechanism and allows a facilitated diffusion in complex cellular environments, where DNA molecules are crowded and coated with DNA-binding proteins[117,119]. Because of the looser conformation of the CRD and the non-rotational sliding in the search mode, electrostatic interactions between the CRD and the DNA backbone are minimal, which allows rapid movement. Upon encountering a target sequence, a conformational change is induced either by the compression of the N-terminal domain in a zipper-like fashion or by any repeat within the CRD[120]. If the repeats match the target DNA sequence, electrostatic interactions between the CRD and the DNA are formed. The TALE is now tightly bound to the major groove of the DNA and more compact in contrast to the search mode. However, the sequence specificity is mainly achieved through negative discrimination arising from steric and electrostatic clashes with non-specific DNA, rather than from a relatively small gain of overall binding energy, when binding to the target sequence[117].

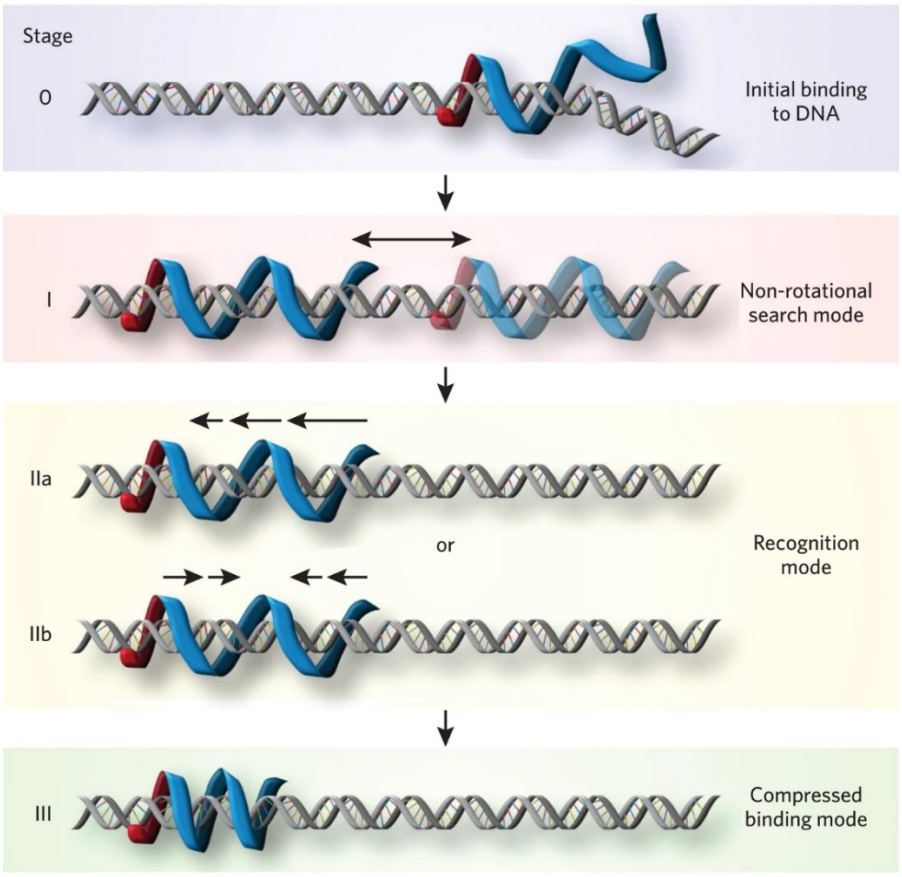


Figure 11. TALE search, recognition and binding modes. DNA-binding of a TALE is initiated (0) by the N-terminal domain (red), which promotes the non-rotational search along the DNA (I), during which the repeat region (blue) is loosely wrapped around the DNA. Upon encountering a target sequence, a conformational change is induced either by the N-terminal domain (IIa) or by any repeat throughout the array (IIb). In case of a matching sequence, the TALE enters the binding mode (III) where it is tightly bound to the major groove of the DNA. Adapted from (Becker et al., 2016) with permission (8.3).

2.3.2 TALE Assembly

Due to the basic code of DNA recognition, in which each RVD corresponds to a specific nucleotide, TALEs can be easily designed to target any desired DNA sequence by assembling the corresponding sequence of RVD repeats. To aid with the repetitive nature of the RVD repeats in TALE assembly, a number of tools are available today for the design and rapid assembly of TALE repeat regions. The most commonly utilized method combines digestion and ligations steps in a single Golden Gate reaction[121,122]. Golden Gate cloning is used to assemble multiple DNA fragments in an ordered fashion. As building blocks, plasmids are used that carry the sequences for the individual RVD repeats. By using type II restriction endonucleases that cleave outside their recognition sites, modules containing the desired RVDs can be generated with unique 4 bp overhangs. The specific design of these cleavage sites guarantees the ligation of the RVDs in a correct order and allows single-reaction assembly due to elimination of the enzyme recognition sites after correct ligation[123]. The assembly is performed in two steps. In the first Golden Gate 1 (GG1) step, multi-repeat modules with up to ten RVD-repeats are assembled in single plasmids. Depending on the number of required RVDs, several plasmids are needed to cover all repeats. In the second Golden Gate 2 (GG2) step, a different type II restriction enzyme is utilized to join the array fragments generated in the GG1 reaction with the last truncated repeat (half repeat) and insert them in the final entry vector. The final entry vector contains the N- and C-terminal domains of the TALE along with the functional domains required for the desired application of the TALE.

2.3.3 TALE-Based Tools and their Applications

By replacing the natural C-terminal activation domain of TALEs with domains that function in mammals, various TALE-based tools have been developed, enabling a wide variety of genome engineering applications including transcriptional modulation, genome and epigenome editing as well as the visualization of genomic elements (Figure 12).

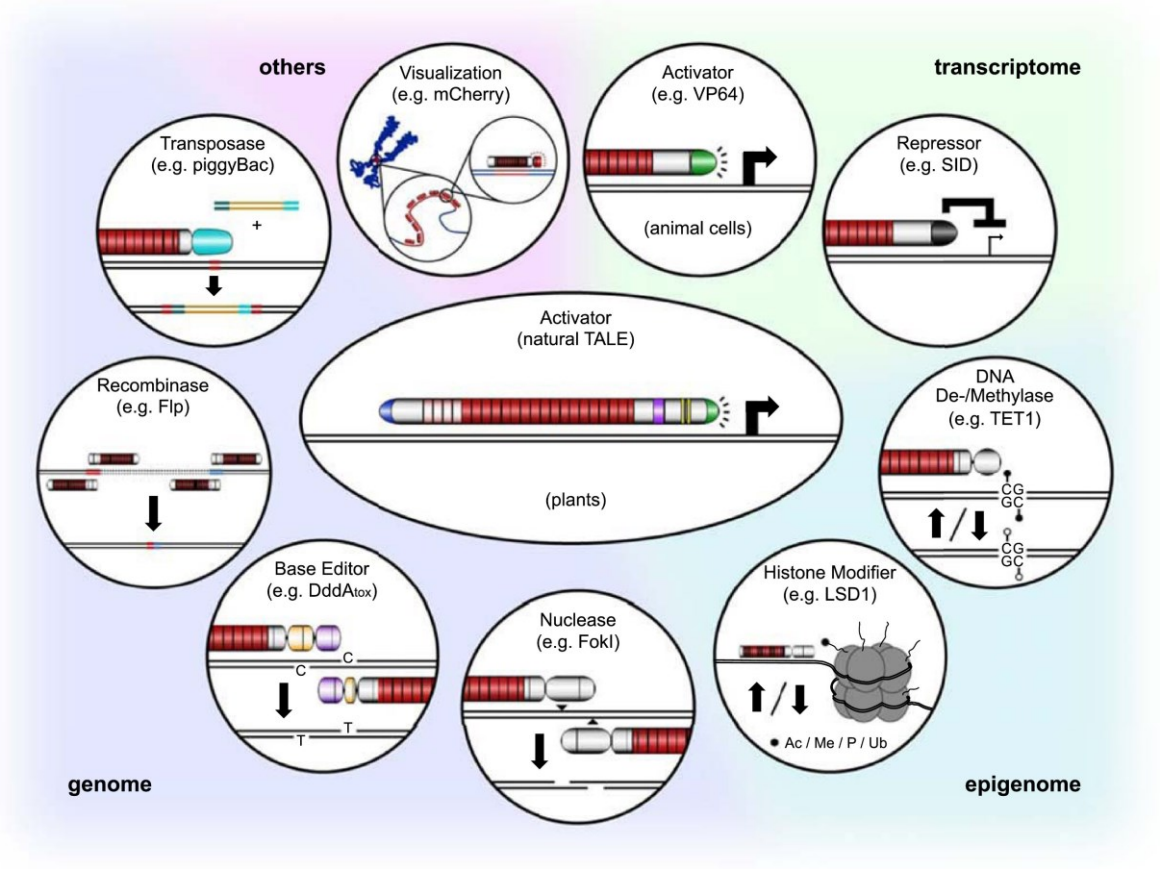


Figure 12. Overview of different TALE-based tools for genome engineering applications and visualization of genomic elements. By fusing functional domains to the DNA-binding domain of TALEs, various TALE-based tools have been generated to modify the transcriptome, epigenome and genome. They can also be used to visualize genomic elements. (Ac: Acetylation, Me: Methylation, P: Phosphorylation, Ub: Ubiquitination). Adapted from (Becker and Boch, 2021) with permission (8.3).

For transcriptional modulation, TALEs have been fused with transcriptional activators or repressors. Targeted gene activation of exogenous and endogenous genes in human and other mammalian cells has been achieved with TALEs fused to the VP16 activation domain from herpes simplex virus or its tetrameric derivative VP64[124,125]. Similarly, targeted gene repression has been shown with TALEs fused to various repressor domains like the Krüppel associated box (KRAB) or mSin interaction domain (SID)[126].

To accomplish targeted genome editing, TALEs have been fused to the catalytic domain of the homodimeric *Fok*I restriction endonuclease to create TALE-nucleases (TALENs) which allow DNA cleavage[127,128]. Because *Fok*I functions as a dimer, two TALENs are required that bind sites on opposing strands of the DNA such that the *Fok*I domains can dimerize and cleave the DNA within a short spacer between the two TALEN binding

sites. The induced double strand breaks activate cellular DNA repair systems and can be repaired via various repair pathways. Whereas non-homologous end joining (NHEJ) results in small variable insertions or deletions which typically lead to gene knockouts, homologous recombination (HR) can be used to replace or insert new DNA segments if a repair template is used[108]. Because double strand breaks are predominantly repaired by NHEJ, TALE nickases have been developed, which cut only a single strand of DNA and induce the HR repair mechanism[129,130]. In an alternate approach, TALEs have been used to guide transposons to a genomic target location. For this, TALEs were fused to a hyperactive variant of the piggyBac transposase to allow the transposition of cargo DNA to the genomic target location[131]. TALE DNA binding domains were further combined with site-specific recombinases such as the tyrosine recombinases Flp and Cre, or the serine recombinase Gin, to achieve targeted genomic rearrangements[132,133]. However, recombinase domains function in a sequence-dependent manner, which is why the sequence specificity of TALE recombinases not only depends on the TALE DNA binding domain, but also on the natural DNA-binding domains of the recombinase itself[134]. Furthermore, engineered TALEs have been used for direct base editing by combining them with the deaminase domain of the bacterial protein DddA (double-stranded DNA deaminase toxin A) that functions as a deaminase and facilitates C-G to T-A conversions. Because fusions of DddA and the DNA-binding domain of a TALE are cell toxic, the deaminase domain was split into two non-toxic halves which were fused to two oppositely binding TALEs. The proximity induced reconstitution led a functional protein with high target specificity[135].

To allow gene expression control without inducing changes on the DNA sequence, DNA-binding domains of TALEs have been fused to enzymes that can modify epigenetic marks. These include epigenetic modifier domains that can methylate or demethylate individual DNA nucleotides and those that can alter histone tail modifications by adding or removing methylation, acetylation, phosphorylation or ubiquitination signals. By altering DNA methylation patterns within promoter or enhancer regions, an increase or decrease of target gene expression can be achieved. This allows the analysis of specific DNA methylation marks in correlation with promoter activity. When fused to DNMT3A or DNMT3L, or their catalytic domains, TALEs provided local CpG methylation resulting in effective and stable downregulation of expression[136-138]. By contrast, the fusion of a TET1 catalytic domain enabled reactivation of endogenous gene expression by targeted

demethylation[139-141]. Several histone-modifying enzymes have been shown to efficiently downregulate target gene expression when fused to DNA-binding domains. The most efficient ones include various histone methyltransferases, demethylases and deacetylases[142-144]. By contrast efficient gene activation has been shown with TALE fusions to the core domain of acetyltransferase p300, leading to acetylation of lysine 27 in histone 3[134,145].

Furthermore, TALEs have been used to allow genome visualization. By fusing TALEs to FPs, repetitive target sequences, such as telomeric repeats or centromere-associated satellite DNA sequences could be visualized in fixed and cultured cells as well as in living organisms[146-149]. This allowed the monitoring of their spatiotemporal organization throughout the cell cycle in a variety of different fluorescent colors.

2.3.4 Sequence-Specific Detection of Methylation Using TALEs

In early studies, a 5mC-sensitivity of TALEs has been reported. This sensitivity was initially described as a bottleneck for genome engineering applications since it hampered the binding of TALEs to their endogenous target sequence when the sequence was methylated[150]. To overcome this drawback, chemical inhibition of DNA methyltransferases by 5-aza-2′-deoxycytidine (5-aza-dC) has been shown to facilitate targeted transcriptional activation due to hypomethylation[150]. However, its cytotoxicity makes it unsuitable for most applications[151]. The sensitivity of TALEs towards 5mC is attributed to the RVD HD, which recognizes C through a hydrogen bond between the aspartate carboxyl group and the C4-amino group[96,113]. The introduction of a methyl group at the C5 position of cytosine reduces the affinity of RVD HD likely due to an increased steric demand in the major groove and shielding of the polar amino group, which negatively affects the hydrogen bond[97,141,152]. By contrast, RVD NG, the natural binder of T, can bind the structurally similar 5mC through a hydrophobic interaction between the 5-methyl group and the Cα-methylene moiety of glycine[109,113]. The missing side chain of glycine provides sufficient space to accommodate the 5-methyl group of T and 5mC, which only differ in their chemical group at the C4 position[96] (Figure 13). By exploiting the different potential of RVD HD and NG to bind 5mC, it has been shown that TALEs have the potential to differentiate 5mC from C[97].

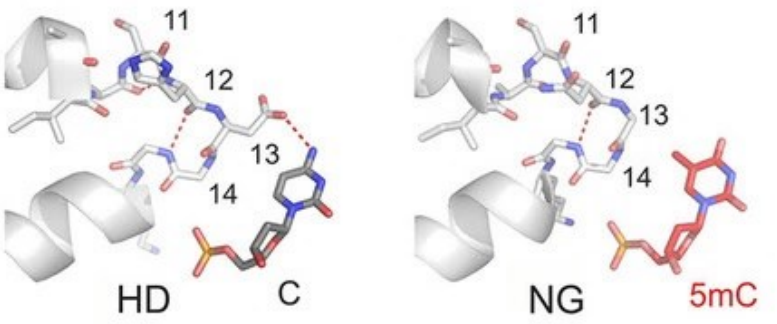


Figure 13. Crystal structures of TALE RVDs HD and NG bound to C or 5mC, respectively. Modified from (Muñoz-López et al., 2020) with permission (8.3).

Furthermore, the natural TALE repeat N*, in which the amino acid at position 13 is deleted, has been shown to bind C and T with similar binding efficiency[107]. This suggested that a missing side chain at position 13 could efficiently accommodate 5mC and led to the identification of TALE repeat N* as universal C and 5mC binding module[95]. Ultimately, the combination of TALEs bearing either the 5mC-selective RVD HD or the universal RVD N* opposite to a potentially methylated cytosine allowed the detection of 5mC with single nucleotide resolution[153]. This concept was further expanded to additionally recognize 5hmC by exploiting the high affinity of RVD HD to bind C but not 5mC or 5hmC, the preferential binding of RVD NG to 5mC but not C or 5hmC, and the increased binding of RVD N* to C and 5mC but not 5hmC[152]. By immobilizing TALEs on a solid phase, affinity enrichment of user-defined DNA sequences from the human genome with selectivity for single epigenetic cytosine nucleobases could be achieved[154]. This allowed the quantification of 5mC levels at single genomic nucleotide positions in a variety of cancer biomarker sequences. Furthermore, genomic target sequences could be isolated with selectivity for single 5hmC, and in combination with sodium borohydride reduction, for single 5fC nucleobases, by using the TALE repeat N*[154].

TALEs recognize nucleotides through direct and sequence-specific contacts in the major groove of DNA. This offers the possibility to engineer TALE repeats for improved or direct recognition of epigenetic nucleobases. Because the backbone loop of TALE repeats is very closely positioned to the C5 position of cytosines and provides insufficient space for the accommodation of larger 5-substituents, size-reduced repeat loops have been screened for their capability to recognize epigenetically modified cytosine[154,155]. This screening identified the repeats G*, S* and T* as universal repeats because they bind C, 5mC, 5hmC, 5fC and 5caC with similar affinities. This is probably attributed to the

strongly reduced RVD loop size and the complete removal of interactions with the nucleobases[154,155].

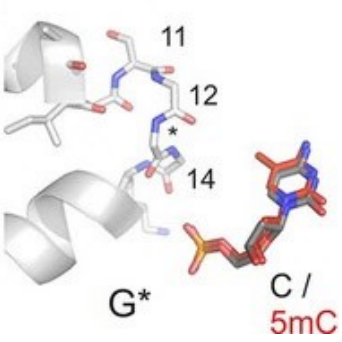


Figure 14. Model of RVD G* bound to C or 5mC. Modified from (Muñoz-López et al., 2020) with permission (8.3).

Interestingly, the repeat P* showed high or moderate affinity to all C bases except for 5caC, making it a negative sensor for 5caC[155]. By contrast, the repeat R**** exhibited strong binding to 5caC but not to any other nucleobase, making it a positive sensor for 5caC[156]. Another study screened a library containing randomized amino acids at RVDs and their neighboring residues for 5mC selectivity[157]. Here, the consensus sequence XXAA showed the highest 5mC selectivity. Amongst them, the ASAA repeat was discovered as the best candidate and has been used to activate endogenous gene expression in a methylation-dependent manner. In another study, all theoretical combinations of amino acids at the RVD (XX) as well as all possible size reduced repeats (X*) were tested to recognize C, 5mC and 5hmC[158,159]. This study identified and later structurally characterized RVD HA as 5mC selective binder, RVD FS as 5hmC binder, RVD RG as 5mC and 5hmC but not C binder, and R* as universal binder. By utilizing these repeats, methylation-dependent gene activation, gene editing and locus-specific 5hmC detection has been achieved.

In an alternative strategy to increase the 5mC selectivity of TALEs, the nonselective binding energy of TALEs has been reduced by substituting several amino acids within the TALE scaffold[160]. In addition to the nucleobase-specific interactions of the RVDs, TALEs interact nonspecifically with phosphates of the DNA backbone via basic amino acids. These include many residues in the NTR and CTR as well as a GG diresidue at positions 14 and 15 and a KQ diresidue at positions 16 and 17 within the repeats of the CRD[109,113]. The substitution of up to two basic residues within the NTR by alanine (K262A, R266A or K171A/R173A) resulted in a moderate increase of C over 5mC

selectivity. Stronger selectivity enhancements were obtained with K16A + Q17A double mutations in the CRD repeats. The combination of mutations in the NTR and CRD was highly synergetic and strongly enhanced the C over 5mC selectivity in genomic affinity enrichments and 5mC selective transcriptional activation *in vivo*.

The intrinsic 5mC sensitivity of RVD HD has further been used for imaging-based single cell analysis of 5mC in defined human DNA sequences[100]. For this, two differentially fluorescently labelled TALEs that otherwise differed only in the presence of either RVD HD or G* opposite to target Cs were used in co-stainings of mammalian cells (Figure 15). Because methylated DNA regions in cells are often bound by 5mC-binding proteins and tend to form heterochromatin, the accessibility for TALEs can be affected[157]. To account for this, a G*-TALE (universal binder of C and 5mC) has been employed to serve as an internal control for the target accessibility, whereas a HD-TALE was used to selectively distinguish between C and 5mC[100]. By analyzing the fluorescence ratio of both TALEs, 5mC levels of human SatIII DNA have been analyzed and correlated with the presence of the SatIII marker protein heat shock factor 1 (HSF1). In a follow up study, the screening of size-reduced TALE repeats resulted in the discovery of TALE repeat NH* which showed a pronounced 5mC-selectivity[99]. The NH* repeat contains a deletion of amino acid 13 and a S11N mutation to allow various polar interactions within the RVD loop. Replacement of the previously used G*-TALE with the NH* TALE enabled the detection of a positive response upon methylation of the target sequence. This allowed its use as a 5mC-reporter in contrast to the mere target accessibility control function of the universal G*-TALE[99].

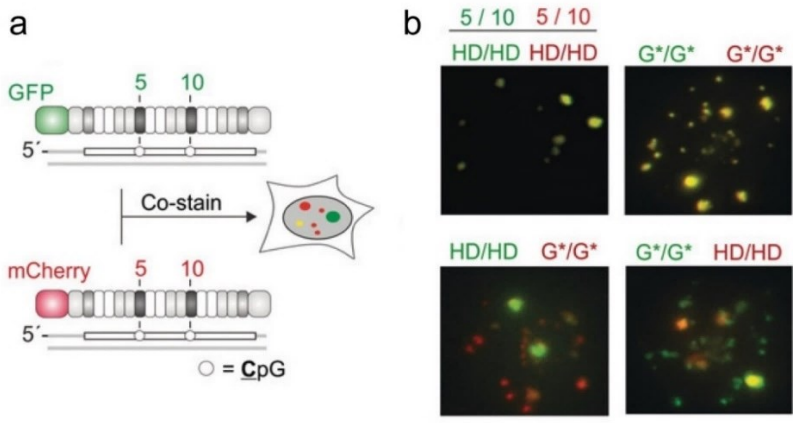


Figure 15. Imaging-based analysis of 5mC in single cells. a) Scheme of co-staining with two differentially fluorescently labelled TALEs targeting the same target sequence with two potentially methylated Cs at positions 5 and 10. b) Fluorescence images of co-stainings with indicated HD or G* repeats at positions 5 and 10. Modified from (Muñoz-López et al., 2020) with permission (8.3).

2.4 Visualization of Genomic DNA Sequences

Microscopic observation is the classic approach to provide morphological information of cellular structures as well as localization information of molecules. Because the most fundamental processes in living organisms like gene expression, DNA replication and genome maintenance are critically regulated by the spatiotemporal organization of the chromatin structure, many imaging approaches have been developed to analyze chromatin architecture and dynamics[161]. Fluorescence in situ hybridization (FISH) is the most commonly used method to visualize the localization of specific DNA sequences[162,163]. It involves the collective hybridization of various dye-labeled oligonucleotide probes to unique sequences within a specific genomic locus. This allows the characterization of its spatial organization also in combination with other genomic loci or FP/antibody-tagged proteins of interest[164-166]. However, FISH requires cell fixation and DNA-denaturation through treatment with formamide and high temperatures. This alters the chromatin structure and prevents studying the chromatin dynamics[167]. To allow live-cell imaging, for many years the labelling of specific genomic loci was constrained by the availability of proteins that are capable to bind specifically to highly repetitive DNA sequences[167]. Accordingly, telomeres and centromeres haven been labeled by fluorescent tagging of their corresponding binding proteins[168-170]. To visualize other genomic loci, initially Lac or Tet operator repeats have been integrated into genomic regions by genome engineering[171]. These artificial sites contain a recombinant locus carrying large arrays of binding sites that can be traced with fluorescently labelled Lac- or Tet-repressor binding proteins[172-174]. However, these systems rely on genome engineering tools and require the introduction of artificial sequences of about 10 kb, which can affect chromatin structure and function of the targeted loci[175]. More flexible approaches utilize programmable DNA-binding proteins such as ZFPs, TALEs or the CRISPR-Cas system. By labelling these systems with FPs, they can guide the fluorescence signal to a specific sequence within the complex genome.

ZFPs are the most common DNA-binding proteins found in eukaryotes[176]. By fusing them to FPs, they were the first constructs that have been applied to label and trace specific DNA sequences in living cells[177,178]. Each individual zinc finger consists of 30 amino acids of which two cysteines and two histidines are coordinated by a zinc ion to form a ββα-configuration. By this, finger-like protrusions are established, which can

make specific contacts with a nucleotide triplet[179-181]. ZFPs belong to one of the most abundant groups of proteins in humans with a wide range in structure and function[182]. Many of these proteins are transcription factors in which the zinc-binding motif mediates direct DNA interactions by the recognition of specific DNA sequences[179,182]. Their binding properties depend on the amino acid sequence of the finger domains, the linker between adjacent fingers and additional interactions with the DNA phosphate backbone[183]. Protein engineering allowed the design of ZFPs that can target user-defined binding sites. In the most common modular assembly process, separate zinc fingers that can each recognize a three-base pair DNA-sequence are combined to generate arrays that can recognize various target sequences[184,185]. However, ZFPs exhibit a context-dependent binding preference since individual zinc finger binding specificities can overlap and depend on neighboring zinc fingers[186]. This makes the assembly of functional ZFPs challenging and requires extensive screening processes.

By contrast TALE proteins contain a modular DNA-binding domain composed of concatenated repeat sequences, each recognizing a unique nucleobase[115]. This modularity has been used to develop engineered, fluorescently labeled TALEs to visualize endogenous repetitive sequences in fixed and cultured cells as well as in living organisms[146-149]. By fusing different FPs with TALEs, centromeres and telomeres have been visualized in the same cell, demonstrating the capability of TALEs to simultaneously image multiple loci[147]. Furthermore, it has been shown that TALEs can be used to monitor the spatiotemporal organization of repetitive sequences during the cell cycle and embryonic development[146,149]. To improve the signal-to-background ratio (SBR) of TALE imaging applications, further studies have applied bimolecular fluorescence complementation (BiFC) to label repetitive genomic loci with TALEs[187,188]. Here, two TALEs fused to non-fluorescent fragments of a FP are brought together by binding to target sequences on different strands of the double-stranded DNA that are separated by a few base pairs. The proximity mediated structural complementation of the two FP fragments into a functional FP leads to reconstitution of fluorescence only at target sites. By this, background fluorescence caused by diffuse full-length FPs could be reduced. Further studies also incorporated dye-modified amino acids (lysine-green[147] or Cy3-lablelled cysteine [189]) into TALEs or combined TALEs with a quantum dot labelling technique to allow live cell imaging of single genomic loci[190].

The CRISPR-Cas9 system is derived from the bacterial immune system against viruses and has been modified as a tool for many purposes like gene regulation and editing, chromatin engineering, epigenetic editing and imaging[191-195]. Engineered CRISPR-Cas9 systems contain a CRISPR-associated endonuclease (Cas9) and a single guide RNA (sgRNA)[196]. The sgRNA is composed of a scaffold sequence necessary for Cas9-binding and a nucleotide sequence that is complementary to a target DNA, in order to direct the Cas9-protein to a genomic target sequence. By altering the sequence of the sgRNA, the Cas9-protein can be easily programmed to target any DNA sequence of interest. To allow a broader spectrum of applications, the endonuclease activity of Cas9 can be catalytically deactivated (dCas9) to avoid DNA cleavage[197]. This enabled the visualization of specific endogenous DNA sequences by using PF-tagged dCas9 proteins[195]. While for highly repetitive sequences a single sgRNA has been shown to generate a sufficient signal for detection, labeling of non-repetitive sequences required the simultaneous expression of several sgRNAs and signal amplification methods[195,198,199].

2.4.1 Signal Amplification Methods

Since the fluorescence of an individual fluorescent molecule is often too faint to be detected by conventional microscopy, signal amplification methods have been developed to improve signal-to-noise ratios (SNRs) and allow imaging of non-repetitive genomic loci. Various methods have been developed in order to recruit multiple fluorescent proteins at a certain genomic locus. One simple approach to achieve this is to tile an array of FP-tagged proteins along the target locus (Figure 16a). With the CRISPR-Cas9 system, this has been achieved by simultaneous expression of 26 unique sgRNAs targeting same region[195]. This allowed labeling of a non-repetitive region in the cell surface associated mucin 4 (MUC4) locus. However, the delivery of many sgRNAs into cells remains challenging and increases the potential of off-target sites[200]. To simplify the delivery of several sgRNAs, a molecular assembly strategy termed chimeric array of gRNA oligonucleotides (CARGO) has been developed[199]. With this strategy, a multiplexed gRNA array can be easily assembled in one plasmid. This has been used to achieve efficient delivery of 12 sgRNAs into cells and recruit dCas9-eGFP to image a genomic region spanning approximately 2 kb[199].

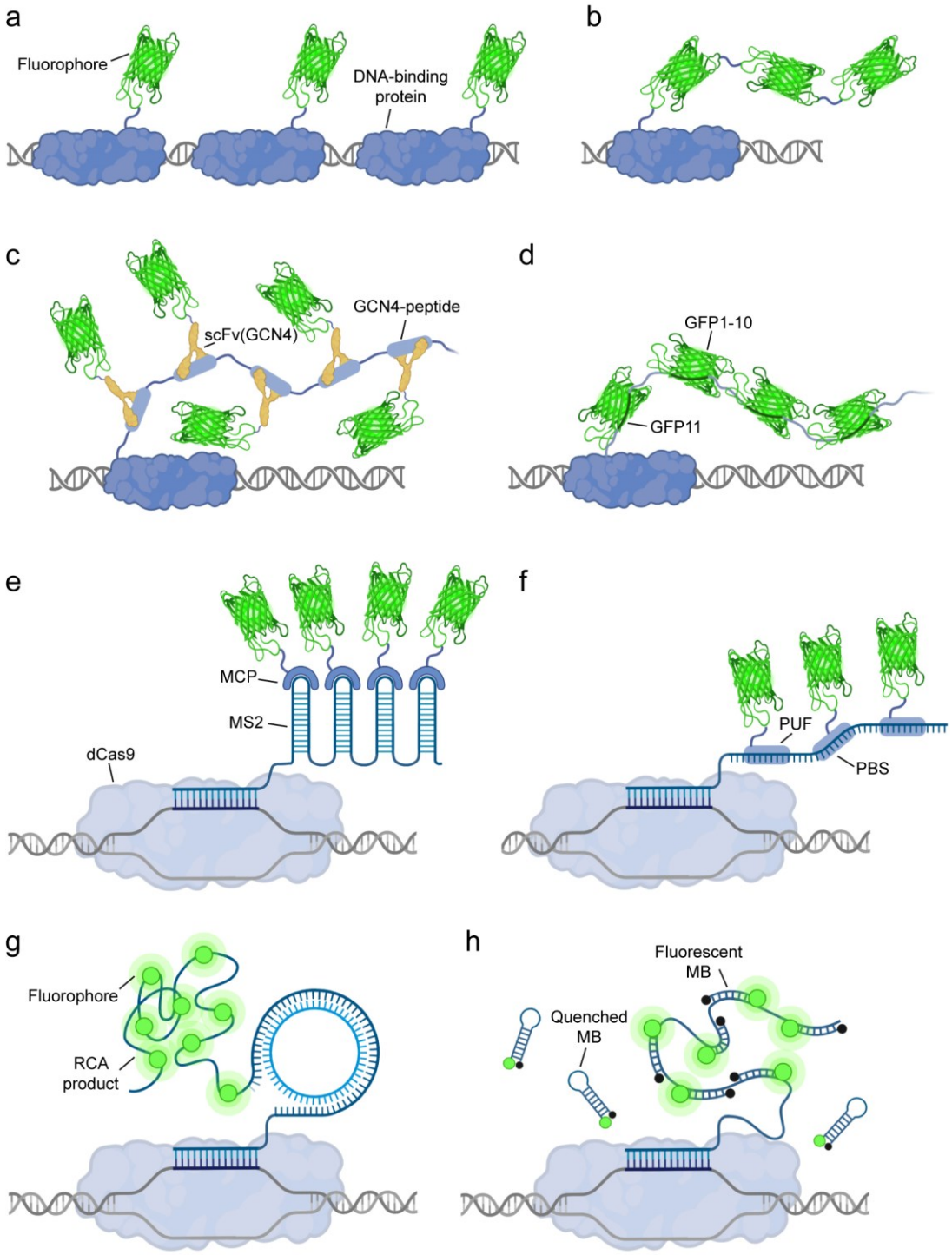


Figure 16. Signal amplification methods to label genomic loci. a) Recruitment of multiple fluorescent proteins to a genomic locus. b) Fusion of three copies of FPs to DNA-binding protein. c) SunTag signal amplification system. d) Self-complementation of split super-folder GFP with tandem array of GFP11 fused to DNA-binding protein, bound by GFP1-10. e) CRISPR-Cas9 system with sgRNA appended with MS2 motifs, bound by fluorescently labeled MS2 capsid protein (MCP). f) Casilio signal amplification with PUF-binding site (PBS) bound by fluorescently labeled Pumilio/Fem3 RNA-binding factor (PUF). g) Rolling circle amplification (RCA) product bound by fluorescent probes. h) Molecular beacons (MBs) conjugated with fluorophore and quencher emit fluorescence only upon hybridization with complementary target sequence of probe.

To reduce the number of FP-tagged proteins required at the target locus and to improve gene detection, proteins have been tagged with several FP molecules. The fusion of three copies of FPs (Figure 16b) has been shown to increase the signal intensity but a further increase of the copy number is challenging due to their relatively large size[201].

To label proteins with an increased number of FPs without drastically increasing their size, various techniques have been developed. Among them, peptide epitope tags have emerged as a convenient tool. Due to their small size, typically between 8-12 amino acids, they can be attached to proteins even in multiple copies without affecting protein folding, targeting or protein-protein interactions[202]. Most peptide epitope tags are antigenic peptide sequences that are bound by tag-specific antibodies. The most common peptide antigens include influenza hemagglutinin (HA), myelocytomatosis viral oncogene (Myc), simian virus 5-derived epitope (V5), the synthetic peptide FLAG and the synthetic streptavidin-binding strep-tag[202]. Due to the availability of many reliable high-affinity primary antibodies against these tags, they are widely used in immunoprecipitation and immunostaining assays. To achieve signal amplification for the detection of low abundant proteins in living cells, the supernova tagging system (SunTag) has been developed as a tool to recruit up to 24 FPs to a single protein (Figure 16c)[203]. In this system, the protein of interest is tagged with a repeating peptide array of 24 copies of the GCN4 peptide. The GCN4 peptide contains 19 amino acids and is derived from the general control noninducible 4 (GCN4) transcription factor. It is bound robustly and specifically by the scFv (GCN4) single-chain variable fragment antibody[204]. scFvs are small-sized artificial constructs composed of the variable regions of the heavy (V_H) and light (V_L) immunoglobulin chains, connected by a flexible peptide linker. This allows their expression in bacterial expression systems and mammalian cells[205,206]. By using fusions of scFv (GCN4) antibodies with FPs, proteins tagged with the GCN4 epitope array can be visualized with very high sensitivity[203,207]. By labeling dCas9 with the SunTag, non-repetitive regions of the MUC4 gene have been imaged with 20 different sgRNAs[208].

