

Isolation of genomic DNA sequences with expanded nucleobase
selectivity using Transcription Activator-Like Effector Proteins

Dissertation

For the academic degree of

Doctor in Natural Sciences

(Dr. rer. nat.)

Chemistry and Chemical Biology department

Technische Universität Dortmund

M.Sc. Preeti Rathi

Dortmund 2017

Eidesstattliche Erklärung

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.

Dortmund, 2017

Publications

The experimental methods and results used in this study are part of the following scientific publications:

Rathi, Preeti; Maurer, Sara; Kubik, Grzegorz; Summerer, Daniel (2016) Isolation of Human Genomic DNA Sequences with Expanded Nucleobase Selectivity. In *Journal of the American Chemical Society* 138 (31), pp. 9910-9918. DOI: 10.1021/jacs.6b04807.

Rathi, Preeti; Maurer, Sara; Summerer, Daniel (2017) Selective Recognition of N4-Methylcytosine in DNA by Engineered Transcription-Activator Like Effectors. *Philosophical Transactions of the Royal Society B* Manuscript ID: RSTB-2017-0078.R1

Rathi, Preeti; Witte, Anna; Summerer, Daniel (2017) Engineering DNA backbone Interactions Results in TALE Scaffolds with Enhanced 5-Methylcytosine Selectivity. *Scientific Reports* DOI: 10.1038/s41598-017-15361-1

Index

Dissertation	1
Eidesstattliche Erklärung	2
Publications	3
Index	4
Acknowledgements	8
List of Abbreviations	9
List of Tables	11
List of Figures	12
Abstract	15
Zusammenfassung	17
I. INTRODUCTION	20
A. The eukaryotic genome	20
B. The central dogma	23
C. Genetics and epigenetics	25
1. Epigenetic modifications	26
2. Writers of methylation: DNA methyltransferases (DNMTs)	27
3. Readers of methylation	30
4. Erasers of methylation	31
5. Occurrence of C variants in cells	33
6. 4mC: An epigenetic C nucleobase found in prokaryotes	35
D. DNA methylation in oncology	36
E. Techniques for detection of epigenetic nucleobases	38
1. Methods involving chemical modification of the target	38
2. Methods involving targeting the modification	41
3. Methods involving sequence specific DNA Binding Domain (DBD)	42
F. Transcription Activator-Like (TAL) Effectors (TALEs)	46
1. Origin	46
2. Structure	46
3. Function	47
4. Interactions between different RVDs and bases	48
5. Non-specific interactions between the TALE protein and DNA	48
6. Mechanism of action	49
7. Factors affecting TALE function	50
8. Applications	50
II. AIM	52
III. MATERIALS AND METHODS	54

A. Materials	54
1. Consumables	54
2. Non-electronic equipments	55
3. Electronic devices	56
4. Chemicals	58
5. Enzymes	60
6. Antibiotics	61
7. Kits	62
8. Buffers	62
9. Media.....	65
10. Bacterial (E. coli) cell strains	66
11. Mammalian cell strains	66
12. Software	66
13. Oligonucleotides.....	67
14. Plasmids	80
15. TALE proteins.....	84
B. Methods.....	88
1. Preparation of gels.....	88
2. Preparation of antibiotic stocks	89
3. Preparation of LB-Agar plates	89
4. Preparation of GH371 chemical competent cells.....	89
5. Preparation of BL21 (DE3) Gold chemical competent cells.....	90
6. Transformation	90
7. Plasmid Isolation	91
8. Agarose gel electrophoresis	91
9. DNA purification.....	92
10. DNA quantification	92
11. DNA Sequencing.....	93
12. Golden Gate (GG) assembly protocol.....	93
13. Expression of TALE proteins.....	97
14. Purification of TALE proteins.....	97
15. Quantification of TALE proteins	98
16. SDS electrophoresis	98
17. Mutagenesis using quikchanges (Agilent)	99
18. Restriction ligation	100
19. Whole genome amplification (WGA).....	100
20. DNA extraction and/or purification, QIAmp mini DNA kit.....	101

21. Random shearing gDNA	101
22. Enzymatic methylation using CpG methyltransferase	101
23. Bisulphite conversion and PCR.....	102
24. Sodium borohydride mediated reduction of fC to hmC.....	102
25. Primer Extension (PEX).....	103
26. Affinity enrichment using TALEs.....	105
Using Ni-NTA magnetic beads	105
Using non-magnetic Ni-NTA resin	105
27. MeDIP	106
28. qPCR	106
29. Electromobility Shift Assay	107
30. Fluorescence Resonance Energy Transfer	107
31. MALDI.....	108
32. Luciferase assay	109
IV. RESULTS AND DISCUSSION	111
A. Isolation of specific gDNA sequences using TALEs.....	112
B. Sequence specific detection of 5mC in gDNA using TALEs	115
C. Quantitative detection of 5mC in gDNA samples.....	119
D. Simultaneous detection of 5mC in different loci of human genome	120
E. Strand specific detection of methylated CpG using TALEs	122
F. Comparison of TALEs and MeDIP for detection of 5mC in gDNA.....	124
G. Using size reduced RVD for detection of oxidative methyl C bases.....	128
H. Selective binder for a prokaryotic epigenetic C nucleobase; 4mC	131
I. Identification of RVD T* as a selective 4mC binder	133
J. Determination of K_i for TALEs Hey2_b (WT) and (T*) using FRET	135
K. Use of RVD HD and T* for discrimination between 5mC and 4mC	139
L. Reduced non-specific interaction improves TALE selectivity.....	140
M. K16A & Q17A mutation TALE CRD repeat reduces binding to 5mC	146
N. Combined mutation in NTR and CRD exhibit highest 5mC sensitivity.....	149
V. SUMMARY AND OUTLOOK.....	156
VI. SUPPLEMENTARY	158
A. Primer Extension gel photos	158
B. Ethidium bromide stained agarose gel images	159
C. Optimization of affinity enrichment using TALEs	160
D. Sequences of amplicons used in spike-in enrichment assay.....	164
E. Plasmid map and sequences of TRX_TALEs	166
F. Plasmid map and sequences of GFP_TALEs	169

G. Elaborate result of the MeDIP assay with internal controls.....	175
H. Bisulphite sequencing result for BRCA1 target site	176
I. Complete scan of TALE_Hey2_b X* library using PEX	177
J. Complete scan of TALE_Hey2_b X* library EMSA	179
K. Result of FRET measurement at different time points.	180
L. Determination of K_i for CDKN2A WT and T* TALEs using FRET	183
M. TALE sequences with regions bearing alanine mutations	185
N. EMSA gel photos for TALEs bearing electrostatic mutations	187
O. Summary of binding by mutated CDKN2A TALEs.....	189
P. Determination of K_i for CDKN2A and ZAP-70 WT and mutated TALE	190
Q. Measurement of in vivo activation assay at different time points	192
R. Maps of plasmids of group A and B used in luciferase assay.....	193
S. Bisulphite sequencing result for CDKN2A target site	195
VII. REFERENCES	196

Acknowledgements

The complete work presented here took about 3.5 years and the genuine effort from many people excluding me. In this final step of my work, I would like to extend my heartfelt thanks to these people. The first person that I am obliged to thank is my Mentor, Prof. Dr. Daniel Summer for giving me this opportunity and always being supportive and encouraging. Not only did he believe in my ability to execute a task but also offered much needed enthusiasm at critical times. The entire work environment designed by him helped me learn and grow scientifically.

My special thanks to Dr. Moritz Johannes Schmidt and Dr. Grzegorz Kubik for helping me practically with everything in my initial days of joining the group. They have been wonderful examples to look up to as well as true friends all through my journey.

No work is enjoyable in a hostile environment and therefore I sincerely thank Dr. Sarah Flade, Sara Maurer, Mario Giess, Anna Witte, Alvaro Munoz and Jan Wolffgramm for the scientific or non-scientific discussions and advice together with the wonderful time I spent in their company.

I thank Katharina Kuhr, Simone Eppmann, Petra Alhorn, Martina Reibner for their perfection in managing everything and being eager to help in any way possible.

I would like to re-thank Dr. Moritz Johannes Schmidt, Dr. Sarah Flade, Sara Maurer for checking, editing and translation of some parts of the thesis.

Heartfelt thanks to my husband Mr. Deep Lomesh Shah for all the expected and unexpected support. Without his unconditional love and believe I would have never made it this far. I would like to thank my late grandmother, my grandfather, my mother, my father and my mother-in-law for providing everything in their power to support me from a long distance.

Finally, I would like to thank the Department of Chemistry and Chemical Biology in Universities of Konstanz and the Technical University of Dortmund for permission to conduct research in their premises. Special thanks to adjoining groups; Max Plank Institute for Molecular Physiology, Universities of Dortmund and Konstanz for permitting the use of different machine/equipment and all the funding agencies for providing the economic means throughout the work.

List of Abbreviations

A	Adenine
APOBEC	Apolipoprotein B mRNA editing enzyme catalytic polypeptide like
APS	Ammonium Persulphate
BER	Base Excision Repair
bp	Base Pair
BRCA1	Breast Cancer 1
BRCA1_a	BRCA1_Antiparallel
BS-seq	Bisulphite Sequencing
C	Cytosine
CAB-seq	Chemical Associated Bisulphite Sequencing
Cys	Cysteine
CpGs	CpG Islands
CDKN2A	Cyclin-dependent kinase Inhibitor 2AC
CLL	Chronic Lymphocytic Leukaemia
CRD	Central Repeat Domain
CTR	C-Terminal Region
Da	Dalton
DNA	Deoxyribonucleic Acid
DNMTs	DNA Methyltransferases
ds	Double Stranded
DSBH	Double Stranded Beta Helix
DTT	Dithiothreitol
EMSA	Electromobility Shift Assay
fCAB-seq	5fC Chemical Assisted BS-seq
FRET	Fluorescence Resonance Energy Transfer
Fw	Forward
gDNA	Genomic DNA
GFP	Green Fluorescent Protein
h	Hour
HD	Histidine aspartate
HDAC	Histone Deacetyltransferase
Hey2	Hairy/enhancer-of-split related with YRPW motif protein 2
His6	Histidine tag
hmU	Hydroxymethyluracil
IF	Initiation Factor
IP	Imaging Plate
k	Kilo
KF	Klenow Fragment
L	Litre/s
MALDI-TOF	Matrix Assisted Laser Desorption/Ionization-Time of Flight
MBD	Methyl Binding Domain
MeDIP	Methylation dependent Immunoprecipitation
mg	Milli Gram
mM	Milli Molar
MGMT	06-Methylguanine-DNA Methyltransferase
min	Minute(s)
miRISC	miRNA Induced Silencing Complex
miRNA	Micro RNA
mRNA	Messenger RNA
NI	Arginine isoleucine
Ni-NTA	Nickel-Nitrilotriacetic acid
ng	Nano gram(s)

nm	Nano meter
nM	Nano Molar
NG	Arginine glycine
NN	Arginine arginine
NGS	Next Generation Sequencing
NTR	N-Terminal Region
Oligo	Oligonucleotides
OxBS-seq	Oxidation Bisulphite Sequencing
PCR	Polymerase Chain Reaction
PEX	Primer Extension
p.p.	Primer Pair(s)
qPCR	Quantitative PCR
redBS-seq	Reduced Bisulphite Sequencing
RLGS	Restriction Landmark Genomic Scanning
rpm	Rotation per Minute
RT	Room Temperature
RNA	Ribonucleic Acid
rRNA	Ribosomal RNA
RVD	Repeat Variable Di-Residue
Rv	Reverse
S	Seconds(s)
SAH	S-adenosylhomocysteine
SAM	S-adenosylmethionine
SMUG1	Single strand specific Monofunctional Uracil DNA
T	Thymine
TAB-seq	Tet Assisted Bisulphite Sequencing
TALEs	Transcription Activator Like Effector(s)
TDG	Thymine DNA Glycosylase
TRD	Transcriptional Repressor Domain
tRNA	Transfer RNA
TRX	Thioredoxin
U	Uracil
UHRF	Ubiquitin-like, containing PHD and RING finger domains
V	Volts
WGA	Whole Genome Amplification
WT	Wild Type
ZAP-70	Zeta-chain Associated Protein Kinase 70
Zf	Zebrafish
Zn	Zinc
ZFP	Zinc Finger Protein
X-gal	5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside
5mC	Methylcytosine
5fC	Formylcytosine
5hmC	Hydroxymethylcytosine
5caC	Carboxylcytosine
4mC	N4 Methylcytosine
$^{\circ}$ C	Degree Celsius

List of Tables

Table III-1: Consumables	55
Table III-2: Non-electronic equipments	55
Table III-3: Electronic devices	58
Table III-4: Chemicals.....	60
Table III-5: Enzymes	61
Table III-6: Antibiotics.....	61
Table III-7: Kits.....	62
Table III-8: Buffers.....	65
Table III-9: Media	66
Table III-10: Bacterial strains.....	66
Table III-11: Mammalian cell line.....	66
Table III-12: Softwares.....	67
Table III-13: Oligonucleotides	79
Table III-14: Plasmids	84
Table III-15: TALEs.....	87

List of Figures

Figure I-1: Eukaryotic cell: structure and composition of DNA	21
Figure I-2: Different configurations of DNA	22
Figure I-3: Protein translation	25
Figure I-4: Modified C variants found in eukaryotes	27
Figure I-5: Formation of 5mC	27
Figure I-6: Eukaryotic DNMTs	29
Figure I-7: Link between DNA methylation and transcription inhibition.....	31
Figure I-8: Active pathway for C demethylation.....	33
Figure I-9: Base pairing between G and different C modifications	34
Figure I-10: Different methylated C in prokaryotes.....	35
Figure I-11: DNA methylation in cancer genes.....	36
Figure I-12: Summary of BS-sequencing methods	40
Figure I-13: Working principle of immunoprecipitation.....	42
Figure I-14: Crystal structure of ZFP bound to its target DNA	43
Figure I-15: Genome engineering using CRISPR-Cas9.....	45
Figure I-16: Crystal structure of TALE bound to its target DNA	46
Figure I-17: Model of TALE protein used in this study.....	47
Figure I-18: Crystal structure of RVD HD interacting with C	49
Figure I-19: Schematic representation of TALEN function.....	49
Figure III-1: Schematic representation of GG assembly.....	94
Figure III-2: Agarose gel photo for cpcr of GG1 plasmid.....	96
Figure III-3: Agarose gel photo for cpcr of GG2 plasmid.....	96
Figure III-4: Agarose gel photo for test digest of final TALE vector	97
Figure III-5: SDS gel photo of purified TALE proteins.....	99
Figure III-6: Working principle of PEX assay	104
Figure III-7: PAGE gel photo of PEX assay with BRCA1 (26) TALE	104
Figure III-8: Working principle of FRET assay	108
Figure IV-1: TALE target sequence in Zf Hey2 locus	113
Figure IV-2: TALE target sequences in human genome.....	113
Figure IV-3: Pictorial representation of TALE based affinity enrichment assay.....	114
Figure IV-4: Isolation of genomic sequences using TALEs	115
Figure IV-5: Comparison of Zf Hey2 and human MGMT target sequence.....	115
Figure IV-6: Working principle behind detection of C/5mC using affinity enrichment...	116
Figure IV-7: Detection of 5mC in Zf genome using affinity enrichment	117
Figure IV-8: Detection of 5mC in cancer relevant genes using affinity enrichment	118
Figure IV-9: Quantification of methylation in CDKN2A locus of human gDNA.....	119
Figure IV-10: Quantification of methylation in Hey2_b locus of Zf gDNA.....	120
Figure IV-11: Working principle of multiplexing (MUX) enrichment assay	121
Figure IV-12: Result of MUX using TALEs targeting different human loci.....	122
Figure IV-13: Working principle and result of strand specific enrichment assay.....	123
Figure IV-14: Working principle of MeDIP assay.....	124
Figure IV-15: Description of sample used for MeDIP assay	125

Figure IV-16: Result of MeDIP assay	126
Figure IV-17: Comparison between MeDIP and TALE based detection of 5mC.....	127
Figure IV-18: Crystal structure of RVD N* binding 5mC	127
Figure IV-19: Affinity enrichment using TALEs containing RVD N*	128
Figure IV-20: Affinity enrichment for detection of 5hmC in Zf gDNA	129
Figure IV-21: Crystal structure of epigenetic C modification in DNA duplexes.....	130
Figure IV-22: Affinity enrichment for detection of C, 5mC, 5hmC, 5fC and 5caC	130
Figure IV-23: NaBH ₄ mediated reduction of 5fC to 5hmC	131
Figure IV-24: Affinity enrichment based discrimination between 5hmC and 5fC	131
Figure IV-25: Crystal structure of RVD N* binding C	132
Figure IV-26: Crystal structure of base pairing between G and 4mC.....	133
Figure IV-27: Hey2_b target sequence containing C, 5mC or 4mC	133
Figure IV-28: Result of PEX assay using TALE Hey2_b T*	134
Figure IV-29: EMSA gel using TALE Hey_2b WT and T*	135
Figure IV-30: FRET assay using TALE Hey2_b WT for C/5mC containing sequence ...	136
Figure IV-31: Scan of TALE_Hey2_b X* library using FRET	137
Figure IV-32: Discriminating between C, 5mC and 4mC in Hey2_b using FRET.....	138
Figure IV-33: Affinity enrichment assay using TALE_Hey2_b WT and T*	140
Figure IV-34: Crystal structure of DNA bound and unbound TALE NTR	141
Figure IV-35: Model of TALEs used in electrostatic mutation studies	142
Figure IV-36: Crystal structure of TALE NTR with mutation sites.....	143
Figure IV-37: C/5mC detection by NTR modified TALE_CDKN2A	145
Figure IV-38: C/5mC detection by NTR modified TALE_ZAP-70	145
Figure IV-39: Crystal structure of TALE CRD with mutation sites	146
Figure IV-40: C/5mC detection by CRD modified TALE_CDKN2A.....	148
Figure IV-41: C/5mC detection by CRD modified TALE_ZAP-70	148
Figure IV-42: C/5mC detection by NTR&CRD modified TALE_CDKN2A & ZAP-70.	150
Figure IV-43: Affinity enrichment using NTR&CRD modified TALEs	152
Figure IV-44: Working principle of TALE based <i>in vivo</i> transcription activation assay..	153
Figure IV-45: Controls of TALE based <i>in vivo</i> transcription activation assay.	155
Figure IV-46: Results of TALE based <i>in vivo</i> transcription activation assay	155
Figure VI-1: PAGE gels with PEX assay using different TALEs.....	158
Figure VI-2: Agarose gel showing randomly sheared Zf gDNA	159
Figure VI-3: Agarose gel showing randomly sheared human gDNA	159
Figure VI-4: Optimization of affinity enrichment assay	163
Figure VI-5: Map of TRX_TALE entry plasmid	166
Figure VI-6: Map of GFP_TALE entry plasmid	169
Figure VI-7: Complete MeDIP assay with internal controls.....	175
Figure VI-8: BS-sequencing of BRCA1 target sequence	176
Figure VI-9: Complete scan of TALE_Hey2_b X* using PEX assay	178
Figure VI-10: Complete scan of TALE_Hey2_b X* using EMSA assay	180
Figure VI-11: FRET positive controls of sequence containing C, 5mC or 4mC	180
Figure VI-12: Complete scan of TALE_Hey2_b X* using FRET assay	182
Figure VI-13: Ki determination for C/5mC/4mC sequences with TALE HD & T*.....	184

Figure VI-14: EMSA gels using TALE_CDKN2A bearing different mutations.....	187
Figure VI-15: EMSA gels using TALE_ZAP-70 bearing different mutations	188
Figure VI-16: Effect of electrostatic mutation on TALE_CDKN2A function.....	189
Figure VI-17: Ki for C/5mC samples with WT & mutated TALE_CDKN2A	191
Figure VI-18: Ki for C/5mC samples with WT & mutated TALE_ZAP-70	192
Figure VI-19: Map of plasmid of group A for <i>in vivo</i> transcription activation assay	192
Figure VI-20: Map of plasmid of group B for <i>in vivo</i> transcription activation assay	194
Figure VI-21: BS-sequencing of CDKN2A target sequence.....	195

Abstract

Deoxyribonucleic acid (DNA) is an information storage molecule that contains the instructions needed for a cell to sustain. The genetic blueprints stored in the sequence of the DNA exhibits the well-established Mendelian inheritance however, epigenetic processes cause cells with same genetic information to follow a different development pathway. In other words, epigenetic modifications can cause changes in the phenotype without altering the genotype. Epigenetic processes are natural and essential for various functions in an organism however if they occur improperly there can be majorly adverse health and behaviour effects.

In mammals, DNA methylation is an epigenetic signal occurring predominantly at the carbon-5-position of cytosine (C) nucleobase existing in a symmetrical CpG (cytosine and guanine separated by a phosphate) dinucleotides. This process gives rise to the most widely characterized epigenetic nucleobase called 5-methylcytosine (5mC) [1]. When located in gene promoters, 5mC causes downregulation of gene transcription [2]. Recent studies have illuminated the essential role of 5mC in cellular events like genomic imprinting, X-chromosome inactivation and repression of transposable elements [3]. Further, aberrant gene regulation due to differences in their epigenetic make-up is associated with pathological states and causes of underlying diseases [4]. DNA methylation was first confirmed to occur in human cancer in 1983. Since then several studies have shown a direct link between the presence of 5mC in promoters of tumour suppressor genes and various cancers [5][6]. Additionally, other studies suggest a role of 5mC in autoimmune diseases, metabolic disorders and neurodegenerative ailments such as Schizophrenia, Fragile X and Prader-Willi syndrome [7–9]. Hence the detection of epigenetic modifications not only plays a crucial role in diagnosis and treatment of several chronic diseases [10,11] but may also provide insights in understanding the role of different epigenetic modifications in regulating various cellular mechanisms. Available methods for the detection of 5mC are either costly, time consuming or inefficient making new tools for efficient detection of different epigenetic modifications highly desirable.

This thesis was focussed on the development of a novel approach for sequence specific detection of epigenetic C modifications in genomic DNA (gDNA) using Transcription Activator-Like Effector (TALE) proteins. TALEs are DNA binding proteins composed of repeats containing conserved amino acid sequences. The 12th and the 13th amino acid of

each repeat denoted as Repeat Variable Di-residue (RVD) however differ depending on the nucleobase targeted. Earlier studies have established the ability of TALE RVD HD (histidine aspartate) to discriminate between DNA sequences differing in harbouring a C or 5mC at a single position [12–14]. Identification of engineered RVDs capable of selectively recognizing other C modifications like hydroxymethyl-, formyl- and carboxylcytosine (5hmC, 5fC and 5caC) have also been reported [15,16].

TALEs were employed in an affinity enrichment assay for modification sensitive binding and isolation of user-defined DNA sequence directly from gDNA samples. This method featuring the use of TALEs as selective sensor for detecting different epigenetic C modifications in gDNA was first of its kind. It enabled quantitation of methylation and detection of hemi-methylation in cancer biomarker loci. In combination with chemical reduction of 5fC to 5hmC, TALEs with a size-reduced RVDs were shown to successfully detect and discriminate between genomic sequence containing a 5hmC or 5fC [17].

N4-methylcytosine (4mC) is a rare epigenetic modification that co-exists with 5mC in some prokaryotic organism. Similar to another epigenetic modification called 6-methyladenine (6mA), 4mC may play a role in host defence and transcription regulation in the bacterial genome [18]. Successful discrimination of C, 5mC and 4mC using the established affinity enrichment assay further highlighted the use of TALEs as tool for studying the individual roles of various epigenetically modified nucleobases found in a genome [19].

Off-target binding exhibited by TALEs can negatively affect their widespread use. Positively charged amino acids in the N-terminal region (NTR) or Central Repeat Domain (CRD) of TALEs are known to bind non-specifically to the negatively charged phosphate backbone of DNA. [20]. To enhance TALE binding sensitivity, these positively charged amino acids were exchanged with alanine. Compared to non-modified TALEs, those with alanine mutations exhibited significantly increased 5mC sensitivity in affinity enrichment assay and in C selective transcription activation in HEK293T cells as demonstrated by the luciferase assay. The results of this study indicated the possibility of developing modified TALEs for enhanced precision and sensitivity in detection of epigenetic modifications [21].

Zusammenfassung

Die Zelle speichert die genetische Information in Form von DNA (engl. desoxyribonucleic acid). Dieses Molekül liefert somit den genetischen Bauplan für RNA und Proteine, die für die Zellfunktion benötigt werden. Die Information auf der DNA ist erblich und folgt den klassischen Vererbungsregeln nach Mendel. Epigenetische Prozesse können bewirken, dass Zellen mit dem gleichen genetischen Material einen anderen Phänotyp entwickeln. Diese Änderung basiert jedoch nicht auf einer Veränderung des Genotyps. Epigenetische Prozesse sind ein natürliches Event und essentiell für zahlreiche Funktionen im Organismus. Fehlerhafte epigenetische Modifikationen können jedoch zu Entwicklungen von Krankheiten und Verhaltensänderungen führen.

In Säugetieren bestimmt die Methylierung von DNA überwiegend an der C5-Position von Cytosin-Nukleobasen, (C) innerhalb von symmetrischen CpG (Cytosin und Guanin getrennt durch ein Phosphat), das epigenetische Signal. In Prozess der Methylierung entsteht die epigenetische Base 5-Methylcytosin (5mC) [1]. Wenn 5mC jedoch in Promotorregionen der DNA präsent ist, führt diese Modifikation zu einer reduzierten Gentranskription [2]. Studien beleuchten die essentielle Rolle von 5mC in zellulären Prozessen wie zum Beispiel X-Chromosom Inaktivierung und Repression von Transposons [3]. Die fehlerhafte Regulierung von Genen basierend auf abweichenden epigenetischen Modifikationen ist oftmals mit pathologischen Symptomen verbunden und kann als direkte Krankheitsursache identifiziert werden [4]. In 1983 wurde erstmals die aberrante Methylierung in menschlichen Krebszellen festgestellt. Seitdem zeigten Studien eine direkte Verbindung zwischen 5mC positioniert innerhalb von Promotorregionen von Tumorsuppressorgenen und zahlreichen Krebsarten [5,6]. Weitere Studien suggerieren die Rolle von 5mC in Autoimmunkrankheiten, Stoffwechselstörungen und neurodegenerative Krankheiten wie zum Beispiel Schizophrenie, fragiles X- und Prader-Wili-Syndrome [7-9]. Die Detektion von epigenetischen Modifikationen spielt nicht nur eine Rolle in Diagnose und Behandlung von zahlreichen chronischen Krankheiten [10,11], sondern könnte auch wichtige Einblicke über epigenetische Modifikationen und ihre regulierende Rolle in zellulären Mechanismen liefern. Gängige Methoden für die Detektion von 5mC sind bisher kostspielig, zeitaufwendig oder ineffizient. Daher besteht weiterhin ein Bedarf an neuen Methoden, welche die verschiedenen epigenetischen Modifikationen effizient detektieren können.

Die vorliegende Arbeit befasst sich mit der Entwicklung eines neuen Ansatzes zur sequenzspezifischen Detektion von epigenetischen C-Modifikationen in genomischer DNA (gDNA) mit Hilfe von TALE Proteinen (Transcription-activator-like effectors). TALEs sind DNA bindende Proteine welche aus sogenannten Repeats bestehen. Diese Repeats beinhalten konservierte Aminosäuren mit Ausnahme von Position 12 und 13. Diese Positionen werden als Repeat Variable Di-residue (RVD) bezeichnet und sind variabel abhängig von ihrer gegenüberliegenden DNA Base. Aktuelle Publikationen zeigen, dass TALE Proteine, die ein Histidin-Aspartate (HD) als RVD tragen, zwischen DNA-Sequenzen mit C oder 5mC an einer einzelnen Position diskriminieren können [12-14]. Es wurde weiterhin über die Identifizierung von modifizierten RVDs, welche weitere Cytosin Modifikationen wie Hydroxymethyl-, Formyl-, und Carboxycytosin (5hmC, 5fC und 5caC) selektiv erkennen können, berichtet [15,16].

In dieser Arbeit wurden TALE Proteine für einen Affinitäts-Anreicherungs-Assay für die Modifikations-sensitive Bindung und Isolierung von spezifischen genomischer DNA-Sequenzen eingesetzt. Diese Methode nutzt erstmals TALE Proteine als selektive Sensoren für die Detektion von epigenetischen Cytosin Modifikationen innerhalb von gDNA. Dieser Assay ermöglicht die Quantifizierung von epigenetischen Methylgruppen sowie die Detektion von hemi-methylierten DNA-Sequenzen in Krebsrelevanten Loci. In Kombination mit chemischen Modifikationen konnte gezeigt werden, dass TALE Proteine mit Größen-reduzierten RVDS zwischen 5hmC und 5fC innerhalb von genomischer DNA Sequenzen unterscheiden können [17].

Neben der epigenetischen Modifikation 5mC gibt es auch die weniger häufig vorkommende DNA Base N4-Methylcytosin (4mC). Beide Modifikationen koexistieren in zahlreichen prokaryotischen Organismen. Ähnlich zu 6-Methyladenine (6mA), welches ebenso in diesen Organismen vorkommt, könnte 4mC eine Rolle in der Genomstabilität und Transkriptionsregulation einnehmen [18]. Die erfolgreiche Diskriminierung zwischen C, 5mC und 4mC mit Hilfe des Affinitäts-Anreicherungs-Assays bestätigt TALE Proteine weiterhin als nützliche Werkzeuge für die Erforschung der individuellen Funktionen zahlreicher epigenetisch modifizierter Nukleinbasen innerhalb des Genoms [19].

Die mögliche unspezifische Bindung durch TALE Proteine könnte ihren weitverbreiteten Einsatz negativ beeinflussen. Positiv geladene Aminosäuren im N-terminalen Bereich oder in der zentralen Repeat-Domäne können unspezifisch an das negativ geladene

Phosphatrückgrad der DNA binden [20]. Um die TALE Bindungssensitivität zu verbessern, wurden diese positiv geladenen Aminosäuren durch neutral geladene Alanin Aminosäuren ersetzt. Diese modifizierten TALE Proteine zeigten eine erhöhte 5mC-Sensitivität im Vergleich zu nicht-modifizierten TALEs. Diese konnte innerhalb des Affinitäts-Anreicherungs-Assay und einem Luciferase-Assay für C-selektive Transkriptionsaktivierung in HEK293T Zellen gezeigt werden. Die Ergebnisse dieser Arbeit beleuchten die Möglichkeit, TALE Proteine weiterhin zu modifizieren, um eine erhöhte Präzision und Sensitivität für die verbesserte Detektion von Epigenetischen Basen zu erzielen [21].

I. INTRODUCTION

A. The eukaryotic genome

The most fundamental property of all living things is their ability to reproduce [22]. The cell, discovered by Robert Hooke in 1665 is the basic structural, functional and biological unit of all living organisms [23]. Each eukaryotic cell consists of an external cytoskeleton and a variety of internal membrane-bound structures called organelles [24]. (Figure I-1). A parent cell replicates to form new 'daughter' cells in a process called cell division [25]. The genetic information of a cell is contained in the linear order of the nucleotides in its DNA which is sequestered in the nucleus of a eukaryotic cell. DNA is organized in chromosomes, each representing one double stranded DNA molecule and the total genetic information stored in the chromosomes of an organism is said to constitute its genome. A genome comprises of all the instructions needed to build, grow and develop any organism [26]. The DNA of all chromosomes is packaged into a compact structure to fit into a cell's nucleus with the aid of specialized proteins. DNA-binding proteins in eukaryotes are classified into two general classes: histones and non-histone chromosomal proteins. The complex of these two classes of proteins with the nuclear DNA of eukaryotic cells is known as chromatin. The nucleosome is the fundamental repeating unit of chromatin. Histones are small, highly basic proteins that are segregated into two main groups – the nucleosomal histones and the H1 histones. The four nucleosomal histones: H2A, H2B, H3 and H4 are responsible for coiling DNA into nucleosomes and histone H1 molecules are thought to be responsible for pulling nucleosomes together into a regular repeating array. The basic chromatin unit consists of a protein octamer containing two molecules of each nucleosomal histones around which is 147 bp of DNA is wrapped [27,28,28].

DNA, an unbranched, linear polymer comprises of two coiled helical strands running in direction opposite to each other. It consists of alternating phosphate and sugar residues on the exterior of the helix with the antiparallel strands joined internally by nucleobase pairing. Externally, the five-carbon sugar (2-deoxyribose) molecules are joined together by phosphodiester bonds, formed between the intermediate phosphate and the third and fifth carbons respectively of the alternate sugars. The hydroxyl group on the first carbon of the sugar on the other hand, is replaced by the internally placed nucleobases [29]. The four nucleobases found in the naturally occurring DNA are classified as purines or pyrimidines.

Purines are formed by fusion of five and six membered heterocycles while the pyrimidines are six-membered heterocycles. Within a DNA all base pairing occurs between a purine and a pyrimidine. Adenine (A) and guanine (G) (purines) on either strand binds respectively to thymine (T) and C (pyrimidines) on the other strand [27]. The nucleobases are joined together by hydrogen bond with double bonds between A and T and triple bonds between G and C. Together with the hydrogen bonds between opposite bases, the base-stacking between the adjacent aromatic bases and the hydrogen bond formed between the bases and the surrounding water molecules plays a key role in stabilizing the DNA double helix [30,31].

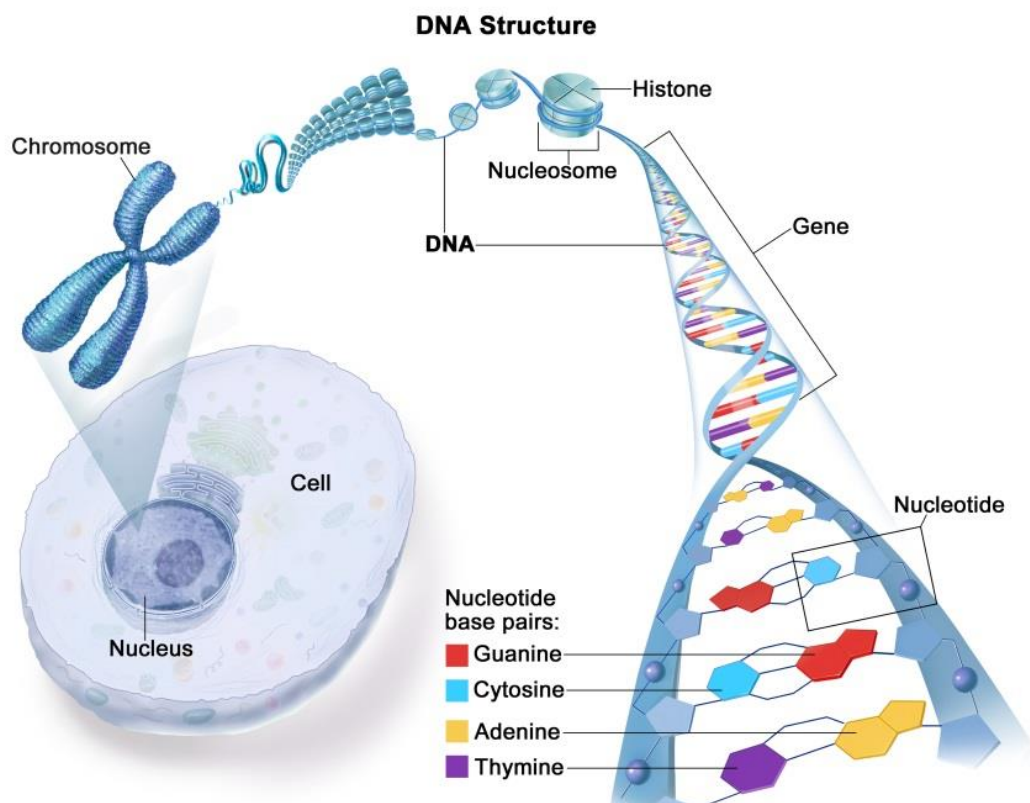


Figure I-1: Eukaryotic Cell. Packaging, structure & composition of nucleic DNA [32].

The two strands of DNA are not in a simple side-by-side arrangement but are instead coiled around the same axis and intertwined with themselves. DNA exists in A, B or Z-forms (Figure I-2). Numerous factors contribute to these different configurations. Amongst them lies the level of hydration, DNA sequence, amount and degree of supercoiling, chemical modification of the nucleobases, presence and concentration of polyamines and metal ions [33]. The A-form is the least abundant and occur in non-physiological, extremely dehydrated

DNA samples or those produced by hybridization of DNA with RNA. The Z-DNA has a left handed helical structure and is formed by stretches of alternating purines and pyrimidines e.g. GCGCGC, especially in negatively supercoiled DNA. Some studies suggest exclusive recognition of the Z-DNA by special proteins like the ones involved in transcription regulation, however conclusive evidence for or against this proposal is not yet available. The B-form is the most common conformation at a neutral pH and in the physiological conditions found in a cell [34]. The B-form of the DNA double helix has about 10 bp per turn with the major groove being wider than the minor groove. As a result, some DNA binding proteins such as transcription factors find bases on the major groove more accessible [35].

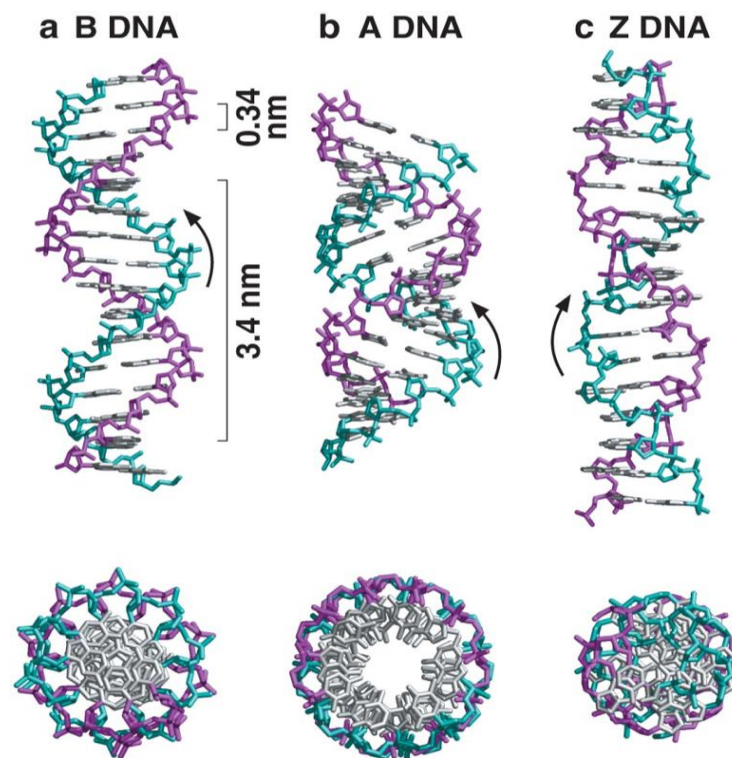


Figure I-2: Different conformations of naturally occurring DNA [36].

The two different nucleic acids found in eukaryotic cells i.e., DNA and RNA differ not only in the type of sugar found on the backbone (the deoxyribose is replaced by a less stable ribose molecule) and the replacement of the pyrimidine base T by uracil in RNA but also in their functions, reactivity and stability. While DNA is responsible for long-term stable storage and transmission of genetic information, the various forms RNA, which exists

mostly as single strands enables the translation of the code stored in DNA into functional proteins that provides the various genotypic and phenotypic traits.

1) Replication of DNA is essentially the process by which it makes a copy of itself. The replication begins with the Initiation step wherein the initiator proteins target certain A-T rich regions in the DNA called 'origins' [37] and fuse together with other recruited proteins to form the pre-replication complex [38]. This complex contains helicases which unzips the double stranded DNA by breaking the intermediate hydrogen bonds starting at the 'origin' to yield branching fork like structure [39,40]. The two single strands hence generated have opposite directionality and each act as a template on which the new strands are created by DNA polymerase; an enzyme capable of adding complementary nucleobases to partially single stranded DNA. Depending on which carbon atom in the deoxyribose attaches itself to the next phosphate, the two DNA strands have 5' to 3' or 3' to 5' direction. During the elongation step, a single RNA primer attaches itself to the 3' end of the leading strand (5' to 3') and the new strand is synthesized continuously in 5' to 3' direction, while the lagging strand receives more than one short RNA primer attachments at different sites which are individually extended by the polymerase giving rise to Okazaki fragments that are then joined together by DNA ligase. Termination of DNA extension occurs at various sites in eukaryotic DNA as the different replication forks cannot meet at a site owing to the non-circular arrangement of DNA in the chromosome. Replication generally aborts close to the telomere region for repetitive DNA due to blocking or stopping of the replication fork.

B. The central dogma

The central dogma encapsulates the processes involved in maintaining and translating the genetic template required for life. The three essential stages are 1) the self-propagation of DNA by semi-conservative replication; 2) unidirectional transcription templated by the genetic code (DNA) for generation of an intermediary messenger RNA (mRNA); 3) translation of the mRNA to produce strings of amino acids colinear with the 5' to 3' order of DNA (Figure I-3).

2) The transcription bubble separates the two strands of the DNA double helix providing RNA polymerases the chance to scan and copy certain segments of antisense DNA strand into complementary RNA strands (generated in 5' to 3' direction). Particular parts of the

DNA called the transcription units encode for one or more genes. A gene that encodes for a protein produces mRNA alternatively, others encode for different forms of RNA i.e. Transfer RNA (tRNA), Ribosomal RNA (rRNA) or microRNA (miRNA) [41]. For the mRNA to proceed to translation, it must leave the nucleus and travel to the cytoplasm where the ribosomes are located. 3) Ribosomes are organelles composed of two subunits which detach initially and close after attachment of the small subunit to mRNA strand with the help of Initiation factors (IFs). The ribosomal subunit consists of specialized proteins and RNA molecules which together manufacture proteins from the sequence information of mRNA. The ribonucleotide sequence is always read in group of three bases called codons. tRNAs acts as an adaptor molecule which has anticodons on their base that recognises complementary codon sequences. Once, specific tRNA attaches to the matching site on the mRNA, tRNA aminoacyl synthetase catalyses the addition of the respective amino acid to the top of the tRNA molecule. This new amino acid is attached to the C-terminus of the previously generated peptide. As tRNAs scan all the codons on the mRNA, respective amino acids keep getting added on the other end till some specific codons called the stop codons are encountered. The combinations of various peptides create complex and functional proteins essential for all living organisms. All possible combinations of ribonucleotides give rise to 64 different codons, of which three are stop codons which terminates the peptide chain extension by recruiting the release factor protein to the ribosome and the releasing the polypeptide chain. The remaining 61 codons can encode for twenty amino acids called standard or canonical amino acids. Other classes of amino acids include those that are not coded by DNA directly but can be incorporated into proteins during translation. These are known as the non-standard proteinogenic amino acids. Selenocysteine found both in prokaryotes and eukaryotes and pyrrolysine found only in some archaea and bacterium are examples of such amino acids. These amino acids are coded via variant codons. Yet another group of non-proteogenic amino acids exists, those that are produced directly isolated from the usual cellular machinery and are incorporated into proteins during post-translational modification. Some examples include carboxy glutamate, 2-aminoisobutyric acid, ornithine etc. [40]. In the past decade 40 non-natural amino acids have been incorporated into proteins using unique codons and corresponding tRNA aminoacyl synthetases [41] (Figure I-3).

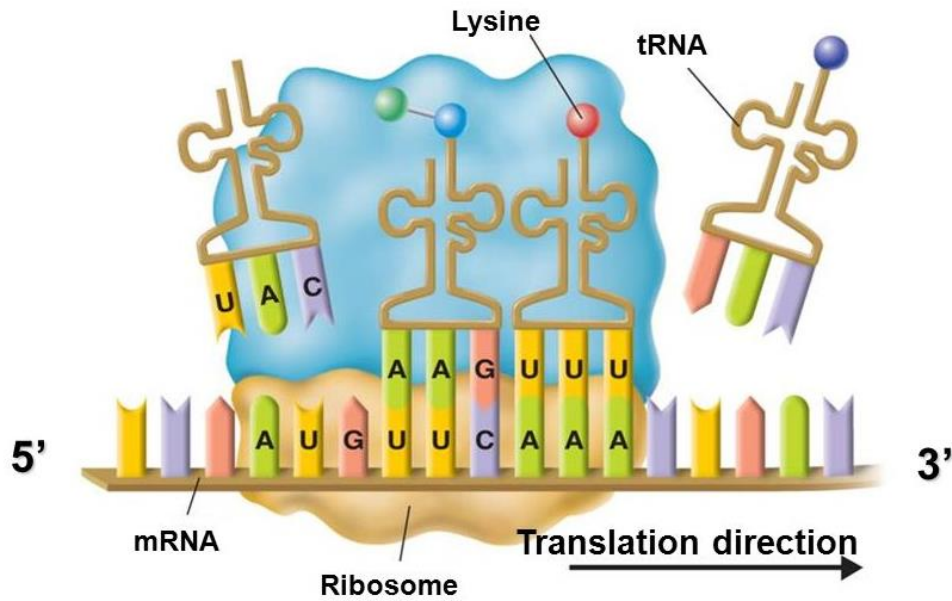


Figure I-3: Translation: method of protein generation in cell [42].

C. Genetics and epigenetics

The fundamental laws of inheritance were established in the 18th century through the work of Johann Gregor Mendel on pea plants. The genetic information stored in DNA is propagated through the germ line to the next generation. Since then the known role of genetics have extended from being responsible for physical resemblance to imparting identical make-up of the DNA sequences that controls almost all the attributes of an individual organism including susceptibility and defence against diseases. Genetics provide the blueprint for manufacture of all functional proteins. The complex process of DNA replication may occasionally bear errors. Often these aberrations are corrected by the standard cellular machinery but sometimes the alterations are permanent and gets incorporated and stored in the gene sequences. Such changes are known as the genetic mutations and can vary drastically in size, regions and extent of influence on the organism. Mutations can cause changes in the nucleotide sequences by insertion, deletion exchange or translocation of bases. Likelihood for many chronic and life-threatening diseases has been linked to a single or combination of genetic mutations.

The term “Epigenetics” was coined in the year 1942 by a British born developmental biologist, Conrad Hal Waddington. Its idea originated due to long-standing studies of

seemingly anomalous (i.e., non-Mendelian) and disparate patterns of inheritance in many organisms. Since then, in the given light of new discoveries, there have been several accepted definitions. Today it can be categorised as stable, heritable changes in the chromosomes that do not change the sequence of nucleotides, instead affect the expression of their genes. Although the genetic make-up of all cells is identical, epigenetics control how, when and where the genetic information should be used or turned off if not required. The greek prefix *epi* mean “on top of” hence, epigenetic features are an addition to the pre-existing genetic basis of inheritance [43]. These changes require unique mechanism for introduction, recognition, removal and enables different regulation of the same genetic region depending on the cell type or physiological and environmental conditions. DNA methylation, small non-coding RNAs, histone modifications and nucleosome positioning, belong to the group of epigenetic modifications since they cause chromatin modification and affect gene transcription. Co-ordinated functioning of epigenetic mechanisms is essential to regulate the epigenetic landscape. Different groups of enzymes capable of either creating (writers), deciphering (readers), changing (editors) or deleting (erasers) the epigenetic marks are crucial members of the family of epigenetic regulators [44].

1. Epigenetic modifications

DNA methylation is yet the most abundantly observed and well characterized epigenetic modification. While 5mC constitutes the earliest found and most frequently occurring nucleobase methylation in eukaryotes, 6-methyladenine (6mA) is additionally known to exist. Recently, oxidized forms of 5mC that is 5hmC, 5fC and 5caC have been discovered (Figure I-4).

DNA methylation along with other epigenetic mechanisms including histone modifications and non-coding RNAs provide a set of interrelated pathways that all create variation in the chromatin polymer. Marks on the chromatin template may be heritable through cell division and collectively contribute to determining cellular phenotype.



Figure I-4: Epigenetic variants of C nucleobase found in eukaryotes [17].

2. Writers of methylation: DNA methyltransferases (DNMTs)

Methylation of DNA has been known since its discovery as a genetic material in 1944. The family of DNMTs are enzymes capable of catalysing the transfer of a methyl group from the co-factor S-adenosylmethionine to the fifth carbon of C (Figure I-5).

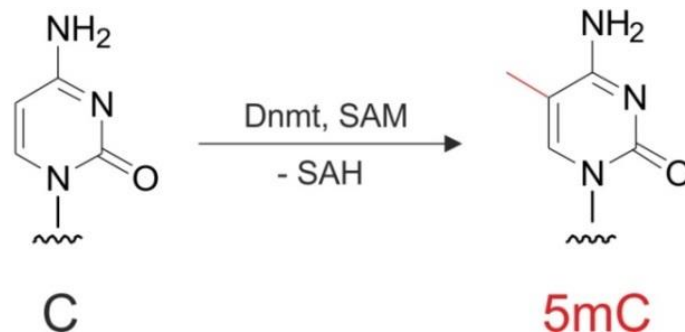


Figure I-5: Formation of 5mC. DNMT catalysed transfer of a methyl group from the cofactor SAM to the 5th position of the C nucleobase [21].

DNMT 1, 2, 3a, 3b and 3L are members reported to exist in mammalian genomes. Structurally DNMTs comprises of a N-terminal regulatory and a C-terminal catalytic domain [45]. Highly conserved regions within their structure carry out catalytic functions (Figure I-6). DNMT1 was the first discovered member and occurs most abundantly in somatic cells

[46]. This enzyme preferentially methylate hemi-methylated DNA [47,48] and is localized via several independent domains to the replication fork where it copies methylation patterns to newly synthesized non-methylated strands after DNA replication [49]. DNMT1 propagates methylation pattern to the next generation and is referred to as the maintenance methyltransferase [50]. In contrast DNMT3a and DNMT3b are known as the de-novo methyltransferase owing to their capability to methylate both native and synthetic DNA and the lack of preference for hemi-methylated DNA [51]. The DNMT3 members lack the nuclear localization signal present in DNMT1 but contains cysteine rich plant homeodomain (PHD) [52]. DNMT3a is produced ubiquitously while DNMT3b is seldom found in differentiated cells with exceptions to thyroid, testes and bone marrow [53]. Knockout studies in mice indicated critical role of DNMT3b in early development and that of DNMT3a in normal cell differentiation. A specific function of DNMT3b is maintenance of methylation of minor satellite repeat adjacent to centromere which is evident by the finding of mutation in their catalytic domain in patients suffering from Immunodeficiency Centromeric instability, Facial anomalies (ICF) syndrome [54,55]. DNMT3a may be involved in methylation of non-CpG Cs [48]. Another member of this family is the DNMT3L which lacks the catalytic domain is expressed mainly during early development and is restricted to germ cells in adults [56,57]. It is hypothesized to assist DNMT3a and DNMT3b methylation and has implicated functions in paternal and maternal imprinting, methylation of re-transposons and compaction of X chromosome [58–62]. The structure of DNMT2 consists of all conserved motifs but exhibits no methyltransferase activity *in vitro* and knocking them out showed no defects in cellular methylation patterns (Figure I-6) [63].

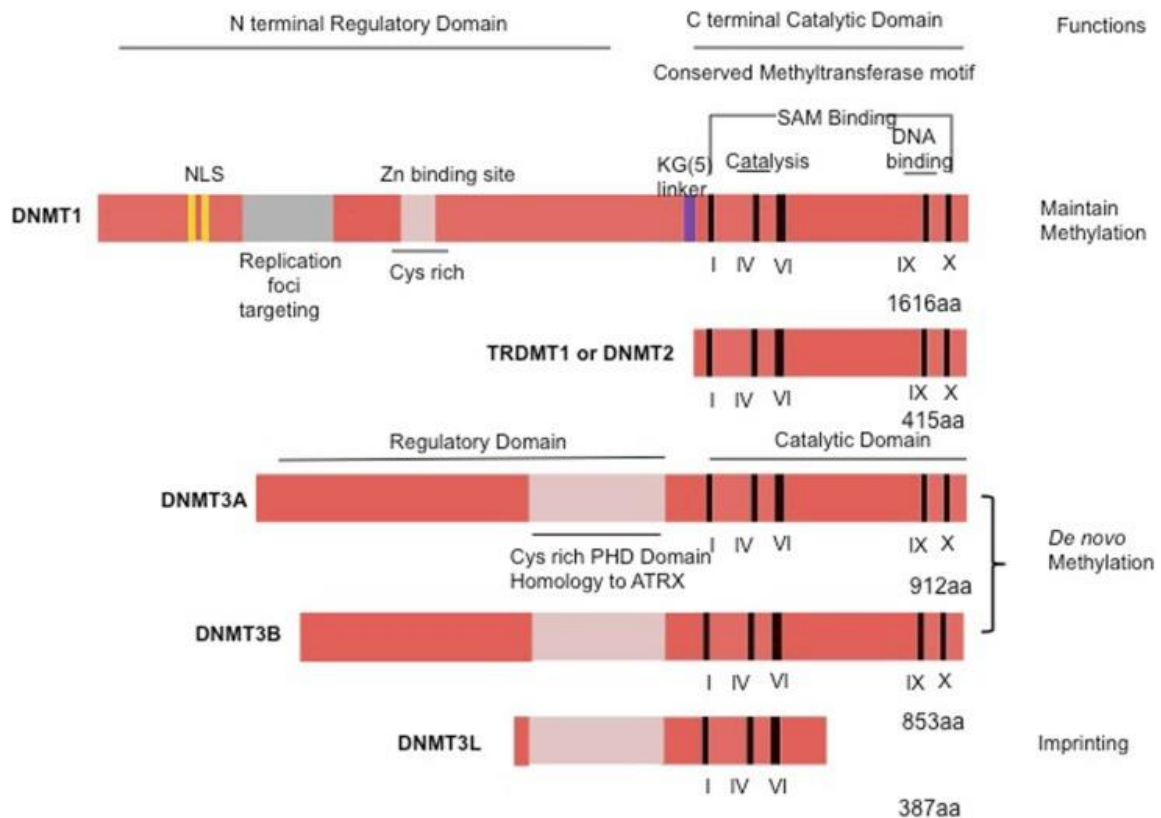


Figure I-6: Eukaryotic DNMTs: Model depicting the features and functions of different DNMTs found in Eukaryotes [64].

CpG islands (CGI) as the name suggests are stretches of DNA sequence with a higher than average CG sequence content. Amongst other characteristics of CpG islands their occurrence in majority gene promoters and highly conserved sequences are some. Lower nucleosome density and less intact histone association due to the presence of CGIs in promoters ensures exposure to transcription factor and hence gene expression. CGIs are generically equipped to influence local chromatin structure and simplify regulation of gene activity. Depending on the location of the CpG, their hypermethylation can either repress or promote gene expression [65–67]. 70% to 80% of the mammalian CpGs are methylated with the non-methylated CpGs being a part of promoter of active genes [68]. Other related and recently discovered region are sequences occurring within approximately 2 kb of the CGIs called CpG shores. Methylation in these regions have been connected to tissue and cancer related methylation as well as age related hypomethylation [69,70].

Methylation of C in mammals is predominant primarily in CpG dyads but they exist additionally in CpA, CpT and CpC contexts. The relevance of non-CpG site methylation is yet inconclusive but they may either be by products of non-specific de novo methylation or play a role in tissue specific gene expression [71–73]. Although rare in most differentiated cells, they are comparatively enriched in ES cells, oocytes, neurons and glial cells [74,75].

3. Readers of methylation: Interplay of different epigenetic mechanisms for transcriptional silencing

The relationship between histone DNA modification and transcription regulation is complex and rather incompletely understood. Hypermethylated DNA packaged into a chromatin structure are resistant to nucleases and enriched in hypoacetylated core histones providing a direct link between DNA methylation and transcriptional silencing [76,77]. C methylation of a gene sequence may interfere with transcription factor binding however most transcription factors do not have a CpG dinucleotide within their binding sites. Further, cases of global silencing of large domains or whole chromosomes like in the case of X chromosome inactivation suggest the interplay of other mechanisms. Methylated DNA binding proteins consists of a methyl binding domain (MBD) and a transcriptional repression domain (TRD). MBD binds to methylated CpG on the major groove of the DNA and the TRD is capable of contacting several regulatory protein and block the access of transcriptional factor binding [78,79]. In the family of methyl CpG binding proteins (MBD's) that consists of MBD1, 2, 3 and 4, MeCP2 was the earliest to be identified (Figure I-7) [80]. All the MBD have been linked to Histone deacetylase (HDAC) for chromatin remodelling [81,82]. HDAC causes a local loss of acetylation of core histone tails which result in tighter packaging of DNA and hence reduced access of transcription factors [83]. Additionally MeCP2 can bind to DNMT1 and recruits it to hemi-methylated DNA [84]. Interestingly, DNMT1 can interact with HDACs however the relevance of the combination is unclear. Other families recognizing 5mC are Ubiquitin-like, containing PHD and RING finger domains, 1 (UHRF) proteins and zinc finger proteins (ZFP). Overall, DNA methylation and histone modifications work together to regulate gene expression. Another epigenetic mediator with probable intervening functions is mature miRNA-induced silencing complex (miRISC) that can inhibit expression by binding and degradation of mRNA [85]. The presence and interactions of so many interdependent functional entities

indicate the existence of very complicated and well-orchestrated regulation machinery which is yet not completely decoded.

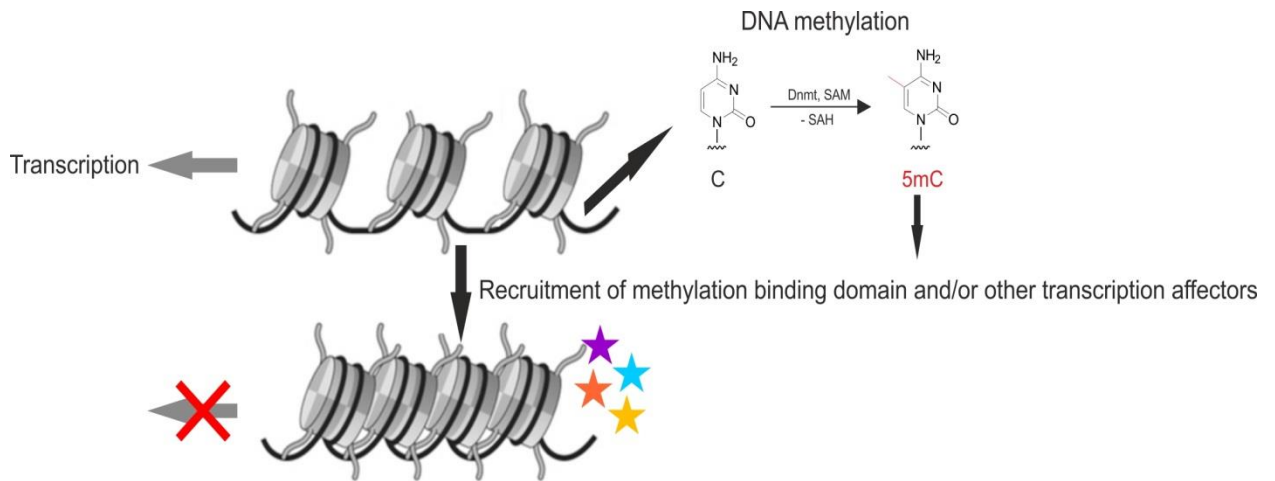


Figure I-7: Link between DNA methylation and transcription: Cartoon depicting the interdependent nature of DNA methylation and other epigenetic effectors causing disruption of gene transcription [14].

4. Erasers of methylation

5mC is both chemically and genetically stable and its formation involves addition of a methyl group on the 5th carbon position of C via a strong covalent carbon-carbon bond making its direct removal energetically unfavourable. There is no known mechanism for direct removal of the methyl group from C. Demethylation process can progress through passive or active pathway. Occurring in the dividing cells, passive demethylation involves dilution of methylation content due dysfunction or absence of maintenance methylation by DNMT1 inhibiting methylation of the newly generated strand after each DNA replication cycle [86–88]. Active demethylation pathway is more complicated, occurs in both dividing and non-dividing cells and includes involvement of enzymatic reactions to catalyse conversion of 5mC to C. The active pathway for removal of methylation can follow either of the two mechanism both concluding in excision by the Base Excision Repair (BER) machinery. The first mechanism includes converting the amine group on C to carbonyl by Activation Induced cytosine Deaminase/

Apolipoprotein B mRNA editing Enzyme Complex (AID/APOBEC) and thus transforming the base C to T. This causes a G/T mismatch at the site which is recognized and repaired by the BER machinery [89]. Although it reduces demethylation, knockout of APOBEC does not prove to be lethal in mice suggesting an alternate active demethylation process [90–92]. The other hypothesized pathway includes a more elaborate oxidation of 5mC by the family of Ten Eleven Translocation dioxygenases (TET) prior its replacement. TET enzymes are α ketoglutarate and Ferrous (II) dependent enzymes composed of double stranded beta helix (DSBH) and a cysteine rich domain. The DSBH brings the Ferrous (II), α ketoglutarate and 5mC group together while the cysteine rich region domain wraps around the DSBH domain to stabilize the overall structure of TET-DNA interaction [93]. This contact does not involve a methyl group which allows accommodation of different C modifications evident by its activity on 5hmC and 5fC [94]. The C terminal of the enzyme alone can localize to the nucleus and oxidize 5mC [95–97]. 5mC is oxidized to 5hmC, which can either act as a substrate for further TET oxidation forming 5fC and 5caC [98] or alternatively, its deamination by AID/APOBEC results in formation of hmU creating a base-pairing error [99]. Thymine DNA glycosylase (TDG) can recognize and cleave hmU, 5fC and 5caC creating abasic sites that are repaired by BER machinery (Figure I-8). One other enzyme belonging to the same family as TET, known as Single strand selective monofunctional uracil DNA glycosylase 1 (SMUG1) may also be involved in active demethylation [100–102].

