



Thèse

2020

Open Access

This version of the publication is provided by the author(s) and made available in accordance with the copyright holder(s).

Synthetic Activation of Aberrantly Silenced Genes in Diseases

Hafner, Nicolas

How to cite

HAFNER, Nicolas. Synthetic Activation of Aberrantly Silenced Genes in Diseases. 2020. doi:
10.13097/archive-ouverte/unige:141767

This publication URL: <https://archive-ouverte.unige.ch//unige:141767>

Publication DOI: [10.13097/archive-ouverte/unige:141767](https://doi.org/10.13097/archive-ouverte/unige:141767)

UNIVERSITÉ DE GENÈVE

FACULTÉ DE MÉDECINE

Section de médecine fondamentale

Département de médecine génétique et développement

Professeur Rabih Murr

SYNTHETIC ACTIVATION OF ABERRANTLY SILENCED GENES IN DISEASES

THÈSE

présentée aux Facultés de médecine et des sciences de l'Université de Genève
pour obtenir le grade de Docteur ès sciences en sciences de la vie,
mention Sciences biomédicales

Par

Nicolas Hafner

de

Mannheim (Allemagne)

Thèse N° 49

GENÈVE

2019



DOCTORAT ÈS SCIENCES EN SCIENCES DE LA VIE DES
FACULTÉS DE MÉDECINE ET DES SCIENCES
MENTION SCIENCES BIOMÉDICALES

Thèse de Mr Nicolas HAFNER

intitulée :

« Synthetic Activation of Aberrantly Silenced Genes in Diseases »

Les Facultés de médecine et des sciences, sur le préavis de Monsieur Rabih MURR, professeur assistant et directeur de thèse (Département de Biologie Moléculaire), Monsieur David SHORE, professeur (Unité des Produits Naturels Bioactifs), Monsieur Mirko TRAJKOVSKI, professeur ordinaire (Département de Physiologie Cellulaire et Métabolisme), Monsieur Zdenko HERCEG (Centre International de Recherche sur le Cancer, Lyon, France) autorisent l'impression de la présente thèse, sans exprimer d'opinion sur les propositions qui y sont énoncées.

Genève, le 20 février 2020

Thèse - 49 -

Le Doyen
Faculté de médecine

Le Décanat
Faculté des sciences

N.B. - La thèse doit porter la déclaration précédente et remplir les conditions énumérées dans les "Informations relatives aux thèses de doctorat à l'Université de Genève".

Table of Contents

Abbreviations	V
Acknowledgements	VIII
Abstract	X
Résumé	XII
1 Introduction	1
1.1 Epigenetics and the Regulation of Gene Expression.....	1
1.1.1 From Information to Function: Gene Expression and its Regulation	1
1.1.2 The Biology of Transcription Factors	7
1.1.3 Chromatin and Epigenetics	12
1.1.4 The 3D Architecture of the Genome	15
1.1.5 Accessing Nucleosomal DNA.....	19
1.1.6 Histone Modifications	27
1.1.7 DNA Methylation.....	32
1.2 CRISPR	66
1.2.1 Exploits from an Ancient War	66
1.2.2 Harnessing CRISPR for Genome Editing	72
1.2.3 No Cutting Needed: dCas9-Effectors.....	75
1.2.4 Opportunities and Challenges	79
2 Aim of the Thesis	83
3 Results	85

3.1	Reactivation of Epigenetically Silenced Tumor Suppressor Genes with Super Pioneer Transcription Factors	85
3.1.1	Validation of the Experimental Approach by Targeting TSGs with dCas9-TET1	87
3.1.2	Phenotypic Changes Subsequent to Targeted p16 Reactivation.....	92
3.1.3	Insertion of the CTCF motif into TSG Promoter Regions.....	96
3.1.4	Insertion of SPF motifs into a Methylation Reporter	102
3.2	Cell Fate of Adipose Tissues: New Targets for Epigenome Editing	110
3.2.1	Identification of Differentially Expressed Genes and Differentially Methylated Regions Upon Browning in VAT Adipocytes.....	112
3.2.3	Differentiation of 3T3L1-preadipocytes under the Influence of 5-aza.....	120
4	Discussion	123
4.1	Experimental Approach.....	123
4.2	Interpretation of the Results.....	130
4.3	Outlook and Perspectives.....	133
5	Materials and Methods	137
5.1	Cell Culture.....	137
5.2	Treatment of Dazl-Snrpn-GFP V6.5 mESCs with 5-azacytidine.....	137
5.3	Cloning of guide RNAs	137
5.4	Lentivirus Production and Transduction	139
5.5	SPF Motif Insertion with CRISPR/Cas9	141
5.6	Genotyping	143

5.7	Bisulfite Conversion, PCR, and Sequencing.....	144
5.8	RNA Extraction and qPCR.....	144
5.9	Protein Extraction and Western Blot.....	146
5.10	Cell Counting Assay.....	147
5.11	Live Cell Imaging.....	148
5.12	Cell Cycle Analysis of Live Cells.....	148
5.13	Senescence β -Galactosidase Staining.....	148
5.14	Flow Cytometry of Dazl-Snrpn Reporter Cells.....	149
5.15	Animal Housing and CAx Treatment.....	149
5.16	Whole Genome Bisulfite Sequencing.....	149
5.17	RNA Sequencing.....	150
5.18	Differentiation of 3T3-L1 Pre-Adipocytes.....	151
6	References.....	152

Abbreviations

3C	Chromosome conformation capture
4C	Circular 3C
5-mC	5-methylcytosine
A	Adenine
aa	Amino acid
AAV	Adeno-associated virus
BAT	Brown adipose tissue
bp	Base pair
BMI	Body-mass index
BS	Bisulfite sequencing
C	Cytosine
Cas	CRISPR associated protein
CAX	Cold/antibiotics treatment
CDK	Cyclin-dependent kinase
cDNA	Cyclic DNA
CGI	CpG island
ChIP	Chromatin immunoprecipitation
Cidea	Cell death activator
CpG	Cytosine followed by a guanine
CRE	Cis-regulatory element
CRISPR	Clustered regularly interspaced short palindromic repeats
CRISPRa	CRISPR activation
CRISPRi	CRISPR interference
crRNA	CRISPR RNA
CTCF	CCCTC-binding factor
Dazl	Deleted in azoospermia-like
DBD	DNA-binding domain
dCas9	Catalytically inactive/dead Cas9
DEG	Differentially expressed gene
DHS	DNase I hypersensitive site
DMR	Differentially methylated region
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DPE	Downstream promoter element
DSB	Double strand break
dsDNA	Double-stranded DNA
Elovl6	Elongation of very long chain fatty acids protein 6
ESC	Embryonic stem cell
FMR	Fully methylated region
FACS	Fluorescence-activated cell sorting
FOXA1	Forkhead box protein A1
fwd	Forward

G	Guanine
gDNA	Genomic DNA
GFP	Green fluorescent protein
gRNA	Guide RNA
GTF	General transcription factor
HAT	Histone acetyltransferase
HCC	Hepatocellular carcinoma
HDAC	Histone deacetylase
HDR	Homology-directed DNA repair
HKMT	Histone lysine methyltransferases
HP1	Heterochromatin protein 1
ICR	Imprinting control region
Indel	Insertion/Deletion
iPS	Induced pluripotent stem cell
kb	kilobase
KI	Knock-in
KLF4	Krüppel-like factor 4
KO	Knock-out
KRAB	Krüppel-associated box
LMR	Low methylated region
lncRNA	long non-coding RNA
LV	Lentivirus
MBD	Methyl-CpG-binding domain
mESC	Mouse embryonic stem cell
MFI	Mean/median fluorescence intensity
ML	Methylation level
mRNA	Messenger RNA
NGS	Next generation sequencing
NHEJ	Non-homologous end joining
NP	Neuronal progenitor
OCT4	Octamer-binding transcription factor 4
PAM	Protospacer-adjacent motif
PCR	Polymerase chain reaction
PF	Pioneer factor
PIC	Preinitiation complex
Pol II	RNA polymerase II
Ppar a/g	Peroxisome proliferator-activated receptor alpha/gamma
Ppargc1a	Ppar g coactivator 1-alpha
pRb	Phosphorylated retinoblastoma protein
PRC	Polycomb repressive complex
qRT-PCR	Quantitative real-time polymerase chain reaction
RAP	Rapamycin
Rassf1a	Ras association domain-containing protein 1A
Rb	Retinoblastoma protein

REST	RE1-silencing transcription factor
rev	Reverse
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
RRBS	Reduced representation bisulfite sequencing
RT	Room temperature
SAM	S-Adenosyl methionine
SAT	Subcutaneous adipose tissue
SC	Scrambled
sgRNA	Single guide RNA
Snrpn	Small nuclear ribonucleoprotein-associated protein N
SOX2	SRY (sex determining region Y)-box 2
SPF	Super pioneer factor
ssODN	Single-stranded oligo DNA nucleotides
T	Thymine
TAD	Topologically-associated domain
TALEN	Transcription activator-like effector nuclease
TAPS	TET-assisted pyridine borane sequencing
TET	Ten-eleven translocation methylcytosine dioxygenase
TF	Transcription factor
tracrRNA	trans-activating crRNA
TRE	Trans-regulatory element
TSG	Tumor suppressor gene
TSS	Transcription start site
UCP1	Uncoupling protein 1
UMR	Unmethylated region
UHRF1	Ubiquitin-like, containing PHD and RING finger domains, 1
VAT	Visceral adipose tissue
WAT	White adipose tissue
WGBS	Whole genome bisulfite sequencing
WT	Wild-type
YY1	Yin Yang 1
ZF	Zinc finger
ZFN	Zink finger nuclease

Acknowledgements

The last five years have been quite an experience; sometimes absolutely amazing, sometimes there were hardships, but definitely educational and even life changing. Working in science brings me great joy and contributing to the advancement of our understanding how the world works has always been a passion of mine. However, science is built on teamwork, and, indeed, many people have contributed to my work, either actively or passively. On this occasion, I would like to express my sincerest gratitude towards them. First, I would like to thank my supervisor, Prof. Rabih Murr, for giving me the opportunity to work in his laboratory. His door has always been open for me and he provided valuable insight, leading to a strong student-mentor relationship. It was a pleasure and a privilege. I would also like to thank Prof. Mirko Trajkovski, with whom we have a still-ongoing collaboration, as well as the other jury members Prof. David Shore from the Science faculty of the University of Geneva, and Prof. Zdenko Herceg from the International Agency for Research on Cancer in Lyon. I would also like to thank the SNF, the Novartis Foundation for Medical-Biological Research, and the Gertrude von Meissner Foundation for funding my research.