For direct labeling of small epitope tags without the need for antibodies, a split super-folder GFP (sfGFP) has been adapted to label proteins of interest by self-complementation (Figure 16d)[209]. For this, the sequence of sfGFP has been cut between the tenth and the eleventh β-strand of the β-barrel into a short 16 ammino acid peptide (GFP11) and a long fragment (GFP1-10). The GFP1-11 fragment is non-fluorescent by itself since the chromophore maturation requires a conserved amino acid located on

GFP11[210]. Both fragments are capable to efficiently self-complement without the need for other protein interactions when they are in close proximity. This results in a reconstituted sfGFP that becomes fluorescent after the chromophore maturation is completed[209,210]. By fusing tandem arrangements of GFP11-tags to a protein of interest, efficient fluorescence signal amplification has been achieved with a high signal-to-noise ratio[209]. This has been used to label repeats in MUC4 and 5S rDNA genes by fusing a repeating array containing 14 copies of GFP11 tags to dCas9[211].

In an alternative strategy, the sgRNA of the CRISPR-Cas9 system has been used for labeling, instead of tagging the protein. Here, the sgRNA has been appended with bacteriophage-derived RNA stem-loop motifs MS2 or PP7[200,212]. MS2 and PP7 are aptamers that are bound by the sequence-specific RNA-binding proteins MS2 capsid protein (MCP) and PP7 coat protein (PCP), respectively (Figure 16e)[213]. The introduction of tandem arrays of these aptamers into sgRNAs allowed the recruitment of multiple fluorescently tagged MCP or PCP proteins to each sgRNA. This resulted in a robust fluorescent signal amplification to monitor individual chromosomal loci with as few as four unique sgRNAs[200,214]. A related approach, termed Casilio, employs a Pumilio/Fem3 RNA-binding factor (PUF) which contains an RNA-binding domain that is programmable to bind a specific 8-mer RNA sequence (PUF-binding site (PBS)) (Figure 16f)[215]. By using a single gRNA appended with an array of 15 BPSs and fluorescently tagged PUF domains, a non-repetitive sequence within the MUC4 gene has been imaged[216].

Another tool for signal amplification is rolling circle amplification (RCA). Here a circular DNA molecule serves as a template for DNA or RNA polymerases to generate long, single-stranded DNA or RNA molecules with tandem repeats. These repeats can contain many different functional units like aptamers or hairpin structures that bind or can be bound by fluorescent probes (Figure 16g)[217]. Also, the incorporation of fluorescent dyes during RCA via fluorophore-conjugated dNTPs or the hybridization of fluorophore-tethered complementary strands to the RCA product allows easy visualization[218]. Similarly molecular beacons (MBs) (Figure 16h) can be used to visualize the RCA product with increased signal-to-noise ratio[219]. MBs are short, single-stranded bi-labeled fluorescent probes. They are held in a hairpin-loop conformation by complementary sequences at both ends of the probe. The 5' and 3' end of the probe each contain a reporter and a quencher molecule whose close proximity causes quenching of

the reporter fluorescence. The middle of the probe contains a DNA sequence complementary to a target sequence. Hybridization of the MB to the target sequence causes the opening of the hairpin-loop structure, which separates the reporter from the quencher and results in detectable fluorescence emission[220]. By combining MBs with the CRISPR-dCas9 system, sgRNAs containing unique MB target sequences have been used to image non-repetitive genomic loci as bright single foci[221,222].

Compared with FPs, organic dyes are generally brighter, more photostable and have a smaller size[167]. Therefore, they are typically used to fluorescently label antibodies. Common organic dyes for this are fluorescein isothiocyanate (FITC), cyanines (Cy2, Cy3, Cy5, Cy7) or Alexa dyes (Alexa Fluor 488, Alexa Fluor 594). The most applied method of antibody labeling is to use an unlabeled primary antibody to target a protein of choice and then targeting a fluorescently labeled secondary antibody to the primary antibody. In order to amplify multiple fluorescence signals, several signal amplification techniques have been developed. One method termed FRACTAL (fluorescent signal amplification via cyclic staining of target molecules) applied cyclic staining of fluorescently labeled secondary antibodies (Figure 17a)[223,224]. Here, the two secondary antibodies bind each other because they originate from host species that are each recognized by the other antibody. This allows alternate labeling between the two secondary antibodies, that can be repeated multiple times to amplify the intensity of the fluorescence signal. Other signal amplification strategies first label antigens with antibodies and then attach multiple fluorophores to or around these antibodies. In the tyramide signal amplification (TSA), proteins are labeled with horseradish peroxidase (HRP)-conjugated antibodies (Figure 17b)[225]. The HRP then catalyzes the conversion of tyramides that are fluorophore-labeled into a highly reactive form that binds covalently to tyrosine residues of proteins. This generates a high density of fluorophores around target proteins and amplifies the fluorescence signal. Another method employs avidin– or streptavidin-biotin complex formations to increase the number of fluorophores at a target protein[226]. Each avidin or streptavidin protein can bind four biotins per molecule with high affinity and selectivity. Typically, the employed secondary antibodies carry multiple biotin molecules. These are then either bound by large strept(avidin)-biotin complexes, linked through reporter enzymes (Figure 17c) or reporter conjugated strept(avidin) proteins. A limitation of this method is that despite blocking, endogenous biotin of the specimen can significantly increase the background signal.

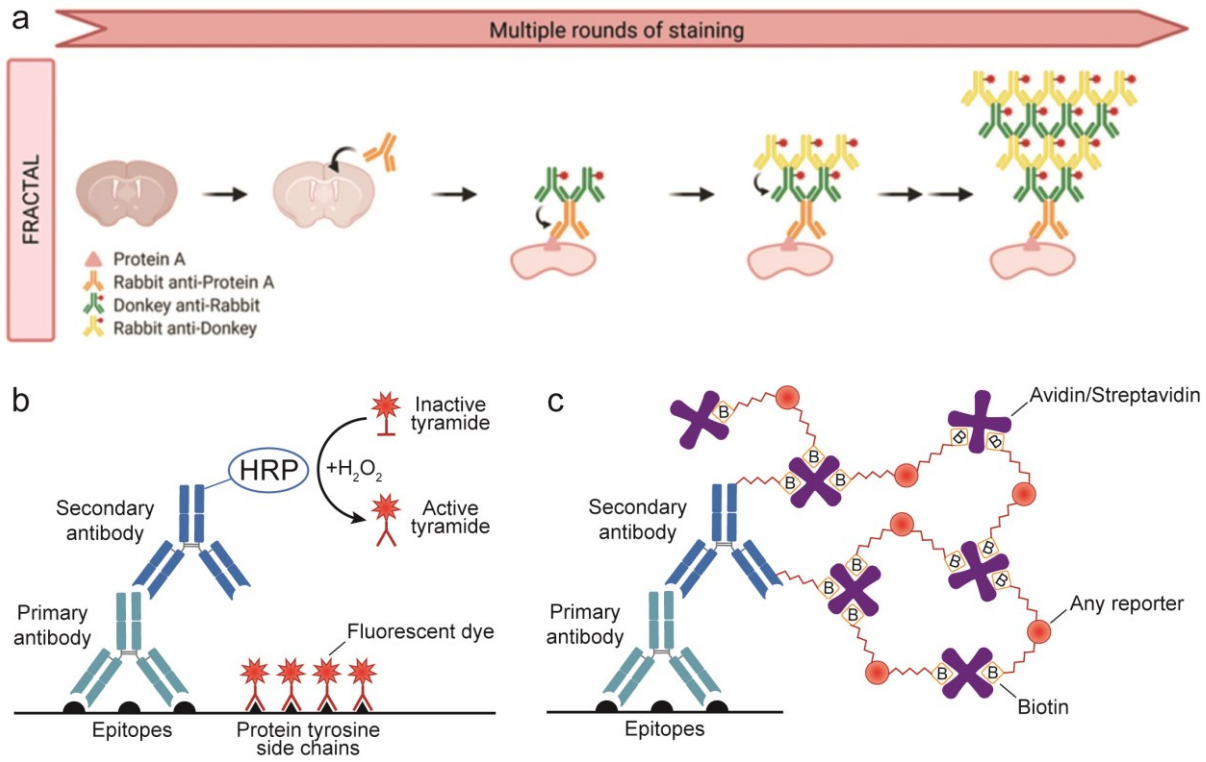


Figure 17. Antibody based signal amplification strategies. Schematics of a) fluorescent signal amplification via cyclic staining of target molecules (FRACTAL), b) tyramide signal amplification (TSA) and c) signal amplification via avidin– or streptavidin-biotin complex formations. a) Modified from (Yeon et al., 2022) with permission (8.3).

2.4.2 MUC4 Gene Locus

The MUC4 gene locus is a popular target for imaging studies because of the repetitive nature of sequences within one of its exons[227]. It encodes the high molecular weight protein MUC4, which is an integral membrane glycoprotein found on the cell surface. MUC4 is mainly expressed in airway epithelial cells for protection and lubrication, and plays important roles in cell signaling, proliferation and differentiation[228]. A variety of carcinomas show aberrant expression of MUC4, often associated with invasive proliferation of tumors[229,230]. The expression of MUC4 is highly regulated by epigenetic mechanisms such as DNA methylation in the promoter region. In tumor patients, MUC4 hypermethylation has been associated with a better overall survival as compared to patients with hypomethylation[231].

The MUC4 gene contains 26 exons (Figure 18a). The first exon comprises the amino-terminal 5'-untranslated region (UTR) of the protein, the second is polymorphic and

encodes the large central repeat domain and the 24 other exons code for the extracellular, transmembrane and carboxyl tail of the protein (Figure 18b). The largest exon 2 contains a 7.5 to 19 kb repeat region with a variable number (>100) of 48 nt tandem repeats[227]. These repeats encode the highly glycosylated central domain of MUC4, which consists of a variable number of tandem repeats rich in serine, threonine, and proline residues[228]. Due to the repetitive nature of the MUC4 locus, it has been a target to study the imaging capabilities for low-repetitive loci by using programmable DNA-binding domains such as TALEs and CRISPRS-dCas9 system[195,200,232].

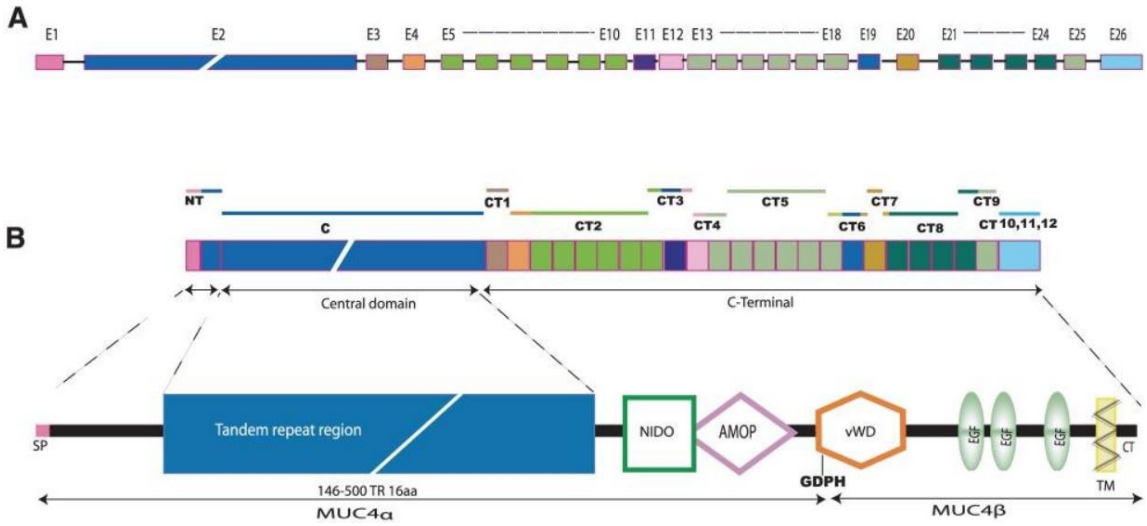


Figure 18. Structure of MUC4 gene and protein. a) 26 exons (E1-E26) of the MUC4 gene. b) The MUC4 protein is divided into an N-terminal (NT), a central domain (C), and a C-terminal (CT) region. The central domain contains the large tandem repeat region. The C-terminal region encodes 12 domains (CT1–CT12), including the nidogen (NIDO)-like, adhesion-associated domain (AMOP), von Willebrand factor (vWD) and 3 carboxyl-terminal epidermal growth factor (EGF)-like domain. The MUC4 protein can be cleaved at GDPH, generating the heavily glycosylated mucin-like subunit MUC4α and the growth factor-like subunit MUC4β. MUC4 is anchored to the cell surface by the transmembrane region (TM). Adapted from (Chaturvedi et al., 2010) with permission (8.3).

3. Aim of the Work

The epigenetic modification 5mC plays essential roles in important cellular processes such as the regulation of gene expression, DNA-protein interactions and chromatin structure[233]. Because methylation patterns are dynamically influenced by developmental and environmental processes, they can vary in different cells and under different environmental conditions[234]. Dysregulation of methylation can lead to aberrant methylation patterns which have been found in many human diseases, including cancer[235]. To better understand the role of 5mC in diseases and other regulatory processes, it is of great interest to study methylation patterns locus-specifically in single cells with high resolution. Previous studies have developed strategies for the *in situ* imaging analysis of cellular 5mC, providing the opportunity to examine direct correlations between locus-specific DNA methylation and other chromatin features. One of these approaches employs FISH probes that carry chelator linkers to enable osmium-mediated cross-linking with 5mC[91,236]. Another method is based on fluorescence complementation between methyl-CpG-binding domains and programmable DNA-binding proteins, each fused to one of the units of a split fluorophore[94]. Although these methods allow imaging of cellular 5mC with locus resolution, they do not provide single nucleotide resolution. In order to overcome this limitation, TALE proteins have recently been utilized as 5mC-sensitive probes in imaging studies. Because TALEs feature a modular repeat-based DNA-binding domain that allows incorporation of 5mC-sensitive repeats, they provide sequence- and 5mC selectivity within a single modular scaffold. In co-staining assays with fluorescently labeled TALEs, this enabled the analysis of cellular 5mC in pericentromeric SATIII DNA[99,100]. However, this method has so far been limited to highly repetitive satellite DNA sequences.

To expand the application scope of this approach, this work aims to develop strategies to enhance the imaging sensitivity with TALEs in order to enable the analysis of 5mC at single gene loci. To develop a straightforward signal amplification strategy compatible with the TALE-based imaging approach, different signal amplification strategies will be evaluated. For this, TALEs will be labeled with repeat arrays of various fluorophore and epitope tags. These will be tested in cell stainings along with a variety of nanobody and antibody signal amplification strategies against the employed tags. The most effective

strategy will then be applied to enable the visualization of only a small number of target sequences in the low-repetitive MUC4 gene locus. To further analyze 5mC at these sequences, co-staining experiments will be performed with a 5mC-selective TALE and a universal control TALE, both targeting the same low-repetitive target sequence. Each TALE will thereby be labeled with a different fluorescent tag whose fluorescence will be enhanced by the developed signal amplification strategy. To enable the detection of 5mC, the signal intensities of both TALEs will be analyzed. This will be applied to compare methylation levels in the MUC4 gene locus between two cell lines with different 5mC perturbations. The successful application of this strategy for a particular low number of target sequences will open up perspectives for the *in situ* analysis of 5mC at coding, non-repetitive gene loci.

4. Results and Discussion

4.1 Fluorescence Signal Amplification Strategies

Previous studies have used fluorescently labeled TALE proteins for the imaging-based analysis of 5mC at highly repetitive satellite DNA sequences[99,100]. These clustered repeats provide straightforward imaging targets, as they allow the recruitment of multiple TALE proteins to generate fluorescent foci with high signal-to-noise ratios. By contrast, single gene loci do not provide repetitive sequences and are thus more challenging to visualize in imaging studies. Therefore, different signal amplification strategies were evaluated to improve the imaging sensitivity and signal-to-noise ratio in order to extend the TALE-based imaging approach for the analysis of 5mC in single gene loci. As reference target, the human SATIII locus was selected, based on its known staining behavior and available reference data. SATIII DNA is located in the pericentromeric regions of chromosomes and contains clustered tandem repeat sequences that vary in length depending on the chromosome[237]. To target this locus, a TALE, termed TALE_SATIII was designed to bind the 18mer sequence ''TGGAACGGAATGGAATGG''. This sequence is present in high copy number in many of the clustered repeats within the SATIII loci, with a maximum of 28k repeats in chromosome 9 (see Table S.1 for copy number analysis of each chromosome). In order to evaluate various signal amplification methods, versions of TALE_SATIII tagged with different N-terminal tags for fluorescence signal amplification were designed (Figure 19a).

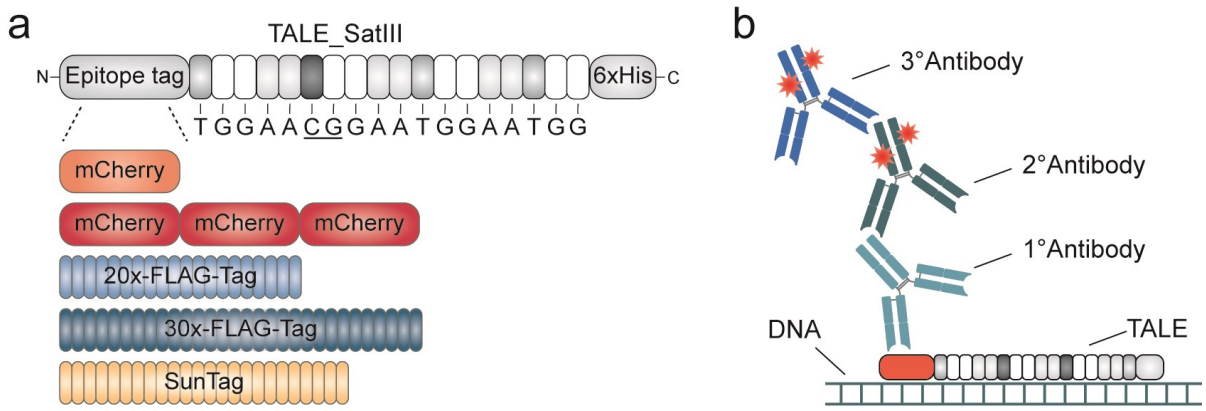


Figure 19. Fluorescence signal amplification scheme. a) TALE structure with different N-terminal epitope tags. b) TALE epitope tag immunostaining with fluorescently labeled antibodies.

A common approach to increase the sensitivity for imaging is to multimerize the number of FPs localized at the target site. A simple technique to achieve this is to label proteins with multiple FPs. Therefore, initially TALE_SATIII versions were designed, carrying either one (1x) or three (3x) N-terminal mCherry proteins (Figure 19a). Since the target sequence of TALE_SATIII contains a single CpG in the 5' region, TALE_SATIII versions were designed with the universally binding G* repeat opposite the C of the CpG to avoid interference of TALE-binding by potentially methylated sequences. To compare different signal amplification strategies, homogeneous staining of all samples and precise control over the TALE protein concentration was required. For this reason, all TALE proteins were recombinantly expressed in *E.coli*, purified by His-tag affinity enrichment and used to stain fixed and permeabilized cells. To verify the ability of the 1x- and 3x-mCherry TALE_SATIII versions to bind and fluorescently label genomic SATIII DNA, U2OS (Human Bone Osteosarcoma Epithelial) cells were stained and analyzed by fluorescence microscopy. Characteristic SATIII staining patterns with expected foci number and morphology[100] were observed, demonstrating the ability of both TALEs to label genomic SATIII DNA in a cellular context (Figure 20).

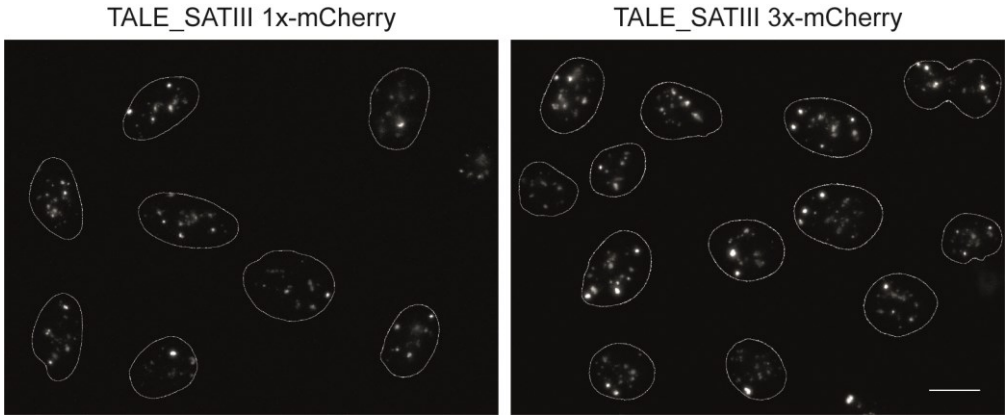


Figure 20. Representative images of SATIII-foci in nuclei of U2OS cells. Cells were stained with TALE_SATIII labeled with 1x-mCherry (left) or 3x-mCherry (right). Scale bar: 15 µm

Analysis of the fluorescence intensity did not show significant differences (1.4-fold) between foci obtained with the 1x- and 3x-mCherry versions of TALE_SATIII (Figure 21b). An explanation for this could be the reduced accessibility of binding sites due to the bulky 3x-mCherry tag, which increases the molecular weight of the protein by 54.4 kDa compared to the 1x-mCherry tag (TALE_SATIII 1x-mCherry: 102.7 kDa, TALE_SATIII 3x-mCherry: 157.1 kDa). This would reduce the number of TALE proteins present at the SATIII loci and therefore not significantly increase the actual number of

FPs present at the target sites. Therefore, additional immunostaining was performed to enhance the signal of the two TALE-mCherry versions (Figure 19b). The fluorophore mCherry emits light in the range of 550-650 nm with an emission maximum at 610 nm[238]. To contribute to this emission, various nanobodies and antibodies labeled with fluorophores that emit light in a similar spectral range were selected. An initially tested anti-mCherry nanobody conjugated with the fluorescent dye ATTO594 (RFP-Booster) did not significantly enhance the fluorescence for any of the TALEs (Figure 21b). However, the combination of a primary anti-mCherry antibody and a secondary antibody labeled with Alexa Fluor plus 594 significantly enhanced the signal for the 1x- and 3x-mCherry TALE version. Compared to TALE staining without immunostaining, a 4.1-fold increase of the absolute fluorescence intensity was obtained for the 1x-mCherry construct and a slightly higher increase of 4.8-fold for the 3x-mCherry construct (Figure 21b). Furthermore, additional immunostaining with a tertiary antibody labeled with the same fluorophore, further increased the absolute fluorescence by 2.1-fold for both mCherry TALE versions. Compared to TALE staining without immunostaining, this resulted in an 8.8-fold and 10.3-fold increase for the 1x- and 3x-mCherry TALE version, respectively (Figure 21b).

Based on these promising results, additional epitope tags were tested for signal amplification by immunostaining. Due to their small size, peptide epitope tags can be attached to proteins in multiple copies without affecting protein folding and targeting[202]. This should allow multimerization of tag-specific antibodies at the labeled protein. Because of its high specificity and the possibility to implement it in a tandem formation[239], the synthetic peptide FLAG was chosen to test this hypothesis. To verify the optimal number of FLAG tag repeats for signal amplification, repeat versions bearing either 20 or 30 concatenated FLAG tags were constructed and fused N-terminally to TALE_SATIII (Figure 19a). In order to avoid steric blocking of adjacent epitopes, the individual FLAG tags were separated by a short linker (see 8.1.4 for sequence information). This should provide sufficient space to allow high saturation of antibodies. Similarly, TALE_SATIII was constructed carrying an N-terminal SunTag consisting of 24 copies of the GCN4 peptide separated by a short linker (Figure 19a, see 8.1.5 for sequence information). As previously, FLAG tag and SunTag TALE_SATIII constructs were recombinantly expressed, purified and used for the staining of U2OS cells. Subsequent immunostaining was performed using primary antibodies against the

respective epitope tags. As secondary and tertiary antibodies, the same Alexa Fluor plus 594 labeled antibodies from previous mCherry stainings were used to allow comparison between conditions. All TALE constructs afforded characteristic SATIII staining patterns as observed with the mCherry constructs (Figure 21a). Fluorescence signal analysis revealed increased intensities for both FLAG tag constructs as well as for the SunTag construct (Figure 21b). Interestingly, the FLAG tag constructs did not show intensity differences between the 20x- and the 30x-FLAG tag version. Both constructs in combination with two antibodies showed an about 3.5-fold increased signal compared to the tested mCherry constructs without immunostaining. However, compared to the mCherry constructs in combination with two antibodies, the fluorescence signals were about 1.4-fold lower (Figure 21b). Additional staining of the two FLAG constructs with the tertiary antibody increased the absolute fluorescence about 2.6-fold compared to immunostaining with two antibodies. This resulted in similar signal intensities as the same signal amplification strategy with the mCherry constructs (Figure 21b). However, large intensity differences between individual foci were observed for both FLAG constructs (Figure 21a). This could be attributed to an unequal accessibility of the FLAG constructs to SATIII loci in different chromosomes. The SunTag construct afforded similar signal intensities as the mCherry-based stainings in combination with two or three antibodies, respectively. In comparison to the staining with two antibodies, the additional staining with the tertiary antibody resulted in a 2.5-fold increased signal, comparable to the observed increases with the mCherry constructs (Figure 21b).

These results indicate that epitope tag mediated antibody staining contributes strongly to the fluorescence signal, especially after multiple rounds of immunostaining. This is likely due to multimerization of antibodies into large complexes that contain a vast number of fluorophores. By contrast, an increase in the number of epitope tags did not contribute significantly to the signal intensity. One reason for this could be a reduced accessibility of genomic TALE binding sites, either due to a restricted mobility of the TALEs or due to steric blocking of adjacent binding sites by already bound TALEs. Both plausible explanations are consequences of the increased size of epitope tags that contain a larger number of repeats.

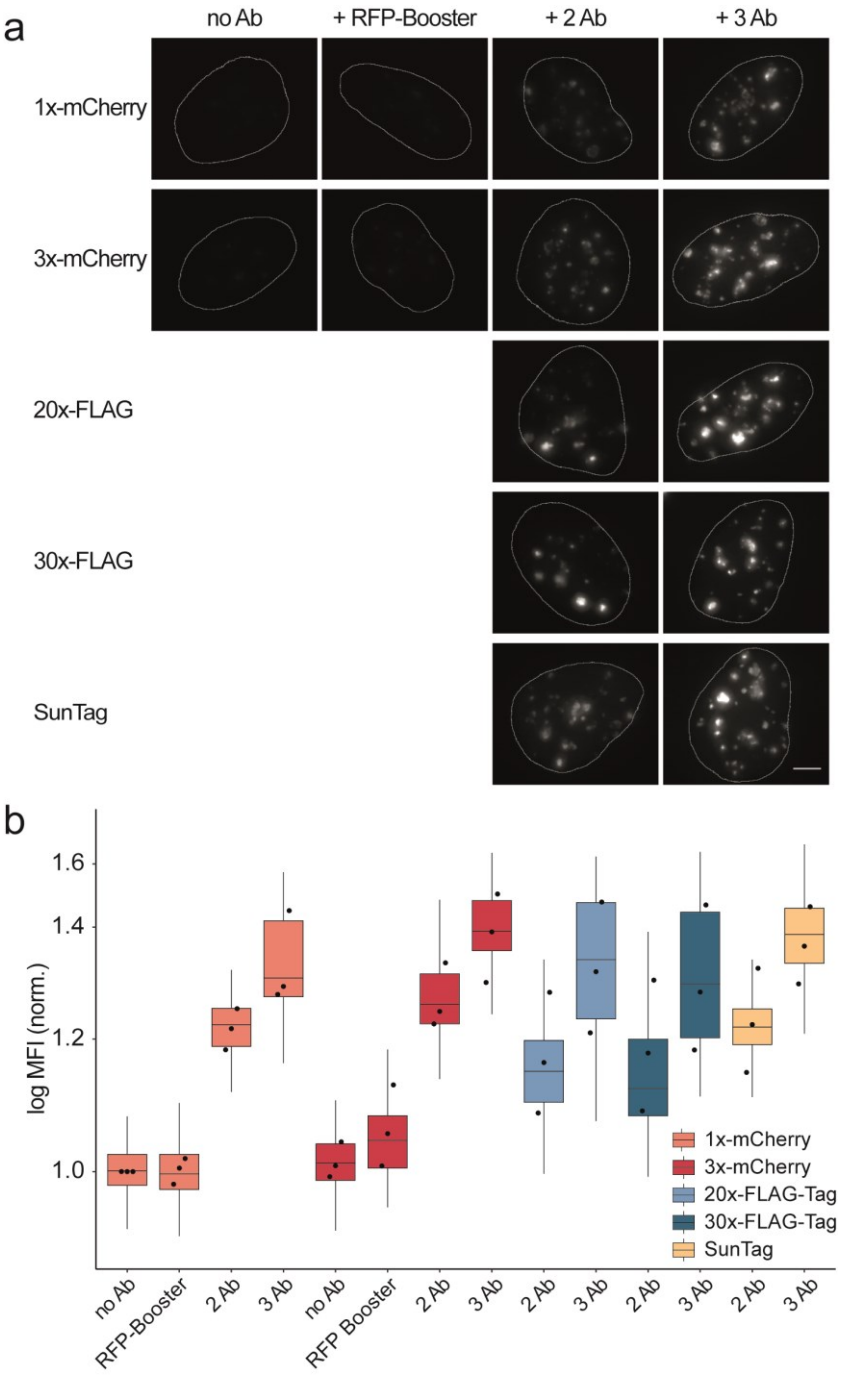


Figure 21. TALE_SATIII staining with different epitope tag and immunostaining combinations. a) Representative images of foci in nuclei of U2OS cells. All images were acquired under identical imaging conditions set at low light intensity and short exposure times to correctly visualize SATIII at the highest signal values. No Ab: No antibody, RFP Booster: RFP-Booster ATTO594, 2 Ab: Primary antibody against epitope tag and secondary Alexa Fluor plus 594 labeled antibody, 3 Ab: Primary antibody against epitope tag, secondary Alexa Fluor plus 594 labeled antibody and tertiary Alexa Fluor plus 594 labeled antibody. Scale bar: 5 µm. b) Mean fluorescence intensities of foci from signal amplification conditions shown in Figure 21a normalized to 1x-mCherry no Ab condition. Dots represent normalized mean values of each experiment with N = 3 totaling 2802 cells and 44038 foci.

To further extend the successful signal amplification strategies to image a genomic locus with a reduced number of target sequences, the MUC4 locus was selected as target locus. This locus contains a central repeat domain consisting of a variable number of tandem repeat sequences. Due to the repetitive nature of sequences, it has also previously been a target in TALE- and CRISPR-dCas9-based imaging studies of low-repetitive genomic loci[195,200]. In this work, the number of theoretical repeat sequences targeted by TALEs is based on the latest genome assembly T2T-CHM13v2.0q[240]. This assembly was generated primarily by long-read sequencing to achieve a complete, gap-free sequence for all 24 human chromosomes. Compared to the previous GRCh38 assembly[241], it contains nearly 200 million base pairs of new DNA sequences, including complex regions such as repetitive loci. In case of the MUC4 locus, the T2T-CHM13v2.0q assembly afforded in an increased number of actual repeat sequences with approximately twice as many repeat sequences compared to the GRCh38 assembly (see Table S.2 for number of theoretical repeat sequences in T2T-CHM13v2.0 and GRCh38 genome assembly). This influences also previous studies based on the MUC4 gene locus.

To study the minimum number of target sequences required for the detection of distinct foci, mCherry-fused TALEs M1-M3 (M for MUC4) were designed with 148, 79 and 32 theoretical repeat target sequences at the MUC4 locus, respectively (Figure 22a). As before, TALEs M1-M3 were recombinantly expressed and tested in imaging studies with U2OS cells. All three TALEs showed two foci per cell after signal amplification with two antibodies, which is consistent with the expected staining pattern of a single gene locus in diploid cells (Figure 22b). Accordingly, dividing cells showed four foci per cell (Figure S.1). To verify selective staining of the MUC4 locus, co-stainings were performed with eGFP labeled versions of TALE_M1-3. After respective immunostaining all possible combinations of the three TALEs showed colocalization of the foci, indicating selective targeting of the MUC4 locus by the employed TALEs (Figure 22c).