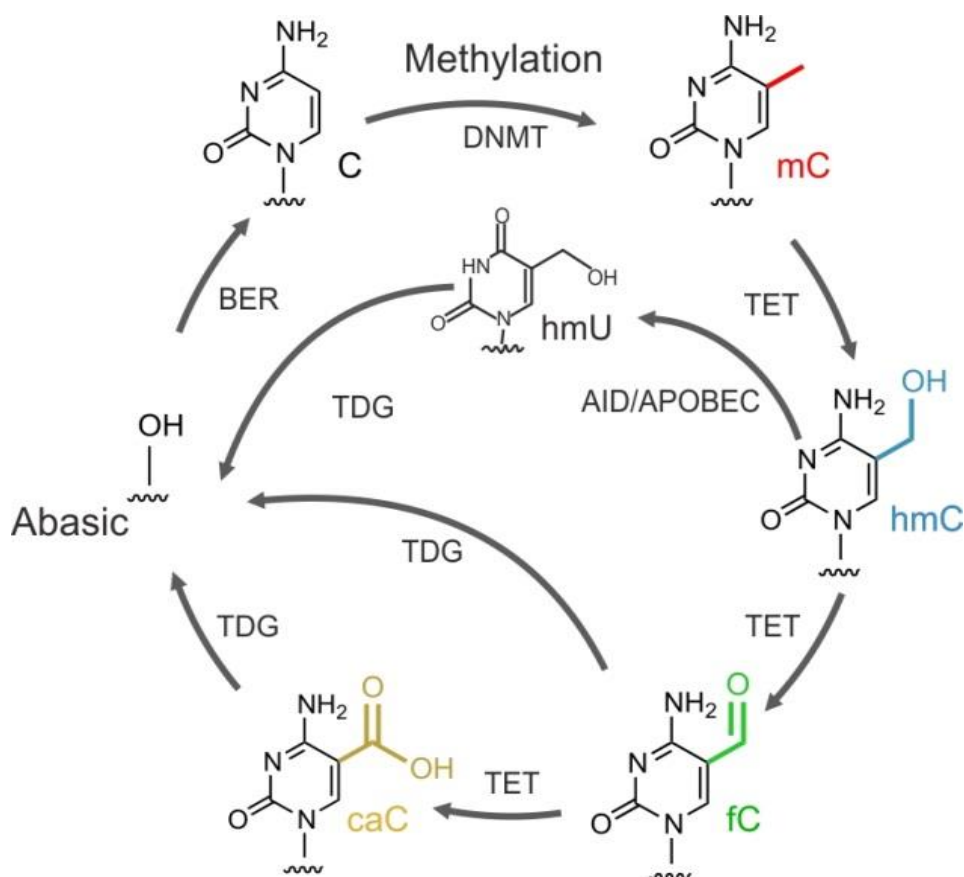


Figure I-8: The active pathway for C demethylation. All the intermediate products and enzymes are mentioned.

5. Occurrence of C variants in cells

Although chemical substitutions on C does not affect DNA base pairing with G, there is change in the steric demand and they provide unique chemical information in the major groove of the DNA (Figure I-9). The size of the chemical group plays a role. A variation in conformational flexibilities has been observed as in the case of 5hmC and 5fC. 5caC bears the disadvantage of harbouring the bulkiest group and being the less studied so far. Compared to 5mC, these oxidation products occur less frequently and their exact roles remain unclear.

All C variants occur at varying abundance depending on the cell type and conditions, with the general trend of occurrence being $5mC > 5hmC > 5fC > 5caC$. For example, a rough estimation suggests that 5hmC marks are as low as 5% of total C in cells of heart, breast and

placenta and as high as upto 65% in that of brain, liver and kidney. In general, it is also less abundant in cancerous cells. Along with fact that 5hmC is not a substrate for TDG, its high presence in promoters of transcriptionally active genes and its stability in DNA places a very compelling argument of it having a distinct regulatory function [72]. Some studies have also revealed unique protein readers specific to 5hmC and 5fC further strengthening their functional significance to more than demethylation intermediates [103]. Techniques for selective and individual detection of each of these variants preferably with sequence specificity and high efficiency is the need of the hour.

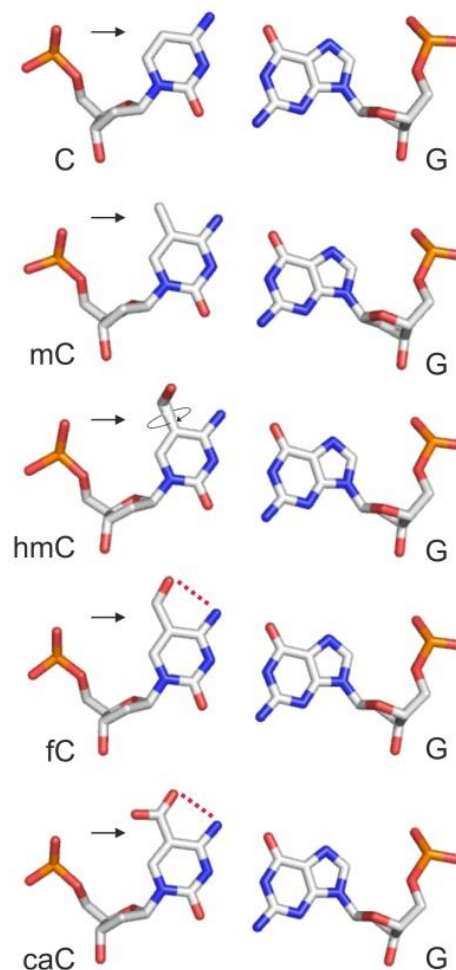


Figure I-9: DNA base pairing between nucleobase G and epigenetic variants of C nucleobase [16].

6. 4mC: An epigenetic C nucleobase found in prokaryotes.

In addition to the well-known 5mC, another epigenetic C base exists in the bacterial genome carrying a methyl group at the fourth amine group of C (4mC) (Figure I-10). This base has been reported to exist only in mitochondrial RNA of some eukaryotes [104]. Like the other epigenetic modifications found in bacteria: 5mC and 6mA (methylation at the 6th amino group of adenine) 4mC is also thought to be a part of the restriction modification system which plays a role in host defence and transcription regulation [18] [105]. These modifications render the bacterial genome safe from restriction by endonucleases which are sensitive to modification on their recognition site. 4mC occurs at significantly lower levels than 5mC and is catalysed by the enzyme S-adenocyl-L-methionine: DNA cytosine N4-methyltransferase. Unlike methylation at the fifth position, *in vitro* methylation at the N4 amino group of C does not occur homogeneously on all CG sequences but is rather sequence specific. For example, BamHI methyltransferase is used to methylate the fourth C of the sequence: 5'GGATCC 3'. Investigating the genomes of various mesophilic bacteria indicated the presence of 4mC. Given that most of the restriction endonucleases come from these bacteria and some of them are not affected by 5mC or 6mA, the base 4mC seems to be a sole member of another restriction modification system [106]. It is of interest to understand the individual roles of the two methylated C nucleobases in the genome of prokaryotes.

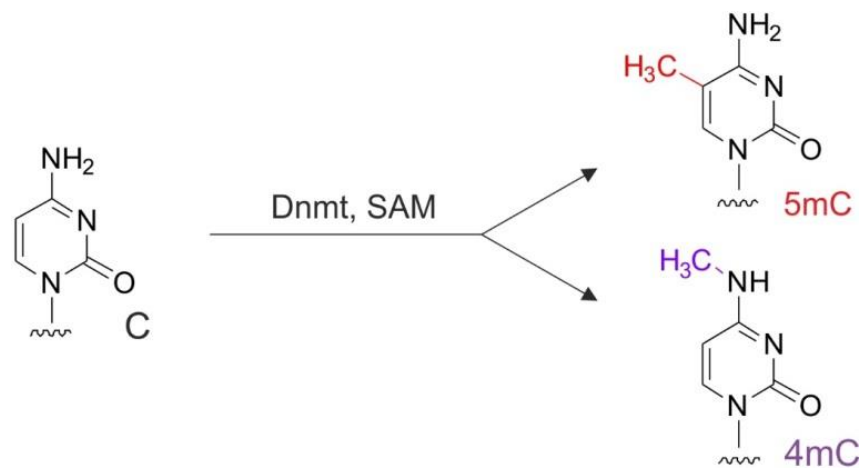


Figure I-10: Methylated C found in prokaryotes. DNMTs are responsible for transfer of a methyl group from the cofactor SAM to either the 4th amine or the 5th carbon position of C [19].

D. DNA methylation in oncology

Cancer encloses a group of diseases marked by uncontrolled, abnormal growth of cells in any part of the body capable of spreading to and invading other parts. Viewed earlier solely as a genetic disease, epigenetic alterations are now being considered as the responsible member for their regulation. Study involving Restriction Landmark Genomic Scanning (RLGS) approach for analysis of 1184 random CGIs of individual tumour types and patients have established aberrant DNA methylation as the hallmark of cancer type [107]. DNA methylation maps (methylomes) of cancer genomes revealed hypermethylation in CGIs of promoters of tumour suppressor genes as well as those involved in cell-cell adhesion and DNA repair [108]. Some cases exhibit loss of one copy of suppressor gene by mutation and transcriptional silencing of the other WT copy (Figure I-11) [109]. Hence, information about DNA methylation serves not only as a diagnostic or prognosis biomarker but also provides insights into the treatment response and so can be put to therapeutic use. These examples may help in elaborating this theory [110].

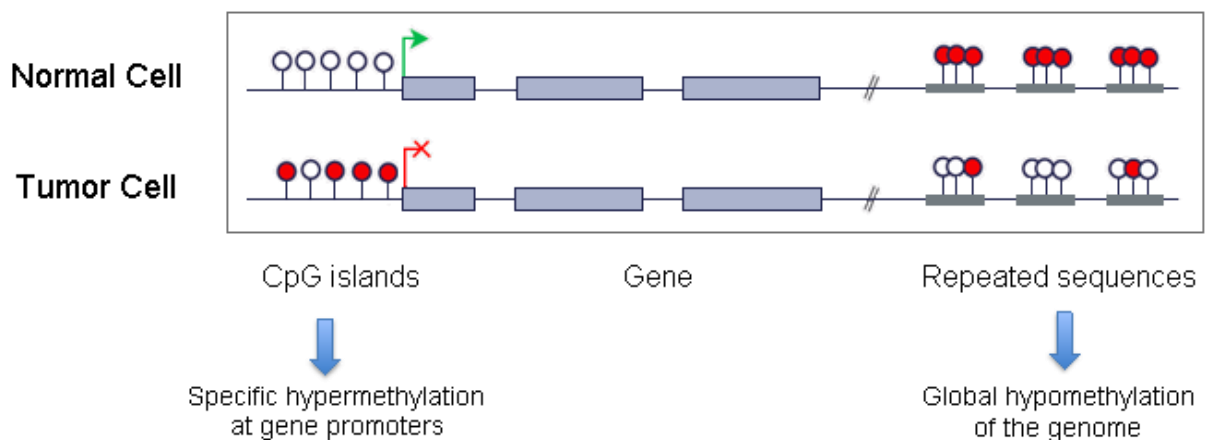


Figure I-11: Cartoon depicting the difference in methylation pattern in normal and tumor cells. Red and white beads indicate methylated and non-methylated CpGs respectively. Green and red arrow indicates transcription activation and inhibition respectively [1].

1) BRCA1 (Breast Cancer gene 1) located in chromosome 17, responsible for production of proteins that play a key role in DNA single strand break repair, is rendered non-functional by hypermethylation of CpG near the transcription start site. This effect is observed in hereditary breast cancer but its mutation is not detected in sporadic breast cancers. Here, DNA methylation could act as a diagnosis biomarker as well as a factor determining the sub type of cancer [111,112].

2) CDKN2A or Cyclin-Dependent Kinase inhibitor 2A gene codes for proteins responsible for tumour suppression. Loss of function due to promoter hypermethylation is associated with familial melanoma, glioblastoma and pancreatic cancers. Location and type of mutation in the CDKN2A gene are deciding factors of the type and intensity of the disease [113].

3) Transcriptional silencing of the O⁶ MethylGuanine DNA MethylTransferases (MGMT) gene causes accumulation of G to A transition and hence mutations in the K-ras and p53 tumour suppressor genes. Demethylation of histone H3 lysine and binding of methyl-CpG binding proteins are observed in MGMT silenced cells. Although MGMT silencing is a poor prognostic factor, it plays a vital role as predictive marker of alkylating agent based chemotherapy [114].

4) Chronic Lymphocytic Leukaemia (CLL) is the most prevalent leukaemia in adults and occurs in two major forms that differ mainly in the average survival rate of patients. Expression of zeta chain-associated protein (ZAP-70) is a crucial factor in determining the type of CLL and predicts remission following chemoimmunotherapy. Methylation at a single CpG present 223 bases downstream of the transcription start site highly regulates gene expression.

The association of mutation in the epigenetic machinery and cancer is now vastly accepted. Methylome studies continue to provide interesting insights into mechanisms of cancer development. However, to fully understand its interplay, and gain further in-depth knowledge, tools for precisely detecting methylation in genomes is highly desirable [115,116].

E. Techniques for detection of epigenetic nucleobases

Since the discovery and characterization of epigenetic modifications, a lot of research has been focused on understanding their role in cellular mechanism. Studies in this field have unravelled the crucial, regulatory roles of certain modifications and it continues to add previously unknown modifications to the list of epigenetic marks. Given the enormity of types of organisms and their different genetic make-up, the possibility of finding novel modifications or precise function of existing ones are very high. The past decade has yielded the proof of existence of several non-canonical C bases and spurred the requirement for tools discriminating them not only from their canonical counterpart but additionally from each other. This means that selective detection of C, 5mC, 5hmC, 5fC, 5caC and 4mC is of very high interest. Although with some shortcomings, previously established methods have proven to be indispensable to detect, measure, and elucidate C derivatives techniques [117].

1. Methods involving chemical modification of the target

Established as one of the earliest techniques for detection of methylated C, sodium bisulphite mediated selective deamination of C but not 5mC to Uracil (U) and subsequent sequencing (*BS-seq*) continues to be by far the most frequently and widely used method [118]. Though reaction with hmC leads to the formation of Cytosine-5-methylsulfonate (CMS) (no deamination to U), both 5fC and 5caC are deaminated to U when heated with sodium bisulphite. Therefore, sequencing following deamination of C 5fC and 5caC yields T while 5mC and 5hmC are read as C. Since being used for the first time, this method has been combined with additional prior chemical conversions or tagging reactions to facilitate parallel detection of all C variants. For detection of 5hmC, TET assisted BS-seq (*TAB-seq*) can be employed. Samples are first treated with T4 β -glucosyltransferase (T4 BGT) for selective glycosylation and hence protection of only 5hmC molecules followed by overall TET 1 based oxidation that converts all non-protected C, 5mC and 5fC members to 5caC. Standard BS-seq reads all C variants except 5hmC as T. Comparison with untreated samples can highlight positions of 5hmC in the sample [119–121]. An extension of this technique has been used for detection of 4mC (*4mC-TAB-seq*). 4mC is a rare epigenetic base found in some bacteria to protect the recognition system of the host. Given the dominance of 5mC in genomes of many organisms, BS-seq is not suitable for accurate mapping of 4mC/5mC in a

population containing both modifications. Conducting and comparing BS-seq and TAB-BS-seq simultaneously however, reveals 4mC containing positions which are unaffected by TET 1 dioxygenases mediated oxidation and relatively tolerant towards sodium bisulphite and hence always read as a mix of C and T both before and after the treatment while C containing samples are oxidized to 5caC and are read as T after treatment [122]. BS-seq following oxidation (*oxBS-seq*) using potassium perrunthenate (K₂Cr₂O₇) can reliably be used for detecting 5hmC in a sample containing both 5mC and 5hmC. While 5hmC gets oxidized to 5fC, 5mC remains unaltered providing different readouts in sequencing [123,124]. Further, to detect 5fC sodium borohydride (NaBH₄) mediated reduction of 5fC to 5hmC before BS-seq (*redBS-seq*) is highly preferred [125]. Detection of fC is additionally possible by using 5fC Assisted Bisulphite sequencing (*fCAB-seq*) wherein, the formyl group of 5fC is first labelled using O-ethylhydroxylamine (EtONH₂), which imparts resistance to deamination by sodium bisulfite. The difference between the number of C readout from BS-Seq and fCABSeq can distinguish the position and number of 5fC molecules present [126]. A similar strategy exists for detection of 5caC, known as the Chemical-modification Assisted Bisulphite sequencing (*CAB-seq*). 5caC can be converted to an amide by reaction with 1-ethyl-3-[3-(dimethylamino) propyl]-carbodiimide hydrochloride (EDC). Subsequent treatment with bisulphite does not cause deamination of the pre-treated 5caC base thus providing different sequencing read-out between EDC treated and not treated 5caC [127]. The summary of different BS-seq is shown below (Figure I-12).

Original base	Treatment	Resulting base	Method
C, mC, hmC, fC, caC ↓ ↓ ↓ ↓ ↓ T C C T T	DNMT mediated methylation	mC, mC, hmC, fC, caC ↓ ↓ ↓ ↓ ↓ C C C T T	BS
C, mC, hmC, fC, caC ↓ ↓ ↓ ↓ ↓ T C C T T	selective oxidation of hmC	C, hmC, fC, fC, caC ↓ ↓ ↓ ↓ ↓ C C T T T	oxBS-seq
C, mC, hmC, fC, caC ↓ ↓ ↓ ↓ ↓ T C C T T	βGT mediated glycosylation and subsequent oxidation	C, hmC, ghmC, caC, caC ↓ ↓ ↓ ↓ ↓ T C C T T	TAB-seq
C, mC, hmC, fC, caC ↓ ↓ ↓ ↓ ↓ T C C T T	NaBH ₄ mediated reduction	C, mC, hmC, hmC, caC ↓ ↓ ↓ ↓ ↓ T C C T T	redBS-seq
C, mC, hmC, fC, caC ↓ ↓ ↓ ↓ ↓ T C C T T	EtONH ₂ mediated fC labelling	C, mC, hmC, fC, caC ↓ ↓ ↓ ↓ ↓ T C C C T	fCAB-seq
C, mC, hmC, fC, caC ↓ ↓ ↓ ↓ ↓ T C C T T	EDC mediated amide conversion of caC	C, mC, hmC, fC, caC ↓ ↓ ↓ ↓ ↓ T C C T C	CAB-seq
mC, 4mC ↓ ↓ C C/T	TET mediated oxidation	caC, 4mC ↓ ↓ T C/T	4mC-TAB-seq

↓ = Sequencing following Sodium Bisulphite treatment

Figure I-12: Table summarising different forms of Bisulphite sequencing techniques. C, 5mC, 5hmC, 5fC, 5caC and 4mC are mentioned in black, red, blue, green, yellow and purple respectively.

The simplicity of design and easy availability of the reagents marks the advantages of these techniques; however, often the DNA is degraded to an extent that makes its subsequent use in PCR or sequencing troublesome. On the contrary incomplete conversions have often been observed leading to unreliable results. High amount of T nucleobases in the DNA reduces its complexity and makes their downstream processing extremely tedious. Further achieving single nucleotide resolution for large samples can be very expensive.

2. Methods involving targeting the modification

Immunoprecipitation assay

One of the simplest and most common approaches for large scale enrichment of methylated DNA sequences is using Methylated DNA Immunoprecipitation (MeDIP). Established first by Weber *et. al.*, in 2005, this method puts to use a monoclonal antibody that specifically recognizes 5mC and enriches DNA sequences containing them [128]. Pre-isolation and sonication of DNA is recommended as the size of DNA fragments affect the overall enrichment efficiency (Figure I-13). The simplicity in the assay principle has encouraged many companies to design products for commercial use. Rabbit polyclonal antibodies for enrichment of 5hmC have been established and use of anti-CMS antibodies for CMS adduct formed upon BS treatment of 5hmC is also known [129]. Another protein capable of specifically recognising the product of T4 BGT mediated glycosylation of 5hmC is the J binding protein 1 (JBP1) used an alternate for detection of genome wide 5hmC content [130]. Antibodies for the other two C5 variants have also been raised. This assay can be coupled with Next Generation Sequencing (NGS) for detailed analysis of the fetched DNA. Following the discovery, naturally occurring MBDs have been utilized to enrich DNA sequences with methylated CGIs. MeCP2 binding is restricted to only the CpG dyad, but has shown to provide better coverage. Some MBD are selective 5mC binders while there are others that bind also to 5hmC [131]. This method provides one of the simplest and most cost-effective approaches for genome wide detection of epigenetic bases and can be utilized for detection of modification inside cells using immunostaining. However, being sequence and strand non-specific it fails to provide the sensitivity of the aforementioned chemical modification techniques.

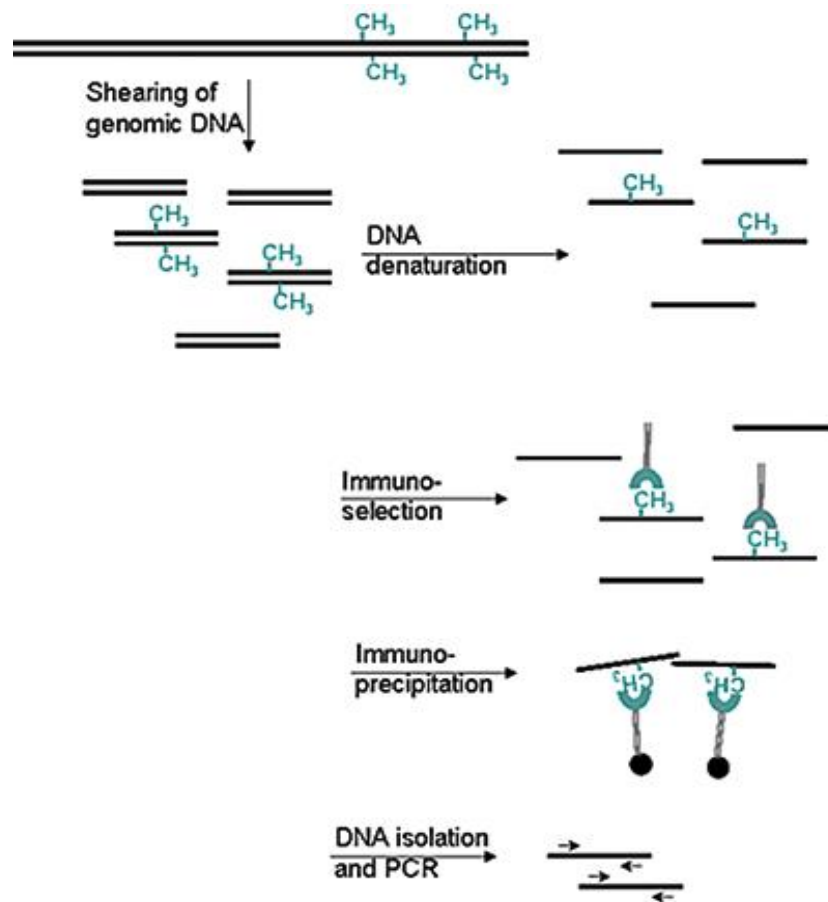


Figure I-13: Schematic representation of the MeDIP technique. The methylation specific monoclonal antibodies are shown to bind to methylated C on both the strands (Diagenode).

3. Methods involving sequence specific DNA Binding Domain (DBD)

Zinc Finger Proteins (ZFP)

ZFP belongs to a very diverse family of proteins with wide variety of functions. They are amongst the most abundant proteins in eukaryotes and are involved in several regulatory functions in the organism [132]. A general requirement for all the members of this massive family of small functional domains is the presence of at least one zinc ion for their structural coordination. Finger indicates the secondary structure formed by α -helix and β -sheet held together by the Zn ion. The domain that contains the Zn finger also acts as the interacting region capable of binding DNA, RNA, proteins or small molecules [133]. They were first observed as repeating domain in the transcription factor of *Xenopus laevis*. The protein uses

a combination of cysteine and histidine residues to coordinate the zinc ion. The method used to classify the ZFP is based on the overall shape of the protein backbone. Cys₂His₂ consisting of nine repeats of thirty amino acids is the most common domain [134]. Each domain forms a left handed ββ α secondary structure and coordinates a Zn ion between two cysteines on the β-sheet hairpin and two histidines in the α-helix. The residues on α-helices make specific contact with the major groove of DNA. Cys₂His₂ that bind DNA tend to have 2-4 tandem domains as part of the protein. It recognizes specific DNA triplet depending on the residues on their α-helix [135,136]. Methylated DNA precipitation combined Luciferase-fused Zinc finger Assay (MELZA) uses MBD and luciferase-fused ZFP to first isolate and amplify methylated DNA and detecting the quantity of specific sequence using luciferase fused to ZF is an example elaborating the application of ZFP for detection of epigenetic marks (Figure I-14) [137].

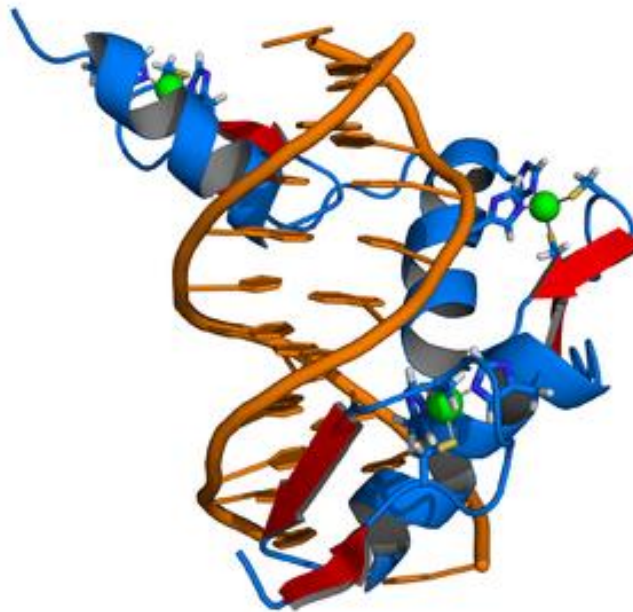


Figure I-14: A model of Zinc finger protein Zif268 binding to its DNA site. DNA is represented in orange, alpha helices in blue, beta sheets in red and the zinc molecule in green. The 2 cysteine and 2 histidine interaction with the zinc ion are shown as purple lines [138].

CRISPR/Cas 9

Genome engineering using Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas (CRISPR associated protein) was awarded the Science breakthrough of the year 2015. They are naturally occurring segments of prokaryotic DNA, belonging to the type II immune system, providing a defence mechanism to a wide range of bacteria [139]. The two main components of this system are guide RNA (gRNA) and Cas9 nuclease. The gRNA which is a short synthetic sequence hosts 1) a scaffold sequence that is necessary for Cas9 binding, 2) a user defined spacer/targeting sequence that inherits the genomic sequence to be targeted [140]. The genomic target is usually 18-25 nucleotides long and is placed upstream to PAM (Protospacer Adjacent Motif). The PAM is an absolute necessity for binding of gRNA and its sequence varies depending on the Cas9 (For example for a Cas9 derived from *Streptococcus Pyrogens*, which is the most commonly used Cas9 for genome studies, the PAM sequence is 5' NGG 3') [141]. Upon expression, the Cas9 binds with gRNA forming a riboprotein complex the changes its conformation and can bind to any PAM sequence but cleavage by Cas9 is executed only when sufficient homology with the target sequence is met [142]. The nuclease can be rendered partially or completely inactive due to mutations in either one or both the two functional endonuclease domains (RuvC and HNH) of Cas9. The partially inactive nuclease is called Cas9 nickase and generates only single stranded break or "nick" The deactivated/dead nuclease is termed as dCas9. The effect of dCas9 on the gene function is reversible as it does not permanently modify the gDNA [143,144]. As a tool for epigenome editing, dCas9 fused to the catalytic Histone AcetylTransferase (HAT) core of p300 was shown to be highly effective in transactivation of genes from proximal and distal enhancers (Figure I-15) [145].

Both ZFP and CRISPR/Cas systems are capable of binding DNA in a sequence specific manner and when coupled with appropriate effector molecules, they can identify, edit or extract specific genomic sequences or manipulate their transcriptional status. A lot of studies have successfully combined DBD with transcriptional activator domains such as VP64 (a tetramer of VP16, which in turn is a viral activation domain and recruits Pol II transcriptional machinery) or p65 (a subunit of the human NF- κ B) and silencing domains such as KRAB (KRüppel-Associated Box, a naturally occurring motif in mammalian zinc finger transcription factors) for gene transcription regulation. Some systems such as the SunTag (co-expression of epitope-tagged dCas9 and antibody-activator effector proteins) or VPR (several different activator domains in series) are often used for greater activation.

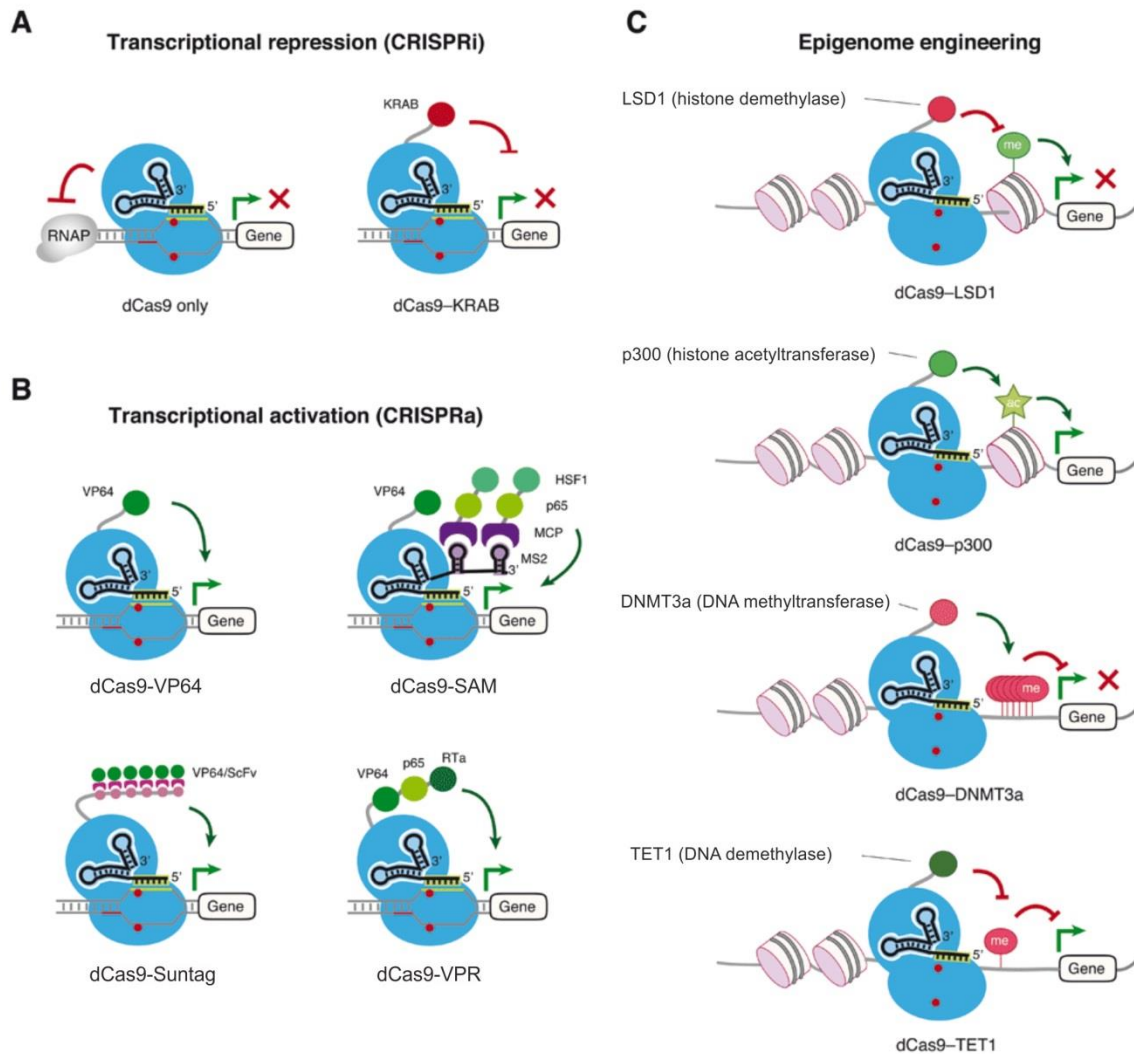


Figure I-15: Different ways of employing Crispr-dCas9 machinery for epigenome engineering [146].

The most commonly used methods have been listed, however other tools for identifying C variants are also present. Most of them rely on advanced sequencing techniques and are usually aimed for detection of C modifications. Use of Maxam and Gilbert sequencing, MRE (methylated Restriction Endonuclease), Teethered Oligonucleotide-Primed sequencing (TOP), Nanopore sequencing and Single Molecule Real Time (SMRT) have additionally been employed.

F. Transcription Activator Like (TAL) Effectors (TALEs)

1. Origin

TALEs are a family of DNA binding proteins found pre-dominantly in the gram negative bacterial species of *Xanthomonas* (Figure I-16). These species are known plant pathogens, causing virulence in more than 200 different plant families. The infection site on plants varies among the various strains but the transmission mode is in general facilitated by the Hrp-type III secretion system (T3S) [147–149]. A hollow conduit called the Hrp-pilus extends from the bacterial membrane to the plant cell wall. The effector protein secreted by the T3S is translocated into the cell cytoplasm via a bacterial translocon complex that inserts into the plant plasma membrane. The *avrBs3* isolated from *Xanthomonas Campestris* (*Xcv*) was the first type III effector protein found [150]. It's a 122-kDa protein (1164 a.a.) [151].

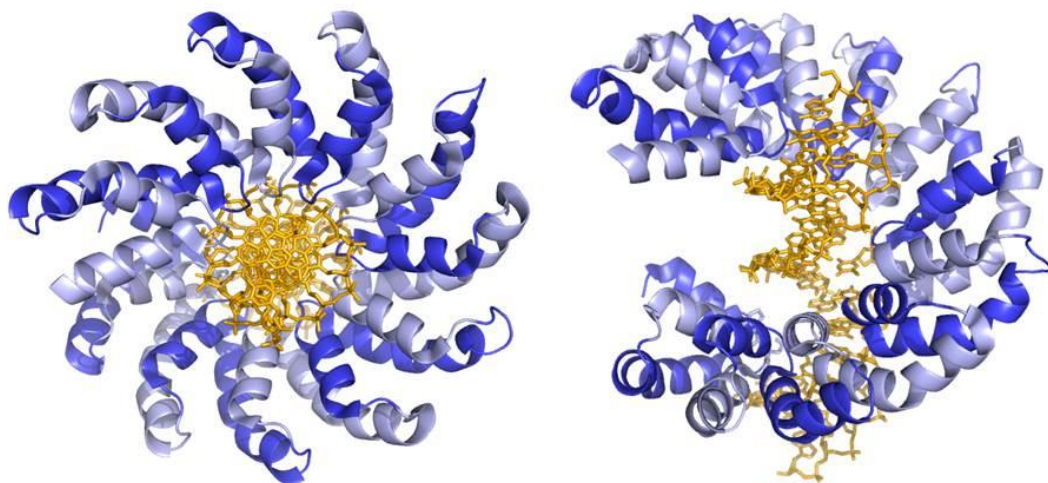


Figure I-16: Crystal structure of TAL PthXo1 binding to its target DNA (Pdb entry: 3UGM)

2. Structure

Structural insights of TALEs revealed a 5' N-terminal region (NTR) that contains the type III translocation signal. The central part is called the Central Repeat Domain (CRD) and is composed of blocks of about 33-35 amino acid sequences termed as the repeats. Each repeat targets one nucleotide on the ds DNA and apart from the beginning (zero repeat and before), the sequence of each repeat is highly repetitive. The number of repeats found in naturally

occurring TALEs range from 1.5 to 33.5. Within the set of mostly conserved amino acids, the duo existing at the 12th and the 13th position, called the Repeat di Variable Residue (RVD) vary between each repeat depending on the nucleobase targeted. A Nuclear Localisation Signal (NLS) and an acidic transcriptional Activation domain (AD) constitute the 3'/C terminal region (CTR) [152]. The crystal structure of a TALE PthXo1 (23.5 repeat) bound to a 36 bp long DNA duplex revealed that this protein wrap helically around the DNA encompassing it and making direct contact to its major groove (Figure I-16). Each repeat forms two helices using amino acids from 3 to 11 and 15 to 33. The RVD is present in a loop between these helices. More than one pair of RVDs has been found targeting each DNA base in nature, however for TALEs expressed for experimental use, the most abundantly used RVDs are NN, NI, NG, HD to target the nucleotides G/A, A, T and C respectively.

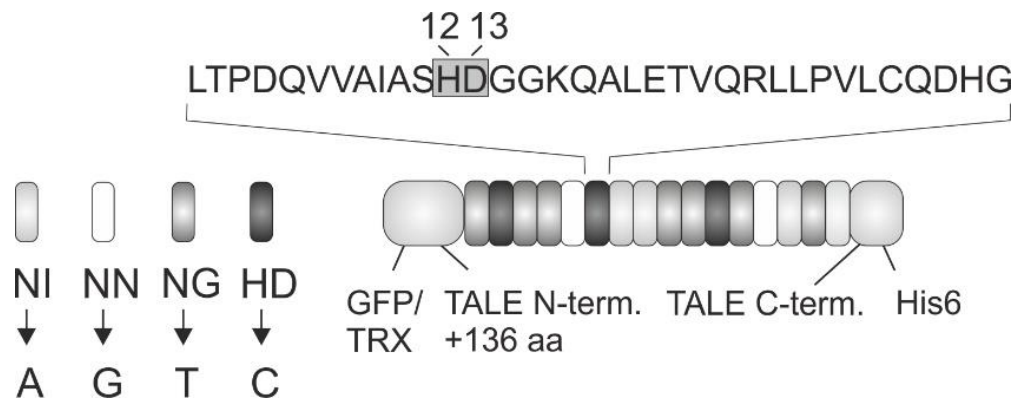


Figure I-17: Model depicting various features of all the TALE proteins used in this study [19]

3. Function

The presence of a NLS and the AD in the C terminal region was intriguing as these were typical eukaryotic motifs. The protein did not lose its function when the whole NLS was replaced by a heterologous NLS (from the SV40 large T-antigen) suggesting its role was driving the protein to the nucleus of a eukaryotic host [147]. Inactivation of AD rendered the protein inactive and function could be partially restored by replacing the inactivated AD by heterologous AD called VP16 (Herpes simplex virus) [153]. This suggested that this protein was transported to the nucleus and was taking part in host transcription regulation.

4. Interactions between different RVDs and bases

The second amino acid in TALE RVD makes direct contact with the nucleobase while the first amino acid (His or Asn) forms a H-bond with carbonyl oxygen of the 8th amino acid in the same repeat. While the RVDs NN and NI are known to be relaxed in their binding specificities, recognition of a base by RVD NG and HD is rather specific [154]. The carboxylate oxygen of aspartate in RVD HD accepts a H-bond from the amine group of C and exhibits van der Waals interaction with the edges of the base (Figure I-18). A non-polar interaction forms between the α -carbon of glycine in NG and the methyl group of the T nucleobase. A similar interaction is observed when T is replaced by 5mC. Comparatively less sensitive recognition of both purines by RVD NN is facilitated by the N7 nitrogen of the bases. The aliphatic side chain of isoleucine in RVD NI can interact with either the carbon 8 position in adenine or the carbon 5 in C via van der Waals contact [155].

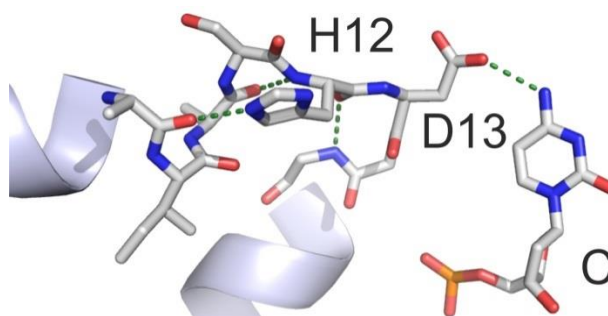


Figure I-18 Interaction of TALE RVD HD with nucleobase C as depicted in the crystal structure of TALE bound to DNA (pdb entry 3V6T). Green dashed line represents the Hydrogen bond between the carboxyl group of aspartate and the amine group of C [156].

5. Non-specific interactions between the TALE protein and DNA

The critical requirement of the nucleobase T 5' to the target sequence (-1) in a DNA is observed in most natural TALEs and recommended for the design of experimental ones. However, TALEs bearing mutation in the NTR prior to the CRD have been designed to recognize other DNA bases at position -1 [20,157]. N-terminal repeats, constituting the amino acid sequence before the CRD, folds similar to the first helices (amino acids 3 to 11)

formed by each repeat. This region converges very close to the start of the DNA. A highly conserved tryptophan (W232) is found on the loop between the helices formed by repeat -1 and 0 and its indole ring is known to extend a van der Waals interaction with the methyl group of T (T₋₁). A proline at 27th position (second helix) of every repeat was crucial for the overall structure and contacting the DNA. Additionally, non-specific contacts to the DNA are made by electrostatic positive charges on the glycine at the 14th, lysine at 16th and glutamine at 17th position of each repeat and the negative phosphate of DNA.

6. Mechanism of action

Many of the DNA binding proteins, search target DNA by employing micro-dissociation/re-association events termed as hopping [158]. They track the major groove while traversing through a DNA by rotating around it via a sliding mechanism [159]. The fundamental mechanism governing the search process for TALE proteins is not fully understood. The single repeat for single base behaviour distinguishes them from other proteins with similar functions.

Study using single molecule techniques suggests that TALEs remain in constant contact with its target while still being able to dissociate efficiently in case the ionic strength is broken. It utilizes a two-state model comprising of a search mode and a recognition mode. The NTR containing positively charged amino acids binds to DNA and exhibits no nucleotide specificity. It facilitates exploratory search via sliding over the DNA and the CRD is capable of halting and searching the DNA sequence and cause longer duration binding with varying frequency (hopping) which is influenced by the affinity of the protein. Another factor hugely affecting the electrostatic interaction and hence the frequency of dissociation is the ionic strength [160,161]. During transiting to a recognition mode with non-target DNA, the steric electrostatic clashes and destabilizes the complex. Upon encountering the correct target sequence, energetically favourable molecular contacts form between the CRD and the DNA base and the TALE undergoes a change in conformation to a more compact form [162].

7. Factors affecting TALE function

Since the successful use of TALE protein for genome engineering, numerous studies have been dedicated to identifying the plausible factors and the extent to which they may alter their binding efficiency and specificity [163]. Different RVDs bind to their respective DNA bases with varying strength where targeting T and C by RVDs NG and HD respectively are robust compared to G and A by NN and NI meaning that the composition of target sequence affects the strength and specificity with which a TALE protein binds to its target sequence. Each position and content contributes differently to the specificity of the protein [164]. The battle between binding strength and specificity has drawn a lot of attention to come up with an optimum size and composition for effective TALE targeting [165]. Some studies have also proposed that the neighbouring RVDs can influence their binding [166] while others indicate reducing the non-specific electrostatic interactions could improve the selective recognition of the TALEs. Interplay of various affecting component meant that there is no one rule for their optimal design, nevertheless these studies help in anticipating their function and provides insight for possible improvement in TALE design.

8. Applications

Naturally occurring TALEs are capable of controlling transcription in the host genome however their use as tools for genome engineering is more diversified.

One of the most popular use of TALEs was in combination with nucleases. TALENs are restriction enzymes employed to achieve restriction of DNA at specific locations. When used with DNA cleavage domain of FokI endonuclease, non-specific DNA cleavage can be achieved. FokI domain functions as a dimer that require two constructs with unique DNA binding domain for sites in the target genome with proper orientation and spacing. (Figure I-19). Such TALENs induces Double-Strand Break (DSB) which are repaired by either Non-Homologous End Joining (NHEJ) or homology directed repair. This creates either indels or when used with ds template can be used to introduce foreign DNA at the DSB.

More recent approaches include combining these proteins with other effector domains such as VP64, p65 or KRAB for transcription regulation. Recent research has also seen their use *in vivo* for epigenome engineering by fusing them with DNMT or TET domains to obtain

methylation or demethylation of sequences occurring in the target vicinity. In addition, coupling them with appropriate fluorophores and employing microscopy has been successfully utilised for tracking DNA inside cells.

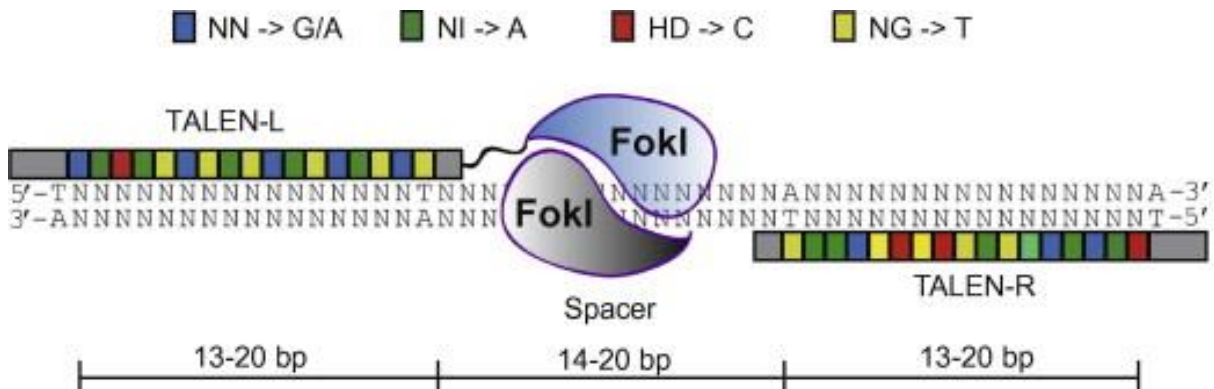


Figure I-19: Schematic representation of TALEN functioning. TALE domain fused with DNA cleavage domain of FokI nuclease causes sequence specific DSB [167].

In the presented study, another unique capability of TALEs to detect chemical modifications on the major groove DNA sequences has been exploited. As described earlier the strong bond between RVD HD and the base C can be perturbed due to presence of a chemical moiety making it a selective sensor of C (Figure I-18). Engineered TALE RVDs have displayed the potential to function as positive and selective binders for many known epigenetic nucleobases. Use of this property of TALEs for sequence specific detection of epigenetic marks on gDNA in an optimized assay could prove to be a strong tool in the field of genome engineering for gaining in-depth knowledge about the function and relevance of different epigenetic mechanisms.

II. AIM

The past decade saw an expanse in the span of epigenetic nucleobases identified in various organisms. To understand how different epigenetic modifications affect genomic processes, information on their location in the genome is important. It is also critical to identify and discriminate between different epigenetic modifications. The main objective of this project was detection of modified C nucleobases at single nucleotide resolution occurring on user-defined gDNA sequences using TALEs.

TALEs have been used to bind chosen DNA sequences [12–18]. TALE RVD recognizes a nucleobase in the major groove of a bound DNA molecule. Hence, they can interact directly with the epigenetic modifications occurring on the nucleobases. Earlier studies have conjointly identified and employed natural and engineered TALE RVDs as selective sensors for the canonical nucleobases as well as for different epigenetic modifications [15,16].

A new affinity enrichment assay was developed for detection of epigenetic modifications in gDNA samples. The rationale was to employ TALEs as programmable affinity probes for isolation of user defined DNA sequences from gDNA samples with selectivity for epigenetic C modification at a single nucleotide resolution. The affinity enrichment assay was first used for detection of 5mC in Zf gDNA and human cancer biomarker sequences. Additional experiments were designed for detection of hemi-methylation and quantitative analysis of 5mC levels in disease-relevant loci of the human genome. To further expand the potential of this method a TALE with size reduced repeat was used in the affinity enrichment assay for selective detection of C, 5mC, 5hmC, 5fC and 5caC and discrimination between 5hmC or 5fC containing sequences.

To obtain TALEs with enhanced epigenetic nucleobase selectivity we aimed at decreasing the non-selective binding energy of TALEs. We hypothesized that a more sensitive discrimination between C and 5mC could be achieved by increasing the relative difference in binding energy between TALEs and C or 5mC containing target sequences. To obtain this positively charged amino acids that interact non-specifically with negatively charged phosphate backbone of DNA were substituted with alanine. To verify improved activity, TALEs bearing alanine mutations were employed for detection of 5mC in target DNA sequences using affinity enrichment assay and for selective recognition and transcription activation *in vivo* using a luciferase assay.

Depending on the site of methylation two different methylated C bases i.e., 5mC and 4mC are known to co-exist in prokaryotic genome. Another part of the work was aimed at identification of a selective 4mC binder to achieve programmable decoding of all C nucleobases present in the bacterial genome. TALEs with size reduced RVDs were screened in a fluorescence based competitive assay to obtain a selective 4mC sensor which was then employed for detection of C, 5mC and 4mC in gDNA sequences using the affinity enrichment assay.

III. MATERIALS AND METHODS

A. Materials

1. Consumables

Item	Supplier
0.2 μ M syringe filter	Sarstedt
3 mm Whatman paper	Whatman
0.1-10 μ L Milipore Ziptips	Sigma
10 μ L, 200 μ L, 1 mL pipette tips	Sarstedt
10 μ L, 200 μ L, 1 mL pipette tips for enrichment assay	Sorenson
10 μ L, 200 μ L, 1 mL pipette tips for qPCR	Brand
200 μ L tubes	Sarstedt
1.5 mL, 2 mL tubes	Sarstedt
1.5 mL, 2 mL tubes for enrichment	Eppendorf
2 mL spin columns	Thermo Fischer
5 mL columns for protein purification	Qiagen
6 tube Magnetic racks	GE Healthcare
12-tube magnetic rack	Qiagen
10 mL, 25 mL serological pipettes	Sarstedt
15 mL, 50 mL falcon tubes	Sarstedt
96 deep well plates	Sarstedt
96 well plates, transparent for BCA	Carl Roth
96 well plates, transparent with lid for cell culture	Sarstedt
96 well plates, white, flat bottom for Luciferase assay	Sarstedt
384 well plate, black for FRET assay	Greiner Bio One
384 well plate, white for qPCR reactions	Bio-rad
Adhesive PCR seal for qPCR plates	Biozym
Amicon Ultra MWCO 3, 50, 100 kDa	Millipore
Cannulas Sterican 0.60 X 80 mm, 0.80 X 120 mm, 0.90 X 40 mm	Braun
Cuvettes semi-micro for OD600 measurement	VWR
Dynabeads His-Tag Isolation and pulldown	Thermo Fischer

Gas	Carl Roth
Glass bead for transformation	Carl Roth
Illustra microspin G-25 columns	GE Healthcare
Ni-NTA Magnetic agarose beads	Qiagen
Ni-NTA Agarose	Qiagen
Nitrile Gloves	VWR
Nuclease free water	Qiagen
Parafilm M	VWR
Petri dishes	Sarstedt
Scalpel	Mediwar
Self-adhesive cover film for plates.	Ratiolab
Syringes	Henke Sass Wolf
UV-transparent disposable Cuvettes for DNA concentration	Sarstedt
ZelluTrans protein dialysis membrane MWCO 3500	Carl Roth

Table III-1: Consumables

2. Non-electronic equipments

Equipment	Specifications	Supplier
MALDI sample plate	MTP 384 target plate ground steel	Bruker
Imaging Plate	BAS-IP2040	Fuji
Pipette	0.5-10 μ L Research Plus	Eppendorf
Pipette	2-20 μ L Research Plus	Eppendorf
Pipette	10-100 μ L Research Plus	Eppendorf
Pipette	20-200 μ L Research	Eppendorf
Pipette	100-1000 μ L Research	Eppendorf
Pipette	0.5-10 μ L multichannel (8) Research	Eppendorf
Pipette	10-100 μ L multichannel (8) Research	Eppendorf
Pipette	Multipipette plus	Eppendorf

Table III-2: Non-electronic equipments

3. Electronic devices

Device	Specifications	Supplier
Agarose gel chamber	SGU 020T-02	C.B.S.Scientific
Agarose gel chamber	Midicell Primo EC-330	Thermo EC
Agarose gel chamber	Sub-cell GT Cell	Bio-rad
Agarose gel chamber	Flat Bed Electrophoresis unit	Roth
Agarose gel visualization system	UV star with powershot G10	Biometra
Autoclave	V-series	Systec
Balance	ABT 220-4M	Kem
Balance	PCJ 3500_2NM	Kem
Balance	SE622	VWR
Centrifuge	5415 R	Eppendorf
Centrifuge	5702 R	Eppendorf
Centrifuge	5810 R	Eppendorf
Centrifuge	5424	Eppendorf
Centrifuge	Sorvall Evolution RC	Thermo Scientific
Electrophoresis power supply	EV233	Consort
Electrophoresis power supply	EV 243	Consort
Electrophoresis power supply	EV 245	Consort
Electrophoresis power supply	PowerPac	Bio-rad
Electroporation device	Eporator	Eppendorf
Heating block	AccuBlock	Labnet
Ice machine	AF20	Scotsman
Incubator	1000	Heidolph
Incubator	Incu-Line	VWR
Incubator	In450	Memmert
Magnetic stirrer	Stir CB161	Stuart
Magnetic stirrer	MR Hei-Standard	Heidolph
Magnetic stirrer	RCT classic	IKA
MALDI-TOF	Ultraflex III	Bruker
Nano Drop	2000	Thermo Scientific

Orbital shaker	STD 1000 shaker	VWR
PAGE gel chamber and unit	Sequi-Gen GT	Bio-rad
PAGE gel chamber and unit	PerfectBlue	Peqlab
Plate reader	infinite M1000	TECAN
PAGE gel drying unit	MGD-4534	VWR
PCR cycler	MyCycler	Bio-Rad
PCR cycler	Peqlab 25 wells	Peqlab
PCR cycler	SimpliAmp	Life technologies
PCR cycler	T1 Thermoblock	Biometra
pH meter	FE20-Basic FiveEasy	Mettler Toledo
pH meter	FE20-Basic SevenEasy	Mettler Toledo
Phospho imager	Molecular imager Chemi-Doc	Bio-rad
Phospho imager	Typhoon FLA9500	GE Healthcare
Photometer	BioPhotometer Plus	Eppendorf
Pipette boy	Rotafiller 3000	Heathrow Scientific
Pipette boy	Easypet 4421	Eppendorf
Pipette boy	Acu Green	Integra Biosciences
qPCR cycler	iCycler	Bio-Rad
qPCR cycler	CFX384 Touch	Bio-Rad
qPCR cycler	Biometra- TOptical	Analytik Jena
qPCR cycler	PikoReal 96	Thermo Scientific
qPCR cycler	LightCycler	Roche
Radioactivity measuring device	Contamat FHT 11M	Thermo Scientific
Refrigerator	Proline 4°C	Liebherr
Refrigerator	Proline -20°C	Liebherr
Refrigerator	Ultra low U725 Innova	New Brunswick Scientific
SDS gel system	Mini Protean Tetra Cell	Bio-rad
Shaking Incubator	Ecotron	Infors HT
Shaking Incubator	I26 New Brunswick	Eppendorf
Sonicator	Digital Sonifier	Branson
Sterile hood	PCR Workstation Pro	Peqlab
Thermomixer	Thermostat Plus	Eppendorf

Thermomixer	Thermomixer comfort	Eppendorf
Thermomixer	PocketBloc	BioEr
UV-lamp	VL-6.LC	VILBER
UV-chamber	UV star (312 nm)	Biometra
Vacuum evaporator	Concentrator plus 5305	Eppendorf
Vortexer	Vortex-Genie 2	Scientific industries
Water bath	JB Aqua 12 Plus	Grant
Water purifier	Millipore filter system	ELGA

Table III-3: Electronic devices

4. Chemicals

Name	Supplier
γ P ³² ATP	Hartmann analytic
2-Mercaptoethanol	Merck
2X qPCR master mix	Qiagen
(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) (HEPES)	Roth
2-log ladder	NEB
2-(N-morpholino) ethanesulfonic acid monohydrate (MES)	Thermo Fischer
Acetic acid	Roth
Acetonitril	Sigma Aldrich
Agar-Agar (Cobe I)	Roth
Ammonium persulfate (APS)	Roth
Ammonium sulfate	Merck
Boric acid	Roth
Bovine serum albumin (BSA)	Cell signalling
Bright-Glo	Promega
Bromophenol blue	Roth
Calcium chloride (CaCl ₂)	Fisher
Color plus pre-stained ladder	NEB

Coomassie brilliant blue G250	Roth
Disodiumhydrogen phosphate	Sigma Aldrich
Dithiothreitol (DTT)	Roth
Dulbecco's modified eagle's medium (DMEM)	PAN biotech
Dulbecco's phosphate (DPBS)	PAN biotech
Ethanol (EtOH) p.a.	Sigma Aldrich
Ethanol (EtOH) 99.9%	Thermo Fischer
Ethidium bromide (EtBr)	Sigma Aldrich
Ethylenediaminetetraacetic acid (EDTA)	Roth
Fetal bovine serum (FBS)	PAN biotech
Form amide	Acros
Formaldehyde aqueous solution (HCHO)	Roth
Glucose	Roth
Glycerol	Roth
Glycine	Roth
Hydrochloric acid (HCl)	VWR
Imidazole	ABCR
Isopropanol	Fisher
Isopropyl β -D-1-thiogalactopyranoside (IPTG)	Roth
L-Glutamine	PAN biotech
Lipofectamine 2000	Thermo Fischer
Lysogeny broth agar (LB-agar)	Roth
Lysogeny broth medium (LB-medium)	Roth
LE Agarose	Roth
Magnesium chloride hexahydrate ($MgCl_2 \cdot 6H_2O$)	Acros
Magnesium sulphate-heptahydrate ($MgSO_4 \cdot 7H_2O$)	Merck
Manganese chloride tetrahydrate ($MnCl_2 \cdot 4H_2O$)	Riedel-de Hæn
Methanol aqueous solution (MeOH)	Sigma Aldrich
Monosodium phosphate (NaH_2PO_4)	Sigma Aldrich
Opti MEM	Gibco
Phenylmethanesulphonylfluoride (PMSF)	Roth
Polyoxyethylenesorbitan monolaurate (Tween-20)	Sigma Aldrich

Potassium chloride (KCl)	Roth
Potassium dihydrogenphosphate-monohydrate (KH ₂ PO ₄)	Sigma Aldrich
Protein marker, broad range (2-212 kDa)	NEB
Rotipherese Gel 40	Roth
S-adenosylmethionine (SAM)	NEB
Salmon sperm DNA	Ambion
Sodium borohydride (NaBH ₄)	Roth
Sodium chloride (NaCl)	Roth
Sodium dodecyl sulphate (SDS)	Roth
Sodium hydroxide (NaOH)	Thermo fischer
Tetramethylethylenediamine (TEMED)	Roth
Trifluoroacetic acid (TFA)	Sigma Aldrich
Tris (hydroxymethyl) aminomethane	Sigma Aldrich
Trinton X-100	Fluka
Tryptone	Roth
Urea	Roth
Xylene cyanol	Roth
Yeast Extract	Roth

Table III-4: Chemicals

5. Enzymes

Name	Function / Application	Supplier
ApaLI	Restriction Zf qPCR amplicon	New England Biolabs
AatII	Restriction final TALE plasmid	New England Biolabs
BsaI	Restriction (GG1)	New England Biolabs
BsmBI	Restriction (GG2)	New England Biolabs
DnaseI	Restriction in FRET based screening	Thermo Scientific
DpnI	Restriction (Quikchange mutagenesis)	New England Biolabs
KF (exo-)	Extension of partial dsDNA (PEX)	New England Biolabs

Lysozyme	Cell lysis	Fluka
MSssI	Methylation	New England Biolabs
NcoI	Restriction (N-terminal His Tag_pPrR_773)	New England Biolabs
NdeI	Restriction Zf qPCR amplicon	New England Biolabs
Pfu	Polymerase	New England Biolabs
Phusion	Polymerase	New England Biolabs
Plasmid safe nuclease	Digestion of non-circular DNA	Biozym
SacI-HF	Restriction BRCA1_a qPCR amplicon	New England Biolabs
Sall	Restriction ligation for p814synC/mC	New England Biolabs
SpeI	Restriction ligation for p814synC/mC	New England Biolabs
StuI	Restriction final TALE plasmid	New England Biolabs
T4 DNA ligase	Ligation	New England Biolabs
T4 polynucleotidekinase	³² P labeling	Fermentas
Taq polymerase	Polymerase	New England Biolabs
XhoI	Restriction (N-terminal His Tag_pPrR_773)	New England Biolabs

Table III-5: Enzymes

6. Antibiotics

Name/Abbreviation	Solvent	Stock Concentration	Supplier
Carbenicillin (Carb)	50 % Ethanol	50 mg/ mL	Roth
Chloramphenicol (Cam)	Ethanol	34 mg/ mL	Roth
Kanamycin (Kan)	MilliQ Water	50 mg/ mL	Roth
Spectinomycin (Spec)	MilliQ Water	100 mg/ mL	Alpha Aesar
Tetracycline (Tet)	Ethanol	12.5 mg/ mL	Sigma

Table III-6: Antibiotics

7. Kits

Name	Purpose	Supplier
GeneJET Plasmid Miniprep kit	Isolation of plasmid from cell cultures	Thermo Scientific
GeneJET Gel Extraction kit	DNA purification and agarose gel extraction	Thermo Scientific
BCA Protein Assay (DTT)	Quantification of proteins	Thermo Scientific
QIAamp DNA Mini kit	Extraction and purification of DNA	Qiagen
EpiTect Bisulphite kit	Deamination of DNA	Qiagen
REPLI-g Amplification kit	Whole genome amplification	Qiagen
EpiTect MSP kit	PCR of bisulphite converted DNA	Qiagen
MeDIP kit	Immunoprecipitation of methylated DNA	Diagenode