I would also like to acknowledge the many people who have contributed to my projects. First, I would like to mention our collaborators from the Trajkovski laboratory, Dr. Nicolas Suarez-Zamorano, who took care of the mice and the CAx treatment as well as RNA extraction and qRT-PCR of 3T3-L1 adipocytes, and Silas Kieser, who performed the expression analysis after CAx treatment. From the Murr laboratory, I would like to thank in particular Victor Ythier, for his contributions in terms of bioinformatics, especially the WGBS analysis, and Hadrien Soldati, for the lentivirus production and the establishment of the dCas9-effector-expressing cell lines, as well as his general support throughout the years. I would also like to thank the other members of the Murr lab, Dr. Ludovica Vanzan, Dr. Ino Karemaker, Dr. Santosh Anand,

and Dr. Muhammad Ahmad Maqbool, for all their support and many contributions, as well as for providing a nice working atmosphere.

I also have to express profound gratitude to Aureliano Stingi, from Prof. Patrick Meraldi's laboratory, and Antonija Sakic, from Prof. Marie-Luce Bochaton-Piallat's laboratory, for their help with life cell imaging and the senescence assay, respectively. Furthermore, I must thank our Flow Cytometry core facility, namely Dr. Jean-Pierre Aubry-Lachainaye, Cécile Gameiro, and Grégory Schneider, for their support, training, and patience over the years.

Finally, I would like to thank my friends and family back home in Mannheim as well as here in Geneva. My parents, Elke and Josef Hafner, certainly deserve the lion's share of my gratitude, since they have been greatly supporting me not only during my PhD, but also during my entire education from the moment I decided to go into Life Sciences, and, in fact, my entire life. During my PhD, I have also made a lot of new friends in Geneva, not only from within the university, but also from all sorts of life, representing nearly every corner of the world. They all have made my stay in Geneva more interesting, exciting, and fun. In particular, I'd like to mention, Meghan Sullivan, who kept me precious company especially during my writing phase, Alexandra and Oliver Greiner-Mahler, who hosted me when I arrived, the Hiking group (Andrea, Friedhelm, Thomas, Ania, Ric, Chris, Justine, Lorraine, Endre, Örs, Mariana, Jagoda, Mathi and Mathi), although we did mostly non-hiking activities, the Canadians (Zack (and Noline), Ben, and Thea), who accepted me into their tribe, my many roommates (Marvin, Jan, Justine, Andrea, Josip, Mario, and Sergio), my brothers (and sisters) in arms at the CMU (Adria, Claire, Ebru, Eva, Cristina, Marta, Sunil, Marion, Miguel, Alexandre, Graziana, Aderonke, Ines, Amy, Vanessa, Caroline, Christopher, Herta, Pauline, Olimpia, Simona, Ola, Piango, Rui and Soner), as well as friends from all over the place (Abbas, An, Sabina, Allegra, Ritwija, Patti, Ryne, Alli, Danielle, Julia, Julian, Anthony, and Kate). You have all contributed in one way or another to an unforgettable experience in Geneva and for that I am deeply thankful.

Abstract

DNA methylation is an important epigenetic modification that could repress gene expression by reducing the accessibility and the affinity of DNA to transcription factors (TFs). Consequently, aberrant distribution of methylation can affect the expression of genes with important functions in healthy cells, thus leading to diseases. One example is hypermethylation and subsequent inhibition of tumor suppressor gene (TSG) promoters in many cancers. As epigenetic modifications are reversible, reactivation of TSGs in cancer cells by targeted demethylation of their promoters, leading to proliferation arrest and/or apoptosis is an attractive therapeutic approach. Our laboratory recently identified a new class of TFs, called super pioneer factors (SPFs), that can bind to methylated DNA, leading to demethylation and chromatin decondensation, thus increasing DNA accessibility.

Here, we explore the possibility of reactivation of methylated promoters of TSGs p16 and Rassf1a in Hepatocellular Carcinoma (HCC) cell lines, by recruitment of SPFs via CRISPR/Cas9-mediated insertion of their binding motifs. We also targeted a methylation sensitive reporter in mouse embryonic stem cells (mESCs). Results show that insertion of binding motifs of two SPFs, CTCF and SOX2, leads in most cases to local demethylation and in some cases to reactivation of target genes. However, neither demethylation nor reactivation seem to be consistent and efficient, suggesting that SPF motif insertion is not a robust technique to accomplish these goals. On the other hand, we show that recruitment of a dCas9-TET1 effector to the methylated p16 promoter leads to robust demethylation and an increase in p16 mRNA resulting in slower growth and senescence of cancer cells. This result confirms the feasibility of our approach and that other SPFs, as well as different insertion locations within the promoters need to be tested.

Other interesting targets of SPF-dependent activation, from a therapeutic point of view, are genes involved in adipocyte browning. This phenomenon involves the emergence of brown-like (beige) adipocytes within white adipose tissue (WAT) and was recently reported to take place in mouse visceral adipose tissue (VAT) after a combined treatment with cold and antibiotics (CAx). Beige adipocytes are of particular interest, since they are associated with higher glucose uptake, insulin sensitivity and a general amelioration of diabetes and obesity. To identify methylation events involved in this process, we performed methylome and transcriptome analysis on VAT in mice after and before CAx treatment. We identified 152 differentially expressed genes (DEGs) and 419 differentially methylated regions (DMRs) with the majority being hypomethylated after CAx treatment. However, no overlap between the DMRs and the regulatory regions of the DEGs could be observed. This could be due to the heterogeneity of the sample that only allows the identification of DMRs and DEGs with the highest differences. Overcoming this heterogeneity could identify relevant DMRs that could be targeted for activation by SPFs or other epigenetic editors.

Résumé

La méthylation de l'ADN, une modification épigénétique importante, peut inhiber l'expression des gènes en réduisant l'accessibilité et l'affinité de l'ADN envers les facteurs de transcription (FT). Par conséquent, la méthylation de l'ADN, lorsqu'elle est mal distribuée, peut entraîner des maladies en inhibant des gènes essentiels pour la fonction des cellules saines. Ainsi, l'inhibition par méthylation des gènes suppresseurs de tumeurs (GST) est une caractéristique de nombreux cancers. Comme les modifications épigénétiques sont réversibles, la réactivation des GST par une déméthylation ciblée de leurs promoteurs afin d'induire l'arrêt de la prolifération et/ou l'apoptose dans les cellules cancéreuses, pourrait être une approche thérapeutique intéressante. Notre laboratoire a récemment identifié une nouvelle classe de FTs, appelés facteurs super pionniers (FSP), pouvant se lier à l'ADN méthylée et induire une déméthylation et la décondensation de la chromatine, augmentant ainsi l'accessibilité de l'ADN.

Ici, nous explorons la possibilité de réactivation ciblée des promoteurs méthylés des GSTs p16 et Rassf1a dans des lignées cellulaires de cancer hépatocellulaire (CHC), par recrutement de FSPs suite à l'insertion de leurs motifs de liaison via CRISPR/Cas9. Nous avons aussi ciblé un rapporteur sensible à la méthylation dans des cellules souches embryonnaires de souris. Les résultats montrent que l'insertion de motifs de liaison de deux FSPs, CTCF et SOX2, conduit dans la plupart des cas à une déméthylation locale et dans certains cas à une réactivation des gènes cibles. Cependant, ni la déméthylation, ni la réactivation ne semblent être reproductibles et efficaces, ce qui suggère que l'insertion de motifs de FSPs ne semble pas être la bonne méthode pour atteindre ces objectifs. Cependant, nous montrons que le recrutement d'un effecteur dCas9-TET1 au promoteur p16 méthylé, conduit à une déméthylation et l'augmentation d'ARNm de p16 ralentissant ainsi la prolifération et induisant la sénescence des cellules cancéreuses. Ce résultat confirme la faisabilité de notre approche et que d'autres SPFs, ainsi que différents emplacements d'insertion dans les promoteurs doivent être testés.

D'autres cibles potentiellement intéressantes d'un point de vue thérapeutique seraient des gènes impliqués dans le "browning" des adipocytes. Ce phénomène implique l'émergence d'adipocytes beiges, ressemblant aux adipocytes bruns, dans le tissu adipeux blanc (WAT), et il a récemment été observé dans le tissu adipeux viscéral (TAV) des souris, suite à leur exposition au froid combiné avec un traitement aux antibiotiques (CAx). Ces adipocytes beiges présentent un intérêt particulier, car ils sont associés à une meilleure absorption de glucose et à une sensibilité à l'insuline ainsi qu'à une amélioration générale de l'état des souris obèses et diabétiques. Pour identifier les changements de méthylation qui seraient impliqués dans ce processus, nous avons effectué une analyse du méthylome et du transcriptome du TAV chez la souris après et avant le traitement CAx. Ces analyses ont révélé 152 gènes différentiellement exprimés (DEG) et 419 régions différentiellement méthylées (DMR), la majorité étant hypométhylée après traitement par CAx. Cependant, aucun chevauchement entre les DMRs et les régions régulatrices des DEGs n'a pu être observé. Cela pourrait être dû à l'hétérogénéité de l'échantillon qui ne permet que l'identification des DMRs et DEGs présentant les différences les plus importantes. Réduire cette hétérogénéité pourrait identifier des DMRs pertinents comme cibles d'activation par les FSPs ou d'autres éditeurs épigénétiques.