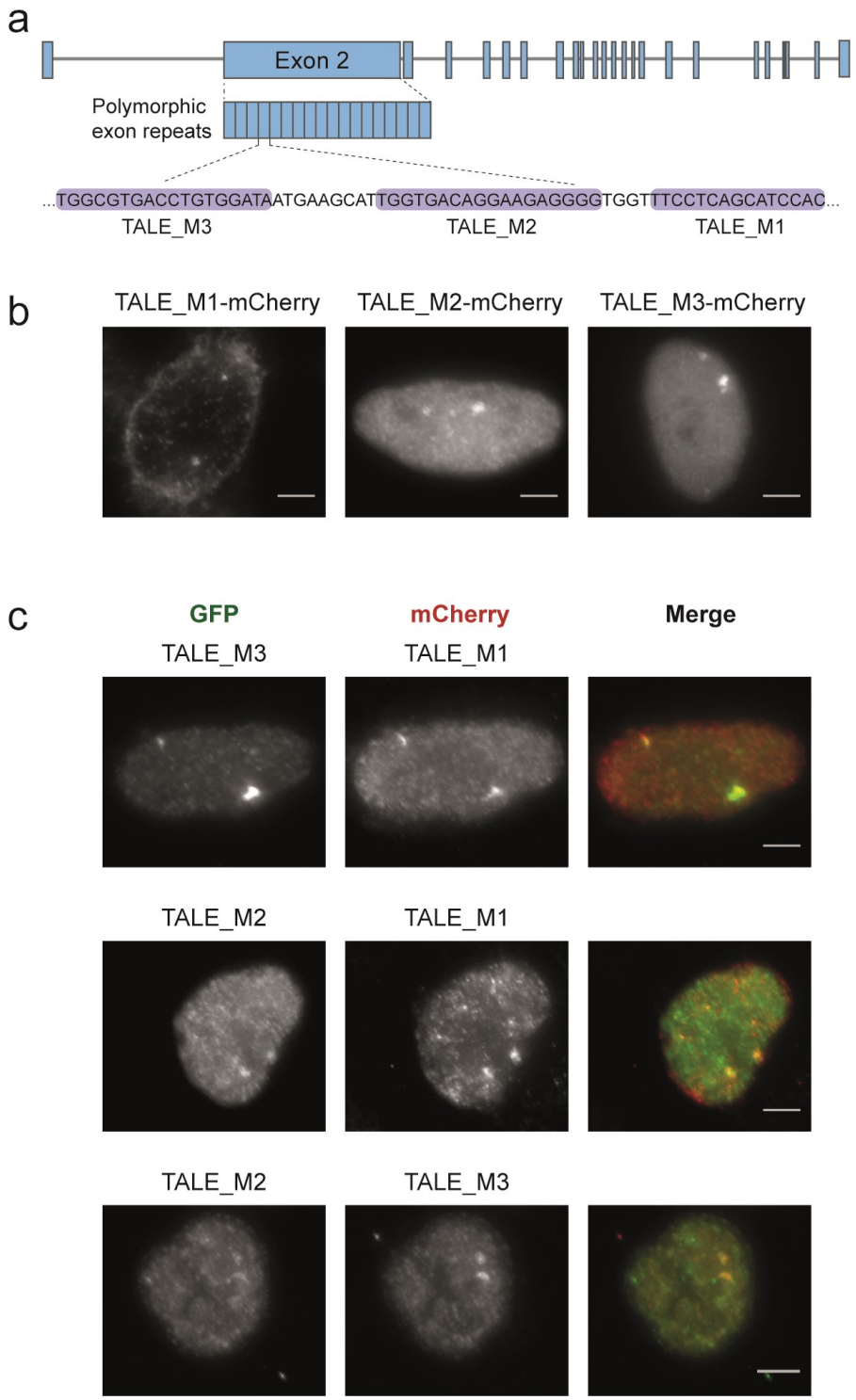


Figure 22. Staining of low repetitive TALE target sequences at the MUC4 locus. a) TALE target sequences (purple) in polymorphic exon 2 repeats of MUC4 locus. b) MUC4-characteristic foci in U2OS cells stained with TALE_M1-3 targeting 148, 79, and 32 target sequences at MUC4 locus, respectively (stains were conducted with TALE_M1–3x-mCherry, TALE_M2–1x-mCherry or TALE_M3–1x-mCherry, followed by immunostaining with a primary anti-mCherry and a secondary Alexa Fluor plus 594 labeled antibody. c) Co-stainings of mCherry labeled TALEs from 22.b and eGFP labeled versions of TALE_M1-3, followed by immunostaining with Mouse anti eGFP, Goat anti-Mouse Alexa Fluor 488 and Rabbit anti-mCherry, Goat anti-Rabbit Alexa Fluor Plus 594. U2OS cells, Scale bars: 5 µm.

In contrast to the highly repetitive SATIII DNA, the MUC4 locus contains a greatly reduced number of repeat sequences. Therefore, its visualization requires much higher sensitivity and lower background signals. To reevaluate previous signal amplification strategies for the visualization of low-repetitive single gene sequences, TALE_M3 was selected since it targeted the lowest number of repeats to visualize the MUC4 locus (32 theoretical repeat sequences). As before, 1x-mCherry, 3x-mCherry, 20x-FLAG, 30x-FLAG and SunTag versions of TALE_M3 were generated and analyzed by fluorescence microscopy.

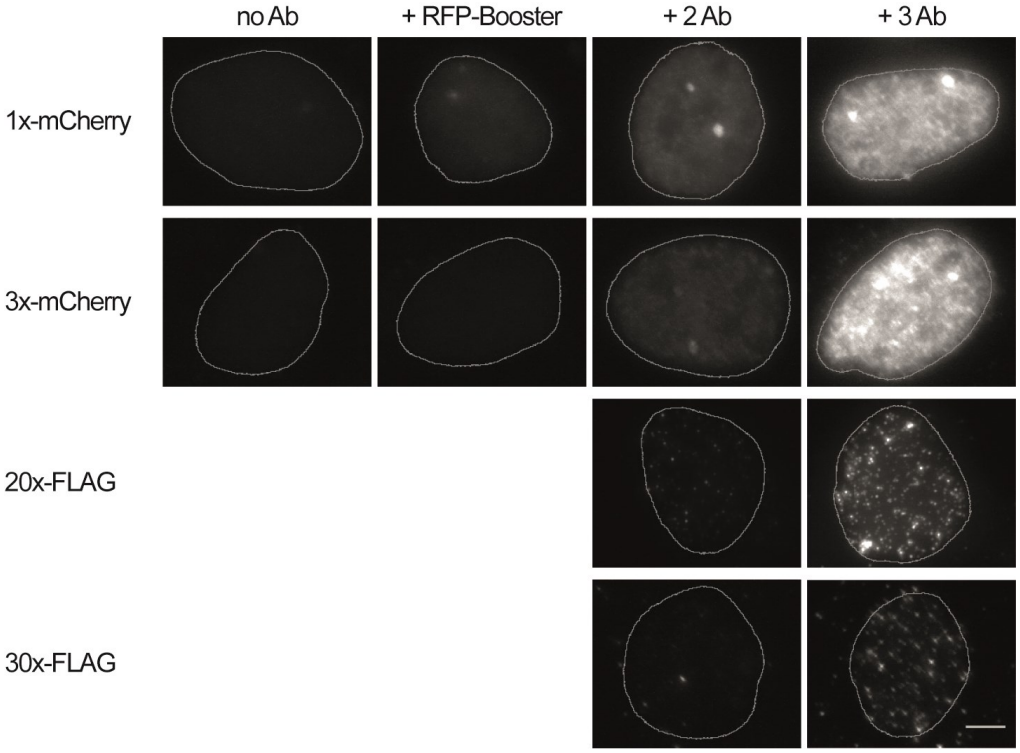


Figure 23. Staining experiments using TALE_M3 bearing different epitope tags in combination with different immunostainings. Representative images of foci in nucleus of U2OS cells are shown. All fluorescence images were acquired under identical imaging conditions. No Ab: No antibody, RFP-Booster: RFP-Booster ATTO594, 2 Ab: Primary antibody against epitope tag and secondary Alexa Fluor plus 594 labeled antibody, 3 Ab: Primary antibody against epitope tag, secondary Alexa Fluor plus 594 labeled antibody and tertiary Alexa Fluor plus 594 labeled antibody. Scale bar: 5 µm.

Due to insufficient signals, no fluorescent foci could be observed for the 1x- and 3x-mCherry constructs without immunostaining (Figure 23). Additional staining with the RFP-Booster did not significantly amplify the signal, however, foci with a low signal intensity could be detected in some of the stained cells (Figure S.2). By contrast, additional immunostaining of the 1x- and 3x-mCherry constructs with two antibodies

resulted in a marked increase in fluorescence intensity. This allowed the detection of two foci in the majority of cells with a high signal-to-noise ratio (Figure 23, S.2). Surprisingly, weaker signals were observed for the 3x-mCherry construct than for the 1x-mCherry construct (Figure 23, S.2). The additional immunostaining with the tertiary antibody revealed a granular nuclear staining. Due to this high background, no distinct MUC4 foci could be quantified under this condition. Unexpectedly, TALE staining with the 20x or 30x FLAG construct in combination with two antibodies did not yield sufficient signals for foci detection. Therefore, although high fluorescence intensities were observed for the SATIII locus, no foci could be detected for the MUC4 locus under this condition. Additional staining with the tertiary antibody again resulted in granular nuclear staining as observed with the mCherry constructs, preventing the detection of distinct MUC4 foci (Figure 23). In case of the SunTag system, off-target staining was observed after immunostaining. This off-target staining showed a MUC4-like staining pattern of a single focus that was not distinguishable from the actual MUC4 foci during foci quantification. Basic local alignment search tool analysis of the GCN4 epitope revealed 88 % identity with the centrosomal protein centriolin. Therefore, co-staining experiments were performed with an anti-centriolin antibody. These confirmed off-target binding of the employed anti-GCN4 antibody to centriolin (Figure S.3). Because of this reason, the SunTag system was excluded from further analysis.

According to these findings, some of the signal amplification strategies that showed successful results for the highly repetitive SATIII locus are ineffective for the detection of the low repetitive MUC4 locus. As opposed to the SATIII experiment, specifically staining with the TALE-constructs carrying either 3x-mCherry, 20x, or 30x-FLAG-tags did not provide the required sensitivity for accurate detection of MUC4 loci. One reason for this might be a difference in accessibility of the two loci by these TALEs. The MUC4 gene encodes a mucin that is expressed by epithelial cells of the respiratory tract, cervix, and colon, but not by bone cells[227]. For this reason, the MUC4 locus in U2OS cells is most likely transcriptionally repressed and present as tightly packed heterochromatin that is less accessible to DNA binding proteins. Compared to the 1x-mCherry construct, the TALE construct labeled with 3x-mCherry has a significantly higher molecular weight (102.7 kDa versus 157.1 kDa), whereas the 20x- and 30x-FLAG tags likely do not adopt an ordered structural conformation. This could impair the mobility of the TALEs during the search mode and prevent their access to less accessible genomic targets. The

repetitive nature of the TALE_SATIII target sequence allows more flexibility in terms of targeting, as TALE_SATIII can access multiple copies of the target sequence at different genomic locations. For TALE_M3, however, the smaller number of target sequences at the single MUC4 gene locus means that TALE_M3 is more likely to be impaired by a low accessibility of the locus. By contrast, staining with 1x-mCherry-labeled TALE_M3 in combination with the two-antibody signal amplification strategy provided a high signal-to-noise ratio and homogeneous staining pattern. This allowed sensitive and precise detection of MUC4 foci in the majority of cells. For this reason, this strategy was selected for further experiments.

4.2 Visualization of 5mC-Differences at the Low-Repetitive Genomic Locus MUC4

In order to visualize 5mC differences at the MUC4 target sequence, a previously developed TALE co-staining approach[99,100] was applied and combined with the developed two-antibody signal amplification strategy. This TALE co-staining approach is based on the sensitivity of the RVD HD towards 5mC and the ability of the RVD G* to recognize any nucleobase (including C and 5mC). For the imaging-based analysis of 5mC, cells are co-stained with two TALEs that target the same target sequence but carry either the selective RVD HD or the universal RVD G* opposite the C nucleobase of the CpG dyad in the target sequence. Both TALE versions are fused to different fluorescent proteins, which allows the detection of 5mC at the target sequence by analyzing the signal intensities of both fluorophores. To apply this strategy, mCherry and eGFP versions of the MUC4-directed TALE_M3 were assembled, bearing either an HD or G* repeat opposite the C at position 4 of the target sequence "TGGCGTGACCTGTGGATA". For signal amplification of eGFP TALEs, the analogous two-antibody signal amplification strategy was applied as for mCherry TALEs. For this, eGFP TALEs were immunostained with a primary anti-eGFP antibody and a secondary antibody labeled with Alexa Fluor 488, which emits fluorescence in the same spectral range as eGFP. This resulted in very similar staining patterns and signal-to-noise ratios as the ones obtained for the mCherry TALEs (Figure S.4). Co-staining experiments with mCherry and eGFP labeled TALE_M3 in combination with immunostaining exhibited clear colocalization, indicating selective

recognition of the MUC4 target sequence by both TALEs (Figure 24). Importantly, no nonspecific off-target binding was observed for the employed antibodies (Figure S.5).

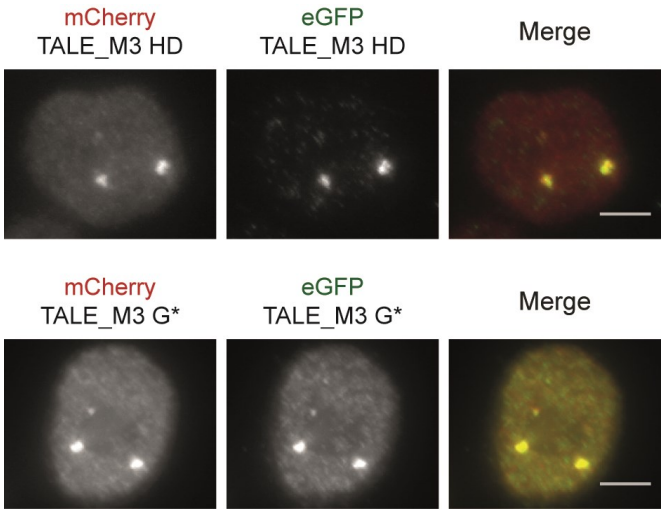


Figure 24. Co-staining of TALE_M3 HD and G* versions labeled with mCherry and eGFP. Cells were stained with primary anti-mCherry and anti-eGFP antibodies and secondary antibodies labeled with Alexa Fluor plus 594 and Alexa Fluor 488, respecively. HCT116 wt cells, scale bars: 5 µm.

For the visualization of 5mC, the human HCT116 colorectal carcinoma cell line was selected due to profound differences in methylation levels between wild type (wt) cells and cells with genetic double-knockouts (DKO) of DNA methyltransferases DNMT1 (-/-) and DNMT3b (-/-). The latter cell line exhibits almost no methyltransferase activity, resulting in a >95 % reduced genomic methylation level compared to the wild type[242]. To quantify the methylation level of both cell lines at the employed target loci, methylated-DNA-immunoprecipitation (MeDIP) experiments followed by real-time quantitative PCR (qPCR) analysis were performed. For the SATIII locus a relatively small (<2-fold) 5mC difference was observed between the two cell lines, whereas for the MUC4 locus an 11.5-fold higher 5mC level was detected for the HCT116 wt cells (Figure 25). Bisulfite sequencing data of the TALE_M3 target sequence at the MUC4 locus confirmed the absence of methylation in HCT116 DKO cells, whereas methylation was detected in HCT116 wt cells (Figure S.6).

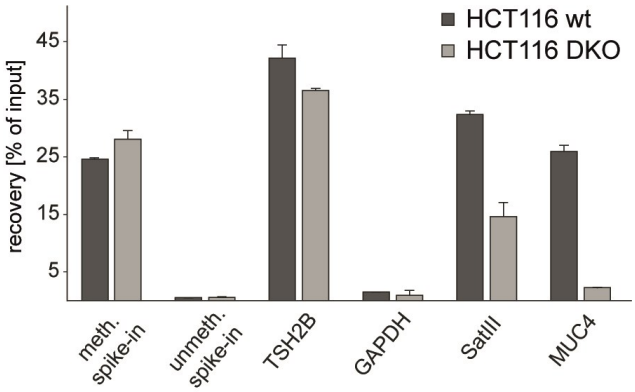


Figure 25. MeDIP analysis of wt and DKO HCT116 cells. Relative recoveries of target DNA from both cell types were assessed by qPCR. Results show means ± SD of technical MeDIP duplicates and qPCR triplicates. Shown are results for methylated and unmethylated spike in controls, for an endogenous control locus with high (TSH2B) and low (GAPDH) 5mC level, and for the SATIII and MUC4 target loci.

For the imaging-based analysis of 5mC, co-stainings of both cell lines were conducted with mCherry- and eGFP-TALE pairs being either both G* TALEs (Figure 26a), or the mCherry version being a G* TALE and the eGFP version being an HD TALE (Figure 26b). Methylation of the target sequence reduces the binding affinity of HD TALEs and could therefore prevent the accurate detection of foci. Therefore, the red channel was chosen for foci detection in image analysis since the universally binding G* TALEs were labeled with mCherry in both co-staining conditions (Figure S.7). Based on the identified foci locations, fluorescence signals were recorded for the red and green fluorescence channel. For ease of comparison, the signals from each TALE were normalized to the mean fluorescence of the foci detected in HCT116 DKO cells.

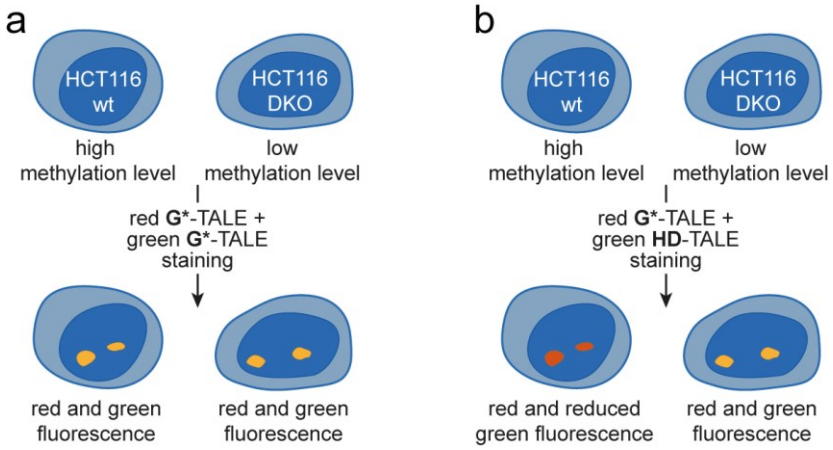


Figure 26. Illustration of TALE co-staining experiment performed with wt and DKO HCT116 cells. a) Co-staining with red G* TALE and green G* TALE. b) Co-staining with red G* TALE and green HD TALE.

In the first co-staining, the signals from both G* TALEs did not differ between the two cell types (Figure 27a). By contrast, in the second co-staining, the HD TALE showed a significantly lower signal in the wt cells, whereas the G* TALE again did not differ between the two cell types (Figure 27a). In general, the signals of mCherry TALEs did not strongly differ between the two cell types even as HD version (Figure S.8a). Therefore, mCherry TALEs were only used as marker TALEs for the MUC4 locus. The eGFP TALE however, showed a decreased signal in the HCT116 wt cells only as HD version but not as G* version. This indicates a selective response of the HD TALE to 5mC differences itself, since the similar signals of the G* TALE indicate similar target accessibility for both cell types (Figure 27a, Figure S.8a). Therefore, the observed signal difference of the HD TALE can be attributed to methylation level differences between the two cell lines, which was in agreement with the 5mC differences observed by MeDIP and Bisulfite sequencing analyses (Figure 25, Figure S6). Interestingly, a similar response of the HD TALE was also observed in the nuclear background (Figure S.9). This suggests a 5mC selectivity also in respect to off-target sequences. To verify this, the nuclear background fluorescence of mCherry-labeled TALEs that did not exhibit fluorescent foci was analyzed. These TALEs were designed to target the MUC4 locus but were identified as weak binders because they did not provide sufficient signal-to-noise ratios for foci detection. Significant differences in the absolute nuclear background intensity were observed between these TALEs, indicating a different number of off-target sequences for each TALE (Figure S.10). Interestingly, for TALEs targeting CpG-containing target sequences, only the HD TALE versions showed increased nuclear background intensities in HCT116 DKO cells compared to the HCT116 wt cells. These results are similar to those observed with TALE_M3 and can be explained by a response to 5mC in off-target sequences that contain a CpG opposite the HD position. For TALEs that were designed to target sequences without CpG, three of four TALEs showed similar nuclear background signals in both cell lines, possibly due to the absence of 5mC in the off-target sequences. However, one TALE showed increased nuclear background signals for HCT116 DKO cells, which could be explained by the presence of 5mC in off-target sequences in HCT116 wt cells.

Similar costaining experiments were further conducted with the corresponding TALE_SATIII pairs. Here, none of the TALEs exhibited marked differences between the two cell lines (Figure 27b, Figure S.8.b). This can be explained by the comparatively small 5mC difference in the SATIII locus between the two cell lines, as revealed by MeDIP analysis (Figure 25). In addition, possible heterogeneity in 5mC levels between the repetitive SATIII target sequences complicates the detection of 5mC differences in this locus.

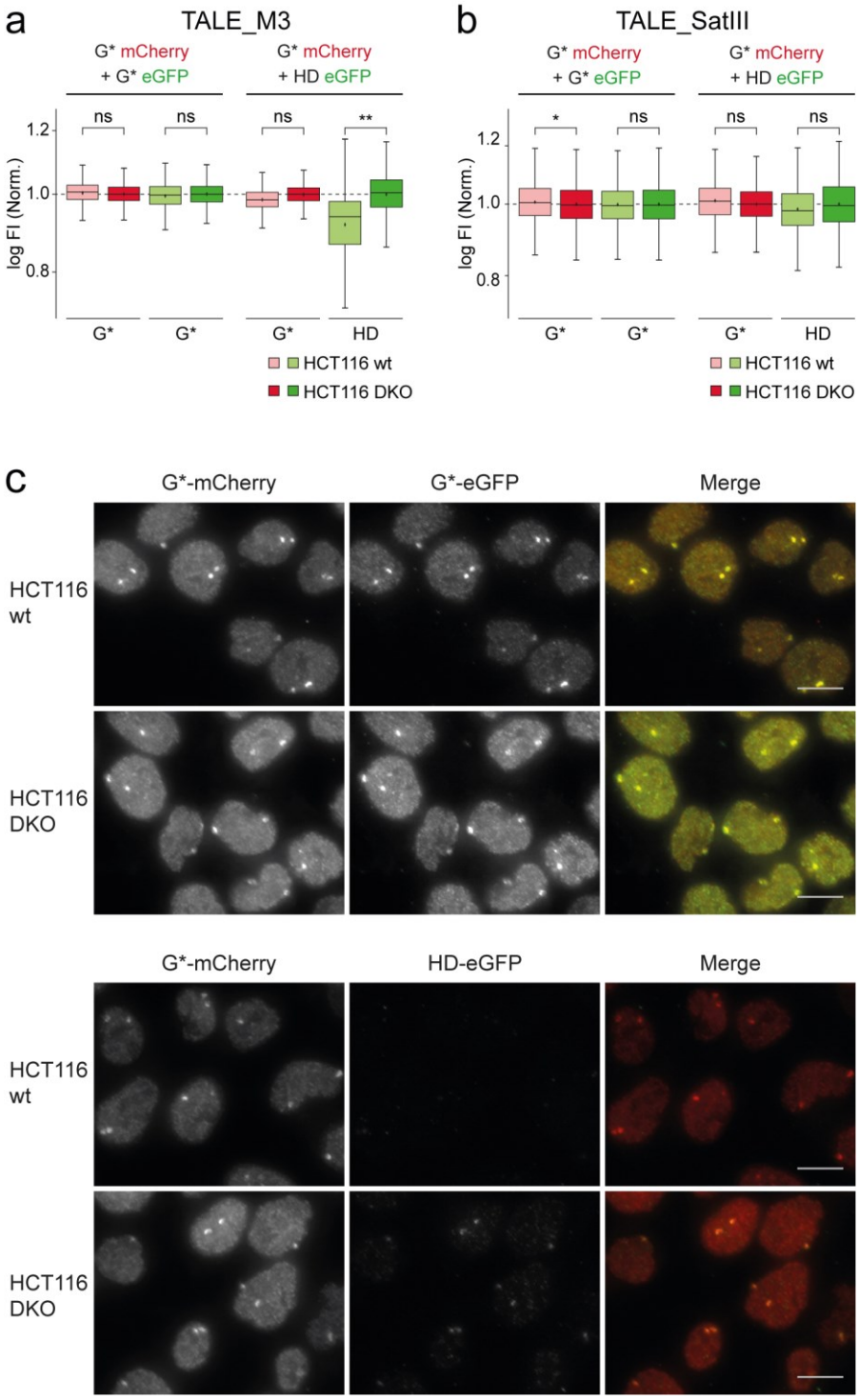


Figure 27. Imaging-based in situ analysis of 5mC by staining with TALE probes. a-b) Fluorescence signal intensities (FI) of foci from HCT116 wt and DKO cells co-stained with G* and HD TALE_M3 (a) or TALE_SATIII versions (b) fused to mCherry or eGFP and immunostained with primary anti mCherry, secondary Alexa Fluor Plus 594, primary anti-eGFP and secondary Alexa Fluor 488. For each TALE, log FI of each focus is normalized to the mean of log FI of all foci from HCT116 DKO cells. N = 4 experiments totaling >2000 foci per condition (a), and >7000 foci per condition (b). P < 0.1*, P < 0.01**(Student's t test). c) Representative images of co-staining (a) with TALE_M3 1x-mCherry and TALE_M3 1x-eGFP bearing either HD or G* repeats opposite the target CpG. Fluorescence images were acquired under identical imaging conditions for each channel. Scale bars: 10 µm.

To further validate that 5mC differences in the MUC4 target sequence were responsible for the observed responses in the TALE staining, an additional co-staining experiment was conducted with an alternative 5mC perturbation. For this, HCT116 DKO cells were transfected with plasmids encoding either a catalytically active DNMT3a3L construct or a catalytically inactive version bearing an E756A mutation. The catalytically active construct should allow effective global *de novo* methylation and thereby restore 5mC in the methylation deficient DKO cells (Figure 28a). MeDIP analysis confirmed that cells transfected with the catalytically active DNMT3a3L construct presented 10-fold increased 5mC levels in the MUC4 locus compared to cells transfected with the catalytically inactive construct (Figure 28b). Indeed, 5mC levels of HCT116 DKO cells transfected with the active DMT3a3L construct were very similar to those detected in HCT116 wt cells. As expected, cells transfected with the inactive DNMT3a3L construct did not show any change in methylation level (Figure 28b, Figure 25). Co-staining of these cells with the same TALE_M3 pairs as employed before revealed highly similar signal intensities as observed in the HCT116 wt versus HCT116 DKO co-staining (Figure 27a, Figure 28c). As before, the signals from both G* TALEs did not differ between HCT116 DKO cells transfected with the active or inactive DNMT3a3L construct. Likewise, co-staining with the G*-mCherry and HD-eGFP TALE_M3 versions showed highly similar signals with the G* TALE but a significantly lower signal of the HD TALE in the cells transfected with active DNMT3a3L. This high similarity between the results of the two co-staining experiments using different perturbations of 5mC confirmed that the applied TALE co-staining approaches indeed report 5mC differences. As a result of signal amplification, the high signal-to-noise ratios provided the necessary sensitivity to detect even changes in only 32 theoretical target sequences at the MUC4 locus. Therefore, the developed TALE co-staining approach in combination with signal amplification provides a straightforward method to identify 5mC differences in low repetitive target sequences.

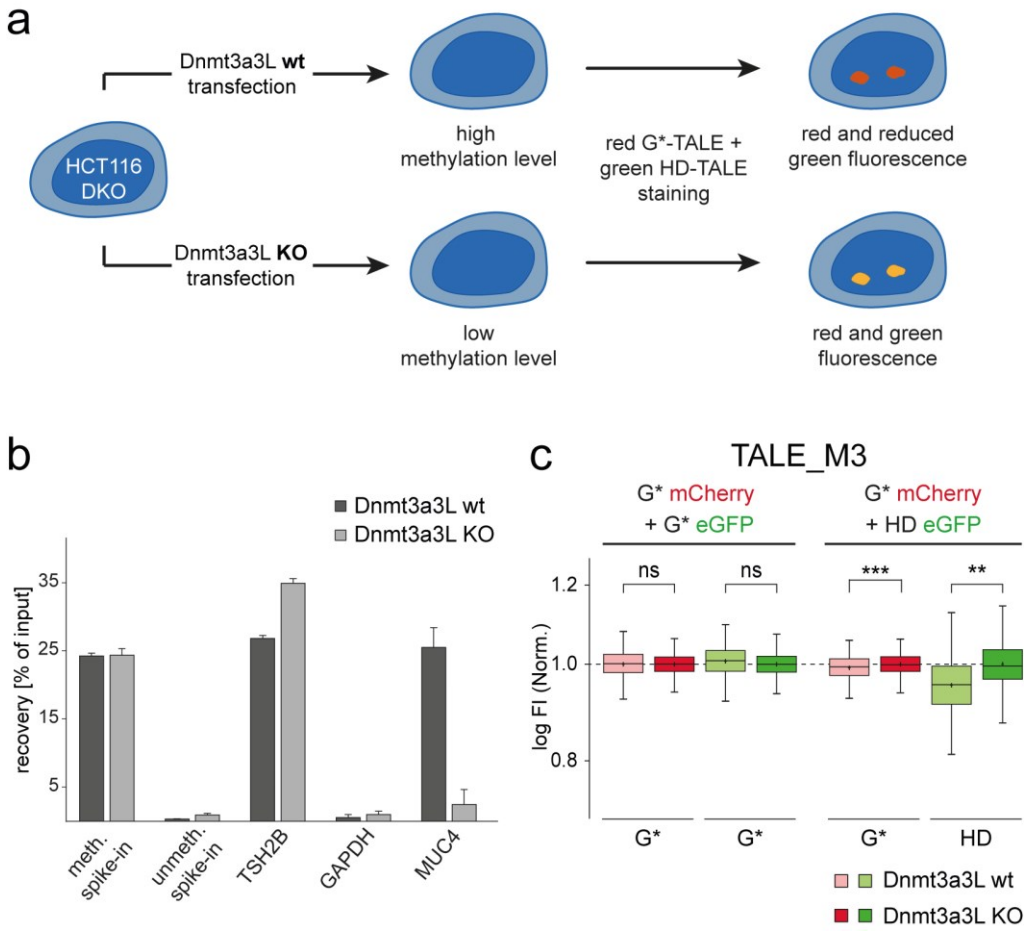


Figure 28. Imaging-based 5mC analysis of HCT116 DKO cells transfected with active or inactive DNMT3a3L. a) Illustration of TALE co-staining experiment with HCT116 DKO cells transfected with Dnmt3a3L or inactive E756A mutant to obtain cells with high and low methylation level, respectively. TALE staining is exemplary shown for red G* TALE and green HD TALE. b) MeDIP analysis of DNA samples from HCT116 DKO cells transfected with either catalytically active Dnmt3a3L or inactive Dnmt3a3L E756A. Relative recoveries of target DNA from both cell types were assessed by qPCR. Results show means ± SD of technical MeDIP duplicates and qPCR triplicates. Shown are results for methylated and unmethylated spike in controls, for an endogenous control locus with high (TSH2B) and low (GAPDH) 5mC level, and for the MUC4 target loci. c) Fluorescence signal intensities (FI) of foci from HCT116 DKO cells transfected with active or inactive Dnmt3a3L. Co-staining was performed with G* and HD TALE_M3 versions fused to mCherry or eGFP and immunostained with primary anti mCherry, secondary Alexa Fluor Plus 594, primary anti-eGFP and secondary Alexa Fluor 488. For each TALE, log FI of each focus is normalized to the mean of log FI of all foci from HCT116 DKO cells. N = 3 experiments totaling >1300 foci per condition. $P < 0.01$**, $P < 0.001$*** (Student's t test).

5. Conclusions and Outlook

This work developed a straightforward signal amplification method for the imaging-based analysis of 5mC in low-repetitive genomic DNA sequences using fluorescent TALE probes. This allowed the visualization of only 32 theoretical repeat sequences in the MUC4 gene locus with high sensitivity and enabled the detection of 5mC differences at these sequences between human cell lines with different levels of methylation.

Based on a previously developed co-staining approach, this method employed TALE probes to analyze 5mC sequence-specifically in imaging studies. To enable this for the analysis of low repetitive genomic loci, this work explored various signal amplification strategies to increase the imaging sensitivity with TALEs. For this, TALE probes were labeled with various fluorophores and epitope tags and tested in cell stainings along with a variety of nanobody and antibody signal amplification strategies against the employed tags. This led to the development of an effective signal amplification protocol that enabled imaging of only 32 theoretical repeat sequences in the MUC4 locus with high signal-to-noise ratio. Furthermore, the application of this signal amplification protocol in combination with the previously described co-staining approach allowed the analysis of 5mC in these sequences. For this, fixed human cells were co-stained with two TALEs targeting the same 32 theoretical repeat sequences in the MUC4 locus, but both being labeled with different fluorophores. One TALE, bearing the 5mC-selective RVD HD, served as a 5mC-sensitive probe that was blocked by 5mC in the target sequence, while the other TALE, bearing the universal RVD G*, was insensitive towards 5mC and served as a universal binder. To amplify the fluorescence signals of both TALEs, additional immunostaining was performed according to the developed signal amplification protocol. By analyzing the fluorescence signal intensities of both TALEs, 5mC differences could be detected between cells with different methylation levels independently of changes in target accessibility. This was demonstrated for samples with different 5mC perturbations, either introduced by DNA methyltransferase knockouts or overexpression. Resulting 5mC-differences were consistent with corresponding MeDIP and Bisulfite sequencing analyses, which indicates reliable detection of methylation by the developed approach. Therefore, it can be concluded that this approach offers a

straightforward and reliable strategy for the *in situ* analysis of 5mC in low-repetitive target sequences.

With the developed signal amplification method, a particularly small number of target sequences could be visualized with high sensitivity. This opens up new possibilities for the analysis of 5mC in non-repetitive single loci, such as gene regulatory regions like promoters. Since the use of only a small subset of specifically designed TALE probes could generate sufficient signal for detection of these regions, direct *in situ* analysis of methylation in key regulatory regions would be possible. This would provide new opportunities to study the epigenetic regulation of a particular gene directly in single cells to gain a deeper understanding of how alterations in methylation patterns contribute to diseases such as cancer.

Furthermore, the TALE-based imaging approach provides positional and topological information about the locus of interest. The combination with additional immunostaining of other chromatin features such as transcription factors, epigenetic readers, or histone modifications could allow for a more comprehensive analysis of the complex interplay between the methylation status of a given genomic locus and other chromatin regulators. This would provide additional information about the regulation of methylation by the local chromatin environment and contribute to a better understanding of epigenetic regulation in gene expression.

To further extend the application scope of the TALE-based imaging approach, other epigenetically modified bases could be targeted. This work focused on the detection of the epigenetic nucleobase 5mC. However, specific engineered RVDs exist that can be used for the detection of the oxidized 5mC bases 5hmC, 5fC and 5caC. By employing these engineered RVDs, the TALE-based imaging approach could be extended to detect these modified nucleobases *in situ*. This would be particularly relevant for tissues that contain high levels of oxidized nucleobases.

6. Material and Methods

Material and Methods were previously published in (A. Jung et al., 2023) (8.3).

6.1 General Information

Synthesis of oligonucleotides

Oligonucleotides listed in Table S1 and Table S2 were synthesized by Merck KgaA (Darmstadt, Germany). Oligonucleotides listed under 7.3.1 were synthesized by Integrated DNA Technologies (Coralville, USA). The desalted oligonucleotides were stored as 100 µM stock in TE buffer.

Sanger sequencing

The sequence of constructed plasmids was checked by Sanger sequencing by Microsynth Seqlab GmbH (Göttingen, Germany) or Eurofins Genomics Germany GmbH (Munich, Germany).

Purification of plasmid DNA and double stranded oligonucleotides

Plasmid DNA was isolated from bacteria strains via silica column purification using NucleoSpin® Plasmid EasyPure kit (MacheryNagel, Düren, Gemany) following the manufacturer's instructions.

Double stranded oligonucleotides digested from plasmids or amplified by PCR were separated and identified by DNA electrophoresis. The crude reaction mixture or agarose gel samples containing the desired products were purified using NucleoSpin® Gel and PCR Clean-Up kit (MacheryNagel).

For DNA electrophoresis, 1% or 2% (w/v) of agarose gel in 0.5× TBE buffer was prepared. The DNA samples were separated using 8 – 12 V/cm. Agarose gels were stained with 0.5 µg/mL ethidium bromide, destained with water, and visualized with UV fluorescence.

Enzymes

The enzymes used in this study are listed below. Corresponding experiments were conducted following the supplier's instructions.

Table 1. List of enzymes.