Table III-7: Kits

8. Buffers

Name and Conditions	Content	Application
Luciferase lysis buffer Stored at RT	NaH ₂ PO ₄ 100 mM Trinton X-100 0.2 % (v/v)	Mammalian HEK293T cell lysis
Agarose gel loading dye (6X) Stored at -20°C	FicollR-400 15% (v/v) EDTA 66 mM Tris 19.8 mM SDS 0.102% (w/v) Bromophenol blue 0.09% (w/v)	Agarose gel sample preparation
SDS gel running buffer	Tris 30.3 g/L Glycerol 144 g/L	SDS gel sample preparation

Stored at RT	SDS	10 g/L	
SDS gel loading buffer (4X) (pH = 6.8) Stored at -20°C	Tris	200 mM	SDS gel running
	Glycerol	40% (v/v)	
	2- Mercaptoethanol	1% (v/v)	
	SDS	8% (v/v)	
	Bromophenol blue	0.008% (w/v)	
Coomassie staining solution	MilliQ water	40% (v/v)	SDS gel staining
	Acetic acid	10% (v/v)	
	Methanol	50% (v/v) Brilliant	
	Blue G250	0.1% (w/v)	
Coomassie de-staining solution	MilliQ water	70% (v/v)	SDS gel de- staining
	Acetic acid	10% (v/v)	
	Methanol	20% (v/v)	
Hybridization buffer A 5XBGrK2 (pH = 8) Stored at -20°C	Tris	100 mM	Hybridization
	NaCl	250 mM	
	MgCl ₂	25 mM	
	Salmon sperm DNA	250 ng/μL	
	BSA	0.5 mg/mL	
	Glycerol	25% (v/v)	
Hybridization buffer B 5BGrK2 w/o BSA (pH = 8)	Tris-HCl	100 mM	EMSA
	NaCl	250 mM	
	MgCl ₂	25 mM	
	Glycerol	25% (v/v)	
Hybridization C buffer 5BGrK1 (pH = 8)	Tris	100 mM	PEX/FRET reactions
	NaCl	250 mM	
	MgCl ₂	25 mM	
	BSA	0.5 mg/mL	
	Glycerol	25% (v/v)	
10 X buffer X (pH = 7.9) Stored at -20°C	NaCl	1.5 M	Affinity Enrichment assay
	Tris-HCl	300 mM	
	MgCl ₂	50mM	
	BSA	5 mg/mL	
	Filtration using 0.2 μm pore size membrane		

	Sterilize and store under UV		
Taq lysis buffer (pH = 9) Stored at 4°C	Tris-HCl NaCl MgCl ₂ Trinton X-100	10 mM 300 mM 2.5mM 0.1% (v/v)	Protein purification
4 X PBS (pH = 8) Stored at 4°C	NaCl KCl Na ₂ HPO ₄ .2H ₂ O KH ₂ PO ₄	548 mM 43 mM 69 mM 24 mM	Protein purification
5 X TALE storage buffer (pH = 7.5) stored at 20°C	NaCl Tris-HCl Glycerol	1M 100 mM 50% (v/v)	TALE protein storage and dialysis
10 X TBE (pH = 8) stored at RT	Tris Boric acid EDTA	890 mM 890 mM 20 mM	PEX gel running
10 X TBE 9M Urea (pH = 8) stored at RT	Tris Boric acid EDTA Urea	890 mM 890 mM 20 mM 9 M	PEX gel running
PEX gel loading buffer Stored at -20°C	Formamide EDTA Bromophenol blue	80% (v/v) 2 mM 1 g/L	PEX sample preparation
Qiagen lysis buffer (pH = 8) Stored at 4°C	NaH ₂ PO ₄ .2H ₂ O NaCl pH adjusted using HCl	50 mM 300 mM	Protein purification
Qiagen lysis buffer 500 mM Imidazole (pH = 8) Stored at 4°C	NaH ₂ PO ₄ .2H ₂ O NaCl Imidazole pH adjusted using HCl	50 mM 300 mM 500 mM	Protein purification
KF ^{exo-} storage buffer (pH = 7.4) Stored at -20°C	Tris DTT EDTA	25 mM 1 mM 0.1 mM	PEX reaction

	Glycerol pH adjusted using HCl	50% (v/v)	
10 X TE DNA storage buffer (pH= 7.4) stored at RT	Tris EDTA pH adjusted using HCl Filtration using 0.2 µm pore size membrane	100 mM 10 mM	Long term plasmid storage
buffer Z (pH = 8) Stored at -20°C	Urea NaCl Hepes Imidazole	8 M 100 mM 20 mM 500 mM	Protein purification
10 X MES Buffer (pH = 5) Stored at RT	MES NaCl pH adjusted using HCl	200 mM 1 M	Ni-NTA bead recycling
10 X TAE buffer (pH = 7.8) stored at RT	Tris EDTA pH adjusted using acetic acid	400 mM 25 mM	EMSA gel running

Table III-8 : Buffers

9. Media

Name	Content	Application	
LB-medium (pH = 7.0)	Tryptone Yeast extract NaCl	10 g/L 5 g/L 5 g/L	Bacterial culture
LB-Agar (pH = 7.0)	Tryptone Yeast extract NaCl Agar-Agar	10 g/L 5 g/L 5 g/L 15 g/L	Bacterial colonies
Super Optimal broth with Catabolite	Tryptone Yeast extract NaCl	2% (w/v) 0.5% (w/v) 10 mM	Bacterial Transformation

repression (SOC) medium (pH = 7.0)	KCl	2.5 mM	
	MgCl ₂	10 mM	
	MgSO ₄	10 mM	
	Glucose	20 mM	
Dulbecco's modified Eagle's medium (DMEM)	FBS	10% (v/v)	HEK293T cell maintainance
	Penicillin/Streptomycin	1% (w/v)	
	L-Glutamine	1% (w/v)	

Table III-9: Media

10. Bacterial (*E. coli*) cell strains

Name and Lab Code	Purpose	Resistance	Supplier
BL21 DE3 sMoS_308	Protein expression	none	Stratagene
BL21 (DE3) Gold sMoS_308	Protein expression	Tetracycline	Agilent
GH371 sDaS_58	Plasmid isolation	none	iGEM
K12 ER 2738 sSaF_375	gDNA extraction	none	NEB

Table III-10: Bacterial (*E. coli*) cell strains

11. Mammalian cell strains

Name and Lab Code	Purpose	Resistance	Supplier
HEK293T	Transfection	none	

Table III-11: Mammalian cell strain

12. Software

Name	Function	Supplier
Serial cloner	Sequence analysis	SerialBasics

Snap gene	Sequence analysis	GSL Biotech
Quantity one	Gel picture analysis	Bio-Rad
Image J	Gel picture analysis	Nat. Institute of Health
Finch TV	Sequencing data	Geospiza
Corel Draw Graphic Suite X6	Figures	Softronic
Origin 8.6 Pro	Data analysis and quantification	OriginLab
MS Office	Word/ Xcel/ Paint	Microsoft
BioDoc Analyse Geldokumentation	Agarose gel image	Analytic Jena
iCycler software	qPCR data analysis	Bio-Rad
CFX-384 software	qPCR data analysis	Bio-Rad

Table III-12: Softwares

13. Oligonucleotides

Non-canonical C nucleobases are marked as following:

V: 5mC **W:** 5hmC **X:** 5fC **Y:** 5caC **Z:** 4mC

Code	Type	Description	Sequence	Supplier
oDas_247	Sequencing	GG2 plasmid rv	GGG TTA TGC TAG TTA TTG CTC AG	Sigma
oDas_315	Sequencing	GG2 plasmid fw	CGT AGA GGA TCG AGA TC	Sigma
oGrK_465	PCR/PEX	Hey2_b C6 to 5mC fw	TGG ATT CCC ACT CTT CAG CCC CAG CGT TAC AGC ATC TTC AGT GGC TTC TTC CAC CGT GAG CTC TTC VGT TTC CAC ATC C	Sigma

oGrK_466	PEX	Hey2_b rv	GGA TGT GGA AAC GGA AGA	Sigma
oGrK_476	PCR/PEX	Hey2_b C fw	TGG ATT CCC ACT CTT CAG CCC CAG CGT TAC AGC ATC TTC AGT GGC TTC TTC CAC CGT GAG CTC TTC CGT TTC CAC ATC C	Sigma
oGrK_520	PCR	Hey2_b C6 to 5hmC fw	TGG ATT CCC ACT CTT CAG CCC CAG CGT TAC AGC ATC TTC AGT GGC TTC TTC CAC CGT GAG CTC TTC W GT TTC CAC ATC C	Metabion
oDas_734	Sequencing	GG1 plasmid fw	TTG ATG CCT GGC AGT TCC CT	Sigma
oDas_735	Sequencing	GG1 plasmid rv	CGA ACC GAA CAG GCT TAT GT	Sigma
oSaB_890	Sequencing	GG2 plasmid fw	AGA TAT GAT TGC GGC CCT G	Sigma
oDas_903	qPCR	Hey2 fw	GGA ACT GAA GTG GGA GCG T	Sigma
oDas_904	qPCR	Hey2 rv	GAT GAT ATC AAA GCA TAC ATC TCA	Sigma
oGrK_1056	PCR	Hey2_b C2/C6/C11 to 5mC fw	TGG ATT CCC ACT CTT CAG CCC CAG CGT TAC AGC ATC TTC AGT GGC TTC TTC CAC CGT GAG CT V TTC V GT TT V CAC ATC C	Sigma
oGrK_1057	PCR	Hey2_b C6/C11 to 5mC fw	TGG ATT CCC ACT CTT CAG CCC CAG CGT TAC AGC ATC TTC AGT GGC	Sigma

			TTC TTC CAC CGT GAG CTV TTC CGT TTV CAC ATC C	
oPrR_1060	qPCR	BRCA1 fw	GGG GAT TGG GAC CTC	Sigma
oPrR_1064	qPCR	CDKN2A fw	TCG AAG CGC TAC CTG AT	Sigma
oPrR_1065	qPCR	CDKN2A rv	TAG AGG AGG TGC GGG	Sigma
oPrR_1076	PEX	BRCA1(26) C13/C20 to 5mC fw	AAA CTC AGG TAG AAT TCT TCC TCT TCC GTC TCT TTC CTT TTA VGT CAT CVG GGG GCA GA	Sigma
oPrR_1077	PEX	BRCA1(26) C fw	AAA CTC AGG TAG AAT TCT TCC TCT TCC GTC TCT TTC CTT TTA CGT CAT CCG GGG GCA GA	Sigma
oPrR_1078	PEX	BRCA1(26) rv	CTG CCC CCG GAT GAC GTA AAA GGA AAG AGA	Sigma
oPrR_1079	PEX	CDKN2A (26) C6 to 5mC fw	ACC CGA CCC CGG GCC GCG GCC GTG GCC AGC CAG TCA GCV GAA GGC TCC ATG CTG CTC CC	Sigma
oPrR_1080	PEX	CDKN2A (26) C fw	ACC CGA CCC CGG GCC GCG GCC GTG GCC AGC CAG TCA GCC GAA GGC TCC ATG CTG CTC CC	Sigma
oPrR_1081	PEX	CDKN2A (30) rv	GCA TGG AGC CTT CGG CTG ACT GGC TGG CCA	Sigma
oPrR_1082	PEX	CDKN2A (26) rv	GGG AGC AGC ATG GAG CCT TCG GCT GA	Sigma
oPrR_1083	PEX	MGMT C11 to 5mC fw	CTT CGG CCG GTA CAA GCC GGG CGG CGC CTT	Sigma

			CCC AGC TTC V G C TG AGG CTC TGT GCC TT	
oPrR_1084	PEX	MGMT C fw	CTT CGG CCG GTA CAA GCC GGG CGG CGC CTT CCC AGC TTC CGC CTG AGG CTC TGT GCC TT	Sigma
oPrR_1085	PEX	MGMT rv	AAG GCA CAG AGC CTC AGG CGG AAG CTG GG A	Sigma
oPrR_1086	sequencing	BRCA1 fw	AGA GCA GAG GGT GAA G	Sigma
oPrR_1087	sequencing	CDKN2A fw	TTC AGG GGT GCC ACA	Sigma
oPrR_1088	sequencing	MGMT fw	AAA CTA AGG CAC AGA GC	Sigma
oPrR_1114	qPCR	BRCA1 rv	AAA GTG CCT GCC CTC TA	Sigma
oGrK_1152	PCR	Hey2_b C rv	GGA TGT GGA AAC GGA AGA GCT CAC GGT GGA AGA AGC CAC TGA AGA TGC TGT AAC GCT GGG GCT GAA GAG TGG GAA TCC A	Sigma
PrR_1184	PEX	BRCA1(26) C13 to 5mC fw	AAA CTC AGG TAG AAT TCT TCC TCT TCC GTC TCT TTC CTT TTA CGT CAT C VG GGG GCA GA	Sigma
oPrR_1185	PEX	BRCA1(26) C20 to 5mC fw	AAA CTC AGG TAG AAT TCT TCC TCT TCC GTC TCT TTC CTT TTA V GT CAT CCG GGG GCA GA	Sigma

oGrK_1236	PEX/FRET/ EMSA/PCR	Hey2_b fw	TGG ATT CCC ACT CTT CAG CCC CAG CGT TAC AGC ATC TTC AGT GGC TTC TTC CAC CGT GAG CTC TTC Z GT TTC CAC ATC C	Iba lifescience s
oPrR_1256	Bisulphite	BRCA1 fw 1	TGG GGG ATT GGG ATT TTT TTT	Sigma
oPrR_1257	Bisulphite	BRCA1 rv 1	TTA ACC ACC CAA TCT ACC CCC	Sigma
oPrR_1307	Bisulphite	BRCA1 fw 2	GAA ATT GGA GAT TTT TAT TAG GG	Sigma
oPrR_1308	Bisulphite	BRCA1 rv 2	TAT CTA AAA AAC CCC ACA ACC TAT C	Sigma
oGrK_1422	PCR	Hey2_b C6 to 5caC fw	TGG ATT CCC ACT CTT CAG CCC CAG CGT TAC AGC ATC TTC AGT GGC TTC TTC CAC CGT GAG CTC TTC Y GT TTC CAC ATC C	Metabion
oGrK_1423	PCR	Hey2_b C6 to 5fC fw	TGG ATT CCC ACT CTT CAG CCC CAG CGT TAC AGC ATC TTC AGT GGC TTC TTC CAC CGT GAG CTC TTC X GT TTC CAC ATC C	Metabion
oPrR_1457	qPCR	MGMT fw	GAA AAG GTA CGG GCC A	Sigma
oPrR_1459	qPCR	MGMT rv	GCC TGA CCC GGA TG	Sigma
oPrR_1514	Sequencing	pAnI521 fw1	GGA ACA ACA CTC AAC CCT	Sigma
oPrR_1515	Sequencing	pAnI521 rv1	CCC GGA AGA GAG TCA AT	Sigma

oGrK_1516	PCR	BRCA1 C fw	CTT CCT CTT CCG TCT CTT TCC TTT TAC GTC ATC CGG GGG CAG ACT	Sigma
oGrK_1518	PCR	BRCA1 C13 to 5mC fw	CTT CCT CTT CCG TCT CTT TCC TTT TAV GTC ATC CGG GGG CAG ACT	Sigma
oGrK_1529	PCR	BRCA1 rv	AGT CTG CCC CCG GAT GAC GTA AAA GGA AAG AGA CGG AAG AGG AAG	Sigma
oPrR_1559	Sequencing	pAnI521 fw 2	TCA GGC AAC TAT GGA TGA A	Sigma
oPrR_1560	Sequencing	pAnI521 rv 2	CCG ATT CAT TAA TGC AGC T	Sigma
oPrR_1561	Sequencing	pAnI521 fw 3	TTG AGT GAG CTG ATA CCG	Sigma
oPrR_1562	Sequencing	pAnI521 rv 3	TGT GAA TCG CTT CAC GAC	Sigma
oGrK_1570	qPCR	BRCA1_ant i fw	CAA TCC AGA GCC CCG	Sigma
oGrK_1577	qPCR	BRCA1_ant i rv	GGA CTC TAC TAC CTT TAC C	Sigma
oGrK_1591	PCR/ EMSA	CDKN2A C fw	GGC CAG CCA GTC AGC CGA AGG CTC CAT GCT GCT CCC CGC CGC CGG C	Sigma
oGrK_1592	PCR/ EMSA	CDKN2A C6 to 5mC	GGC CAG CCA GTC AGC VGA AGG CTC CAT GCT GCT CCC CGC CGC CGG C	Sigma
oGrK_1597	PCR	Hey2_b rv	TTG TAT AAA TCG GAG AAG ATC CTG TCT CCC CCC AAA CAA ACA GTA G	Sigma

oGrK_1599	PCR	BRCA1_ant i fw	GCA CAG GTC TCC AAT CTA TCC ACT GGA TTT CCG TGA GAA TTG TGC CCG	Sigma
oGrK_1600	PCR	BRCA1_ant i C2 to 5mC rv	AGT CTG CCC CCG GAT GAV GTA AAA GGA AAG AGA CGG AAG AGG AAG	Sigma
oGrK_1601	PCR	BRCA1_ant i rv	GTC AAA GAA TAC CCA TCT GTC AGC TTC GGA AAT CCA CTC TCC CAC	Sigma
oGrK_1617	PCR/ EMSA	CDKN2A rv	GCC GGC GGC GGG GAG CAG CAT GGA GCC TTC GGC TGA CTG GCT GGC C	Sigma
oPrR_1748	qPCR	ZCH3H13 fw	TCT CGG TCC ACT CGT GAT G	Sigma
oPrR_1749	qPCR	ZCH3H13 rv	CCG GGA TTC TTC TGG ATA TG	Sigma
oPrR_1759	qPCR	ZAP-70 fw	GCT CTG GGA GAC CTG	Sigma
oPrR_1764	qPCR	ZAP-70 rv	CGG TTC TGG GTG AGG	Sigma
oPrR_1803	PEX	ZAP-70 (26) rv	ATG TCA CAC CCC GCC AGG GTT TCT CA	Sigma
oPrR_1805	PEX	ZAP-70 (23) C fw	GAG GAT GAA CCC CCC CCA TCC ATG AGT GAG AAA CCC TGG CGG GGT GTG ACA TCCTCC CC	Sigma
oPrR_1826	PEX	ZAP-70 (26) C14 to 5mC	TTC AGC CCC AGC GTT ACA GCA TCT TCA GTG GCT TCT TCC ACC GTG AGC TCT TCV GTT TCC ACA TCC ACC ACA TCC CAA C	Sigma

oPrR_1827	PEX	ZAP-70 (26) C fw	TTC AGC CCC AGC GTT ACA GCA TCT TCA GTG GCT TCT TCC ACC GTG AGC TCT TCC GTT TCC ACA TCC ACC ACA TCC CAA C	Sigma
oPrR_1828	PEX	ZAP-70 (23) rv	GTT GGG ATG TGG TGG ATG TGG AAA CGG	Sigma
oPrR_1835	PEX	ZAP-70 (23) C4 to 5mC	GAG GAT GAA CCC CCC CCA TCC ATG AGT GAG AAA CCC TGG V GG GGT GTG ACA TCC TCC CC	Sigma
oSaM_1892	FRET	Hey2_b rv	Cy5 GGA TGT GGA AAC GGA AGA Cy3	Metabion
oPrR_1916	Restriction/ digestion	pAnI521 N- terminal His tag fw	TTT TCA TAT GGG CCA TCA CCA TCA CCA TCA CAG CAA GGG CGA GGA GCT GTT CAC CG	Sigma
oPrR_1917	Restriction/ digestion	pAnI521 N- terminal His tag rv	AAA AGC GGC CGC TCA TTC CAG CAG ATG GTC GTT GGA GAC GAC C	Sigma
oPrR_1918	Mutation quikchange	N-terminal K171A fw	GCA ACA GGA GAA AAT CAA GCC TGC CGT CAG GAG CAC CGT CGC GCA AC	Sigma
oPrR_1919	Mutation quikchange	N-terminal K171A rv	GTG TTG CGC GAC GGT GCT CCT GAC GGC AGG CTT GAT TTT CTC CTG TTG	Sigma
oPrR_1920	Mutation quikchange	N-terminal R173A fw	GGA GAA AAT CAA GCC TAA GGT CGC CAG CAC CGT CGC GCA ACA CCA C	Sigma

oPrR_1921	Mutation quikchange	N-terminal R173A rv	CGT GGT GTT GCG CGA CGG TGC TGG CGA CCT TAG GCT TGA TTT TCT C	Sigma
oPrR_1922	Mutation quikchange	N-terminal R236A fw	GTA AAC AGT GGT CGG GAG CGG CCG CAC TTG AGG CGC TGC TGA CTG	Sigma
oPrR_1923	Mutation quikchange	N-terminal R236A rv	GTC AGC AGC GCC TCA AGT GCG GCC GCT CCC GAC CAC TGT TTA CC	Sigma
oPrR_1924	Mutation quikchange	N-terminal R266A fw	CAG CTG CTG AAG ATC GCG AAG GCC GGG GGA GTA ACA GCG GTA G	Sigma
oPrR_1925	Mutation quikchange	N-terminal R266A rv	CTC TAC CGC TGT TAC TCC CCC GGC CTT CGC GAT CTT CAG CAG C	Sigma
oPrR_1926	Mutation quikchange	N-terminal K171A/R17 3A fw	AGG AGA AAA TCA AGC CTG CCG TCG CCA GCA CCG TCG CGC AAC ACC A	Sigma
oPrR_1927	Mutation quikchange	N-terminal K171A/R17 3A rv	GTG TTG CGC GAC GGT GCT GGC GAC GGC AGG CTT GAT TTT CTC CTG TTG	Sigma
oPrR_1928	Mutation quikchange	N-terminal K262A fw	CAC CGG GCA GCT GCT GGC CAT CGC GAA GAG AGG GGG AG	Sigma
oPrR_1929	Mutation quikchange	N-terminal K262A rv	CCC CTC TCT TCG CGA TGG CCA GCA GCT GCC CGG TGT CG	Sigma
oPrR_1930	Mutation quikchange	N-terminal K262A/K26 5A/R266A fw	CAG CTG CTG GCC ATC GCG GCC GCC GGG GGA GTA ACA GCG GT	Sigma

oPrR_1931	Mutation quikchange	N-terminal K262A/K26 5A/R266A rv	CGC TGT TAC TCC CCC GGC GGC CGC GAT GGC CAG CAG CTG C	Sigma
oPrR_1932	Mutation quikchange	pNN3 K16A/Q17 A fw	CGC CAG CAA CAA TGG CGG CGC CGC CGC GCT CGA AAC GGT GCA GCG	Sigma
oPrR_1933	Mutation quikchange	pNN3 K16A/Q17 A rv	GCT GCA CCG TTT CGA GCG CGG CGG CGC CGC CAT TGT TGC TGG CGA TAG	Sigma
oPrR_1934	Mutation quikchange	pNN7 K16A/Q17 A fw	CGC CAG CAA CAA TGG CGG CGC CGC CGC GCT CGA AAC GAG ACC CTC	Sigma
oPrR_1935	Mutation quikchange	pNN7 K16A/Q17 A rv	GGG TCT CGT TTC GAG CGC GGC GGC GCC GCC ATT GTT GCT GGC GAT AG	Sigma
oPrR_1936	Mutation quikchange	pHD3 K16A/Q17 A fw	CGC CAG CCA CGA TGG CGG CGC CGC CGC GCT CGA AAC GGT GCA GC	Sigma
oPrR_1937	Mutation quikchange	pHD3 K16A/Q17 A rv	CTG CAC CGT TTC GAG CGC GGC GGC GCC GCC ATC GTG GCT GGC GAT AG	Sigma
oPrR_1938	Mutation quikchange	pHD5 K16A/Q17 A fw	CGC CAG CCA CGA TGG CGG CGC CGC CGC GCT CGA AAC GGT GCA GG	Sigma
oPrR_1939	Mutation quikchange	pHD5 K16A/Q17 A rv	CTG CAC CGT TTC GAG CGC GGC GGC GCC GCC ATC GTG GCT GGC GAT AG	Sigma

oPrR_1963	Sequencing	pAnI521 N terminal His tag fw	AGA CCC CAA CGA GAA GCG CGA	Sigma
oPrR_1964	Sequencing	pAnI521 N terminal His tag rv	GAC CAT GAT TAC GCC AAG CGC	Sigma
oAnW_2016	Luciferase assay	CDKN2A insert C fw	TTT TGT CGA CTC AGC CGA AGG CTC CAT GCT GCT CCC ACT AGT TTT T	Sigma
oAnW_2017	Luciferase assay	CDKN2A insert rv	AAA AAC TAG TGG GAG CAG CAT GGA GCC TTC GGC TGA GTC GAC AAA A	Sigma
oPrR_2053	qPCR	CDKN2A fw	GGC TCC TCA TTC CTC TTC CTT G	Sigma
oPrR_2054	qPCR	CDKN2A rv	GTC GGG TAG AGG AGG TGC G	Sigma
oPrR_2121	EMSA	ZAP-70 C fw	CAT CCA TGA GTG AGA AAC CCT GGC GGG GTG TGA CAT CCT CCC CCG G	Sigma
oPrR_2122	EMSA	ZAP-70 rv	CCG GGG GAG GAT GTC ACA CCC CGC CAG GGT TTC TCA CTC ATG GAT G	Sigma
oPrR_2123	EMSA	ZAP-70 C14 to 5mC fw	CAT CCA TGA GTG AGA AAC CCT GG V GGG GTG TGA CAT CCT CCC CCG G	Sigma
oSAM_2170	EMSA	Hey2_b rv	Cy5 GGA TGT GGA AAC GGA AGA GCT CAC GGT GGA AGA AGC CAC TGA	Metabion

			AGA TGC TGT AAC GCT GGG GCT GAA GAG TGG GAA TCC A	
oPrR_2248	Luciferase assay	CDKN2A insert C6 to 5mC fw	TTT TGT CGA CTC AGC V GA AGG CTC CAT GCT GCT CCC ACT AGT TTT T	Sigma
oPrR_2278	FRET	ZAP-70 C fw	CCC CAG GGG GCT CTG GGA GAC CTG GCA GAG GAT GAA CCC CCC CCA TCC ATG AGT GAG AAA CCC TGG CGG GGT GTG ACA T	Sigma
oPrR_2279	FRET	ZAP-70 rv	Cy5 ATG TCA CAC CCC GCC AGG GTT TCT CA Cy3	Metabion
oPrR_2280	FRET	CDKN2A C fw	GCA GCG CCC GCA CCT CCT CTA CCC GAC CCC GGG CCG CGG CCG TGG CCA GCC AGT CAG CCG AAG GCT CCA TGC TGC TCC C	Sigma
oPrR_2281	FRET	CDKN2A rv	Cy5 GGG AGC AGC ATG GAG CCT TCG GCT GA Cy3	Metabion
oPrR_2282	FRET	ZAP-70 C14 to 5mC	CCC CAG GGG GCT CTG GGA GAC CTG GCA GAG GAT GAA CCC CCC CCA TCC ATG AGT GAG AAA CCC TGG V GG GGT GTG ACA T	Sigma
oPrR_2283	FRET	CDKN2A C6 to 5mC	GCA GCG CCC GCA CCT CCT CTA CCC GAC CCC GGG CCG CGG CCG TGG	Sigma

			CCA GCC AGT CAG CVG AAG GCT CCA TGC TGC TCC C	
oPrR_2284	Mutation quikchange	pAnW750 K262A fw	CAC AGG CCA ACT TGT GGC CAT TGC AAA ACG TGG CGG CG	Sigma
oPrR_2285	Mutation quikchange	pAnW750 K262A rv	CGC CAC GTT TTG CAA TGG CCA CAA GTT GGC CTG TGT CC	Sigma
oPrR_2423	Bisulphite	pAnW814 fw 1	ATA TTT TGG ATT ATT AGT TTT ATA GG	Sigma
oPrR_2424	Bisulphite	pAnW814 rv 1	AAC ATC TTC CAT AAT AAA TAC CTC CT	Sigma
oPrR_2425	Bisulphite	pAnW814 fw 2	TAG TGG TTA TAA TAG TAG TTT TAG TT	Sigma
oPrR_2426	Bisulphite	pAnW814 rv 2	ATA ATT TAT ATT CAA CCC ATA A	Sigma
oPrR_2427	Bisulphite	pAnW814 fw 3	GAG GTT TTA AGG GGT TAT GTT ATT AAT	Sigma
oPrR_2428	Bisulphite	pAnW814 rv 3	ACT AAT ACC CAT ACT ATT CAA CAA CT	Sigma
oPrR_2503	FRET	CDKN2A (18) rv	Cy5 GGG AGC AGC ATG GAG CCT TCG GCT GA Cy3	
oPrR_2534	FRET	CDKN2A fw	GCA GCG CCC GCA CCT CCT CTA CCC GAC CCC GGG CCG CGG CCG TGG CCA GCC AGT CAG CZG AAG GCT CCA TGC TGC TCC C	

Table III-13: Oligonucleotides

14. Plasmids

Code	Backbone	Description
pSaB_339	pET-TRX-6His	GG2 final entry vector
pGrK_430	pET-TRX-TALE	TALE Hey2_a (20)
pIrT_446	pET-TRX-TALE	TALE Hey2_b (18)
pGrK_471	pET-TRX-TALE	TALE Hey2_c (18)
pPrR_515	pET-TRX-TALE-6His	TALE BRCA1_(30)
pPrR_516	pET-TRX-TALE-6His	TALE BRCA1_(26)
pPrR_517	pET-TRX-TALE-6His	TALE CDKN2A_(30)
pPrR_518	pET-TRX-TALE-6His	TALE CDKN2A_(26)
pPrR_519	pET-TRX-TALE-6His	TALE MGMT_(30)
pAnI_520	pET-TRX-TALE-6His	TALE MGMT_(26)
pPrR_521	pET-GFP-6His	GG2 final entry vector
pSaM_586	pET-GFP-TALE-6His	TALE Hey2_b HD5
pPrR_651	pET-GFP-TALE-6His	TALE Hey2_b HD5 to N*
pPrR_652	pET-GFP-TALE-6His	TALE BRCA1_anti_(26)
pPrR_673	pET-GFP-TALE-6His	TALE_Hey2_b N*(2) HD(6) N*(11)_(26)
pPrR_674	pET-GFP-TALE-6His	TALE_Hey2_b N*(2) HD(6) N*(11)_(21)
pPrR_675	pET-GFP-TALE-6His	TALE_CDKN2A_N*(2) N*(19)_(26)
pPrR_676	pET-GFP-TALE-6His	TALE ZAP-70_(23)
pPrR_677	pET-GFP-TALE-6His	TALE ZAP-70_(26)
pSaM_738	pET-GFP-TALE-6His	TALE Hey2_b HD5 to A*
pSaM_739	pET-GFP-TALE-6His	TALE Hey2_b HD5 to C*
pSaM_740	pET-GFP-TALE-6His	TALE Hey2_b HD5 to D*
pSaM_741	pET-GFP-TALE-6His	TALE Hey2_b HD5 to E*
pSaM_742	pET-GFP-TALE-6His	TALE Hey2_b HD5 to F*
pSaM_743	pET-GFP-TALE-6His	TALE Hey2_b HD5 to G*
pSaM_744	pET-GFP-TALE-6His	TALE Hey2_b HD5 to H*
pSaM_745	pET-GFP-TALE-6His	TALE Hey2_b HD5 to I*
pSaM_746	pET-GFP-TALE-6His	TALE Hey2_b HD5 to K*
pSaM_747	pET-GFP-TALE-6His	TALE Hey2_b HD5 to L*
pSaM_748	pET-GFP-TALE-6His	TALE Hey2_b HD5 to M*

pSaM_749	pET-GFP-TALE-6His	TALE Hey2_b HD5 to P*
pAnW_750	pcDNA3.1-VP64	Mammalian GG2 final entry vector
pAnW_751	pcDNA3.1	pAnW 750 without VP64
pAnW_755	pBS-MinCMV-luc	Mammalian reporter plasmid
pPrR_762	RVD pNN3	Mutation K16A&Q17A
pPrR_763	RVD pNN7	Mutation K16A&Q17A
pPrR_764	RVD pHD3	Mutation K16A&Q17A
pPrR_765	RVD pHD7	Mutation K16A&Q17A
pPrR_766	pET-GFP-6His	Mutation K171A
pPrR_767	pET-GFP-6His	Mutation R173A
pPrR_768	pET-GFP-6His	Mutation R236A
pPrR_769	pET-GFP-6His	Mutation R266A
pPrR_770	pET-GFP-6His	Mutation K171A&R173A
pPrR_771	pET-GFP-6His	Mutation K262A
pPrR_772	pET-GFP-6His	Mutation K262A&K265A&R266A
pPrR_773	pET-6His-GFP	GG2 entry vector with N-terminal His tag
pSaM_774	pET-GFP-TALE-6His	TALE Hey2_b HD5 to Q*
pSaM_775	pET-GFP-TALE-6His	TALE Hey2_b HD5 to R*
pSaM_776	pET-GFP-TALE-6His	TALE Hey2_b HD5 to S*
pSaM_777	pET-GFP-TALE-6His	TALE Hey2_b HD5 to T*
pSaM_778	pET-GFP-TALE-6His	TALE Hey2_b HD5 to V*
pSaM_779	pET-GFP-TALE-6His	TALE Hey2_b HD5 to X*
pSaM_780	pET-GFP-TALE-6His	TALE Hey2_b HD5 to Y*
pPrR_788	pET-GFP-TALE-6His	TALE_CDKN2A_(26) in pPrR766 entry vector
pPrR_789	pET-GFP-TALE-6His	TALE_CDKN2A_(26) in pPrR767 entry vector
pPrR_790	pET-GFP-TALE-6His	TALE_CDKN2A_(26) in pPrR768 entry vector
pPrR_791	pET-GFP-TALE-6His	TALE_CDKN2A_(26) in pPrR769 entry vector
pPrR_792	pET-GFP-TALE-6His	TALE_CDKN2A_(26) in pPrR770 entry vector

pPrR_793	pET-GFP-TALE-6His	TALE_CDKN2A_(26) in pPrR771 entry vector
pPrR_794	pET-GFP-TALE-6His	TALE_CDKN2A_(26) in pPrR772 entry vector
pPrR_795	pET-6His-GFP-TALE	TALE_CDKN2A_(26) in pPrR773
pPrR_796	pET-6His-GFP-TALE	TALE_ZAP-70_(26) in pPrR773
pAnW_814	pBS-MinCMV-luc	pAnW_755 with CDKN2A (26) binding site
pPrR_814s ynC	pBS-MinCMV-luc	pAnW_755 with CDKN2A_(26) binding site restriction/ligation C
pPrR_814s ynmC	pBS-MinCMV-luc	pAnW_755 with CDKN2a_(26) binding site restriction/ligation C6 to 5mC
pAnW_818	pcDNA3.1-TALE-VP64	TALE_CDKN2A_(26)
pAnW_820	pcDNA3.1-TALE	TALE_CDKN2A_(26)
pAnW_893	pcDNA3.1-TALE-VP64	TALE_CDKN2A_(26) HD6 to G*
pPrR_897	pET-GFP-TALE-6His	TALE_CDKN2A_(26) NN4 with K16A&Q17A
pPrR_898	pET-GFP-TALE-6His	TALE_CDKN2A_(26) HD6 with K16A&Q17A
pPrR_899	pET-GFP-TALE-6His	TALE_CDKN2A_(26) NN18 with K16A&Q17A
pPrR_900	pET-GFP-TALE-6His	TALE_CDKN2A_(26) HD24 with K16A&Q17A
pPrR_901	pET-GFP-TALE-6His	TALE_CDKN2A_(26) HD6&NN8 with K16A&Q17A
pPrR_902	pET-GFP-TALE-6His	TALE_CDKN2A_(26) NN4&HD6 with K16A&Q17A
pPrR_903	pET-GFP-TALE-6His	TALE_CDKN2A_(26) NN4&HD6&NN18&HD24 with K16A&Q17A
pPrR_904	pET-GFP-TALE-6His	TALE_ZAP-70_(26) NN4 with K16A&Q17A
pPrR_905	pET-GFP-TALE-6His	TALE_ZAP-70_(26) HD14 with K16A&Q17A

pPrR_906	pET-GFP-TALE-6His	TALE_ZAP-70_(26) NN18 with K16A&Q17A
pPrR_907	pET-GFP-TALE-6His	TALE_ZAP-70_(26) HD24 with K16A&Q17A
pPrR_908	pET-GFP-TALE-6His	TALE_ZAP-70_(26) NN4&HD14 with K16A&Q17A
pPrR_909	pET-GFP-TALE-6His	TALE_ZAP-70_(26) NN4&HD14&NN18 with K16A&Q17A
pPrR_910	pET-GFP-TALE-6His	TALE_ZAP-70_(26) NN4&HD14&NN18&HD24 with K16A&Q17A
PrR_944	pET-GFP-TALE-6His	TALE_ZAP-70_(26) in pPrR766 entry vector
pPrR_945	pET-GFP-TALE-6His	TALE_ZAP-70_(26) in pPrR767 entry vector
pPrR_946	pET-GFP-TALE-6His	TALE_ZAP-70_(26) in pPrR768 entry vector
pPrR_947	pET-GFP-TALE-6His	TALE_ZAP-70_(26) in pPrR769 entry vector
pPrR_948	pET-GFP-TALE-6His	TALE_ZAP-70_(26) in pPrR770 entry vector
pPrR_949	pET-GFP-TALE-6His	TALE_ZAP-70_(26) in pPrR771 entry vector
pPrR_950	pET-GFP-TALE-6His	TALE_ZAP-70_(26) in pPrR772 entry vector
pPrR_951	pET-GFP-TALE-6His	CDKN2A_(26) in pPrR_770 NN4 with K16A&Q17A
pPrR_952	pET-GFP-TALE-6His	CDKN2A_(26) in pPrR_771 HD6 with K16A&Q17A
pPrR_953	pET-GFP-TALE-6His	CDKN2A_(26) in pPrR_770 NN4 with K16A&Q17A
pPrR_954	pET-GFP-TALE-6His	CDKN2A_(26) in pPrR_771 HD6 with K16A&Q17A

pPrR_955	pET-GFP-TALE-6His	ZAP-70_(26) in pPrR_770 NN4 with K16A&Q17A
pPrR_956	pET-GFP-TALE-6His	ZAP-70_(26) in pPrR_770 HD14 with K16A&Q17A
pPrR_957	pET-GFP-TALE-6His	ZAP-70_(26) in pPrR_771 NN4 with K16A&Q17A
pPrR_958	pET-GFP-TALE-6His	ZAP-70_(26) in pPrR_770 HD14 with K16A&Q17A
pPrR_1134	pcDNA3.1- VP64	pAnW_750 with N-terminal K262A mutation
pPrR_1149	pcDNA3.1-TALE-VP64	CDKN2A_(26)_in pPrR_1134
pPrR_1198	pET-GFP-TALE-6His	BRCA1_(26)
pPrR_1199	pET-GFP-TALE-6His	CDKN2A_(26)
pPrR_1200	pET-GFP-TALE-6His	MGMT_(26)

Table III-14: Plasmids

15. TALE proteins

Number	Target	Target sequence	Plasmid
064	BRCA1 (26)	TCTTTCCTTTTACGTCATCCGGGGGC	pPrR_515
065	BRCA1 (30)	TCTCTTTCCTTTTACGTCATCCGGGGG CAG	pPrR_516
066	CDKN2A (26)	TCAGCCGAAGGCTCCATGCTGCTCCC	pPrR_517
067	CDKN2A (30)	TGGCCAGCCAGTCAGCCGAAGGCTCC ATGC	pPrR_518
068	MGMT (26)	TCCCAGCTTCCGCCTGAGGCTCTGTG	pPrR_519
069	MGMT (30)	TCCCAGCTTCCGCCTGAGGCTCTGTG CCTT	pPrR_520
137	Hey2_b (18)	TCTTCCGTTTCCACATCC	pPrR_651

138	BRCA1_anti	TGCCCCCGGATGACGTAAAAGGAAA G	pPrR_652
143	Hey2_b (30)	TCTTCCGTTTCCACATCCACCACATCC CAA	pPrR_673
144	Hey2_b (25)	TCTTCCGTTTCCACATCCACCACAT	pPrR_674
145	Hey2_b (21)	TCTTCCGTTTCCACATCCACC	pPrR_675
146	CDKN2A (26)	TCAGCCGAAGGCTCCATGCTGCTCCC	pPrR_676
147	ZAP-70 (23)	TGGCGGGGTGTGACATCCTCCCC	pPrR_677
148	ZAP-70 (26)	TGAGAAACCCTGGCGGGGTGTGACAT	pPrR_678
171	Hey2_b (18)	TCTTCCGTTTCCACATCC	pPrR_738
172	Hey2_b (18)	TCTTCCGTTTCCACATCC	pPrR_739
173	Hey2_b (18)	TCTTCCGTTTCCACATCC	pPrR_740
174	Hey2_b (18)	TCTTCCGTTTCCACATCC	pPrR_741
175	Hey2_b (18)	TCTTCCGTTTCCACATCC	pPrR_742
176	Hey2_b (18)	TCTTCCGTTTCCACATCC	pPrR_743
177	Hey2_b (18)	TCTTCCGTTTCCACATCC	pPrR_744
178	Hey2_b (18)	TCTTCCGTTTCCACATCC	pPrR_745
179	Hey2_b (18)	TCTTCCGTTTCCACATCC	pPrR_746
180	Hey2_b (18)	TCTTCCGTTTCCACATCC	pPrR_747
181	Hey2_b (18)	TCTTCCGTTTCCACATCC	pPrR_748
182	Hey2_b (18)	TCTTCCGTTTCCACATCC	pPrR_749
183	Hey2_b (18)	TCTTCCGTTTCCACATCC	pPrR_774
184	Hey2_b (18)	TCTTCCGTTTCCACATCC	pPrR_775
185	Hey2_b (18)	TCTTCCGTTTCCACATCC	pPrR_776
186	Hey2_b (18)	TCTTCCGTTTCCACATCC	pPrR_777
187	Hey2_b (18)	TCTTCCGTTTCCACATCC	pPrR_778
188	Hey2_b (18)	TCTTCCGTTTCCACATCC	pPrR_779
189	Hey2_b (18)	TCTTCCGTTTCCACATCC	pPrR_780
190	CDKN2A (26)	TCAGCCGAAGGCTCCATGCTGCTCCC	pPrR_788
191	CDKN2A (26)	TCAGCCGAAGGCTCCATGCTGCTCCC	pPrR_789
192	CDKN2A (26)	TCAGCCGAAGGCTCCATGCTGCTCCC	pPrR_790
193	CDKN2A (26)	TCAGCCGAAGGCTCCATGCTGCTCCC	pPrR_791
194	CDKN2A (26)	TCAGCCGAAGGCTCCATGCTGCTCCC	pPrR_792

195	CDKN2A (26)	TCAGCCGAAGGCTCCATGCTGCTCCC	pPrR_793
196	CDKN2A (26)	TCAGCCGAAGGCTCCATGCTGCTCCC	pPrR_794
197	CDKN2A (26)	TCAGCCGAAGGCTCCATGCTGCTCCC	pPrR_795
198	CDKN2A (26)	TCAGCCGAAGGCTCCATGCTGCTCCC	pPrR_796
242	ZAP-70 (26)	TGAGAAACCCTGGCGGGGTGTGACAT	pPrR_944
243	ZAP-70 (26)	TGAGAAACCCTGGCGGGGTGTGACAT	pPrR_945
244	ZAP-70 (26)	TGAGAAACCCTGGCGGGGTGTGACAT	pPrR_946
245	ZAP-70 (26)	TGAGAAACCCTGGCGGGGTGTGACAT	pPrR_947
246	ZAP-70 (26)	TGAGAAACCCTGGCGGGGTGTGACAT	pPrR_948
247	ZAP-70 (26)	TGAGAAACCCTGGCGGGGTGTGACAT	pPrR_949
248	ZAP-70 (26)	TGAGAAACCCTGGCGGGGTGTGACAT	pPrR_950
249	CDKN2A (26)	TCAGCCGAAGGCTCCATGCTGCTCCC	pPrR_951
250	CDKN2A (26)	TCAGCCGAAGGCTCCATGCTGCTCCC	pPrR_952
251	CDKN2A (26)	TCAGCCGAAGGCTCCATGCTGCTCCC	pPrR_953
252	CDKN2A (26)	TCAGCCGAAGGCTCCATGCTGCTCCC	pPrR_954
253	ZAP-70 (26)	TGAGAAACCCTGGCGGGGTGTGACAT	pPrR_955
254	ZAP-70 (26)	TGAGAAACCCTGGCGGGGTGTGACAT	pPrR_956
255	ZAP-70 (26)	TGAGAAACCCTGGCGGGGTGTGACAT	pPrR_957
256	ZAP-70 (26)	TGAGAAACCCTGGCGGGGTGTGACAT	pPrR_958
257	CDKN2A (26)	TCAGCCGAAGGCTCCATGCTGCTCCC	pPrR_897
258	CDKN2A (26)	TCAGCCGAAGGCTCCATGCTGCTCCC	pPrR_898
259	CDKN2A (26)	TCAGCCGAAGGCTCCATGCTGCTCCC	pPrR_899
260	CDKN2A (26)	TCAGCCGAAGGCTCCATGCTGCTCCC	pPrR_900
261	CDKN2A (26)	TCAGCCGAAGGCTCCATGCTGCTCCC	pPrR5_901
262	CDKN2A (26)	TCAGCCGAAGGCTCCATGCTGCTCCC	pPrR_902
263	CDKN2A (26)	TCAGCCGAAGGCTCCATGCTGCTCCC	pPrR_903
264	ZAP-70 (26)	TGAGAAACCCTGGCGGGGTGTGACAT	pPrR_904
265	ZAP-70 (26)	TGAGAAACCCTGGCGGGGTGTGACAT	pPrR_905
266	ZAP-70 (26)	TGAGAAACCCTGGCGGGGTGTGACAT	pPrR_906
267	ZAP-70 (26)	TGAGAAACCCTGGCGGGGTGTGACAT	pPrR_907
268	ZAP-70 (26)	TGAGAAACCCTGGCGGGGTGTGACAT	pPrR_908
269	ZAP-70 (26)	TGAGAAACCCTGGCGGGGTGTGACAT	pPrR_909

270	ZAP-70 (26)	TGAGAAACCCTGGCGGGGTGTGACAT	pPrR_910
477	BRCA1 (26)	TCTTTCCTTTTACGTCATCCGGGGGC	pPrR_1198
478	CDKN2A (26)	TCAGCCGAAGGCTCCATGCTGCTCCC	pPrR_1199
479	MGMT (26)	TCCCAGCTTCCGCCTGAGGCTCTGTG	pPrR_1200

Table III-15: TALEs

B. Methods

1. Preparation of gels

Denaturing PAGE gel

For all Primer extension reactions, denaturing PAGE gel was prepared by adding 8 mL acrylamide solution (25%), 2 mL of 10 X TBE with 9 M urea, 10 mL of 9 M urea, 160 μ L of 10 % APS and 8 μ L of TEMED together and pouring the solution between the glass plates. The required combs were inserted and the set-up was left undisturbed till the gel polymerized. The combs were carefully removed and the set-up was immersed in PAGE running buffer and pre-run for 30 min at 120 V before use.

Agarose gel

Depending on the purpose, gels with different percentages of agarose were prepared. Agarose was weighed in a glass bottle and dissolved in 0.5 X TBE buffer (For a 1% gel, 1 g agarose gel is dissolved with 100 ml of buffer) by heating using microwave till the liquid was clear. Molten agarose was cooled and poured on a horizontal gel casting chamber with appropriate combs placed prior to adding the agarose solution.

SDS gel

SDS gels were prepared consisting of a lower resolving region (8%) and an upper stacking region (5%). After assembly of the gel casting apparatus, a solution was prepared by adding 3808 μ L milliQ water, 1092 μ L of 40 % acrylamide concentrate, 420 μ L of Tris-HCl (pH 8.8), 66.7 μ L of 10 % SDS solution, 66.7 μ L of 10% APS and 6.67 μ L of TEMED. The solution was mixed by shaking and poured in between the two glass plates and layered with isopropanol. After polymerization of the resolving gel, isopropanol was drained out and a solution containing 1000 μ L milliQ water, 208 μ L of 40 % acrylamide concentrate, 420 μ L of Tris-HCl (pH 6.8), 16.7 μ L of 10 % SDS solution, 16.7 μ L of 10% APS and 1.67 μ L of TEMED was prepared and poured on top. The combs were carefully inserted and the set-up was left undisturbed.

EMSA gel

Prepared using the same casting apparatus as the SDS gel, one EMSA gel (10 mL) was prepared by adding 7.4 mL water, 500 μ L of 10 X buffer TAE, 2 mL of 40% acrylamide gel,

100 μ L of 10% APS and 10 μ L of TEMED. The solution was slightly shaken and poured between the glass plates. Appropriate combs were inserted and the set-up was left undisturbed till the gel polymerized following which the combs were removed.

2. Preparation of antibiotic stocks

A 1000X stock of antibiotics were prepared by dissolving them in appropriate solvents (see antibiotics table) and passing the solution through a sterile filter using syringes. They were distributed in 1.5 mL tubes and stored at -20°C .

3. Preparation of LB-Agar plates

Sterile LB-Agar was heated in a microwave till it melted completely. It was then cooled down till RT and the required antibiotic/antibiotics at correct concentration (see list of chemicals) was added. Additionally, when required, 50 $\mu\text{g}/\text{mL}$ solution of X-gal () was also added. After thorough mixing, the liquid agar was poured into plastic petri dishes next to a flame. The plates were kept partly closed till the agar solidified. These plates were then immediately used for plating a culture or closed and stored at 4°C until further use.

4. Preparation of GH371 chemical competent cells

Overnight cultures prepared from single colony of the bacterial cell were diluted 100 times with fresh LB medium and incubated at 37°C , 200 rpm till the O.D. at 600 nm reached 0.5. The culture was cooled on ice centrifuged at 4°C at 4000 rpm for 10 min and hereafter the culture is always kept cold on ice. The supernatant was discarded and the pellet was resuspended in (1/10 volume of the culture) cold 100 mM MgCl_2 solution followed by another centrifugation at 4°C for 15 min. After discarding the supernatant, the pellet was resuspended in same volume of cold 50 mM CaCl_2 solution and centrifuged at 4°C for 15 min. After final centrifugation at similar conditions, the pellet was dissolved in (1/20 of the volume of the resuspension solutions) 50 mM CaCl_2 containing 15% glycerol. The entire

volume was distributed into 50 μL aliquots, snap frozen in liquid nitrogen and stored at -80°C .

5. Preparation of BL21 (DE3) Gold chemical competent cells

Overnight cultures prepared from single colony of the bacterial cell were diluted 160 times with fresh LB medium containing 100 $\mu\text{g}/\text{mL}$ of tetracycline and incubated at 28°C , 200 rpm till the O.D. at 600 nm reaches 0.5. The culture was cooled on ice and centrifuged at 4°C at 4000 rpm for 10 min and hereafter the culture is always kept cold on ice. The supernatant was discarded and the pellet was re-suspended in (1/4 volume of the culture) cold TB buffer and incubated for 10 min on ice followed by another centrifugation at 4°C for 10 min. After discarding the supernatant, the pellet was resuspended in (1/27 volume of the culture) cold TB buffer and (1/257 volume of the culture) DMSO and incubated for 10 min on ice. The entire volume was distributed into 100 μL aliquots, snap frozen in liquid nitrogen and stored at -80°C .

6. Transformation

In chemical competent cells by heat shock:

Cells stored at -80°C were thawed on ice and plasmids were mixed by careful pipetting. The samples were incubated on ice for 30 min following which it was exposed for 40 s to 42°C in a thermo mixer. The samples were immediately transferred back to ice for another 2-min following which 1 mL pre-warmed SOC was added to each tube. The tubes were kept at 37°C , shaking at 700 rpm in a thermo-mixer for 1 h. LB-Agar plates containing the required antibiotic were prepared in advance and desired volume of the culture was plated on them and spread using glass beads next to a flame. The plates were left open in an incubator till the liquid soaked in. The beads were removed; the plates were inverted and incubated overnight at 37°C .

In electrocompetent cells by electroporation:

Cells stored at -80°C were thawed on ice and plasmids were mixed by careful pipetting. The sample was transferred to a pre-cooled cuvette, placed in the electroporation device and the

cells were exposed to a 5 ms pulse of 1800 V. 1 ml of pre-warmed SOC was immediately added to the samples and from there it was handled in the same way as mentioned above.

7. Plasmid Isolation

5 mL overnight cultures were spun down at 4000 rpm for 15 min and the remnant media was discarded. Plasmid isolation was performed using a kit as per manufacturer's protocol. The pellet was completely dissolved in 250 μ L re-suspension buffer and transferred to a 1.5 mL tube. 250 μ L of lysis solution was added and mixed by inverting the tubes 5-6 times. 350 μ L of neutralization solution was added and mixed by inverting the tubes. Samples were centrifuged for 5 min at 12000 rpm. The clear solution was carefully transferred to columns provided in the kit and the columns were spun in the same condition for 1 min. The columns were washed two times using 500 μ L of wash buffer and centrifugation following which the columns were spun additionally to get rid of any trace liquid residue. DNA was eluted in a total of 50 μ L pre-heated water in two steps.

8. Agarose gel electrophoresis

Polymerised gel was transferred to the running chamber and samples containing 1 X loading dye was pipetted into the wells. 2 log ladders was loaded alongside as control to estimate the size of DNA in the sample and the gel was run on a horizontal agarose gel chamber at 70 V for 1.5 h. Gel was stained externally by immersing in ethidium bromide solution for 15 min and then destained in water for another 15 min. Chambers illuminated with UV light was used to analyse the gel (SI Figure VI-3). For extraction of DNA from agarose gels, UV protection helmets were used and correct size bands were cut out using scalpel and stored at -20°C until further processing.

9. DNA purification

Using a kit:

For purification of DNA from a reaction, binding buffer was added 1:1 (v/v) to the samples. If the desired DNA was smaller than 500 bp or bigger than 10 kb, equal volume (1:1:1) of 100% isopropanol or water was also added. Solution was then transferred to the provided columns and spinned down at 12000 rpm for 1 min. The columns were emptied and if the samples were to be sent for sequencing, 100 μ L of binding buffer was carefully added on the column bed and centrifuged for another minute. The columns were washed twice with 500 μ L wash buffer and after its removal another step of centrifugation was performed to get rid of any remnant liquid. DNA was eluted in 50 μ L of pre-warmed water in two steps.

Samples sliced from the gels were dissolved in equal volume (100 μ L for 100 mg gel) of binding buffer with or without isopropanol or water and treated as mentioned above.

Using Isoamyl alcohol chloroform phenol:

To each sample, equal volume of solution containing Phenol chloroform isoamylalcohol in a ratio of 25:24:1 was added and mixed by vortexing for 15 s. The samples were then centrifuged for 5 min at RT at 16000 rpm. Out of the two different phases obtained, the aqueous phase was collected in a separate tube and twice the volume of pre-chilled ethanol, 10% volume of 3 M sodium acetate was added and the samples were kept overnight at -20°C. The following day, samples were stored at -80°C for 1 h then centrifuged at 16000 rpm and 4°C for 30 min. Supernatant was removed without disturbing the pellet and 150 μ L of 70% ethanol was added. Following another centrifugation step at same conditions for 2 min and removing the supernatant, the dried pellets were diluted in water and stored at -20°C until further use.

10. DNA quantification

The concentration of DNA in any sample was measured using a spectrophotometer or a nano-drop. Cuvettes containing water was used as blank and consecutive measurement for the samples were carried out using fresh cuvettes. For measurement of samples with limited volume, 1 μ L of water was first added to the loading point in the nano-drop and the machine

was blanked. All the samples were then quantified with intermediate cleaning of the loading point.

11. DNA Sequencing

All sequencing samples were prepared according to the stated requirement of GATC. Approximately, 800 ng of the DNA to be sequenced was mixed with 2.5 nM of the respective primer in a total volume of 10 μ L and sent in a 1.5 mL tube with the barcode stuck on the tubes. The results were downloaded from the GATC website and viewed on Finch TV or snap gene.

12. Golden Gate (GG) assembly protocol.

TALE proteins used in this study was assembled using the GG assembly protocol described by Cermak et al [168]. This protocol involves systematic cloning of individual module plasmids to achieve TALEs of desired sequences and sizes in two steps. The first step involves designing sub-plasmids which are all combined in the next step in the final construct containing the complete sequence. Once the target DNA sequence is chosen, care has to be taken to avoid including the first T (T0) and the last base in the design of the first and the last sub-plasmid. While the repeat for T0 is pre-included in the plasmid containing the final sequence, the repeat for the last base is always added as an individual plasmid in the second step. Assembly of 12-31 repeats containing TALE is possible using this protocol, however there is slight variance depending on the desired length. Sequences containing 12-21 repeats use two sub-plasmids in the first step, FusA and FusB(n) (where n= total number of repeats - 11). The first ten repeats are included in a FusA vector and the remaining repeats (excluding the last repeat) are combined in respective FusB. For TALEs targeting DNA sequences >21 bases, three sub-plasmids are used in the first step. The first 10 repeats are included in plasmid FusA30A, the next ten in FusA30B and the remaining in FusB(n) (where n= total number of repeats - 21). In the next step, the sub-plasmids and the plasmid with the respective last repeat (pLr) are combined in the final plasmid to obtain the final construct used for expression of the designed TALEs.

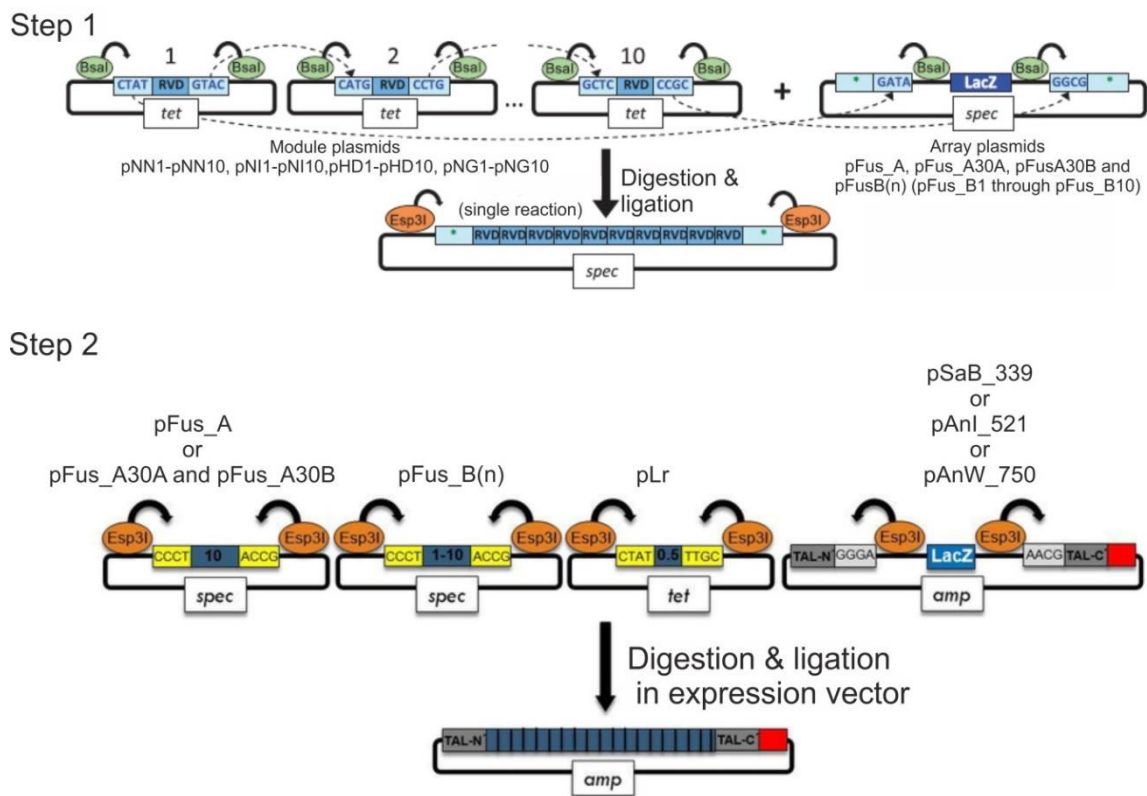


Figure III-1: Photographic depiction of the golden gate assembly protocol. Restriction ligation for creation of plasmids; using GG1 assembly/intermediate vectors (Step 1) and GG2 assembly/Final TALE containing vector (Step 2) [168].