1 Introduction

1.1 Epigenetics and the Regulation of Gene Expression

When taking a close look at two distinct cell types within an organism, e.g. a mature white adipocyte and a dopaminergic neuron, one can notice striking differences in both form and function. The adipocyte stores energy in form of lipids and possesses a simple round shape, while the neuron displays exotic features such as branching dendrites and an axon, which enables information transfer in an almost transistor-like manner. Astonishingly, both cells have the exact identical genotype, the set of genetic information stored within the DNA (deoxyribonucleic acid). Yet, how is it possible that both cells present such vastly different features? The answer, at least in part, can be found in differences in gene expression and one of its major regulators, the epigenetic code.

1.1.1 From Information to Function: Gene Expression and its Regulation

The genome contains all the information that is needed to build, maintain, and regulate an organism. The genetic code, an arrangement of the four nucleobases: adenine (A), cytosine (C), guanine (G), and thymine (T), evolved over the last four billion years¹ through random mutation and natural selection, and thus, gave every species and almost every non-asexual individual its unique genetic information.

Information by itself, however, is only part of the equation, as it needs to be converted into some type of function. In 1956, Francis Crick proposed the 'Central Dogma of Molecular Biology'^{2,3}, which states that DNA is first transcribed into RNA (ribonucleic acid), and RNA is translated into protein, and thus, function (Fig. 1.1A). Furthermore, the information flow may never go backwards from protein to RNA or DNA. We now know that sequential information can be transferred from nucleic acids to nucleic acids (from DNA to DNA during DNA replication, from DNA to RNA during transcription, and in special cases from RNA to DNA

during reverse transcription), as well as from nucleic acids to amino acid chains during translation. However, once the information is inside the amino acid sequence, it can never be transferred back to a nucleic acid sequence. While highly simplified, the Central Dogma still holds true today and gives a framework for all processes involving gene expression.

The flow of information is an essential multistep process, and at its beginning lies the gene, a physical and functional unit of DNA within the genome, which gives rise to transcripts that can be either protein-coding or non-coding. The exact number of protein-coding genes in humans has been subject to constant updates and has been corrected downwards over the decades from initially 50,000-100,000 genes⁴⁻⁶, to 30,000-40,000 when the human genome was published in 2001⁷, and further down to 20,000-25,000⁸. Currently, this number is believed to be 21,306, with about the same number of non-coding genes⁹. Furthermore, in eukaryotes, each gene may be the template for multiple transcripts (~7.5 on average in humans⁹), due to alternative splicing^{10,11}. However, only a subset of these genes is getting expressed in a given cell at a given time. This subset, or gene expression profile, may depend on the acquired cell type within an organism, on external *stimuli*, on pathologies, or on internal cellular mechanisms such as imprinting or X chromosome inactivation. Spatiotemporal control, i.e. if, to what level, when, and where a gene is expressed, is of utmost importance. Therefore, gene expression is a highly regulated process, both on the transcription and translation level. In this work, we mostly deal with the first step of information flow: the regulation of transcription of DNA into messenger RNA (mRNA).

The Transcription Machinery

The quintessential enzyme in the transcription process is the RNA polymerase, and, consequently, there is a number of processes involved in recruiting (or preventing to recruit) it to the transcription start site (TSS) of a gene. All organisms have, therefore, developed highly

sophisticated mechanisms to control the initiation of transcription, employing a vast array of cis- and trans-regulatory elements.

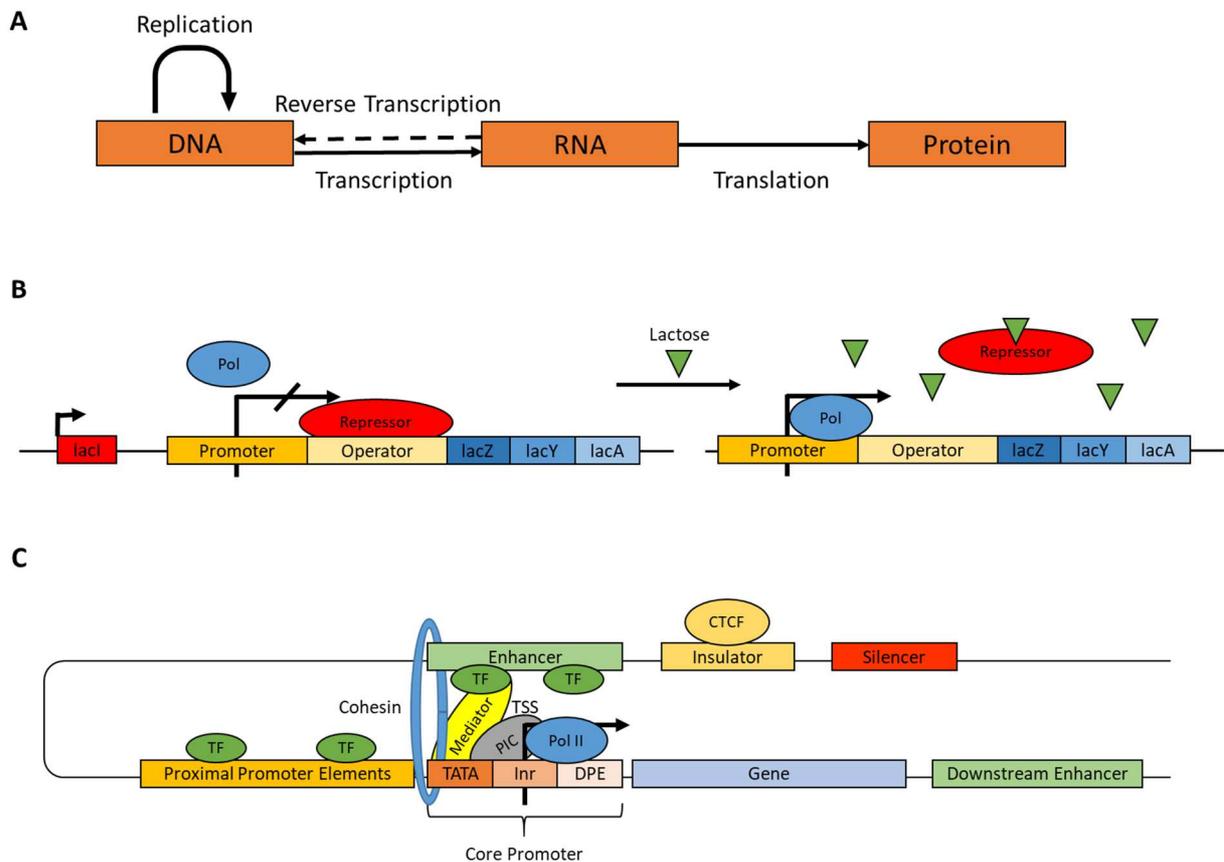


Figure 1.1: Regulation of Gene Expression. **A)** The Central Dogma of Molecular Biology: DNA is transcribed into RNA, and RNA is translated into proteins, but the flow of information may never go back from protein to nucleic acid. **B)** Example of transcription regulation in prokaryotes: Regulation of the *lac* operon by coordination of cis- and trans-regulatory elements (CREs and TREs, respectively), such as promoters, operators, and repressors. Left: *lac* operon in repressed state. Right: Active *lac* operon in the presence of lactose. **C)** Initiation of gene expression in eukaryotes via the coordination between distal and proximal CREs, the core promoter, and transcription factors (TFs).

Transcription Regulation in Prokaryotes

A common feature of gene expression regulation in prokaryotic genomes is the operon. Here, a number of functionally related genes share the same regulatory elements and are therefore co-regulated and co-expressed. A prime example for an operon, and also the first one to be described, is the *lac* operon in *E. coli*^{12,13} (Fig. 1.1.B). Three genes necessary for lactose

utilization, *lacZ*, *lacY*, and *lacA*, are being co-expressed. To initiate transcription, RNA polymerase needs to bind to the promoter upstream from the gene cluster. Between the promoter and the genes is the operator sequence, which may be bound by a repressor, a protein with an allosteric site that, upon binding, represses RNA polymerase activity. In case of the *lac* operon, the repressor is expressed by *lacI* and its allosteric site may be occupied by lactose, which in turn prevents its binding to the operator, and, therefore, enables *lacZYA* expression in the presence of lactose^{12,13}. The cis-regulatory elements (CREs), i.e. regions of non-coding DNA that regulate transcription, are in this case the promoter and the operator, while *lacI* acts as a trans-regulatory element, i.e. a gene that encodes for a transcription factor (TF).

Transcription Regulation in Eukaryotes

While operons in rare cases also exist in eukaryotes^{14,15}, they are being increasingly replaced in higher organisms by more sophisticated architecture and topological complexity, which enables the recruitment of more distal CREs, as well as the involvement of an ever larger number of transcription factors (Fig. 1.1.C).

In eukaryotes, RNA polymerase II (Pol II) is responsible for the synthesis of mRNA and its activity is highly regulated by the coordination of three major classes of CREs: (1) the core promoter, (2) the proximal promoter, and (3) distal CREs, such as enhancers, silencers, and insulators. Core promoters are short elements of DNA that stretch roughly 50 base pairs (bp) upstream and downstream from the TSS. Their main objective is to serve as a binding platform for Pol II and the general transcription factors (GTFs), which together form the preinitiation complex (PIC)¹⁶. The core promoter contains several motifs in close proximity to the TSS, such as a TATA-box, an initiator (Inr) motif, or a downstream promoter element (DPE)¹⁷. The TATA-box, a sequence rich in Ts and As, is usually found 30bp upstream of the TSS¹⁸ and represents a recognition site for the TATA-box-binding protein (TBP)¹⁹, which in turn is a subunit of TFIID, one of the GTFs. While the TATA-box is a very well-known motif²⁰ and also

highly conserved in eukaryotes, it is only present in a minority of core promoters²¹. The more frequent, but also less well preserved Inr motif overlaps with the TSS²² and presents further elements for TFIID binding^{23,24}. The DPE is often present in core promoters containing an Inr motif, but lacking a TATA-box²⁵. In coordination with the Inr motif, the DPE is bound by subunits of TFIID, which requires appropriate spacing between the two motifs^{26,27}. It is, therefore, as the name suggests, located downstream of the TSS²⁵. The already mentioned GTFs, in order of assembly, are TFIID, TFIIA, TFIIB, TFIIF, TFIIE, and TFIIH, which themselves consist of multiple subunits such as TBPs, TBP-associated factors (TAFs), or ATPases²⁸. Typically, assembly of the PIC starts with the binding of TFIID to the appropriate motif in the core promoter. It then recruits TFIIA and TFIIB, followed by Pol II and TFIIF. Finally, TFIIE binds to the complex and recruits TFIIH, and both phosphorylate Pol II and unwind DNA due to their helicase activity^{29,30}. Subsequently, the PIC opens and the first nucleotides are synthesized^{27,31}.