Enzyme	Company
AcsI	NEB (New England Biolabs)
AgeI	NEB (New England Biolabs)
BamHI	NEB (New England Biolabs)
BsaI	NEB (New England Biolabs)
BsmBI (Esp3I)	NEB (New England Biolabs)
KOD hot start DNA polymerase	Merck Millipore
Lysozyme	Sigma Aldrich
NdeI	NEB (New England Biolabs)
NEBuilder® HiFi DNA assembly cloning kit	NEB (New England Biolabs)
NheI	NEB (New England Biolabs)
NotI	NEB (New England Biolabs)
OneTaq DNA polymerase	NEB (New England Biolabs)
Q5 HiFi DNA polymerase	NEB (New England Biolabs)
Plasmid-Safe DNase	Epicentre
RcoRI	NEB (New England Biolabs)
T4-DNA Ligase	NEB (New England Biolabs)
Taq-DNA ligase	NEB (New England Biolabs)
XbaI	NEB (New England Biolabs)

Antibodies

Table 2. List of antibodies.

Antibody	**Host**	**Reactivity**	**Fluorophore**	**Company**
GFP-Booster Atto488	Alpaca	AcGFP, Citrine, CFP, eCFP, eGFP, eYFP, GFP S65T, mCerulean, pHluorin, sfGFP, mClover (Clover A206K), TagGFP, tagGFP2, Venus, wtGFP, YFP	Atto488	Chromotek
RFP-Booster Atto594	Alpaca	DsRed, mCherry, mPlum, mRFP, mRFPruby	Atto594	Chromotek

Anti-GFP from mouse IgG1κ	Mouse	GFP		Sigma-Aldrich
Anti-mCherry antibody	Rabbit	mCherry		Abcam
Alexa Fluor 488 goat anti-mouse	Goat	Mouse	Alexa Fluor 488	Thermo Scientific
Goat anti-Rabbit IgG (H+L) Alexa Fluor Plus 594	Goat	Rabbit	Alexa Fluor plus 594	Thermo Fisher
Anti-Flag-tag	Rabbit	Flag-tag (DYKDDDDK peptide)		Sigma
Anti-GCN4, Rabbit IgG	Rabbit	GCN4 (HLENEVARLKK peptide)		Biozol
Centriolin sc-365521	Mouse	Centriolin (human)		Santa Cruz Biotechnology

Bacteria strains

Table 3. *E.coli* strains used in this study.

Strain	Genotype	Supplier
GH371	F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(araleu)7697 galU galK rpsL (StrR) endA1 nupG fhuA::IS2 (confers phage T1 resistance),upp-	iGEM
DH5α	F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF) U169, hsdR17(rK- mK+), λ–	Invitrogen™ (Thermo Fisher Scientific)
BL21-Gold DE3	B F–ompT hsdS(rB–mB–) dcm+ Tetr gal λ(DE3) endA Hte	Agilent

Mammalian cell lines

Table 4. Mammalian cell lines used in this study.

Cell line	Origin	Supplier
U2OS	Human Osteosarcoma	Sigma Aldrich
HCT116	Colorectal carcinoma	Horizon Discovery Ltd.
HCT116 DKO DNMT1 (Δexons3-5/Δexons3-5), DNMT3B (-/-)	Colorectal carcinoma	Horizon Discovery Ltd.

Biological reagents and chemicals

Table 5. List of biological reagents and chemicals.

Name	Supplier
2-Amino-2-(hydroxymethyl)propane-1,3-diol (Tris), buffer grade	Carl Roth
2-Log DNA ladder	NEB
Acetic acid	Carl Roth
Agarose LE, molecular biology grade	Biozym Scientific
Amicon™ Ultra-0.5 Centrifugal Filter units	Merck
Bovine serum albumin (BSA)	Cell Signaling Technology
Carbenicillin, disodium salt	Carl Roth
Dimethyl sulfoxide (DMSO; ≥99.7 %)	Merck
DMEM	Pan Biotech
dNTPs	NEB
DPBS	Pan Biotech
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
EDTA	Pan Biotech
Ethanol (EtOH; ≥99.8 %)	VWR Chemicals
Ethidium bromide	Carl Roth
Fetal bovine serum (FBS; Premium South America)	PAN Biotech
FuGENE HD transfection reagent	Promega
Gibson Assembly® Master Mix	NEB
HisPur™ Ni-NTA Resin	Thermo Fisher Scientific
Isopropanol	Fisher Scientific
Lauroyl sarcosinate	AppliChem
LB-Agar	Roth
LB-Medium	Roth
L-Glutamine	PAN-Biotech
Lipofecatmine 2000 Reagent	Invitrogen
MagMeDIP qPCR Kit	Diagenode
Microplate BCA Protein Assay Kit	Thermo Fisher Scientific
Monosodium phosphate (NaH_2PO_4)	Roth
Opti-MEM® I	Gibco
Penicillin-Streptomycin (10,000 U/mL Penicillin; 10 mg/mL Streptomycin)	PAN Biotech
QIAmp DNA Mini Kit	Qiagen
RPMI 1640	PanBiotech
Sodium chloride (NaCl; ≥99.5%)	Merck
Sodium hydroxide	Merck
Spectinomycin	Alfa Aesar
Tetracycline	Sigma-Aldrich
Triton® X-100	Fluka Chemika
Trypsin/EDTA (w/o Ca_2+, Mg_2+)	PAN Biotech
Tween® 20	Fisher Bioreagents

Vectashield with DAPI	Vector Laboratories
X-Gal	Thermo Fisher Scientific

Software

Table 6. List of software.

Name	Company
Adobe Illustrator 2022 v26.4	Adobe
cellSens Dimension Version 3	Olympus
ChemDraw Professional v21.0.28	PerkinElmer
Fiji	Wayne Rasband
ImageJ	Wayne Rasband
NanoDrop 2000 v1.6	Thermo Fisher Scientific
Office 365	Microsoft
Cell sorter software v.2.1.5	Sony Biotechnology
SnapGene v4.3	Dotmatics

6.2 Plasmid Cloning

Plasmid pAnJ1861, for expression of TALEs fused to three N-terminal mCherry, was cloned by Gibson assembly[243]. For this, TALE expression vector pAlM1577[100], containing N-terminal mCherry, was linearized using primer o3661 and o3662. The two inserts mCherry_2 and mCherry_3 were amplified using o3657 and o3658 for mCherry_2 and o3659 and o3660 for mCherry_3. Gibson assembly was performed with a 2:1 insert:vector molar ratio using Gibson assembly master mix (NEB).

To generate plasmids for expression of TALEs fused to 20 or 30 N-terminal FLAG tags, restriction sites AgeI and NheI were introduced into vector pAnI521[244] via QuikChange site-directed mutagenesis (Agilent) using primers o4468 and o4469 for AgeI and o4470 and o4471 for NheI. A pre-existing AgeI site was removed using o4472 and o4473 via site-directed mutagenesis. To generate pPiB2670, containing a 10x-FLAG array, two inserts containing 5x-FLAG sequences each (see 8.1.4 for sequences), were amplified using o4493 and o4494 for insert_1 and o4495 and o4496 for insert_2. For following ligation, modified vector pAnI521 was restricted with AgeI and NheI, insert_1 with AgeI and BamHI and insert_2 with BamHI and XbaI. Vector and insets were ligated with T4 ligase (New England Biolabs, #M0202T) using a 3:1 insert:vector molar ratio. Site-

directed mutagenesis using o4587 and o4588 was performed to introduce a start codon. For generation of pPiB2683 containing a 20x-FLAG array, modified pPiB2670 was restricted with AgeI and NheI and ligated with restricted insert_1 and 2 as described above. For cloning of pPiB2700 containing a 30x-FLAG array, the before described restriction and ligation was repeated with pPiB2683 as vector.

The plasmid pCrW2056 for expression of TALEs fused to 24 N-terminal GCN4 tags was cloned by amplification of the 24x-GCN4 insert from pAlM1103 (Adgene, #60910) using primers o3935 and o3936. The vector pAnI521 was linearized with restriction enzymes NdeI and NotI and ligated to the insert by Gibson assembly with a 2:1 insert:vector molar ratio.

TALEs were assembled as previously described by Golden Gate Assembly[123] (see Table S7 for detailed RVD composition). To generate plasmids coding for TALE proteins in frame with a C-terminal His6-Tag and different N-terminal tags, the plasmids pAnI521 for 1x-GFP, pAlM1577 for 1x-mCherry, pAnJ1861 for 3x-mCherry, pPiB2683 for 20x-FLAG, pPiB2700 for 30x-FLAG and pCrW2056 for 24x-GCN4 were used as entry vectors in Golden Gate 2 reactions.

To generate the active DNMT3a3L vector pCoT3181, plasmids pAlH1894 and pJaW876[100] were restricted with AcsI and RcoRI and ligated with T4 ligase using a 2:1 insert:vector molar ratio. Generated plasmid pCoT3180 was linearized with NotI and ligated by Gibson assembly with the CMV-EBFP2 insert, amplified from plasmid EBFP2-N1 (Adgene, #54595) with primers o5028 and o5029, resulting in pCoT3181. Mutation E756A for catalytically inactive DNMT3a3L was introduced by site-directed mutagenesis using primers o2038 and o2039 to generate pAnJ3188.

6.3 TALE Expression and Purification

TALEs were expressed and purified as described previously[153]. Briefly, TALE plasmids were transformed in electrocompetent BL21 DE3 Gold *E.coli* cells and grown on LB carbenicillin (Carb, 100 mg/mL) agar plates at 37 °C overnight. 5 mL LB medium supplemented with 100 mg/mL carbenicillin were inoculated with a single colony and incubated for 4 h at 37 °C and 220 rpm. This starter culture was transferred to a flask containing 100 mL of LB + Carb and incubated under the same conditions until a OD_{600}

of 0.6 arbitrary units (au) was reached. TALE expression was induced by addition 0.4 mM IPTG. For expression of TALEs fused to mCherry tags, cultures were incubated at 18 °C and 220 rpm overnight. Expression cultures of TALEs fused to GFP tag, FLAG tags or SunTag were incubated at 37 °C and 220 rpm for 4 h. Cells were harvested by centrifugation at 3000 g at 4 °C for 20 min. The pellet was kept at 20 °C for 2 h and resuspended in 10 mL Deep Lysis Buffer (10 mM Tris-HCl, 300 mM NaCl, 2.5 mM $MgCl_2$, 5 % DMSO, 0.2 % sodium lauroyl sarcosinate (AppliChem), 0.1 % Triton X-100, pH = 9) containing 1 mM PMSF, 1 mM DTT and 50 µg/mL lysozyme (Sigma Aldrich). Cell lysis was aided by sonication on ice (3 min; 20 % amplitude; 4 sec on, 2 sec off). Samples were centrifuged at 14000 g at 4 °C for 20 min to remove cell debris. The supernatant was incubated with 0.5 mL HisPur™ Ni-NTA Resin (ThermoFisher Scientific, #88221) overnight at 4 °C spinning on a rotating wheel. The beads were collected and washed with PBS, twice with Lysis buffer (10 mM Tris-HCl, 300 mM NaCl, 2.5 mM $MgCl_2$, 0.1 % Triton X-100, pH = 9) + 20 mM imidazole + 1mM DTT and three times with Lysis buffer + 50 mM imidazole + 1mM DTT. TALEs were eluted by incubating the beads with 1 mL Lysis buffer + 500 mM imidazole + 1 mM DTT shaking at 800 rpm at 4 °C overnight. Samples were centrifuged at 12000 g for 5 min and the supernatant was purified with Amicon™ Ultra-0.5 Centrifugal Filter units (Merck, MWCO: 100 kDa, #UFC510024) by centrifugation at 14000 g for 10 min at 4 °C. For washing, the volume of the sample was filled up to 500 µL with TALE Storage buffer (200 mM NaCl, 20 mM Tris, 10% glycerol, pH = 7.5) + 1 mM DTT and centrifuged at 14000 g for 10 min at 4 °C. Washing was repeated three times. Samples were recovered by centrifugation at 1000 g for 2 min and filled up with TALE Storage Buffer to a volume of 500 µL. Samples were centrifuged at 14000 g at 4 °C for 5 min and aliquots were snap-frozen with liquid nitrogen and stored at -80 °C. Protein concentrations were measured by BCA using Microplate BCA Protein Assay Kit – Reducing Agent Compatible (ThermoFisher Scientific, #23252) following manufacturer's instructions.

6.4 Mammalian Cell Transfection with DNMT3a3L

1 million HCT116 DKO cells (DNMT1 (Δexons3-5/Δexons3-5), DNMT3B (-/-), Horizon Discovery Ltd., #HD R02-022) were seeded on 10 cm diameter dishes (Sarstedt, #83.3902.300) in 10 mL full RPMI medium (RPMI 1640 (with L-Glu) (PanBiotech, #P04-

16500) + 10% FBS (PanBiotech, #P30-3302) + 1% Pen/Strep (PanBiotech, #P06-07050)) and incubated at 37 °C and 5% CO_2 for three days. For transfection, 36 µL FuGENE HD transfection reagent (Promega, #E2311) was mixed with 564 µL OptiMEM (Gibco) and incubated for 5 min at RT. 12 µg (120 µL of 100 ng/µL solution) plasmid pCoT3181 or pAnJ3188 was added, mixed by pipetting up and down and kept for 15 min at RT. The medium of the cells was replaced with 2500 µL fresh prewarmed full RPMI medium and the transfection mixture was added dropwise to the cells. The cells were placed for 1.5 h in the incubator, then 8 mL pre-warmed full RPMI medium was added. The cells were further incubated for 48 h at 37 °C and 5% CO_2 before sorting.

6.5 Flow Cytometry and Cell Sorting

Transfected HCT116 DKO cells were washed with DPBS (PanBiotech, P04-361000), trypsinized with Trypsin 0.05% / EDTA 0.02% (PanBiotech, #P10-038100) for 3 min at 37 °C and blocked with full RPMI medium. Cells were pelleted by centrifugation at 200 g for 5 min at 37 °C, washed with 5 mL DPBS and resuspended in 1.5 mL prewarmed DPBS + 1% BSA. Cells were sorted with a Sony Cell Sorter model LE-SH800SFP in targeted mode using the 405nm laser (filter FL1 450/50, Optical Filter Pattern 2) to detect EBFP2 transfection control from DNMT3a3L plasmids. Gates were set to assure similar expression levels of fluorescent protein in active and inactive DNMT3a3L transfected samples. EBFP2+ cells were collected in tubes containing prewarmed full RPMI medium. For microscopy, 40.000 cells/well were seeded on µ Plate 96 Well Black ibiTreat tissue culture treated plates (ibidi, #89626) (coated with 0.01 % poly L-lysine in DPBS for 1 h at 37°C) and incubated for 16 h at 37 °C and 5% CO_2.

6.6 TALE and Antibody Staining

19.000 U2OS (Sigma Aldrich, #92022711-1VL) cells, 32.000 HCT116 wt cells or 32.000 HCT116 DKO cells per well were seeded on µ Plate 96 Well Black ibiTreat tissue culture treated plates and incubated overnight at 37 °C and 5% CO_2 in growth medium (DMEM (w/o L-Glu) (PanBiotech, #P04-03609) + 10% FBS + 1% Pen/Strep + 1% L-Glu (PanBiotech, #P04-80100) for U2OS cells or RPMI 1640 medium (with L-Glu) + 10% FBS + 1% Pen/Strep for HCT116 cells). On the next day cells were washed with DPBS

and fixed with ice-cold methanol at -20 °C for 10 min. Cells were washed with DPBS and treated with 2N HCl for 5 min at RT followed by three washing steps with DPBS at 450 rpm for 5 min each. Cells were blocked with Blocking Buffer (DPBS + 1% BSA + 0.1% Tween20) at 450 rpm at RT for 2 h. TALE staining was performed with 200 µL of 1 nM purified TALE in Blocking Buffer for 30 min at 450 rpm at RT. Cells were washed three times with Blocking Buffer for 5 min at RT shaking at 450 rpm and kept in Blocking Buffer at 4 °C, shaking at 450 rpm overnight. For antibody staining, all primary antibodies were diluted 1:1000 in Blocking Buffer, 200 µL was added to each well and incubated for 1 h, shaking at 450 rpm at RT. Cells were washed three times with Blocking Buffer for 5 min, 450 rpm each. All secondary antibodies were diluted 1:2000 in Blocking Buffer and 200 µL were added to each well for 1 h at 450 rpm, RT. Cells were washed three times with Blocking Buffer for 5 min, 450 rpm each. For anti-RFP Booster staining, RFP Booster was diluted 1:200 in Blocking Buffer. Staining and washing was performed as described before. Nucleus staining was performed by incubating the samples with 2 µL per well of Vectashield with DAPI (Vector Laboratories, #H-1200) in 200 µl DPBS for 10 min at 450 rpm, RT. Each well was washed with DPBS for 5 min at 450 rpm, RT and kept in DPBS for microscopy.

6.7 Microscopy

Experiments were performed using an Olympus IX81 microscope equipped with LEDs as excitation light source (150–750 mW) coupled with a Hamamatsu model C10600-10B-H camera. Images were acquired using a 60x oil objective with compatible immersion oil (ibidi, #50101). Z-stack images (0.3 µm/step, range: 8 µm) for DAPI (excitation filter 395/25 nm, emission filter 474/27 nm), EGFP (excitation filter 475/28 nm, emission filter 554/23 nm) and mCherry (excitation filter 555/28 nm, emission filter 635/18 nm) were taken.

6.8 Image Processing and Analysis

Image processing and analysis was performed as described previously[100]. The intensity and subcellular localization of foci was analyzed from z projections of image stacks (1344 × 1024 pixels, 12 bits) with maximal intensity using the FIJI distribution of

ImageJ. To subtract the background, the mean intensity of an out-of-interest region was measured from each channel and subtracted from the stack. Nuclear regions were selected from DAPI images (10 μm^2 minimum area, circularity between 0.5-1.0). To analyze intensity and size of the foci in mCherry images, the "GaussFit OnSpot" plugin was applied, using elliptical shape and Levenberg Marquard fit mode with a rectangle half size of 10 pixel. Spots larger than 12 pixel or outside the nuclear regions were excluded and the prominence (signal-to-noise ratio) was adjusted for each condition to only select foci-like objects. The generated mask from mCherry images was applied to the eGFP images to measure the mean fluorescence intensity. Image processing was performed in batch, utilizing an ImageJ macro script. For each nucleus, the number, size, and intensity of the associated foci in the mCherry and eGFP images was recorded.

6.9 Data Analysis and Statistics

Data analysis and plotting was performed with R as described previously[100]. For each TALE, the log transformed mean fluorescence intensity of each focus was normalized to the average fluorescence intensity of all foci from the associated HCT116 DKO or DNMT3a3L KO transfected sample for each experiment. Graphs were plotted using the ggplot2 library. For statistical analysis a Stutdent's t test was applied with GraphPad, considering the number of independent experiments as sample size (N ≥ 3 independent experiments in every case).

6.10 Methylated DNA Immunoprecipitation and qPCR

DNA from HCT116 cells was isolated and purified using the QIAmp DNA Mini Kit (Qiagen, #51304) following the manufacturer's instructions. 300 µL (10 ng/µL) purified DNA was sheared into fragments around 400 bp, using the Bioruptor Pico (Diagenode, #B01080010) in 1.5 mL Bioruptor Microtubes (Diagenode, #C30010016). Samples were sonicated for 11 cycles (30 sec on, 30 sec off) at 4 °C. Fragment size was analyzed with the BioAnalyzer (Agilent) and DNA was concentrated to a concentration of 100 ng/µL using the Concentrator plus (Eppendorf). For methylated DNA immunoprecipitation and qPCR, the MagMeDIP qPCR Kit (Diagenode, #C02010021) was used according to the

manufacturer's instructions. qPCR was performed on the CFX384 Touch real-time PCR detection system (Bio-Rad).

6.11 gDNA Isolation, Bisulfite Conversion and Sanger Sequencing

DNA from HCT116 wt and DKO cells was isolated and purified using the QIAmp DNA Mini Kit (Qiagen, #51304) following the manufacturer's instructions. Bisulfite conversion was conducted with 1 µg DNA per reaction, using the EpiTect Bisulfite Kit (Qiagen #59104) according to the manufacturer's instructions. A 157 bp locus within the MUC4 locus was amplified from bisulfite-converted gDNA using OneTaq DNA polymerase (NEB #M0480S) and primers o5262 and o5263. Amplified DNA was purified with the Monarch PCR & DNA Cleanup Kit (NEB #T1030L) and sequenced using primer o5262 as forward primer and o5262 as reverse primer.

7. Supplementary Information

7.1 Supplementary Figures

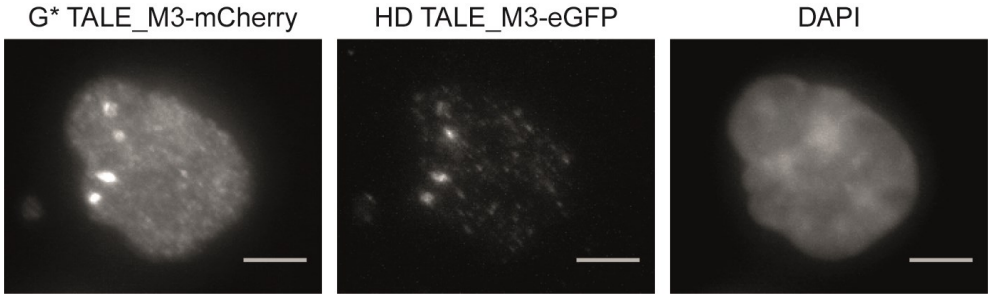


Figure S1. Dividing HCT116 DKO cell shows MUC4 foci in mirrored conformation. Cells were co-stained with G* TALE_M3-mCherry and HD TALE_M3-eGFP. Immunostaining with Ms anti-eGFP, Goat anti-Mouse Alexa Fluor 488 and Rb anti-mCherry, Goat anti-Rabbit Alexa Fluor Plus 594. Scale bars: 5 µm.

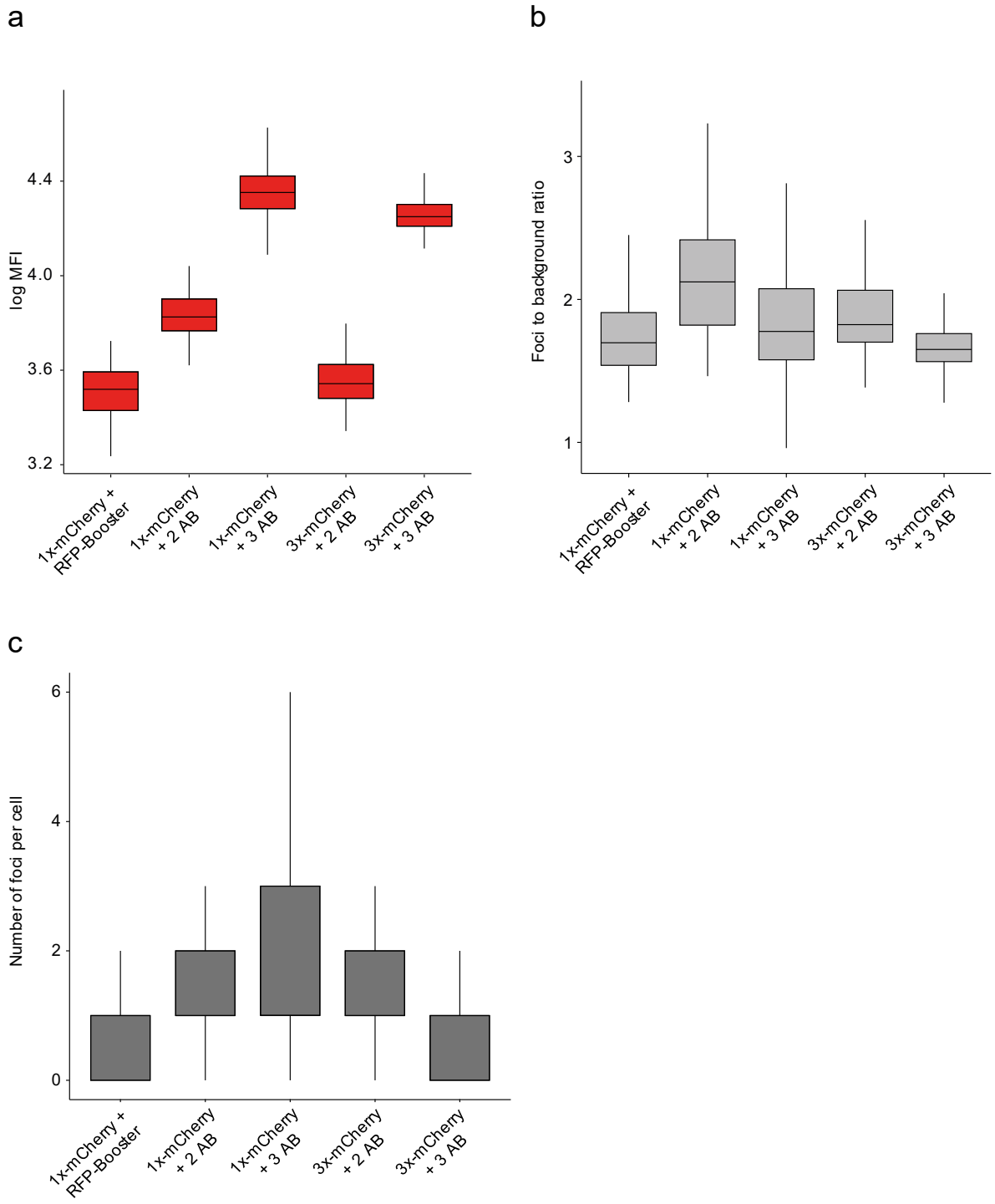


Figure S2. Analysis of fluorescence signal amplification with TALE_M3. Fluorescence intensities of foci and nuclear background from experiment shown in Fig 23. a) log MFI of foci from TALE_M3 staining with different epitope tag and immunostaining combinations. b) MFI ratio of foci to nuclear background. c) Average number of foci per cell.

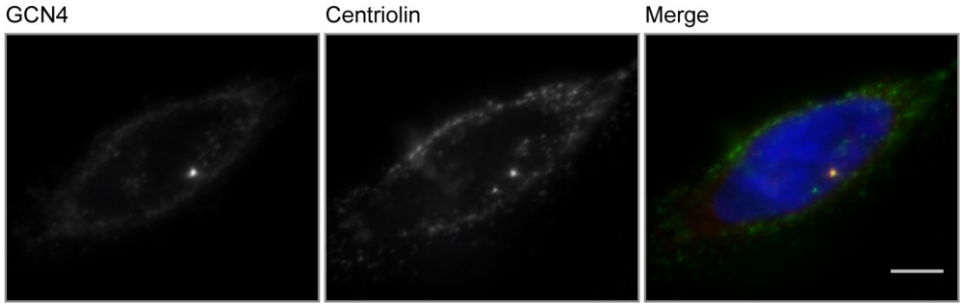


Figure S3. GCN4 – Centriolin co-staining. U2OS cell stained with Rb anti-GCN4 (1:500), anti-Rabbit Alexa Fluor plus 594 (1:1000) and Ms anti-Centriolin (1:50), anti-Mouse Alexa Fluor 488 (1:1000). Scale bar: 5 µm.

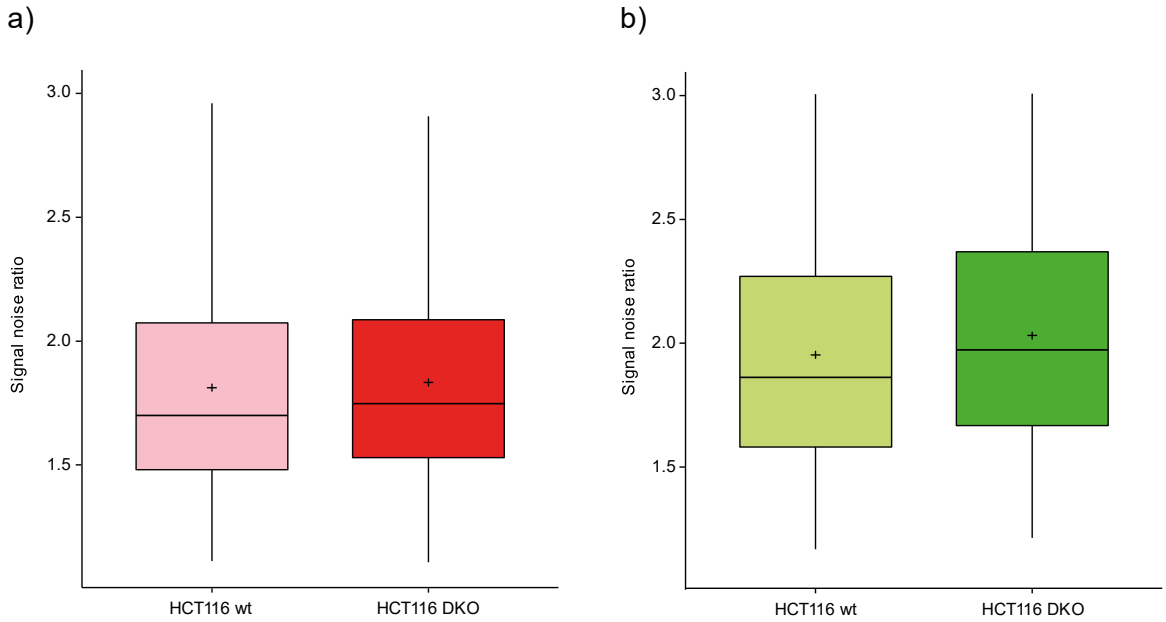


Figure S4. Signal-to-noise ratio in red and green fluorescence channel. HCT 116 wt and DKO cells were co-stained with G* TALE_M3-mCherry and G* TALE_M3-GFP and immunostained with Ms anti-eGFP, Goat anti-Mouse Alexa Fluor 488 and Rb anti-mCherry, Goat anti-Rabbit Alexa Fluor Plus 594. Ratio of mean fluorescence intensity of all foci to mean fluorescence intensity of nuclei background is shown for the a) red and b) green channel.

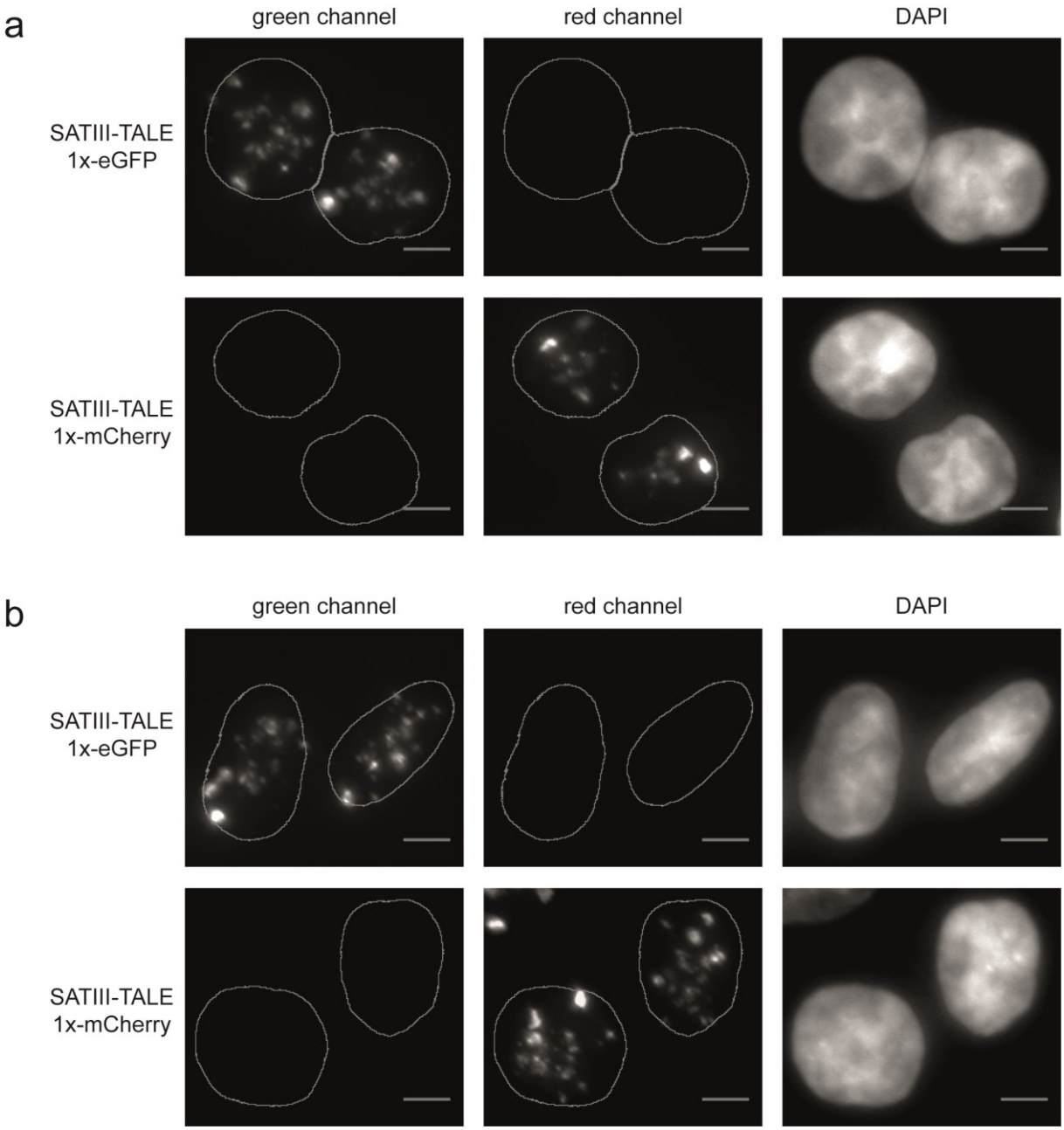


Figure S5. Employed antibodies in co-staining experiments do not show unspecific off-target binding. HCT116 wt (a) and HCT116 DKO (b) cells stained with TALE_SatIII 1x-eGFP or TALE_SatIII 1x-mCherry. Cells were immunostained with Mouse anti-eGFP, Goat anti-Mouse Alexa Fluor 488 and Rabbit anti-mCherry, Goat anti-Rabbit Alexa Fluor Plus 594. Fluorescence images of were acquired under the same imaging conditions for each channel. Scale bars: 5 µm.