The first step was selection of a target DNA sequence and size. Using the accepted recognition pairs of RVD and bases (NN for G, NI for A, NG for T and HD for C) and taking into account the above-mentioned rules, 0.5 μL of 150 $\text{ng}/\mu\text{L}$ of each repeat modules and respective array plasmids were first combined in a single restriction/ligation step containing units of restriction endonuclease BsaI and units of T4 DNA ligase in 1 X T4 DNA ligase buffer. This was called the Golden gate 1 (GG1) reaction (Figure III-1, Step 1) and was performed in a thermocycler using the following program: 10 cycles of 37°C for 5 min and 16°C for 10 min followed by 1 cycle each of 50°C for 5 min and 80°C for 5 min. The reaction was then stored at 4°C until further use. The generated plasmids were called the GG1 plasmids. 0.5 μL of plasmid safe containing 5 mM ATP was added to each GG1

reaction and incubated at 37°C for 1 h and at 70°C for 20 min. 5 µL of the reaction was then transformed into GH371 cells and 400 µL of the culture was plated on LB-Agar plates containing 100 mg/L Spectinomycin and 50 mg/L X-gal. The plates were inverted and incubated overnight at 37°C. Colony PCR was performed for white colonies by first suspending the colonies in water and performing PCR using the suspension along with 0.1 mM of each primer o734 and o735, 1 X ThermoPol buffer, 0.2 mM of dNTPs and Taq DNA polymerase using the program: one cycle of 94°C for 2 min, 35 cycles of 94°C for 30 s, 55°C for 20 s and 68°C for 2 min followed by one cycle of 68°C for 2 min. The reaction was loaded on 1% Agarose gel and the gel was run at 70 V for 1.5 h following which it was stained with Ethidium Bromide solution and analysed using UV light (Figure III-2). For the colonies showing the correct length on the gel, the suspension was inoculated with 5 ml LB cultures containing 100 mg/L Spectinomycin and the culture were incubated at 37°C, shaking at 180 rpm overnight. The correct colonies were also saved by spotting it on a fresh LB-Agar plate called the safe plate containing the correct antibiotic and storing the plate at 4°C. Plasmids were isolated the following day using the mentioned protocol and their concentrations were set at 150 ng/µL. Respective GG1 plasmids were pooled together along with the correct last repeat plasmid and the entry vector (pSaB_339 or pAnI_521) to perform the second step called the Golden Gate 2 (Figure III-1, Step 2). Reaction was performed like GG1 but with BsmBI (Esp31) restriction enzyme in the thermocycler as follows: 1 cycle each of 37°C for 10 min, 16°C for 15 min, 37°C for 15 min and 80°C for 5 min. For transformation, 5 µL of the reaction volume was used along with 50 µL of GH371 chemical competent cells and 400 µL of the LB culture was plated on LB-Agar plates containing 50 mg/L Carb. The next day colony PCR was performed as described above but with extension time of 4 min instead of 2 min. Colonies showing the correct size on Agarose gel (Figure III-3) was saved on a safe plate and overnight cultures was started with their suspension. Test digest was performed with GG2 final vector using AatII and StuI endonuclease to generate two fragments each bigger than 10 kbp (Figure III-4). Following isolation, approximately 500 to 1000 ng of the plasmids were sent for sequencing using 2.5 mM of primers o247, o890, and o315 as three different samples. 100 ng of plasmids displaying correct sequences were transformed in 50 µL of chemical competent BL21DE3 Gold cells and 100 µL of the inoculum was plated on a Carb containing LB-Agar plate and kept at 37°C overnight.

All the TALEs used in this study were assembled according to the GG assembly protocol based on *Xanthomonas axonopodis* scaffold and possesses a N-terminal TRX or GFP tag, a shortened AvrBs3-type sequence (136 amino acid preceding the first canonical repeat) and a C-terminus His6 tag (Figure I-17).

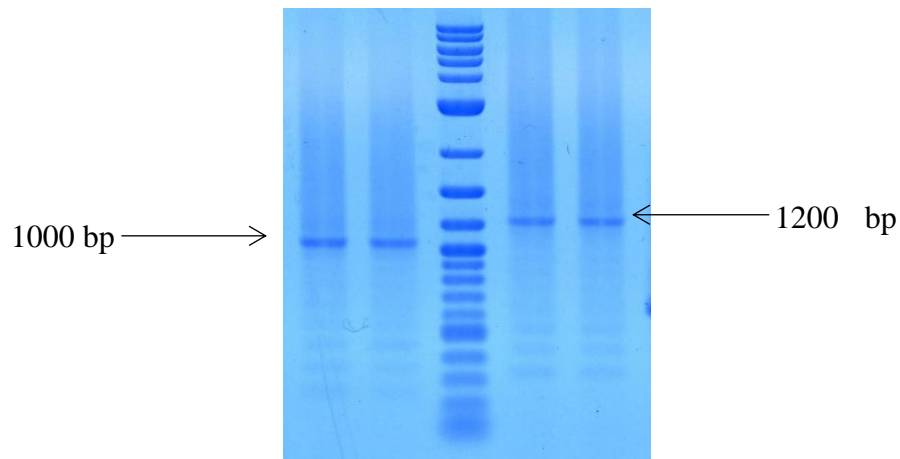


Figure III-2: Agarose gel showing the result of colony PCR of GG1 plasmids. Wells 1&2: Plasmid containing 8 RVDs. Well 3: 2-log ladder. Wells 4 and 5 Plasmid containing 10 RVDs

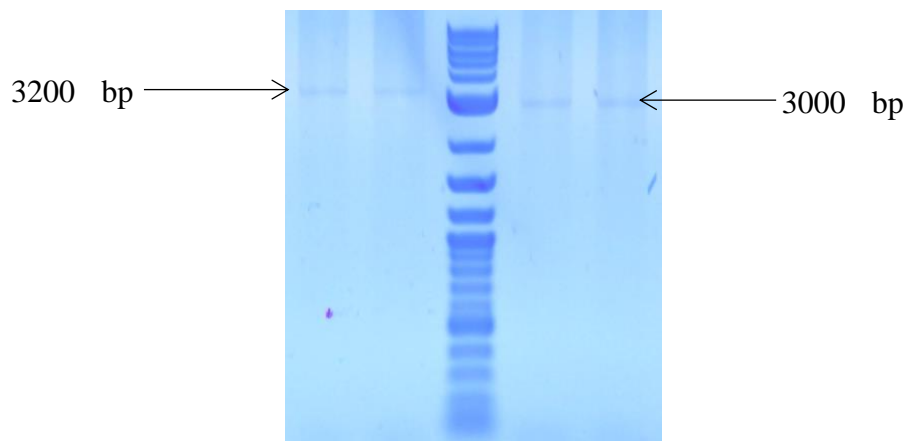


Figure III-3: Agarose gel showing the result of colony PCR of GG2 final vectors. Wells 1&2: TALE plasmid containing 30 RVDs. Well 3: 2-log ladder. Wells 4 and 5 TALE plasmids containing 26 RVDs

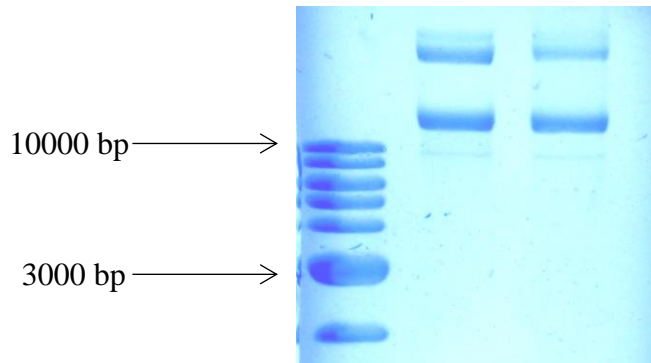


Figure III-4: Agarose gel showing the result of restriction digestion of GG2 final vector using restriction enzymes AflII and XbaI

13. Expression of TALE proteins

Overnight cultures from a single colony of plasmid transformed in *E. coli* BL21DE3 Gold cells were dissolved 50 times with fresh LB media containing 50 µg/mL of carbenicillin in an Erlenmeyer flask and the culture was grown at 37°C shaking till the OD at 600 nm reached 0.4. Expression was induced with 0.2 mM IPTG and incubated for additional 5 h. The culture was centrifuged at 4°C, 4000 rpm for 20 min. The pellets were either stored at -20°C until further purification.

14. Purification of TALE proteins

Freshly centrifuged or frozen pellets from expression cultures were dissolved in taq lysis buffer containing 50 µg/mL lysozyme and 1 mM PMSF. The solution was kept shaking at 100 rpm and RT for at least 30 min following which it was sonicated using the following cycle: The culture was then centrifuged for 30 min at 4°C and 9000 rpm.

- For TALEs with a TRX tags, the resultant supernatant was discarded and buffer Z was added to the pellet and mixed vigorously. Ni-NTA beads were then added and the solution was shaken for additional 3 h at RT. The solution was then loaded on purification column and the flow through was discarded. The resin was washed twice with 4XPBS containing 8 mM urea, 4 times with

buffer Z containing 20 mM imidazole and once with the buffer Z containing 50 mM imidazole. Finally the protein was eluted thrice in buffer containing 500 mM imidazole.

- For TALEs with a GFP tag, the supernatant was collected in a fresh falcon and mixed with Ni-NTA beads for atleast 3 h at 4°C. The solution was then loaded onto a purification column and washed twice with cold 4XPBS buffer containing 1 mM DTT, 4 times with Qiagen lysis buffer containing 20 mM imidazole and 1 mM DTT and once with Qiagen lysis buffer containing 50 mM imidazole and 1 mM DTT. The protein was finally eluted in Qiagen lysis buffer containing 500 mM imidazole and 1 mM DTT thrice.

Combined elution was packed in a dialysis membrane (10kDa molecular weight threshold) and dialysed in 1 X TALE storage buffer containing 1 mM DTT at 4°C. The dialysis buffer was changed till imidazole amount in the sample was reduced to picomolar range. The proteins were then collected, distributed in aliquots, snap frozen in liquid nitrogen and stored at -80°C until further use.

15. Quantification of TALE proteins

Bicinchoninic acid assay (BCA) was used to estimate the concentration of the proteins. The procedure was carried out according to manufacturer's protocol wherein a series of known concentration of a standard protein is plotted against the readout from the TECAN and the equation generated by the linear regression curve is used to calculate the concentration of unknown proteins.

16. SDS electrophoresis

For analysis of protein size and quality 1 X SDS-PAGE loading dye was added to each protein samples and they were heated at 95°C for 5 min. The samples were loaded on the SDS gel which was run in a mini protean vertical electrophoresis cell unit at 70 V and 30 mA for 1.5 h. The gel was carefully taken out from between the plates, stained in brilliant blue solution for 1 h and destained in destaining solution for 1 h (Figure III-5).

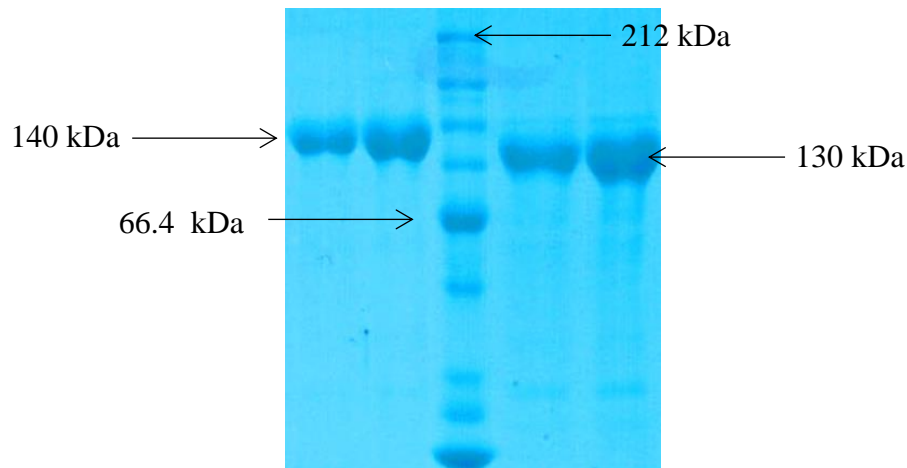


Figure III-5: SDS showing purified TALEs. Wells 1&2: Second and first elution of a TALE containing 30 RVDs. Well 3: Protein marker. Wells 4&5 Second and first elution of a TALE containing 26 RVDs.

17. Mutagenesis using quikchanges (Agilent)

For insertion deletion or change of upto 6 continuous bases in any DNA sequence, primers were designed according to supplier's protocol. Essentially, both the forward and reverse primers targeted the same region and were shifted only by 2 to 3 base pairs. The region in which the mutation was to be introduced was designed to be approximately at the centre of each primer. C/G was always selected as the beginning and end bases of the primers and their size ranged from 30 to 40 base pairs. Care was taken to have the melting temperature above 72°C. 10 ng of template, 0.4 µM of each primer, 1 X Pfu polymerase buffer and 2 units of Pfu polymerase were mixed in a final volume of 25 µL. The samples were run in a cyclor using the program: 95°C for 30 s followed by 20 cycles of 95°C for 30 s, 55°C for 30 s, 68°C for 8 min and a final extension at 72°C for 10 min. After addition of 1 µL of DpnI endonuclease, the reaction was incubated at 37°C for 1 h and at 80°C for 20 min. The samples were then transformed into GH371 cells. Next day 3 colonies were picked for o/v cultures and sent for sequencing as described.

18. Restriction ligation

For creating longer or discontinuous mutations in any vector, primers with restriction sites of particular endonucleases were used in a PCR reaction to introduce these sites on the sides of the region where the changed sequence was to be inserted. In the same time, complementary oligonucleotide bearing the desired mutation and sites for the same restriction enzyme as the vector were ordered. Oligonucleotides were hybridized in 1 X of buffer compatible with the endonucleases by heating at 95°C and incubating for 30 min at RT. Both the plasmid and the oligonucleotides were restricted in parallel using the respective endonucleases and were ligated o/v at 4°C using 10 units of T4 DNA ligase. The ligation product was transformed into GH371 cells and plated on agar plates containing the correct antibiotics. Next day 3 colonies were picked for o/v cultures and sent for sequencing as described.

19. Whole genome amplification (WGA)

Human gDNA (Yoruban male individual, encode entry NA18507) was obtained from Coriell Institute whereas the gDNA of Zebrafish (Zf) and bacteria was extracted from either the tissue of a single individual (Konstanz wildtype) or from the single clone of *E. coli* strain K12 ER2738. To obtain modification free gDNA, the WGA kit was used as per the supplier's protocol. buffer DLB was prepared in advance and buffers D1, N1 were prepared according to the sample size. To each 5 µL DNA sample (~100ng) in TE buffer, 5 µL of buffer D1 was added and mixed by vortexing and short centrifugation. The samples were incubated for 3 min at RT following which 10 µL of buffer N1 was added and mixed well. 30 µL of a master mix containing nuclease free water, REPLI-g mini reaction buffer and REPLI-g mini DNA polymerase was then added and after mixing the samples were incubated for 16 h at 30°C. For inactivation of the polymerase, the samples were heated at 65°C for 3 min and the samples were purified using QIAmp DNA mini kit (see Methods DNA extraction and cleaning using QIAmp DNA mini kit).

20. DNA extraction and/or purification, QIAmp mini DNA kit

For extraction and purification of DNA, 200 μL of the samples were mixed with 20 μL of proteinase K solution and 200 μL of buffer AL and mixed by pulse-vortexing for 15 s followed by incubation at 56°C for 10 min. 200 μL of 100 % ethanol was added and mixed thoroughly and the entire volume was transferred to the column provided in the kit. Centrifugation was performed at RT, 8000 rpm for 1 min and the flow-through was discarded. The columns were washed using 500 μL of each buffer AW1 and AW2 by centrifugation at 8000 rpm for 1 or 3 min respectively. DNA samples were eluted in 50 μL of pre-warmed water in two steps.

21. Random shearing gDNA

To obtain gDNA sample containing different size distribution, the WGA samples were addressed to random shearing by sonication. Samples contained in a 2-ml tube were placed in a flask containing ice to avoid over heating of the samples. To avoid splashing of the samples, the tube opening was covered by paraffin film. Probe was introduced into the tube by making a small hole on the film. 20 cycles of 30 s on/ 30 s off pulse at 20 % power resulted in a broad fragment size range of 100 – 1000 bp as evident by visualization on 1 % agarose gel (SI Figure VI-2).

22. Enzymatic methylation using CpG methyltransferase

To perform enrichment with genomic samples containing uniformly methylated CpGs, a part of the randomly sheared gDNA was treated with 20 units of M. SssI in a 200 μL reaction volume containing 1 X NEB buffer 2 and 640 μM of the S-adenosylmethionine. Control samples (non-methylated) were treated in parallel without the enzyme. The samples were incubated at 37°C for 14 h in total including replenishment with 4 μL of 32 μM SAM after 4 h. To verify the efficiency of methylation, both non-methylated and methylated samples were treated with sodium bisulphite as described below.

23. Bisulphite conversion and PCR

Deamination of DNA samples using sodium bisulphite was performed as per manufacturer's protocol. In short, DNA (100ng) in 20 μ L of water was mixed with 35 μ L of DNA protect buffer and 85 μ L of freshly dissolved sodium bisulphite solution. The tubes containing the reaction were then placed in a PCR cycler and the following program was performed: (Denaturation at 95°C for 5 min, incubation at 60°C for 25 min, 95°C for 5 min, 60°C for 85 min, 95°C for 5 min, 60°C for 175 min and 20°C overnight). The reaction was immediately cleaned using the buffers and instructions provided in the kit. 560 μ L of freshly prepared buffer BL was added and mixed by vortexing. The samples were briefly centrifuged followed by transferring the entire volume to a column provided in the kit and spinned at 14000 rpm for 1 min. The columns were washed once with 500 μ L of buffer BW incubated for 15 min with buffer BD followed by washing twice with 500 μ L of buffer BW and centrifuged empty to get rid of any liquid. DNA was eluted in 50 μ L elution buffer and stored at -20°C until further use. Bisulphite specific primers were often used in nested PCR reactions where the first PCR reaction is performed using the converted DNA while the following reactions used the product generated in prior PCR reactions as templates. For higher sensitivity in primer binding, the master mix from the EpiTect MSP kit containing HotStar Taq d-Tect Polymerase was used with template and primers. The PCR reaction were performed as follows: (Activation at 95°C for 10 min followed by 40 cycles of denaturing at 95°C for 15 s, annealing at 55°C for 30 s and extension at 72°C for 30 s. A final extension step was performed at 72°C for 10 min after which the samples were stored at 4°C). The final reaction product was cleaned and sent for sequencing.

24. Sodium borohydride mediated reduction of fC to hmC

For discriminating between sequences containing a single hmC or fC, 50 μ L samples were treated by addition of 150 μ L freshly prepared 1 M solution of the reducing agent; sodium borohydride while continuous shaking for 1 h at RT in dark with mixing, briefly centrifuging and opening the tubes for release of hydrogen every 15 min. The reaction was quenched using 600 μ L of 1 M sodium acetate (pH 5) for 10 min at RT followed by purification using DNA purification kit. Both treated samples and untreated samples were used in parallel in an enrichment assay using TALEs bearing N* RVD targeting the modified C.

25. Primer Extension (PEX)

PEX assay was developed and established in our group by Dr. Grzegorz Kubik [12]

5' γ -³²P-ATP labelling

Prior to all PEX reactions, the reverse primers were labelled at the 5' with ³²P by incubating 10 pmol of the primer with 1 μ L of γ -³²P-ATP and 1 μ L of T4 polynucleotide kinase in 1 X kinase buffer at 37°C for 1 h and then heat inactivating the enzyme at 75°C for 10 min. The samples were cleaned using G-25 columns that were prepared in advance by vortexing the resin, removing the bottom closure, opening the column cap slightly and centrifuging at 3200 rpm for 1 min. Samples were loaded at the centre of the resin bed and centrifuged. The collected 50 μ L samples were diluted with 20 μ L of respective unlabelled primer (10 μ M) to obtain 70 μ L of 3 μ M solution. Handling of ³²P containing samples were always undertaken behind radiation protective glass and wearing appropriate lab coat, gloves, eye protection gear and a measuring device which was investigated and refreshed monthly.

Reaction

Pair consisting of a longer forward and a shorter reverse oligonucleotide were hybridized in 2 X hybridization buffer A by heating at 95°C for 5 min and cooling for 30 min at RT to obtain a final concentration of 33.3 nM ds DNA sequence consisting the respective TALE target sequence. 3 μ L of this template was incubated with 3 μ L of TALEs in 1 X Tale storage buffer for 30 min at RT. 6 μ L of a mix containing 1 X buffer hybridization buffer B, 200 μ M of dNTP mix and different units of Klenow Fragment was added to these samples and incubated further for 15 min following which 12 μ L of PAGE loading buffer was added to each sample and heated for 5 min at 95°. The samples were cooled on ice and loaded on a pre-run gel with. The gel was run in 1 X TBE at 120 V for 1 h and was transferred to a Wattmann paper which was dried for atleast 1 h at 80°C. Later the gels were placed overnight on Phospho Imaging screen (IP) and after the gel was removed, the IP was read on a phosphor imager (Figure III-7) and analysed using software quantity one.

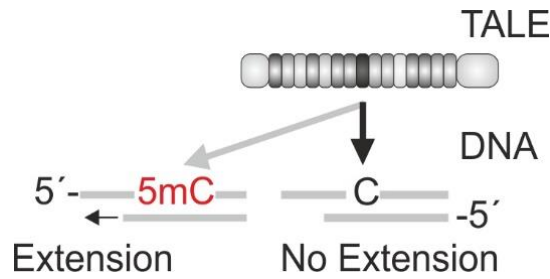


Figure III-6: Schematic representation of the PEX assay. TALE binding is inhibited by presence of 5mC but not C in the target sequence. Extension by polymerase is disrupted in TALE bound samples [19].

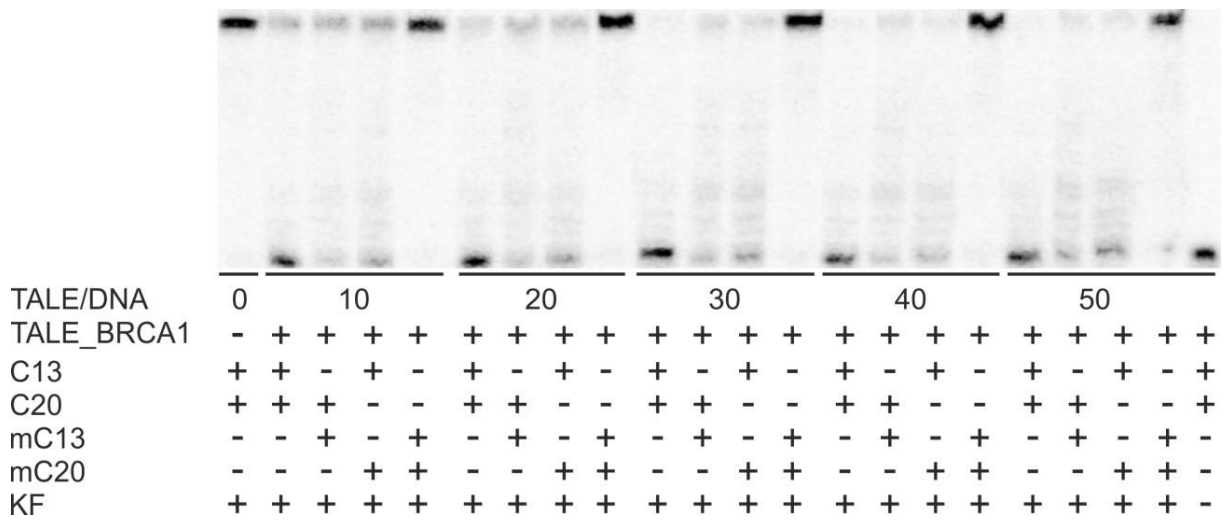


Figure III-7: Example of a PAGE gel containing samples from a PEX assay. The first well contains sample with no TALE and the last well contains sample with no polymerase. DNA, TALE samples and their ratios are mentioned below. PEX was conducted using fw oligos o1077 (C13 C20), o1184 (C13 mC20), o1185 (mC13 C20) and o1076 (mC13 mC20) and the common radiolabeled rv primer o1078.

26. Affinity enrichment using TALEs

Using Ni-NTA magnetic beads

In a 1.5 mL tube, desired gDNA was mixed with calculated quantity of respective TALE protein, 1.2 μL of 10 X buffer X with 5mM DTT added freshly and nuclease free water in a final volume of 12 μL . The samples were kept for 30 min at 25°C and on ice for 30 min. Meanwhile Ni-NTA magnetic beads were prepared by pipetting the required amount of beads in a fresh 1.5 mL tube and placing it on a magnetic stand. Entire volume of the storage liquid was removed and the same volume of buffer X without DTT was added. The beads were also stored on ice until further use. After incubation, 5 μL of the beads and 483 μL of pre-cooled buffer X with fresh 0.5 mM DTT were added to each tube and the tubes were incubated at 4°C and 700 rpm shaking for 1 h. All the tubes were placed on the magnetic stand and after 2 min the liquid was removed carefully without disturbing the beads. 500 μL of cold buffer X was added and the beads were washed carefully by pipetting up and down. The beads were allowed to settle for additional 2 min. Buffer X was completely removed and 500 μL of nuclease free water was added and the samples were heated and shaken at 95°C and 1400 rpm for 5 min. The tubes were allowed to cool and placed on the magnetic rack. Liquid from each tube was collected carefully in a fresh 1.5 mL tube and were dried using a vacuum drier at RT for 3 h. 50 μL of nuclease free water was added to each tube, mixed and centrifuged briefly. 7.5 μL of these samples were then used in a qPCR reaction.

Spike-ins were prepared in advance (see list of oligonucleotides) via PCR with respective primers and purified via agarose gel extraction. Background gDNA to be used was also restricted using appropriate restriction endonuclease (see results) and cleaned using PCR purification.

Using non-magnetic Ni-NTA resin

Experiments using non-magnetic beads were conducted like the above described method, however instead of 5 μL of magnetic beads, 50 μL of Ni-NTA resin (diluted 20 times with buffer X) was added to the samples and the samples were loaded on a spin column and spinned for 1 min at 1000 rpm. The beads in the column were washed as described earlier and for the final step, the beads were eluted in 500 μL of water and the tube was heated at 95°C for 5 min at 1400 rpm and cooled down at RT. The bead solution was re-loaded into the same columns and the spinning step was repeated. Hereafter samples were treated in the

same way as described earlier. For crosslinking experiments, following the incubation of TALE and DNA for 1 h and addition of 353 μL of buffer X, 5 μL of formaldehyde (0.37%) was added to the samples and they were shaken at RT for 10 min. 125 μL of 1M glycine solution was added to each tube and the sample was shaken for another 5 min. Samples were then treated the same way as described earlier.

27. MeDIP

MeDIP assay was performed parallel to enrichment assay using TALEs for samples bearing varying levels of 5mC. Protocol was followed as per provider's instructions. In short, samples to be enriched were prepared and separately mixed with antibody mix which contained positive (methylated) and negative (non-methylated) DNA controls. After treatment and washing in the recommended condition, samples were used for conducting qPCR using the control primers provided in the kit as well as the primers designed for the samples. The controls indicated the immunoprecipitation efficiency and the quantity of the target samples enriched were calculated similar to the enrichment assay as described below.

28. qPCR

All reaction components were collected under the sterile air-flow hood. The reactions contained 12.5 μL (1 X) of qPCR master mix from Promega and 7.5 μL of the template. Primers were added to a final concentration of 0.2 μM (5 μL). All the components were added in a 384 well plate and RT-PCR was performed on CFX-384 TOUCH 1000 cycler. For each primer pair (p.p.), a linear curve was obtained upon plotting Ct value obtained v/s concentration of template. The equations generated were then used to determine the concentration of the target in experimental samples. The curves generated by templates differing only in C modifications (C, 5mC, 5hmC, 5fC, 5CaC and 4mC) were similar.

29. Electromobility Shift Assay

The oligonucleotide pairs consisting of a forward primer bearing the target sequence and the reverse primer with or without a Cy5 fluorescence tag were hybridized in 2 X hybridization buffer A by heating at 95°C for 5 min and kept at RT for 30 min. 500 nM of the hybrid was then incubated with 100 nM of respective TALEs in 1 X TALE storage buffer in a final volume of 10 μ L. The reaction was kept for 30 min at RT followed by incubation on ice for another 30 min. The EMSA gel was pre-run at 70 V and 4°C for 30 min in a Mini Protean vertical electrophoresis cell unit. Following addition of 5 μ L of sample into each well, the gel was run for another 90 min in the same conditions. The intact glass plates were placed in a Typhoon FLA-9500 and the fluorescence for samples containing the Cy5 tag, emission was read using 647 nm band filters and for those containing no additional tag was read at 525 nm. Results were analysed using the software Image J. TALE-DNA complex formation was calculated by building the ratios of the shifted band (TALE-DNA complex) and the complete area of the same lane.

30. Fluorescence Resonance Energy Transfer

FRET assay was first applied in our group by Sara Maurer. It was optimized and used for determination of inhibition constant (K_i) for TALEs as mentioned in the publications [19].

Templates containing the TALE binding site and differing only on the modifications on the targeted C were hybridized with a common reverse primer covering the target sequence and containing a 5'Cy5 and a 3'Cy3 fluorescent tags were hybridized in hybridization buffer A by heating the mix at 95°C for 5 min and incubating at RT for 30 min. In a 384 well plate 3 μ L of the hybridized DNA (200 nM) was mixed with different concentrations of respective TALE proteins in 3 μ L of 1 X TALE storage buffer and the mixture was allowed to stand at RT for 30 min. 6 μ L of a solution containing 2 X Dnase I buffer and 1 U of Dnase I was added to each well and the plate was shortly spun and immediately placed in a TECAN M1000 plate reader which was pre-heated at 37°C. Fluorescence by Cy5 emission was measured at 665 nm for every 5 min. Cy5 fluorescence values of a control containing no TALE was subtracted from the data and the ratio of Cy5 fluorescence of samples with TALEs to that of control without Dnase I was plotted as relative Cy5 fluorescence. In K_i measurements, data were analysed on the software Origin using Hill fit.

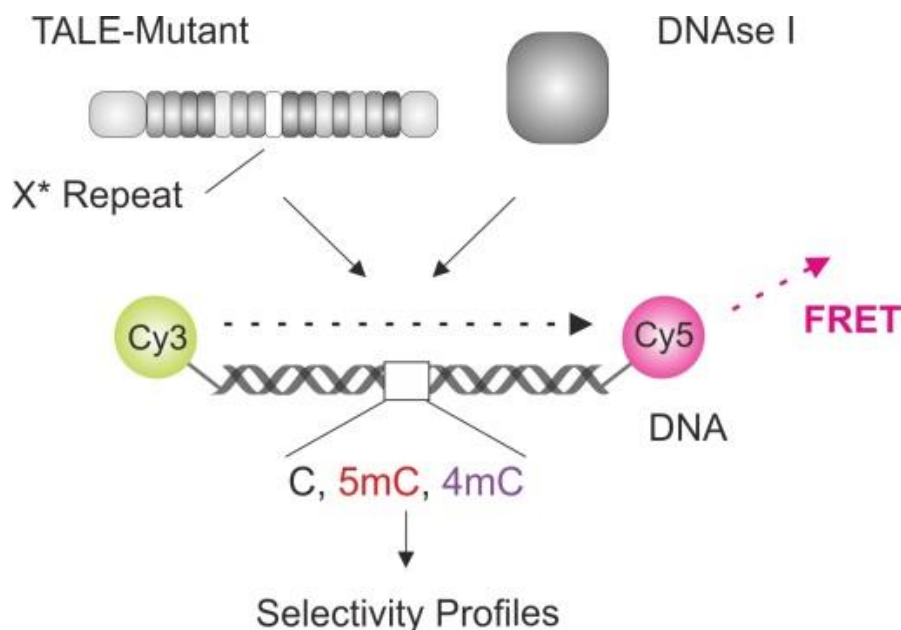


Figure III-8: Pictorial representation of the set-up used for screening library of engineered TALEs using FRET with DNA containing samples bearing either C, 5mC or 4mC. The fluorophore Cy5 is excited (552 nm) by emission of Cy3 only when they lie in close proximity uncut by DnaseI. Emission by Cy5 at 665 nm was recorded [19].

31. MALDI

Detection of modified bases on oligonucleotides was first established in our group by Mario Giess.

Oligonucleotides o476, o465 and o1236 each containing the Hey2_b binding site and either a C, 5mC or 4mC on the C of the CG respectively, were hybridized with o1152 in separate reactions in 1 X NEB CutSmart buffer by heating at 95°C for 5 min followed by incubation at RT for 30 min. Each hybrid was then restricted by addition of 10 units of SacI-HF and incubating at 37°C for 4 h followed by heat deactivation of the enzyme for 20 min at 65°C. DNA was purified using Phenol/Chloroform extraction followed by ethanol precipitation with 10% v/v of 3 M NaOAc (pH=5.2). The pellets were dissolved in 10 µL of aqueous solution containing 1% v/v Acetonitrile and 0.1% v/v TFA_{aq}. Zip-tips were wetted with Acetonitrile and equilibrated with 0.1% v/v TFA_{aq}. After equilibration, the sample was bound to the tip resin by repeatedly pipetting up and down. The bound DNA was washed twice with 0.1% v/v TFA_{aq} and then co-eluted with the matrix solution containing 50 mg/mL

HPA and 10 mg/mL D-ammonium citrate in 50% v/v acetonitrile_{aq}. The samples were spotted on the MALDI measuring plate and left to dry. Masses were measured on MALDI-TOF (Bruker) using linear positive mode and analysed using the flexAnalysis software.

32. Luciferase assay

Luciferase assay was first established in our group by Anna Witte. It was optimized and adapted for conducting methylation sensitive *in vivo* transcription activation by TALEs as mentioned in the publication [21].

Plasmids pPrR814synC and pPrR814synmC bearing a single C or mC in the TALE binding site were constructed as follows: Oligo pairs oAnW2016/oAnW2017 and oPrR2248/oAnW2017 were hybridized (4 μ M each) in 1 X CutSmart buffer by heating at 95 °C for 5 min and incubating at RT for 30 min. Plasmid pAnW755 and the hybridized oligonucleotide duplexes containing or not containing the single mC position were digested with each 10 units of *Sall* HF and *SpeI* and purified using a PCR purification kit. Plasmids p814synC and p814synmC were constructed by ligation of the digested pAnW755 with the digested oligonucleotide duplexes using T4 DNA ligase. Ligation reactions were purified via PCR purification and quantified by UV absorption measurement at 260 nm on nano-drop 2000. HEK293T cells were maintained in DMEM media supplemented with 1 % penicillin/streptomycin, 10 % FBS and 1% L-Glutamine. 10^4 cells were cultured in a 96 well plate overnight prior transfection. Opti MEM and lipofectamin 2000 were mixed according to the manufacturer's protocol. 25 ng of plasmid from group A (encoding a TALE binding site and a minCMV promoter upstream of a firefly luciferase gene) and 175 ng of plasmid from group B (encoding the TALE-VP64 fusion constructs; for plasmid maps, see the SI Figure VI-20) were pipetted in pairs as follows: pAnW755/pAnW818, pAnW814/pAnW818, pAnW814/pAnW893, pPrR814synC/p818, pPrR814synmC/pAnW818, pPrR814synC/pAnW893, pPrR814synmC/pAnW893, pPrR814synC/pPrR1149 and pPrR814synmC/pPrR1149. Transfection mix was added to each plasmid pair and incubated for 20 min at RT. The solution was added to wells of the 96 well plate and incubated at 37 °C and 5% CO₂ for 48 hr. Each well was then washed with 20 μ l of DPBS and the supernatant removed. 40 μ l of lysis buffer was added to each well and mixed vigorously. The plate was then incubated on ice for 20 min. After incubation, 20 μ L of the lysis solution from each well was combined with each 90 μ L of Bright-Glo in a

second 96 well plate. The plate was quickly spun down and the luminescence was immediately measured on a TECAN M1000 plate reader (wavelength 380-600 nM). Ratio of luminescence from each sample to that of a sample transfected with TALE_CDKN2A (WT) and a nonmethylated luciferase reporter plasmid pAnW814 (that had not been ligated with a synthetic DNA oligonucleotide duplex) was plotted as relative luminescence. The error bars represent standard errors from triplicate experiments. For analysing the integrity of the synthetically introduced 5mC position in plasmids over the duration of the experiment, the samples were collected after 48 h of incubation. DNA was extracted with a QiaAmp DNA mini kit and converted with bisulphite. Nested PCR was performed on these samples with p.p. oPrR2425/oPrR2426 followed by oPrR2424/oPrR2427 using the Epiect MSP kit. The PCR product was purified by agarose gel electrophoresis, extracted with a PCR purification kit and sequenced with primer oPrR2427 (SI Figure VI-21).

IV. RESULTS AND DISCUSSION

The ability of TALE repeats to bind differently to its target sequence containing either canonical nucleobases or their epigenetic counterparts is already established [12–15]. The otherwise strong hydrogen bond between TALE RVD HD and DNA nucleobase C can be perturbed to varying extent depending on the epigenetic modification present (5mC, 5hmC, 5fC, 5caC or 4mC) (Figure I-18). This effect is evident by lower binding or reduced preference towards sequences bearing these marks. Although methylation of C has been observed to occur universally, in eukaryotic mature cells it is found almost exclusively in CpG context. To test various factors that could affect the overall binding strength, sequence specificity and sensitivity towards epigenetic modifications, TALEs containing RVD HD targeting canonical or modified C were designed for sequences occurring in different genomes, of different lengths or bearing different number of CGIs and their function for detection of different epigenetic C nucleobase was observed in various assays.

All the TALEs designed and used in this study were first tested for specific binding and sensitivity using the primer extension experiments previously established and described by Dr. Grzegorz Kubik (Figure III-6) [12]. Degree of TALEs binding to synthetic oligonucleotides bearing either no, one or more 5mC on one strand and hybridized to ³²P labelled shorter reverse primer was observed based on the amount of completely extended ds oligonucleotide generated due to the action of KF (polymerase) on TALE bound/unbound, partially dsDNA sequence. A ratio of extended to non-extended product was indicative of the extent to which TALEs could compete with KF to bind to a sequence. All the ratios were normalized to control samples containing no TALE and showing complete extension (assigned a value 1). A clear band of high intensity with samples containing one or more 5mC and a comparatively faint or no band in ones with no 5mC indicated that TALE RVD HD bound stronger to target sequence with C compared to those with any modification (Figure III-7, SI Figure VI-1).

For affinity enrichment assay, qPCR reactions for Zf loci Hey2_a, Hey2_b, Hey2_c was conducted using p.p. o903/o904 while those for human genomic loci BRCA1, BRCA1_a, CDKN2A, MGMT, ZAP-70 were conducted using the p.p. o1060/o1114, o1570/o1577, o1064/o1065, o1457/o1459 and o1759/o1764 respectively.

A. Isolation of specific gDNA sequences using TALEs

Establishment of successful TALE based affinity enrichment of desired genomic sequences required various optimization. Experiments were conducted with different Ni-NTA beads, buffers, TALE and DNA concentrations, position of His6 tag in TALE protein, different size of TALE protein and in combination with cross-linking using Formaldehyde (HCHO) (SI Figure VI-4).

The first step in designing of an assay using solid phase bound TALEs to enrich specific sequences from a pool of gDNA was to test for off-target sequence binding by TALEs. TALE proteins targeting sequences in *Zf* genome (Hey2_a, Hey2_b and Hey2_c) (Figure IV-1) were expressed and used individually with 200 ng of randomly sheared, non-methylated gDNA (SI Figure VI-2). The TALE-DNA complex was then immobilized by magnetic Ni-NTA beads binding of the TALE His6 tag. Following washing and separation from the protein, the enriched gDNA sequence was quantified using data from qPCR with respective p.p. that recognized regions close to the target sequence.

As off-target gDNA controls, these TALEs were used with randomly sheared human gDNA (SI Figure VI-3) and enrichment of promoter sequence of two unrelated loci (BRCA1 and MGMT) (Figure IV-2) was observed. TALEs targeting the mentioned human loci were also tested for enrichment of Hey2_b sequence from *Zf* gDNA.

While on-target sequences were enriched in a range of 600 to 1000 copies, isolation of only negligible quantities of off-target sequences were observed by all TALEs. TALE_MGMT was seen to enrich slightly higher copies of Hey2 sequences. This was considered an artefact owing to similarity between the two target sequences (Figure IV-5). To ensure that sequences were enriched solely bound to TALE protein, control with no TALE was used in parallel. qPCR based quantification indicated non-significant amounts of target DNA sequences in samples without TALE proteins (Figure IV-4).

D. rerio

HEY2a: 5'-TACTGCTGCTCC**CG**CTGCTC-3'
 HEY2b: 5'-TCTT**CG**TTTCCACATCC-3'
 HEY2c: 5'-TTTACTGCTGCTCC**CG**CT-3'

nt: 0 5 10 15 20

Figure IV-1: TALE target sequences in Zf Hey2 loci as mentioned on the left. Sequences are always displayed from 5' to 3' direction. CpG sites are indicated in bold and underlined. The scale under the sequence indicates the length of the target sequence in bp [17].

H. sapiens

BRCA1: 5'-TCTTTCCTTTTA**CG**TCATC**CG**GGGGC-3'
 MGMT: 5'-TCCCAGCTTC**CG**CCTGAGGCTCTGTG-3'
 CDKN2A: 5'-TCAGC**CG**AAGGCTCCATGCTGCTCCC-3'
 ZAP-70: 5'-TGAGAAACCCTGG**CG**GGGTGTGACAT-3'
 BRCA1_a: 5'-TGCCCC**CG**GATGACGTAAAAGGAAAG-3'

nt: 0 5 10 15 20 25

Figure IV-2: TALE target sequences in human genome. Loci are mentioned on the left. Sequences are always displayed from 5' to 3' direction. CpG sites are indicated in bold and underlined. The scale under the sequence indicates the length of the target sequence in bp [17].

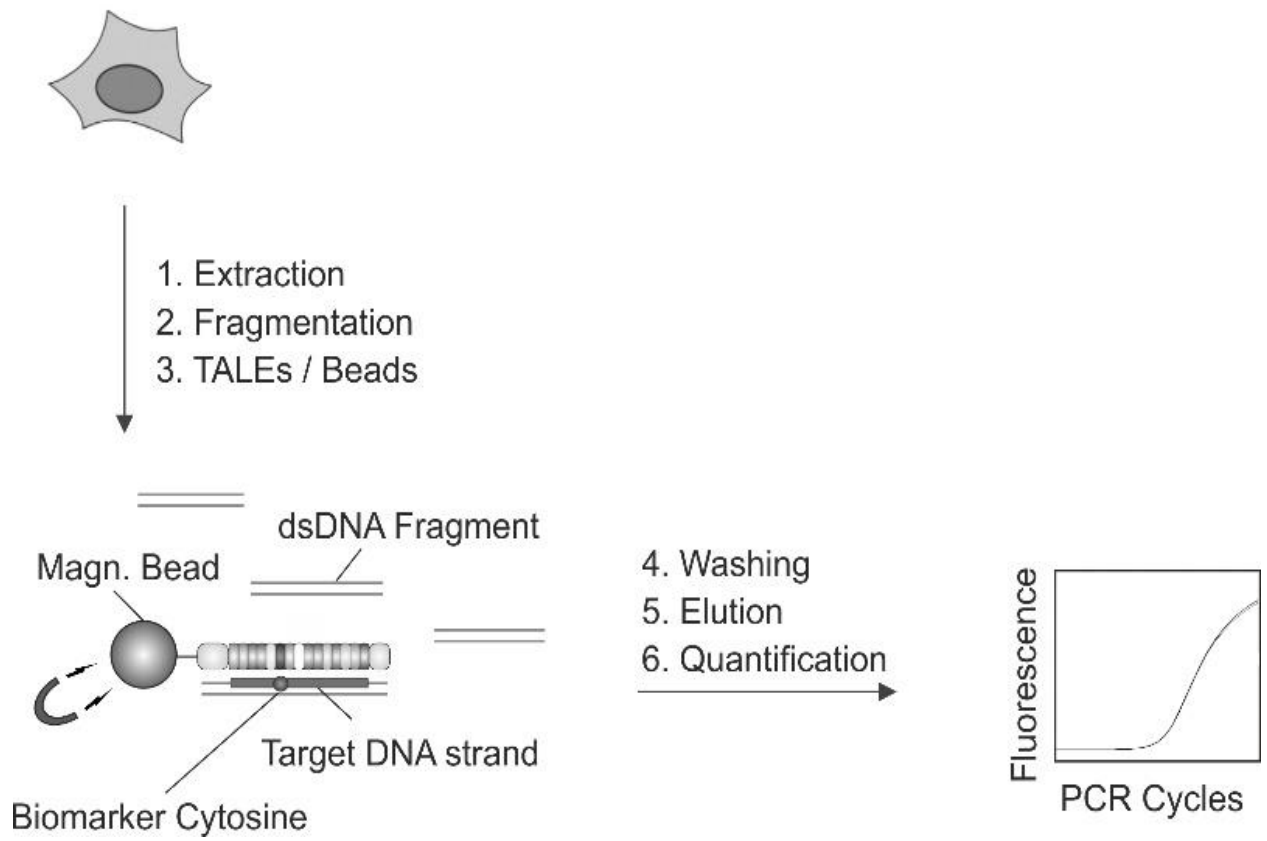


Figure IV-3: Pictorial representation of TALE based affinity enrichment assay [17].

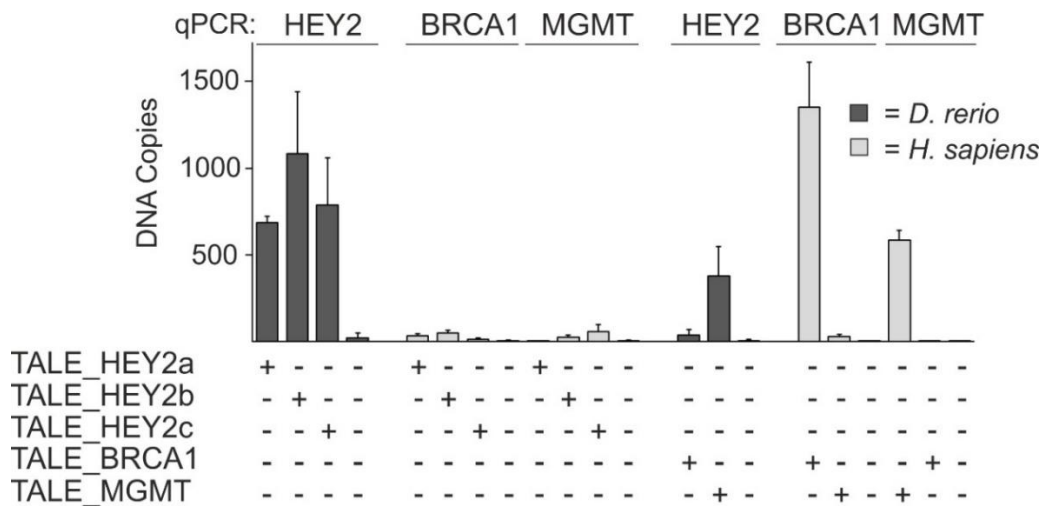


Figure IV-4: Result indicating enrichment of Zf (dark grey) and human gDNA sequences (light grey) using one target and several off-target TALEs. Y-axis depicts the quantity of gDNA enriched being bound to TALEs mentioned below, calculated as per data from qPCR performed using p.p. for loci mentioned above. Error bars indicate standard error from experimental and qPCR duplicates [17].

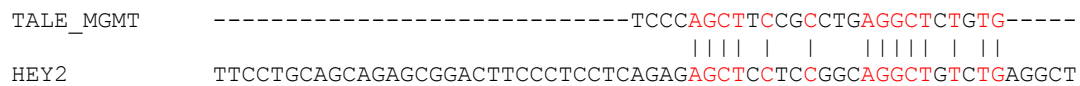


Figure IV-5: Similarity between TALE target sequence in human MGMT and Zf Hey2 loci [17].

B. Sequence specific detection of 5mC in gDNA using TALEs

Following isolation of specific target sequences from gDNA with little or no non-specific binding activity, the next focus was to use affinity enrichment assay for detection of 5mC in gDNA samples. This meant using TALEs to target specific sequence and RVD HD in TALEs to differentiate between C and 5mC. To achieve this, WGA and randomly sheared Zf gDNA was treated with CpG methyltransferase (M. SssI) to homogeneously add methyl groups on all C of CGIs. Both methylated and non-methylated gDNA was then used separately but simultaneously with the same TALEs targeting sequences harbouring the CpG site of interest and affinity enrichment was conducted as described earlier (Figure I-6).

Isolation of significantly higher quantities of non-methylated target sequence was observed using three TALEs targeting three distinct sequences in Zf Hey2 locus (Figure IV-1). Before being used for the assay, qPCR was performed with dilution series of both non-methylated and M. SssI treated WGA gDNA samples. Both type of samples generated similar threshold cycle (Ct values) and linear regression curves. Although all TALEs bound better to non-methylated DNA, TALE_Hey2_a exhibited the highest difference in binding between C/5mC containing target sequence (Figure IV-7).

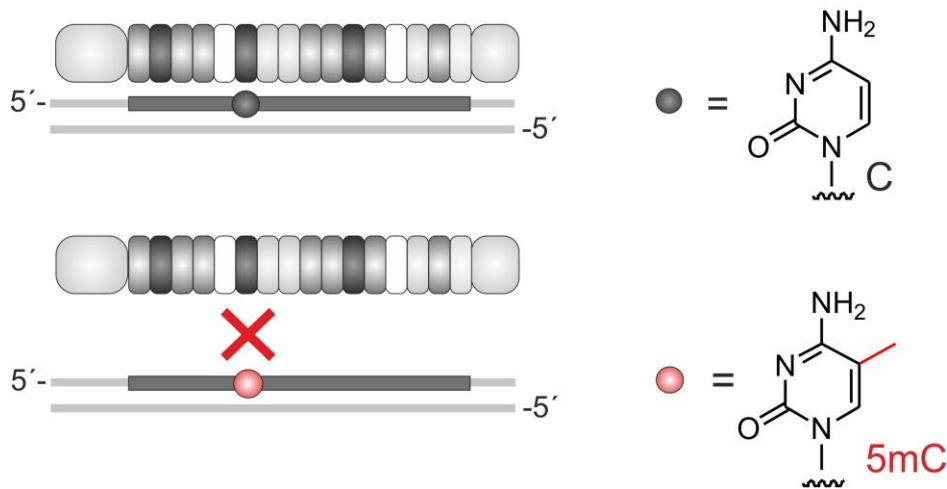


Figure IV-6: Schematic representation of principle behind differential binding of TALEs to C or 5mC containing samples. Comparatively lower binding to 5mC containing samples in affinity enrichment is evident by a higher Ct value or lower amount of DNA enriched [17].

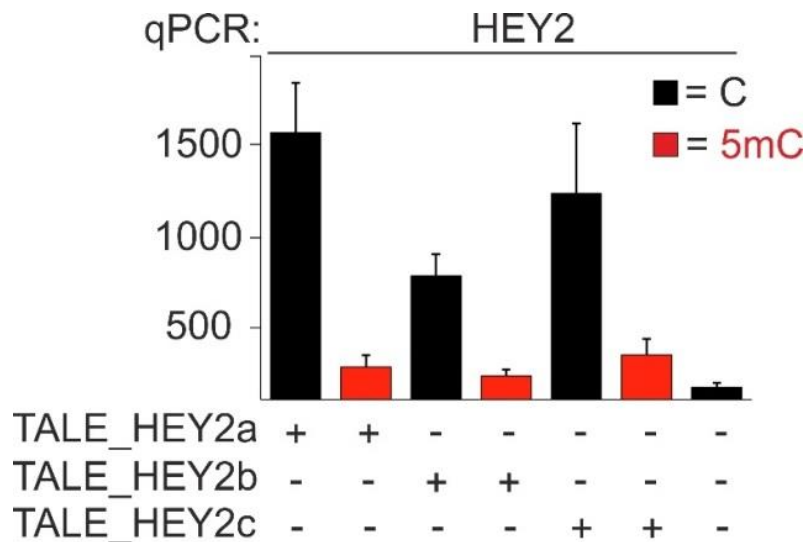


Figure IV-7: Result indicating enrichment of target sequence from *M. SssI* treated (red) or untreated (black) *Zf* gDNA sample using TALEs mentioned below. Each target sequence contained one CpG (Figure IV-1). Error bars indicate standard error from experimental and qPCR duplicates [17].

An Analogous assay was conducted to observe the isolation of C or 5mC containing target sequence in four different cancer biomarkers (Figure IV-2). Despite differing in extent to distinguish between C/5mC, all the candidates displayed direct sequence isolation in combination with 5mC selectivity. TALE_CDKN2A exhibited both highest enrichment and sensitivity while TALE_MGMT distinguished least between C/5mC owing probably to an overall lower enrichment efficiency (Figure IV-8). While all these proteins targeted 26 bp sequences, the placement of the target CpG was different. Also, unlike the other three, BRCA1 target sequence had 2 CpG sites in the target sequence (Figure IV-2).

To ensure the stability of the 5mC in the target sequence, sodium bisulphite based deamination was performed for both *M. SssI* treated and non-treated human gDNA samples. Sequencing with a reverse primer specific for BRCA1 target site showed replacement of G with A only in the non-treated samples which is consistent with 5mC being unaffected by deamination (SI Figure VI-8).

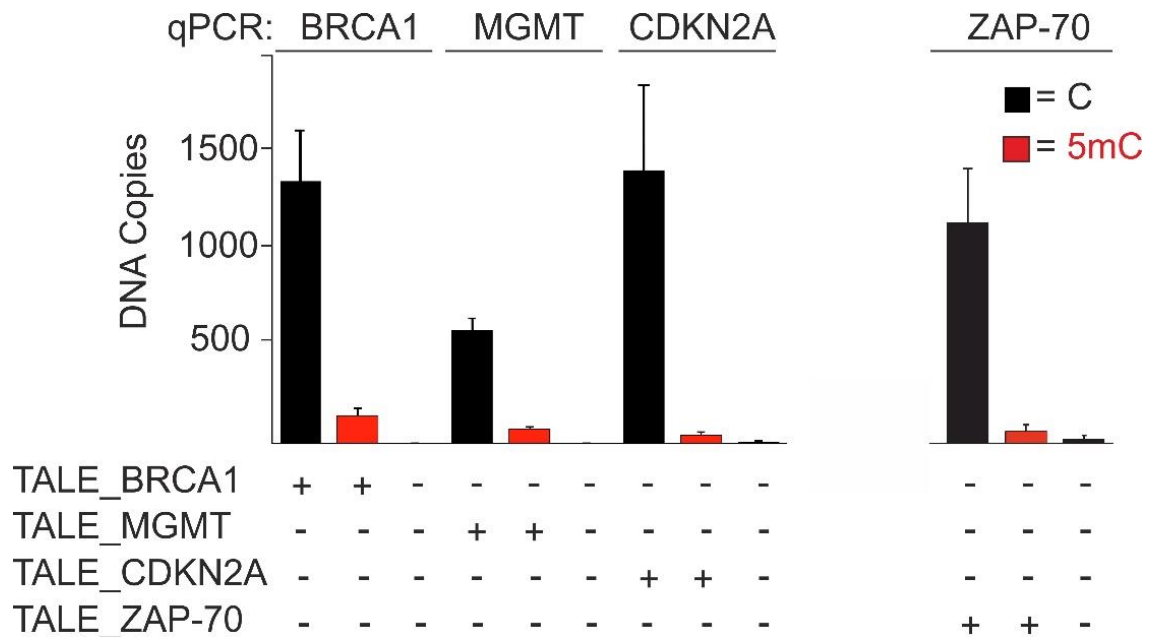


Figure IV-8: Result indicating enrichment of target sequence from M. SssI treated (red) or untreated (black) WGA human gDNA sample using TALEs mentioned below. Target sequences contained one or two CpG sites (Figure IV-2). qPCR was performed for amplifying loci mentioned above the columns. Break in the X-axis indicates experiments conducted using the same set-up at separate occasions. Error bars indicate standard error from experimental and qPCR duplicates [17].

The possibility of using TALEs in an enrichment assay to detect sequence specific, single nucleotide methylation in genomic sequences, tested using various sequences prompted their possible use as a sensitive technique for detecting methylation in a gene relevant for diagnosis of certain diseases in humans. Methylation at a single CG in the ZAP-70 locus (occurring 223 bp downstream of the transcription start site) in human genome serves as a clinical biomarker for prognosis of the type of chronic lymphocytic leukaemia in adults. A TALE targeting a 26 bp sequence containing the mentioned CG was designed and an affinity enrichment assay was performed with C or 5mC containing human gDNA. A clear preference for binding and isolating samples with no modification was observed also for this medically relevant and important target sequence (Figure IV-8). These results also paved the way to conduct more advanced experiments requiring higher sensitivity and selectivity including detection of C modifications other than 5mC.

C. Quantitative detection of 5mC in gDNA samples

It was evident from the experiments so far, that TALEs containing RVD HD enriched higher copies of non-modified C containing sequence compared to those bearing 5mC. However, gDNA samples typically do not exhibit a homologous pattern or content of 5mC occurrence. To reveal if this experimental set up was sensitive enough to be able to detect the extent of methylation present, a levelling assay was designed with samples containing varying percentages of 5mC. Enrichment was performed with a total of 200 ng of randomly sheared DNA as described earlier. M. SssI treated and non-treated gDNA were mixed to yield experimental sample carrying varying percentage of methylation (from 0% to 100%). Based on prior observation, the most promising candidate i.e., TALE_CDKN2A was chosen for the experiment. Quantification of the amount of target DNA isolated from different samples using qPCR showed a near linear, inversely proportional relation between the level of 5mC present in the target sequence and their binding by the target TALE protein (Figure IV-9).

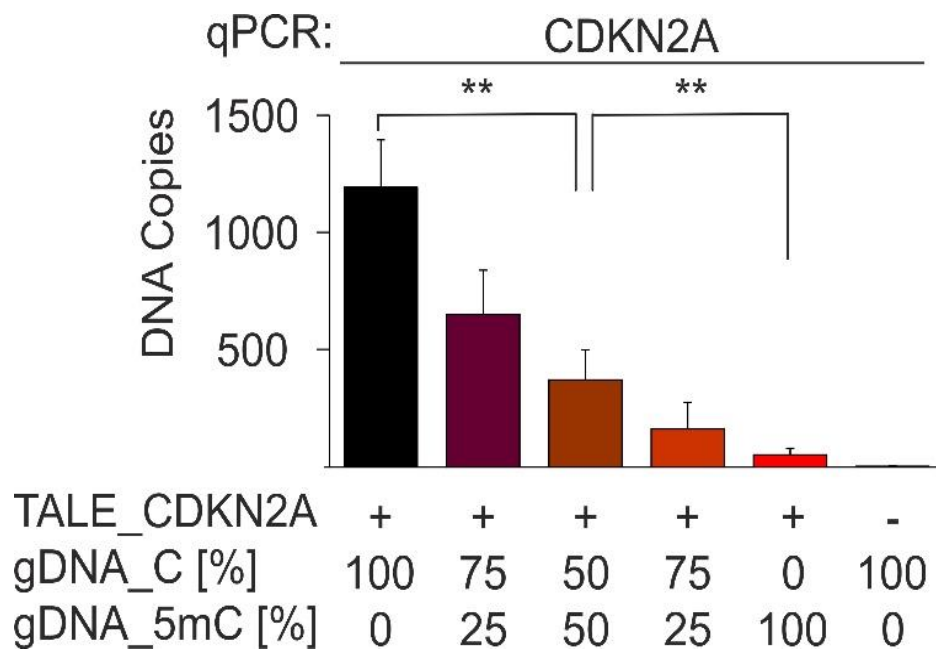


Figure IV-9: Affinity enrichment using TALE_CDKN2A and samples of human gDNA containing varying percentage of the targeted CpG methylated as mentioned below. Error bars indicate standard error from experimental and qPCR duplicates. ** indicate significance difference as calculated by student t-test with $p \leq 0.01$ [17].

To be able to verify the universality of this approach, an analogous experiment was performed with the TALE_Hey2_b and Zf gDNA in a similar set-up. Although less drastic, similar effect of methylation on TALE binding and isolation was observed (Figure IV-10).

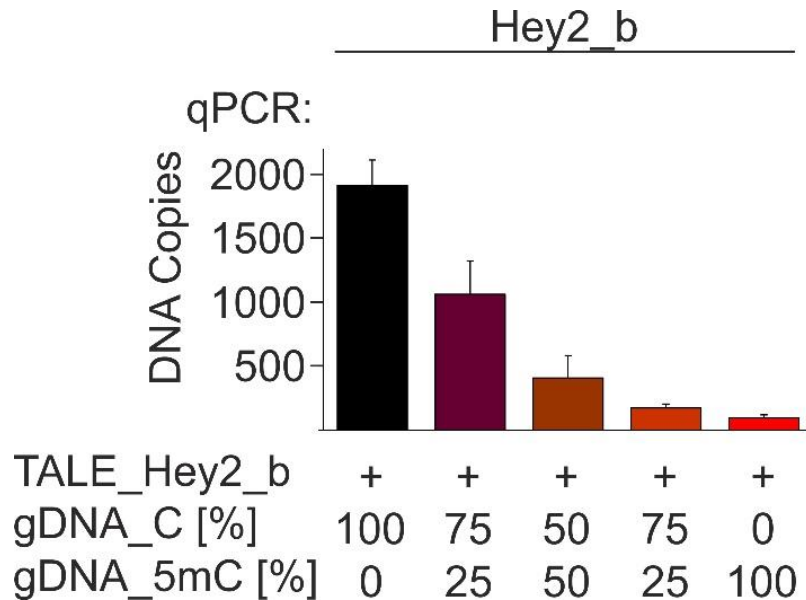


Figure IV-10: Affinity enrichment using TALE_Hey2_b and samples of Zf gDNA containing varying percentage of the targeted CpG methylated as mentioned below. Error bars indicate standard error from experimental and qPCR duplicates.

D. Simultaneous detection of 5mC in different loci of human genome

The complete protocol for enrichment of desired sequences using TALEs requires one day and a few preparations in advance. It would be highly beneficial and reasonable if a customized mix of different TALEs could be used together for simultaneous detection of methylation at more than one target sequence in one genome (Figure IV-11).

A multiplexing experiment was designed for detection of 5mC in human gDNA using TALEs BRCA1, CDKN2A and MGMT. Samples containing one, two or all the three TALEs in equal amounts with 200 ng of human gDNA were used in an enrichment assay. The isolated DNA sample were quantified in a qPCR reaction to reveal the number of DNA copies enriched for each individual target sequence. Discrimination in enrichment of

samples containing C and 5mC containing gDNA was observed in every single combination, irrespective of the number of TALEs used together and without the need to increase the DNA concentration (Figure IV-12). The results for control samples containing no TALE were also in agreement with the previously observed effect and exhibited negligible non-specific binding by the TALEs in this assay (result not displayed). The conditions employed are well suited for the sample size of laboratory purposes but since the process does not require expensive reagents or instruments, it can be scaled up and customized according to requirement.