While core promoters by themselves already exhibit weak basal activity³², coordination with other CREs is necessary for robust gene expression. Both proximal promoters and distal CREs contain multiple binding sites for sequence-specific transcription factors (TFs), that interact in some way with the general transcription machinery and Mediator, a protein complex consisting of a variable number of subunits, which regulates Pol II activity as well as chromatin architecture³³. Proximal promoters are typically found 100 to 1000 bp upstream from the TSS, whereas distal CREs can be both upstream and downstream from the TSS, often tens or hundreds of kilobases (kb) distant from the TSS³⁴. In order for distal CREs to bridge the distance, DNA-loops need to be established. These depend on proteins such as cohesin, Mediator, Yin Yang 1 (YY1), and CCCTC-binding factor (CTCF)^{35,36}. For example, both YY1 and CTCF bind to their target sites within a CRE and subsequently form homodimers with another CTCF or YY1 molecule at another CRE. YY1 seems to be mostly responsible for

enhancer-promoter interactions, while CTCF primarily establishes larger insulated neighborhoods binding insulators³⁶. On a lower level of its hierarchy, DNA architecture, especially promoter-enhancer interactions, is cell type specific and depends on many factors like epigenetics and cell-specific TF availability. Some enhancers may also interact with multiple genes. Among these count the so-called super enhancers, which are larger in size and contain more TF binding sites, especially for TFs that are markers for certain cell identities, such as the Yamanaka pluripotency factors^{37,38}. Conversely, several enhancers are often needed to regulate the expression of one gene³⁹⁻⁴¹.

Profiling Regulatory Elements

Proximal regulatory elements are easily identifiable due to their proximity to their target genes. On the other hand, distal regulatory elements are difficult to locate as they could be located at any distance from their target genes. Currently, several approaches are used to identify enhancers, assess their activity and identify their target genes. One approach used to identify an enhancer consists in the identification of enhancer-related compaction profiles of the chromatin, a higher-order structure composed of the DNA and histone proteins. Chemical modifications of the DNA or the histones, called epigenetic modifications, could affect the accessibility of the DNA. These notions will be detailed in the chapters 1.1.3 to 1.17. Enrichment for enhancer-specific histone marks can be determined genome-wide with chromatin immunoprecipitation (ChIP) sequencing (ChIP-seq)⁴² using antibodies that recognize these marks. ChIP will be further explained in the chapter 1.1.2. While specific chromatin status correlates very well with enhancer positions, it does not allow to identify target genes and to assess the functionality of the putative enhancers. Techniques based on chromosome conformation capture (3C) make it possible to identify enhancer-promoter loops, as well as other higher order chromatin structures⁴³⁻⁴⁶. All 3C methods have in common that DNA gets first cross-linked *in vivo*, and afterwards digested with restriction enzymes at a low concentration and for a short period of

time. This allows for intramolecular ligation, connecting exclusively pieces of DNA, which were originally several kilobases apart from one another, but are brought together due to looping. After de-crosslinking, the fragments around the ligation sites are further processed and analyzed, for example by sequencing. Sequences that result from a chimera of two sequences that are not in proximity on the linear DNA sequence define two interacting regions. To identify a single, often predicted, enhancer-promoter connection, 3C that uses oligos hybridizing to pre-determined sequences is sufficient⁴³, while circularized chromosome conformation capture (4C) can determine all connections with one specific locus⁴⁵, and Hi-C identifies in an unbiased fashion all connections within the genome of a given cell type⁴⁴. While these methods will yield mostly cell type-specific interactions, they cannot determine whether these interactions are functionally important for gene transcription. Self-transcribing active regulatory region sequencing (STARR-seq) can identify enhancer activity for millions of arbitrary DNA sequences (from sheared genomic DNA), by cloning them downstream of a minimal promoter into a reporter vector⁴⁷. The vector library will be transfected into cells. If the region possesses enhancer activity, it will be able to transcribe itself. The resulting mRNA can be isolated and reverse-transcribed into cDNA (cyclic DNA). Fragments which were able to amplify themselves can now be detected by deep sequencing approaches⁴⁷. Finally, mutational analysis using CRISPR-Cas9 based approaches is used to assess the functionality of a given enhancer in its endogenous context⁴⁸.

Taken together, gene expression is regulated on multiple levels, which involves CREs and TFs as well as their ability to interact with each other, i.e. their accessibility, which will be explored in the upcoming chapters.

1.1.2 The Biology of Transcription Factors

One integral part of the complex system that controls the gene expression profile of a cell in a given state are transcription factors (TFs). These are proteins capable of interacting directly

with the DNA in a sequence-specific manner, in order to positively or negatively regulate transcription. They form complex networks that control processes like differentiation⁴⁹, proliferation, or pathways downstream of external *stimuli*. They can often be regarded as markers for cell types or certain cell states⁵⁰.

For a TF to conduct both DNA binding and transcription regulation, it has to contain two main components: one or more DNA-binding domains (DBD) and one or more effector domains (Fig. 1.2A). Out of the roughly 100 known DBDs in eukaryotes^{51,52}, C2H2-zinc finger (ZF), Homeodomain, basic helix-loop-helix (bHLH), basic leucine zipper (bZIP), and nuclear hormone receptor (NHR) are the most abundant domains, and TFs containing such domains are classified into families⁵³. As with the number of genes, the actual number of human TFs is also continuously updated. However, here the number is ever increasing. The current census predicts 1639 human TFs⁵⁴, more than half contain either C2H2-ZF domains or homeodomains (Fig. 1.2B)⁵⁴. To add to the complexity, some DBDs can bind multiple distinct sequences. Zfp335, for example, a TF with a C2H2-ZF domain is able to independently interact with DNA using either the N- or C-terminal section of its DBD⁵⁵. On the other hand, there are also proteins that contain regions homologous to DBDs, however they do not seem to interact with DNA at all. Finally, in some cases the DBD does not seem to be essential for the function of the TF. For example, it has been shown that complete deletion of the homeodomain of CERS class TFs, which is embedded within multiple transmembrane regions, does not affect their function⁵⁶.

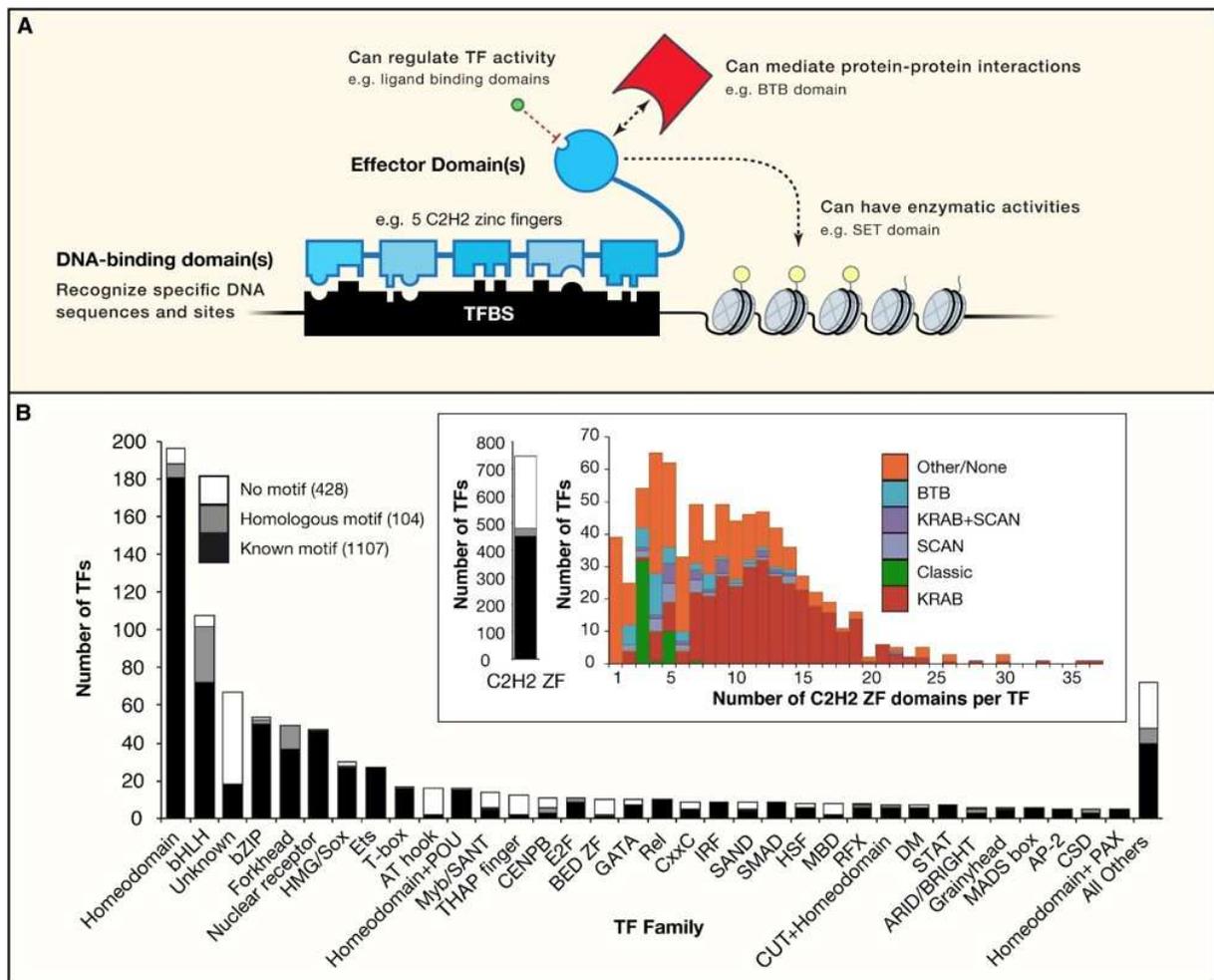


Figure 1.2: Human Transcription Factors⁵⁴. **A)** Schematic representation of a transcription factor (TF), containing a DNA-binding domain (DBD) and an effector domain, whose mechanism and impact on transcription may vary drastically between different TFs. **B)** Number of known TFs by DBD and status of motif identification. Insert: Distribution of effector domain classes per number of C2H2-zinc finger (ZF) domains.