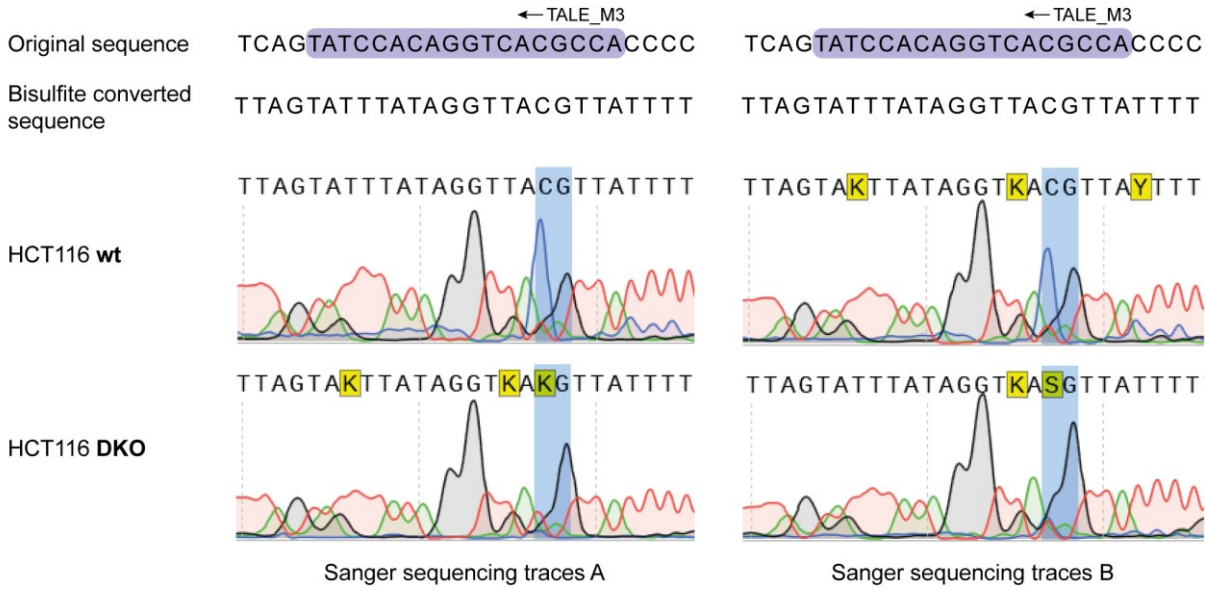


Figure S6. Sanger sequencing traces of PCR product from bisulfite converted DNA from HCT116 wt and HCT116 DKO cells. Sanger sequencing traces A and B were obtained from different PRC products. TALE_M3 target sequence highlighted in purple. CpG position highlighted in light blue.

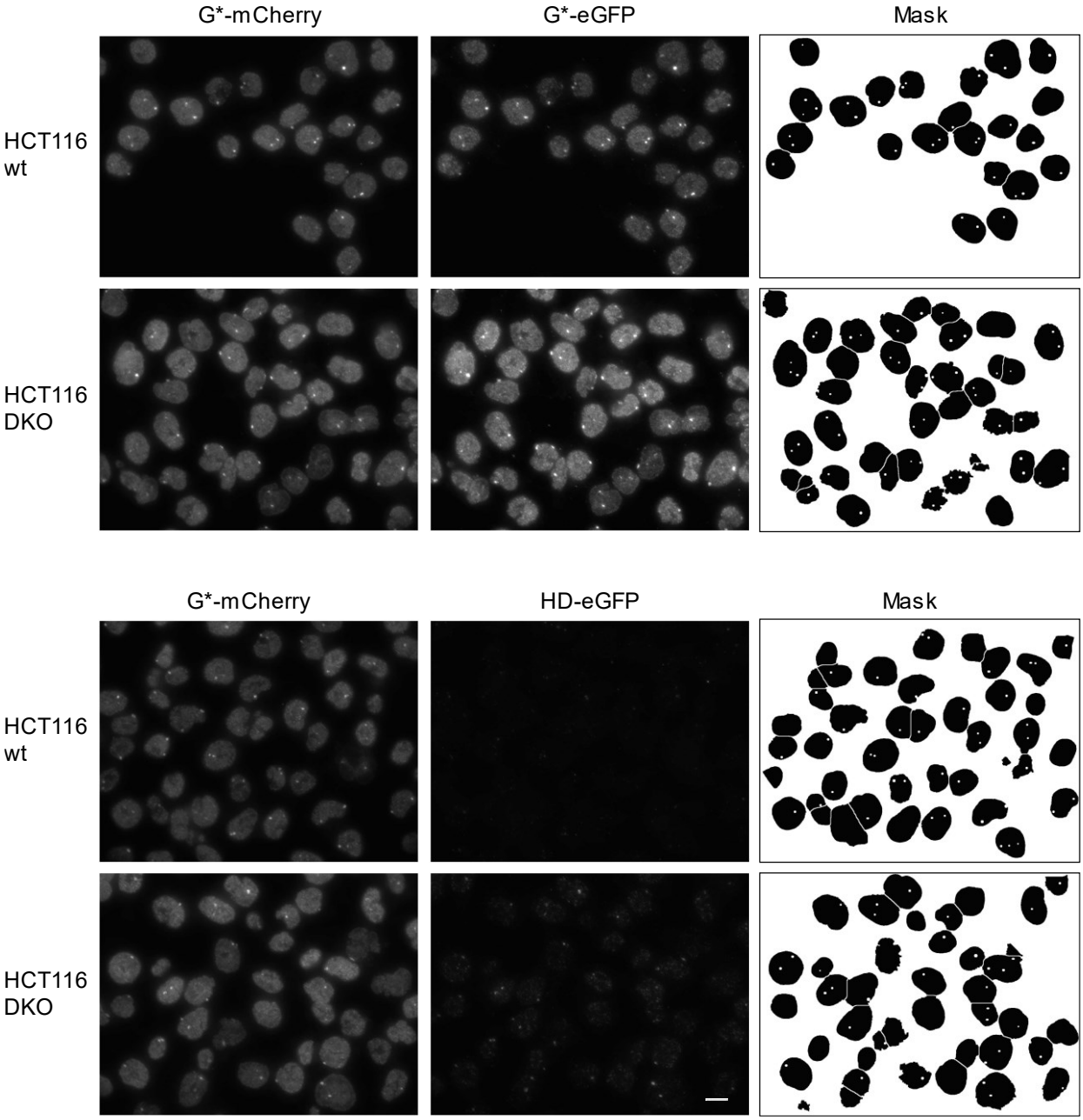


Figure S7. TALE_M3 co-staining of HCT116 wt and HCT116 DKO cells with generated masks from mCherry images. Co-staining of G*-mCherry and G*-eGFP TALE on top. Co-staining of G*-mCherry and HD-eGFP TALE below. Fluorescence images of were acquired under the same imaging conditions for each channel. Scale bar: 15 µm.

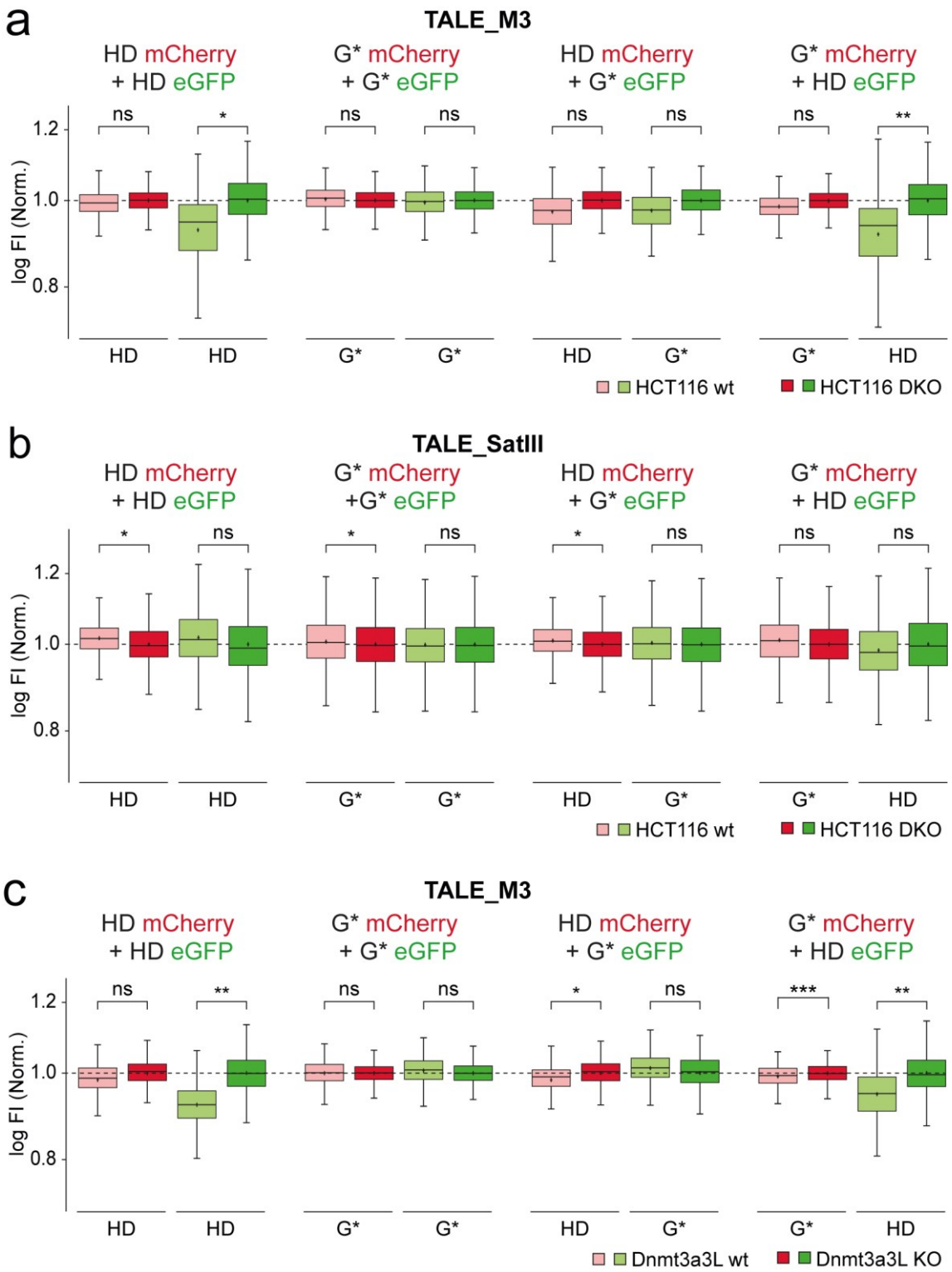


Figure S8. Fluorescence signal intensities of foci from cells co-stained with G* and HD TALE fused to mCherry or eGFP. a) TALE_M3 co-staining of HCT116 wt and DKO cells. N = 4 experiments totaling > 2000 foci per condition b) TALE_SatIII co-staining of HCT116 wt and DKO cells. N = 4 experiments totaling > 7000 foci per condition c) TALE_M3 co-staining of HCT116 DKO cells transfected with Dnmt3a3L wt or KO. N = 3 experiments totaling > 1300 foci per condition. For each TALE, log FI of each focus is normalized to the mean of log FI of all foci from HCT116 DKO cells (a, b) or HCT116 DKO cells transfected with Dnmt3a3L KO (c). $P < 0.1$*, $P < 0.01$**, $P < 0.001$*** (Student's *t*-test).

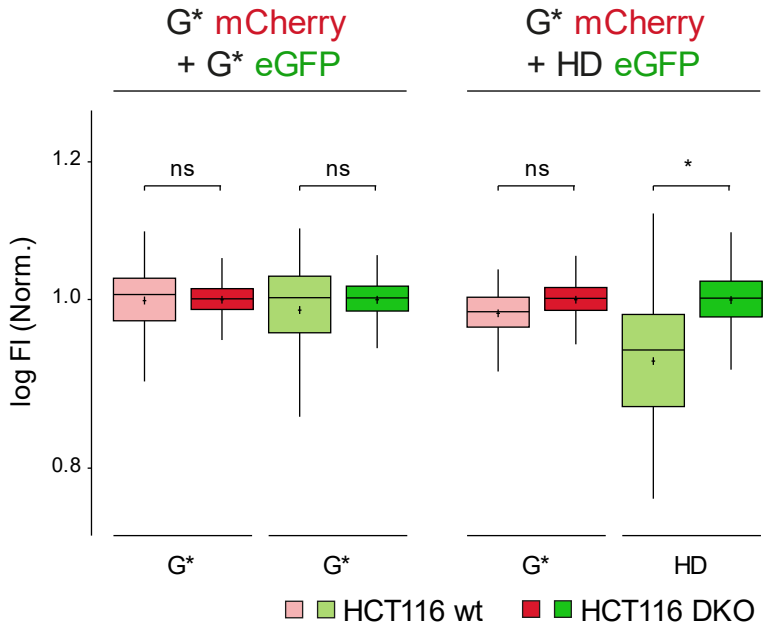


Figure S9. Fluorescence signal intensities of nuclei background from cells co-stained with G* and HD TALE fused to mCherry or eGFP. TALE_M3 co-staining of HCT116 wt and DKO cells. N = 4 experiments totaling > 2000 nuclei per condition. For each TALE, log FI of each nucleus is normalized to the mean of log FI of all nuclei from HCT116 DKO cells. $P < 0.1$* (Student's *t*-test).

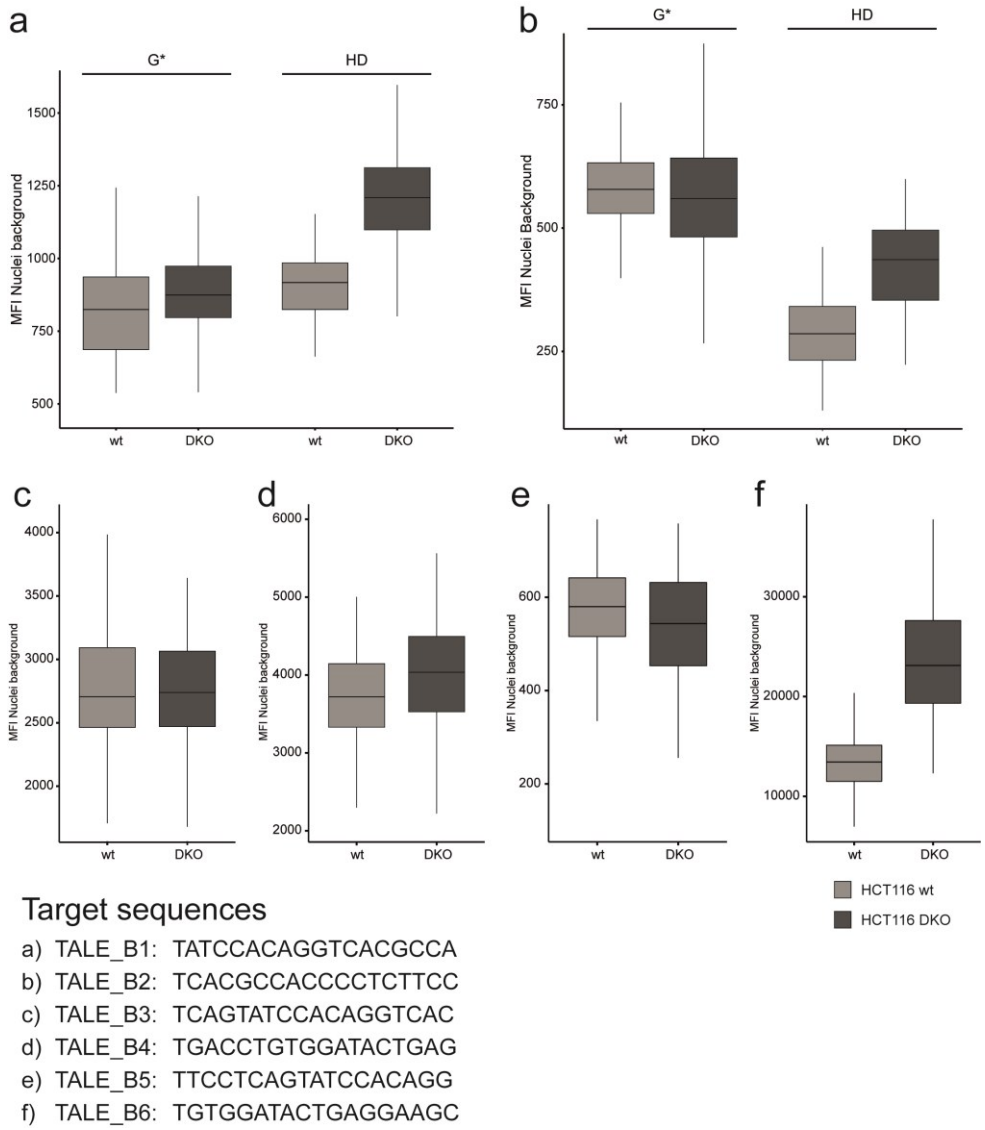


Figure S10. Nuclei background fluorescence intensities of HCT116 wt and DKO cells stained with G* and HD TALEs not showing MUC4 foci. Staining with single mCherry TALEs of identical size and overall design but different CpG-containing or CpG-free MUC4 target sequences of identical lengths (see table). TALE single-stainings were followed by immunostaining with Rabbit anti-mCherry and Goat anti-Rabbit Alexa Fluor Plus 594 prior data acquisition.

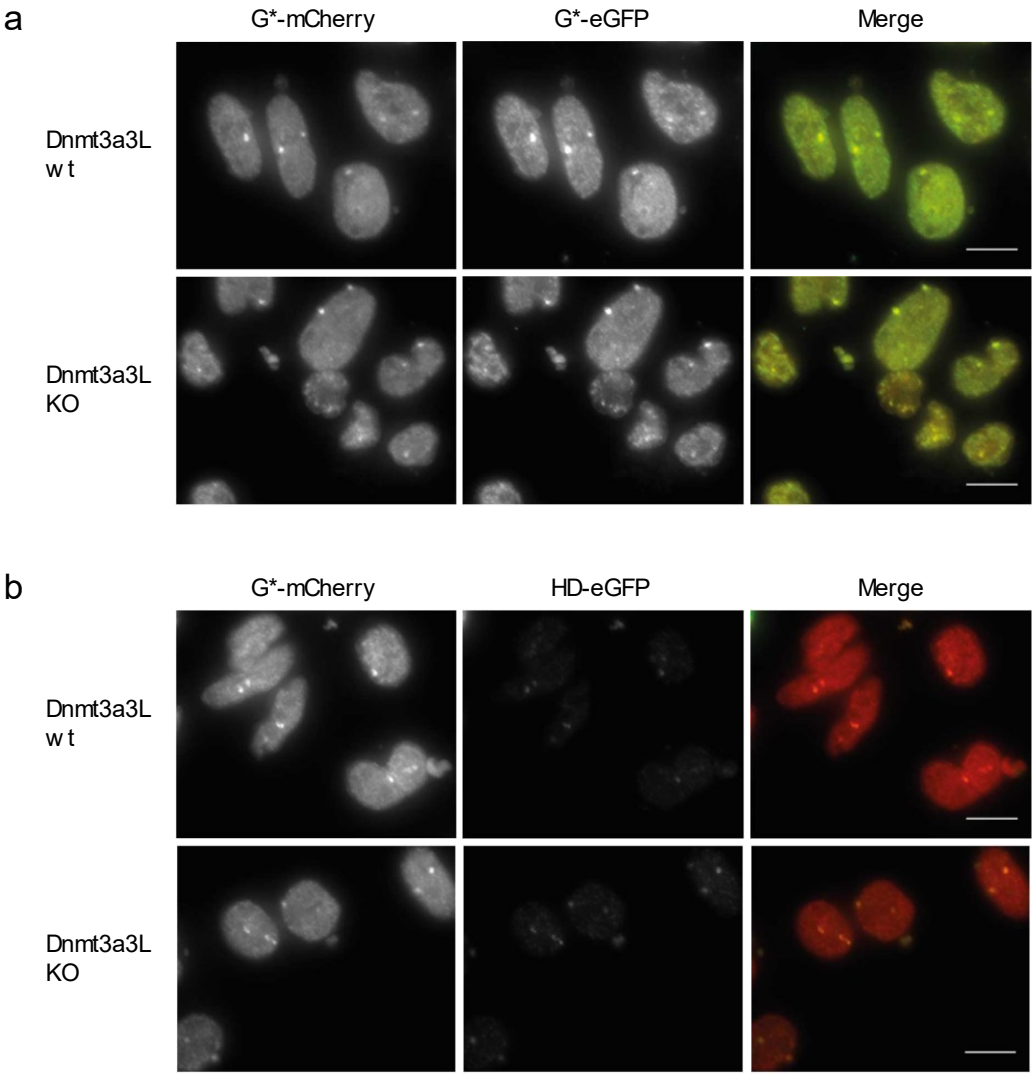


Figure S11. TALE_M3 1x-mCherry and TALE_M3 1x-eGFP co-staining of HCT116 DKO cells transfected with DNMT3a3L wt or KO. a) Co-staining with G*-mCherry and G* eGFP TALE. b) Co-staining with G*-mCherry and HD eGFP TALE. Cells were stained with primary anti-mCherry and anti-eGFP antibodies and secondary Alexa Fluor plus 594 labeled and Alexa Fluor 488 labeled antibodies. Fluorescence images of were acquired under the same imaging conditions for each channel. Scale bars: 10 μm.

7.2 Supplementary Tables

Table S1. Theoretical target sequences of TALE_SatIII in each chromosome.

Chromosome	Target sequences TALE_SatIII
1	213
2	29
3	30
4	92
5	289
6	0
7	175
8	0
9	27997
10	231
11	0
12	2
13	625
14	1127
15	4297
16	3
17	443
18	0
19	0
20	395
21	1022
22	1197
x	0
y	451

Table S2. Theoretical repeats of sequences in MUC4 gene based on GRCh38.p14 and T2T-CHM13v2.0 genome assembly.

MUC4 TALE target sequences used in this work

Target sequence	GRCh38.p14	T2T-CHM13v2.0
TGGCGTGACCTGTGGATA	16	32
TGGTGACAGGAAGAGGGG	46	79
TTCCTCAGCATCCAC	95	148
TATCCACAGGTCACGCCA	16	32
TCACGCCACCCCTCTTCC	25	49
TCAGTATCCACAGGTCAC	18	36
TGACCTGTGGATACTGAG	18	38
TTCCTCAGTATCCACAGG	43	76
TGTGGATACTGAGGAAGC	9	20

MUC4 TALE or CRISPR-dCas9 target sequences used by others

Target sequence	GRCh38.p14	T2T-CHM13v2.0
CCTGTCACCGACACTTCC[227,232]	43	67
GACCTGTGGATGCTGAGGAA[200]	55	81
CTTCCTGTCACCGACACTTC[198]	42	65
CAGCATCCACAGGTCACGCCAC[198]	32	44
GTCACCGACACTTCCTCAGCATCCAC[195]	28	43
TCTTCCTGTCACCGACACTTC[195]	39	62
CAGCATCCACAGGTCACGCCAC[195]	32	44

Table S3. Statistical data of fluorescence signal amplification experiment from Figure 21b.

	3x-mCh	1x-mCh + Booster	3x-mCh + Booster	1x-mCh + 2AB	3x-mCh + 2AB	1x-mCh + 3AB	3x-mCh + 3AB	20x-FLAG + 2AB	30x-FLAG + 2AB	20x-FLAG + 3AB	30x-FLAG + 3AB	SunTag + 2AB	SunTag + 3AB
1x-mCh	ns	ns	ns	***	**	**	**	*	*	**	*	*	***
3x-mCh		ns	ns	**	**	**	**	*	*	*	*	*	**
1x-mCh + Booster			ns	***	**	**	**	*	*	**	*	*	**
3x-mCh + Booster				*	*	*	**	ns	ns	*	*	*	**
1x-mCh + 2AB					ns	*	*	ns	ns	ns	ns	ns	*
3x-mCh + 2AB						ns	ns	ns	ns	ns	ns	ns	ns
1x-mCh + 3AB							ns	ns	ns	ns	ns	ns	ns
3x-mCh + 3AB								*	*	ns	ns	*	ns
20x-FLAG + 2AB									ns	ns	ns	ns	*
30x-FLAG + 2AB										ns	ns	ns	*
20x-FLAG + 3AB											ns	ns	ns
30x-FLAG + 3AB												ns	ns
SunTag + 2AB													ns

Student's *t*-test: $P < 0.1$*, $P < 0.01$**, $P < 0.001$***

Table S4. Oligonucleotides for cloning.

Name	Sequence (5'→ 3')
o2038	CCCTTCTTCTGGCTCTTTGCCAATGTGGTGGCCATGGGCG
o2039	CCATGGCCACCACATTGGCAAAGAGCCAGAAGAAGGGGCG
o3579	TTTAACTTTAAGAAGGAGATATACAGGATCCGGAAGTATGCGTAAAG
o3580	CGAATAACCGAGTGTCTTATCTAGATTACCTGCCTCGAGTTTGTACAGT
o3657	TCTGGCGGCCGCTCTGGCATGGTGAGCAAGGGCGAGGAGG
o3658	GCTACCATCGATGCTACCCTTGTACAGCTCGTCCATGCCG
o3659	GGTAGCATCGATGGTAGCATGGTGAGCAAGGGCGAGGAGG
o3660	TTGCGAATAACCGAGTGTCTTGTACAGCTCGTCCATGCCG
o3661	ACACTCGGTTATTCGCAA
o3662	GCCAGAGCGGCCGCCAGACTTGTACAGCTCGTCCATGCCG
o3935	TTTAACTTTAAGAAGGAGATATACAGGATCCAACGGTCCGGG
o3936	CGAATAACCGAGTGTCTTACCCGAGCCAGAACCCTTT
o4468	GAAGGAGATATACCGGTGAGCAAGGGCGAGGAGCTGTTCACCGGG
o4469	CCCGGTGAACAGCTCCTCGCCCTTGCTCACCGGTATATCTCCTTC
o4470	CTCTCGGCATGGACGAGCTAGCCAAGACACTCGGTTATTCGCAAC
o4471	GTTGCGAATAACCGAGTGTCTTGGCTAGCTCGTCCATGCCGAGAG
o4472	GCTCACTGCCCGCTTTCCACTGGTCGTCTCCAACGACCATCTGCT
o4473	AGCAGATGGTCGTTGGAGACGACCAGTGGAAAGCGGGCAGTGAGC
o4493	ATCCCGACCGGTAAGCTGCT
o4494	TAATCGGATCCGCCTTTGTC
o4495	AAGGCGGATCCGATTACAAA
o4496	CGGCTCTAGACCTTTGTCAT
o4587	GTTTAACTTTAAGAAGGAGATATACCATGAAGCTGCTAGCCTCGG
o4588	CCGAGGCTAGCAGCTTCATGGTATATCTCCTTCTTAAAGTTAAAC
o5028	CGACGGATCGGGAGCCGTTACATAACTTACGGTAAATGGCCC
o5029	GATGTCTGCTCGAAGCGCGGCCGCTTACT

Table S5. PCR primer.

qPCR

Locus	Name	Sequence (5'→ 3')
SatIII	o2419_fw	AATCAACCCGAGTGCAATCGAATGGAATCG
	o2420_rv	TCCATTCCATTCCTGTACTCGG
MUC4	o5181_fw	CCTCTTCATGTCACCGATGCTTCCTC
	o5182_rv	GGTGTGACCTGTGGATACTGAGGAAGG

Bisulfite PCR

Locus	Name	Sequence (5'→ 3')
MUC4	o5262_fw	ATAGGTTATATTATTTTTTTTTATGTTATCGA
	o5263_rv	TAATATAACCTATAAATACTAAAAAAAAACTAAT

Table S6. Plasmids.

Plasmid	Description / Gene	Resistance
pAnJ1861	TALE entry vector 3x-mCherry	Amp
pAlM1577	TALE entry vector 1x-mCherry	Amp
pAnI521	TALE entry vector 1x-GFP	Amp
pPiB2670	TALE entry vector 10x-FLAG	Amp
pPiB2683	TALE entry vector 20x-FLAG	Amp
pPiB2700	TALE entry vector 30x-FLAG	Amp
pCrW2056	TALE entry vector 24x-GCN4	Amp
pAlM1103	dCas9-24xGCN4	Amp
pCoT3181	DNMT3a3L + EBFP2	Amp
pAlH1894	DNMT3L	Amp
pJaW876	DNMT3a3L	Amp
pCoT3180	DNMT3a3L	Amp
pAlM1100	CMV-EBFP2	Kan
pAnJ3188	DNMT3a3L KO E756A	Amp
pAnJ3213	TALE_M1 in p521	Amp
pAnJ2842	TALE_M1 in p1861	Amp
pAnJ2807	TALE_M1 in p1577	Amp
pAnJ2803	TALE_M1 in p2683	Amp
pAnJ2805	TALE_M1 in p2700	Amp
pAnJ3281	TALE_M2 in p521	Amp
pAnJ3280	TALE_M2 in 1861	Amp
pAnJ3279	TALE_M2 in p1577	Amp
pAnJ3277	TALE_M2 in p2683	Amp
pAnJ3278	TALE_M2 in p2700	Amp
pAnJ2779	TALE_M3 HD in p521	Amp
pAnJ2780	TALE_M3 G* in p521	Amp
pAnJ3259	TALE_M3 HD in p1861	Amp
pAnJ3260	TALE_M3 G* in p1861	Amp
pAnJ2781	TALE_M3 HD in p1577	Amp
pAnJ2782	TALE_M3 G* in p1577	Amp
pAnJ3255	TALE_M3 HD in p2683	Amp
pAnJ3256	TALE_M3 HD in p2700	Amp
pAnJ3257	TALE_M3 G* in p2700	Amp

pAnJ3240	TALE_SATIII HD in p521	Amp
pAlM2873	TALE_SATIII G* in p521	Amp
pAnJ3237	TALE_SATIII HD in p1861	Amp
pAnJ3238	TALE_SATIII G* in p1861	Amp
pAnJ3248	TALE_SATIII HD in p1577	Amp
pAlM2355	TALE_SATIII G* in p1577	Amp
pAnJ3241	TALE_SATIII HD in p2683	Amp
pAnJ3244	TALE_SATIII G* in p2683	Amp
pAnJ3242	TALE_SATIII HD in p2700	Amp
pAnJ3245	TALE_SATIII G* in p2700	Amp
pAnJ3236	TALE_SATIII HD in p2056	Amp
pAnJ3243	TALE_SATIII G* in p2056	Amp
pAnJ3388	TALE_B1 HD in p1861	Amp
pAnJ3389	TALE_B1 G* in p1861	Amp
pAnJ3386	TALE_B1 HD in p1577	Amp
pAnJ3387	TALE_B1 G* in p1577	Amp
pAnJ3402	TALE_B2 HD in p1861	Amp
pAnJ3403	TALE_B2 G* in p1861	Amp
pAnJ3400	TALE_B2 HD in p1577	Amp
pAnJ3401	TALE_B2 G* in p1577	Amp
pAnJ3391	TALE_B3 in p1861	Amp
pAnJ3390	TALE_B3 in p1577	Amp
pAnJ3393	TALE_B4 in p1861	Amp
pAnJ3392	TALE_B4 in p1577	Amp
pAnJ3395	TALE_B5 in p1861	Amp
pAnJ3394	TALE_B5 in p1577	Amp
pAnJ3397	TALE_B6 in p1861	Amp
pAnJ3396	TALE_B6 in p1577	Amp

7.3 TALE Assembly

Table S7. TALE RVD composition.

TALE_M1

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
NG	NG	HD	HD	NG	HD	NI	NN	HD	NI	NG	HD	HD	NI	HD

TALE_M2

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
NG	NN	NN	NG	NN	NI	HD	NI	NN	NN	NI	NI	NN	NI	NN	NN	NN	NN

TALE_M3

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
HD	NG	NN	NN	HD	NN	NG	NN	NI	HD	HD	NG	NN	NG	NN	NN	NI	NG	NI
G*	NG	NN	NN	G*	NN	NG	NN	NI	HD	HD	NG	NN	NG	NN	NN	NI	NG	NI

TALE_SatIII

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
HD	NG	NN	NN	NI	NI	HD	NN	NN	NI	NI	NG	NN	NN	NI	NI	NG	NN	NN
G*	NG	NN	NN	NI	NI	G*	NN	NN	NI	NI	NG	NN	NN	NI	NI	NG	NN	NN

TALE_B1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
HD	NG	NI	NG	HD	HD	NI	HD	NI	NN	NN	NG	HD	NI	HD	NN	HD	HD	NI
G*	NG	NI	NG	HD	HD	NI	HD	NI	NN	NN	NG	HD	NI	HD	NN	HD	HD	NI

TALE_B2

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
HD	NG	HD	NI	HD	NN	HD	HD	NI	HD	HD	HD	HD	NG	HD	NG	NG	HD	HD
G*	NG	HD	NI	HD	NN	HD	HD	NI	HD	HD	HD	HD	NG	HD	NG	NG	HD	HD

TALE_B3

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
NG	HD	NI	NN	NG	NI	NG	HD	HD	NI	HD	NI	NN	NN	NG	HD	NI	HD

TALE_B4

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
NG	NN	NI	HD	HD	NG	NN	NG	NN	NN	NI	NG	NI	HD	NG	NN	NI	NN

TALE_B5

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
NG	NG	HD	HD	NG	HD	NI	NN	NG	NI	NG	HD	HD	NI	HD	NI	NN	HD

TALE_B6

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
NG	NN	NG	NN	NN	NI	NG	NI	HD	NG	NN	NI	NN	NN	NI	NI	NN	HD

8. Appendix

8.1 Epitope Tag Sequences

8.1.1 mCherry Sequence

ATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGT
GCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCC
CTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTG
GGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACAT
CCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGA
GGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCACTGCAGGACGGCGAGTTCATCTACAA
GGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGG
CTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCA
GAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAA
GAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAA
CGAGGACTACACCATCGTGGAACAGTACGAACGCGCTGAGGGCCGCCACTCCACCGGCGGCAT
GGACGAGCTGTACAAG

8.1.2 EGFP Sequence

ATGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGAC
GTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTATGGCAAGCTG
ACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCC
TGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCA
AGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACT
ACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGG
GCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCC
ACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCC
ACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCG
ACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCC
CAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGG
CATGGACGAGCTGTACAAG

8.1.3 3x-mCherry Sequence

Linker sequences are highlighted in gray.

5’ATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGG
GCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCC
CTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTG
GGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACAT
CCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGA
GGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCACTGCAGGACGGCGAGTTCATCTACAA
GGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGG
CTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCA
GAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAA
GAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAA
CGAGGACTACACCATCGTGGAACAGTACGAACGCGCTGAGGGCCGCCACTCCACCGGCGGCAT
GGACGAGCTGTACAAGTCTGGCGGCCGCTCTGGCATGGTGAGCAAGGGCGAGGAGGATAACA
TGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACG
AGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGA
AGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACG
GCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGA
GGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGA
CTCCTCACTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTC
CGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCC
CGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACT
ACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACA
ACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACG
AACGCGCTGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGGGTAGCATCGATGG
TAGCATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGC
ACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAG
GGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCC
CCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTG
TCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACC
CAGGACTCCTCACTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCC
GACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGAC
GGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGT
CAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGA
CATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCTGAGGGCCGCCACTCCAC
CGGCGGCATGGACGAGCTGTACAAG-3’

8.1.4 5x-FLAG Tag Sequence

FLAG sequences are highlighted in gray.