This experiment specially highlighted the potential of TALEs for application as a diagnosis tool in the medicine industry.

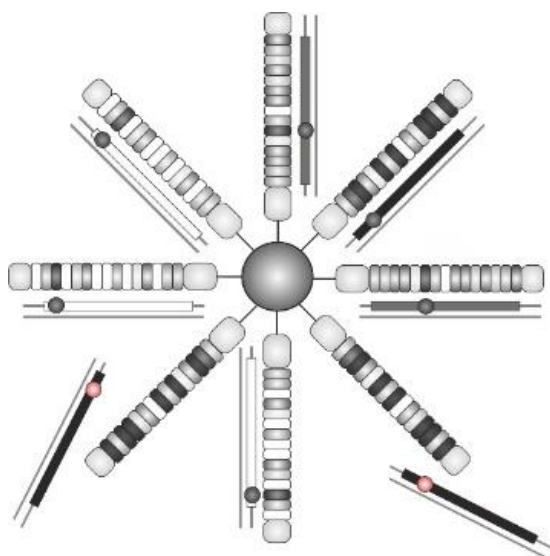


Figure IV-11: Pictorial representation of the multiplexing affinity enrichment assay. TALEs targeting sequences in different loci of the same genome are used together for parallel detection of methylation in more than one desired sequence. Ni-NTA bead shown in grey in the centre is bound to the His6 tag of different TALEs. Randomly sheared dsDNA with different target sequence depicted as white, grey or black bars. Grey and red spheres on the target sequence represent non-methylated and methylated CpG respectively. Please note that each gDNA sample contained either C or 5mC in the target sequence and not both [17].

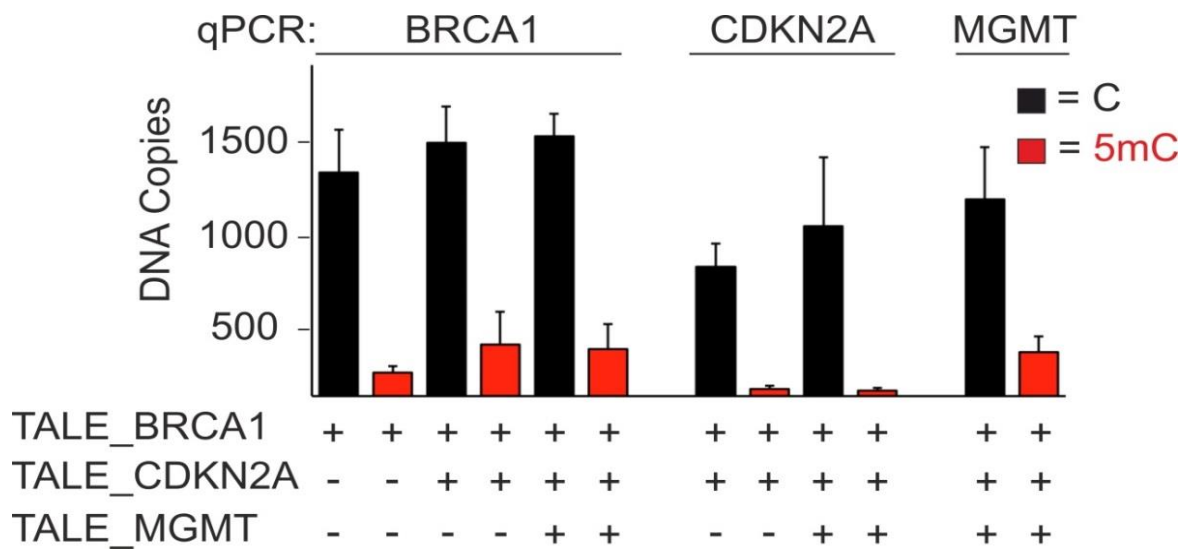


Figure IV-12: Result of multiplexing affinity enrichment assay using TALEs targeting three different human genomic loci as mentioned below. Quantification of DNA enriched in qPCR using p.p. for loci mentioned above the columns. Error bars indicate standard error from experimental and qPCR duplicates [17].

E. Strand specific detection of methylated CpG using TALEs

The level and pattern of CpG methylation in DNA does not always remain constant. At any given time, DNA in a cell can consist of methylated or non-methylated C. Due to the replication mechanism that uses a single strand as templates and extends DNA by pairing both C and 5mC with G, the methylation is diluted and DNA can therefore exist in a third state called a hemi-methylated state where only one of the strand bears a methylated C on the CpG while the C of the CpG on the complementary strand remains free of methylation. To understand different mechanisms occurring at various time points, the ability to detect 5mC in a strand specific manner is of high biological interest. Since TALE protein modules contact the DNA nucleobases only on the target strand that too strictly in the 5' to 3' directions, simultaneous detection of methylation on CG of both the strand using the same TALE was not possible. To find out if TALEs could detect methylation exclusively on only one strand using our approach, two different TALEs targeting opposite strands of the same sequence would be required. TALE_BRCA1 and TALE_BRCA1_a (BRCA1 antiparallel)

targeting the sense and the antisense strand respectively of BRCA1 locus were therefore expressed. To ensure methylation only at the required position, PCR products were prepared with single C or 5mC containing target sequences on both strands (SI Sequence 1). P.p. o1516/o1601 or o1518/o1601 was used to generate C or 5mC containing BRCA1 PCR product and p.p. o1599/o1529 or o1600/o1529 was used C or 5mC containing BRCA1_a PCR product. To keep the complexity of the samples is unaffected, the amplicons were used along with randomly sheared human gDNA restricted additionally with SacI restriction endonuclease (with recognition sequence within the BRCA1_a target sequence). Input templates with or without the PCR products spiked-in were first tested in qPCR reaction with appropriate p.p. and significantly low amplification of samples containing only the background gDNA was observed. Obtained results depicted successful discrimination by TALE_BRCA1_a between C and 5mC occurring within its target sequence but not when it occurred on the complementary strand (Figure IV-13). It was hence concluded that strand specific detection of methylation was possible using TALEs. From earlier observations, it was clear that TALE_BRCA1 could detect 5mC on its target, therefore results of only TALE_BRCA1_a is shown here.

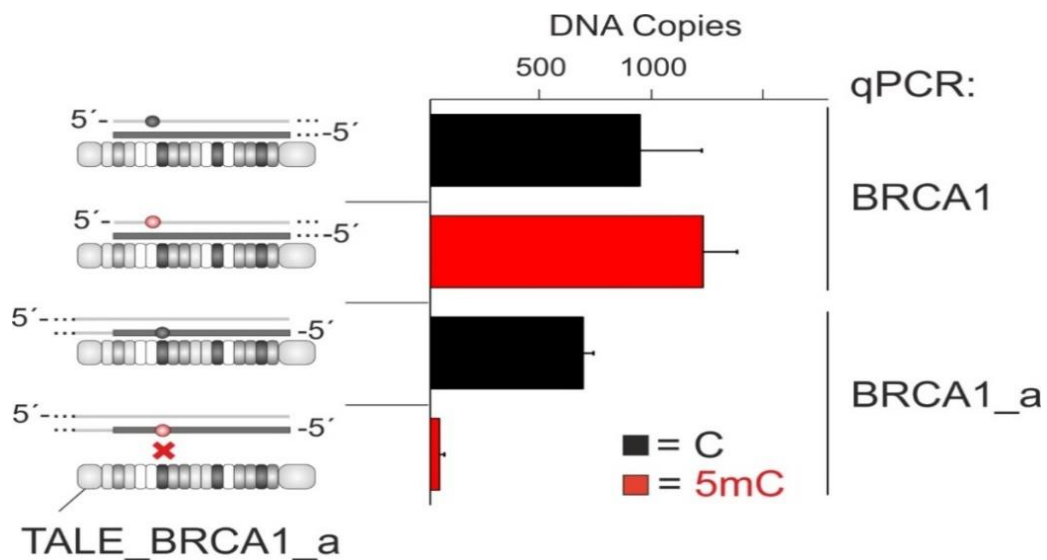


Figure IV-13: Pictorial representation of DNA and TALE samples used (left) and result of strand specific detection of 5mC by TALE_BRCA1_a (right). Red and gray beads represent 5mC and C respectively occurring on the target sequence. The amplified locus is mentioned on the right. Error bars indicate standard error from experimental triplicates and qPCR duplicates [17].

F. Comparison of TALEs and MeDIP for detection of 5mC in gDNA

Previously conducted experiment established that TALE binding was sensitive to a single 5mC in the target sequence and could be used as a tool for successful discrimination between modified and non-modified samples. The next idea was to compare the efficiency and sensitivity of this approach with the most commonly employed MeDIP (Methylation Dependent Immunoprecipitation) technique (Figure IV-14) for quantitative detection of epigenetic modification on DNA.

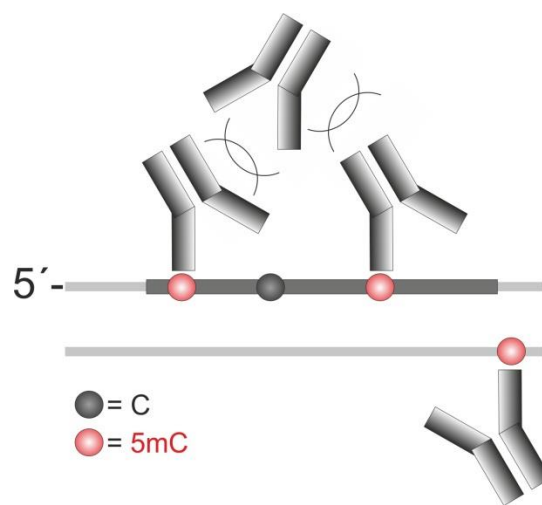


Figure IV-14: Schematic representation showing the working principle of MeDIP assay using monoclonal antibodies. The gray bar represents a target sequence occurring on the sense strand of Zf Hey2b locus locus and the gray and red beads represent non-methylated and methylated CpG respectively [17].

Amplification products were obtained by conducting PCR on the Zf Hey2 gene with p.p. o476/o1597, o465/o1597, o1057/o1597 and o1056/o1597 to incorporate none, one, two or three 5mC in the target site of TALE_Hey2_b (Figure IV-15) and used with the monoclonal 5mC-antibody supplied in the MeDIP kit. The sample mix was supplemented with non-methylated and methylated control DNA and their primers were provided to track and verify the overall efficiency of immunoprecipitation by the antibodies in the MeDIP kit. After the assay, the isolated product was tested and quantified in a qPCR reaction using primers

provided in the kit as well as those designed specifically for Zf Hey2 sequence. While the quantity of enriched control DNA (positive and negative) were consistent with the values mentioned in the kit manual, (SI Figure VI-7) only negligible amounts of Hey2 sequence was recovered for all different type samples containing varying number of 5mC in the target sequence (SI Sequence 2). To verify that this result was precisely due to low level of methylation on the target sequence, this experiment was repeated with an additional sample carrying 30 5mCs (Figure IV-15). The sample was prepared by treating the PCR product with M.SssI which would methylate the 15 CGs occurring in each strand of the amplicon. Like earlier, amplification of the controls sequences was obtained in an expected range while Hey2 sequences with upto 3 5mCs in the TALE target sequence were isolated in negligible amounts. Samples containing all methylated CGs in the Hey2 locus were however enriched in reasonable amounts (~7% of the input) (Figure IV-16). This indicated that in addition to lacking sequence selectivity and strand specificity, this method also suffers bias due to DNA fragment size and methylation levels.

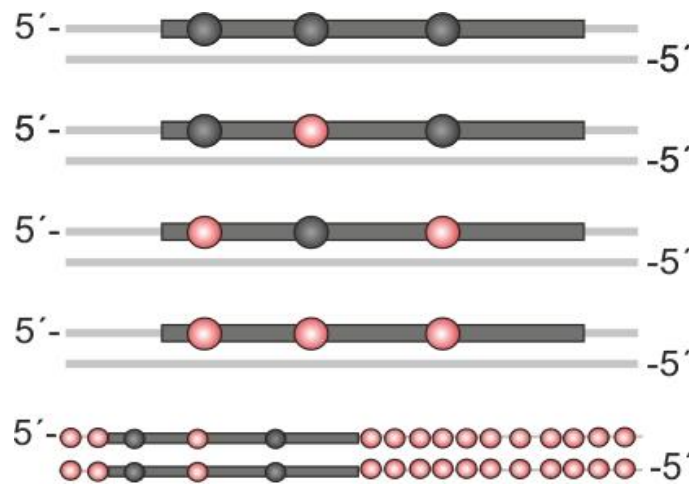


Figure IV-15: Pictorial representation of PCR amplicons containing 0, 1, 2, 3 or 30 methylated CpGs used in the experiment. Gray solid bars represent the TALE target sequence on Zf Hey2_b locus. Gray and red beads indicate C and 5mC respectively.

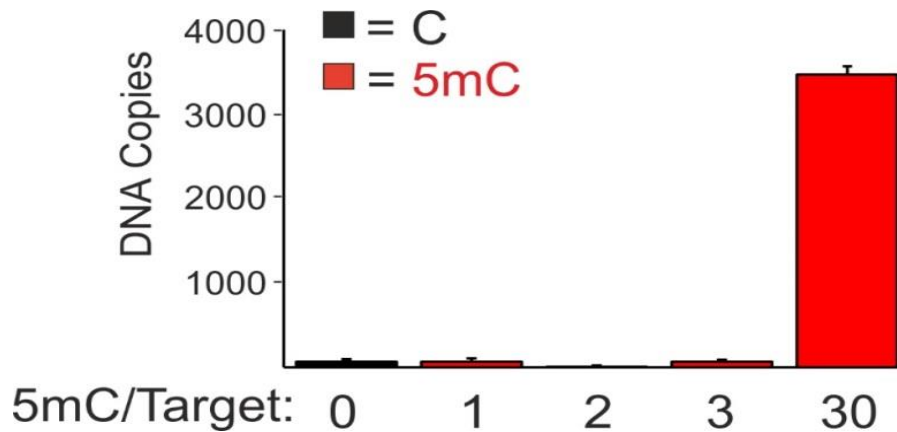


Figure IV-16: Results obtained from conducting the immuniprecipitation assay (Figure IV-14) using DNA samples containing the same target sequence but varying in the number of 5mC present which is denoted under the columns (Figure IV-15). The Y-axis indicates the quantity of amplicon enriched using the protocol. Error bars indicate standard error from experimental and qPCR replicates [17].

For comparison, a TALE_Hey2_b HD1, HD10 to N* was designed and expressed. It contained RVD HD targeting Con two positions, one on the left and other on right of the targeted CG, replaced by N*. Here, N represents the amino acid arginine and * indicates deletion of an amino acid at position 13th of the repeat. Due to reduction in size, the RVD N* can accommodate the methyl group of C and hence binds with similar affinity to C and 5mC (Figure IV-18). When used with the PCR products containing 0 to 3 5mCs the TALE showed highest enrichment for samples containing no 5mC, lowest for the sample containing a single 5mC at the CG and intermediate for the other samples containing methylation at either the non-CG C or at all the three C (Figure IV-19). Overall, PCR products treated with M.SssI was enriched the least. Additional advantage of the enrichment assay was the possibility to conduct a complete experiment in 7.5 h whereas the MeDIP protocol required 24 h and more hostile conditions to be completely executed. These data therefore suggested superior sensitivity of TALEs over antibodies for detection of 5mC.

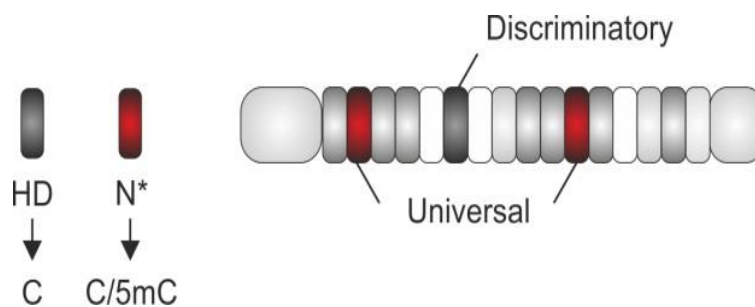


Figure IV-17: Model of the engineered TALE Hey2_b used in the experiment with two HD repeats at position 1 and 10 replaced by N* repeats. The N* repeat is known to bind to both C and 5mC with similar affinity and hence denoted as an universal sensor [17].

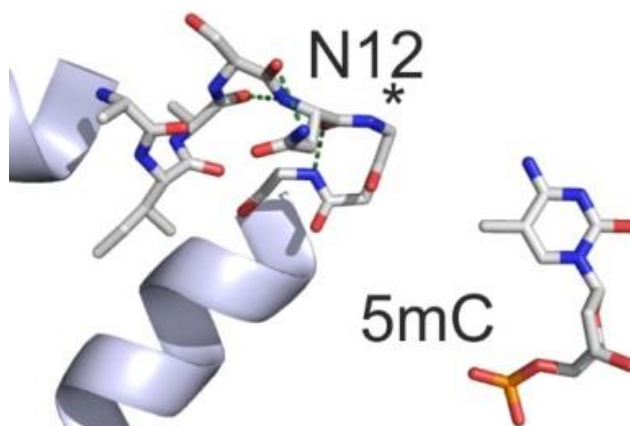


Figure IV-18: Crystal structure of TALE repeat with RVD N* binding to 5mC (pdb entry: 3V6T for N* binding to C and 4GJP for structure of 5mC). Reduction in size due to deletion of the 13th amino acid allows the accommodation of the methyl group, the presence of which disrupts binding by the RVD HD to 5mC [17].

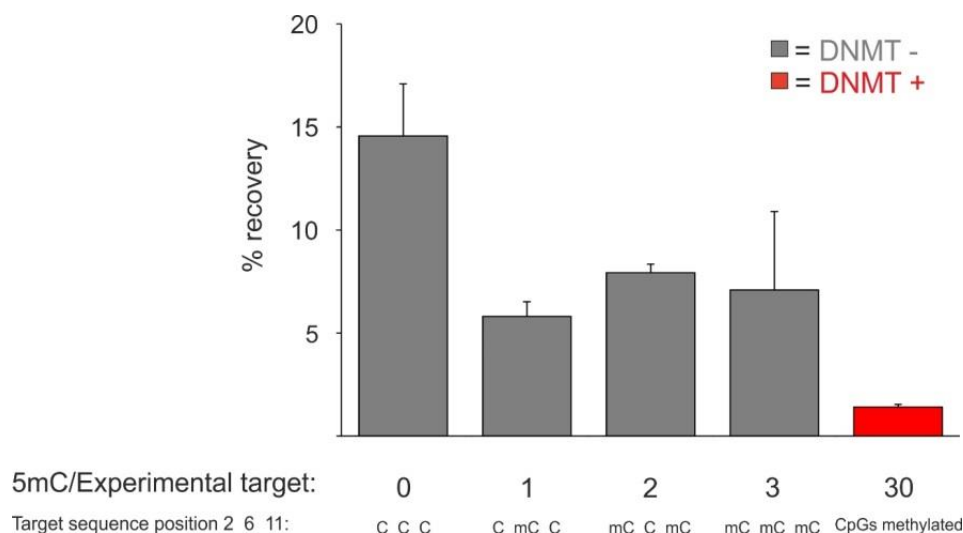


Figure IV-19: Results obtained from affinity enrichment using engineered TALE_Hey2_b (Figure IV-17) and DNA samples containing different amount of 5mC (Figure IV-15) as mentioned below the columns. Y-axis indicates the percentage of input DNA sample recovered by enrichment. Error bars indicate standard error from experimental triplicates and qPCR duplicates.

G. Using size reduced RVD for detection of oxidative methyl C bases.

After observation of successful 5mC discrimination in various experiments, the obvious question raised was if this technique could be extended for detection of oxidative methyl C epigenetic modifications. Parallel work in our group by Sara Maurer included screening of TALEs with engineered RVDs for discovery of selective sensors for each of the C modifications reported in eukaryotes. While all different C variants exhibit base pairing with a nucleobase G in DNA (Figure I-9), the chemical substitutions on C affect the contact between DNA nucleobases and TALE RVDs due to their intrinsic capability to accept or donate a hydrogen atom. Moreover, they display a difference in steric demand and conformational flexibility.

Earlier work in our group has established discrimination between 5mC and 5hmC using TALE with RVD N* but a positive sensor for selective detection of epigenetic nucleobases 5fC and 5caC has not yet been achieved. RVD HD binds to C but not to 5mC due to perturbation of the hydrogen bond by the presence of a methyl group (Figure I-18). Similarly, RVD NG that is found in natural TALEs targeting DNA nucleobase T is known

to bear affinity to 5mC but not to C due to similar chemistry of the methyl groups in T and 5mC. Another naturally occurring TALE RVD N* (Figure IV-18) has been reported to bind both C and 5mC but not 5hmC thus serving as a discriminatory sensor for 5hmC. Amplicons containing a single C, 5mC, 5hmC, 5fC and 5caC were generated by performing PCR on Zf genome using p.p o476/o1597, o465/o1597, o465/o1597, o465/o1597 and o465/o1597. TALE based affinity enrichment assay using TALE_Hey2_b with RVD HD targeting the 5th C changed to N* and PCR products containing the Hey2_b binding site (SI Sequence 2) bearing a single C, 5mC or 5hmC on the CpG opposite RVD N* confirmed the previously observed effect [15]. While DNA containing C or 5mC was enriched in similar amounts, significantly lower amounts of samples containing a single 5hmC was obtained (Figure IV-20). Next the experiment was repeated and samples containing 5fC or 5caC were also included. It was observed that in addition to 5hmC, 5caC containing samples were also bound less by N* containing TALEs. 5fC containing samples were enriched in quantities comparable to 5mC (Figure IV-24) perhaps owing to the similarity in structure of these modified nucleobases facing the RVD (Figure IV-21). The result was consistent with the previously seen effect in primer extension experiments using synthetic oligonucleotides [16].

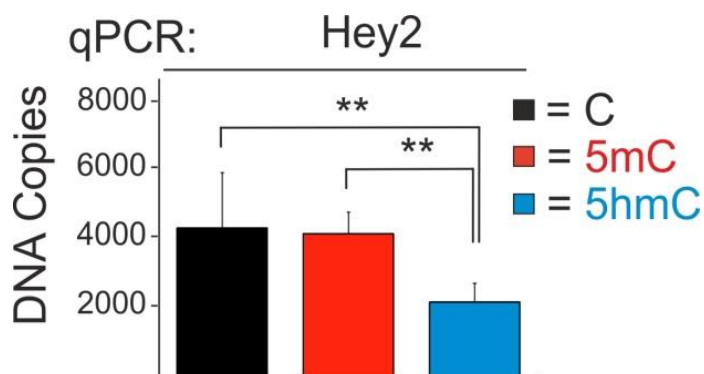


Figure IV-20: Result of affinity enrichment experiment conducted using engineered TALE_Hey2_b with original RVD HD5 replaced by N*. The DNA samples used contained a single C or 5mC or 5hmC opposite RVD N*. Y-axis denotes the quantity of DNA samples enriched. Error bars indicate standard error from experimental triplicates and qPCR duplicates. ** represents significant difference when calculated using Student t-test with $p \leq 0.01$ [17].



Figure IV-21: Epigenetic modifications of C nucleobase. 5mC, 5hmC, 5fC and 5caC as found in crystal structure of DNA duplexes. (pdb entry: 4GJP, 4I9V, 4QC7 and 4PWM respectively) [17].

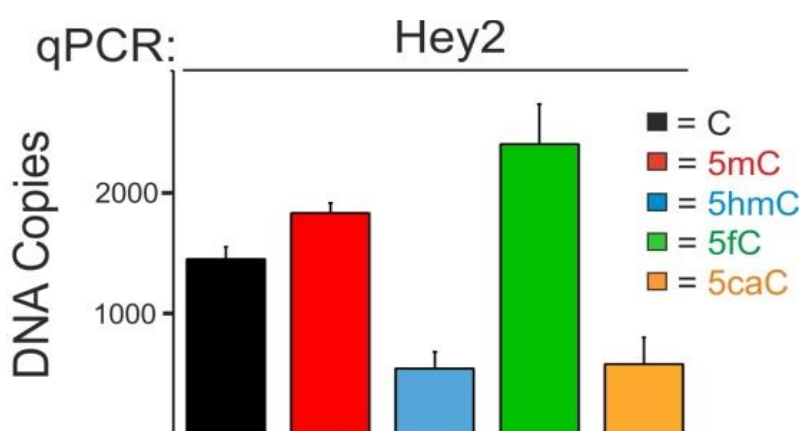


Figure IV-22: Result of affinity enrichment assay conducted using engineered TALE_Hey2_b with N* targeting a single C, 5mC, 5hmC, 5fC or 5caC containing samples. Y-axis denotes the quantity of DNA enriched. Error bars indicate standard error from experimental triplicates and qPCR duplicates [17].

To further verify that the difference in enrichment is solely due to the single modifications, prior enrichment, the spike-in containing 5fC was reduced using sodium borohydride (NaBH_4). Both NaBH_4 treated and non-treated samples were used along with 5hmC containing amplicons in the experiment. While enrichment of 5hmC containing samples were similar in both cases, isolation of NaBH_4 treated 5fC samples reduced significantly and reached values like that of 5hmC containing samples. This showed that a combination of chemical reduction and affinity enrichment using TALEs with engineered RVDs could be used for selective detection of 5fC in gDNA samples (Figure IV-24).

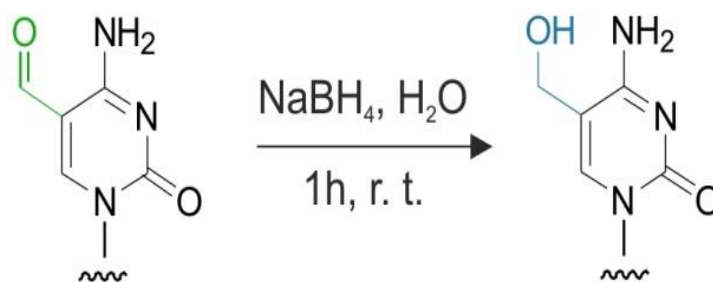


Figure IV-23: Reaction showing NaBH_4 mediated reduction of 5fC to 5hmC [17].

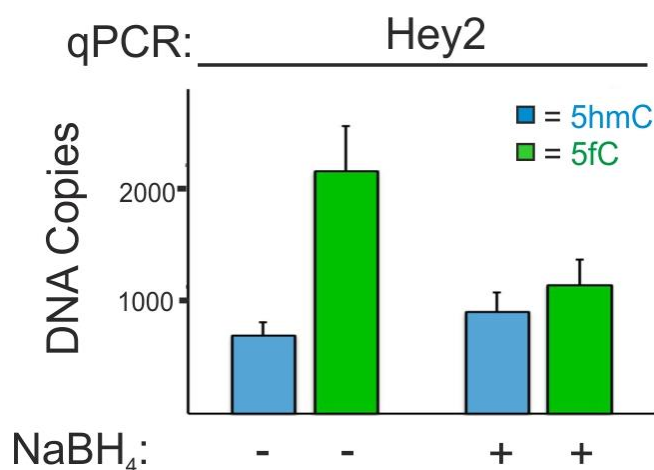


Figure IV-24: Result of affinity enrichment assay conducted using engineered TALE_Hey2_b with N* targeting samples containing a single 5hmC or 5fC. Results before (left) and after (right) the NaBH_4 mediated reduction (Figure IV-23) are shown. Y-axis denotes the quantity of DNA enriched. Error bars indicate standard error from experimental triplicates and qPCR duplicates [17].

H. Selective binder for a prokaryotic epigenetic C nucleobase; 4mC

In some bacteria methylation of C nucleobase is more versatile than in eukaryotes and their genome consists of both 5mC and 4mC (Figure I-10). Unlike 5mC, 4mC is a less abundant base and not much is known about its exact role in the bacterial genome. Approaches like SMRT-Seq, TAB-Seq and use of selective restriction enzymes provide option for detection

of modified bases throughout the bacterial genome. However, they are expensive, time consuming and require special machinery. The existence of 4mC in the commercially purchased synthetic oligonucleotides was confirmed using mass spectrometry. Because the spectrum indicated an increase in weight equivalent to a methyl group for both 5mC and 4mC containing samples, differential detection of the two methylated C nucleobases could not be achieved using this technique.

To provide a complementary method, we decided to use TALEs for selective detection of 4mC. A library of modified TALEs targeting the *Hey2_b* sequence in *Zf* genome was prepared [16]. Repeat 5 originally carrying RVD HD was replaced by X* where X denotes all 20 naturally occurring amino acids and * denotes a deletion. This TALE library was scanned in different competitive or binding assays for selective binding to target sequence carrying either a C, 5mC or 4mC against the engineered RVD X* (Figure IV-27)

The deletion of the 13th amino acid was thought to provide some flexibility to accommodate the N4 methyl group similar to the binding observed between RVD N* and 5mC nucleobase (Figure IV-25, Figure IV-18).

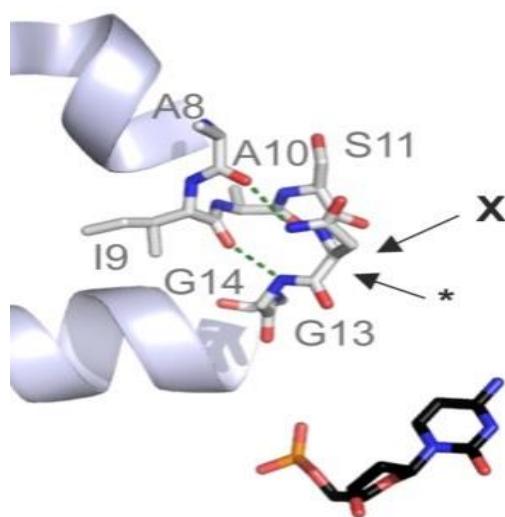


Figure IV-25: Crystal structure of a TALE-DNA complex showing RVD N* binding to the nucleobase C (pdb entry 3V6T) [19].

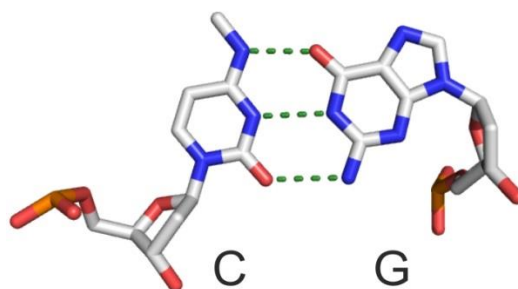


Figure IV-26: Crystal structure showing base pairing in Z-DNA between nucleobases G and 4mC (pdb entry: 133D) [19].

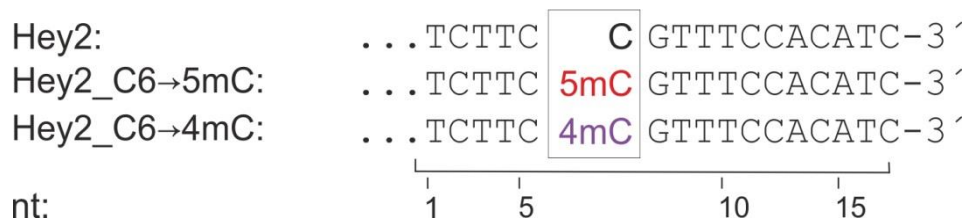


Figure IV-27: TALE target sequence Hey2_b containing a C, 5mC or 4mC at the fifth position targeted by the library of TALEs bearing the engineered RVD X* where X denotes all naturally occurring amino acids and * denotes deletion of the 13th amino acid in the TALE repeat [19].

I. Identification of RVD T* as a selective 4mC binder

Primer extension assay as described before was conducted using TALEs from the X* library with synthetic oligonucleotides o476, o465 or o1236 bearing the Hey2_b TALE target sequence with the C, 5mC or 4mC positioned opposite the engineered RVD and the common ³²P labelled rv primer o466. Result produced from evaluation of the assay indicated a few mutants that displayed preferred binding to 4mC compared to C or 5mC containing oligonucleotides (Figure IV-28). What was also evident was that the WT TALE with RVD HD which did not bind 5mC in all the previously conducted experiments, exhibited poor binding also towards 4mC (SI Figure VI-9). Amongst the most promising candidates binding exclusively to 4mC was TALEs with RVD T* (Figure IV-28).

Using two mutants that both did not bind 5mC and bound either C or 4mC, discrimination of 4mC from 5mC was possible.

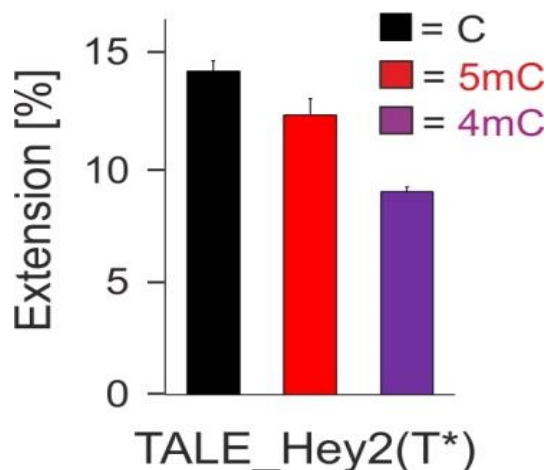


Figure IV-28: Result of PEX assay conducted as shown in Figure III-6 using engineered TALE_Hey2_b with RVD HD replaced by T*. The Y axis represents the percentage of completely extended ds product. Error bars indicate standard error from experimental replicates [19].

Considering that the observation could be assay dependent it was necessary to use the same TALE in another method to judge the universality of the effect. Hence electromobility shift assay was conducted also with the whole library (SI Figure VI-10) as described for the PEX assay. dsDNA containing C, 5mC or 4mC were produced by hybridizing oligonucleotides o465, o476 and o1236 respectively with the common Cy5 labelled primer o2170. Assessment of the result from this method produced conclusions consistent with the previously observed effect. While a clear shift band of DNA-TALE hybrid was seen when TALE containing HD was used with C containing template only a faint or none was observed for 5mC or 4mC containing DNA (Figure IV-29). Quantification and plotting of the result also conveyed T* as a selective binder of 4mC (Figure IV-29).

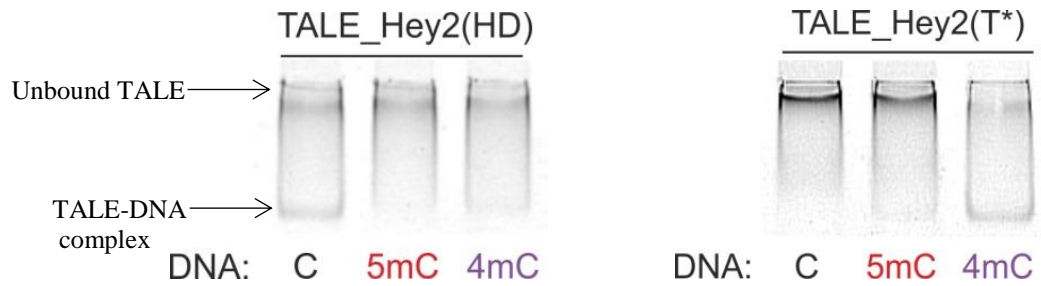


Figure IV-29: Pictures of EMSA gels showing binding by TALE_Hey2_b with RVD HD (left) or RVD T* (right) at the 5th position to DNA containing a single C, 5mC or 4mC in the target sequence [19].

J. Determination of K_i for TALEs Hey2_b (WT) and (T*) using FRET

A new homogeneous fluorescence assay based on competitive binding between TALEs and DnaseI to a synthetic ds oligonucleotide was optimised and used [19]. The duplex contained a donor/acceptor dye pair (Cy3/Cy5) capable of undergoing fluorescence resonance energy transfer (FRET) (Figure III-8) such that restriction by DnaseI due to low binding of TALEs lead to increase in the distance between the donor/acceptor dye and hence reduced fluorescence by emission of the acceptor Cy5.

Partially double stranded oligos containing a single C, 5mC or 4mC were obtained by hybridizing oligos o476, o465 or o1236 respectively with a common 5'Cy5 and 3' Cy3 labelled rv primer o1892. When incubated with TALE_Hey2_b (WT) with RVD HD in repeat 5, fluorescence for C containing oligonucleotide were reduced due to restriction by DnaseI only after 50 min of incubation while 5mC or 4mC containing templates exhibited extremely low values of fluorescence starting 10 min after addition of DnaseI (Figure IV-30). This indicated that TALE binding protected the DNA from digestion by DnaseI and this effect was quantifiable using this FRET assay.

Screening of the entire mutant library in this assay was performed for various time points (SI Figure VI-12). Controls without DnaseI for C, 5mC or 4mC containing samples were similar and were used for normalization of fluorescence from each of the respective samples (SI Figure VI-11).

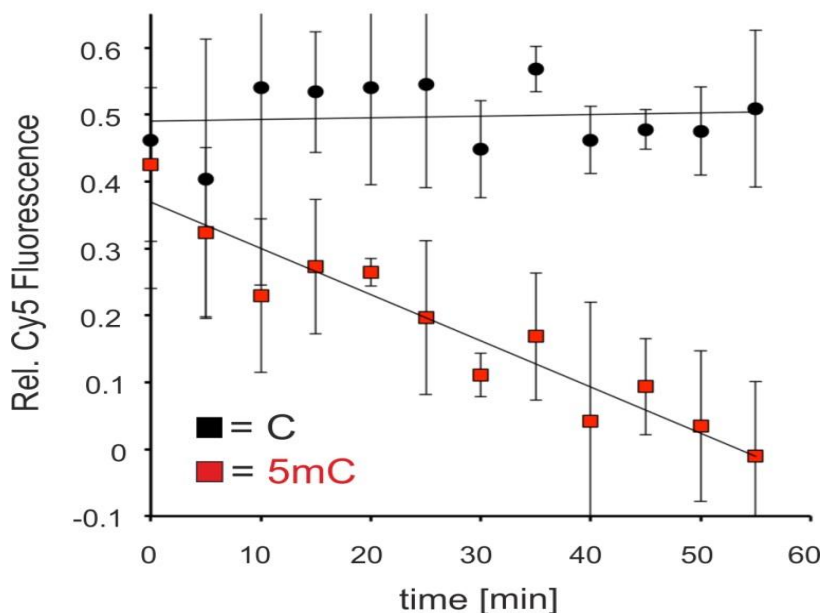


Figure IV-30: Results from the FRET assay conducted using TALE_Hey2_b with RVD HD at position 5 targeting either a C or 5mC containing sequence. Y-axis indicates the normalized fluorescence values at different time points mentioned on the X-axis. Error bars indicate standard error from experimental duplicates [19].

To summarize the whole scan, engineered TALEs with aromatic amino acids H*, F* and W* exhibited an overall weak binding with an exception of Y* showing high affinity and slight selectivity towards 5mC. P* also showed general weak binding. Mutants bearing the polar RVDs N* and Q* as well as G*, M*, V*, L*, I* and C* bound to all the three C variants in a similar fashion. The A* and K* showed weakest binding to C. Acidic amino acids D* and E* showed slight selectivity for 4mC. The basic repeat R* displayed strong binding to C. TALEs with hydroxyl bearing S* and T* bound strongly to all C variants but only T* showed positive selectivity to 4mC. However, both HD and T* containing TALEs, exhibited reduced fluorescence for oligonucleotide containing 5mC (Figure IV-31). Hence, RVD HD and T* were recognized as positive, selective sensors respectively for C and 4mC that both did not bind to 5mC.

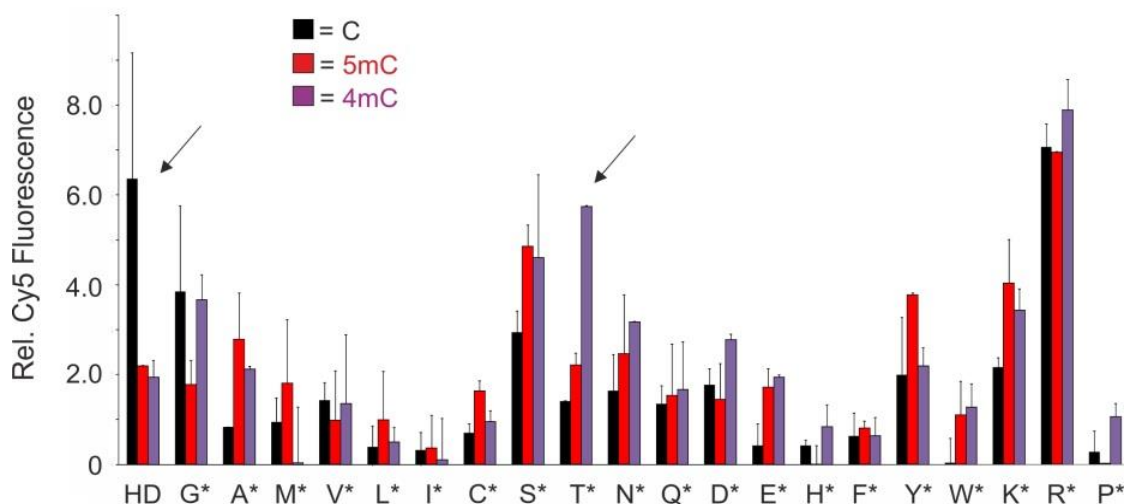


Figure IV-31: Result from scanning the TALE_Hey2_b X* library using FRET values after 60 min of DnaseI addition. TALEs containing RVD HD or T* at position 5 are marked by arrow. The Y-axis indicates the normalized fluorescence values. Error bars indicate standard error from experimental duplicates [19].

To further characterize the selectivity of RVDs HD and T* a FRET assay was conducted with different concentrations of TALEs containing these RVDs while the DNA amount was kept constant to get different TALE/DNA ratios. The Cy5 emission fluorescence values were plotted against the TALE concentrations in Origin and Ki values for each TALE with each C variant were obtained using the Hill fit (Figure IV-32 a, b). The maximal Cy5 fluorescence and inhibition constant for each modification with both TALEs Hey2_(HD) and Hey2_(T*) are shown (Figure IV-32 c).

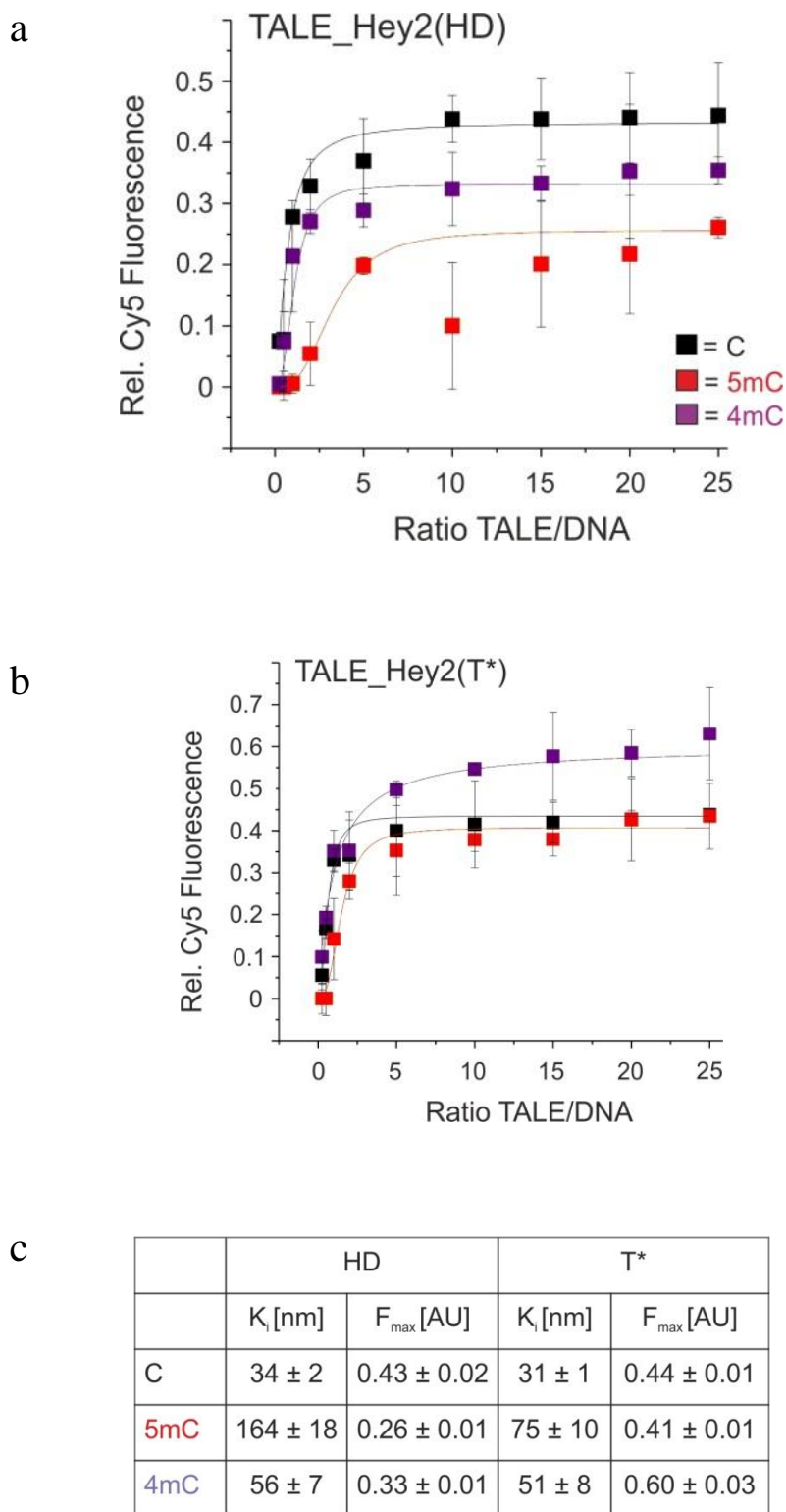


Figure IV-32: Fits generated using a FRET assay for TALE_Hey2_b with HD (a) or T* (b) at the fifth position with C, 5mC or 4mC containing sequence. Both TALEs show least binding to 5mC with TALE_HD and T* binding strongly to C and 4mC containing samples respectively. Error bars indicate standard error from experimental duplicates. The highest fluorescence value and inhibition constant for each combination is listed in the table (c) [19].

Although the values of K_i and highest fluorescence recorded were not very different for different modifications using either TALE_Hey2_b HD or TALE_Hey2_b T*, the values were lowest for 5mC in both cases. Further the curves indicate preferential binding of C or 4mC by TALE HD or TALE T* respectively.

K. Use of RVD HD and T* for discrimination between 5mC and 4mC

After establishing TALE RVD T* as selective 4mC binder using *in vitro* binding and competition assays, we wanted to observe if the effect was also produced in a more complex, whole gDNA samples containing a single C, 5mC or 4mC in the target sequence. Zf spike-ins were generated using p.p o476/o1597, o465/o1597 and o1236/o1597 respectively (SI Sequence 2). These were then added in a concentration representing a single genomic locus to WGA and randomly sheared Zf gDNA. The background gDNA was additionally restricted using NdeI and ApaLI restriction endonucleases which cleave in the middle of the product amplified by the qPCR primers. Both TALEs Hey2_b HD and T* enriched extremely low quantities of samples containing the 5mC spike-in and clearly favoured the enrichment of C and 4mC containing samples respectively (Figure IV-33). Based on this result it was evident that simultaneous use of TALEs containing RVD HD or RVD T* targeting C, 5mC or 4mC containing target sequence could be successfully used for discrimination between 5mC and 4mC, the two methylated C nucleobases occurring in certain prokaryote genomes. It further strengthens the efficient use of TALEs for detection of rare epigenetic modifications for which a lot is not yet known.

All the above discussed experiments for selective detection of 4mC were conducted for the same, 18 bp long, Zf Hey2_b target sequence. To eliminate the observed effect being sequence or target length specific, another 26 bp long CDKN2A target sequence having a single C, 5mC or 4mC was designed to be used with TALE_CDKN2A (26) HD and TALE_CDKN2A (26) T*. A FRET assay with different TALE concentrations was conducted for determination of K_i . T* containing TALE was seen to exhibit preferential binding to 4mC while HD containing TALE continued to show preference to C evident by the respective fits, K_i and FRET saturation values (SI Figure VI-13).

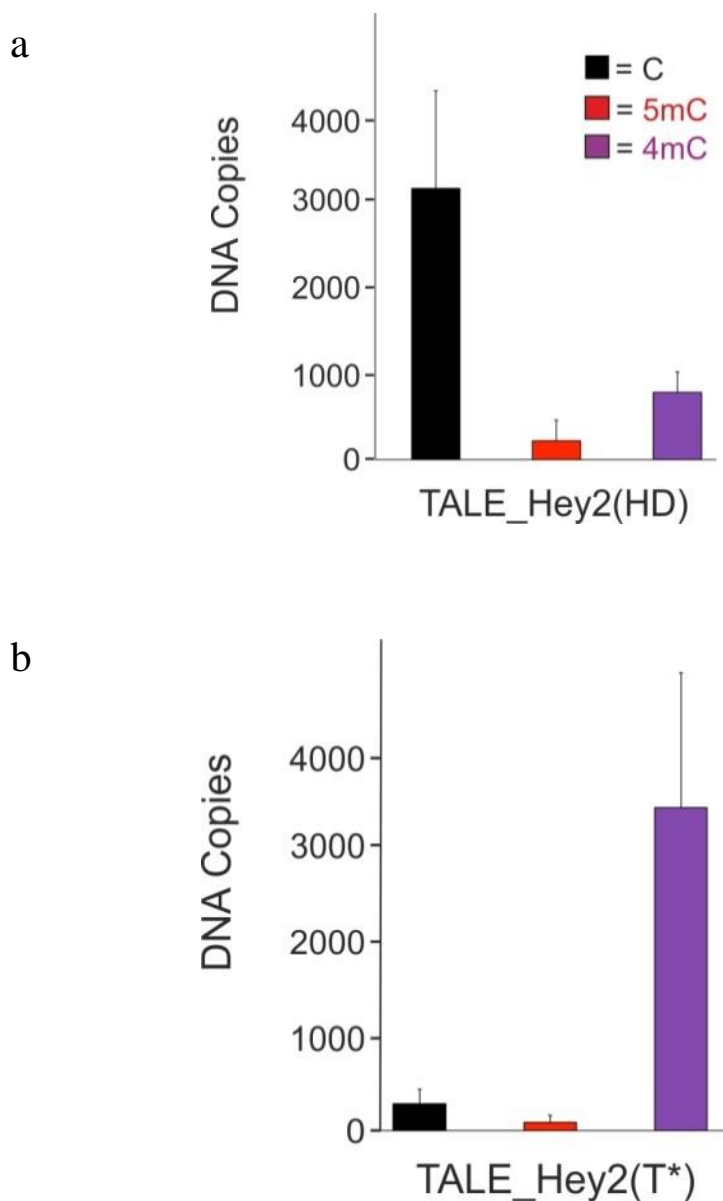


Figure IV-33: Result of affinity enrichment assay conducted using C, 5mC or 4mC containing DNA with TALE_Hey2_b HD (a) or T*5 (b). Y-axis indicates the enriched target sequence quantified by qPCR. Error bars indicate standard error from experimental triplicates and qPCR duplicates [19].

L. Reduced non-specific interaction improves TALE selectivity

TALE protein interacts with dsDNA such that each repeat of the TALE CRD contacts one nucleobase in the DNA major groove; the site where all the epigenetic DNA modifications exists. This makes TALEs a useful tool for detection of epigenetic C modifications in a

chosen sequence. Several studies have recognized different plausible factors affecting the stringency with which TALEs recognize their specific targets. The non-specific interactions extended by all regions of TALEs may account for their binding strength. However, it could also sometimes be the source for off target binding by these proteins. One such example is the cloud of well conserved, positively charged amino acids in the N-terminal region (repeats -3, -2, -1 and 0) of the TALEs whose position suggest non-selective interaction with the backbone of a negatively charged DNA (Figure IV-34) [20].

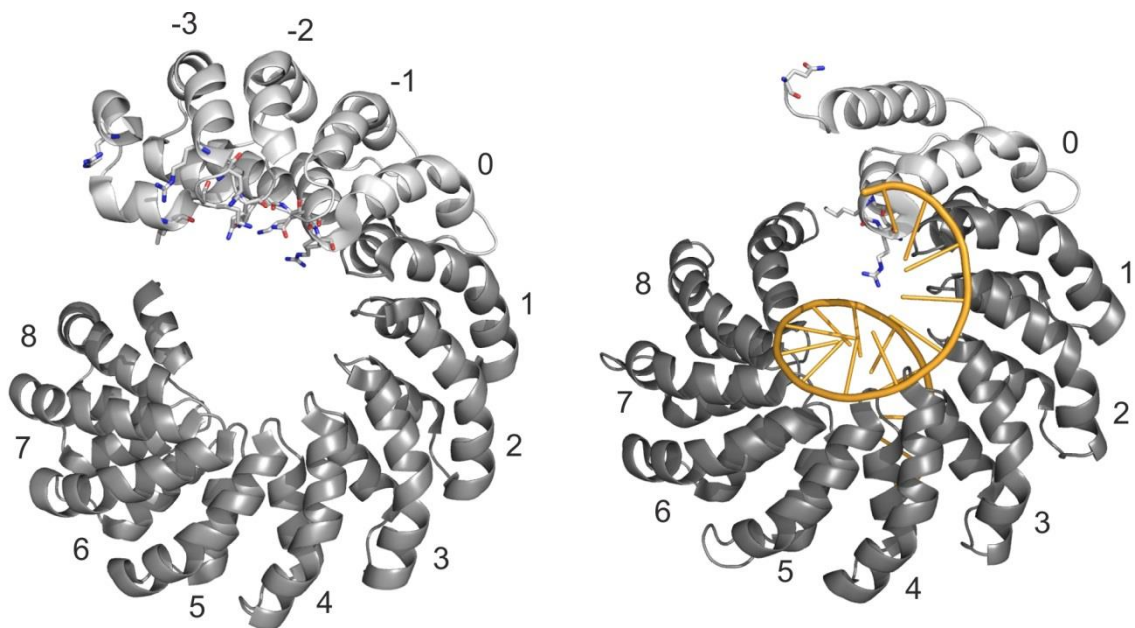


Figure IV-34: Crystal structure of DNA unbound TALE complex (left) (pdb entry 4HPZ) showing repeats on N-terminal relative to T0 (-3, -2 and -1) in light and CRD repeats (1-8) in dark grey. The same structure is shown in a DNA (orange) bound state on the right (pdb entry 3V6T). Amino acids contributing to non-specific interaction are shown as sticks [21].

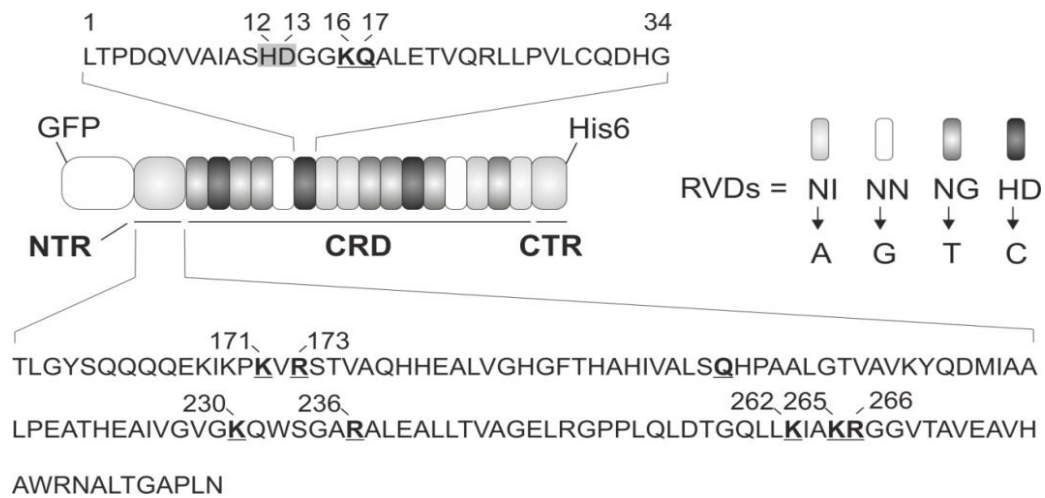


Figure IV-35: Model of TALE bearing electrostatic mutation at the N-terminal. The position of the amino acids relative to the start site is mentioned and they are shown in bold and underlined. On the top the targeted amino acids of the CRD; K16 and Q17 of one of the repeat is underlined [21].

One of the ideas for improving selectivity and hence sensitivity of TALEs for detection of epigenetic modified nucleobases included reducing these non-specific interactions between the protein and DNA sequences. Several studies have proved that TALE NTR contributes the most to its binding strength and a mutation within this part could render them non-functional. TALEs with no CRD or CTR and only 140 amino acids in the NTR were capable to binding to dsDNA in a non-specific manner [154]. Changes on the CTR however were highly tolerated without any loss of function.

To characterize the effect of reduction in non-specific interactions on TALE function, individual basic amino acids in the NTR were exchanged with alanine. The chosen amino acid and position from the NTR start site were lysine at positions 171, 262 and arginine at 173, 236, 266 denoted as K171A, K262A, R173A, R236A and R266A. In addition, TALEs with more than one mutation i.e., K171A&R173A and K262A&K265A&R266A were also created (Figure IV-36).

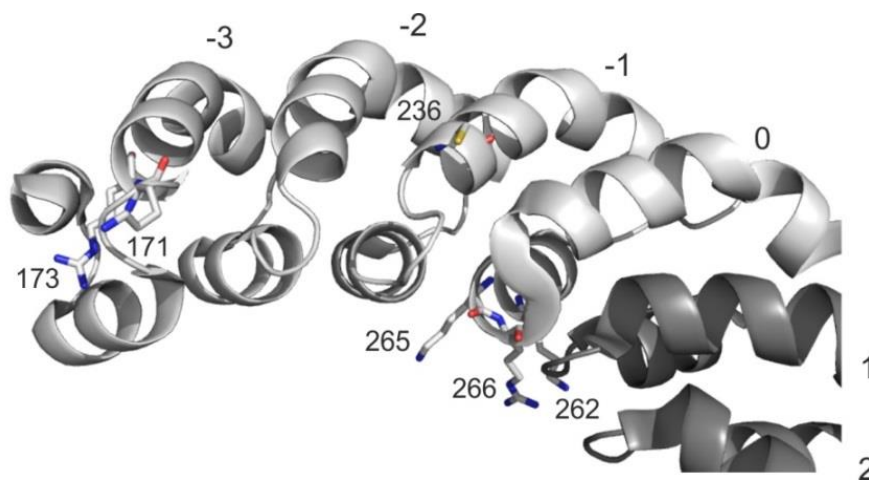


Figure IV-36: Crystal structure of a TALE zoomed in at the N-terminal region (pdb entry 4HPZ) to show the position of amino acids chosen to be exchanged with alanine. TALE NTR repeats are shown in light grey and those in the CRD in dark grey [21].

TALE CDKN2A was already tested to function successfully in various *in vitro* assays for detection of 5mC hence seven TALEs all targeting the CDKN2A site but bearing the above described NTR mutation/s were expressed (SI Sequence 7). For comparison, these proteins were always used alongside WT CDKN2A TALE (with no modifications).

EMSA assay was performed as described previously, where the TALEs bearing RVD HD targeted a C or 5mC within the target sequence and the hybrid was tracked via the fluorescence readout of the TALE-GFP domain in an EMSA gel. The TALE-DNA complex formed a single defined band which exhibited higher mobility compared to protein without DNA (Figure VI-14 a). Analysis of data revealed that WT CDKN2A TALE showed 2.0 fold greater binding to sequence that contained a C as opposed to that which contained a 5mC. In comparison to other assays (PEX, Enrichment) differentiation between C and 5mC seemed to be diminished in this case probably because the absence of any non-specific DNA caused the TALEs to remain bound to the DNA longer than usual. However, in comparison to the WT every modified TALE bearing alanine mutation/s displayed better discrimination between C and 5mC containing oligonucleotides (Figure IV-37). Starting from the farthest from the CRD, TALEs with mutations K171A and R173A (located in the -3 repeat showed a 1.4 and 1.5-time better discrimination between C from 5mC as compared to the WT. When both the mutations were present together, the number increased to 2.9-folds. The mutation

R236A located in the -1-repeat caused a decrease in overall binding strength, yet increased selectivity by 1.4 folds. The next three mutations were present in the 0 repeat of which the basic amino acids at positions K265 and R266 are shown to interact with the DNA phosphate backbone in the crystal structure [156]. TALEs containing either one of these amino acids exchanged with alanine showed an approximately 3 folds increased selectivity for C over 5mC. A triple mutation of K262A&K265A&R266A in a TALE reduced DNA binding drastically while still exhibiting C/5mC selectivity (SI Figure VI-14 d) (Figure IV-37).

To see if the same effect could be seen in another sequence context, TALEs bearing the same set of mutations were expressed to target the ZAP-70 locus. The sequence contains the variable C or 5mC site at position 13 as opposed to CDKN2A sequence where the C/5mC position of interest occurs at the 5th position. Since the triple mutation caused extremely low binding, this mutation was left out of the set. Conduction and analysis of EMSA (SI Figure VI-15 a) exhibited a similar trend of improved specificity compared to the WT TALE. K171A, R173A containing ZAP-70 TALEs showed 1.1 and 1.3 folds increase respectively for C/5mC discrimination compared to the WT TALE while a combination of these mutations (K171A&R173A) did not grant more benefit. The highest difference was observed for TALE bearing the R266A mutation (Figure IV-38). These observations led to the conclusion that reduced binding to 5mC can be achieved by reducing the non-specific interaction extended by basic amino acids in the TALE NTR repeats.