While the purpose of a DBD is fairly apparent, the impact of the effector can vary drastically between TFs. In some cases, TFs simply prevent other proteins from binding through steric hindrance. One such example is the aforementioned *lac* repressor¹³. Usually, however, TFs recruit additional cofactors, which may form vast complexes and influence transcription in a multitude of pathways including modulation of histone modifications, nucleosome occupancy, phosphorylation of proteins, or DNA architecture. TFs with a Krüppel-associated box (KRAB) domain, for example, recruit KRAB-associated protein-1 (KAP1), which itself recruits Histone-lysine N-methyltransferase (SETDB1) and Heterochromatin protein-1 (HP1), resulting in

restrictive histone mark enrichment, and, consequently, closing of the surrounding chromatin⁵⁷. Other TFs may recruit p300 histone acetyltransferase, which in turn leads to open chromatin⁵⁸. As many TFs recruit multiple cofactors, sometimes with opposite effect, a strict separation between ‘activators’ and ‘repressors’ is not recommended^{59,60}. Their ultimate effect is in the end also determined by availability of cofactors and the local genomic context^{61,62}.

The preferred DNA sequence or, ‘motif’, to which a TF binds throughout the genome can be determined with *in vitro* as well as *in vivo* techniques. Systematic evolution of ligands through exponential enrichment (SELEX)^{63,64} is a method in which TFs are added *in vitro* to a pool of randomized DNA sequences. The bound sequences are being separated from the unbound ones, amplified by PCR, and both sequenced as well as reintroduced to a new pool of sequences for a new round of SELEX. While SELEX has the benefit of being high-throughput, it cannot identify the genome-wide binding locations of a given TF. For that purpose, *in vivo* techniques based on chromatin immunoprecipitation (ChIP) can be performed. Here, proteins are cross-linked to DNA using formaldehyde or UV-light, and subsequently precipitated with the corresponding antibody. The enriched DNA, bound to the protein, is then isolated and analyzed by qPCR, microarray (ChIP-chip), sequencing (ChIP-seq)⁴², or the related, but more precise, ChIP-exo⁶⁵, which employs exonucleases to increase resolution. However, ChIP-based techniques also come with a number of drawbacks. If the determination of a motif is desired, the obtained consensus sequence might be skewed, depending on the chromatin state and the resulting TF-binding capacity inside the given cell type. The data is also highly dependent on antibody quality, and, it is impossible to distinguish between indirect and direct binding⁶⁶. A position weight matrix (PWM) can be created from the pool of sequences identified as bound by the given TF⁶⁷. For each given position within the motif a score for the probability of occurrence of the four bases is calculated, giving the PWM as a result. The score indicates the

affinity of a TF for the given sequence. The obtained motifs can then be represented in form of a sequence.

There is, however, only a partial overlap between the predicted motif and the actual occupation of all possible TF-binding-sites in the genome⁶⁸. In fact, aside from CTCF^{69,70} that may not be regarded as a *bona fide* TF, the vast majority of TFs only bind to a small fraction of their respective motif matches in a cell-type specific fashion. These observations indicate that TF-binding is influenced by their ability to access the DNA and to cooperate with each other, i.e. by changing the chromatin state or through cooperative binding⁷¹. In a study using fusion protein pairs, in which TFs or transcriptional cofactors were coupled to a GAL4 DBD, in different enhancer contexts, TFs were categorized into 15 clusters based on their functional profiles. This allowed the prediction of functional distinct TF-TF and TF-cofactor pairs⁷². In terms of binding affinity, TF-TF pairs with different individual binding characteristics and from different structural classes have been identified by consecutive affinity-purification SELEX (CAP-SELEX)⁶⁴. The majority of identified pairs share structural traits with their partners, whilst the minority are structurally distinct from one another. This study has also shown that spacing and motif orientation is critical for cooperative binding⁶⁴. The pluripotency factor c-Myc, which contains a bHLH domain, is an example for cooperativity in terms of accessing its nucleosomal target, as it needs to associate with other factors, such as Oct4 or other homeodomain-containing proteins, to bind a degenerate E-box motif (CANNTG instead of CACGTG)⁷³.

While TFs and the genomic context of regulatory regions of a gene play an important role in the regulation of gene expression, the accessibility of these sites is also critical to its spatiotemporal control. Factors influencing this accessibility will be further discussed in the following chapters.

1.1.3 Chromatin and Epigenetics

So far, for the first two chapters, this work has treated the relationship between CREs and TFs, as if the DNA was ‘naked’ or unobstructed. However, information needed for transcription regulation is, in its physiological form, not just DNA, but a sophisticatedly orchestrated complex of a plethora of proteins with DNA, as well as direct chemical modifications on the DNA molecule itself. This complex is defined as chromatin and its composition and architecture are keys to the understanding of the regulation of gene expression. Even cytologically, regions of distinct chromatin condensation within the nucleus are visible. These regions have been described already in the early 20th century, defining loosely condensed regions as ‘euchromatin’ and densely compacted regions as ‘heterochromatin’⁷⁴. From early on there have been speculations of a connection between the activity of genes and the level of condensation. Euchromatin is usually associated with active gene expression, and heterochromatin with gene silencing. In fact, the status of the chromatin is highly distinct in cells or organisms that share an identical genotype but express a variant phenotype.

In 1942, Conrad Waddington introduced the term ‘epigenetics’ as the part of biology which, in a nutshell, explains how different phenotypes can arise from same genotypes⁷⁵⁻⁷⁷. Later, the definition became a bit tighter, only considering heritable changes in the phenotype without changes in the genotype⁷⁸ as *bona fide* epigenetic changes. The structure of chromatin is highly dynamic and hierarchical, and epigenetic mechanisms act on each level⁷⁹ (Fig. 1.3). After the DNA double helix (~2nm), the smallest unit of DNA packaging is the nucleosome (~10nm), which consists of a histone protein core, around which the DNA strand is wrapped. This core, in turn, consists of two copies of each of the four core histone proteins (H2A, H2B, H3 and H4), forming the canonical histone octamer. These nucleosomes form ‘beads-on-a-string’ and can be either loosely (euchromatin) or tightly (heterochromatin) packed, forming chromatin fibers (~30nm). Heterochromatin can be further divided into constitutive and facultative

heterochromatin, with the former being variable between the different cell types of an organism, while the latter is densely packed in all cell types. Finally, chromatin is organized into topologically associated regions (TAD) and chromosomal territories.

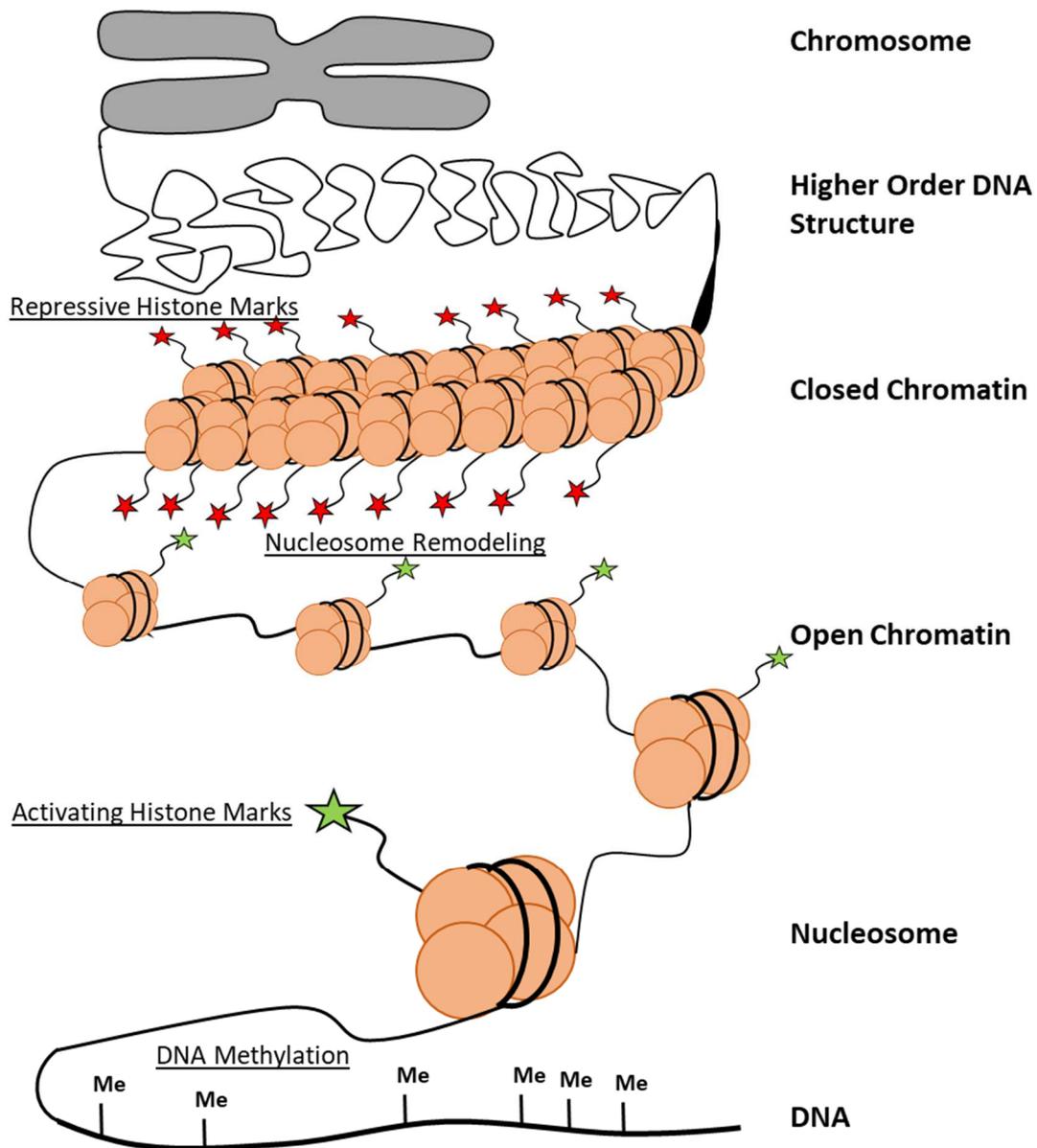


Figure 1.3: Epigenetic Mechanisms and Hierarchical Structure of the Chromatin. The DNA molecule is wrapped around a histone protein core, forming a nucleosome, the smallest unit of chromatin packaging. Nucleosomes can be loosely or densely packed, establishing regions of open chromatin with higher gene activity (euchromatin) and of closed chromatin in which gene activity is repressed (heterochromatin). Finally, chromatin fibers are organized in higher order structures. On each level, epigenetic mechanisms, such as DNA methylation, histone modifications, or nucleosome rearrangements are involved.