Insert_1

5’ATCCCGACCGGTAAGCTGCTAGCCTCGGACTACAAAGACGACGACGACAAGGGTGGTTCTG
ATTATAAGGACGATGACGATAAAGGAGGTTCCGACTATAAAGATGATGATGACAAAGGAGGG
TCAGATTACAAGGATGACGACGATAAGGGGGGCTCGGATTACAAGGACGACGATGACAAAGG
CGGATCCGATTA-3’

Insert_2

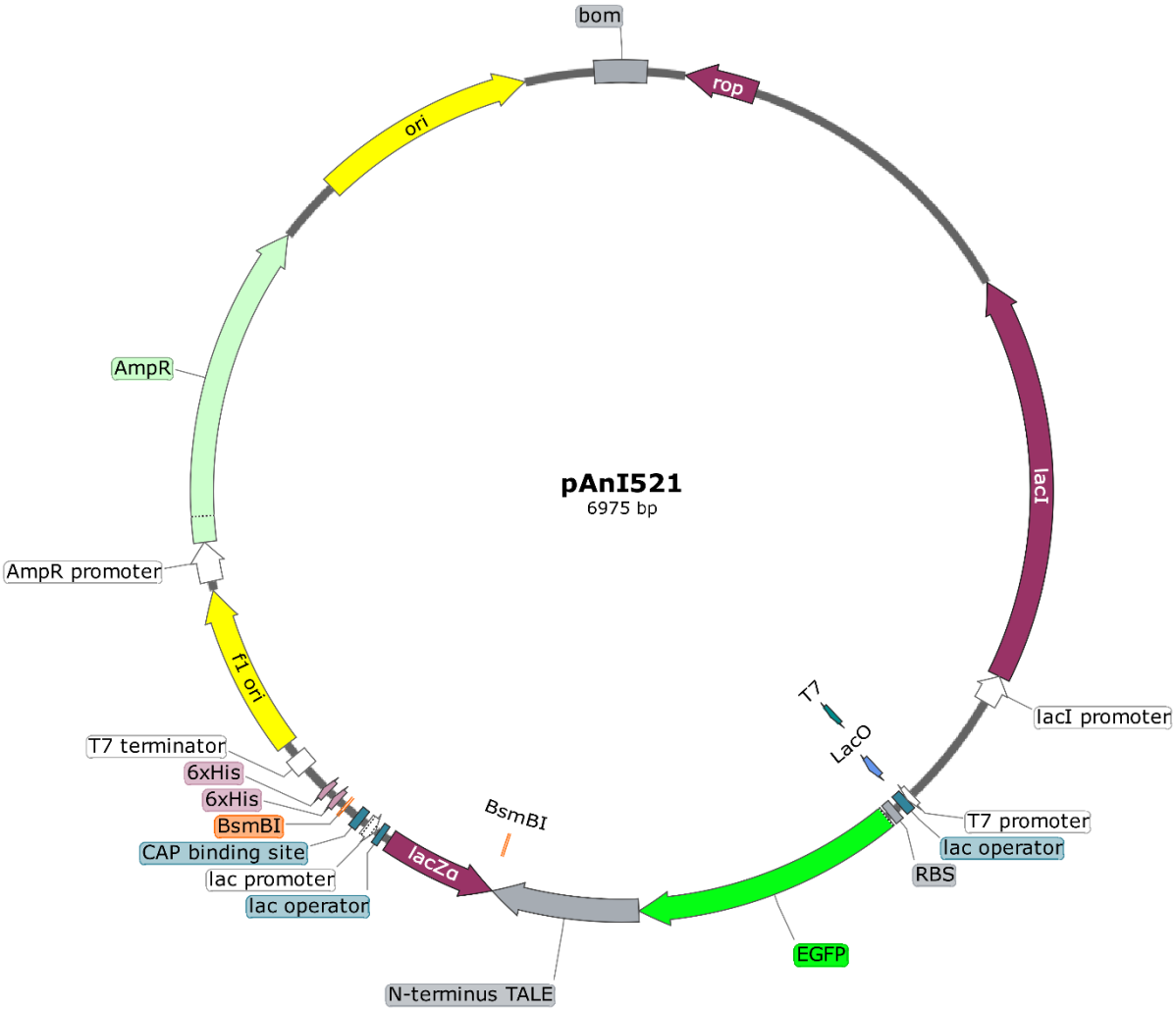
5’AAGGCGGATCCGATTACAAAGACGATGACGACAAGGGTGGATCAGATTATAAGGACGATGA
TGATAAAGGCGGGTCTGACTATAAAGATGATGATGACAAAGGGGGCTCGGACTACAAGGATG
ACGACGATAAGGGCGGCTCAGACTATAAGGACGACGATGACAAAGGTCTAGAGCCG-3’

8.1.5 24x-GCN4 (SunTag) Sequence

Linker sequences are highlighted in gray.

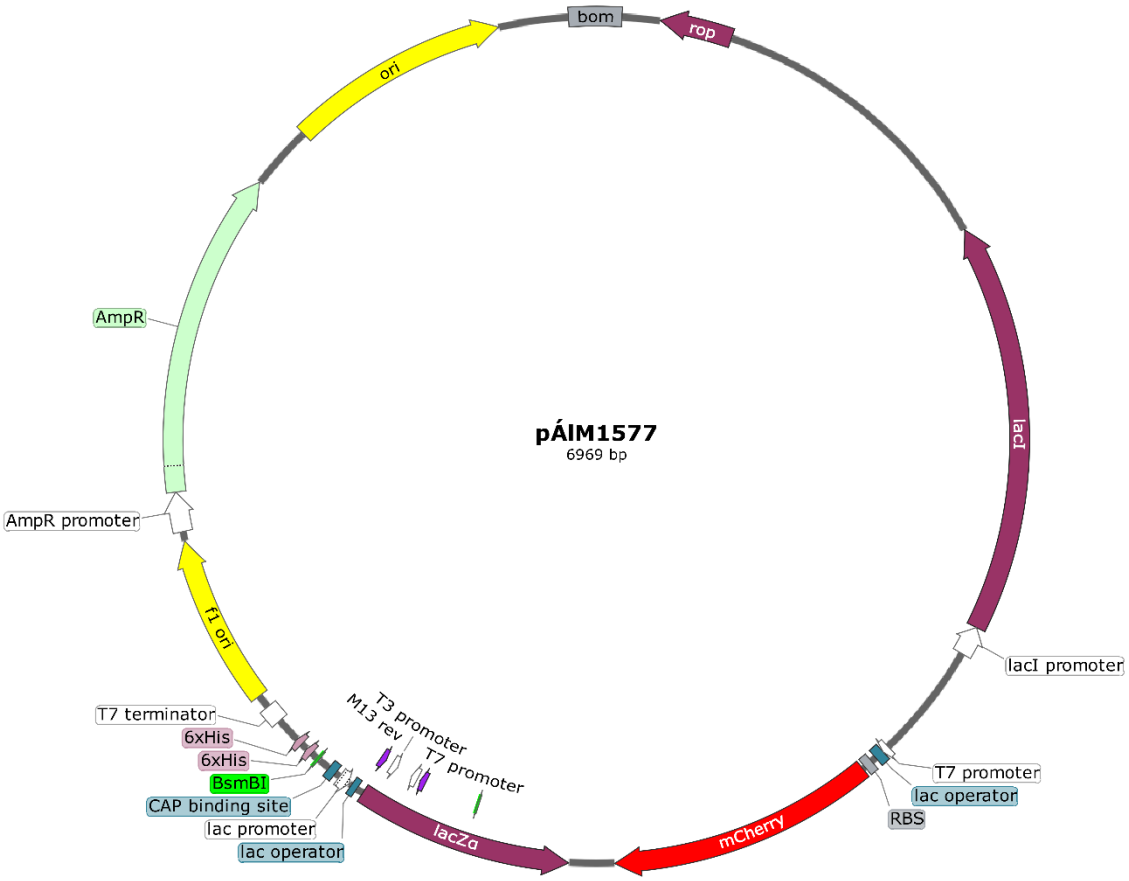
5’GAAGAACTTTTGAGCAAGAATTATCATCTTGAGAACGAAGTGGCTCGTCTTAAGAAAGGTT
CTGGCAGTGGAGAAGAACTGCTTTCAAAGAATTACCACCTGGAAAATGAGGTAGCTAGACTG
AAAAAGGGGAGCGGAAGTGGGGAGGAGTTGCTGAGCAAAAATTATCATTTGGAGAACGAAGT
AGCACGACTAAAGAAAGGGTCCGGATCGGGTGAGGAGTTACTCTCGAAAAATTATCATCTCG
AAAACGAAGTGGCTCGGCTAAAAAAGGGCAGTGGTTCTGGAGAAGAGCTATTATCTAAAAAC
TACCACCTCGAAAATGAGGTGGCACGCTTAAAAAAGGGAAGTGGCAGTGGTGAAGAGCTACT
ATCCAAGAATTATCATCTTGAGAACGAGGTAGCGCGTTTGAAGAAGGGTTCCGGCTCAGGAG
AGGAACTGCTCTCGAAGAACTATCATCTTGAAAATGAGGTCGCTCGATTAAAAAAGGGATCG
GGCAGTGGTGAGGAACTACTTTCAAAGAATTACCACCTCGAAAACGAAGTAGCTCGATTAAA
GAAAGGTTCAGGGTCGGGTGAAGAATTACTGAGTAAAAATTATCATCTGGAAAATGAGGTAG
CGAGACTAAAAAAGGGGAGTGGTTCTGGCGAAGAGTTGCTATCGAAAAATTATCATCTTGAG
AACGAAGTTGCTAGGCTCAAAAAGGGCTCAGGCTCAGGCGAGGAGTTGCTCTCGAAAAACTA
CCACTTGGAAAATGAGGTCGCGAGGTTGAAAAAGGGGAGCGGGTCGGGCGAGGAGTTATTGA
GCAAAAACTATCATTTAGAGAACGAAGTCGCGCGCTTAAAGAAAGGCTCGGGCTCGGGCGAA
GAACTCTTATCGAAGAACTACCACCTCGAAAATGAGGTCGCCAGGTTGAAAAAGGGCAGTGG
CAGCGGGGAGGAACTCTTGAGCAAGAACTACCACTTGGAGAATGAGGTCGCGAGATTGAAGA
AAGGGTCGGGGAGCGGCGAGGAATTGCTCAGCAAGAATTATCATTTGGAGAACGAAGTCGCC
AGGCTCAAGAAAGGCTCGGGGTCGGGGGAGGAATTGTTGAGTAAAAACTACCACTTGGAAAA
TGAAGTCGCCAGGCTCAAAAAAGGGAGTGGGAGCGGCGAAGAGTTATTGAGCAAAAATTACC
ACTTGGAGAACGAAGTGGCAAGGCTCAAGAAAGGGAGCGGCAGCGGGGAGGAGCTCTTATCG
AAGAACTACCACTTAGAGAATGAAGTCGCCCGCTTGAAGAAAGGCTCGGGGAGCGGGGAAGA
GCTCTTGAGCAAGAACTACCACTTGGAAAATGAGGTGGCGCGCTTGAAGAAAGGGAGCGGGA
GCGGGGAAGAGTTACTATCTAAGAATTATCATCTCGAGAACGAGGTGGCTCGACTAAAGAAG
GGCTCCGGCAGTGGGGAGGAACTCCTGTCGAAGAACTATCATCTTGAAAATGAGGTTGCAAG
ACTTAAAAAGGGGTCCGGATCAGGTGAGGAACTACTCAGTAAGAATTACCACCTGGAAAACG
AAGTTGCACGTTTGAAGAAAGGATCAGGATCAGGCGAAGAACTGCTCTCAAAAGATTATCAT
TTGGAAAATGAGGTTGCACGTTTAAAAAAGGGAAGTGGCAGTGGTGAGGAACTTCTGTCGAA
AAATTATCATCTCGAGAATGAAGTAGCCCGACTTAAAAAGGGTTCTGGCTCGGGT-3’

8.2 Plasmid Maps



pAnI521

GG2 Entry Vector for bacterial expression of TALEs fused to N-terminal EGFP and C-terminal 6xHis for protein purification.

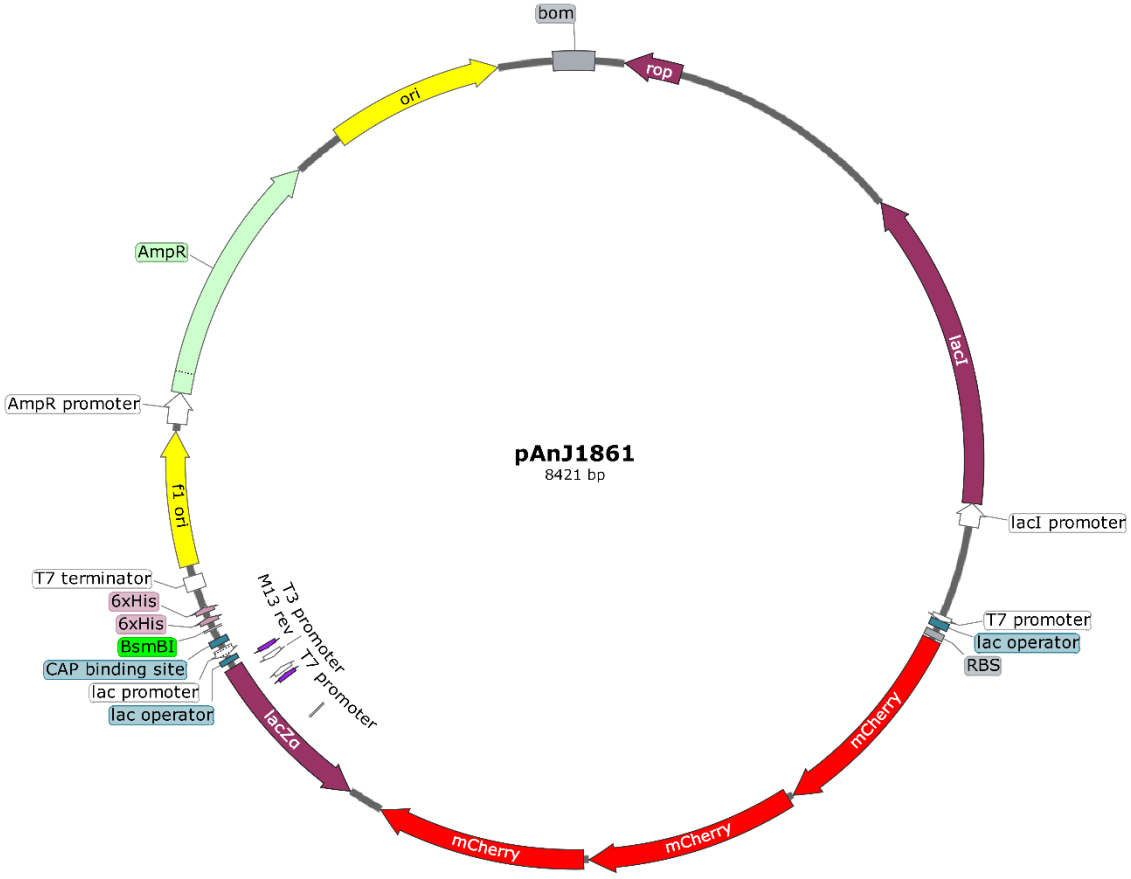


pAlM1577

Vector for bacterial expression of TALEs fused to N-terminal mCherry and C-terminal 6xHis for protein purification.

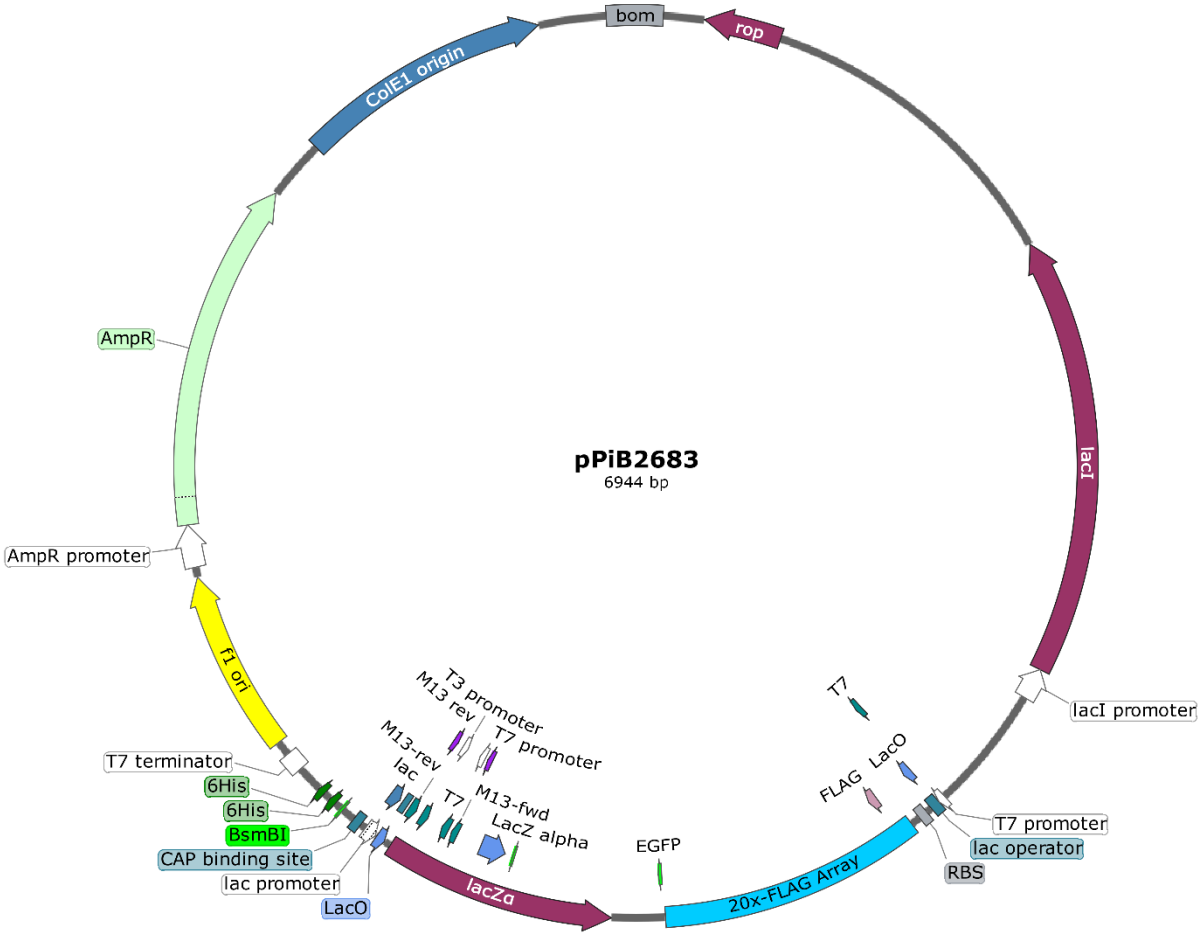
pAnJ1860

Vector for bacterial expression of TALEs fused to three N-terminal sfGFP and C-terminal 6xHis for protein purification.



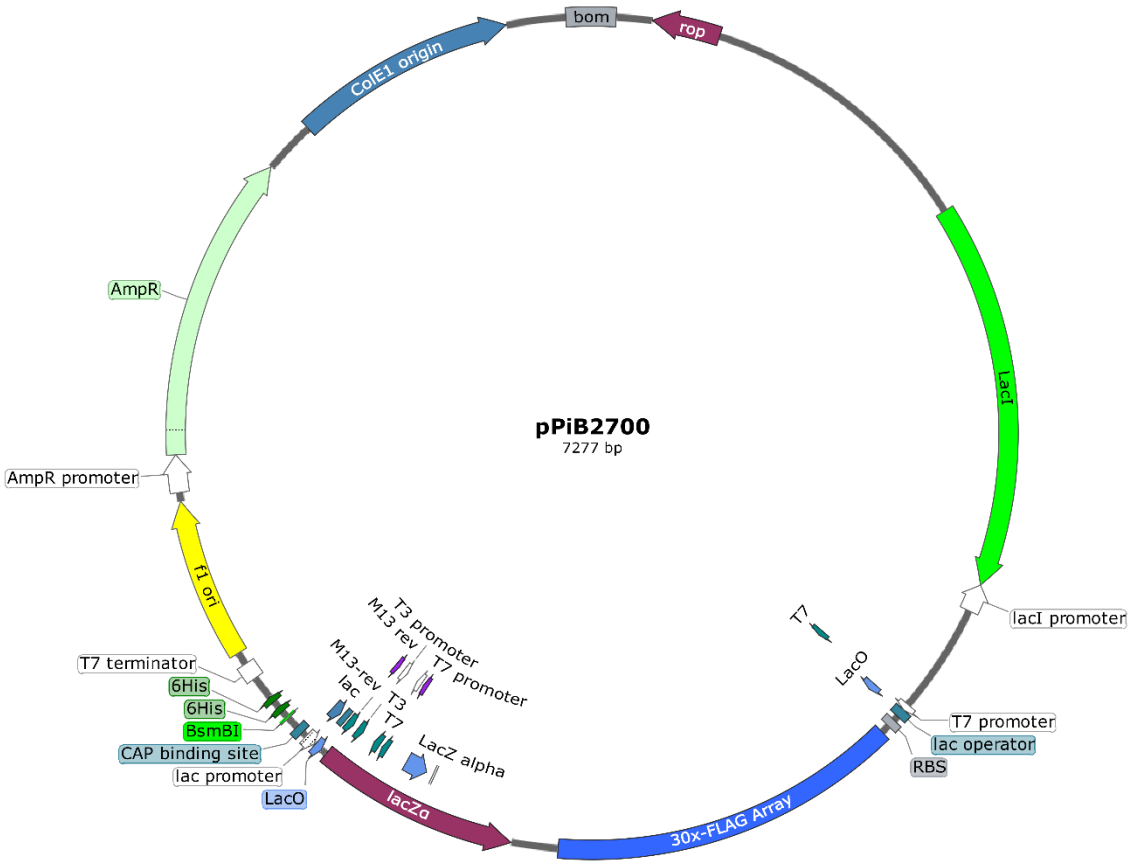
pAnJ1861

Vector for bacterial expression of TALEs fused to three N-terminal mCherry and C-terminal 6xHis for protein purification.



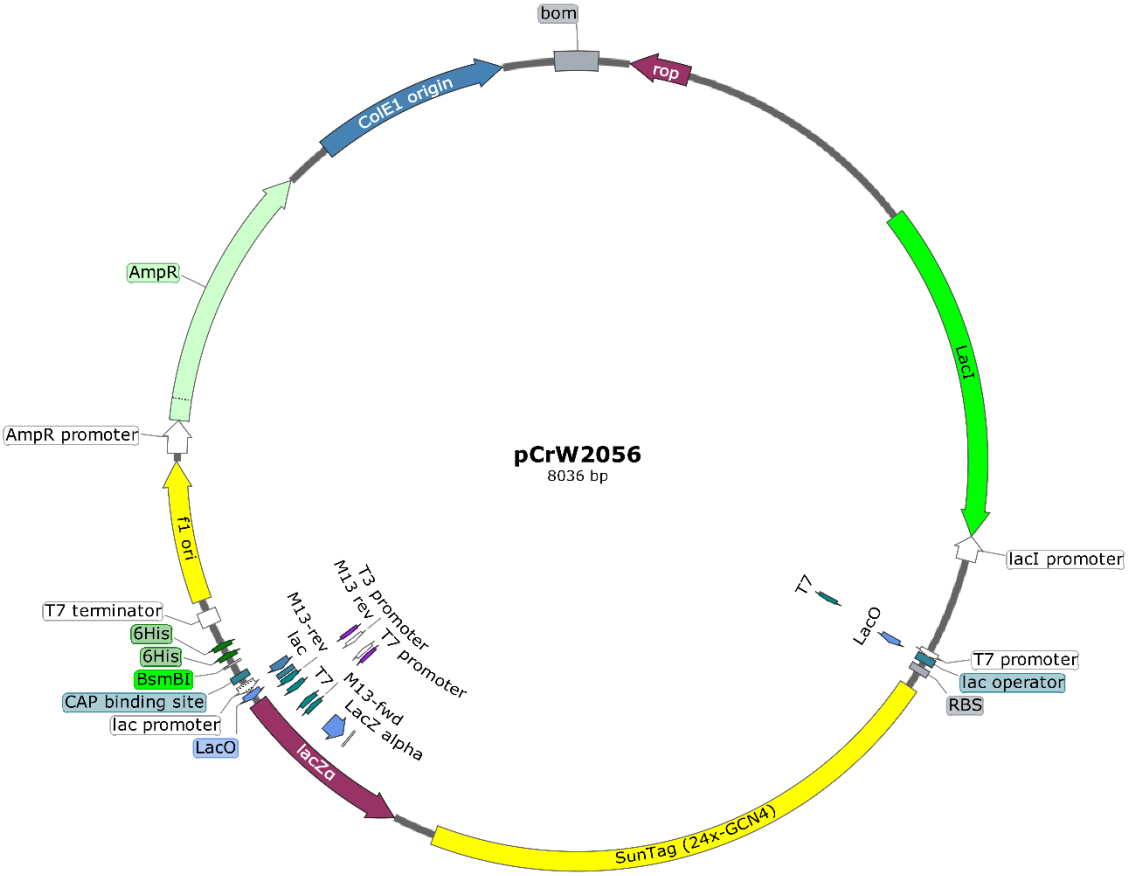
pPiB2683

Vector for bacterial expression of TALEs fused to 20 N-terminal FLAG tags and C-terminal 6xHis for protein purification.



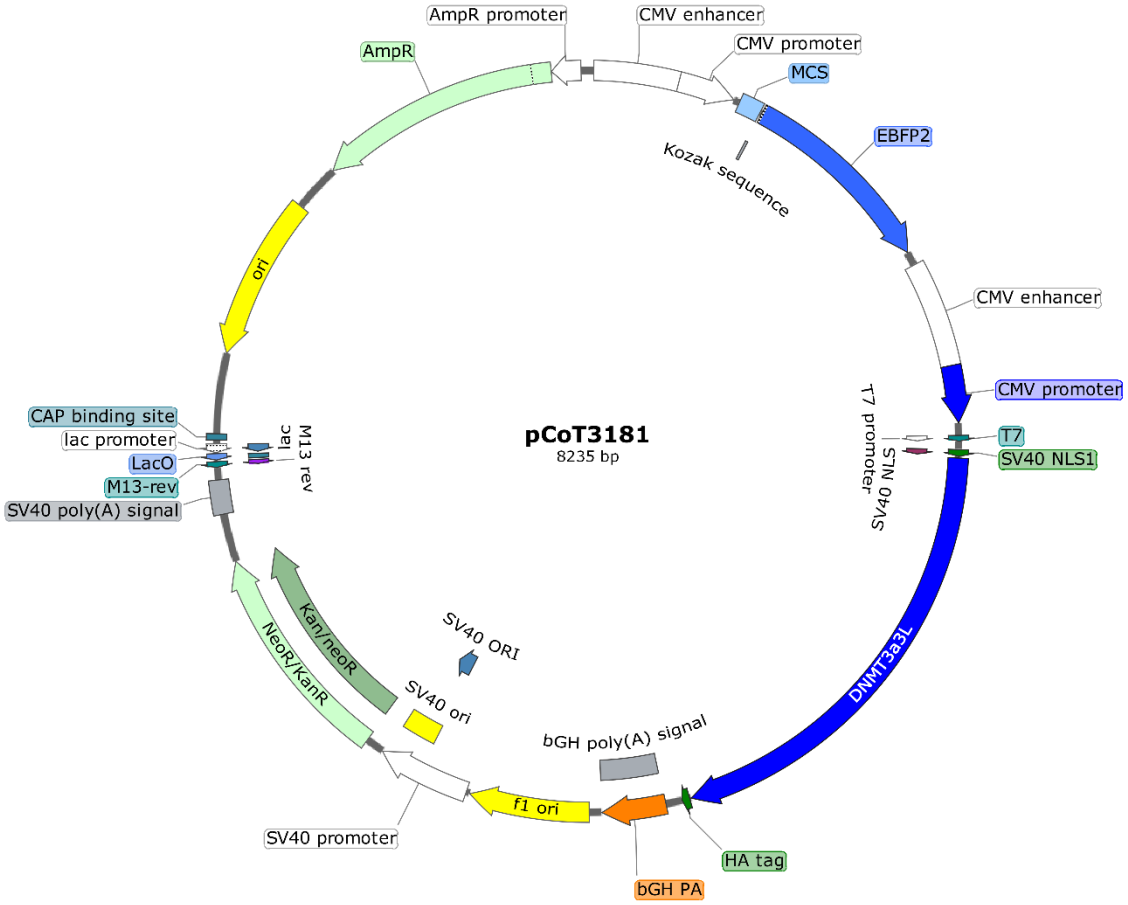
pPiB2700

Vector for bacterial expression of TALEs fused to 30 N-terminal FLAG tags and C-terminal 6xHis for protein purification.



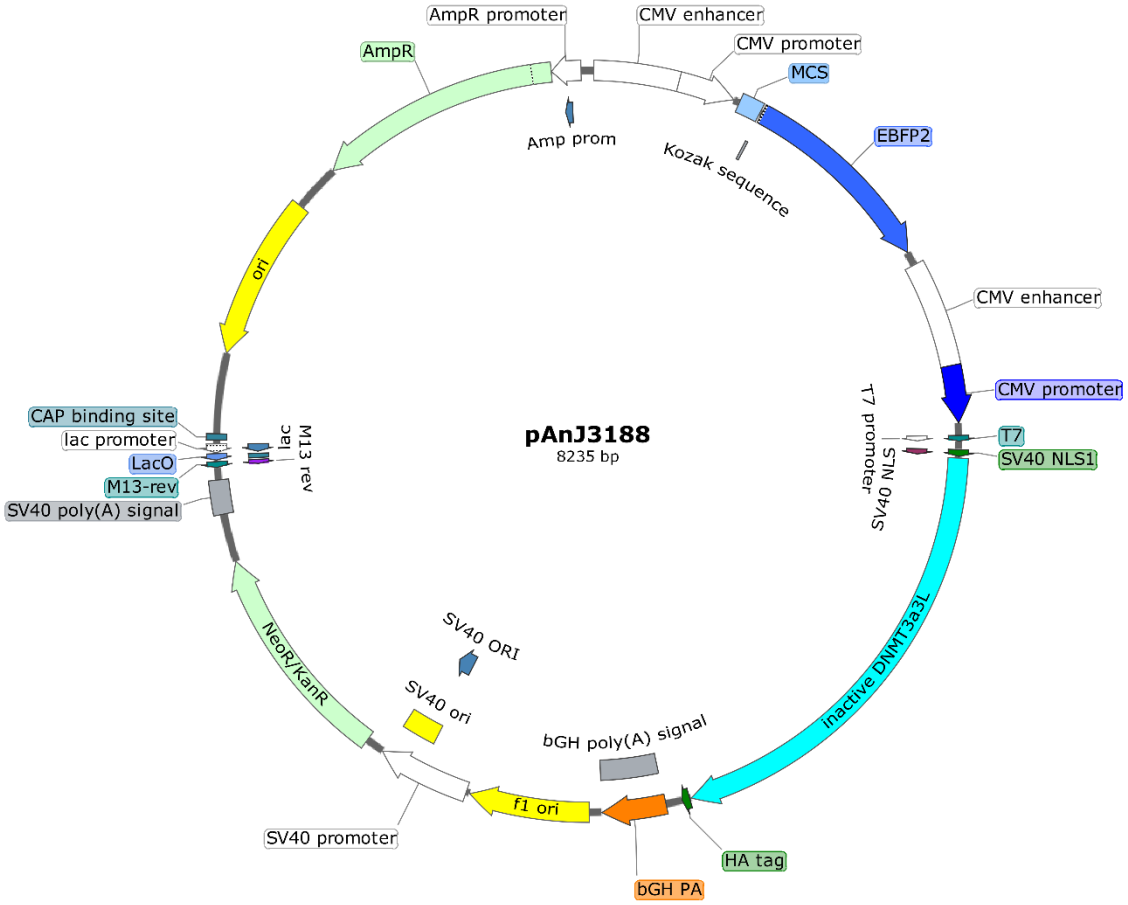
pCrW2056

Vector for bacterial expression of TALEs fused to N-terminal SunTag (24x-GCN4) and C-terminal 6xHis for protein purification.



pCoT3181

Vector for mammalian expression of active DNMT3a3L for live-cell DNA methylation. As transfection control, the EBFP2 fluorophore stands under an independent promoter (CMV).



pAnJ3188

Vector for mammalian expression of inactive DNMT3a3L E756A for live-cell DNA methylation. As transfection control, the EBFP2 fluorophore stands under an independent promoter (CMV).

8.3 Credits and Copyright Licenses

The content of Chapter 4 and Supplementary has been published in (A. Jung et al., 2023) and licensed under CC BY 4.0, which allows to copy and redistribute the material in any medium or format and remix, transform as well as build upon the material for any purpose, even commercially. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/. The reprinted figures are listed below:

[1] **Figure 19, Figures 21-28,** and **Figures S1-S11** are reprinted from (A. Jung, 2023). Copyright © 2023 The Authors, published by American Chemical Society.

[2] **Figure 2** is reprinted with permission from (L. Pray, 2008). Copyright © 2013 Nature Education.

[3] **Figure 3** is modified from (Pierce, 2013). Copyright © 2013 Nature Education.

[4] **Figure 4** is adapted with permission from (Bhaumik et al., 2007) Copyright © 2007, Nature Publishing Group (license number 5491510363459)

[5] **Figure 6** is adapted with permission from (Huntriss, 2021). Copyright © 2021 Elsevier Inc. (License number 5465941160580)

[6] **Figure 7** is created based on (Muñoz-López et al., 2017) Copyright © 2018 The Chemical Society of Japan & Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim. (License number 5471331037787)

[7] **Figure 8** is created based on (Muñoz-López et al., 2020) Copyright © 2020 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. The original figure is licensed under CC BY 4.0.

[8] **Figure 10** is adapted with permission from (L. DeFrancesco, 2012). Copyright © 2012, Nature Publishing Group, a division of Macmillan Publishers Limited. (License number 5440261453932)

[9] **Figure 11** is adapted with permission from (Becker et al., 2016). Copyright © 2016, Nature Publishing Group, a division of Macmillan Publishers Limited. (License number 5444220682609)

[10] **Figure 12** is adapted with permission from (Becker and Boch, 2021). Copyright © 1969, Elsevier. The original figure is licensed under CC BY 4.0.

[11] **Figure 13** is modified with permission from (Muñoz-López et al., 2020). Copyright © 2020, The Authors. ChemBioChem published by Wiley-VCH GmbH. The original figure is licensed under CC BY 4.0.

[12] **Figure 13** is modified with permission from (Muñoz-López et al., 2020). Copyright © 2020, The Authors. ChemBioChem published by Wiley-VCH GmbH. The original figure is licensed under CC BY 4.0.

[13] **Figure 15** is modified with permission from (Muñoz-López et al., 2020). Copyright © 2020, The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. The original figure is licensed under CC BY 4.0.