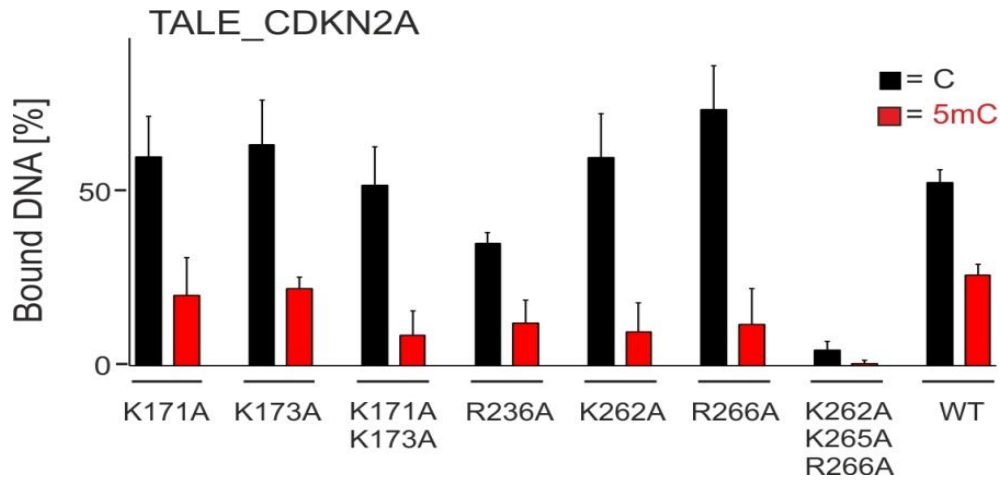


Figure IV-37: Quantification of binding by TALE_CDKN2A carrying different mutation (mentioned below). The Y-axis denotes the percentage of DNA bound by each TALE. Error bars indicate standard error from experimental replicates [21].

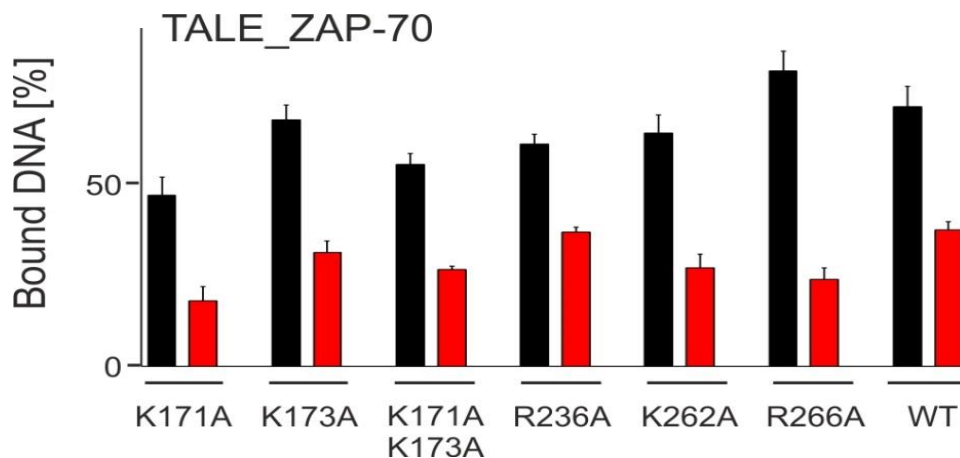


Figure IV-38: Quantification of binding by TALE_ZAP-70 carrying different mutation (mentioned below). The Y-axis denotes the percentage of DNA bound by each TALE. Error bars indicate standard error from experimental replicates. TALEs with triple mutation that showed extremely reduced binding earlier was not used [21].

M. K16A & Q17A mutation TALE CRD repeat reduces binding to 5mC

Charge based non-specific interactions between TALE and DNA are not solely exhibited by the NTR repeats. All repeats in the CRD additionally consists of basic amino acids lysine and glutamine at position 16 and 17 which are also known to bind independent of the sequence context to the DNA backbone directly and via water mediated bond respectively (Figure IV-39).

To extend the analysis of the mutation studies, K16 and Q17 in one, two or three repeats in the CRD of TALE_CDKN2A were changed to alanine. Repeats harbouring these mutations were randomly distributed over the length of the protein including the one opposite the C/5mC variable position (5* for CDKN2A and 13* for ZAP-70) (Figure IV-2). Other selected repeats bearing K16A&Q17A mutations included 3rd, 17th and 23rd.

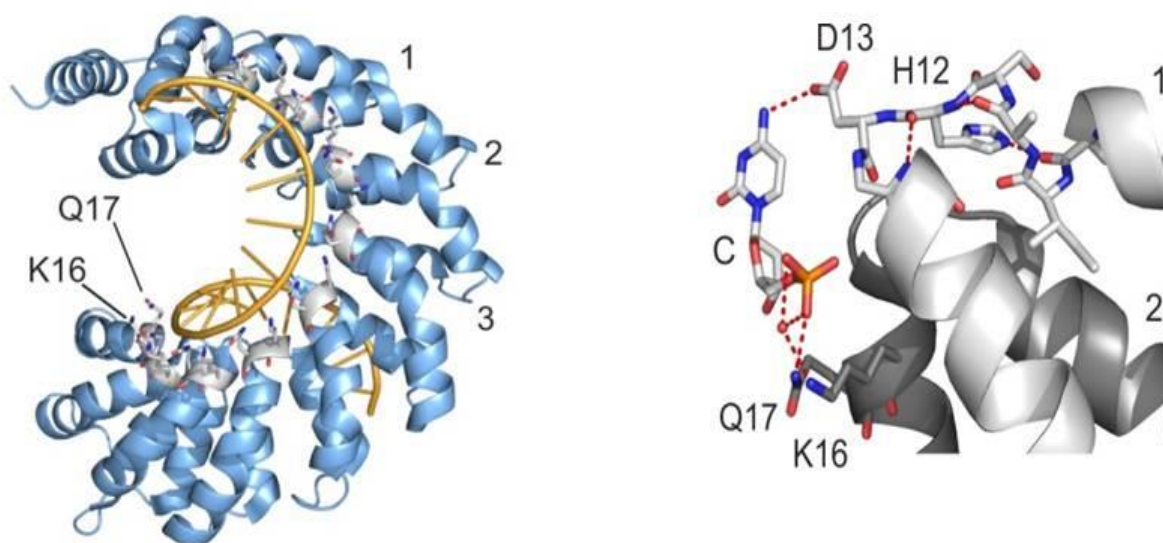


Figure IV-39: Crystal structure of TALE-DNA complex (pdb entry 3V6T) showing the interaction between K16 & Q17 of each repeat with DNA backbone (left). On the right is shown a closer image from the crystal structure for a repeat harbouring RVD HD displaying H-bond between TALE repeat amino acid D13 and DNA nucleobase C in red dashed lines and water molecules as red spheres [21].

In addition, TALEs with K16A&Q17A mutation in two (5th and 17th) and three repeats (3rd, 5th and 17th) were also created. Mutations in single repeats closer to the NTR (positions 3

and 5) enhanced C/5mC discrimination by 5.7 and 7.2 folds compared to WT while those more towards the C-terminal (positions 17 and 23) showed an increase of 1.2 and 2.2 folds over WT respectively (Figure IV-40, SI Figure VI-14 b). This difference was expected given that TALEs wound around the DNA starting from 5' and extended till the 3' ends. Changing K16 and Q17 in two or three repeats simultaneously caused extremely reduced binding by the TALEs.

Once again TALEs bearing K16A&Q17A mutations in repeats 3, 13, 17 and 23 were designed for the ZAP-70 sequence context. Due to abolished activity observed in presence of more than one repeat bearing K16A&Q17A mutation, double and triple repeats bearing the modifications were left out. A similar trend of increase in improvement of C/5mC selectivity was observed such that alteration in the 3rd position showed the highest (2.5 folds) and that on the 23rd position showed the lowest increase in selectivity over the WT TALEs (Figure IV-41, SI Figure VI-15 b).

Hence elimination of basic amino acids occurring not only in NTR but also in CRD of TALEs had a positive outcome for detection of 5mC. Between the two regions, mutations in CRD close to the NTR led to a higher increase in selectivity as compared to the NTR mutations. Moreover, an overall increased discriminatory effect was observed for TALE_CDKN2A probably due the C/5mC variable position being situated more towards the start of the CRD. This difference has also previously been observed in enrichment assay where amongst TALEs targeting four different human loci, TALE_CDKN2A performed the best (Figure IV-8).

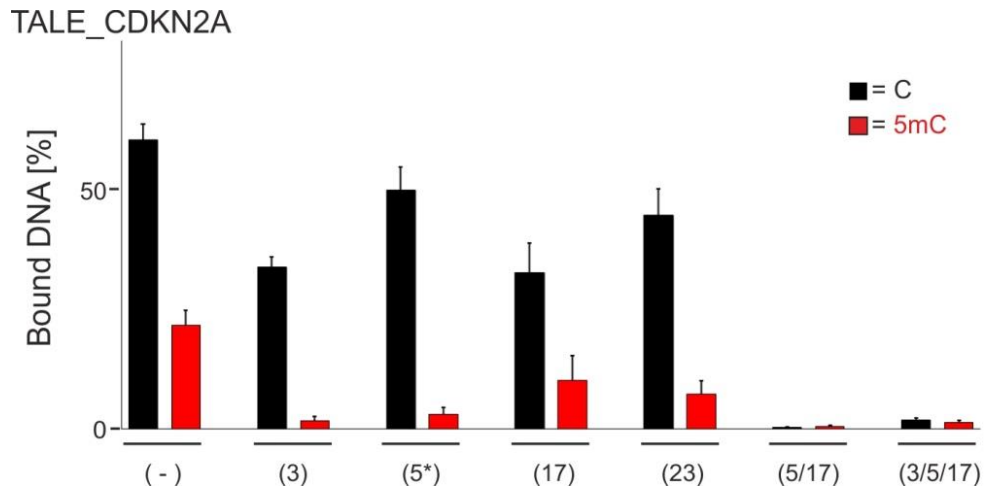


Figure IV-40: Quantification of binding by TALE_CDKN2A carrying K16A&Q17A mutations at different repeats (mentioned below). The Y-axis denotes the percentage of DNA bound by each TALE. Error bars indicate standard error from experimental replicates [21].

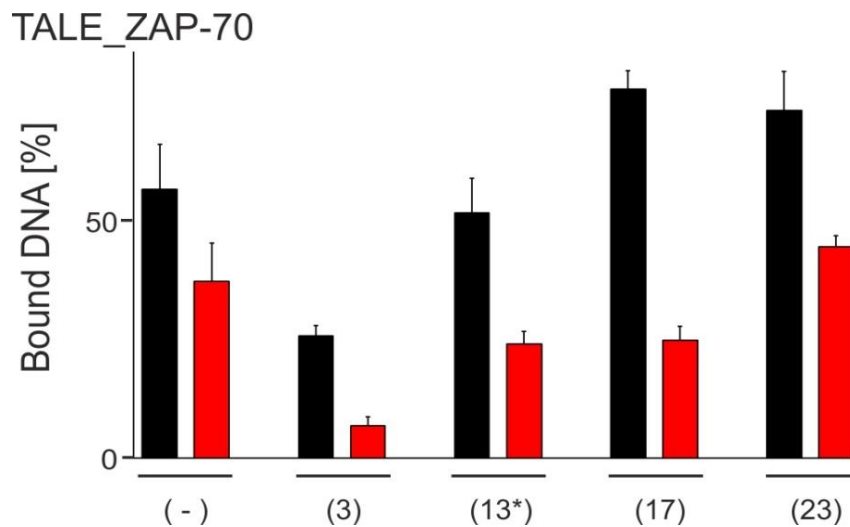


Figure IV-41: Quantification of binding by TALE_ZAP-70 carrying K16A&Q17A mutations at different repeats (mentioned below). The Y-axis denotes the percentage of DNA bound by each TALE. Error bars indicate standard error from experimental replicates. TALEs with K16A&Q17A mutation in multiple repeats that earlier displayed extremely reduced binding were not used [21].

N. Combined mutation in NTR and CRD exhibit highest 5mC sensitivity

The next obvious step to further understand the role of basic amino acid interactions in TALE functionality was to use a TALE with simultaneous mutations in the NTR and CRD. Conditions that performed best in each category were selected and TALEs CDKN2A and ZAP-70 carrying the combinations of mutations in the NTR i.e., K171A&R173A or K262A and the CRD i.e., K16A&Q17A changes in either repeat 3 or 5 of CDKN2A and repeat 3 or 13 of ZAP-70 were respectively created. The selectivity displayed by these combination mutants was higher than those with mutation in either the NTR or CRD of TALEs. However, in comparison to the WT TALE there was a drastic improvement in C/5mC selectivity. The four CDKN2A TALEs with mutation in both NTR and CRD showed an increase in selective binding to C from 11.0 folds for TALEs with NTR K171A&R173A and K16A&Q17A mutations in the 5th repeat to 18.8 folds for TALEs with NTR K262A and K16A&Q17A mutations in the 3rd repeat (Figure IV-42 a, SI Figure VI-14 c). The increase in C/5mC selectivity for ZAP-70 TALEs with mutations in only either TALE NTR or CRD compared to its WT was not as pronounced as those for the CDKN2A context but combined mutations in both regions showed an improvement ranging from 4.0 folds for TALEs with NTR K171A&R173A and K16A&Q17A mutations in the 13th repeat to 50.6 folds for TALEs with NTR K262A and K16A&Q17A mutations in the 3rd repeat (Figure IV-42 b, SI Figure VI-15). It was interesting to note that while mutation in more than one repeat of the CRD inhibits function, a combination of NTR and CRD mutations displayed drastically improved selectivity. Further, since the effect was observed in different repeat positions and sequences, it cannot be considered restricted to a single region, position or sequence context.

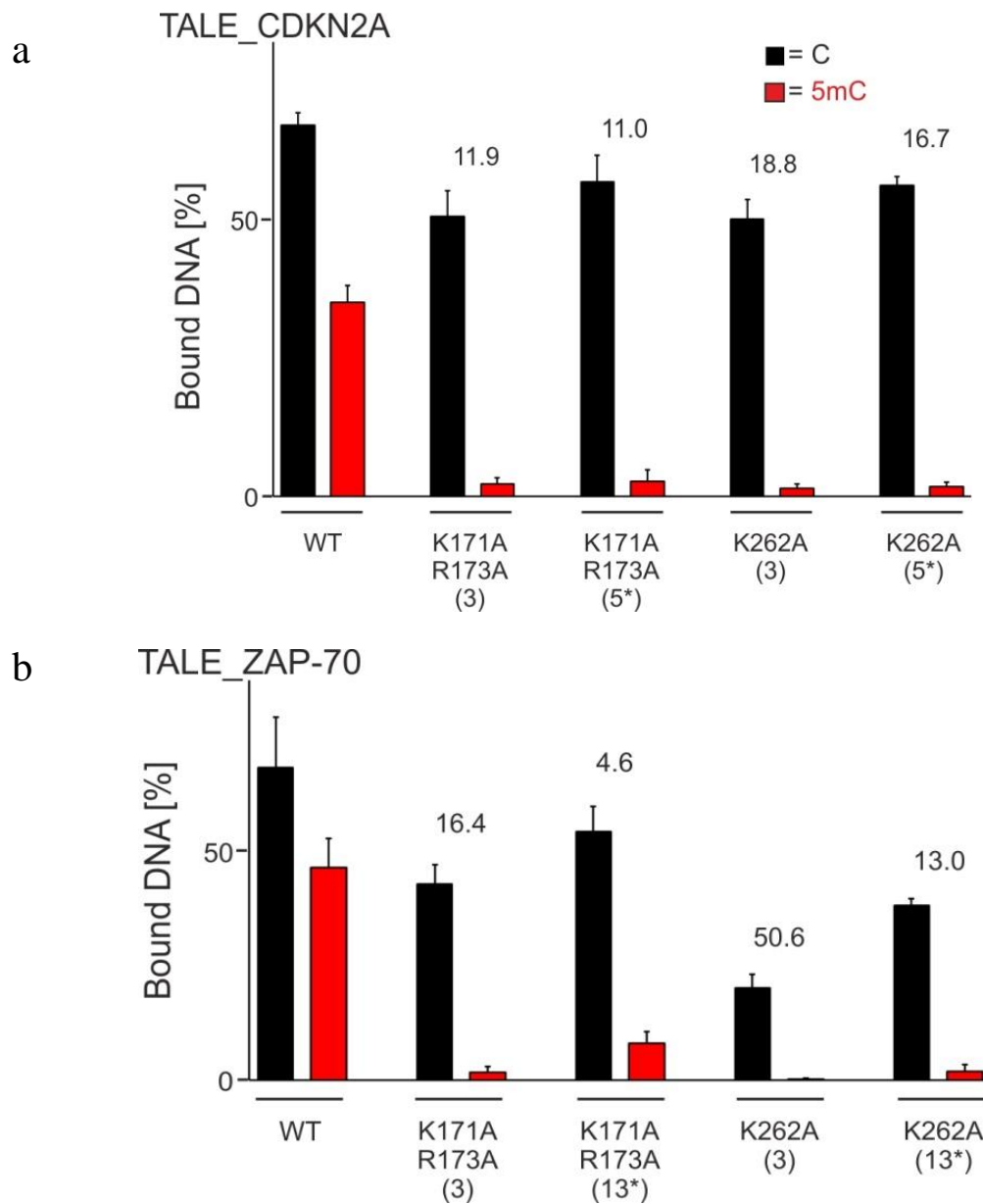


Figure IV-42: Quantification of binding by TALE_CDKN2A (a) and TALE_ZAP-70 (b) carrying mutations at both TALE NTR and CRD repeats (mentioned below). The Y-axis denotes the percentage of DNA bound by each TALE. * next to the number represent the repeat targeting the C site of interest. Error bars indicate standard error from experimental replicates. TALEs with multiple repeats bearing mutation that showed extremely reduced binding earlier was not used in (b) [21].

Detection of 5mC in genomic population at single nucleotide resolution and in the sequence of choice was already shown using the enrichment assay. To test whether the identified electrostatic mutants could also enhance epigenetic base detection in this assay, candidates bearing mutation both in TALE NTR (K262A) and CRD (K16A&Q17A) of the repeat

targeting the C/5mC variable position in the sequence were chosen for both CDKN2A and ZAP-70 TALEs. When used in an enrichment assay with 200 ng of either unmodified or CpG methylated human gDNA, CDKN2A_TALE_K262A and K16A&Q17A (repeat 5) and ZAP-70_TALE_K262A and K16A&Q17A (repeat 13) enriched amount of target DNA containing no methylation comparable with that enriched by their respective WT TALEs. However, the mutant TALEs CDKN2A and ZAP-70 enriched lower amount of methylated DNA resulting in 2 folds or 4.6 folds better C/5mC discrimination respectively than exhibited by their WT counterparts (Figure IV-43). This led to the conclusion that the increased selectivity displayed by the electrostatically mutated TALEs was not confined to one assay and can be used to achieve better detection of methylation in genomic samples using affinity enrichment assay as well.

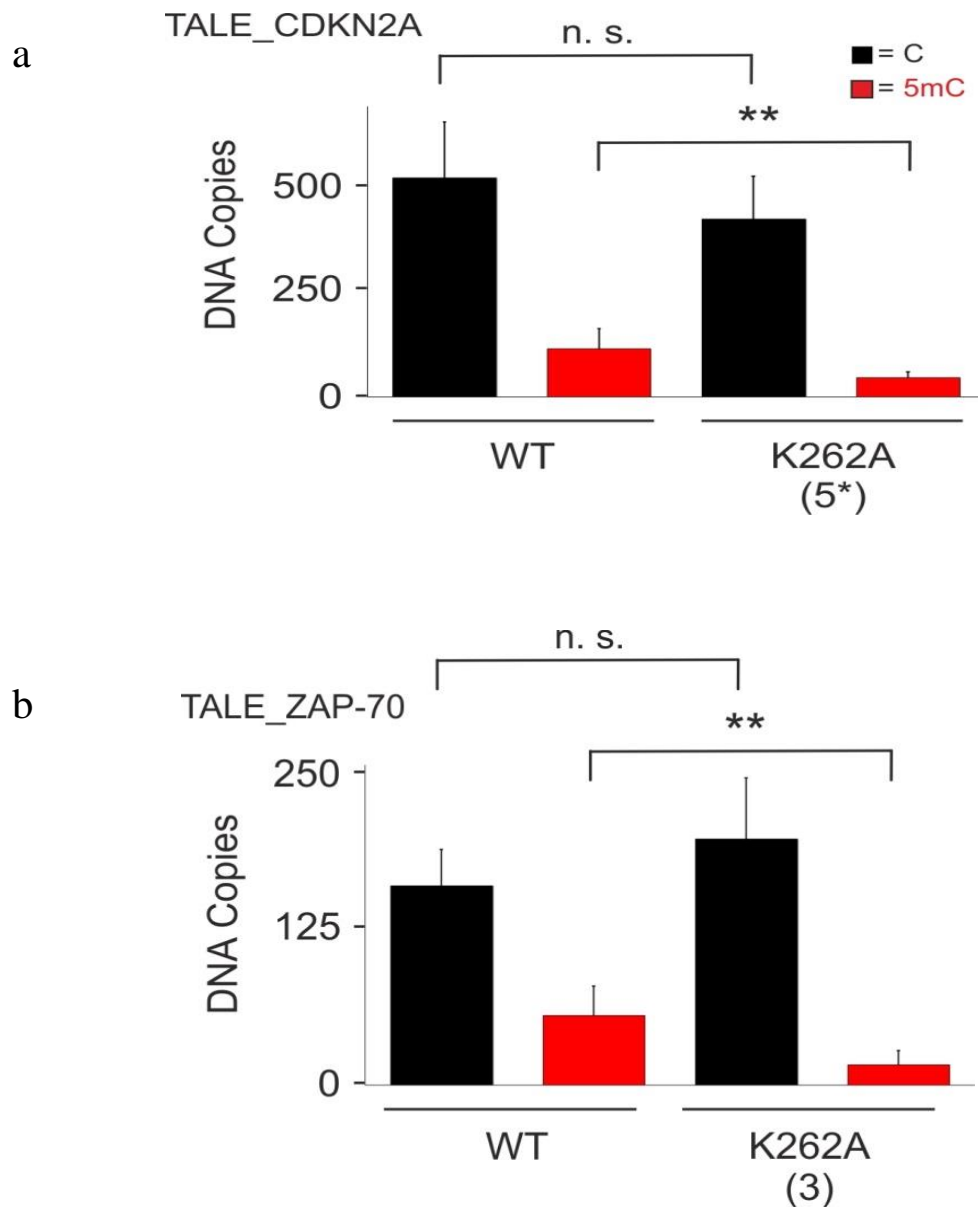


Figure IV-43: Result of affinity enrichment assay using WT TALEs or with mutation (mentioned below) exhibiting the highest C/5mC selectivity for CDKN2A (a) and ZAP-70 (b) target sequence. Error bars indicate standard error from experimental triplicates and qPCR duplicates. ** denotes significant difference calculated using Student t-test with $p \leq 0.01$. n.s. = not significant [21].

One of the most desired goals for developing epigenome engineering tools is to extend their application for detection of epigenetic modifications inside cells. TALEs have been utilized for transcription regulation [169,170], chromatin viewing [171] and combined with nucleases (TALENs) for site specific genomic editing. All the experiments conducted in this

work previously were aimed for *in vitro* detection of 5mC using TALEs hence the next experiment was designed to test if electrostatic mutations in TALEs had a beneficial effect on sensing 5mC inside cells. TALE_CDKN2A WT and modified TALE_CDKN2A_K262A and K16A&Q17A at position 5 were employed in a reporter gene activation assay where two plasmids were co-transfected in HEK293T cells. Plasmids of group A carried the TALE binding site between an upstream mini-CMV promoter and a firefly luciferase gene and that of group B contained the TALE sequence fused to VP64 transcriptional activation domain (SI Figure VI-20).

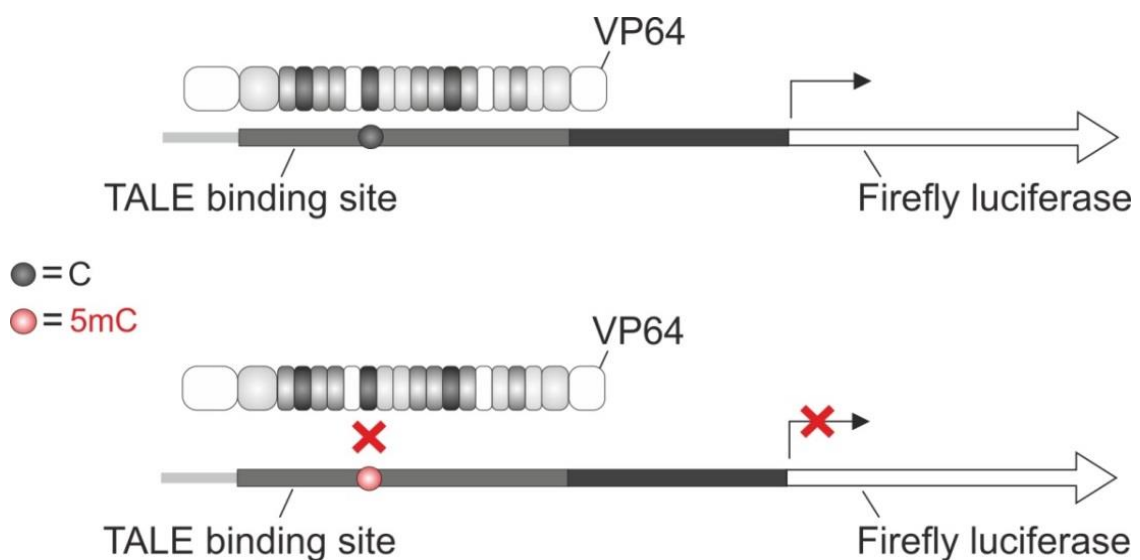


Figure IV-44: Pictorial representation demonstrating the principle behind TALE based *in vivo* transcription activation assay. Grey bar represents the target site and a single grey or red bead represents non-methylated or methylated C of interest. TALE binding and placement of the activator domain VP64 facilitates transcription of the luciferase gene (above). On the other hand, inhibition of TALE binding due to the presence of 5mC, inhibits transcription (below) [21].

Control experiment with group A plasmid containing all CpGs methylated by *in vitro* Dnmt treatment resulted in complete inhibition of luciferase activity probably due to promoter methylation prohibiting any transcriptional factor binding.

To perform unbiased comparative experiments plasmid of group A were created by restriction ligation to contain a single C or 5mC in the binding site and was not transformed

into bacterial cell prior transfection to avoid any uncontrolled methylation or demethylation. Experimental design prompted luminescence because of a VP64 fused TALE binding to the target site bringing about transcription and production of luciferase gene product. Relative luminescence for all samples was calculated in comparison to a positive control (plasmids pAnW814 & pAnW818) displaying highest luminescence. The result showed a selective binding based transcription activation with samples containing no binding site exhibiting the lowest values (Figure IV-45). Evaluation of different controls and experimental samples revealed that samples containing a single 5mC base showed reduced luciferase activity with TALEs bearing RVD HD at repeat 5 but not for those carrying a G* at the same position. Since RVD G* has been earlier shown to bind C or 5mC with equal affinity [16], this meant that the reduced binding by the two HD containing TALEs were solely due to 5mC selectivity. The WT CDKN2A TALE exhibited a difference in luminescence of 1.8 times between C and 5mC containing target site. While luminescence value for C containing samples were similar for WT or mutant CDKN2A TALEs, those for 5mC containing samples were significantly reduced for the mutant TALE increasing the C/5mC selectivity difference to 6.3 folds. Upon subtraction of background the mutant TALE displayed a difference of 102 folds (Figure IV-46).

Sequencing data collected for bisulphite converted samples post transfection confirmed that the inserted 5mC was not removed by transfection into the HEK293T cells and that the observed decrease in luminescence was solely due to the presence of a single 5mC on the target sequence (SI Figure VI-21).

To further characterize binding by the WT TALEs and those bearing alanine mutations, FRET assay was conducted with different TALE concentrations while keeping the DNA concentration constant to determine their K_i . The Cy5 emission fluorescence values were plotted against the TALE concentrations in Origin and K_i values were obtained using the dose response curve with hill fit. Although the K_i values do not indicate a significant difference, the curves show clear difference in saturations establishing clear superiority of selectivity exhibited by the electrostatic mutation for both CDKN2A and ZAP-70 sequence context (SI Figure VI-17, SI Figure VI-18).

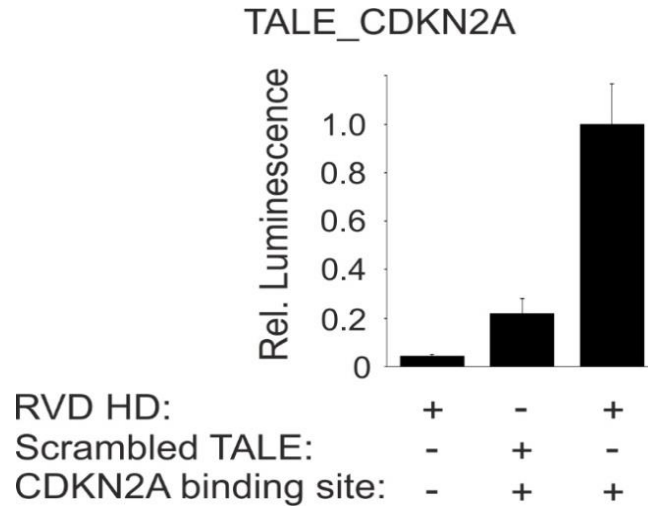


Figure IV-45: Controls used for *in vivo* activation of transcription by TALEs. Samples with no binding site and WT TALE and those with HD containing binding site and WT TALE showed lowest (negative control) and highest (positive control) luminescence respectively. Relative luminescence was calculated relative to the positive control [16].

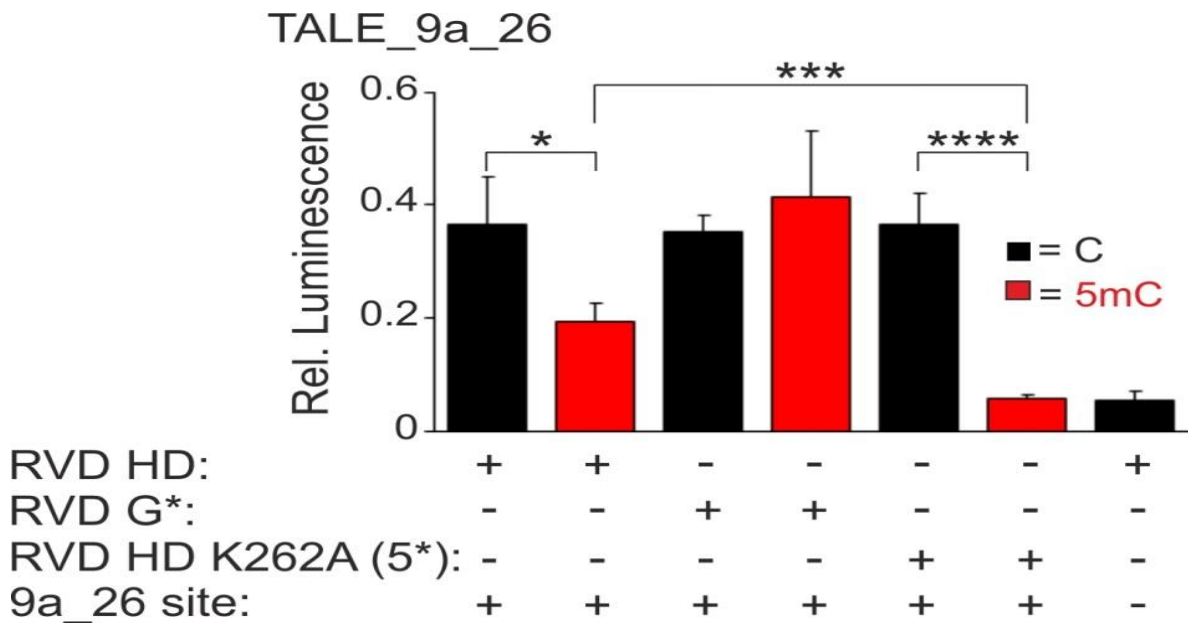


Figure IV-46: Result from TALE based *in vivo* transcription activation assay. Y-axis represents the relative luminescence values. Error bars represent standard error from experimental triplicates. significant difference was calculated using Student t-test with * = $p \leq 0.05$, *** = $p \leq 0.005$ and **** = $p \leq 0.001$ [16].

V. SUMMARY AND OUTLOOK

Various epigenetic processes are critical to regulate normal cellular development and differentiation. This fact is supported by the observation of abnormal epigenetic regulatory controls in disease states. Understanding how different epigenetic processes are synchronized will likely have important and far-reaching implications for human biology and diseases in this post-genomic era. Epigenomic data can be useful in the clinical management of human diseases and the potential applications of epigenetic biomarkers in optimizing customized drug and therapies.

The past decade has experienced a huge boost in the development of methods and techniques for genome engineering. Amongst other tools, the use of TALE proteins has been a popular choice due to their ability to target a DNA sequence of choice and direct access to epigenetic modifications on DNA nucleobases. Ease in design and production and successful *in vivo* applications strengthened their position as a technique of choice for sequence specific detection or editing in genomes. Moreover, discovery of naturally occurring and engineered TALE RVDs as discriminatory sensors for various epigenetic C modifications further supported the idea of utilizing TALEs for detection of epigenetically modified DNA nucleobases.

A novel method was designed and optimized to isolate sequences of user-defined loci directly from whole genomic sample while exhibiting sensitivity towards epigenetic modification occurring on a single nucleotide resolution. This affinity enrichment assay was shown to perform better than state-of-the-art immunoprecipitation assays for detection of hemi-methylation or in samples bearing extremely low percentage of 5mC. Successful detection of a single methylated CpG in the ZAP-70 target sequence that is indicative of the type of CLL in adults serves as an excellent example to highlight the potential of TALEs as a sensitive diagnostic tool in diseases linked to epigenetic modifications [17].

TALEs bearing alanine mutations displayed drastic improvement in C/5mC selectivity in various competitive, binding assays, genomic enrichment assay as well as in activation of transcription in mammalian cells. Enhanced TALE function obtained by controlled amino acid mutations highlight the potential of improvement in TALE design to reduce non-specific interaction and hence decrease off-target binding [21].

Additionally, TALEs with engineered RVDs were shown to successfully detect and discriminate between prokaryotic epigenetic C modification. 4mC is an extremely rare epigenetic modification and its detection in the prokaryotic genome is troublesome. BS-seq used for to differentiate 5mC and 4mC is not very efficient and the use of SMRT is extremely expensive. In comparison use of TALEs for detection of low abundant epigenetic nucleobases in user-defined genomic sequence provides a more efficient and simpler solution.

Certain aspects of TALE functioning have been extensively studied like those pertaining to selectivity extended by naturally occurring TALE RVDs to different canonical nucleobases, effect of length or nucleotide composition of target site on TALE binding and influence of different regions of TALE protein on the strength and sensitivity with which TALEs bind to their target sequence. However, research on their binding mechanism remains very limited. Yet there are no studies that establish clear rules followed by TALEs for recognition and binding to their target sequences. Getting in depth knowledge of their mechanism of action will surely shed light on the way these proteins work and the numerous factors and conditions that affect their function [19].

TALEs have proven to be a powerful tool and have the potential for improvement in various aspects. Recent advances in research combine the use of DNA binding domain such as TALEs with various effectors such as VP64 or enzymes such as DNMTs for site specific engineering of genomes. This suggests continued use of TALEs in future alongside other techniques such as CRISPR-Cas9 system for detection of epigenetic modifications. While discovery of modification specific antibodies for every epigenetic base provides a simplified way of isolation and detection, revolutionary Next Generation Sequencing (NGS) techniques mark the logical method of choice for unravelling genome wide genetic and epigenetic landscape of complex genomes. Understanding the reason behind their existence and roles played by the epigenetic machinery is now the key factor to solve the mystery behind most critical diseases. Hence, these powerful tools mark their position in our lives and in the world of medical research.

VI. SUPPLEMENTARY

A. Primer Extension gel photos

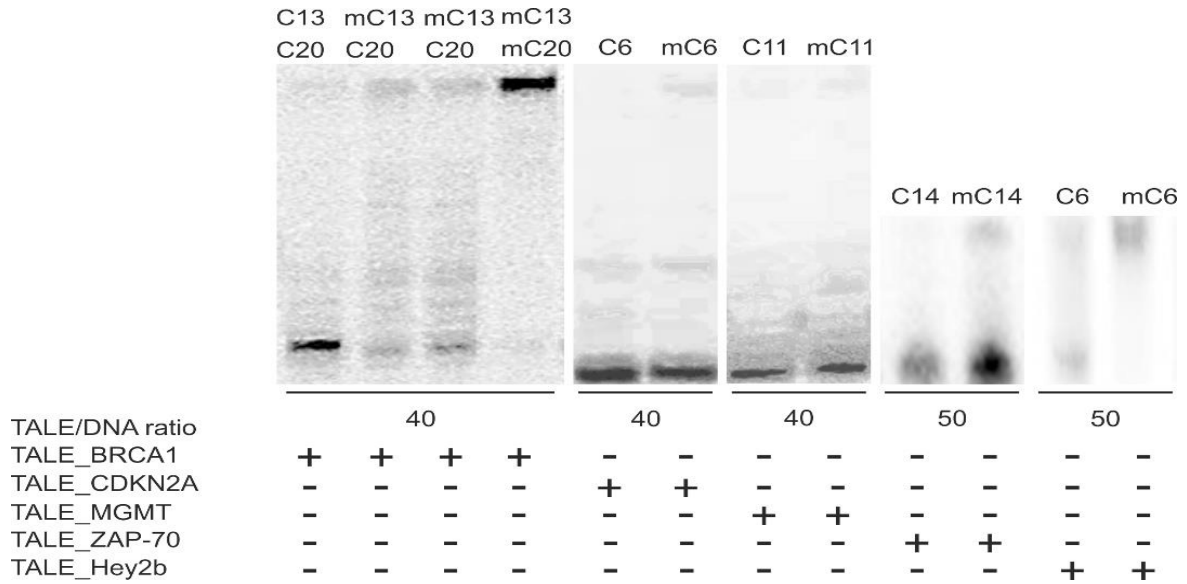


Figure VI-1: PAGE gels with PEX reactions for the mentioned loci. The oligonucleotides used for BRCA1 has been mentioned before (Figure III-7). Oligo pairs o1080 (C6)/o1082, o1079 (mC6)/o1082f or CDKN2A, o1084 (C11)/o1085, o1083 (mC11)/o1085 for MGMT, o1805 (C14)/o1803, o1835 (mC14)/o1803 for ZAP-70 and o1826 (C6)/o1828, o1827 (C6)/o1828 for Hey2_b was used.

B. Ethidium bromide stained agarose gel images

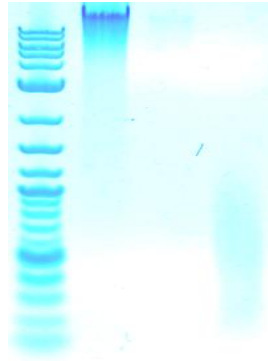


Figure VI-2: 1% Agarose gels showing randomly sheared Zf gDNA. Well 1: 2-log ladder, Well 2: Whole genomic amplified Zf DNA, Well 3: empty and Well 4 Zf DNA sheared randomly by sonication.

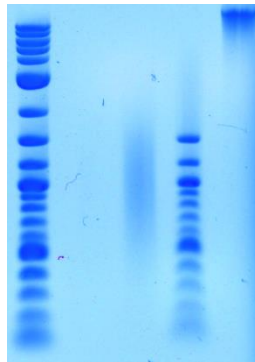
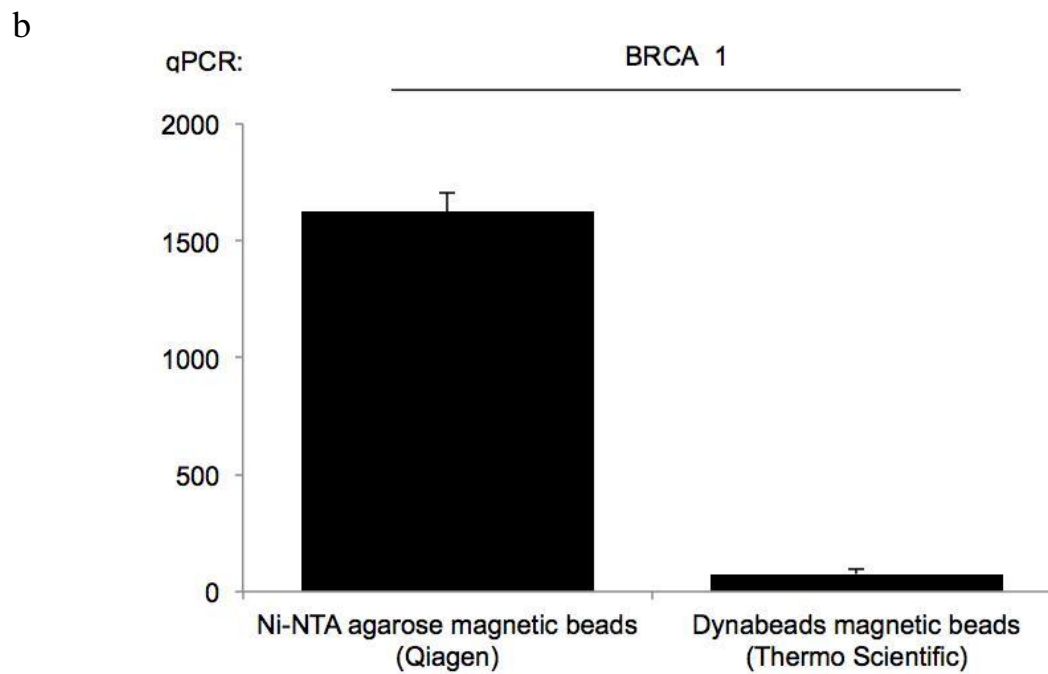
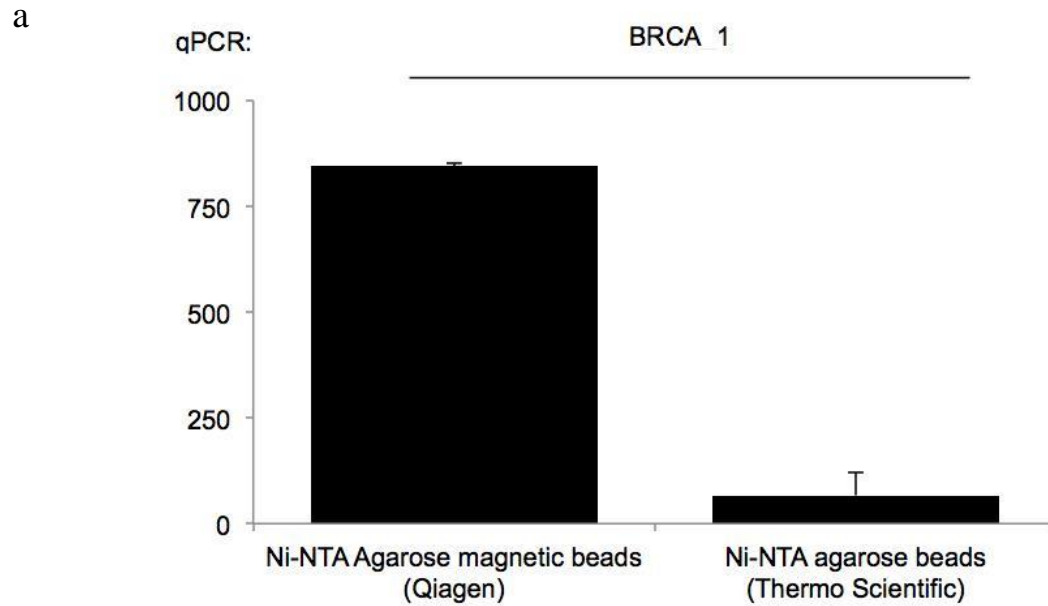
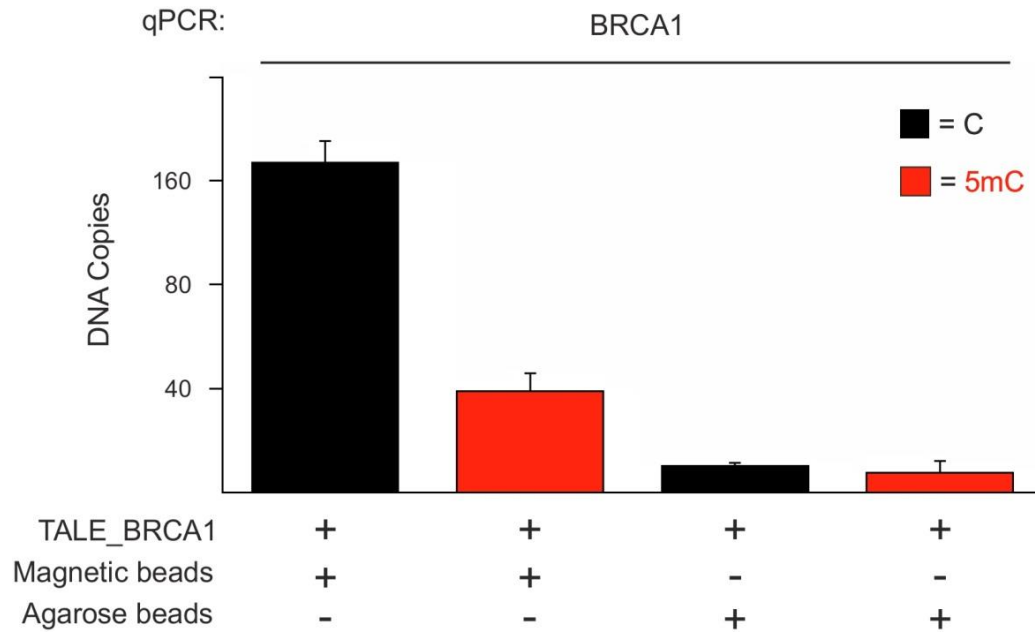


Figure VI-3: 1% Agarose gels showing randomly sheared human gDNA. Well 1: 2-log ladder, well 2: empty, Well 3: WGA human gDNA sheared randomly by sonication, Well 4: 100 bp ladder and Well 5: WGA human gDNA.

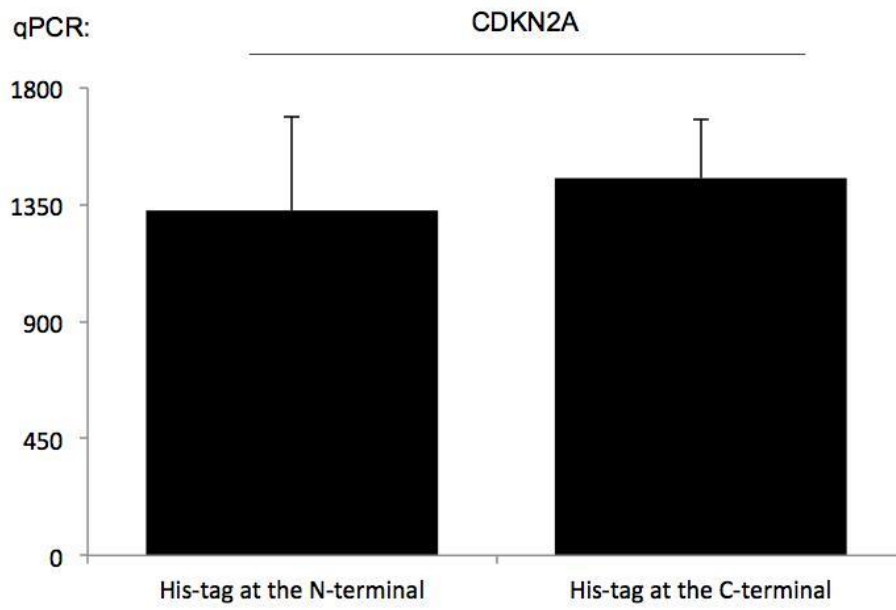
C. Optimization of affinity enrichment using TALEs



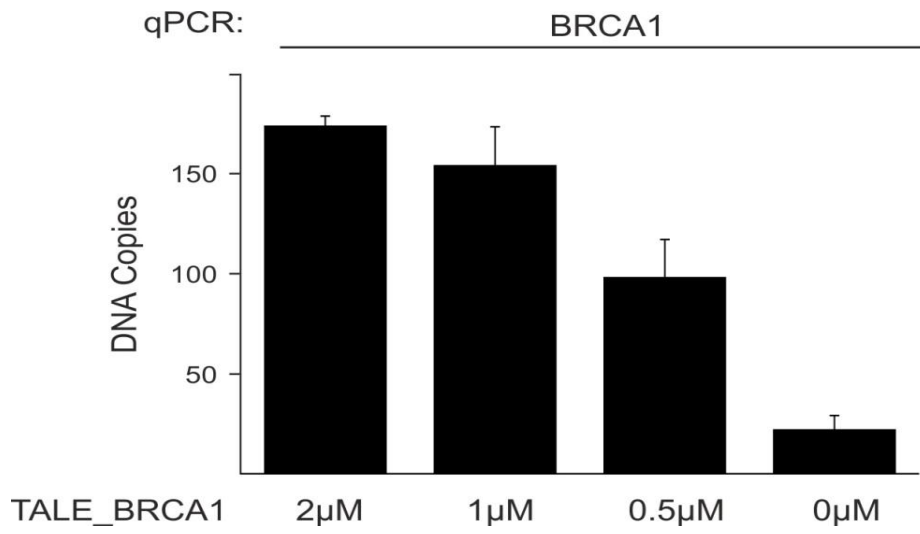
c



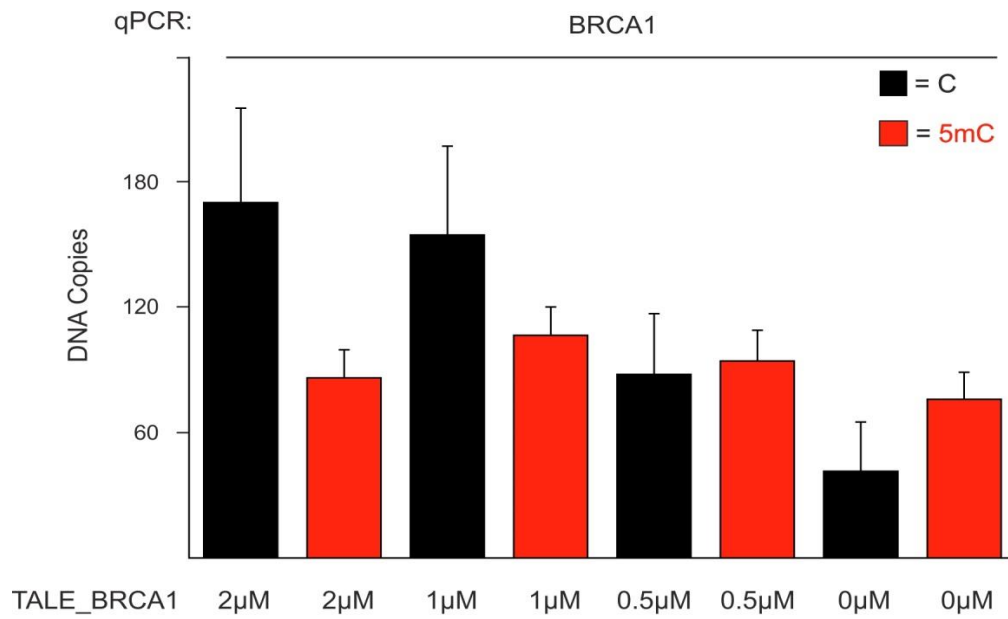
d



e



f



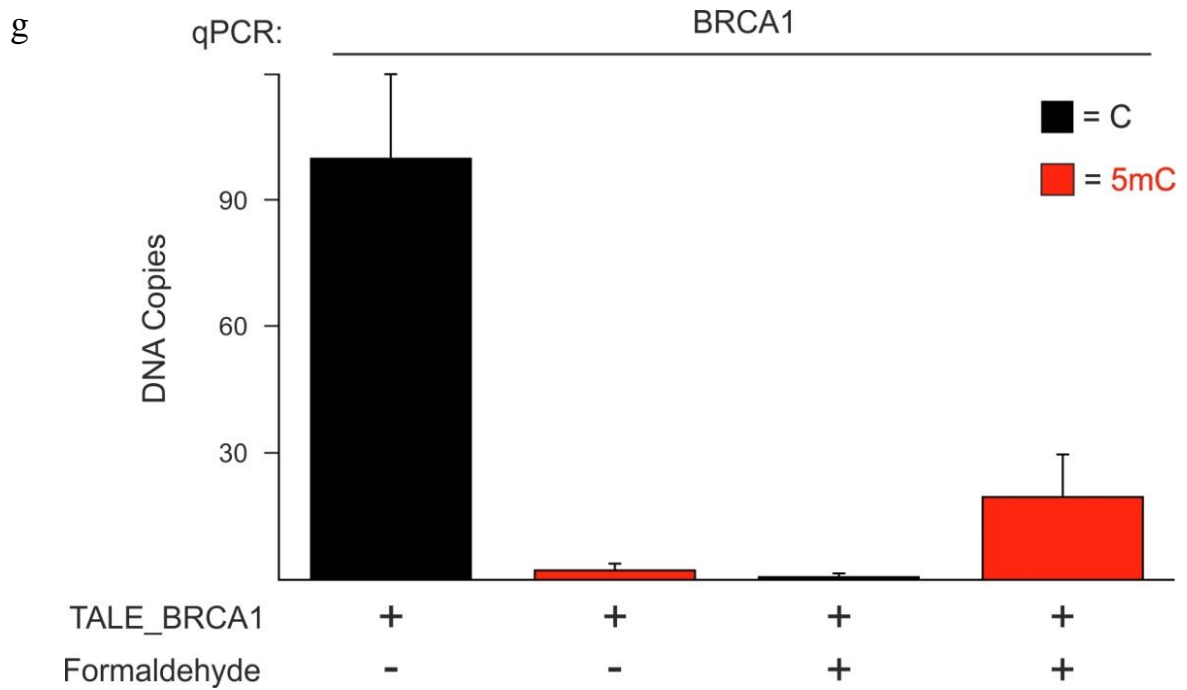


Figure VI-4: Adjusting various factors to optimize efficiency and selectivity of affinity enrichment assay using TALEs. Enrichment assay performed with BRCA1 TALE using one non-magnetic and two magnetic Ni-NTA beads. Results displaying highest enrichment and 5mC selectivity by Ni-NTA agarose magnetic beads for TALE_BRCA1 (a, b, c). Result displaying the effect of position of His6 tag on enrichment (d) using TALE_CDKN2A (Sequence 5, Sequence 6). Experiment conducted using different TALE_BRCA1 concentration displayed increase in DNA sample enriched upon increasing the concentration of TALEs only for non-methylated samples (e, f). Using HCHO to crosslink DNA to TALE protein negatively affected the efficiency and 5mC selectivity by TALE_BRCA1 (g) (Sequence 5). Error bars represent standard error from experimental and qPCR duplicates. Input DNA concentration for each experiment was 200 ng.

D. Sequences of amplicons used in spike-in enrichment assay

BRCA1

CTTCCTCTTCCGTCTCTTTCCTTTTAC**CGTCATCCGGGGGC**AGACTGGGTGGCCAATCCAGA
GCCCCGAGAGACGCTTGGCTCTTTCTGTCCCTCCCATCCTCTGATTGTACCTTGATTTTCGT
ATTCTGAGAGGCTGCTGCTTAGCGGTAGCCCCTTGGTTTCCGTGGCAACGGAAAAGCGCGG
GAATTACAGATAAATTAAACTGCGACTGCGCGGCGTGAGCTCGCTGAGACTTCCTGGACG
GGGGACAGGCTGTGGGGTTTCTCAGATAACTGGGCCCTGCGCTCAGGAGGCCCTCACCCCT
CTGCTCTGGGTAAAGGTAGTAGAGTCCCGGGAAAGGGACAGGGGGCCCAAGTGATGCTCTG
GGTACTGGCGTGGGAGAGTGGATTTCCGAAGCTGACAGATGGGTATTCTTTGAC

BRCA1_a

GTCAAAGAATACCCATCTGTCAGCTTCGGAAATCCACTCTCCCACGCCAGTACCCCAGAGC
ATCACTTGGGCCCCCTGTCCCTTTCCCGGACTCTACTACCTTTACCCAGAGCAGAGGGTG
AAGGCCTCCTGAGCGCAGGGGCCAGTTATCTGAGAAACCCACAGCCTGTCCCCGTCCA
GGAAGTCTCAGCGAGCTCACGCCGCGCAGTCGCAGTTTTAATTTATCTGTAATTCGCGC
TTTTCCGTTGCCACGGAAACCAAGGGGCTACCGCTAAGCAGCAGCCTCTCAGAATACGAAA
TCAAGGTACAATCAGAGGATGGGAGGGACAGAAAGAGCCAAGCGTCTCTCGGGGCTCTGGA
TTGGCCACCCAGTCTGCCCC**CG**GATGACGTAAAAGGAAAGAGACGGAAGAGGAAG

Sequence 1: PCR amplicons from reaction on human gDNA generated for experiment showing strand specific detection of 5mC using TALEs BRCA1, BRCA1_a. The TALE target sequences are highlighted in gray and the CpG of interest is underlined and shown in bold.

Zf Hey2_b PCR product

TGGATTCCCACTCTTCAGCCCCAG**CG**TTACAGCATCTTCAGTGGCTTCTTCCAC**CG**TGAGC
TC**TT****C****CG****TTTCCACATCC**ACCACATCCCAACAGAGCAG**CG**GGAGCAGCAGTAAACCATAC**C**
GAC**CG**TGGGGAAGTGAAGTGGGAG**CG**TTTTAAATGTTGGATTTAAATGTTGGA**CG**TCTTCC
 ATGCTTTGTACATAAAGGAAAGCAG**CG**GCTATTGTGCCTGCTT**CG**GTCAGCAGCATGGGCT
 TTTGTCTTCTCTACACTTGTGCACATATGCAG**CG**TCAAACCTAAGCCAACATTCTGGGAA
 GAAAAGAAAGAGTTTTTACAC**CGT****CG**CACTGTGTTGGAAAC**CG**TAAAGGAAGTTTGTCTG
 TTTTAACAGTGCCTGCATAAACACTGCTAACATGCTGCATTTGAGATGTATGCTTTGATAT
 CATCTGACTTCCACAAACACCCAACAGCAGCTTTAGAGTGAACAGCTTGTCTGAAACAAA
 CCAAAGTTTTGCAGATAATCACTAAAGTGAGGTGTTTGTTTTTTTATCTCTGATTTAACAA
 TCCAGTTTGTAATCTGTACATGTGTAAGATTGTAAGTAGAGTTTATATTGAAATTAGTTC
 ATTGGTATGATGCACTTCAATCACTACTGTTTGTGGGGGAGACAGGATCTTCTC**CG**AT
 TTATACAA

Sequence 2: Amplicon obtained from PCR on Zf DNA generated for conducting all spike-in affinity enrichment assays. Zf Hey2_b target sequence is highlighted in yellow. For MeDIP comparison assay all relevant C/5mC positions occurring in the target sequence are shown in bold and highlighted. M. SssI treated samples bearing 15 C/5mC on each strand are highlighted in green. The relevant CpG for experiments with TALE_Hey2_b T*5 or TALE_Hey2_b N*5 bearing a C, 5mC, 5hmC, 5fC, 5caC or 4mC within the target sequence is shown in red and underlined.

E. Plasmid map and sequences of TRX_TALEs

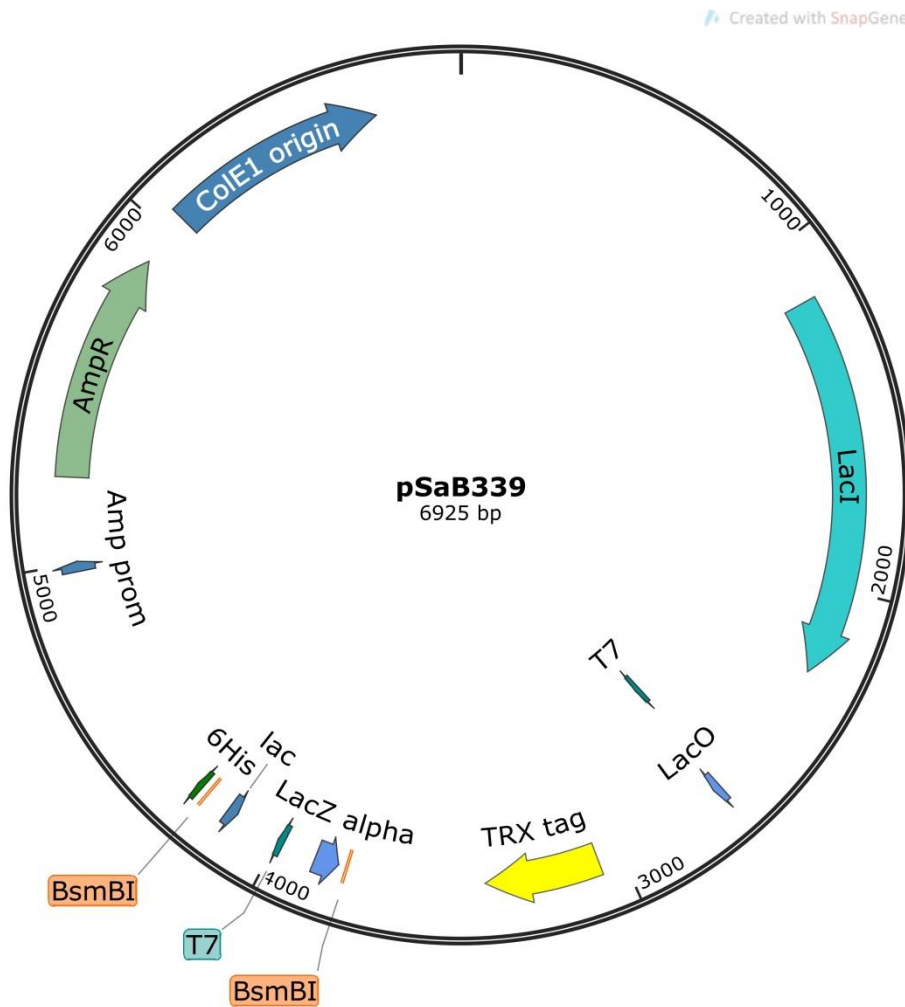


Figure VI-5: Plasmid map showing all the features of the final TALE entry plasmid with TRX tag. The complete TALE sequence is cloned between the two BsmBI sites.