The exact definition of the term “epigenetics” and what it exactly covers have been frequent subjects for debate. However, among its chief mechanisms are almost always counted (1) nucleosome remodeling, (2) histone modifications, and (3) DNA methylation⁷⁹. DNA architecture and higher order chromatin organization can both be seen as an epigenetic mechanism or as consequence of such. At times, the line between cause and consequence of certain marks in epigenetics becomes blurry and is in many cases not fully understood. Ultimately, they all have in common to be part of the intricate machinery involving regulatory elements, which in the end determines gene expression. This being said, epigenetic modifications have been involved in the regulation of virtually all processes that use DNA as a template, which include transcription, DNA replication and repair. Furthermore, epigenetic marks contribute to the stability of the genome by silencing genomic regions like telomeres, centromeres, and transposable elements. These marks reduce recombination events of repetitive regions and ensure correct microtubule attachment⁸⁰⁻⁸².

Lastly, long non-coding RNAs (lncRNAs) are often considered as part of epigenetics⁷⁹. These are RNA fragments of more than 200bp, which are not translated into proteins^{83,84}. While an increasing number of lncRNAs is getting identified, their exact function remains elusive. It is, however, evident that they are involved in a multitude of biological processes^{84,85}. Of the ones with clearly defined function, some target gene transcription by acting on Pol II⁸⁶, others are involved in epigenetic regulation. For example, Xist, one of the most prominent lncRNAs, is involved in the epigenetic inactivation of the X-chromosome⁸⁷. Kcnqot1, another lncRNA, directs certain histone modifications, responsible for imprinting, i.e. the allele-specific silencing of genes⁸⁸. However, this thesis will not cover lncRNAs in a dedicated chapter, since this work is mostly about chromatin and the relationship between the different CREs and TFs. Their impact on regulation of gene expression, however, is acknowledged, and they will be brought up in the appropriate context.

In the following chapters, epigenetic mechanisms from DNA architecture to DNA methylation will be explored. It has to be noted that, since most of the mechanisms are linked to one another, a clear separation by chapter without mentioning the others is hardly possible.

1.1.4 The 3D Architecture of the Genome

Although at first glance it might appear that way, chromatin in the nucleus does not exist in random tangles but is organized in highly ordered structures. Chromosome painting and Hi-C studies have shown that interphase chromosomes occupy preferred territories within the nucleus (Fig. 1.4A)^{44,89,90}. These so-called chromosome territories⁹¹ are themselves highly organized, with gene-rich, more likely to be transcribed regions towards the territorial border and gene deserts mostly in the center⁹²⁻⁹⁴. Chromosomal regions of similar activity also tend to interact with one another in an inter-chromosomal manner, forming compartments on the next, smaller scale of organization (Fig. 1.4B)⁴⁴. Active compartments, termed compartment A, are usually in the center of the nucleus, while inactive compartments, termed compartment B, tend to locate closer to the nuclear lamina or the nucleolus^{44,95}.

On the next lower level of chromatin organization are topologically associated domains (TADs)⁹⁶, which are self-interacting domains, with the purpose to confine regulatory interactions within their boundaries, but limit them between domains (insulation) (Fig. 1.4C)⁹⁷. These domains of insulated interactions can range from tens to hundreds and thousands of kilobases⁹⁶⁻¹⁰⁰. Regarding gene expression, TADs represent a module where genes within their boundaries tend to be co-regulated^{99,101,102}. This can be explained due to the restriction of contacts between enhancers and promoters within the same TAD¹⁰³. TAD boundaries are enriched in CTCF and cohesin^{96,104,105}, that, together, promote the establishment of loops, with cohesin extruding the loop until it reaches a CTCF anchor point¹⁰⁶⁻¹⁰⁸. The genomic context at these anchor points is also of importance for determining in which direction the loop will face. Indeed, at TAD boundaries, CTCF often binds two asymmetrical motifs facing each other.

Changing the direction of these motifs by genome engineering, leads to the disruption of the TAD and changes of enhancer-promoter interaction within its boundaries¹⁰⁹. This also emphasizes the importance of CTCF for TAD formation and maintenance. In general, disruption of TAD boundaries may lead to aberrant interactions between regulatory elements, and, therefore, can result in abnormal gene expression, which in turn may be a cause for disease¹¹⁰⁻¹¹³.

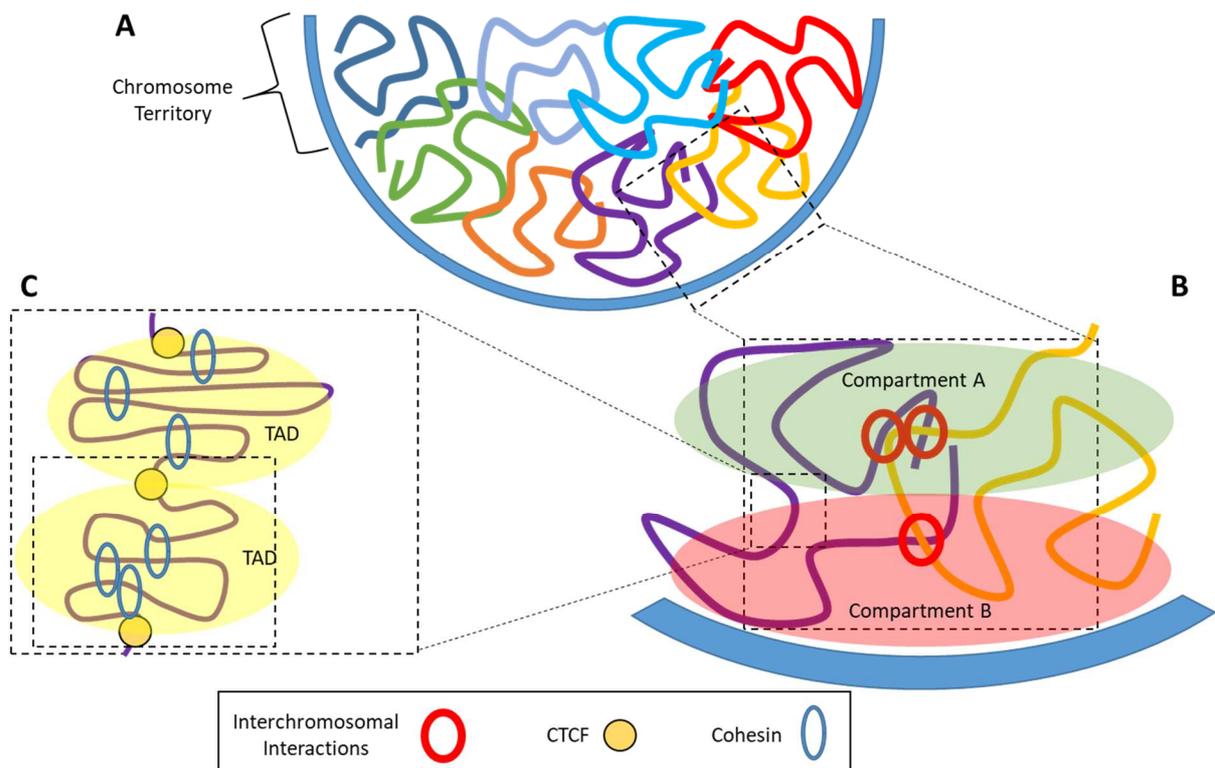


Figure 1.4: Hierarchy of 3D Chromatin Architecture. **A)** Chromosomes occupy specific territories within the nucleus. Active regions tend to be located towards the border of the chromosomal territory, while inactive regions tend to locate towards the center of the territory. **B)** Regions belonging to different chromosomes, but with similar transcriptional activity may interact with one another, forming compartments. **C)** Topologically associated domains (TADs) are formed to insulate loop domains between regulatory elements on the lowest scale of the hierarchy within specific boundaries. Anchor points of TADs are bound by CTCF, while cohesin is responsible for loop extrusion.

TADs have been shown to be consistent across cell types with very little cell type-specific variation^{96,114}. In fact, TADs have not only been shown to be present in most metazoans¹¹⁵, but

are also evolutionary conserved across species^{96,105}. Mice, for example, share about three quarters of their TAD boundaries with humans. CTCF binding sites at TAD boundaries are also less frequently subject to change within the boundaries, underlining the importance of conserving these sites¹⁰⁵. Some TADs harbor large amounts of conserved noncoding elements (CNEs) which are often developmental long-range enhancers, hinting at their role in conserving certain building blocks or modules within the genome over hundreds of millions of years¹¹⁶. This limitation to an insulated domain could allow changes in the composition of CREs during evolution, without affecting too many genes at once^{115,116}. Another observation supporting the modular function of TADs during evolution, is the closer proximity of genes to each other when they are within a TAD boundary than in neighboring TADs¹¹⁷.