[14] **Figure 17a** is modified with permission from (Yeon et al., 2022). Copyright © 2022, The Author(s). The original figure is licensed under CC BY 4.0.

[15] **Figure 18** is adapted with permission from (Chaturvedi et al., 2010). Copyright © 2007, FASEB. (License number 5476450182833)

9. References

(1) Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. *Molecular Biology of the Cell*; Garland Science, **2002**; Vol. 4th edition.

(2) Watson, J. D.; Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature 1953 171:4356* **1953**, *171* (4356), 737–738.

(3) Nirenberg, M. W.; Matthaei, J. H. The Dependence of Cell-Free Protein Synthesis in E. Coli upon Naturally Occurring or Synthetic Polyribonucleotides. *Proc Natl Acad Sci U S A* **1961**, *47* (10), 1588–1602.

(4) Pauling, L.; Corey, R. B. A Proposed Structure For The Nucleic Acids. *Proc Natl Acad Sci U S A* **1953**, *39* (2), 84.

(5) Pray, L. Discovery of DNA Structure and Function: Watson and Crick. *Nature Education* **2008**, *1* (1), 1–100.

(6) Svozil, D.; Kalina, J.; Omelka, M.; Schneider, B. DNA Conformations and Their Sequence Preferences. *Nucleic Acids Res* **2008**, *36* (11), 3690–3706.

(7) Wing, R.; Drew, H.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. Crystal Structure Analysis of a Complete Turn of B-DNA. *Nature 1980 287:5784* **1980**, *287* (5784), 755–758.

(8) Chargaff, E. Chemical Specificity of Nucleic Acids and Mechanism of Their Enzymatic Degradation. *Experientia* **1950**, *6* (6), 201–209.

(9) Pauling, L.; Corey, R. B. Specific Hydrogen-Bond Formation between Pyramidines and Purines in Deoxyribonucleic Acids. *Arch Biochem Biophys* **1956**, *65* (1), 164–181.

(10) Tjio, J. H.; Levan, A. The Chromosome Number Of Man. *Hereditas* **1956**, *42* (1–2), 1–6.

(11) Ford, C. E.; Hamerton, J. L. The Chromosomes of Man. *Nature* **1956**, *178* (4541).

(12) National Research Council (US) Committtee. *Mapping and Sequencing the Human Genome*; National Academies Press (US), **1988**.

(13) Sun, H. bin; Shen, J.; Yokota, H. Size-Dependent Positioning of Human Chromosomes in Interphase Nuclei. *Biophys J* **2000**, *79* (1), 184–190.

(14) Maeshima, K.; Ide, S.; Babokhov, M. Dynamic Chromatin Organization without the 30-Nm Fiber. *Curr Opin Cell Biol* **2019**, *58*, 95–104.

(15) McGinty, R. K.; Tan, S. Nucleosome Structure and Function. *Chem Rev* **2015**, *115* (6), 2255–2273.

(16) Mariño-Ramírez, L.; Kann, M. G.; Shoemaker, B. A.; Landsman, D. Histone Structure and Nucleosome Stability. *Expert Rev Proteomics* **2005**, *2* (5), 719–729.

(17) van Holde, K. E. *Chromatin. Series in Molecular Biology.*, 3rd ed.; Morton Bradbury, E., Ed.; Springer-Verlag: New York, **1989**; Vol. 2.

(18) Pierce, B. *Genetics: A Conceptual Approach*, 2nd ed.; Macmillan, P., Ed.; W. H. Freeman, **2013**.

(19) Tamaru, H. Confining Euchromatin/Heterochromatin Territory: Jumonji Crosses the Line. *Genes Dev* **2010**, *24* (14), 1465.

(20) Thakur, J.; Packiaraj, J.; Henikoff, S. Sequence, Chromatin and Evolution of Satellite Dna. *Int J Mol Sci* **2021**, *22* (9), 4309.

(21) Bannister, A. J.; Kouzarides, T. Regulation of Chromatin by Histone Modifications. *Cell Research 2011 21:3* **2011**, *21* (3), 381–395.

(22) Bhaumik, S. R.; Smith, E.; Shilatifard, A. Covalent Modifications of Histones during Development and Disease Pathogenesis. *Nature Structural & Molecular Biology 2007 14:11* **2007**, *14* (11), 1008–1016.

(23) Roth, S. Y.; Denu, J. M.; Allis, C. D. Histone Acetyltransferases. *Annu Rev Biochem* **2003**, *70*, 81–120.

(24) Greer, E. L.; Shi, Y. Histone Methylation: A Dynamic Mark in Health, Disease and Inheritance. *Nat Rev Genet* **2012**, *13* (5), 343.

(25) Rossetto, D.; Avvakumov, N.; Côté, J. Histone Phosphorylation: A Chromatin Modification Involved in Diverse Nuclear Events. *Epigenetics* **2012**, *7* (10), 1098.

(26) Cao, J.; Yan, Q. Histone Ubiquitination and Deubiquitination in Transcription, DNA Damage Response, and Cancer. *Front Oncol* **2012**, *2* (26).

(27) Yun, M.; Wu, J.; Workman, J. L.; Li, B. Readers of Histone Modifications. *Cell Res* **2011**, *21* (4), 564–578.

(28) Jenuwein, T.; Allis, C. D. Translating the Histone Code. *Science* **2001**, *293* (5532), 1074–1080.

(29) Strahl, B. D.; Allis, C. D. The Language of Covalent Histone Modifications. *Nature* **2000**, *403* (6765).

(30) Rando, O. J. Combinatorial Complexity in Chromatin Structure and Function: Revisiting the Histone Code. *Curr Opin Genet Dev* **2012**, *22* (2), 148–155.

(31) Jin, B.; Li, Y.; Robertson, K. D. DNA Methylation: Superior or Subordinate in the Epigenetic Hierarchy? *Genes Cancer* **2011**, *2* (6), 607.

(32) Moore, L. D.; Le, T.; Fan, G. DNA Methylation and Its Basic Function. *Neuropsychopharmacology* **2013**, *38* (1), 38.

(33) Medvedeva, Y. A.; Khamis, A. M.; Kulakovskiy, I. v.; Ba-Alawi, W.; Bhuyan, M. S. I.; Kawaji, H.; Lassmann, T.; Harbers, M.; Forrest, A. R. R.; Bajic, V. B. Effects of Cytosine Methylation on Transcription Factor Binding Sites. *BMC Genomics* **2014**, *15* (1), 1–12.

(34) Nan, X.; Ng, H. H.; Johnson, C. A.; Laherty, C. D.; Turner, B. M.; Eisenman, R. N.; Bird, A. Transcriptional Repression by the Methyl-CpG-Binding Protein MeCP2 Involves a Histone Deacetylase Complex. *Nature 1998 393:6683* **1998**, *393* (6683), 386–389.

(35) Suzuki, M. M.; Bird, A. DNA Methylation Landscapes: Provocative Insights from Epigenomics. *Nature Reviews Genetics 2008 9:6* **2008**, *9* (6), 465–476.

(36) Rose, N. R.; Klose, R. J. Understanding the Relationship between DNA Methylation and Histone Lysine Methylation. *Biochim Biophys Acta Gene Regul Mech* **2014**, *1839* (12), 1362–1372.

(37) Handy, D. E.; Castro, R.; Loscalzo, J. Epigenetic Modifications: Basic Mechanisms and Role in Cardiovascular Disease. *Circulation* **2011**, *123* (19), 2156–2156.

(38) Vaissière, T.; Sawan, C.; Herceg, Z. Epigenetic Interplay between Histone Modifications and DNA Methylation in Gene Silencing. *Mutat Res* **2008**, *659* (1–2), 40–48. https://doi.org/10.1016/J.MRREV.2008.02.004. Y.; Chen, X.; Lu, C. The Interplay between DNA and Histone Methylation: Molecular Mechanisms and Disease Implications. *EMBO Rep* **2021**, *22* (5), e51803.

(40) Wu, J. C.; Santi, D. v. Kinetic and Catalytic Mechanism of HhaI Methyltransferase. *Journal of Biological Chemistry* **1987**, *262* (10), 4778–4786.

(41) Gowher, H.; Jeltsch, A. Mammalian DNA Methyltransferases: New Discoveries and Open Questions. *Biochem Soc Trans* **2018**, *46* (5), 1202.

(42) Coulondre, C.; Miller, J. H.; Farabaugh, P. J.; Gilbert, W. Molecular Basis of Base Substitution Hotspots in Escherichia Coli. *Nature* **1978**, *274* (5673), 775–780.

(43) Bird, A. P. DNA Methylation and the Frequency of CpG in Animal DNA. *Nucleic Acids Res* **1980**, *8* (7), 1504.

(44) Ziller, M. J.; Gu, H.; Müller, F.; Donaghey, J.; Tsai, L. T. Y.; Kohlbacher, O.; de Jager, P. L.; Rosen, E. D.; Bennett, D. A.; Bernstein, B. E.; Gnirke, A.; Meissner, A. Charting a Dynamic DNA Methylation Landscape of the Human Genome. *Nature* **2013**, *500* (7463), 477–481.

(45) Ehrlich, M.; Gama-Sosa, M. A.; Huang, L. H.; Midgett, R. M.; Kuo, K. C.; Mccune, R. A.; Gehrke, C. Amount and Distribution of 5-Methylcytosine in Human DNA from Different Types of Tissues of Cells. *Nucleic Acids Res* **1982**, *10* (8), 2709–2721.

(46) Deaton, A. M.; Bird, A. CpG Islands and the Regulation of Transcription. *Genes Dev* **2011**, *25* (10), 1022.

(47) Gardiner-Garden, M.; Frommer, M. CpG Islands in Vertebrate Genomes. *J Mol Biol* **1987**, *196* (2), 261–282.

(48) Illingworth, R. S.; Gruenewald-Schneider, U.; Webb, S.; Kerr, A. R. W.; James, K. D.; Turner, D. J.; Smith, C.; Harrison, D. J.; Andrews, R.; Bird, A. P. Orphan CpG Islands Identify Numerous Conserved Promoters in the Mammalian Genome. *PLoS Genet* **2010**, *6* (9).

(49) Tazi, J.; Bird, A. Alternative Chromatin Structure at CpG Islands. *Cell* **1990**, *60* (6), 909–920.

(50) Ramasamy, D.; Rao, A. K. D. M.; Rajkumar, T.; Mani, S. Experimental and Computational Approaches for Non-CpG Methylation Analysis. *Epigenomes 2022, Vol. 6, Page 24* **2022**, *6* (3), 24.

(51) Ramsahoye, B. H.; Biniszkiewicz, D.; Lyko, F.; Clark, V.; Bird, A. P.; Jaenisch, R. Non-CpG Methylation Is Prevalent in Embryonic Stem Cells and May Be Mediated by DNA Methyltransferase 3a. *Proc Natl Acad Sci U S A* **2000**, *97* (10), 5237–5242.

(52) de Mendoza, A.; Poppe, D.; Buckberry, S.; Pflueger, J.; Albertin, C. B.; Daish, T.; Bertrand, S.; de la Calle-Mustienes, E.; Gómez-Skarmeta, J. L.; Nery, J. R.; Ecker, J. R.; Baer, B.; Ragsdale, C. W.; Grützner, F.; Escriva, H.; Venkatesh, B.; Bogdanovic, O.; Lister, R. The Emergence of the Brain Non-CpG Methylation System in Vertebrates. *Nature Ecology & Evolution 2021 5:3* **2021**, *5* (3), 369–378.

(53) Lee, J. H.; Saito, Y.; Park, S. J.; Nakai, K. Existence and Possible Roles of Independent Non-CpG Methylation in the Mammalian Brain. *DNA Research* **2020**, *27* (4), 1–11.

(54) Villicaña, S.; Bell, J. T. Genetic Impacts on DNA Methylation: Research Findings and Future Perspectives. *Genome Biology 2021 22:1* **2021**, *22* (1), 1–35.

(55) Huntriss, J. Epigenetic Reprogramming in the Embryo. In *Epigenetics and Reproductive Health*; Academic Press, 2021; pp 97–116. https://doi.org/10.1016/B978-0-12-819753-0.00006-4.

(56) Santos, F.; Dean, W. Epigenetic Reprogramming during Early Development in Mammals. *Reproduction* **2004**, *127* (6), 643–651.

(57) Tucker, K. L.; Beard, C.; Dausman, J.; Jackson-Grusby, L.; Laird, P. W.; Lei, H.; Li, E.; Jaenisch, R. Germ-Line Passage Is Required for Establishment of Methylation and Expression Patterns of Imprinted but Not of Nonimprinted Genes. *Genes Dev* **1996**, *10* (8), 1008–1020.

(58) Mayer, W.; Niveleau, A.; Walter, J.; Fundele, R.; Haaf, T. Demethylation of the Zygotic Paternal Genome. *Nature 2000 403:6769* **2000**, *403* (6769), 501–502.

(59) Santos, F.; Hendrich, B.; Reik, W.; Dean, W. Dynamic Reprogramming of DNA Methylation in the Early Mouse Embryo. *Dev Biol* **2002**, *241* (1), 172–182.

(60) Dean, W.; Santos, F.; Stojkovic, M.; Zakhartchenko, V.; Walter, J.; Wolf, E.; Reik, W. Conservation of Methylation Reprogramming in Mammalian Development: Aberrant Reprogramming in Cloned Embryos. *Proc Natl Acad Sci U S A* **2001**, *98* (24), 13734–13738.

(61) SanMiguel, J. M.; Bartolomei, M. S. DNA Methylation Dynamics of Genomic Imprinting in Mouse Development. *Biol Reprod* **2018**, *99* (1), 252–262.

(62) Okano, M.; Xie, S.; Li, E. Cloning and Characterization of a Family of Novel Mammalian DNA (Cytosine-5) Methyltransferases. *Nat Genet* **1998**, *19* (3), 219–220.

(63) Okano, M.; Bell, D. W.; Haber, D. A.; Li, E. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for de Novo Methylation and Mammalian Development. *Cell* **1999**, *99* (3), 247–257.

(64) Mao, S. Q.; Cuesta, S. M.; Tannahill, D.; Balasubramanian, S. Genome-Wide DNA Methylation Signatures Are Determined by DNMT3A/B Sequence Preferences. *Biochemistry* **2020**, *59* (27), 2541–2550.

(65) Xie, S.; Wang, Z.; Okano, M.; Nogami, M.; Li, Y.; He, W. W.; Okumura, K.; Li, E. Cloning, Expression and Chromosome Locations of the Human DNMT3 Gene Family. *Gene* **1999**, *236* (1), 87–95.

(66) Aapola, U.; Shibuya, K.; Scott, H. S.; Ollila, J.; Vihinen, M.; Heino, M.; Shintani, A.; Kawasaki, K.; Minoshima, S.; Krohn, K.; Antonarakis, S. E.; Shimizu, N.; Kudoh, J.; Peterson, P. Isolation and Initial Characterization of a Novel Zinc Finger Gene, DNMT3L, on 21q22.3, Related to the Cytosine-5-Methyltransferase 3 Gene Family. *Genomics* **2000**, *65* (3), 293–298.

(67) Saitou, M.; Kagiwada, S.; Kurimoto, K. Epigenetic Reprogramming in Mouse Pre-Implantation Development and Primordial Germ Cells. *Development* **2012**, *139* (1), 15–31.

(68) Veland, N.; Lu, Y.; Hardikar, S.; Gaddis, S.; Zeng, Y.; Liu, B.; Estecio, M. R.; Takata, Y.; Lin, K.; Tomida, M. W.; Shen, J.; Saha, D.; Gowher, H.; Zhao, H.; Chen, T. DNMT3L Facilitates DNA Methylation Partly by Maintaining DNMT3A Stability in Mouse Embryonic Stem Cells. *Nucleic Acids Res* **2019**, *47* (1), 152–167.

(69) Pradhan, S.; Bacolla, A.; Wells, R. D.; Roberts, R. J. Recombinant Human DNA (Cytosine-5) Methyltransferase. I. Expression, Purification, and Comparison of de Novo and Maintenance Methylation. *J Biol Chem* **1999**, *274* (46), 33002–33010.

(70) Hermann, A.; Goyal, R.; Jeltsch, A. The Dnmt1 DNA-(Cytosine-C5)-Methyltransferase Methylates DNA Processively with High Preference for Hemimethylated Target Sites. *J Biol Chem* **2004**, *279* (46), 48350–48359.

(71) Leonhardt, H.; Page, A. W.; Weier, H. U.; Bestor, T. H. A Targeting Sequence Directs DNA Methyltransferase to Sites of DNA Replication in Mammalian Nuclei. *Cell* **1992**, *71* (5), 865–873.

(72) Mortusewicz, O.; Schermelleh, L.; Walter, J.; Cardoso, M. C.; Leonhardt, H. Recruitment of DNA Methyltransferase I to DNA Repair Sites. *Proc Natl Acad Sci U S A* **2005**, *102* (25), 8905–8909.

(73) Jin, B.; Robertson, K. D. DNA Methyltransferases, DNA Damage Repair, and Cancer. *Adv Exp Med Biol* **2013**, *754*, 3–29.

(74) Rougier, N.; Bourc'his, D.; Molina Gomes, D.; Niveleau, A.; Plachot, M.; Pàldi, A.; Viegas-Péquignot, E. Chromosome Methylation Patterns during Mammalian Preimplantation Development. *Genes Dev* **1998**, *12* (14), 2108–2113.

(75) Kohli, R. M.; Zhang, Y. TET Enzymes, TDG and the Dynamics of DNA Demethylation. *Nature* **2013**, *502* (7472), 479.

(76) He, Y. F.; Li, B. Z.; Li, Z.; Liu, P.; Wang, Y.; Tang, Q.; Ding, J.; Jia, Y.; Chen, Z.; Li, N.; Sun, Y.; Li, X.; Dai, Q.; Song, C. X.; Zhang, K.; He, C.; Xu, G. L. Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science (1979)* **2011**, *333* (6047), 1303–1307.

(77) Inoue, A.; Zhang, Y. Replication-Dependent Loss of 5-Hydroxymethylcytosine in Mouse Preimplantation Embryos. *Science (1979)* **2011**, *334* (6053), 194.

(78) Inoue, A.; Shen, L.; Dai, Q.; He, C.; Zhang, Y. Generation and Replication-Dependent Dilution of 5fC and 5caC during Mouse Preimplantation Development. *Cell Research 2011 21:12* **2011**, *21* (12), 1670–1676.

(79) Lv, H.; Dao, F. Y.; Zhang, D.; Yang, H.; Lin, H. Advances in Mapping the Epigenetic Modifications of 5-Methylcytosine (5mC), N6-Methyladenine (6mA), and N4-Methylcytosine (4mC). *Biotechnol Bioeng* **2021**, *118* (11), 4204–4216.

(80) Frommer, M.; McDonald, L. E.; Millar, D. S.; Collis, C. M.; Watt, F.; Grigg, G. W.; Molloy, P. L.; Paul, C. L. A Genomic Sequencing Protocol That Yields a Positive Display of 5-Methylcytosine Residues in Individual DNA Strands. *Proc Natl Acad Sci U S A* **1992**, *89* (5), 1827–1831.

(81) Li, Y.; Tollefsbol, T. O. DNA Methylation Detection: Bisulfite Genomic Sequencing Analysis. *Methods Mol Biol* **2011**, *791*, 11–21.

(82) Tanić, M. Epigenome-Wide Association Study (EWAS): Methods and Applications. *Epigenetics Methods* **2020**, 591–613.

(83) Liu, Y.; Siejka-Zielińska, P.; Velikova, G.; Bi, Y.; Yuan, F.; Tomkova, M.; Bai, C.; Chen, L.; Schuster-Böckler, B.; Song, C. X. Bisulfite-Free Direct Detection of 5-Methylcytosine and 5-Hydroxymethylcytosine at Base Resolution. *Nat Biotechnol* **2019**, *37* (4), 424–429.

(84) Tang, L. Bisulfite-Free Epigenetic Sequencing. *Nature Methods 2019 16:4* **2019**, *16* (4), 286–286.

(85) Vaisvila, R.; Ponnaluri, V. K. C.; Sun, Z.; Langhorst, B. W.; Saleh, L.; Guan, S.; Dai, N.; Campbell, M. A.; Sexton, B. S.; Marks, K.; Samaranayake, M.; Samuelson, J. C.; Church, H. E.; Tamanaha, E.; Corrêa, I. R.; Pradhan, S.; Dimalanta, E. T.; Evans, T. C.; Williams, L.; Davis, T. B. Enzymatic Methyl Sequencing Detects DNA Methylation at Single-Base Resolution from Picograms of DNA. *Genome Res* **2021**, *31* (7), 1280–1289.

(86) Flusberg, B. A.; Webster, D. R.; Lee, J. H.; Travers, K. J.; Olivares, E. C.; Clark, T. A.; Korlach, J.; Turner, S. W. Direct Detection of DNA Methylation during Single-Molecule, Real-Time Sequencing. *Nat Methods* **2010**, *7* (6), 461–465.

(87) Liu, Q.; Fang, L.; Yu, G.; Wang, D.; Xiao, C. le; Wang, K. Detection of DNA Base Modifications by Deep Recurrent Neural Network on Oxford Nanopore Sequencing Data. *Nat Commun* **2019**, *10* (1), 2449.

(88) Tse, O. Y. O.; Jiang, P.; Cheng, S. H.; Peng, W.; Shang, H.; Wong, J.; Chan, S. L.; Poon, L. C. Y.; Leung, T. Y.; Chan, K. C. A.; Chiu, R. W. K.; Lo, Y. M. D. Genome-Wide Detection of Cytosine Methylation by Single Molecule Real-Time Sequencing. *Proceedings of the National Academy of Sciences* **2021**, *118* (5), e2019768118.

(89) Kobayakawa, S.; Miike, K.; Nakao, M.; Abe, K. Dynamic Changes in the Epigenomic State and Nuclear Organization of Differentiating Mouse Embryonic Stem Cells. *Genes Cells* **2007**, *12* (4), 447–460.

(90) Yamagata, K. DNA Methylation Profiling Using Live-Cell Imaging. *Methods* **2010**, *52* (3), 259–266.

(91) Li, Y.; Miyanari, Y.; Shirane, K.; Nitta, H.; Kubota, T.; Ohashi, H.; Okamoto, A.; Sasaki, H. Sequence-Specific Microscopic Visualization of DNA Methylation Status at Satellite Repeats in Individual Cell Nuclei and Chromosomes. *Nucleic Acids Res* **2013**, *41* (19).

(92) Shiura, H.; Okamoto, A.; Sasaki, H.; Abe, K. Whole-Mount MeFISH: A Novel Technique for Simultaneous Visualization of Specific DNA Methylation and Protein/RNA Expression. *PLoS One* **2014**, *9* (4), e95750.

(93) Buchmuller, B.; Jung, A.; Muñoz-López, Á.; Summerer, D. Programmable Tools for Targeted Analysis of Epigenetic DNA Modifications. *Curr Opin Chem Biol* **2021**, *63*, 1–10.

(94) Lungu, C.; Pinter, S.; Broche, J.; Rathert, P.; Jeltsch, A. Modular Fluorescence Complementation Sensors for Live Cell Detection of Epigenetic Signals at Endogenous Genomic Sites. *Nature Communications 2017 8:1* **2017**, *8* (1), 1–13.

(95) Valton, J.; Dupuy, A.; Daboussi, F.; Thomas, S.; Maréchal, A.; Macmaster, R.; Melliand, K.; Juillerat, A.; Duchateau, P. Overcoming Transcription Activator-like Effector (TALE) DNA Binding Domain Sensitivity to Cytosine Methylation. *Journal of Biological Chemistry* **2012**, *287* (46), 38427–38432..

(96) Deng, D.; Yin, P.; Yan, C.; Pan, X.; Gong, X.; Qi, S.; Xie, T.; Mahfouz, M.; Zhu, J. K.; Yan, N.; Shi, Y. Recognition of Methylated DNA by TAL Effectors. *Cell Research 2012 22:10* **2012**, *22* (10), 1502–1504.

(97) Kubik, G.; Schmidt, M. J.; Penner, J. E.; Summerer, D. Programmable and Highly Resolved In Vitro Detection of 5-Methylcytosine by TALEs. *Angewandte Chemie International Edition* **2014**, *53* (23), 6002–6006.

(98) Rathi, P.; Maurer, S.; Kubik, G.; Summerer, D. Isolation of Human Genomic DNA Sequences with Expanded Nucleobase Selectivity. *J Am Chem Soc* **2016**, *138* (31), 9910–9918.

(99) Muñoz-López, Á.; Jung, A.; Buchmuller, B.; Wolffgramm, J.; Maurer, S.; Witte, A.; Summerer, D. Engineered TALE Repeats for Enhanced Imaging-Based Analysis of Cellular 5-Methylcytosine. *ChemBioChem* **2021**, *22* (4), 645–651.

(100) Muñoz-López, Á.; Buchmuller, B.; Wolffgramm, J.; Jung, A.; Hussong, M.; Kanne, J.; Schweiger, M. R.; Summerer, D.; Buchmuller, B.; Wolffgramm, J.; Jung, A.; Summerer, D.; Hussong, M.; Kanne, J.; Schweiger, M. R. Designer Receptors for Nucleotide-Resolution Analysis of Genomic 5-Methylcytosine by Cellular Imaging. *Angewandte Chemie International Edition* **2020**, *59* (23), 8927–8931.

(101) Boch, J.; Bonas, U. Xanthomonas AvrBs3 Family-Type III Effectors: Discovery and Function. *Annual review of phytopathology* **2010**, *48*, 419–436.

(102) White, F. F.; Potnis, N.; Jones, J. B.; Koebnik, R. The Type III Effectors of Xanthomonas. *Mol Plant Pathol* **2009**, *10* (6), 749–766.

(103) Römer, P.; Hahn, S.; Jordan, T.; Strauß, T.; Bonas, U.; Lahaye, T. Plant Pathogen Recognition Mediated by Promoter Activation of the Pepper Bs3 Resistance Gene. *Science (1979)* **2007**, *318* (5850), 645–648.

(104) Bogdanove, A. J.; Voytas, D. F. TAL Effectors: Customizable Proteins for DNA Targeting. *Science (1979)* **2011**, *333* (6051), 1843–1846.

(105) Boch, J.; Scholze, H.; Schornack, S.; Landgraf, A.; Hahn, S.; Kay, S.; Lahaye, T.; Nickstadt, A.; Bonas, U. Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science (1979)* **2009**, *326* (5959), 1509–1512.

(106) Sanjana, N. E.; Cong, L.; Zhou, Y.; Cunniff, M. M.; Feng, G.; Zhang, F. A Transcription Activator-Like Effector (TALE) Toolbox for Genome Engineering. *Nat Protoc* **2012**, *7* (1), 171–192.

(107) Moscou, M. J.; Bogdanove, A. J. A Simple Cipher Governs DNA Recognition by TAL Effectors. *Science (1979)* **2009**, *326* (5959), 1501.

(108) Doyle, E. L.; Stoddard, B. L.; Voytas, D. F.; Bogdanove, A. J. TAL Effectors: Highly Adaptable Phytobacterial Virulence Factors and Readily Engineered DNA Targeting Proteins. *Trends Cell Biol* **2013**, *23* (8), 390–398.

(109) Deng, D.; Yan, C.; Pan, X.; Mahfouz, M.; Wang, J.; Zhu, J. K.; Shi, Y.; Yan, N. Structural Basis for Sequence-Specific Recognition of DNA by TAL Effectors. *Science* **2012**, *335* (6069), 720–723.

(110) Streubel, J.; Blücher, C.; Landgraf, A.; Boch, J. TAL Effector RVD Specificities and Efficiencies. *Nat Biotechnol* **2012**, *30* (7), 593–595.

(111) Meckler, J. F.; Bhakta, M. S.; Kim, M. S.; Ovadia, R.; Habrian, C. H.; Zykovich, A.; Yu, A.; Lockwood, S. H.; Morbitzer, R.; Elsäesser, J.; Lahaye, T.; Segal, D. J.; Baldwin, E. P. Quantitative Analysis of TALE–DNA Interactions Suggests Polarity Effects. *Nucleic Acids Res* **2013**, *41* (7), 4118–4128.

(112) Wicky, B. I. M.; Stenta, M.; Dal Peraro, M. TAL Effectors Specificity Stems from Negative Discrimination. *PLoS One* **2013**, *8* (11), e80261.

(113) Mak, A. N. S.; Bradley, P.; Cernadas, R. A.; Bogdanove, A. J.; Stoddard, B. L. The Crystal Structure of TAL Effector PthXo1 Bound to Its DNA Target. *Science (1979)* **2012**, *335* (6069), 716–719.

(114) Mak, A. N. S.; Bradley, P.; Bogdanove, A. J.; Stoddard, B. L. TAL Effectors: Function, Structure, Engineering and Applications. *Curr Opin Struct Biol* **2013**, *23* (1), 93–99.

(115) Juillerat, A.; Beurdeley, M.; Valton, J.; Thomas, S.; Dubois, G.; Zaslavskiy, M.; Mikolajczak, J.; Bietz, F.; Silva, G. H.; Duclert, A.; Daboussi, F.; Duchateau, P. Exploring the Transcription Activator-like Effectors Scaffold Versatility to Expand the Toolbox of Designer Nucleases. *BMC Mol Biol* **2014**, *15* (1), 1–10.

(116) Gao, H.; Wu, X.; Chai, J.; Han, Z. Crystal Structure of a TALE Protein Reveals an Extended N-Terminal DNA Binding Region. *Cell Res* **2012**, *22* (12), 1716–1720.

(117) Cuculis, L.; Abil, Z.; Zhao, H.; Schroeder, C. M. Direct Observation of TALE Protein Dynamics Reveals a Two-State Search Mechanism. *Nature Communications 2015 6:1* **2015**, *6* (1), 1–11.

(118) Moore, R.; Chandrahas, A.; Bleris, L. Transcription Activator-like Effectors: A Toolkit for Synthetic Biology. *ACS Synth Biol* **2014**, *3* (10), 708–716.

(119) Cuculis, L.; Abil, Z.; Zhao, H.; Schroeder, C. M. TALE Proteins Search DNA Using a Rotationally Decoupled Mechanism. *Nature Chemical Biology 2016 12:10* **2016**, *12* (10), 831–837.

(120) Becker, S.; Boch, J. TALEs Spin along, but Not Around. *Nature Chemical Biology 2016 12:10* **2016**, *12* (10), 766–768.

(121) Engler, C.; Gruetzner, R.; Kandzia, R.; Marillonnet, S. Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. *PLoS One* **2009**, *4* (5), e5553.

(122) Engler, C.; Kandzia, R.; Marillonnet, S. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLoS One* **2008**, *3* (11), e3647.

(123) Cermak, T.; Doyle, E. L.; Christian, M.; Wang, L.; Zhang, Y.; Schmidt, C.; Baller, J. A.; Somia, N. v.; Bogdanove, A. J.; Voytas, D. F. Efficient Design and Assembly of Custom TALEN and Other TAL Effector-Based Constructs for DNA Targeting. *Nucleic Acids Res* **2011**, *39* (12).

(124) Geißler, R.; Scholze, H.; Hahn, S.; Streubel, J.; Bonas, U.; Behrens, S. E.; Boch, J. Transcriptional Activators of Human Genes with Programmable DNA-Specificity. *PLoS One* **2011**, *6* (5), e19509.

(125) Zhang, F.; Cong, L.; Lodato, S.; Kosuri, S.; Church, G. M.; Arlotta, P. Programmable Sequence-Specific Transcriptional Regulation of Mammalian Genome Using Designer TAL Effectors. *Nat Biotechnol* **2011**, *29* (2), 149–153.

(126) Cong, L.; Zhou, R.; Kuo, Y. C.; Cunniff, M.; Zhang, F. Comprehensive Interrogation of Natural TALE DNA Binding Modules and Transcriptional Repressor Domains. *Nat Commun* **2012**, *3*, 968.

(127) Christian, M.; Cermak, T.; Doyle, E. L.; Schmidt, C.; Zhang, F.; Hummel, A.; Bogdanove, A. J.; Voytas, D. F. Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. *Genetics* **2010**, *186* (2), 756–761.

(128) Li, T.; Huang, S.; Jiang, W. Z.; Wright, D.; Spalding, M. H.; Weeks, D. P.; Yang, B. TAL Nucleases (TALNs): Hybrid Proteins Composed of TAL Effectors and FokI DNA-Cleavage Domain. *Nucleic Acids Res* **2011**, *39* (1), 359–372.

(129) Gabsalilow, L.; Schierling, B.; Friedhoff, P.; Pingoud, A.; Wende, W. Site- and Strand-Specific Nicking of DNA by Fusion Proteins Derived from MutH and I-SceI or TALE Repeats. *Nucleic Acids Res* **2013**, *41* (7), e83.

(130) Luo, Y.; Wang, Y.; Liu, J.; Cui, C.; Wu, Y.; Lan, H.; Chen, Q.; Liu, X.; Quan, F.; Guo, Z.; Zhang, Y. Generation of TALE Nickase-Mediated Gene-Targeted Cows Expressing Human Serum Albumin in Mammary Glands. *Sci Rep* **2016**, *6*, e20657.

(131) Owens, J. B.; Mauro, D.; Stoytchev, I.; Bhakta, M. S.; Kim, M. S.; Segal, D. J.; Moisyadi, S. Transcription Activator like Effector (TALE)-Directed PiggyBac Transposition in Human Cells. *Nucleic Acids Res* **2013**, *41* (19), 9197–9207.

(132) Mercer, A. C.; Gaj, T.; Fuller, R. P.; Barbas, C. F. Chimeric TALE Recombinases with Programmable DNA Sequence Specificity. *Nucleic Acids Res* **2012**, *40* (21), e11163.

(133) Voziyanova, E.; Li, F.; Shah, R.; Voziyanov, Y. Genome Targeting by Hybrid Flp-TAL Recombinases. *Sci Rep* **2020**, *10* (1), 17479. https://doi.org/10.1038/S41598-020-74474-2.

(134) Becker, S.; Boch, J. TALE and TALEN Genome Editing Technologies. *Gene and Genome Editing* **2021**, *2*, 100007.

(135) Mok, B. Y.; de Moraes, M. H.; Zeng, J.; Bosch, D. E.; Kotrys, A. v.; Raguram, A.; Hsu, F. S.; Radey, M. C.; Peterson, S. B.; Mootha, V. K.; Mougous, J. D.; Liu, D. R. A Bacterial Cytidine Deaminase Toxin Enables CRISPR-Free Mitochondrial Base Editing. *Nature 2020 583:7817* **2020**, *583* (7817), 631–637.

(136) Rivenbark, A. G.; Stolzenburg, S.; Beltran, A. S.; Yuan, X.; Rots, M. G.; Strahl, B. D.; Blancafort, P. Epigenetic Reprogramming of Cancer Cells via Targeted DNA Methylation. *Epigenetics* **2012**, *7* (4), 350–360.