TRX_TALE_Hey2_a (Molecular weight: 97430 Da)

MSDKI IHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNI
 DQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGERQHMDSP
 DLGTVDLRTLGYSSQQQEKIKPKVRSVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVK
 YQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVT
 AVEAVHAWRNALTGAPLNLTDPQVVAIAS **NI**GGKQALETVQRLLPVLCQDHGLTPDQVVAI
 AS **HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NG**GGKQALETVQRLLPVLCQDHGLT
 PDQVVAIAS **NN**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **HD**GGKQALETVQRLLPV
 CQDHGLTPDQVVAIAS **NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NN**GGKQALET
 QRLLPVLCQDHGLTPDQVVAIAS **HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NG**GG
 KQALETVQRLLPVLCQDHGLTPDQVVAIAS **HD**GGKQALETVQRLLPVLCQDHGLTPDQVVA
 IAS **HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **HD**GGKQALETVQRLLPVLCQDHGL
 TPDQVVAIAS **NN**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **HD**GGKQALETVQRLLPV
 LCQDHGLTPDQVVAIAS **NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NN**GGKQALET
 VQRLLPVLCQDHGLTPDQVVAIAS **HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NN**G
 GKQALETVQRLLPVLCQDHGLTPDQVVAIAS **HD**GGKQALESIVAQLSRPDPALAALTNDHL
 LE **HHHHHH***

TRX_TALE_Hey2_b (Molecular weight: 90150 Da)

MSDKI IHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNI
 DQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGERQHMDSP
 DLGTVDLRTLGYSSQQQEKIKPKVRSVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVK
 YQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVT
 AVEAVHAWRNALTGAPLNLTDPQVVAIAS **HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAI
 AS **NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NG**GGKQALETVQRLLPVLCQDHGLT
 PDQVVAIAS **HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **HD**GGKQALETVQRLLPV
 CQDHGLTPDQVVAIAS **NN**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NG**GGKQALET
 QRLLPVLCQDHGLTPDQVVAIAS **NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NG**GG
 KQALETVQRLLPVLCQDHGLTPDQVVAIAS **HD**GGKQALETVQRLLPVLCQDHGLTPDQVVA
 IAS **HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NI**GGKQALETVQRLLPVLCQDHGL
 TPDQVVAIAS **HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NI**GGKQALETVQRLLPV
 LCQDHGLTPDQVVAIAS **NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **HD**GGKQALET
 VQRLLPVLCQDHGLTPDQVVAIAS **HD**GGKQALESIVAQLSRPDPALAALTNDHLLLE **HHHHH**
H*

TRX_TALE_Hey2_c (Molecular weight: 90130 Da)

MSDKI IHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNI
 DQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSQGLKEFLDANLAGSGSGERQHMDSP
 DLGTVDLRTLGYSSQQQEKIKPKVRSVTAQHHEALVGHGFTHAHIVALSQHPAALGTVAVK
 YQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVT
 AVEAVHAWRNALTGAPLNLTDPQVVAIAS **NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAI
 AS **NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NI**GGKQALETVQRLLPVLCQDHGLT
 PDQVVAIAS **HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NG**GGKQALETVQRLLPVLC
 QDHGLTPDQVVAIAS **NN**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **HD**GGKQALETV
 QRLLPVLCQDHGLTPDQVVAIAS **NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NN**GG
 KQALETVQRLLPVLCQDHGLTPDQVVAIAS **HD**GGKQALETVQRLLPVLCQDHGLTPDQVVA
 IAS **NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **HD**GGKQALETVQRLLPVLCQDHGL
 TPDQVVAIAS **HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **HD**GGKQALETVQRLLPV
 LCQDHGLTPDQVVAIAS **NN**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **HD**GGKQALET
 VQRLLPVLCQDHGLTPDQVVAIAS **NG**GGKQALE SIVAQLSRPDPALAAALTNDHLLLE **HHHHH**
H*

Sequence 3: Amino acid sequences and molecular weights of TRX-TALE_Hey2_a, b and c used in enrichment assay for different target sequences in the Zf Hey2 gene. N-terminal TRX sequence and all RVDs are highlighted in green with the RVD targeting the CpG of interest marked in red and underlined. C-terminal His6 tag is highlighted in blue.

F. Plasmid map and sequences of GFP_TALEs

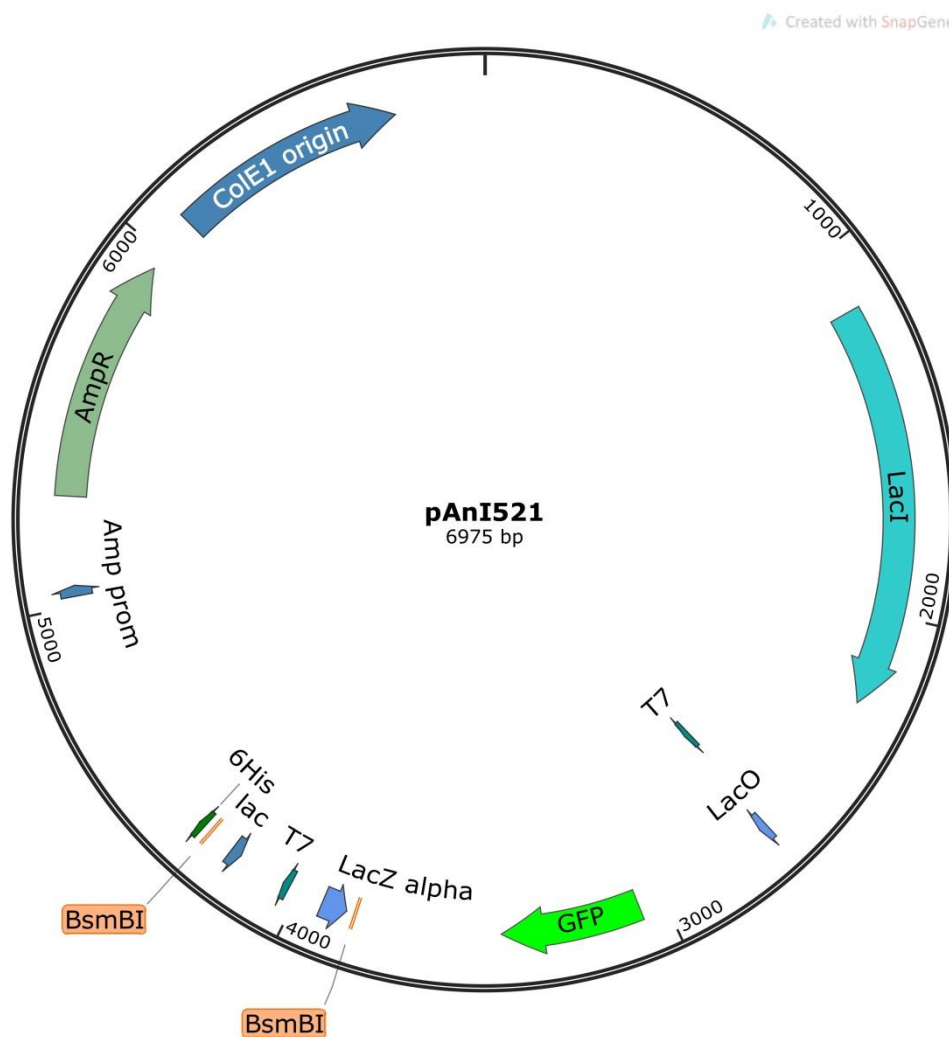


Figure VI-6: Plasmid map showing all the features of the final TALE entry plasmid with TRX tag. The complete TALE sequence is cloned between the two BsmBI sites.

GFP_TALE_Hey2_b (Molecular weight: 102880 Da)

```

MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTTLKFICTTGKLPVPWPTLV
TTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNR
IELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKANFKIRHNIEDGSVQLADHYQ
QNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLEFVTAAGITLGMDELYKTLGYSQ
QQQEKIKPKVIRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHE
AIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKI AKRGGVTAVEAVHAWRNALTG
APLNLTPDQVVAIAS HDGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS NGGGKQALETVQ

```

RLLPVLCQDHGLTPDQVVAIAS **NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **HD**GGK
QALETVQRLLPVLCQDHGLTPDQVVAIAS **HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAI
AS **NN**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NG**GGKQALETVQRLLPVLCQDHGLT
PDQVVAIAS **NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NG**GGKQALETVQRLLPVL
CQDHGLTPDQVVAIAS **HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **HD**GGKQALETV
QRLLPVLCQDHGLTPDQVVAIAS **NI**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **HD**GG
KQALETVQRLLPVLCQDHGLTPDQVVAIAS **NI**GGKQALETVQRLLPVLCQDHGLTPDQVVA
IAS **NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **HD**GGKQALETVQRLLPVLCQDHGLT
PDQVVAIAS **NG**GGKQALESIVAQLSRPDPALAALTNDHLLLE **HHHHHH***

Sequence 4: Amino acid sequences and molecular weights of GFP-TALE_Hey2_HD used in various assays. N-terminal GFP sequence and all RVDs are highlighted in green with RVD HD5 targeting the CpG of interest shown in red and underlined. Amino acids H and D at this position is replaced by any of the 20 natural amino acids to obtain the X* TALEs. The weights of TALE_Hey2_b N*5 and TALE_Hey2_b T*5 are 102740 Da and 102730 Da respectively. C-terminal His6 tag is highlighted in blue.

GFP_TALE_BRCA1 (26) (Molecular weight: 131350 Da)

MSKGEELFTGVVPILEVELDGDVNGHKFSVSGEGEGDATYGKLTCLKFICTTGKLPVPWPPTLV
 TTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNR
 IELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKANFKIRHNIEDGSVQLADHYQ
 QNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMVLLFVTAAGITLGMDELYKTLGYSQ
 QQQEKIKPKVVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHE
 AIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKI AKRGGVTAVEAVHAWRNALTG
 APLNLTDPQVVAIAS **HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NG**GGKQALETVQ
 RLLPVLCQDHGLTPDQVVAIAS **NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NG**GGK
 QALETVQRLLPVLCQDHGLTPDQVVAIAS **HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAI
 AS **HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NG**GGKQALETVQRLLPVLCQDHGLT
 PDQVVAIAS **NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NG**GGKQALETVQRLLPVL
 CQDHGLTPDQVVAIAS **NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NI**GGKQALETV
 QRLLPVLCQDHGLTPDQVVAIAS **HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NN**GG
 KQALETVQRLLPVLCQDHGLTPDQVVAIAS **NG**GGKQALETVQRLLPVLCQDHGLTPDQVVA
 IAS **HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NI**GGKQALETVQRLLPVLCQDHGL
 TPDQVVAIAS **NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **HD**GGKQALETVQRLLPV
 LCQDHGLTPDQVVAIAS **HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NN**GGKQALET
 VQRLLPVLCQDHGLTPDQVVAIAS **NN**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NN**GG
 GKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NN**GGKQALETVQRLLPVLCQDHGLTPDQVV
 AIAS **NN**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **HD**GGKQALE SIVAQLSRPDPALA
 ALTNDHILLE **HHHHHH***

GFP_TALE_BRCA1 (30) (Molecular weight: 145590 Da)

MSKGEELFTGVVPILEVELDGDVNGHKFSVSGEGEGDATYGKLTCLKFICTTGKLPVPWPPTLV
 TTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNR
 IELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKANFKIRHNIEDGSVQLADHYQ
 QNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMVLLFVTAAGITLGMDELYKTLGYSQ
 QQQEKIKPKVVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHE
 AIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKI AKRGGVTAVEAVHAWRNALTG
 APLNLTDPQVVAIAS **HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NG**GGKQALETVQ
 RLLPVLCQDHGLTPDQVVAIAS **HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NG**GGK
 QALETVQRLLPVLCQDHGLTPDQVVAIAS **NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAI
 AS **NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **HD**GGKQALETVQRLLPVLCQDHGLT
 PDQVVAIAS **HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NG**GGKQALETVQRLLPVL
 CQDHGLTPDQVVAIAS **NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NG**GGKQALETV
 QRLLPVLCQDHGLTPDQVVAIAS **NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NI**GG
 KQALETVQRLLPVLCQDHGLTPDQVVAIAS **HD**GGKQALETVQRLLPVLCQDHGLTPDQVVA
 IAS **NN**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NG**GGKQALETVQRLLPVLCQDHGL
 TPDQVVAIAS **HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NI**GGKQALETVQRLLPV
 LCQDHGLTPDQVVAIAS **NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **HD**GGKQALET
 VQRLLPVLCQDHGLTPDQVVAIAS **HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS **NN**GG

RLLPVLCQDHGLTPDQVVAIAS^{HD}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NI}GGK
 QALETVQRLLPVLCQDHGLTPDQVVAIAS^{NN}GGKQALETVQRLLPVLCQDHGLTPDQVVAI
 AS^{HD}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NG}GGKQALETVQRLLPVLCQDHGLT
 PDQVVAIAS^{NG}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{HD}GGKQALETVQRLLPVL
 CQDHGLTPDQVVAIAS^{HD}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NN}GGKQALETV
 QRLLPVLCQDHGLTPDQVVAIAS^{HD}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{HD}GG
 KQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NG}GGKQALETVQRLLPVLCQDHGLTPDQVVA
 IAS^{NN}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NI}GGKQALETVQRLLPVLCQDHGL
 TPDQVVAIAS^{NN}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NN}GGKQALETVQRLLPV
 LCQDHGLTPDQVVAIAS^{HD}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NG}GGKQALET
 VQRLLPVLCQDHGLTPDQVVAIAS^{HD}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NG}
 GKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NN}GGKQALETVQRLLPVLCQDHGLTPDQVV
 AIAS^{NG}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NN}GGKQALESIVAQLSRPDPALA
 ALTNDHLLLE^{HHHHHH}*

GFP_TALE_ZAP-70 (Molecular weight: 131550 Da)

^{MSKGEELFTGVVPILEVELDGDVNGHKFSVSGEGEGDATYGKLTCLKFICTTGKLPVPWPTLV}
^{TTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNR}
^{IELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKANFKIRHNIEDGSVQLADHYQ}
^{QNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMVLLFVTAAGITLGMDELYK}TLGYSQ
 QQQEKIKPKVRSSTVAQHHEALVGHGFTHAHIVALSQHPAALGTAVVKYQDMI AALPEATHE
 AIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKI AKRGGVTAVEAVHAWRNALTG
 APLNLT PDQVVAIAS^{NN}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NI}GGKQALETVQ
 RLLPVLCQDHGLTPDQVVAIAS^{NN}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NI}GGK
 QALETVQRLLPVLCQDHGLTPDQVVAIAS^{NI}GGKQALETVQRLLPVLCQDHGLTPDQVVAI
 AS^{NI}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{HD}GGKQALETVQRLLPVLCQDHGLT
 PDQVVAIAS^{HD}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{HD}GGKQALETVQRLLPVL
 CQDHGLTPDQVVAIAS^{NG}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NN}GGKQALETV
 QRLLPVLCQDHGLTPDQVVAIAS^{NN}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{HD}GG
 KQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NN}GGKQALETVQRLLPVLCQDHGLTPDQVVA
 IAS^{NN}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NN}GGKQALETVQRLLPVLCQDHGL
 TPDQVVAIAS^{NN}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NG}GGKQALETVQRLLPV
 LCQDHGLTPDQVVAIAS^{NN}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NG}GGKQALET
 VQRLLPVLCQDHGLTPDQVVAIAS^{NN}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NI}G
 GKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{HD}GGKQALETVQRLLPVLCQDHGLTPDQVV
 AIAS^{NI}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NG}GGKQALESIVAQLSRPDPALA
 ALTNDHLLLE^{HHHHHH}*

GFP_TALE_BRCA1_a (Molecular weight: 131690 Da)

MSKGEELFTGVVPILEVELDGDVNGHKFSVSGEGEGDATYGKLTCLKFICTTGKLPVPWPPTLV
 FTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFE^{HD}GDTLVNR
 IELKIDFKEDGNILGHKLEYNYN^{NI}SHNVYIMADKQKNGIKANFKIRHNIEDGSVQLADHYQ
 QNTPIGDGPVLLPDNHYLSTQSALS^{NI}KDPNEKRDH^{HD}MVLL^{NI}EFVTAAGITLGMDELYK^{NI}TLGYSQ
 QQQEKIKPKV^{HD}RSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHE
 AIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKI^{HD}AKRGGVTAVEAVHAWRNALTG
 APLNLT^{NI}PDQVVAIAS^{NI}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{HD}GGKQALETVQ
 RLLPVLCQDHGLTPDQVVAIAS^{HD}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{HD}GGK
 QALETVQRLLPVLCQDHGLTPDQVVAIAS^{HD}GGKQALETVQRLLPVLCQDHGLTPDQVVAI
 AS^{HD}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NI}GGKQALETVQRLLPVLCQDHGLT
 PDQVVAIAS^{NI}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NI}GGKQALETVQRLLPVL
 CQDHGLTPDQVVAIAS^{NI}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NI}GGKQALETV
 QRLLPVLCQDHGLTPDQVVAIAS^{NI}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{HD}GG
 KQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NI}GGKQALETVQRLLPVLCQDHGLTPDQVVA
 IAS^{NI}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NI}GGKQALETVQRLLPVLCQDHGL
 TPDQVVAIAS^{NI}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NI}GGKQALETVQRLLPV
 LCQDHGLTPDQVVAIAS^{NI}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NI}GGKQALET
 VQRLLPVLCQDHGLTPDQVVAIAS^{NI}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NI}GG
 KQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NI}GGKQALETVQRLLPVLCQDHGLTPDQVV
 AIAS^{NI}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NI}GGKQALETVQRLLPVLCQDHGLTPDQV
 ALTNDHLLLE^{HHHHHH}*

Sequence 5: Amino acid sequences and molecular weights of GFP-TALEs targeting human genomic sequences. N-terminal GFP sequence and all RVDs are highlighted in green with RVD HD targeting the CpG of interest shown in red and underlined. C-terminal His6 tag is highlighted in blue.

GFP_TALE_CDKN2A_N-terminal His-Tag (Molecular weight: 131760 Da)

MG^{HHHHHH}SKGEELFTGVVPILEVELDGDVNGHKFSVSGEGEGDATYGKLTCLKFICTTGKLP
 VPWPTLVTTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFE
 GDTLVNR^{NI}IELKIDFKEDGNILGHKLEYNYN^{NI}SHNVYIMADKQKNGIKANFKIRHNIEDGSV
 QLADHYQ^{NI}QNTPIGDGPVLLPDNHYLSTQSALS^{NI}KDPNEKRDH^{HD}MVLL^{NI}EFVTAAGITLGMDELY
 K^{NI}TLGYSQQQQEKIKPKV^{HD}RSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAA
 LPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKI^{HD}AKRGGVTAVEAVHA
 WRNALTGAPLNLT^{NI}PDQVVAIAS^{HD}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NI}GGK
 QALETVQRLLPVLCQDHGLTPDQVVAIAS^{NI}GGKQALETVQRLLPVLCQDHGLTPDQVVAI
 AS^{HD}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{HD}GGKQALETVQRLLPVLCQDHGLT
 PDQVVAIAS^{NI}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NI}GGKQALETVQRLLPVL
 CQDHGLTPDQVVAIAS^{NI}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{NI}GGKQALETV
 QRLLPVLCQDHGLTPDQVVAIAS^{NI}GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS^{HD}GG

H. Bisulphite sequencing result for BRCA1 target site

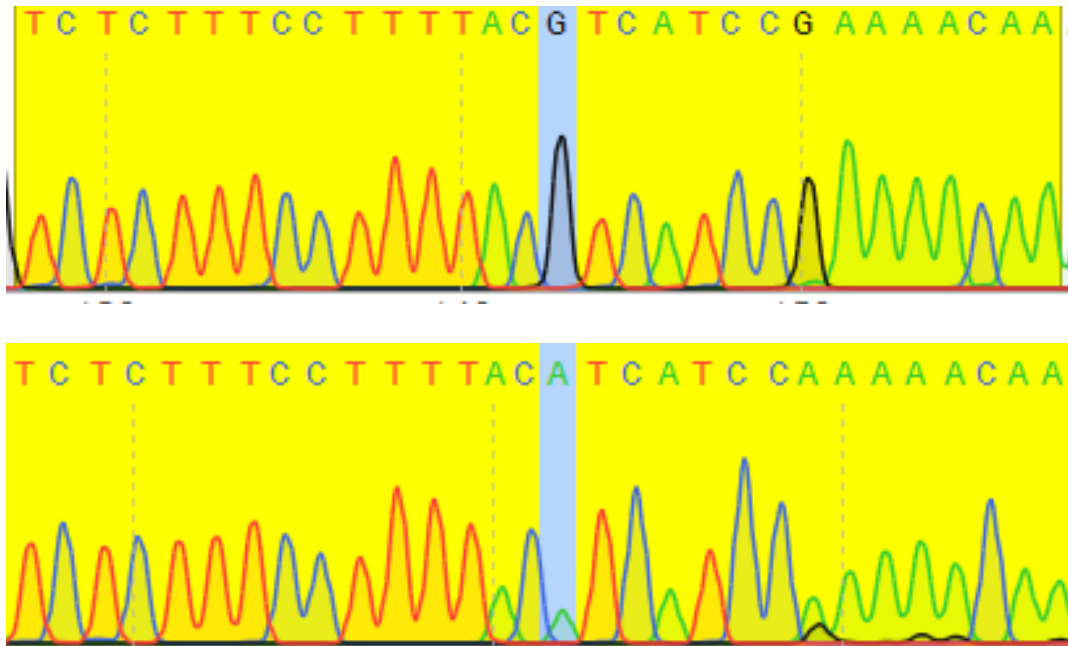
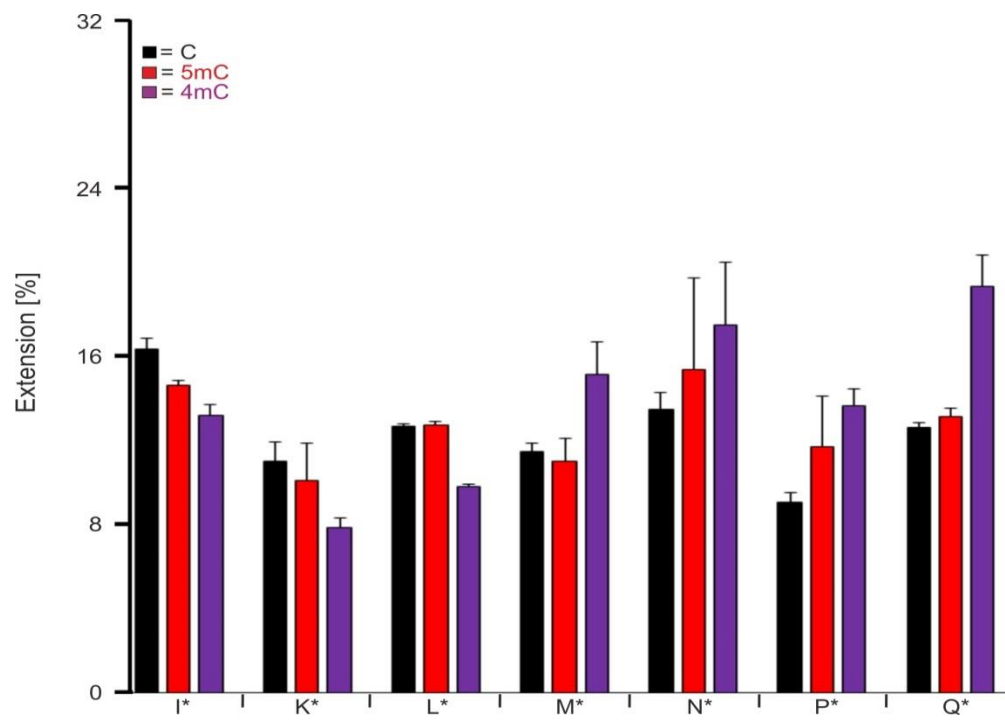
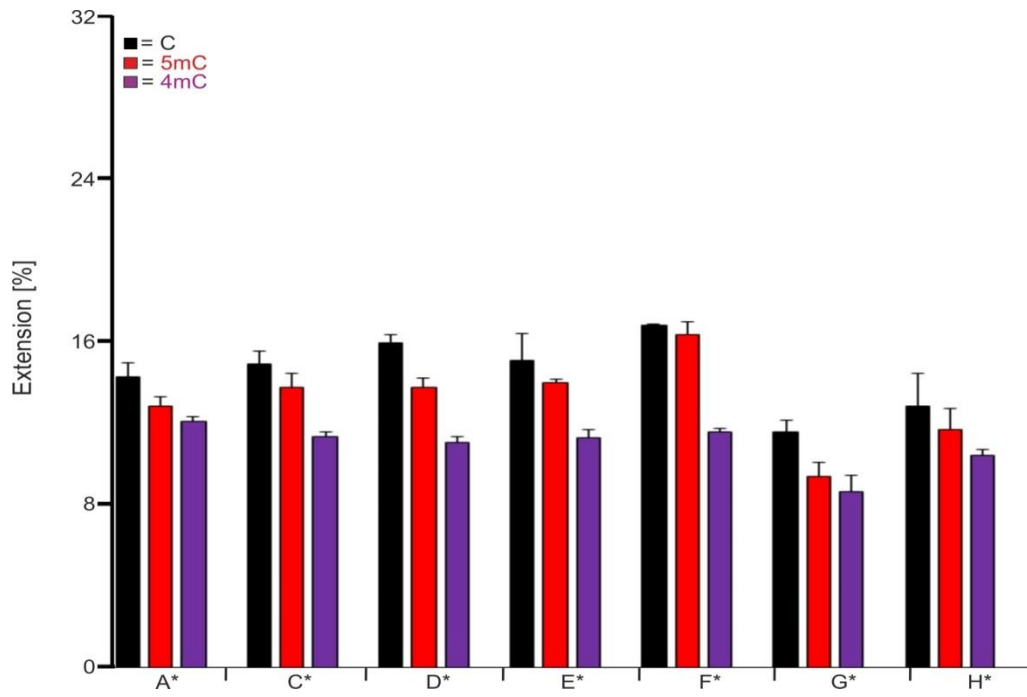


Figure VI-8: Sequencing trace for bisulphite treated sample using primers amplifying the antisense strand of BRCA1 locus. The TALE target sequence is highlighted in yellow and the CpG of interest is highlighted in blue. Comparison of *M. SssI* treated (above) and untreated samples (below) indicated replacement of A (due to presence of C) by G (due to presence of 5mC) in methylated samples.

I. Complete scan of TALE_Hey2_b X* library using PEX



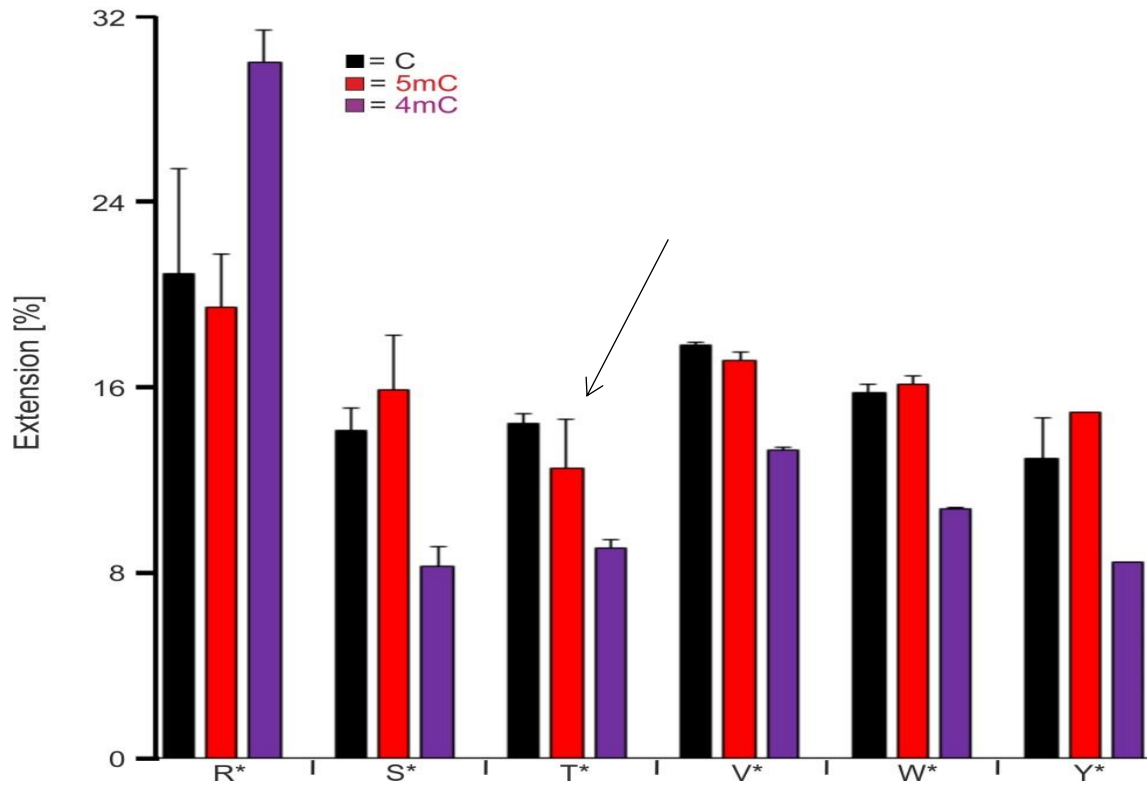
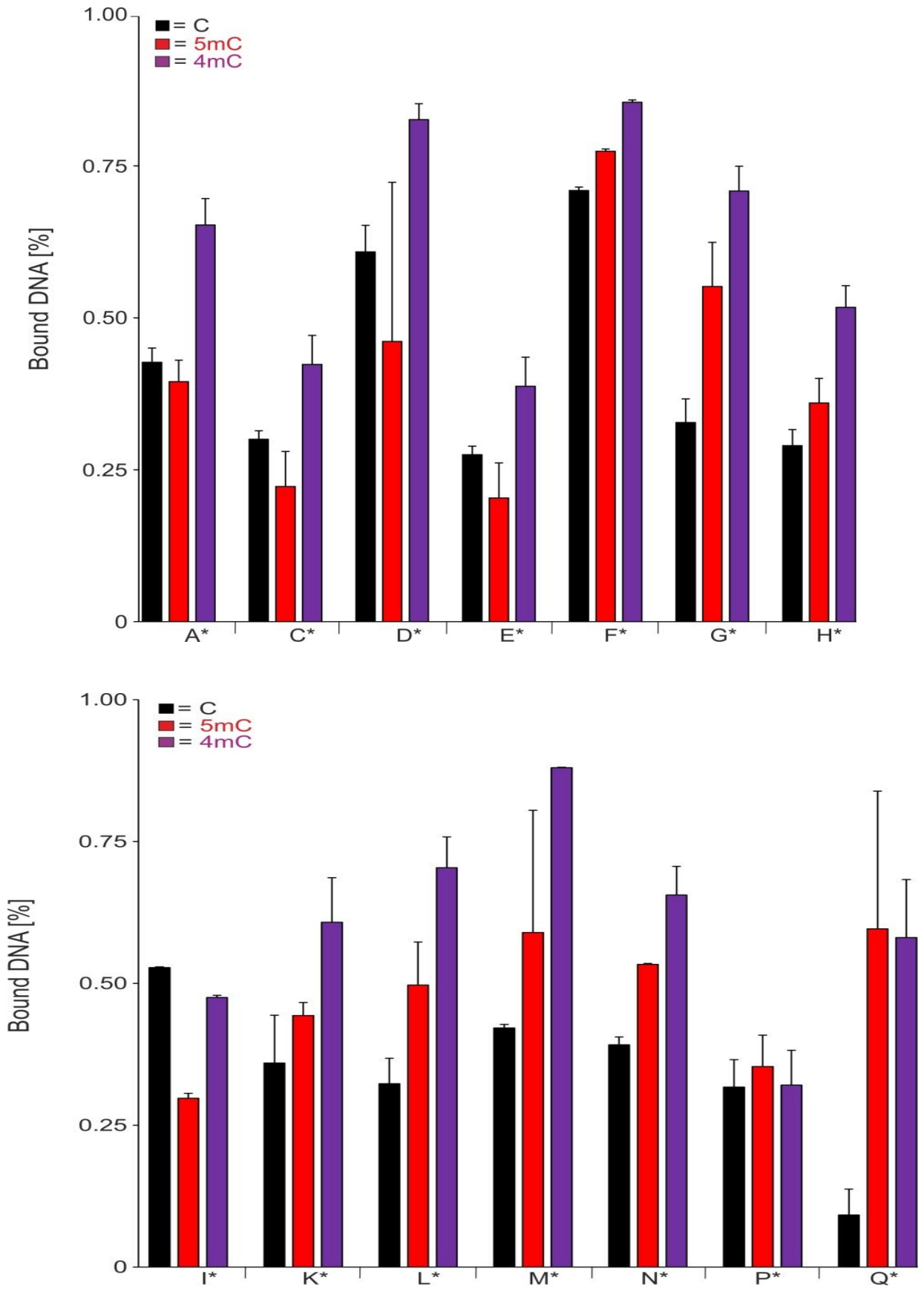


Figure VI-9: Complete scan of Hey2_b X* TALEs (mentioned below) using PEX to find a candidate that binds selectively to 4mC. TALE_Hey2_b T* showed selective binding to 4mC.

J. Complete scan of TALE_Hey2_b X* library EMSA



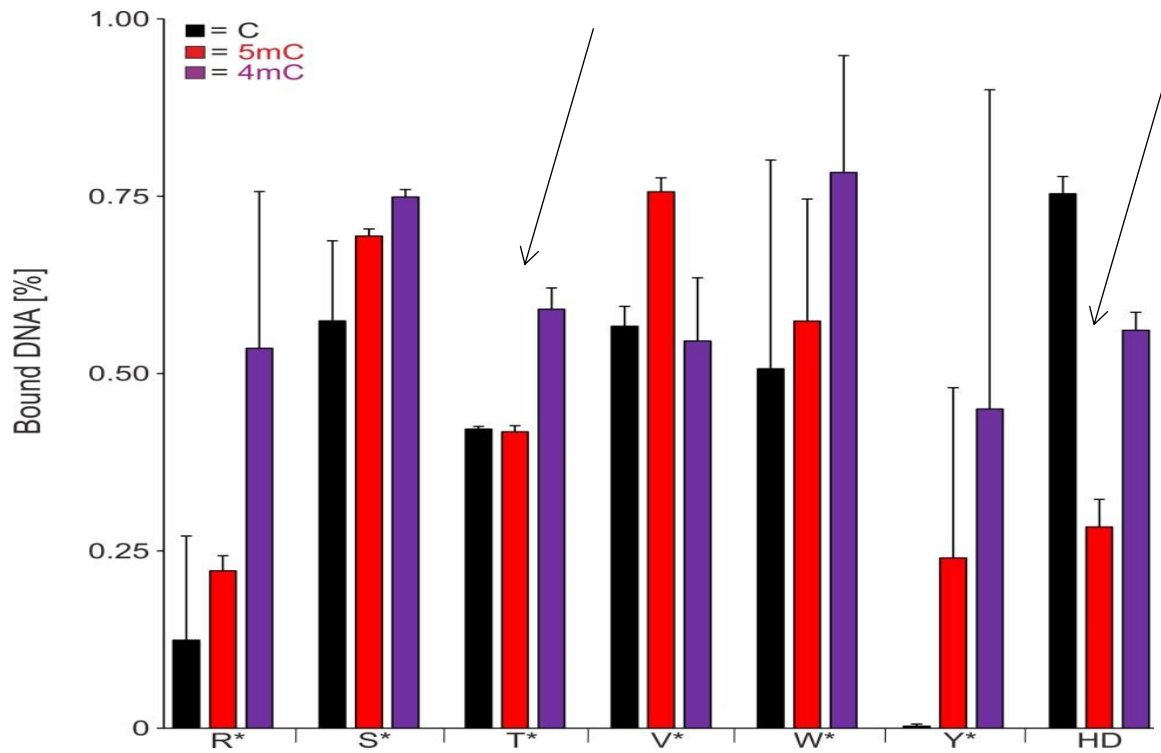


Figure VI-10: Complete scan of Hey2_b X* TALEs (mentioned below) using EMSA to confirm that 4mC shows preferential binding to 4mC whereas TALE Hey2_b HD binds strongly to C.

K. Result of FRET measurement at different time points.

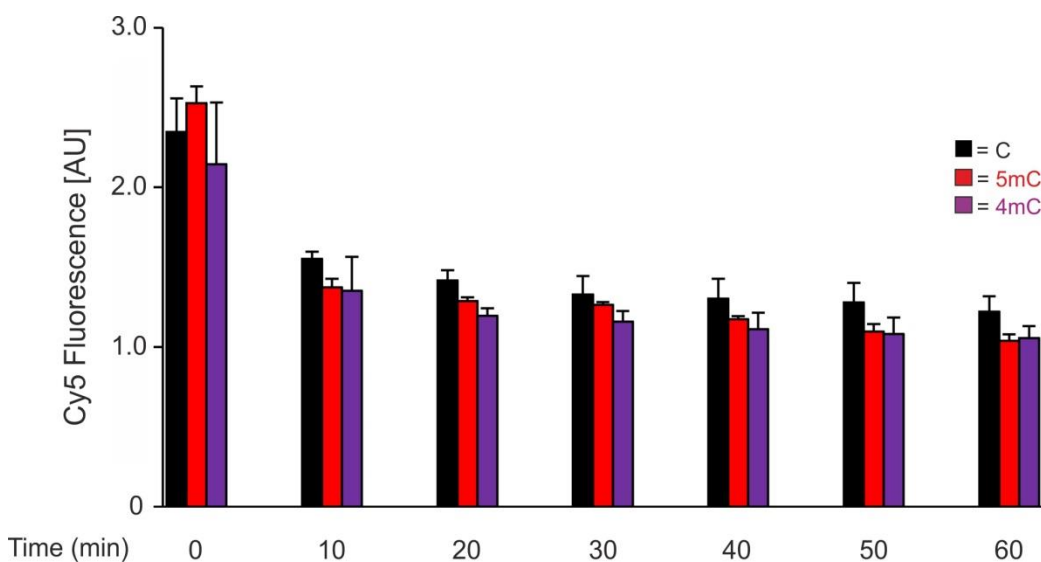
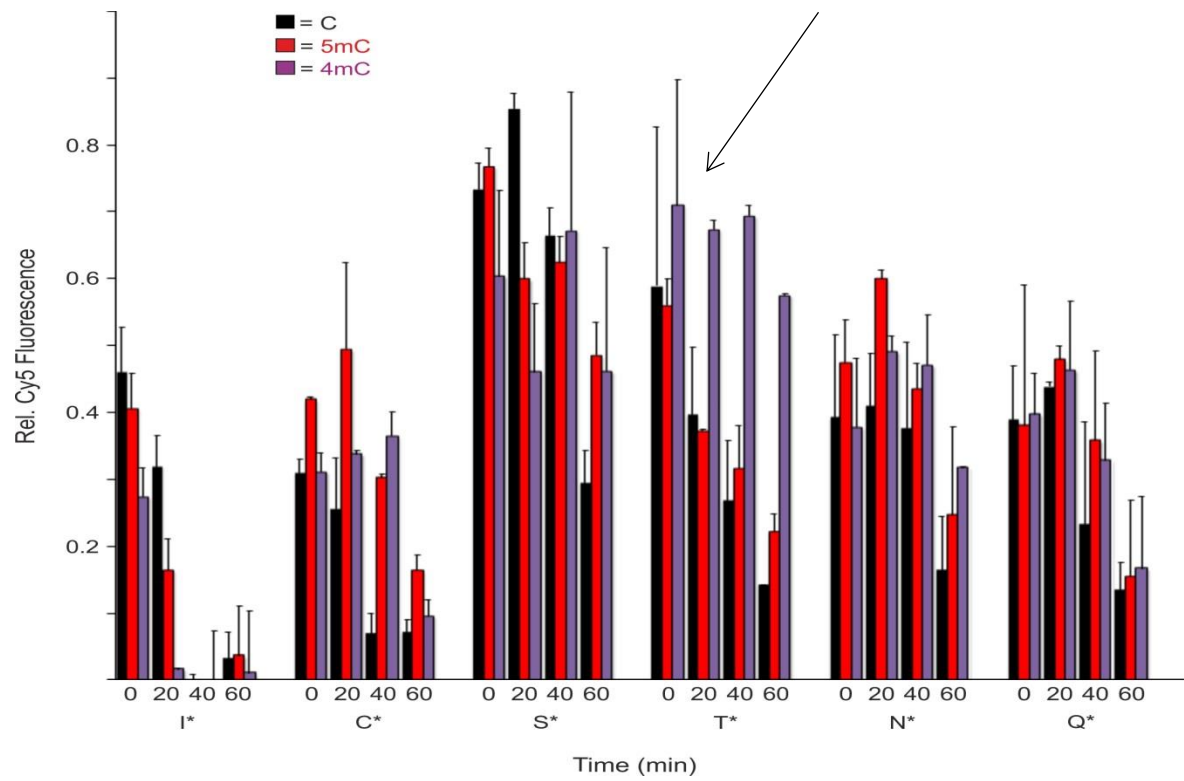
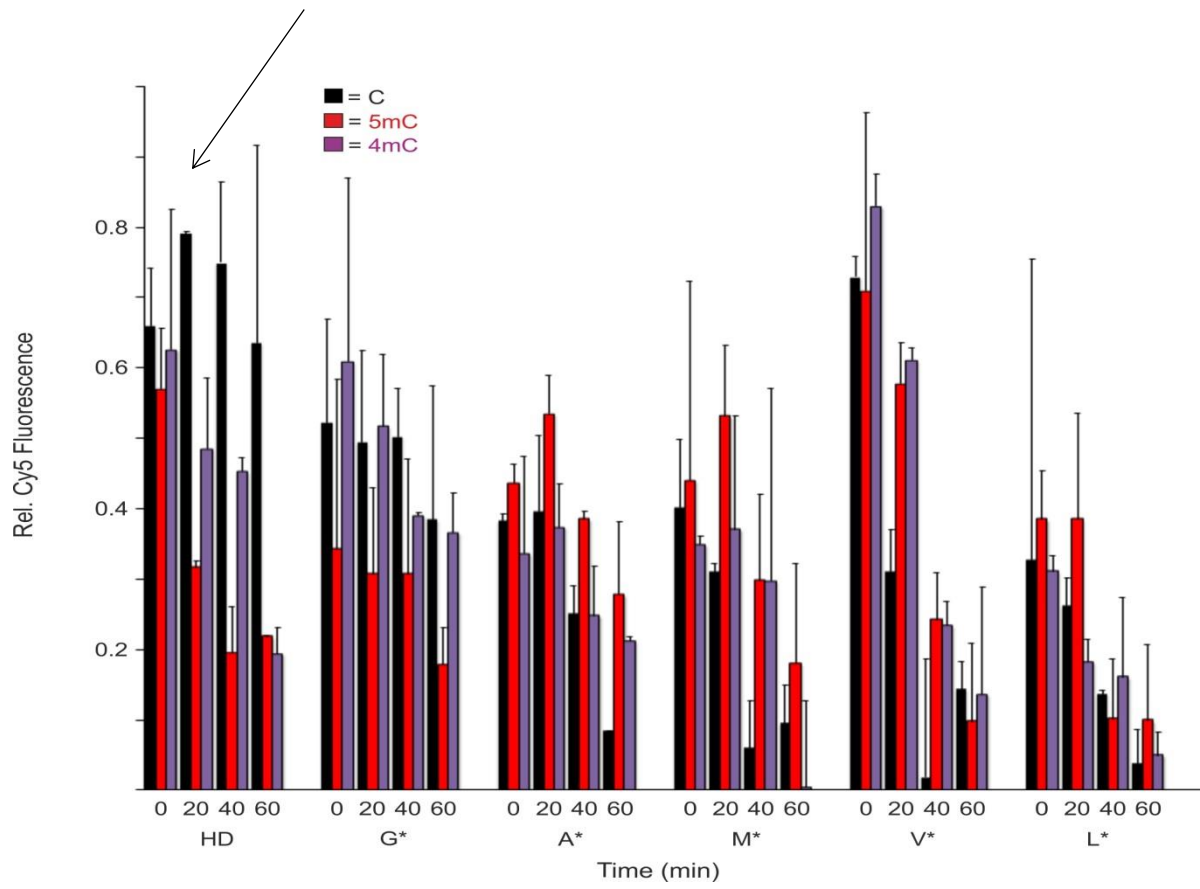


Figure VI-11: FRET control showing highest fluorescence by samples containing C, 5mC or 4mC measured at different time points.



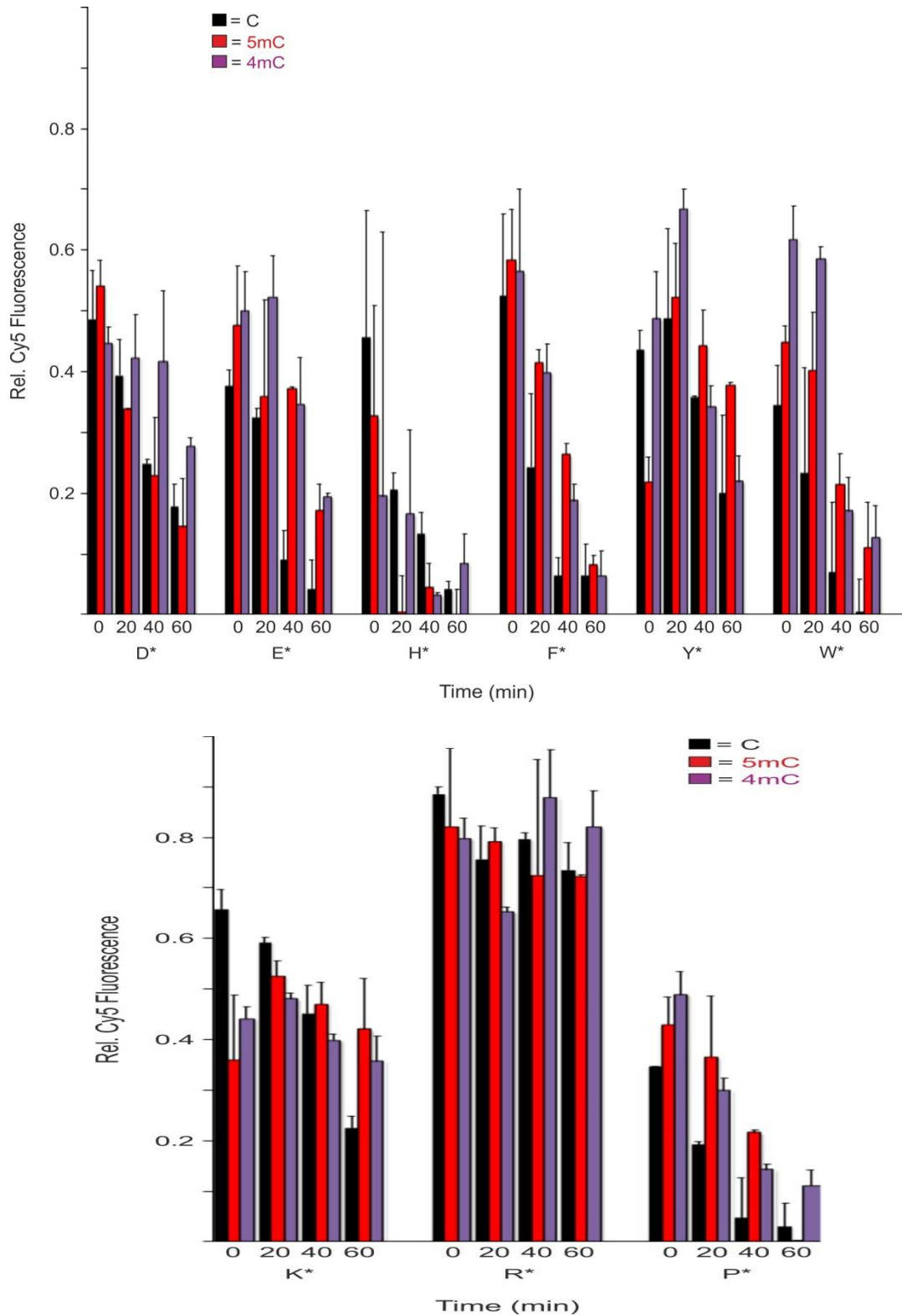
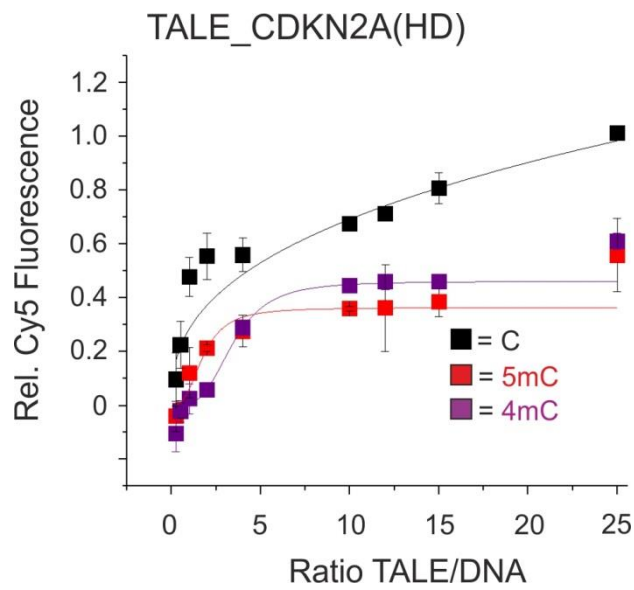


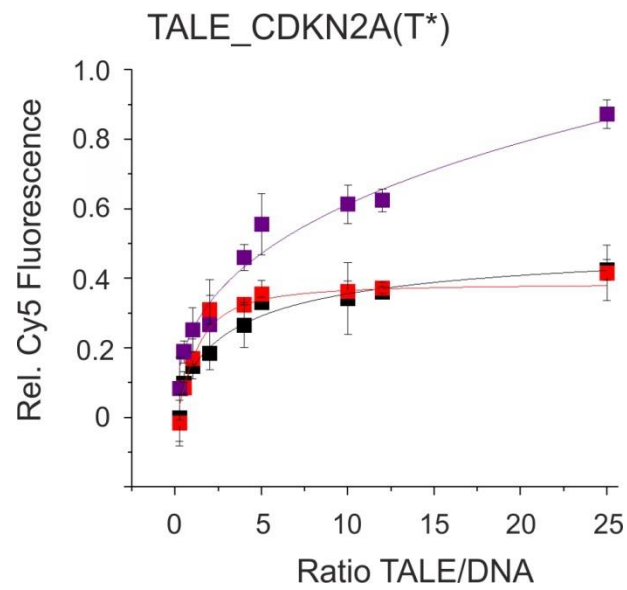
Figure VI-12: Complete scan of Hey2_b X* TALEs (mentioned below) using FRET. Measurements at four different time points after Dnase I addition are shown. The earlier observed effect in PEX assay as well as EMSA assay was confirmed and TALEs Hey2_b HD5 and T*5 were established as selective binders of C and 4mC respectively.

L. Determination of K_i for CDKN2A WT and T* TALEs using FRET

a



b



c

	HD		T*	
	K_i [nm]	F_{max} [AU]	K_i [nm]	F_{max} [AU]
C	1060 ± 50	1.79 ± 2.1	125 ± 40	0.45 ± 0.051
5mC	451 ± 50	0.72 ± 0.4	54 ± 6.5	0.38 ± 0.01
4mC	257 ± 50	0.59 ± 0.0	1322 ± 250	1.73 ± 1.04

Figure VI-13: Fits generated using FRET for TALE_CDKN2A (26) with HD (a) or T* (b) at the fifth position targeting C, 5mC or 4mC containing sequence. Both TALEs show least binding to 5mC with TALE_HD and T* binding strongly to C and 4mC containing samples respectively. Error bars indicate standard error from experimental duplicates. The highest fluorescence value and inhibition constant for each combination is listed in the table (c).

M. TALE sequences with regions bearing alanine mutations

GFP_TALE_CDKN2A (26) used for Electrostatic mutation studies

MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTCLKFICTTGKLPVPWPPTLV
 TTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNR
 IELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKANFKIRHNIEDGSVQLADHYQ
 QNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLEFVTAAGITLGMDELYKTLGYSQ
 QQQEKIKPKV^RSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHE
 AIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKI^AK^RGGVTAVEAVHAWRNALTG
 APLNLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQ
 RLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGK
 QALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAI
 ASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLT
 PDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLC
 QDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETV
 QRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGG
 KQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVA
 IASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGL
 TPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
 LCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALET
 VQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGG
 GKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVV
 AIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALE SIVAQLSRPDPALA
 ALTNDHLLLEHHHHHH*

GFP_TALE_ZAP-70 (26) used for Electrostatic mutation studies

MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTCLKFICTTGKLPVPWPPTLV
 TTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNR
 IELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKANFKIRHNIEDGSVQLADHYQ
 QNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLEFVTAAGITLGMDELYKTLGYSQ
 QQQEKIKPKV^RSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHE
 AIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKI^AK^RGGVTAVEAVHAWRNALTG
 APLNLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQ
 RLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGK
 QALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAI
 ASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLT
 PDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLC
 QDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETV
 QRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGG
 KQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVA
 IASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGL

TPDQVVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVVAIASNNGGKQALETVQRLLPV
LCQDHGLTPDQVVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVVAIASNNGGKQALET
VQRLLPVLCQDHGLTPDQVVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVVAIASNIG
GKQALETVQRLLPVLCQDHGLTPDQVVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVV
AIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVVAIASNNGGKQALESIVAQLSRPDPALA
ALTNDHLLLEHHHHH*

Color code:

K171, **R173**, **R236**, **K262**, **K265**, **R266**, K16A, **Q17A**

Sequence 7: Amino acid sequences of TALE_CDKN2A and TALE_ZAP-70 used in electrostatic mutation studies. All the N-terminal and the CRD amino acids in the repeat targeting the CpG of interest that were changed to alanine are colour coded. In case of TALE_ZAP-70, the best performing candidate had K16A and Q17A mutation on the third repeat as shown in bold and underlined.

N. EMSA gel photos for TALEs bearing electrostatic mutations

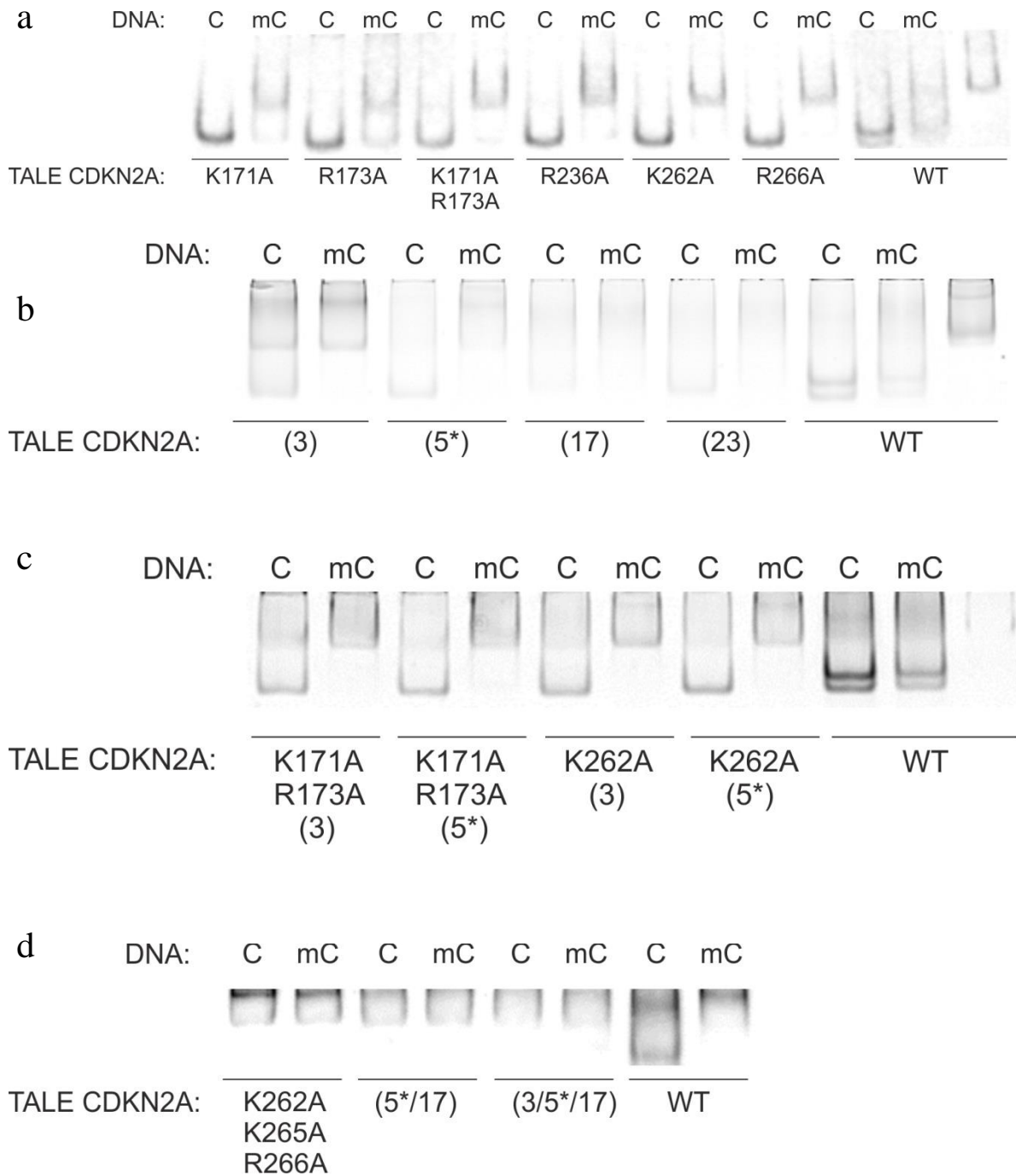


Figure VI-14: Images of EMSA gels showing binding by CDKN2A TALEs to C or 5mC containing target sequence used in the assay. Amino acids replaced by alanine in the N-terminal (a), K16A and Q17A mutations within the repeats (b) and combination of mutation is mentioned below. Samples with mutations that adversely affects binding are also mentioned (d).

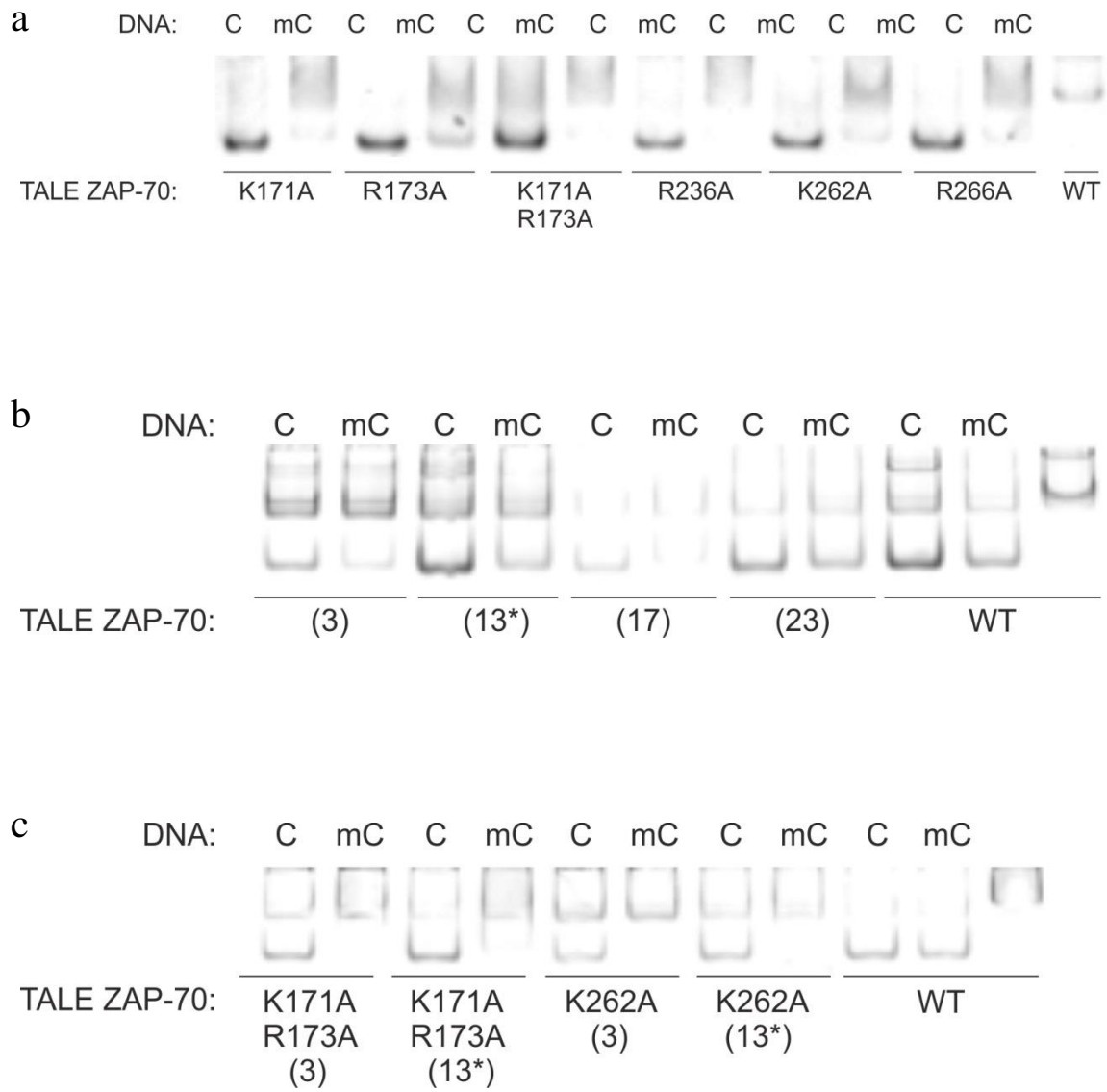


Figure VI-15: Images of EMSA gels showing binding by ZAP-70 TALEs to C or 5mC containing target sequence used in the assay. Amino acids replaced by alanine in the N-terminal (a), K16A and Q17A mutations within the repeats (b) and combination of mutation as mentioned below.