It is fairly well established that the strength and organization of the 3D chromatin structure increases during development. Initially, chromatin in fertilized oocytes seems dispersed, but drastically compacts during the first transition to the 2-cell state¹¹⁸. Hi-C studies on preimplantation embryos, however, have shown that at least until the 8-cell state, chromatin is more “relaxed” due to weaker chromosome compartments and TADs¹¹⁹⁻¹²¹. Establishment of the chromatin structure during development, or, more precisely, reestablishment to the conditions of a fully functional organism of a species, is slower during early development with increasing kinetics while chromatin transitions from a relaxed to a more rigid state^{90,120-122}. The exact reasons for this initial weakening of the structure remain elusive, but emphasizes the special status of chromatin architecture during development¹²³.

During cell differentiation, the changes that occur on the different levels of the structural hierarchy are quite diverse. On the one hand, TADs remain more stable than compartments or intradomain loops^{96,114,124}, with roughly 20% TAD reorganization between embryonic stem cells (ESCs) and neuronal progenitors (NPs), for example¹²⁴. On the other hand, about one third of the genome undergoes a switch from compartment A to B or vice versa¹¹⁴, although this

switch leaves individual genes largely unaffected in regards to expression¹¹⁴. Changes in enhancer-promoter looping, however, are very distinct between different cell types and are a hallmark of differentiation^{125,126}. One of the key differences is, that in contrast to long-range interactions, which are facilitated by CTCF, other proteins are involved³⁶. As already mentioned, YY1 seems to be more involved than CTCF in the formation of enhancer-promoter loops³⁶. Moreover, loop formation does not necessarily require cohesion either¹²⁷, making CRE interactions more independent from key architectural proteins.

The 3D architecture of chromatin is highly ordered on across multiple levels and its correct organization is crucial in many processes involving development, differentiation, as well as diseases. Chromatin structures are shown to be both conserved at higher scales of hierarchy and highly modular on its lower levels. This is an important point to keep in mind when adding or removing potential boundary sites by genome editing.

1.1.5 Accessing Nucleosomal DNA

The degree of compaction of the chromatin could be dictated by the strength of the wrapping around the nucleosomes and the frequency of nucleosomes within a given stretch of DNA. A main player in promoting nucleosome-nucleosome interactions and subsequently nucleosomal frequency is the linker histone H1 that sits at the entry and exit point of the DNA coil from the nucleosome¹²⁸.

One of the main functions of chromatin and nucleosome dynamics is the stabilization of the genome during processes such as DNA damage repair, DNA replication, and chromosome segregation¹²⁹. However, along with nucleosome occupation, inevitably comes occlusion of the DNA, making it more or less accessible to TFs and members of the transcription machinery. Nonetheless, evolution took this occlusion and made it into a powerful tool to regulate gene expression. Densely packed chromatin with high nucleosome density is considered repressive, while open chromatin correlates with active gene expression. The state of the chromatin is, therefore, highly dynamic, and subject to changes that depend on epigenetic mechanisms.

Occupation of nucleosomes can be measured with several techniques that take advantage of the ‘protected’ status of the nucleosomal DNA¹³⁰. During DNase I hypersensitive site sequencing (DNase-seq)^{131,132}, DNA that is not protected by a nucleosome or a DNA-binding factor, is digested by the DNase I, a type II endonuclease. Depending on the protocol, the resulting fragments can be either labeled with a barcode¹³¹ or selected by size¹³². Subsequently, this enriches for a library of protected fragments whose ends can be sequenced. The information is then used to identify the borders of the unprotected regions, called DNase I hypersensitive sites (DHSs). Two types of DHSs corresponding to two levels of resolution can be identified: open, nucleosome free/poor-regions constitute the first level. Within these regions, DNA-binding factor occupancy modulates the level of sensitivity to DNaseI, thus leaving patches of less sensitive *loci* that correspond to the footprints of binding of these factors. In general, the vast

majority of DHSs can be found at regulatory elements: promoters and enhancers¹³¹⁻¹³³. An alternative method to detect open DNA is the assay for transposase-accessible chromatin using sequencing (ATAC-seq)^{134,135}. Here, hyperactive transposase Tn5 both cuts accessible DNA, and ligates adapters to the resulting fragments. The efficacy of the transposase allows the ATAC-seq protocol to be performed with both smaller samples (as little as 500 cells) and in less time than DNase-seq. Conversely, micrococcal nuclease (MNase) cleaves and eliminates internucleosomal DNA with both endo- and exonuclease ability. In contrast to DNase- and ATAC-seq, MNase-seq identifies DNA fragments spanning one nucleosome, and represents another method to measure chromatin accessibility^{136,137}. To test both chromatin accessibility and the methylation status of the DNA, nucleosome occupancy and methylome sequencing (NOME-seq) uses GpC methyltransferase M.CviPI to methylate GpC sites not occluded by nucleosomes^{138,139}. Subsequently, the DNA undergoes bisulfite treatment and can be sequenced to identify the regions that got methylated and are therefore considered accessible. This method is particularly interesting to assess methylation status and nucleosome occupancy of CpG-poor enhancers. Both DNA methylation and methods for its assessment will be covered in more details in Chapter 1.1.7.

Nucleosome occupancy is usually the lowest close to the promoter regions and TSS of active genes, followed by active, poised, and inactive enhancers¹³⁰. In heterochromatin the occupancy is the highest in constitutive chromatin, i.e. chromatin which is closed throughout the majority of cell types, followed by facultative chromatin.

Chromatin Remodeling

As it became clear through the discussion, nucleosome positioning is dynamic. The key mechanisms influencing nucleosome dynamics involve assembly, editing, and accessibility of nucleosomes¹⁴⁰ (Fig. 1.5A). In charge of those mechanisms are, for the most part, four families of adenosine triphosphate (ATP) -dependent protein complexes, commonly referred to as

remodelers¹⁴⁰. These families include: imitation switch (ISWI), chromodomain helicase DNA-binding (CHD), INO80, and switch/sucrose non-fermentable (SWI/SNF)¹⁴⁰ and are classified by their domain structure as well as the size of their transposase lobe separating insert. Although each family covers a preferential mechanism of nucleosome remodeling, all these complexes have a number of traits in common: (1) higher affinity for nucleosomal DNA over 'naked' DNA, (2) the possession of an ATP-dependent transposase domain, consisting of two RecA-like lobes, (3) an effector domain or subunit responsible for facilitating its mechanism of action, and (4) a domain or subunit regulating the ATPase¹⁴⁰. DNA translocation can be imagined as a sliding movement of the remodeler complex along the DNA (Fig. 1.5C), breaking its contact with the histone core: lobe 1 binds to the DNA leaving a small space between the two lobes. Subsequently, lobe 2 binds DNA in the presence of ATP, which triggers DNA release and a sliding movement to bridge the distance by lobe 1. Hydrolysis of ATP to ADP triggers lobe 2 to release the DNA and lobe 1 to bind it again. Lobe 2 now moves one to two nucleotides from 3' to 5' and ADP is released causing forcible DNA translocation¹⁴¹⁻¹⁴⁵. It is important to note that the translocation domain remains fixed in relation to the histone, causing the DNA also to translocate relative to the octamer. This is most likely facilitated by a histone-binding-domain (HBD), one of which has been proven to be present in SWI/SNF¹⁴⁶. Although this mechanism is widely recognized, other mechanisms were proposed to explain the mechanism of action of ATP-dependent nucleosome remodelers.