(137) Siddique, A. N.; Nunna, S.; Rajavelu, A.; Zhang, Y.; Jurkowska, R. Z.; Reinhardt, R.; Rots, M. G.; Ragozin, S.; Jurkowski, T. P.; Jeltsch, A. Targeted Methylation and Gene Silencing of VEGF-A in Human Cells by Using a Designed Dnmt3a–Dnmt3L Single-Chain Fusion Protein with Increased DNA Methylation Activity. *J Mol Biol* **2013**, *425* (3), 479–491.

(138) Bernstein, D. L.; le Lay, J. E.; Ruano, E. G.; Kaestner, K. H. TALE-Mediated Epigenetic Suppression of CDKN2A Increases Replication in Human Fibroblasts. *J Clin Invest* **2015**, *125* (5), 1998–2006.

(139) Maeder, M. L.; Angstman, J. F.; Richardson, M. E.; Linder, S. J.; Cascio, V. M.; Tsai, S. Q.; Ho, Q. H.; Sander, J. D.; Reyon, D.; Bernstein, B. E.; Costello, J. F.; Wilkinson, M. F.; Joung, J. K. Targeted DNA Demethylation and Activation of Endogenous Genes Using Programmable TALE-TET1 Fusion Proteins. *Nature Biotechnology 2013 31:12* **2013**, *31* (12), 1137–1142.

(140) Li, K.; Pang, J.; Cheng, H.; Liu, W. P.; Di, J. M.; Xiao, H. J.; Luo, Y.; Zhang, H.; Huang, W. T.; Chen, M. K.; Li, L. Y.; Shao, C. K.; Feng, Y. H.; Gao, X. Manipulation of Prostate Cancer Metastasis by Locus-Specific Modification of the CRMP4 Promoter Region Using Chimeric TALE DNA Methyltransferase and Demethylase. *Oncotarget* **2015**, *6* (12), 10030.

(141) Kubik, G.; Summerer, D. TALEored Epigenetics: A DNA-Binding Scaffold for Programmable Epigenome Editing and Analysis. *ChemBioChem* **2016**, *17* (11), 975–980.

(142) Konermann, S.; Brigham, M. D.; Trevino, A. E.; Hsu, P. D.; Heidenreich, M.; Cong, L.; Platt, R. J.; Scott, D. A.; Church, G. M.; Zhang, F. Optical Control of Mammalian Endogenous Transcription and Epigenetic States. *Nature 2013 500:7463* **2013**, *500* (7463), 472–476.

(143) Mendenhall, E. M.; Williamson, K. E.; Reyon, D.; Zou, J. Y.; Ram, O.; Joung, J. K.; Bernstein, B. E. Locus-Specific Editing of Histone Modifications at Endogenous Enhancers. *Nature Biotechnology 2013 31:12* **2013**, *31* (12), 1133–1136.

(144) Cho, H. S.; Kang, J. G.; Lee, J. H.; Lee, J. J.; Jeon, S. K.; Ko, J. H.; Kim, D. S.; Park, K. H.; Kim, Y. S.; Kim, N. S. Direct Regulation of E-Cadherin by Targeted Histone Methylation of TALE-SET Fusion Protein in Cancer Cells. *Oncotarget* **2015**, *6* (27), 23837–23844.

(145) Hilton, I. B.; D'Ippolito, A. M.; Vockley, C. M.; Thakore, P. I.; Crawford, G. E.; Reddy, T. E.; Gersbach, C. A. Epigenome Editing by a CRISPR-Cas9-Based Acetyltransferase Activates Genes from Promoters and Enhancers. *Nature Biotechnology* **2015**, *33* (5), 510–517.

(146) Miyanari, Y.; Ziegler-Birling, C.; Torres-Padilla, M. E. Live Visualization of Chromatin Dynamics with Fluorescent TALEs. *Nature Structural & Molecular Biology 2013 20:11* **2013**, *20* (11), 1321–1324.

(147) Ma, H.; Reyes-Gutierrez, P.; Pederson, T. Visualization of Repetitive DNA Sequences in Human Chromosomes with Transcription Activator-like Effectors. *Proc Natl Acad Sci U S A* **2013**, *110* (52), 21048–21053.

(148) Thanisch, K.; Schneider, K.; Morbitzer, R.; Solovei, I.; Lahaye, T.; Bultmann, S.; Leonhardt, H. Targeting and Tracing of Specific DNA Sequences with DTALEs in Living Cells. *Nucleic Acids Res* **2014**, *42* (6), e38.

(149) Yuan, K.; Shermoen, A. W.; O'Farrell, P. H. Illuminating DNA Replication during Drosophila Development Using TALE-Lights. *Curr Biol* **2014**, *24* (4), R144.

(150) Bultmann, S.; Morbitzer, R.; Schmidt, C. S.; Thanisch, K.; Spada, F.; Elsaesser, J.; Lahaye, T.; Leonhardt, H. Targeted Transcriptional Activation of Silent Oct4 Pluripotency Gene by Combining Designer TALEs and Inhibition of Epigenetic Modifiers. *Nucleic Acids Res* **2012**, *40* (12), 5368–5377.

(151) Palii, S. S.; van Emburgh, B. O.; Sankpal, U. T.; Brown, K. D.; Robertson, K. D. DNA Methylation Inhibitor 5-Aza-2'-Deoxycytidine Induces Reversible Genome-Wide DNA Damage That Is Distinctly Influenced by DNA Methyltransferases 1 and 3B. *Mol Cell Biol* **2008**, *28* (2), 752–771.

(152) Kubik, G.; Batke, S.; Summerer, D. Programmable Sensors of 5-Hydroxymethylcytosine. *J Am Chem Soc* **2015**, *137* (1), 2–5.

(153) Kubik, G.; Summerer, D. Achieving Single-Nucleotide Resolution of 5-Methylcytosine Detection with TALEs. *ChemBioChem* **2015**, *16* (2), 228–231.

(154) Rathi, P.; Maurer, S.; Kubik, G.; Summerer, D. Isolation of Human Genomic DNA Sequences with Expanded Nucleobase Selectivity. *J Am Chem Soc* **2016**, *138* (31), 9910–9918.

(155) Maurer, S.; Giess, M.; Koch, O.; Summerer, D. Interrogating Key Positions of Size-Reduced TALE Repeats Reveals a Programmable Sensor of 5-Carboxylcytosine. *ACS Chem Biol* **2016**, *11* (12), 3294–3299.

(156) Maurer, S.; Buchmuller, B.; Ehrt, C.; Jasper, J.; Koch, O.; Summerer, D. Overcoming Conservation in TALE–DNA Interactions: A Minimal Repeat Scaffold Enables Selective Recognition of an Oxidized 5-Methylcytosine. *Chem Sci* **2018**, *9* (36), 7247–7252.

(157) Tsuji, S.; Futaki, S.; Imanishi, M. Sequence-Specific Recognition of Methylated DNA by an Engineered Transcription Activator-like Effector Protein. *Chemical Communications* **2016**, *52* (99), 14238–14241.

(158) Zhang, Y.; Liu, L.; Guo, S.; Song, J.; Zhu, C.; Yue, Z.; Wei, W.; Yi, C. Deciphering TAL Effectors for 5-Methylcytosine and 5-Hydroxymethylcytosine Recognition. *Nature Communications 2017 8:1* **2017**, *8* (1), 1–9.

(159) Liu, L.; Zhang, Y.; Liu, M.; Wei, W.; Yi, C.; Peng, J. Structural Insights into the Specific Recognition of 5-Methylcytosine and 5-Hydroxymethylcytosine by TAL Effectors. *J Mol Biol* **2020**, *432* (4), 1035–1047.

(160) Rathi, P.; Witte, A.; Summerer, D. Engineering DNA Backbone Interactions Results in TALE Scaffolds with Enhanced 5-Methylcytosine Selectivity. *Scientific Reports 2017 7:1* **2017**, *7* (1), 1–13.

(161) Sato, H.; Das, S.; Singer, R. H.; Singer, R. H.; Vera, M.; Vera, M. Imaging of DNA and RNA in Living Eukaryotic Cells to Reveal Spatio-Temporal Dynamics of Gene Expression. *Annu Rev Biochem* **2020**, *89*, 159–187.

(162) Schwarzacher, T.; Heslop-Harrison, J. S. Direct Fluorochrome-Labeled DNA Probes for Direct Fluorescent in Situ Hybridization to Chromosomes. *Methods Mol Biol* **1994**, *28*, 167–176.

(163) Nath, J.; Johnson, K. L. A Review of Fluorescence in Situ Hybridization (FISH): Current Status and Future Prospects. *Biotech Histochem* **2000**, *75* (2), 54–78.

(164) Fujita, N.; Takebayashi, S.; Okumura, K.; Kudo, S.; Chiba, T.; Saya, H.; Nakao, M. Methylation-Mediated Transcriptional Silencing in Euchromatin by Methyl-CpG Binding Protein MBD1 Isoforms. *Mol Cell Biol* **1999**, *19* (9), 6415–6426.

(165) Koo, D. H.; Han, F.; Birchler, J. A.; Jiang, J. Distinct DNA Methylation Patterns Associated with Active and Inactive Centromeres of the Maize B Chromosome. *Genome Res* **2011**, *21* (6), 908–914.

(166) He, J.; Mo, D.; Chen, J.; Luo, L. Combined Whole-Mount Fluorescence in Situ Hybridization and Antibody Staining in Zebrafish Embryos and Larvae. *Nat Protoc* **2020**, *15* (10), 3361–3379.

(167) Wu, X.; Mao, S.; Ying, Y.; Krueger, C. J.; Chen, A. K. Progress and Challenges for Live-Cell Imaging of Genomic Loci Using CRISPR-Based Platforms. *Genomics Proteomics Bioinformatics* **2019**, *17* (2), 119.

(168) Bronshtein, I.; Kepten, E.; Kanter, I.; Berezin, S.; Lindner, M.; Redwood, A. B.; Mai, S.; Gonzalo, S.; Foisner, R.; Shav-Tal, Y.; Garini, Y. Loss of Lamin A Function Increases Chromatin Dynamics in the Nuclear Interior. *Nature Communications 2015 6:1* **2015**, *6* (1), 1–9.

(169) Mattern, K. A.; Swiggers, S. J. J.; Nigg, A. L.; Löwenberg, B.; Houtsmuller, A. B.; Zijlmans, J. M. J. M. Dynamics of Protein Binding to Telomeres in Living Cells: Implications for Telomere Structure and Function. *Mol Cell Biol* **2004**, *24* (12), 5587–5594.

(170) Gasser, S. M. Nuclear Architecture: Visualizing Chromatin Dynamics in Interphase Nuclei. *Science (1979)* **2002**, *296* (5572), 1412–1416.

(171) Chen, B.; Guan, J.; Huang, B. Imaging Specific Genomic DNA in Living Cells. *Annu Rev Biophys* **2016**, *45*, 1.

(172) Robinett, C. C.; Straight, A.; Li, G.; Willhelm, C.; Sudlow, G.; Murray, A.; Belmont, A. S. In Vivo Localization of DNA Sequences and Visualization of Large-Scale Chromatin Organization Using Lac Operator/Repressor Recognition. *J Cell Biol* **1996**, *135* (6), 1685–1700.

(173) Marshall, W. F.; Straight, A.; Marko, J. F.; Swedlow, J.; Dernburg, A.; Belmont, A.; Murray, A. W.; Agard, D. A.; Sedat, J. W. Interphase Chromosomes Undergo Constrained Diffusional Motion in Living Cells. *Current Biology* **1997**, *7* (12), 930–939.

(174) Michaelis, C.; Ciosk, R.; Nasmyth, K. Cohesins: Chromosomal Proteins That Prevent Premature Separation of Sister Chromatids. *Cell* **1997**, *91* (1), 35–45.

(175) Shaban, H. A.; Seeber, A. Monitoring the Spatio-Temporal Organization and Dynamics of the Genome. *Nucleic Acids Res* **2020**, *48* (7), 3423–3434.

(176) Tupler, R.; Perini, G.; Green, M. R. Expressing the Human Genome. *Nature 2001 409:6822* **2001**, *409* (6822), 832–833.

(177) Casas-Delucchi, C. S.; Becker, A.; Bolius, J. J.; Cristina Cardoso, M. Targeted Manipulation of Heterochromatin Rescues MeCP2 Rett Mutants and Re-Establishes Higher Order Chromatin Organization. *Nucleic Acids Res* **2012**, *40* (22), e176.

(178) Lindhout, B. I.; Fransz, P.; Tessadori, F.; Meckel, T.; Hooykaas, P. J. J.; van der Zaal, B. J. Live Cell Imaging of Repetitive DNA Sequences via GFP-Tagged Polydactyl Zinc Finger Proteins. *Nucleic Acids Res* **2007**, *35* (16), e107.

(179) Cassandri, M.; Smirnov, A.; Novelli, F.; Pitolli, C.; Agostini, M.; Malewicz, M.; Melino, G.; Raschellà, G. Zinc-Finger Proteins in Health and Disease. *Cell Death Discovery 2017 3:1* **2017**, *3* (1), 1–12.

(180) Klug, A. Co-Chairman's Remarks: Protein Designs for the Specific Recognition of DNA. *Gene* **1993**, *135* (1–2), 83–92.

(181) Pavletich, N. P.; Pabo, C. O. Zinc Finger-DNA Recognition: Crystal Structure of a Zif268-DNA Complex at 2.1 Å. *Science (1979)* **1991**, *252* (5007), 809–817.

(182) Laity, J. H.; Lee, B. M.; Wright, P. E. Zinc Finger Proteins: New Insights into Structural and Functional Diversity. *Curr Opin Struct Biol* **2001**, *11* (1), 39–46.

(183) Pabo, C. O.; Peisach, E.; Grant, R. A. Design and Selection of Novel Cys2His2 Zinc Finger Proteins. *Annu Rev Biochem* **2001**, *70*, 313–340.

(184) Dreier, B.; Fuller, R. P.; Segal, D. J.; Lund, C. v.; Blancafort, P.; Huber, A.; Koksch, B.; Barbas, C. F. Development of Zinc Finger Domains for Recognition of the 5′-CNN-3′ Family DNA Sequences and Their Use in the Construction of Artificial Transcription Factors. *Journal of Biological Chemistry* **2005**, *280* (42), 35588–35597.

(185) Bae, K. H.; Kwon, Y. do; Shin, H. C.; Hwang, M. S.; Ryu, E. H.; Park, K. S.; Yang, H. Y.; Lee, D. ki; Lee, Y.; Park, J.; Kwon, H. S.; Kim, H. W.; Yeh, B. il; Lee, H. W.; Sohn, S. H.; Yoon, J.; Seol, W.; Kim, J. S. Human Zinc Fingers as Building Blocks in the Construction of Artificial Transcription Factors. *Nature Biotechnology 2003 21:3* **2003**, *21* (3), 275–280.

(186) Mueller, A. L.; Corbi-Verge, C.; Giganti, D. O.; Ichikawa, D. M.; Spencer, J. M.; MacRae, M.; Garton, M.; Kim, P. M.; Noyes, M. B. The Geometric Influence on the Cys2His2 Zinc Finger Domain and Functional Plasticity. *Nucleic Acids Res* **2020**, *48* (11), 6382–6402.

(187) Hu, H.; Zhang, H.; Wang, S.; Ding, M.; An, H.; Hou, Y.; Yang, X.; Wei, W.; Sun, Y.; Tang, C. Live Visualization of Genomic Loci with BiFC-TALE. *Scientific Reports 2017 7:1* **2017**, *7* (1), 1–8.

(188) Hu, H.; Yang, X.; Tang, C. Visualization of Genomic Loci in Living Cells with BiFC-TALE. *Curr Protoc Cell Biol* **2019**, *82* (1), e78.

(189) Schuller, K. G.; Mitra, J.; Ha, T.; Barrick, D. Functional Instability Allows Access to DNA in Longer Transcription Activator-Like Effector (TALE) Arrays. *Elife* **2019**, *8*, e38298.

(190) Ma, Y.; Wang, M.; Li, W.; Zhang, Z.; Zhang, X.; Tan, T.; Zhang, X. E.; Cui, Z. Live Cell Imaging of Single Genomic Loci with Quantum Dot-Labeled TALEs. *Nature Communications 2017 8:1* **2017**, *8* (1), 1–8.

(191) Cong, L.; Ran, F. A.; Cox, D.; Lin, S.; Barretto, R.; Habib, N.; Hsu, P. D.; Wu, X.; Jiang, W.; Marraffini, L. A.; Zhang, F. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **2013**, *339* (6121), 819.

(192) Jinek, M.; East, A.; Cheng, A.; Lin, S.; Ma, E.; Doudna, J. RNA-Programmed Genome Editing in Human Cells. *Elife* **2013**, *2* (2), 471.

(193) Mali, P.; Esvelt, K. M.; Church, G. M. Cas9 as a Versatile Tool for Engineering Biology. *Nat Methods* **2013**, *10* (10), 957–963.

(194) Gilbert, L. A.; Larson, M. H.; Morsut, L.; Liu, Z.; Brar, G. A.; Torres, S. E.; Stern-Ginossar, N.; Brandman, O.; Whitehead, E. H.; Doudna, J. A.; Lim, W. A.; Weissman, J. S.; Qi, L. S. CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell* **2013**, *154* (2), 442–451.

(195) Chen, B.; Gilbert, L. A.; Cimini, B. A.; Schnitzbauer, J.; Zhang, W.; Li, G. W.; Park, J.; Blackburn, E. H.; Weissman, J. S.; Qi, L. S.; Huang, B. Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. *Cell* **2013**, *155* (7), 1479–1491.

(196) Adli, M. The CRISPR Tool Kit for Genome Editing and Beyond. *Nature Communications 2018 9:1* **2018**, *9* (1), 1–13.

(197) Qi, L. S.; Larson, M. H.; Gilbert, L. A.; Doudna, J. A.; Weissman, J. S.; Arkin, A. P.; Lim, W. A. Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **2013**, *152* (5), 1173–1183.

(198) Chen, B.; Hu, J.; Almeida, R.; Liu, H.; Balakrishnan, S.; Covill-Cooke, C.; Lim, W. A.; Huang, B. Expanding the CRISPR Imaging Toolset with Staphylococcus Aureus Cas9 for Simultaneous Imaging of Multiple Genomic Loci. *Nucleic Acids Res* **2016**, *44* (8), e75.

(199) Gu, B.; Swigut, T.; Spencley, A.; Bauer, M. R.; Chung, M.; Meyer, T.; Wysocka, J. Transcription-Coupled Changes in Nuclear Mobility of Mammalian Cis-Regulatory Elements. *Science* **2018**, *359* (6379), 1050–1055.

(200) Qin, P.; Parlak, M.; Kuscu, C.; Bandaria, J.; Mir, M.; Szlachta, K.; Singh, R.; Darzacq, X.; Yildiz, A.; Adli, M. Live Cell Imaging of Low- and Non-Repetitive Chromosome Loci Using CRISPR-Cas9. *Nature Communications 2017 8:1* **2017**, *8* (1), 1–10.

(201) Shao, S.; Xue, B.; Sun, Y. Intranucleus Single-Molecule Imaging in Living Cells. *Biophys J* **2018**, *115* (2), 181–189.

(202) Viswanathan, S.; Williams, M. E.; Bloss, E. B.; Stasevich, T. J.; Speer, C. M.; Nern, A.; Pfeiffer, B. D.; Hooks, B. M.; Li, W. P.; English, B. P.; Tian, T.; Henry, G. L.; Macklin, J. J.; Patel, R.;

Gerfen, C. R.; Zhuang, X.; Wang, Y.; Rubin, G. M.; Looger, L. L. High-Performance Probes for Light and Electron Microscopy. *Nat Methods* **2015**, *12* (6), 568–576.

(203) Tanenbaum, M. E.; Gilbert, L. A.; Qi, L. S.; Weissman, J. S.; Vale, R. D. A Protein Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging. *Cell* **2014**, *159* (3), 635.

(204) Wörn, A.; der Maur, A. A.; Escher, D.; Honegger, A.; Barberis, A.; Plückthun, A. Correlation between in Vitro Stability and in Vivo Performance of Anti-GCN4 Intrabodies as Cytoplasmic Inhibitors. *J Biol Chem* **2000**, *275* (4), 2795–2803.

(205) Colby, D. W.; Garg, P.; Holden, T.; Chao, G.; Webster, J. M.; Messer, A.; Ingram, V. M.; Wittrup, K. D. Development of a Human Light Chain Variable Domain (V(L)) Intracellular Antibody Specific for the Amino Terminus of Huntingtin via Yeast Surface Display. *J Mol Biol* **2004**, *342* (3), 901–912.

(206) Lecerf, J. M.; Shirley, T. L.; Zhu, Q.; Kazantsev, A.; Amersdorfer, P.; Housman, D. E.; Messer, A.; Huston, J. S. Human Single-Chain Fv Intrabodies Counteract in Situ Huntingtin Aggregation in Cellular Models of Huntington's Disease. *Proc Natl Acad Sci U S A* **2001**, *98* (8), 4764–4769.

(207) Ye, H.; Rong, Z.; Lin, Y. Live Cell Imaging of Genomic Loci Using DCas9-SunTag System and a Bright Fluorescent Protein. *Protein Cell* **2017**, *8* (11), 853–855.

(208) Shao, S.; Chang, L.; Sun, Y.; Hou, Y.; Fan, X.; Sun, Y. Multiplexed SgRNA Expression Allows Versatile Single Nonrepetitive DNA Labeling and Endogenous Gene Regulation. *ACS Synth Biol* **2018**, *7* (1), 176–186.

(209) Kamiyama, D.; Sekine, S.; Barsi-Rhyne, B.; Hu, J.; Chen, B.; Gilbert, L. A.; Ishikawa, H.; Leonetti, M. D.; Marshall, W. F.; Weissman, J. S.; Huang, B. Versatile Protein Tagging in Cells with Split Fluorescent Protein. *Nature Communications 2016 7:1* **2016**, *7* (1), 1–9.

(210) Cabantous, S.; Terwilliger, T. C.; Waldo, G. S. Protein Tagging and Detection with Engineered Self-Assembling Fragments of Green Fluorescent Protein. *Nat Biotechnol* **2005**, *23* (1), 102–107.

(211) Chen, B.; Zou, W.; Xu, H.; Liang, Y.; Huang, B. Efficient Labeling and Imaging of Protein-Coding Genes in Living Cells Using CRISPR-Tag. *Nature Communications 2018 9:1* **2018**, *9* (1), 1–9.

(212) Fu, Y.; Rocha, P. P.; Luo, V. M.; Raviram, R.; Deng, Y.; Mazzoni, E. O.; Skok, J. A. CRISPR-DCas9 and SgRNA Scaffolds Enable Dual-Colour Live Imaging of Satellite Sequences and Repeat-Enriched Individual Loci. *Nat Commun* **2016**, *7*.

(213) Wang, S.; Su, J. H.; Zhang, F.; Zhuang, X. An RNA-Aptamer-Based Two-Color CRISPR Labeling System. *Scientific Reports 2016 6:1* **2016**, *6* (1), 1–7.

(214) Ma, H.; Tu, L. C.; Naseri, A.; Chung, Y. C.; Grunwald, D.; Zhang, S.; Pederson, T. CRISPR-Sirius: RNA Scaffolds for Signal Amplification in Genome Imaging. *Nat Methods* **2018**, *15* (11), 928–931.

(215) Cheng, A. W.; Jillette, N.; Lee, P.; Plaskon, D.; Fujiwara, Y.; Wang, W.; Taghbalout, A.; Wang, H. Casilio: A Versatile CRISPR-Cas9-Pumilio Hybrid for Gene Regulation and Genomic Labeling. *Cell Research 2016 26:2* **2016**, *26* (2), 254–257.

(216) Clow, P. A.; Du, M.; Jillette, N.; Taghbalout, A.; Zhu, J. J.; Cheng, A. W. CRISPR-Mediated Multiplexed Live Cell Imaging of Nonrepetitive Genomic Loci with One Guide RNA per Locus. *Nature Communications 2022 13:1* **2022**, *13* (1), 1–10.

(217) Yue, S.; Li, Y.; Qiao, Z.; Song, W.; Bi, S. Rolling Circle Replication for Biosensing, Bioimaging, and Biomedicine. *Trends Biotechnol* **2021**, *39* (11), 1160–1172.

(218) Ali, M. M.; Li, F.; Zhang, Z.; Zhang, K.; Kang, D. K.; Ankrum, J. A.; Le, X. C.; Zhao, W. Rolling Circle Amplification: A Versatile Tool for Chemical Biology, Materials Science and Medicine. *Chem Soc Rev* **2014**, *43* (10), 3324–3341.

(219) Smolina, I. v.; Demidov, V. v.; Cantor, C. R.; Broude, N. E. Real-Time Monitoring of Branched Rolling-Circle DNA Amplification with Peptide Nucleic Acid Beacon. *Anal Biochem* **2004**, *335* (2), 326–329.

(220) Mao, S.; Ying, Y.; Wu, R.; Chen, A. K. Recent Advances in the Molecular Beacon Technology for Live-Cell Single-Molecule Imaging. *iScience* **2020**, *23* (12), 101801.

(221) Wu, X.; Mao, S.; Yang, Y.; Rushdi, M. N.; Krueger, C. J.; Chen, A. K. A CRISPR/Molecular Beacon Hybrid System for Live-Cell Genomic Imaging. *Nucleic Acids Res* **2018**, *46* (13), E80.

(222) Mao, S.; Ying, Y.; Wu, X.; Krueger, C. J.; Chen, A. K. CRISPR/Dual-FRET Molecular Beacon for Sensitive Live-Cell Imaging of Non-Repetitive Genomic Loci. *Nucleic Acids Res* **2019**, *47* (20), e131.

(223) Cho, Y.; Seo, J.; Sim, Y.; Chung, J.; Park, C. E.; Park, C. G.; Kim, D.; Chang, J. B. FRACTAL: Signal Amplification of Immunofluorescence via Cyclic Staining of Target Molecules. *Nanoscale* **2020**, *12* (46), 23506–23513.

(224) Yeon, H.; Cho, Y.; Seo, J.; Sim, Y.; Chang, J. B. Simultaneous Amplification of Multiple Immunofluorescence Signals via Cyclic Staining of Target Molecules Using Mutually Cross-Adsorbed Antibodies. *Scientific Reports 2022 12:1* **2022**, *12* (1), 1–10.

(225) Faget, L.; Hnasko, T. S. Tyramide Signal Amplification for Immunofluorescent Enhancement. *Methods Mol Biol* **2015**, *1318*, 161–172.

(226) Yang, Z.; Luo, S.; Dai, H.; Li, J.; Jiao, X.; Hu, X. A Biotin–Streptavidin Signal Amplification Strategy for a Highly Sensitive Chemiluminescent Immunoassay for Chicken Interferon-γ. *RSC Adv* **2013**, *3* (45), 22868–22871.

(227) Athmane, N.; Williamson, I.; Boyle, S.; Biddie, S. C.; Bickmore, W. A. MUC4 Is Not Expressed in Cell Lines Used for Live Cell Imaging. *Wellcome Open Res* **2021**, *6*, 265.

(228) Chaturvedi, P.; Singh, A. P.; Batra, S. K. Structure, Evolution, and Biology of the MUC4 Mucin. *The FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **2008**, *22* (4), 966.

(229) Carraway, K. L.; Theodoropoulos, G.; Kozloski, G. A.; Carothers Carraway, C. A. Muc4/MUC4 Functions and Regulation in Cancer. *Future Oncol* **2009**, *5* (10), 1631.

(230) Chakraborty, S.; Jain, M.; Sasson, A. R.; Batra, S. K. MUC4 as a Diagnostic Marker in Cancer. *Expert Opin Med Diagn* **2008**, *2* (8), 891–910.

(231) Yokoyama, S.; Higashi, M.; Kitamoto, S.; Oeldorf, M.; Knippschild, U.; Kornmann, M.; Maemura, K.; Kurahara, H.; Wiest, E.; Hamada, T.; Kitazono, I.; Goto, Y.; Tasaki, T.; Hiraki, T.; Hatanaka, K.; Mataki, Y.; Taguchi, H.; Hashimoto, S.; Batra, S. K.; Tanimoto, A.; Yonezawa, S.; Hollingsworth, M. A. Aberrant Methylation of MUC1 and MUC4 Promoters Are Potential Prognostic Biomarkers for Pancreatic Ductal Adenocarcinomas. *Oncotarget* **2016**, *7* (27), 42553.

(232) Ren, R.; Deng, L.; Xue, Y.; Suzuki, K.; Zhang, W.; Yu, Y.; Wu, J.; Sun, L.; Gong, X.; Luan, H.; Yang, F.; Ju, Z.; Ren, X.; Wang, S.; Tang, H.; Geng, L.; Zhang, W.; Li, J.; Qiao, J.; Xu, T.; Qu, J.; Liu, G. H. Visualization of Aging-Associated Chromatin Alterations with an Engineered TALE System. *Cell Research 2017 27:4* **2017**, *27* (4), 483–504.

(233) Allis, C. D.; Jenuwein, T. The Molecular Hallmarks of Epigenetic Control. *Nat Rev Genet* **2016**, *17* (8), 487–500.

(234) Kumar, S.; Chinnusamy, V.; Mohapatra, T. Epigenetics of Modified DNA Bases: 5-Methylcytosine and Beyond. *Front Genet* **2018**, *9*, 640.

(235) Heyn, H.; Esteller, M. DNA Methylation Profiling in the Clinic: Applications and Challenges. *Nat Rev Genet* **2012**, *13* (10), 679–692.

(236) Tanaka, K.; Tainaka, K.; Umemoto, T.; Nomura, A.; Okamoto, A. An Osmium-DNA Interstrand Complex: Application to Facile DNA Methylation Analysis. *J Am Chem Soc* **2007**, *129* (46), 14511–14517.

(237) Kanne, J.; Hussong, M.; Isensee, J.; Muñoz-López, Á.; Wolffgramm, J.; Heß, F.; Grimm, C.; Bessonov, S.; Meder, L.; Wang, J.; Reinhardt, H. C.; Odenthal, M.; Hucho, T.; Büttner, R.; Summerer, D.; Schweiger, M. R. Pericentromeric Satellite III Transcripts Induce Etoposide Resistance. *Cell Death & Disease* **2021**, *12* (6), 1–15.

(238) Shu, X.; Shaner, N. C.; Yarbrough, C. A.; Tsien, R. Y.; Remington, S. J. Novel Chromophores and Buried Charges Control Color in MFruits. *Biochemistry* **2006**, *45* (32).

(239) Ueda, M.; Manabe, Y.; Mukai, M. The High Performance of 3XFLAG for Target Purification of a Bioactive Metabolite: A Tag Combined with a Highly Effective Linker Structure. *Bioorg Med Chem Lett* **2011**, *21* (5), 1359–1362.

(240) Nurk, S.; Koren, S.; Rhie, A.; Rautiainen, M.; Bzikadze, A. v.; Mikheenko, A.; Vollger, M. R.; Altemose, N.; Uralsky, L.; Gershman, A.; Aganezov, S.; Hoyt, S. J.; Diekhans, M.; Logsdon, G. A.; Alonge, M.; Antonarakis, S. E.; Borchers, M.; Bouffard, G. G.; Brooks, S. Y.; Caldas, G. v.; Chen, N. C.; Cheng, H.; Chin, C. S.; Chow, W.; de Lima, L. G.; Dishuck, P. C.; Durbin, R.; Dvorkina, T.; Fiddes, I. T.; Formenti, G.; Fulton, R. S.; Fungtammasan, A.; Garrison, E.; Grady, P. G. S.; Graves-Lindsay, T. A.; Hall, I. M.; Hansen, N. F.; Hartley, G. A.; Haukness, M.; Howe, K.; Hunkapiller, M. W.; Jain, C.; Jain, M.; Jarvis, E. D.; Kerpedjiev, P.; Kirsche, M.; Kolmogorov, M.; Korlach, J.; Kremitzki, M.; Li, H.; Maduro, V. v.; Marschall, T.; McCartney, A. M.; McDaniel, J.; Miller, D. E.; Mullikin, J. C.; Myers, E. W.; Olson, N. D.; Paten, B.; Peluso, P.; Pevzner, P. A.; Porubsky, D.; Potapova, T.; Rogaev, E. I.; Rosenfeld, J. A.; Salzberg, S. L.; Schneider, V. A.; Sedlazeck, F. J.; Shafin, K.; Shew, C. J.; Shumate, A.; Sims, Y.; Smit, A. F. A.; Soto, D. C.; Sovi, I.; Storer, J. M.; Streets, A.; Sullivan, B. A.; Thibaud-Nissen, F.; Torrance, J.;

Wagner, J.; Walenz, B. P.; Wenger, A.; Wood, J. M. D.; Xiao, C.; Yan, S. M.; Young, A. C.; Zarate, S.; Surti, U.; McCoy, R. C.; Dennis, M. Y.; Alexandrov, I. A.; Gerton, J. L.; O'Neill, R. J.; Timp, W.; Zook, J. M.; Schatz, M. C.; Eichler, E. E.; Miga, K. H.; Phillippy, A. M. The Complete Sequence of a Human Genome. *Science (1979)* **2022**, *376* (6588), 44–53.

(241) Schneider, V. A.; Graves-Lindsay, T.; Howe, K.; Bouk, N.; Chen, H.-C.; Kitts, P. A.; Murphy, T. D.; Pruitt, K. D.; Thibaud-Nissen, F.; Albracht, D.; Fulton, R. S.; Kremitzki, M.; Magrini, V.; Markovic, C.; McGrath, S.; Steinberg, K. M.; Auger, K.; Chow, W.; Collins, J.; Harden, G.; Hubbard, T.; Pelan, S.; Simpson, J. T.; Threadgold, G.; Torrance, J.; Wood, J.; Clarke, L.; Koren, S.; Boitano, M.; Li, H.; Chin, C.-S.; Phillippy, A. M.; Durbin, R.; Wilson, R. K.; Flicek, P.; Church, D. M. Evaluation of GRCh38 and de Novo Haploid Genome Assemblies Demonstrates the Enduring Quality of the Reference Assembly. *bioRxiv* **2016**, 072116.

(242) Rhee, I.; Bachman, K. E.; Park, B. H.; Jair, K. W.; Yen, R. W. C.; Schuebel, K. E.; Cui, H.; Feinberg, A. P.; Lengauer, C.; Kinzler, K. W.; Baylin, S. B.; Vogelstein, B. DNMT1 and DNMT3b Cooperate to Silence Genes in Human Cancer Cells. *Nature* **2002**, *416* (6880), 552–556.

(243) Gibson, D. G.; Young, L.; Chuang, R. Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nat Methods* **2009**, *6* (5), 343–345.

(244) Gieß, M.; Muñoz-López, Á.; Buchmuller, B.; Kubik, G.; Summerer, D. Programmable Protein-DNA Cross-Linking for the Direct Capture and Quantification of 5-Formylcytosine. *J Am Chem Soc* **2019**, *141*, 9453–9457.

Eidesstattliche Versicherung (Affidavit)

Jung, Anne
Name, Vorname
(Surname, first name)

149875
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):

Imaging-Based Analysis of 5-Methylcytosine at Low-Repetitive Genomic Loci Using Transcription Activator-Like Effector Probes

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)