O. Summary of binding by mutated CDKN2A TALEs

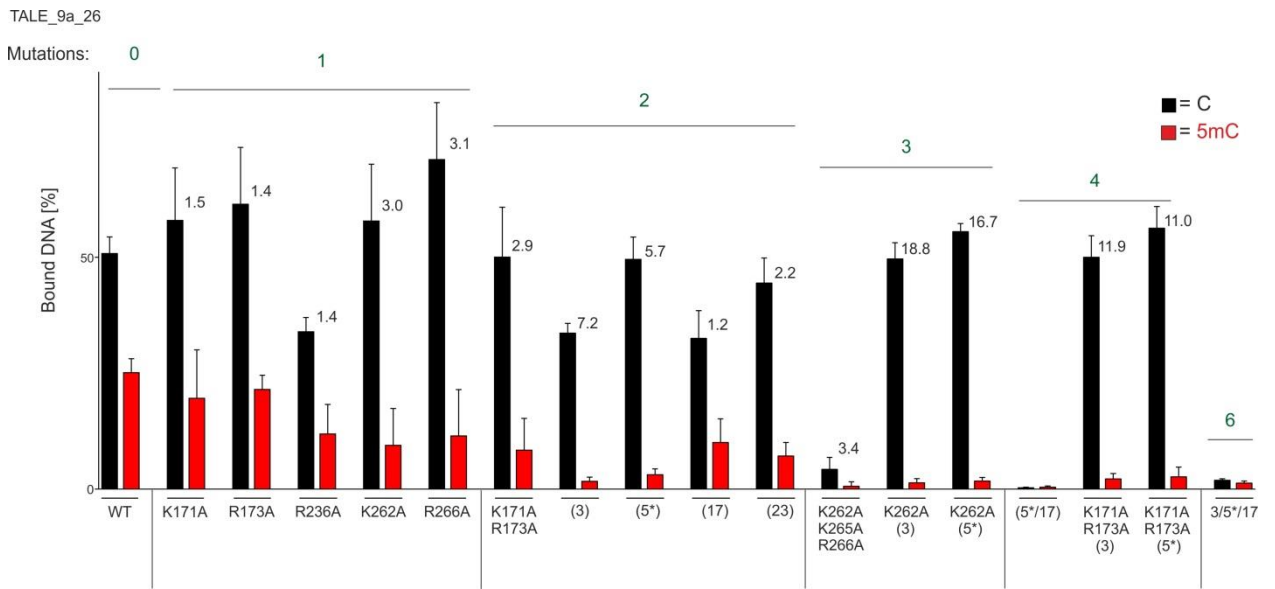
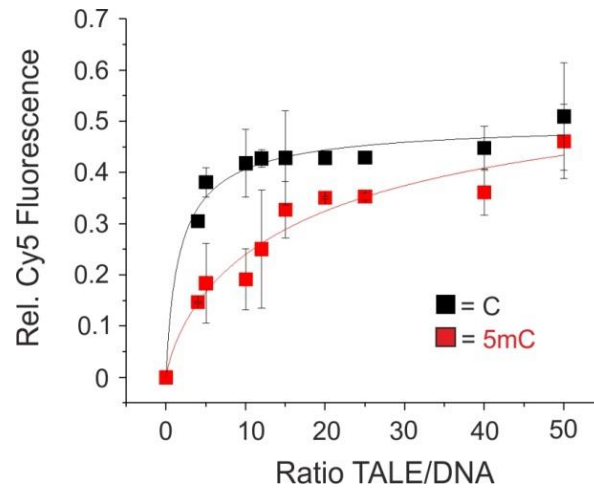


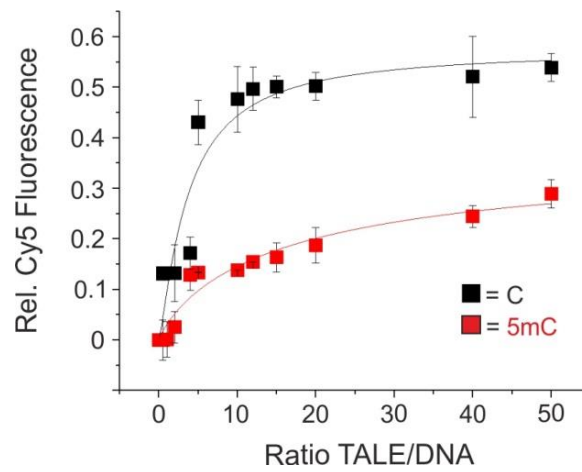
Figure VI-16: Summary of results obtained from quantifying binding by various CDKN2A TALEs in an EMSA assay categorized based on the number of mutation present on each.

P. Determination of K_i for CDKN2A and ZAP-70 WT and mutated TALE using FRET

a



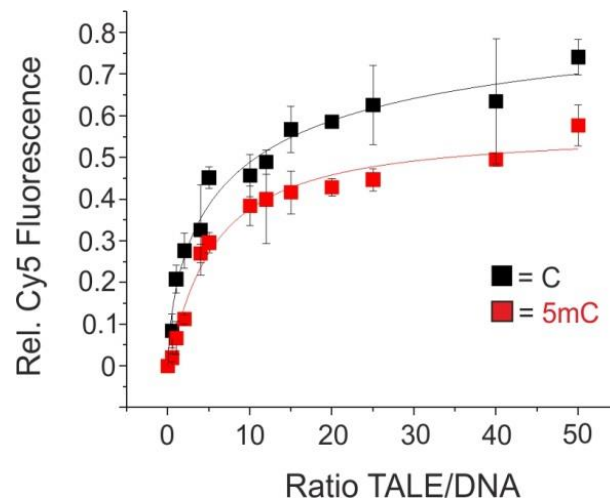
b



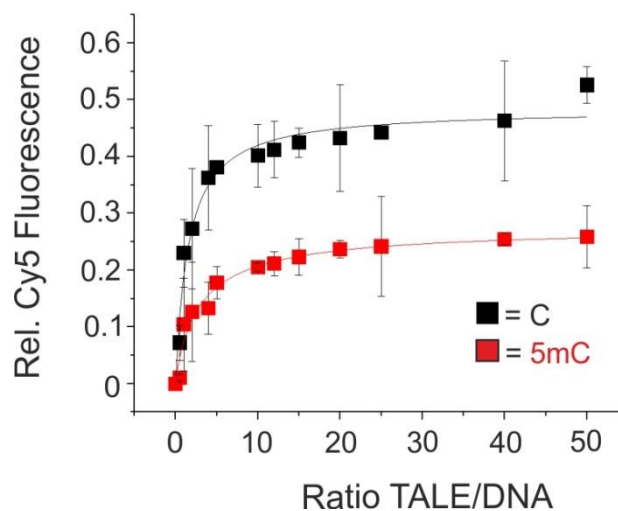
CDKN2A	HD		K262A/(5*)	
	K_i [nm]	F_{max} [AU]	K_i [nm]	F_{max} [AU]
C	91 ± 32	0.50 ± 0.05	186 ± 59	0.58 ± 0.68
5mC	991 ± 1400	0.65 ± 0.33	876 ± 850	0.38 ± 0.14

Figure VI-17: Fits generated using FRET for WT (a) and mutated (b) TALEs CDKN2A targeting sequence containing C or 5mC. In comparison to the WT TALE, one bearing alanine mutations (K262A and repeat 5 K16A Q17A) clearly shows a bigger difference in binding between C and 5mC containing samples. The values generated using Hill equation is shown in the box. Error bars indicate standard error from experimental duplicates. The highest fluorescence value and inhibition constant for each combination is listed in the table.

a



b



ZAP-70	HD		K262A/(3)	
	K_i [nm]	F_{max} [AU]	K_i [nm]	F_{max} [AU]
C	399 ± 235	0.91 ± 0.15	77 ± 13	0.48 ± 0.02
5mC	266 ± 50	0.56 ± 0.04	151 ± 41	0.27 ± 0.02

Figure VI-18: Fits generated using FRET for WT (a) and mut (b) TALEs ZAP-70 targeting sequence containing C or 5mC. In comparison to the WT TALE, one bearing alanine mutations (K262A and repeat 3 K16A Q17A) clearly shows a bigger difference in binding between C and 5mC containing samples. The values generated using Hill equation is shown in the box. Error bars indicate standard error from experimental duplicates. The highest fluorescence value and inhibition constant for each combination is listed in the table.

Q. Measurement of *in vivo* activation assay at different time points

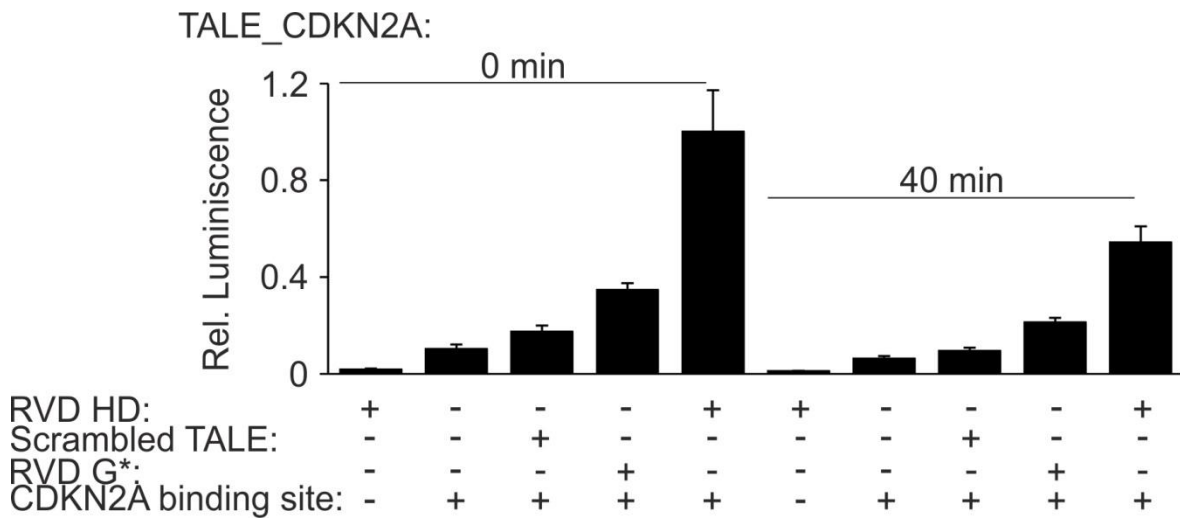
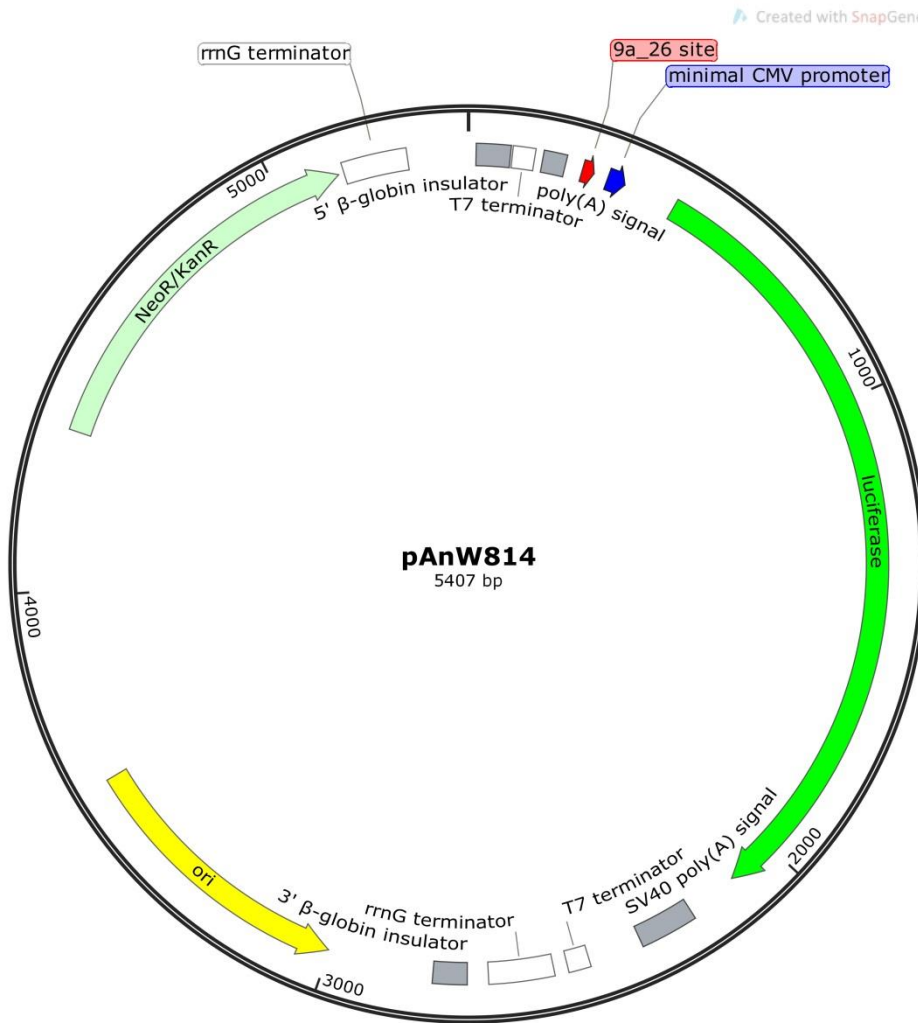
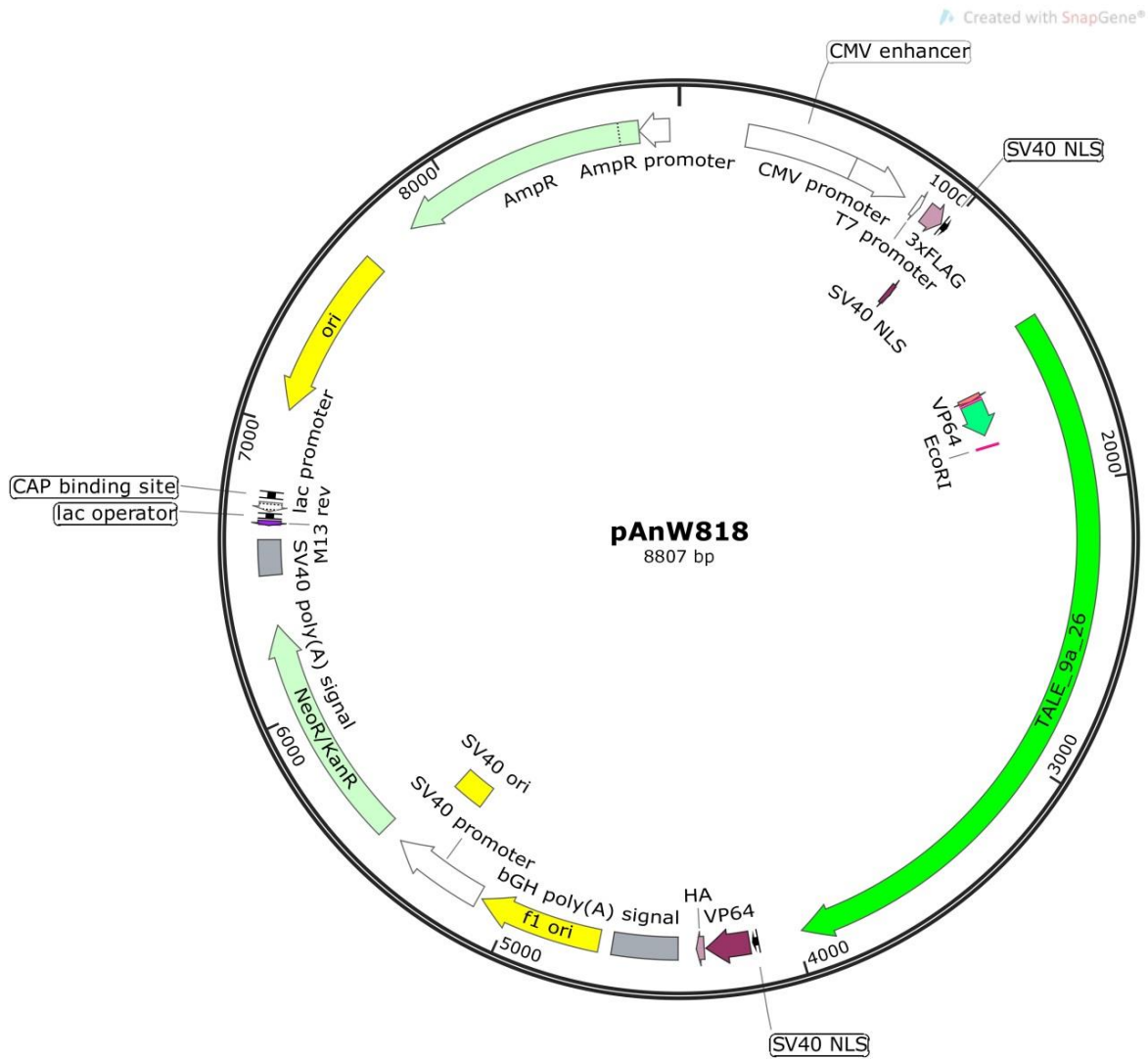


Figure VI-19: Measurement of luminescence by controls used in *in vivo* transcription assay at different time points indicated degradation of the luminescence substrate over time but did not show any change in the observed effect.

R. Maps of plasmids of group A and B used in luciferase assay



Group A plasmid



Group B plasmid

Figure VI-20: Maps showing all the features of plasmids of group A (above) and group B (below) used in *in vivo* transcription activation assay. In the target sequence carrying (plasmid group A), CDKN2A TALE binding site is highlighted in red and the TALE sequence in the expression plasmid (group B) is highlighted in green.

S. Bisulphite sequencing result for CDKN2A target site

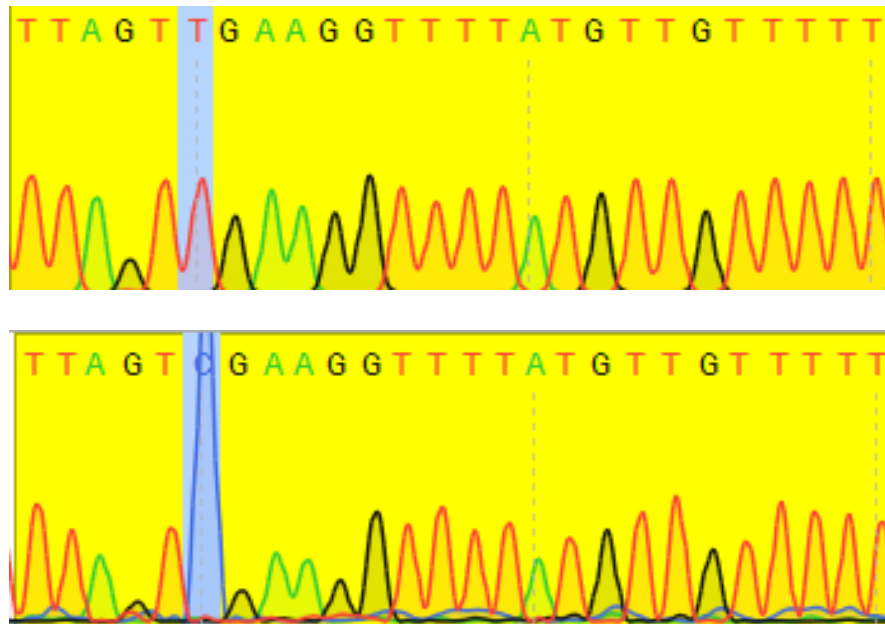


Figure VI-21: Sequencing trace of samples extracted from HEK293T cells following *in vivo* transcription activation assay. The samples were treated with sodium bisulphite and sequencing was performed using primers amplifying the sense strand of CDKN2A locus. The TALE target sequence is highlighted in yellow and the CpG of interest is highlighted in blue. Comparison of C (above) or 5mC (below) bearing samples indicated replacement of C by T at the CpG of interest.

VII. REFERENCES

- [1] H. Y. Zoghbi and A. L. Beaudet, “Epigenetics and Human Disease,” *Cold Spring Harbor perspectives in biology*, vol. 8, no. 2, a019497, 2016.
- [2] L. Wang, J. Zhang, J. Duan et al., “Programming and inheritance of parental DNA methylomes in mammals,” *Cell*, vol. 157, no. 4, pp. 979–991, 2014.
- [3] A. Bird, “DNA methylation patterns and epigenetic memory,” *Genes & development*, vol. 16, no. 1, pp. 6–21, 2002.
- [4] T. K. Kelly, D. D. de Carvalho, and P. A. Jones, “Epigenetic modifications as therapeutic targets,” *Nature biotechnology*, vol. 28, no. 10, pp. 1069–1078, 2010.
- [5] A. P. Feinberg and B. Vogelstein, “Hypomethylation distinguishes genes of some human cancers from their normal counterparts,” *Nature*, vol. 301, no. 5895, pp. 89–92, 1983.
- [6] M. Kulis and M. Esteller, “DNA methylation and cancer,” *Advances in genetics*, vol. 70, pp. 27–56, 2010.
- [7] J. Mill, T. Tang, Z. Kaminsky et al., “Epigenomic profiling reveals DNA-methylation changes associated with major psychosis,” *American journal of human genetics*, vol. 82, no. 3, pp. 696–711, 2008.
- [8] A. J. Verkerk, M. Pieretti, J. S. Sutcliffe et al., “Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome,” *Cell*, vol. 65, no. 5, pp. 905–914, 1991.
- [9] K. Buiting, “Prader-Willi syndrome and Angelman syndrome,” *American journal of medical genetics. Part C, Seminars in medical genetics*, 154C, no. 3, pp. 365–376, 2010.
- [10] A. Portela and M. Esteller, “Epigenetic modifications and human disease,” *Nature biotechnology*, vol. 28, no. 10, pp. 1057–1068, 2010.
- [11] M. Jakovcevski and S. Akbarian, “Epigenetic mechanisms in neurological disease,” *Nature medicine*, vol. 18, no. 8, pp. 1194–1204, 2012.
- [12] G. Kubik, M. J. Schmidt, J. E. Penner et al., “Programmable and highly resolved in vitro detection of 5-methylcytosine by TALEs,” *Angewandte Chemie (International ed. in English)*, vol. 53, no. 23, pp. 6002–6006, 2014.
- [13] G. Kubik and D. Summerer, “Deciphering Epigenetic Cytosine Modifications by Direct Molecular Recognition,” *ACS chemical biology*, vol. 10, no. 7, pp. 1580–1589, 2015.

- [14] G. Kubik and D. Summerer, “Achieving single-nucleotide resolution of 5-methylcytosine detection with TALEs,” *Chembiochem : a European journal of chemical biology*, vol. 16, no. 2, pp. 228–231, 2015.
- [15] G. Kubik, S. Batke, and D. Summerer, “Programmable sensors of 5-hydroxymethylcytosine,” *Journal of the American Chemical Society*, vol. 137, no. 1, pp. 2–5, 2015.
- [16] S. Maurer, M. Giess, O. Koch et al., “Interrogating Key Positions of Size-Reduced TALE Repeats Reveals a Programmable Sensor of 5-Carboxylcytosine,” *ACS chemical biology*, vol. 11, no. 12, pp. 3294–3299, 2016.
- [17] P. Rathi, S. Maurer, G. Kubik et al., “Isolation of Human Genomic DNA Sequences with Expanded Nucleobase Selectivity,” *Journal of the American Chemical Society*, vol. 138, no. 31, pp. 9910–9918, 2016.
- [18] S. Flade, J. Jasper, M. Gieß et al., “The N6-Position of Adenine Is a Blind Spot for TAL-Effectors That Enables Effective Binding of Methylated and Fluorophore-Labeled DNA,” *ACS chemical biology*, vol. 12, no. 7, pp. 1719–1725, 2017.
- [19] Preeti Rathi, Sara Maurer, Daniel Summerer, *Selective Recognition of N4-Methylcytosine in DNA by Engineered Transcription Activator-Like Effectors*, 2017.
- [20] H. Gao, X. Wu, J. Chai et al., “Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region,” *Cell research*, vol. 22, no. 12, pp. 1716–1720, 2012.
- [21] P. Rathi, A. Witte, and D. Summerer, “Engineering DNA Backbone Interactions Results in TALE Scaffolds with Enhanced 5-Methylcytosine Selectivity,” *Scientific reports*, vol. 7, no. 1, p. 15067, 2017.
- [22] G. M. Cooper, *The cell: A molecular approach*, ASM Press [u.a.], Washington, DC, 2000.
- [23] G. Karp, *Cell and molecular biology: Concepts and experiments*, John Wiley, Hoboken NJ, 2010.
- [24] A. Maton, *Cells: Building blocks of life*, Prentice-Hall, Englewood Cliffs, N.J., 1997.
- [25] R. Hine, *A dictionary of biology: Robert S. Hine*, Oxford University Press, Oxford, 2008.
- [26] A. J. F. Griffiths, *Introduction to genetic analysis*, W.H. Freeman and Co, New York, 2012.
- [27] B. Alberts, *Molecular biology of the cell*, Garland Science, New York, 2002.
- [28] C. D. Allis, T. Jenuwein, D. Reinberg, et al., eds., *Epigenetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 2009.

- [29] J. M. Berg, J. L. Tymoczko, and L. Stryer, *Biochemistry*, W. H. Freeman and CO, New York, 2001.
- [30] P. Yakovchuk, E. Protozanova, and M. D. Frank-Kamenetskii, “Base-stacking and base-pairing contributions into thermal stability of the DNA double helix,” *Nucleic acids research*, vol. 34, no. 2, pp. 564–574, 2006.
- [31] B. E. Tropp, *Molecular biology: Genes to proteins*, Jones & Bartlett Learning, Sudbury, Mass., 2012.
- [32] H. Lee, J. Palm, S. M. Grimes et al., “The Cancer Genome Atlas Clinical Explorer: a web and mobile interface for identifying clinical-genomic driver associations,” *Genome medicine*, vol. 7, p. 112, 2015.
- [33] H. S. Basu, B. G. Feuerstein, D. A. Zaring et al., “Recognition of Z-RNA and Z-DNA determinants by polyamines in solution: Experimental and theoretical studies,” *Journal of biomolecular structure & dynamics*, vol. 6, no. 2, pp. 299–309, 1988.
- [34] A. Ghosh and M. Bansal, “A glossary of DNA structures from A to Z,” *Acta crystallographica. Section D, Biological crystallography*, vol. 59, Pt 4, pp. 620–626, 2003.
- [35] R. Wing, H. Drew, T. Takano et al., “Crystal structure analysis of a complete turn of B-DNA,” *Nature*, vol. 287, no. 5784, pp. 755–758, 1980.
- [36] Macmillan Publishers Ltd, 1/23/2003, pp. 448–453.
- [37] J. M. Berg, J. L. Tymoczko, L. Stryer et al., *Biochemistry*, W.H. Freeman and Company, New York, 2012.
- [38] H. F. Lodish, *Molecular cell biology*, W.H. Freeman, New York, Basingstoke, 1999.
- [39] E. P. Solomon, L. R. Berg, and D. W. Martin, *Biology*, Brooks/Cole Thomson Learning, Belmont, CA, 2005.
- [40] A. Klug, “Francis Crick (8th June 1916 - 28th July 2004),” *Nature cell biology*, vol. 6, no. 9, pp. 799–800, 2004.
- [41] Q. Wang, A. R. Parrish, and L. Wang, “Expanding the genetic code for biological studies,” *Chemistry & biology*, vol. 16, no. 3, pp. 323–336, 2009.
- [42] www.protein synthesis.org.
- [43] V. E. A. Russo, A. D. Riggs, and R. A. Martienssen, *Epigenetic mechanisms of gene regulation*, Cold Spring Harbor Laboratory Press, Plainview, N.Y., 1996.
- [44] T. Witte, C. Plass, and C. Gerhauser, “Pan-cancer patterns of DNA methylation,” *Genome medicine*, vol. 6, no. 8, p. 66, 2014.

- [45] R. W. Yen, P. M. Vertino, B. D. Nelkin et al., "Isolation and characterization of the cDNA encoding human DNA methyltransferase," *Nucleic acids research*, vol. 20, no. 9, pp. 2287–2291, 1992.
- [46] K. D. Robertson, E. Uzvolgyi, G. Liang et al., "The human DNA methyltransferases (DNMTs) 1, 3a and 3b: Coordinate mRNA expression in normal tissues and overexpression in tumors," *Nucleic acids research*, vol. 27, no. 11, pp. 2291–2298, 1999.
- [47] S. Pradhan, A. Bacolla, R. D. Wells et al., "Recombinant Human DNA (Cytosine-5) Methyltransferase," *Journal of Biological Chemistry*, vol. 274, no. 46, pp. 33002–33010, 1999.
- [48] B. H. Ramsahoye, D. Biniszkievicz, F. Lyko et al., "Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 10, pp. 5237–5242, 2000.
- [49] H. Leonhardt, A. W. Page, H. U. Weier et al., "A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei," *Cell*, vol. 71, no. 5, pp. 865–873, 1992.
- [50] A. Hermann, R. Goyal, and A. Jeltsch, "The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites," *Journal of Biological Chemistry*, vol. 279, no. 46, pp. 48350–48359, 2004.
- [51] M. Okano, D. W. Bell, D. A. Haber et al., "DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development," *Cell*, vol. 99, no. 3, pp. 247–257, 1999.
- [52] R. Aasland, T. J. Gibson, and A. F. Stewart, "The PHD finger: implications for chromatin-mediated transcriptional regulation," *Trends in biochemical sciences*, vol. 20, no. 2, pp. 56–59, 1995.
- [53] S. Xie, Z. Wang, M. Okano et al., "Cloning, expression and chromosome locations of the human DNMT3 gene family," *Gene*, vol. 236, no. 1, pp. 87–95, 1999.
- [54] G. L. Xu, T. H. Bestor, D. Bourc'his et al., "Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene," *Nature*, vol. 402, no. 6758, pp. 187–191, 1999.
- [55] R. S. Hansen, C. Wijmenga, P. Luo et al., "The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome," *Proceedings of the National*

- Academy of Sciences of the United States of America*, vol. 96, no. 25, pp. 14412–14417, 1999.
- [56] U. Aapola, K. Kawasaki, H. S. Scott et al., “Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family,” *Genomics*, vol. 65, no. 3, pp. 293–298, 2000.
- [57] K. Hata, M. Okano, H. Lei et al., “Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice,” *Development (Cambridge, England)*, vol. 129, no. 8, pp. 1983–1993, 2002.
- [58] D. Bourc'his, G. L. Xu, C. S. Lin et al., “Dnmt3L and the establishment of maternal genomic imprints,” *Science*, vol. 294, no. 5551, pp. 2536–2539, 2001.
- [59] D. Bourc'his and T. H. Bestor, “Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L,” *Nature*, vol. 431, no. 7004, pp. 96–99, 2004.
- [60] K. E. Webster, M. K. O'Bryan, S. Fletcher et al., “Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 11, pp. 4068–4073, 2005.
- [61] T. C. Shovlin, D. Bourc'his, S. La Salle et al., “Sex-specific promoters regulate Dnmt3L expression in mouse germ cells,” *Human reproduction (Oxford, England)*, vol. 22, no. 2, pp. 457–467, 2007.
- [62] N. M. Zamudio, H. S. Scott, K. Wolski et al., “DNMT3L is a regulator of X chromosome compaction and post-meiotic gene transcription,” *PloS one*, vol. 6, no. 3, e18276, 2011.
- [63] M. Okano, S. Xie, and E. Li, “Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells,” *Nucleic acids research*, vol. 26, no. 11, pp. 2536–2540, 1998.
- [64] D. Subramaniam, R. Thombre, A. Dhar et al., “DNA methyltransferases: A novel target for prevention and therapy,” *Frontiers in oncology*, vol. 4, p. 80, 2014.
- [65] R. Lister, M. Pelizzola, R. H. Dowen et al., “Human DNA methylomes at base resolution show widespread epigenomic differences,” *Nature*, vol. 462, no. 7271, pp. 315–322, 2009.
- [66] X. Yang, H. Han, D. D. de Carvalho et al., “Gene body methylation can alter gene expression and is a therapeutic target in cancer,” *Cancer cell*, vol. 26, no. 4, pp. 577–590, 2014.
- [67] S. Shukla, E. Kavak, M. Gregory et al., “CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing,” *Nature*, vol. 479, no. 7371, pp. 74–79, 2011.

- [68] J. K. Christman, “5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy,” *Oncogene*, vol. 21, no. 35, pp. 5483–5495, 2002.
- [69] R. A. Irizarry, C. Ladd-Acosta, B. Wen et al., “The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores,” *Nature genetics*, vol. 41, no. 2, pp. 178–186, 2009.
- [70] N. Shimoda, T. Izawa, A. Yoshizawa et al., “Decrease in cytosine methylation at CpG island shores and increase in DNA fragmentation during zebrafish aging,” *Age (Dordrecht, Netherlands)*, vol. 36, no. 1, pp. 103–115, 2014.
- [71] J. A. Law and S. E. Jacobsen, “Establishing, maintaining and modifying DNA methylation patterns in plants and animals,” *Nature reviews. Genetics*, vol. 11, no. 3, pp. 204–220, 2010.
- [72] M. Bachman, S. Uribe-Lewis, X. Yang et al., “5-Hydroxymethylcytosine is a predominantly stable DNA modification,” *Nature chemistry*, vol. 6, no. 12, pp. 1049–1055, 2014.
- [73] H. Ma, R. Morey, R. C. O'Neil et al., “Abnormalities in human pluripotent cells due to reprogramming mechanisms,” *Nature*, vol. 511, no. 7508, pp. 177–183, 2014.
- [74] R. Lister, M. Pelizzola, R. H. Dowen et al., “Human DNA methylomes at base resolution show widespread epigenomic differences,” *Nature*, vol. 462, no. 7271, pp. 315–322, 2009.
- [75] L. Laurent, E. Wong, G. Li et al., “Dynamic changes in the human methylome during differentiation,” *Genome research*, vol. 20, no. 3, pp. 320–331, 2010.
- [76] P. A. Jones and P. W. Laird, “Cancer epigenetics comes of age,” *Nature genetics*, vol. 21, no. 2, pp. 163–167, 1999.
- [77] S. Eden, T. Hashimshony, I. Keshet et al., “DNA methylation models histone acetylation,” *Nature*, vol. 394, no. 6696, p. 842, 1998.
- [78] R. I. Wakefield, B. O. Smith, X. Nan et al., “The solution structure of the domain from MeCP2 that binds to methylated DNA,” *Journal of molecular biology*, vol. 291, no. 5, pp. 1055–1065, 1999.
- [79] X. Nan, F. J. Campoy, and A. Bird, “MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin,” *Cell*, vol. 88, no. 4, pp. 471–481, 1997.
- [80] J. D. Lewis, R. R. Meehan, W. J. Henzel et al., “Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA,” *Cell*, vol. 69, no. 6, pp. 905–914, 1992.

- [81] P. L. Jones, G. J. Veenstra, P. A. Wade et al., “Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription,” *Nature genetics*, vol. 19, no. 2, pp. 187–191, 1998.
- [82] X. Nan, H. H. Ng, C. A. Johnson et al., “Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex,” *Nature*, vol. 393, no. 6683, pp. 386–389, 1998.
- [83] K. D. Robertson and A. P. Wolffe, “DNA methylation in health and disease,” *Nature reviews. Genetics*, vol. 1, no. 1, pp. 11–19, 2000.
- [84] H. Kimura, T. Nakamura, T. Ogawa et al., “Transcription of mouse DNA methyltransferase 1 (Dnmt1) is regulated by both E2F-Rb-HDAC-dependent and -independent pathways,” *Nucleic acids research*, vol. 31, no. 12, pp. 3101–3113, 2003.
- [85] E. Berezikov, “Evolution of microRNA diversity and regulation in animals,” *Nature reviews. Genetics*, vol. 12, no. 12, pp. 846–860, 2011.
- [86] W. Mayer, A. Niveleau, J. Walter et al., “Demethylation of the zygotic paternal genome,” *Nature*, vol. 403, no. 6769, pp. 501–502, 2000.
- [87] J. Oswald, S. Engemann, N. Lane et al., “Active demethylation of the paternal genome in the mouse zygote,” *Current biology : CB*, vol. 10, no. 8, pp. 475–478, 2000.
- [88] X. Zhang, J. Yazaki, A. Sundaresan et al., “Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis,” *Cell*, vol. 126, no. 6, pp. 1189–1201, 2006.
- [89] K. Rai, I. J. Huggins, S. R. James et al., “DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45,” *Cell*, vol. 135, no. 7, pp. 1201–1212, 2008.
- [90] N. Bhutani, J. J. Brady, M. Damian et al., “Reprogramming towards pluripotency requires AID-dependent DNA demethylation,” *Nature*, vol. 463, no. 7284, pp. 1042–1047, 2010.
- [91] M. Muramatsu, K. Kinoshita, S. Fagarasan et al., “Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme,” *Cell*, vol. 102, no. 5, pp. 553–563, 2000.
- [92] C. Popp, W. Dean, S. Feng et al., “Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency,” *Nature*, vol. 463, no. 7284, pp. 1101–1105, 2010.

- [93] W. A. Pastor, L. Aravind, and A. Rao, “TETonic shift: biological roles of TET proteins in DNA demethylation and transcription,” *Nature reviews. Molecular cell biology*, vol. 14, no. 6, pp. 341–356, 2013.
- [94] L. Hu, Z. Li, J. Cheng et al., “Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation,” *Cell*, vol. 155, no. 7, pp. 1545–1555, 2013.
- [95] M. Tahiliani, K. P. Koh, Y. Shen et al., “Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1,” *Science (New York, N.Y.)*, vol. 324, no. 5929, pp. 930–935, 2009.
- [96] S. Ito, A. C. D'Alessio, O. V. Taranova et al., “Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification,” *Nature*, vol. 466, no. 7310, pp. 1129–1133, 2010.
- [97] H. Zhang, X. Zhang, E. Clark et al., “TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine,” *Cell research*, vol. 20, no. 12, pp. 1390–1393, 2010.
- [98] M. Y. Liu, H. Torabifard, D. J. Crawford et al., “Mutations along a TET2 active site scaffold stall oxidation at 5-hydroxymethylcytosine,” *Nature chemical biology*, vol. 13, no. 2, pp. 181–187, 2017.
- [99] J. U. Guo, Y. Su, C. Zhong et al., “Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain,” *Cell*, vol. 145, no. 3, pp. 423–434, 2011.
- [100] S. Cortellino, J. Xu, M. Sannai et al., “Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair,” *Cell*, vol. 146, no. 1, pp. 67–79, 2011.
- [101] J. U. Guo, D. K. Ma, H. Mo et al., “Neuronal activity modifies the DNA methylation landscape in the adult brain,” *Nature neuroscience*, vol. 14, no. 10, pp. 1345–1351, 2011.
- [102] H. M. Hassan, B. Kolendowski, M. Iovic et al., “Regulation of Active DNA Demethylation through RAR-Mediated Recruitment of a TET/TDG Complex,” *Cell reports*, vol. 19, no. 8, pp. 1685–1697, 2017.
- [103] M. Iurlaro, G. Ficz, D. Oxley et al., “A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation,” *Genome biology*, vol. 14, no. 10, R119, 2013.

- [104] A. Janulaitis, S. Klimašauskas, M. Petrušyte et al., “Cytosine modification in DNA by Bcn I methylase yields N4 -methylcytosine,” *FEBS Letters*, vol. 161, no. 1, pp. 131–134, 1983.
- [105] K. Vasu and V. Nagaraja, “Diverse functions of restriction-modification systems in addition to cellular defense,” *Microbiology and molecular biology reviews : MMBR*, vol. 77, no. 1, pp. 53–72, 2013.
- [106] M. Ehrlich, G. G. Wilson, K. C. Kuo et al., “N4-methylcytosine as a minor base in bacterial DNA,” *Journal of Bacteriology*, vol. 169, no. 3, pp. 939–943, 1987.
- [107] J. F. Costello, M. C. Frühwald, D. J. Smiraglia et al., “Aberrant CpG-island methylation has non-random and tumour-type-specific patterns,” *Nature genetics*, vol. 24, no. 2, pp. 132–138, 2000.
- [108] J. R. Graff, J. G. Herman, R. G. Lapidus et al., “E-cadherin expression is silenced by DNA hypermethylation in human breast and prostate carcinomas,” *Cancer research*, vol. 55, no. 22, pp. 5195–5199, 1995.
- [109] S. K. Myöhänen, S. B. Baylin, and J. G. Herman, “Hypermethylation can selectively silence individual p16ink4A alleles in neoplasia,” *Cancer research*, vol. 58, no. 4, pp. 591–593, 1998.
- [110] M. Esteller, “CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future,” *Oncogene*, vol. 21, no. 35, pp. 5427–5440, 2002.
- [111] J. Veeck, S. Roper, F. Setien et al., “BRCA1 CpG island hypermethylation predicts sensitivity to poly(adenosine diphosphate)-ribose polymerase inhibitors,” *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, vol. 28, no. 29, e563-4; author reply e565-6, 2010.
- [112] J. C. Rice, “Methylation of the BRCA1 promoter is associated with decreased BRCA1 mRNA levels in clinical breast cancer specimens,” *Carcinogenesis*, vol. 21, no. 9, pp. 1761–1765, 2000.
- [113] R. Zhao, B. Y. Choi, M.-H. Lee et al., “Implications of Genetic and Epigenetic Alterations of CDKN2A (p16(INK4a)) in Cancer,” *EBioMedicine*, vol. 8, pp. 30–39, 2016.
- [114] H. Soejima, W. Zhao, and T. Mukai, “Epigenetic silencing of the MGMT gene in cancer,” *Biochemistry and cell biology = Biochimie et biologie cellulaire*, vol. 83, no. 4, pp. 429–437, 2005.
- [115] R. Claus, D. M. Lucas, S. Stilgenbauer et al., “Quantitative DNA methylation analysis identifies a single CpG dinucleotide important for ZAP-70 expression and

- predictive of prognosis in chronic lymphocytic leukemia,” *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, vol. 30, no. 20, pp. 2483–2491, 2012.
- [116] R. Claus, D. M. Lucas, A. S. Ruppert et al., “Validation of ZAP-70 methylation and its relative significance in predicting outcome in chronic lymphocytic leukemia,” *Blood*, vol. 124, no. 1, pp. 42–48, 2014.
- [117] M. J. Booth, E.-A. Raiber, and S. Balasubramanian, “Chemical methods for decoding cytosine modifications in DNA,” *Chemical reviews*, vol. 115, no. 6, pp. 2240–2254, 2015.
- [118] H. Hayatsu, Y. Wataya, K. Kai et al., “Reaction of sodium bisulfite with uracil, cytosine, and their derivatives,” *Biochemistry*, vol. 9, no. 14, pp. 2858–2865, 2002.
- [119] M. Yu, G. C. Hon, K. E. Szulwach et al., “Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome,” *Cell*, vol. 149, no. 6, pp. 1368–1380, 2012.
- [120] Lunt MR, Siebke JC, Burton K, “Glucosylated nucleotide sequences from T2-bacteriophage deoxyribonucleic acid.,” 92(1), pp. 27–36, 1964.
- [121] M. Yu, G. C. Hon, K. E. Szulwach et al., “Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine,” *Nature protocols*, vol. 7, no. 12, pp. 2159–2170, 2012.
- [122] M. Yu, L. Ji, D. A. Neumann et al., “Base-resolution detection of N4-methylcytosine in genomic DNA using 4mC-Tet-assisted-bisulfite- sequencing,” *Nucleic acids research*, vol. 43, no. 21, e148, 2015.
- [123] M. J. Booth, M. R. Branco, G. Ficz et al., “Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution,” *Science (New York, N.Y.)*, vol. 336, no. 6083, pp. 934–937, 2012.
- [124] M. J. Booth, T. W. B. Ost, D. Beraldi et al., “Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine,” *Nature protocols*, vol. 8, no. 10, pp. 1841–1851, 2013.
- [125] M. J. Booth, G. Marsico, M. Bachman et al., “Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution,” *Nature chemistry*, vol. 6, no. 5, pp. 435–440, 2014.
- [126] C.-X. Song, K. E. Szulwach, Q. Dai et al., “Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming,” *Cell*, vol. 153, no. 3, pp. 678–691, 2013.

- [127] X. Lu, C.-X. Song, K. Szulwach et al., “Chemical modification-assisted bisulfite sequencing (CAB-Seq) for 5-carboxylcytosine detection in DNA,” *Journal of the American Chemical Society*, vol. 135, no. 25, pp. 9315–9317, 2013.
- [128] M. Weber, J. J. Davies, D. Wittig et al., “Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells,” *Nature genetics*, vol. 37, no. 8, pp. 853–862, 2005.
- [129] W. A. Pastor, U. J. Pape, Y. Huang et al., “Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells,” *Nature*, vol. 473, no. 7347, pp. 394–397, 2011.
- [130] A. B. Robertson, J. A. Dahl, C. B. Vågbo et al., “A novel method for the efficient and selective identification of 5-hydroxymethylcytosine in genomic DNA,” *Nucleic acids research*, vol. 39, no. 8, e55, 2011.
- [131] N. Li, M. Ye, Y. Li et al., “Whole genome DNA methylation analysis based on high throughput sequencing technology,” *Methods (San Diego, Calif.)*, vol. 52, no. 3, pp. 203–212, 2010.
- [132] J. H. Laity, B. M. Lee, and P. E. Wright, “Zinc finger proteins: new insights into structural and functional diversity,” *Current opinion in structural biology*, vol. 11, no. 1, pp. 39–46, 2001.
- [133] “Zinc finger proteins: new insights into structural and functional diversity - ScienceDirect,”
<http://www.sciencedirect.com/science/article/pii/S0959440X00001676?via%3Dihub>.
- [134] “Review article,” *Linguistics*, vol. 29, no. 1, 1991.
- [135] “Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus oocytes*,” European Molecular Biology Organization, 1/1/1985, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC554390/>.
- [136] M. Lee, G. Gippert, K. Soman et al., “Three-dimensional solution structure of a single zinc finger DNA-binding domain,” *Science*, vol. 245, no. 4918, pp. 635–637, 1989.
- [137] D. Hiraoka, W. Yoshida, K. Abe et al., “Development of a method to measure DNA methylation levels by using methyl CpG-binding protein and luciferase-fused zinc finger protein,” *Analytical chemistry*, vol. 84, no. 19, pp. 8259–8264, 2012.
- [138] M. Elrod-Erickson, M. A. Rould, L. Nekludova et al., “Zif268 protein–DNA complex refined at 1.6Å: A model system for understanding zinc finger–DNA interactions,” *Structure*, vol. 4, no. 10, pp. 1171–1180, 1996.

- [139] G. Gasiunas, R. Barrangou, P. Horvath et al., “Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 39, E2579-86, 2012.
- [140] P. Mali, L. Yang, K. M. Esvelt et al., “RNA-guided human genome engineering via Cas9,” *Science (New York, N.Y.)*, vol. 339, no. 6121, pp. 823–826, 2013.
- [141] Le Cong, F. A. Ran, D. Cox et al., “Multiplex genome engineering using CRISPR/Cas systems,” *Science (New York, N.Y.)*, vol. 339, no. 6121, pp. 819–823, 2013.
- [142] S. W. Cho, S. Kim, J. M. Kim et al., “Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease,” *Nature biotechnology*, vol. 31, no. 3, pp. 230–232, 2013.
- [143] M. Jinek, K. Chylinski, I. Fonfara et al., “A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity,” *Science (New York, N.Y.)*, vol. 337, no. 6096, pp. 816–821, 2012.
- [144] L. S. Qi, M. H. Larson, L. A. Gilbert et al., “Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression,” *Cell*, vol. 152, no. 5, pp. 1173–1183, 2013.
- [145] I. B. Hilton, A. M. D'Ippolito, C. M. Vockley et al., “Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers,” *Nature biotechnology*, vol. 33, no. 5, pp. 510–517, 2015.
- [146] A. Lo and L. Qi, “Genetic and epigenetic control of gene expression by CRISPR-Cas systems,” *F1000Research*, vol. 6, 2017.
- [147] U. Bonas, R. E. Stall, and B. Staskawicz, “Genetic and structural characterization of the avirulence gene *avrBs3* from *Xanthomonas campestris* pv. *vesicatoria*,” *Molecular & general genetics : MGG*, vol. 218, no. 1, pp. 127–136, 1989.
- [148] S. Cunnac, A. Occhialini, P. Barberis et al., “Inventory and functional analysis of the large Hrp regulon in *Ralstonia solanacearum*: identification of novel effector proteins translocated to plant host cells through the type III secretion system,” *Molecular microbiology*, vol. 53, no. 1, pp. 115–128, 2004.
- [149] K. Herbers, J. Conrads-Strauch, and U. Bonas, “Race-specificity of plant resistance to bacterial spot disease determined by repetitive motifs in a bacterial avirulence protein,” *Nature*, vol. 356, no. 6365, pp. 172–174, 1992.

- [150] H.-J. Cho, Y.-J. Park, T.-H. Noh et al., “Molecular analysis of the hrp gene cluster in *Xanthomonas oryzae* pathovar *oryzae* KACC10859,” *Microbial pathogenesis*, vol. 44, no. 6, pp. 473–483, 2008.
- [151] V. Knoop, B. Staskawicz, and U. Bonas, “Expression of the avirulence gene *avrBs3* from *Xanthomonas campestris* pv. *vesicatoria* is not under the control of hrp genes and is independent of plant factors,” *Journal of Bacteriology*, vol. 173, no. 22, pp. 7142–7150, 1991.
- [152] A. N.-S. Mak, P. Bradley, R. A. Cernadas et al., “The crystal structure of TAL effector PthXo1 bound to its DNA target,” *Science (New York, N.Y.)*, vol. 335, no. 6069, pp. 716–719, 2012.
- [153] G. van den Ackerveken, E. Marois, and U. Bonas, “Recognition of the Bacterial Avirulence Protein AvrBs3 Occurs inside the Host Plant Cell,” *Cell*, vol. 87, no. 7, pp. 1307–1316, 1996.
- [154] J. F. Meckler, M. S. Bhakta, M.-S. Kim et al., “Quantitative analysis of TALE-DNA interactions suggests polarity effects,” *Nucleic acids research*, vol. 41, no. 7, pp. 4118–4128, 2013.
- [155] B. Yang, W. Zhu, L. B. Johnson et al., “The virulence factor AvrXa7 of *Xanthomonas oryzae* pv. *oryzae* is a type III secretion pathway-dependent nuclear-localized double-stranded DNA-binding protein,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 17, pp. 9807–9812, 2000.
- [156] D. Deng, C. Y. Yan, X. J. Pan et al., *Crystal structure of the DNA-bound dHax3, a TAL effector, at 1.85 angstrom*, 2012.
- [157] B. M. Lamb, A. C. Mercer, and C. F. Barbas, “Directed evolution of the TALE N-terminal domain for recognition of all 5' bases,” *Nucleic acids research*, vol. 41, no. 21, pp. 9779–9785, 2013.
- [158] F. Wang, S. Redding, I. J. Finkelstein et al., “The promoter-search mechanism of *Escherichia coli* RNA polymerase is dominated by three-dimensional diffusion,” *Nature structural & molecular biology*, vol. 20, no. 2, pp. 174–181, 2013.
- [159] P. C. Blainey, G. Luo, S. C. Kou et al., “Nonspecifically bound proteins spin while diffusing along DNA,” *Nature structural & molecular biology*, vol. 16, no. 12, pp. 1224–1229, 2009.
- [160] L. Cuculis, Z. Abil, H. Zhao et al., “Direct observation of TALE protein dynamics reveals a two-state search mechanism,” *Nature communications*, vol. 6, p. 7277, 2015.

- [161] L. Cuculis, Z. Abil, H. Zhao et al., “TALE proteins search DNA using a rotationally decoupled mechanism,” *Nature chemical biology*, vol. 12, no. 10, pp. 831–837, 2016.
- [162] A. N.-S. Mak, P. Bradley, A. J. Bogdanove et al., “TAL effectors: Function, structure, engineering and applications,” *Current opinion in structural biology*, vol. 23, no. 1, pp. 93–99, 2013.
- [163] B. P. Hubbard, A. H. Badran, J. A. Zuris et al., “Continuous directed evolution of DNA-binding proteins to improve TALEN specificity,” *Nature methods*, vol. 12, no. 10, pp. 939–942, 2015.
- [164] A. Juillerat, G. Dubois, J. Valton et al., “Comprehensive analysis of the specificity of transcription activator-like effector nucleases,” *Nucleic acids research*, vol. 42, no. 8, pp. 5390–5402, 2014.
- [165] F. C. Rinaldi, L. A. Doyle, B. L. Stoddard et al., “The effect of increasing numbers of repeats on TAL effector DNA binding specificity,” *Nucleic acids research*, 2017.
- [166] T. Sakuma, H. Ochiai, T. Kaneko et al., “Repeating pattern of non-RVD variations in DNA-binding modules enhances TALEN activity,” *Scientific reports*, vol. 3, p. 3379, 2013.
- [167] P. Yi, W. Li, and G. Ou, “The application of transcription activator-like effector nucleases for genome editing in *C. elegans*,” *Methods (San Diego, Calif.)*, vol. 68, no. 3, pp. 389–396, 2014.
- [168] T. Cermak, E. L. Doyle, M. Christian et al., “Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting,” *Nucleic acids research*, vol. 39, no. 12, e82, 2011.
- [169] J. Boch and U. Bonas, “*Xanthomonas* AvrBs3 family-type III effectors: discovery and function,” *Annual review of phytopathology*, vol. 48, pp. 419–436, 2010.
- [170] A. J. Bogdanove and D. F. Voytas, “TAL effectors: customizable proteins for DNA targeting,” *Science (New York, N.Y.)*, vol. 333, no. 6051, pp. 1843–1846, 2011.
- [171] B. Chen, J. Guan, and B. Huang, “Imaging Specific Genomic DNA in Living Cells,” *Annual review of biophysics*, vol. 45, pp. 1–23, 2016.

Preeti Rathi, M.Sc.

Schillerstr. 6B
13156 Berlin
Deutschland
Email: preetideepshah@gmail.com
Mobil: 0176-72 41 48 67



Persönliche Daten

Geboren am: 21. Sept 1988
Geburtsort: Kalkutta, Indien
Nationalität: Indisch
Familienstand: Verheiratet
Sprachkenntnisse: Deutsch-B2 (Mittelstufe), Englisch-fließend

Ausbildung

Seit 02/2014 **Doktorandin (Dr. rer. nat.)**
an der Universität Konstanz und TU Dortmund in den Bereichen Chemie und Chemische Biologie
unter Anleitung von Prof. Dr. Daniel Summerer
„Detection of Epigenetic Modifications in genomic DNA using TALE proteins“

10/2011 – 01/2014 **Masters of Molecular Medicine (1.8 „gut“)**
am Friedrich-Schiller-Universität, FSU Jena

07/2007 - 08/2011 **Bachelor Biotechnik (1.3 „sehr gut“)**
am M.S Ramaiah Technische Universität, Bangalore, Indien

06/2005 - 04/2007 **Abitur (1.5)**
Physik, Chemie, Biologie und Mathematik, C.B.S.C Board,
Kalkutta, Indien

Praktische Erfahrung

11/2012 - 02/2014 **Praktikum und Masterarbeit (1.3 „sehr gut“)**
an der MPI für Chemie ecologie, Jena
Master Thesis Thema: „ Nicotine detoxification by Manduca Sexta“
- Gene knockdown in *M. sexta* using RNAi
- Analysierung mRNA Niveau durch qRT-PCR
- Nicotin Metabolismus Auswertung durch LC-MS

08/2012 – 11/2012 **Master Forschung Projekt Module 1 (1,3“ sehr gut“)**
Abteilung: Human Genetik, FSU Jena
Thema: „ Role of HTERT receptor in Prostate Cancer“
- Untersuchung der Wildtyp-Androgenrezeptor-basierten Inhibition von hTERT in HeLA-Zelllinien mittels qRT-PCR

- 05/2011 - 08/2011 **Master Forschung Projekt Module 2 (1,3“ sehr gut“)**
 Abteilung: Human Nephrologie, FSU Jena
 Thema : „ Identification of H1F1a binding sites using qRT-PCR“
 - Identifizierung von Genen, die durch H1F-1a-Bindung mittels qRT-PCR aktiviert werden
- 12/2010 – 05/2011 **Bachelorarbeit (1, 0 “sehr gut“)**
 am M.S.Ramaiah medical college, Bangalore, Indien
 Thema: “Identification & Characterization of Enterococcus Faecalis from Dental cavity“
 - Identifizierung einer unterschiedlichen Kombination von Antibiotika zur Hemmung von E. faecalis-Infektionen beim Menschen
- 07/2010 – 10/2010 **Freiwilliger Praxiskurs (1, 0 “sehr gut“)**
 am Maharani Laxmi Ammani college, Bangalore, India
 Thema: “ “Effect of Punica granatum on Alloxan induced Diabetic Rats”
 - Verabreichung von P. granatum-Extrakt an diabetische Ratten und Beobachtung seiner Wirkung durch Analyse des Extrakts aus Organen

Scientific Expertise und Fähigkeiten

Molekularbiologie:	Genetisches Klonen, Gibson Assembly, Quick changes, Bisulphite treatment und DNA-Sequenzanalyse, Proteinreinigung, -expression und -modifikation,, Transcription Activator Like Effector (TALE) proteinassemblierung, Nachweis von epigenetischen Markierungen auf DNA
BiochemischMethoden / Assays :	Gen-Knockout mit RNAi, Gel-Elektrophorese (EMSA, PAGE), Primer-Extension (PEX) mit radioaktiv markierter DNA, Einbau nichtnatürlicher Aminosäuren, Genomische Anreicherung mit TALE-Proteinen, Nachweis von modifizierter Nukleobase mittels in vivo-Luciferase-Assay
Zellkultur / Modellorganismus :	Microbielles modelle (<i>E. coli</i> , <i>E. Faecalis</i> , <i>A. Tumifaciens</i>), Zelllinie (HeLa, HEK293T) Pflanzenmodelle (<i>N. attenuata</i>); Tiermodelle (<i>M. Sexta</i> , <i>M. musculus</i> , <i>D. rerio</i>)
Nukleinsäure / Proteinstrukturanalyse :	RT-qPCR, HPLC, GCMS, MALDI, Fluoreszenz- und Radioaktivitäts-Quantifizierungstests an Tecan M1000 und Typhoon
Software Kenntnisse :	Vector NTI, Serial Cloner, Snap gene, Origin, Corel Draw, Quantity one, Image J, Pymol, Logo, Basic

Wissenschaftliche Veröffentlichung

- **P. Rathi**, S. Maurer, D. Summerer
 “Selective recognition of N4-Methylcytosine in DNA by Engineered Transcription-Activator-Like Effectors”
Philosophical Transactions of the Royal Society B, **2017**, RSTB-2017-0078
- **P. Rathi**, A. Witte, D. Summerer
 “Engineered DNA Backbone Interactions Results in TALE Scaffolds with Enhanced Epigenetic Nucleobase Selectivity” *Scientific Reports*, **2017**, 7: 15067, 10.1038/s41598-017-15361-1

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

• P. Kumar, **P. Rathi**, M. Schöttner, I.T. Baldwin, S. Pandit
“Differences in Nicotine Metabolism of Two *Nicotiana attenuata* Herbivores Render Them Differentially Susceptible to a Common Native Predator”
Plos One, **2014**, (4): e95982

• **P. Rathi**, P. Kumar, S. Pandit, Roger Estrada-Tejedor, Matthias Schöttner, Ian T Baldwin
“Sweet detoxification: Novel neuronal effects of insects’ novel nicotine detoxification product, nicotine N-D-glucoside”
Manuskript in Vorbereitung

Wissenschaftliche Poster Präsentation

• **P. Rathi**, D. Summerer
“Isolation of human genomic DNA sequences with expanded nucleobase selectivity: Detection of 5-methylcytosine with improved selectivity” Poster presentation at: Frontiers in epigenetic chemical biology, The Royal Society meet; **2015** May 22-23; London, United Kingdom

• **P. Rathi**, D. Summerer
“Programmable Affinity Enrichment Probes for the Direct Quantification of Genomic 5-Methylcytosine”
Poster presentation at: Genome Engineering and Synthetic Biology: Tools and Technologies.
VIB conference series; **2016** Jan 28-29; Ghent, Belgium

• **P. Rathi**, D. Summerer
“Programmable Affinity Enrichment Probes for the Direct Quantification of Genomic 5-Methylcytosine”
Poster presentation at: Shaping the Molecules of Life: Chemical Biology of Nucleic Acid and Protein Modifications. Gesellschaft Deutscher Chemiker GDCh Annual meet; **2016** Jul 6-8; Frankfurt, Germany

• **P. Rathi**, D. Summerer
“TALEs as a potential tool for detection of promoter hyper-methylation in cancer”
Poster presentation at: Challenges in chemical biology. Royal Society of Chemistry meet; **2015** Jun 15-18; Zurich, Switzerland

Kontakt Referenz

• **Prof. Dr. Daniel Summerer**
Abteilung Chemie und Chemische Biologie, Technische Universität, Dortmund, Deutschland
(PhD Supervisor) E-mail: daniel.summerer@tu-dortmund.de



Berlin, 16/11/2017