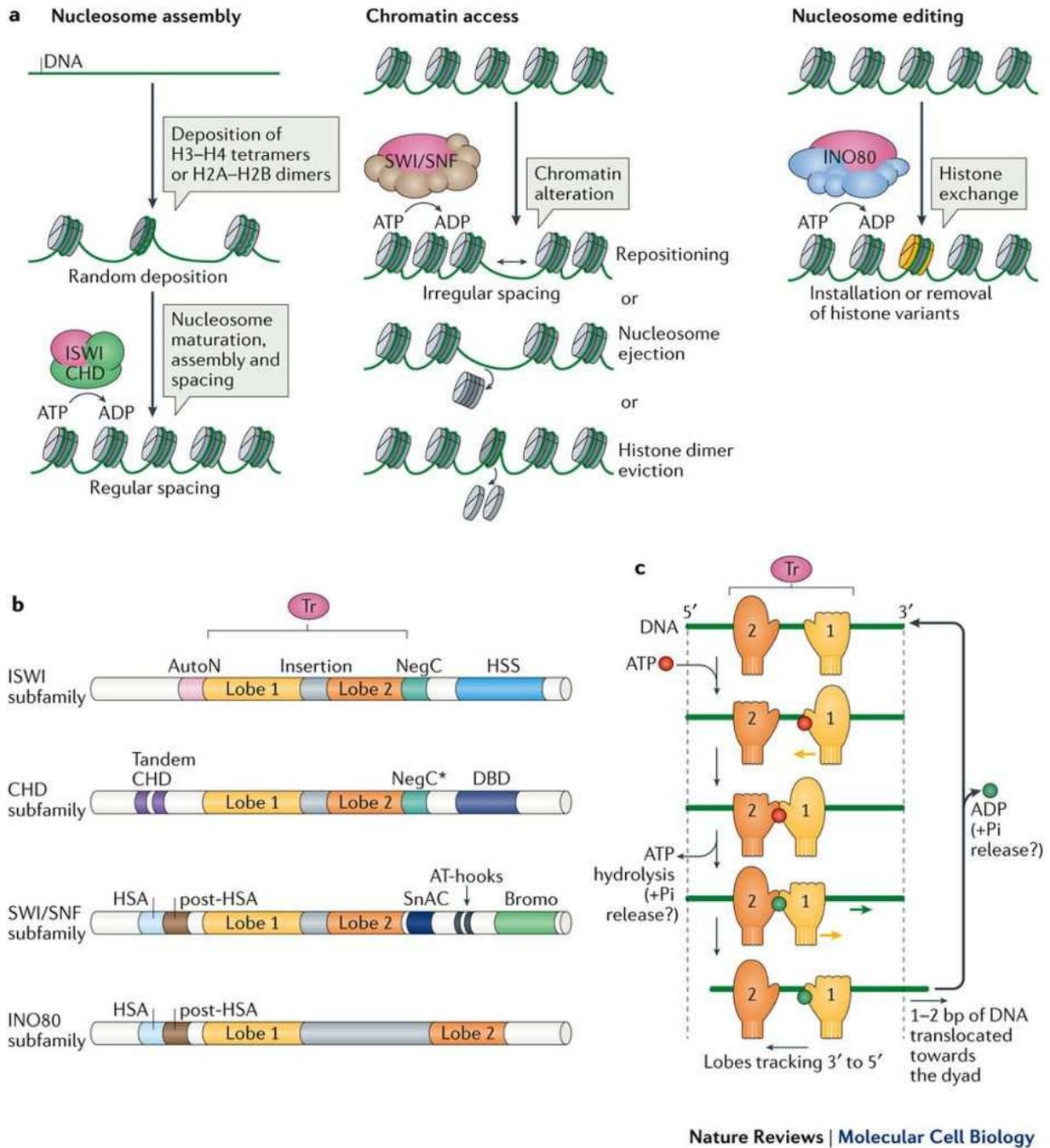


Figure 1.5: Nucleosome Dynamics by ATP-dependent Remodelers¹⁴⁰. A) The three different mechanisms of nucleosome remodeling involve assembly, editing, and accessibility. The latter of which can be broken down into repositioning and ejection of whole nucleosomes, as well as eviction of histone dimers. Members of the four families of chromatin remodelers preferentially carry out one of the remodeling processes, but exceptions, where one factor can use more than one type of remodeling, do exist. B) Schematic of the different domains/subunits defining the individual remodeler families. C) Mechanism of DNA translocation during one cycle of ATP hydrolysis.

Chromatin assembly and proper spacing is mostly carried out by members of the ISWI and CHD families. In ISWI complexes, the two RecA-like lobes of the translocase (Tr) domain are separated by a small insertion¹⁴⁷ and flanked by two domains regulating the ATPase activity, autoinhibitory N terminal (AutoN) and negative regulator of coupling (NegC)¹⁴⁸. Furthermore, it contains a C-terminal HAND-SANT (Swi3, Ada2, N-Cor, and TFIIB)-SLIDE (HSS) domain¹⁴⁹. CHD complexes share similarities with ISWI regarding the Tr insertion and NegC domain¹⁵⁰. However, they lack AutoN and HSS domains and possess a DBD (only consisting of SANT and SLIDE)¹⁵¹ as well as two N-terminal chromodomains^{152,153}. After replication, pro-nucleosomes, which do not contain the proper canonical octamer and around which DNA is not properly wrapped, are deposited with random spacing. The above mentioned remodelers of the ISWI and CDH families bind extranucleosomal DNA with their HSS and DBDs, and the histones with a histone binding domain, creating a measuring stick for spacing¹⁵⁴. In ISWI, the Tr-domain is inhibited by NegC and AutoN. However, binding of HSS to linker DNA and the Arg17 to Arg19 residues of the H4 histone tail, in turn inhibit NegC and AutoN, respectively, allowing translocase activity to happen¹⁴⁸.

Nucleosome editing, which involves removal and replacement of particular histones within the octamer with canonical or variant histones, is regulated by members of the INO80 family. Histone variants, such as H2A.Z can have a multitude of effects for example on gene expression¹⁵⁵ or repression of non-coding RNAs¹⁵⁶. Unlike with the other remodelers, the lobes of the Tr-domain are separated by a much larger insert of around 250 amino acids in yeast and more than 1000 amino acids in mammals^{157,158}, which, in yeast, binds further subunits, such as ruvB-like protein 1 (Rvb1) and Rvb2, actin-related protein (ARP), and YL-1^{159,160}. Other domains in INO80 are the helicase/SANT-associated (HSA) and post-HSA domains¹⁶¹. Examples for well-studied INO80 complexes in yeast are SWR1C for the exchange of the canonical H2A-H2B dimer with the variant H2A.Z-H2B¹⁶², and INO80C, which facilitates,

among other functions, the reverse exchange from H2A.Z-H2B to H2A-H2B¹⁶³. This dimer replacement is conducted by disruption of DNA-histone connections due to ATPase-mediated DNA translocation, which subsequently results in the release of the canonical or variant dimer and the loading with the respective dimer¹⁶⁴. For disruption of the DNA-histone connections it is crucial that the histone core remains in its position relative to the remodeler to establish sufficient tension on the DNA. In INO80C, this is made possible by ARP-facilitated binding of the HSA-domain to the histones with high affinity^{159,161}. After loading, DNA tension is released and DNA-histone connections are reestablished.

Finally, chromatin accessibility, which involves nucleosome sliding, eviction of histone dimers, or ejection of whole nucleosomes, is mostly regulated by members of the SWI/SNF family. Like INO80 complexes, SWI/SNF contains HSA and post-HSA domains, however, the amino acid insert between the Tr-domain lobes is more similar in size to ISWI and CHD^{165,166}. In addition, it contains a Snf2 ATP coupling (SnAC), which serves as HBD, AT-hooks and a bromodomain, which act as DBD^{165,166}. The exact mechanisms of sliding and ejection are still not fully understood. Clapier et al. propose two non-mutually exclusive models¹⁴⁰: (1) forcible translocation of the DNA around the octamer can be either weak or strong, which may result in simple sliding of the DNA around the core bound by the remodeler, or the complete ejection of the octamer, respectively. (2) ejection of the nucleosome adjacent to the one bound by the remodeler, by sliding of the linker DNA until there is no space between the two nucleosomes left, and subsequent spooling-off of the DNA around the adjacent octamer.

Especially in higher eukaryotes SWI/SNF complexes are highly modular molecular machines which show tissue specific changes in composition¹⁶⁷. This is of particular interest for the study of mammalian SWI/SNF (mSWI/SNF), which are an assembly of roughly 29 subunits including either SMARCA4 (BRG1) or SMARCA2 (BRM) as ATPases^{168,169}. A comprehensive study, using genetic manipulation of the subunits, mass spectrometry (MS), and crosslinking MS, has

identified three major classes of mSWI/SNF, canonical BRG1/BRM-associated factor (BAF), polybromo-associated BAF (PBAF), and non-canonical BAF (ncBAF)¹⁷⁰. These three classes show tissue specific activity and follow a strict order of assembly around a core BAF module¹⁷⁰. Interestingly, mutations in these subunits may impair the assembly and are major hallmarks in certain diseases¹⁷⁰, being present in 20% of all cancers¹⁶⁹ and up to 100% in some pediatric cancers¹⁷¹.

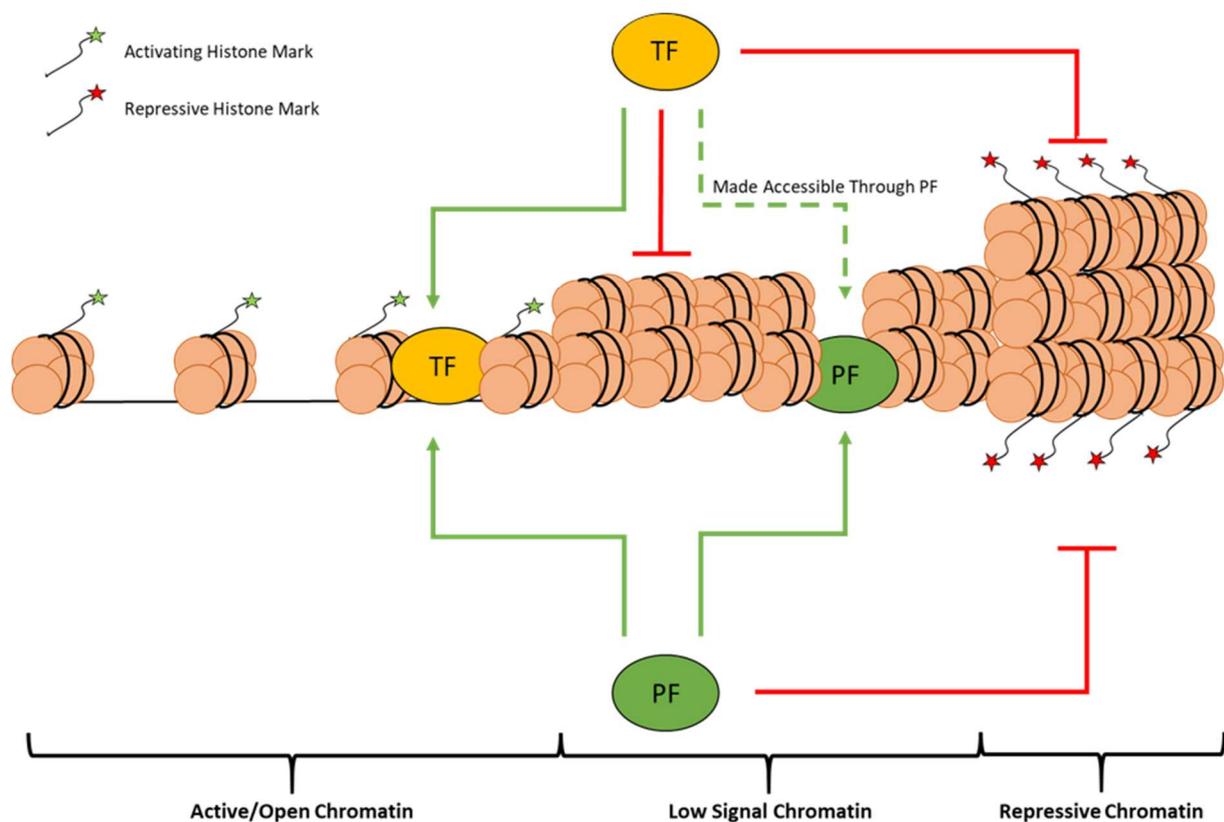


Figure 1.6: Pioneer Transcription Factors. Transcription factors (TFs) can usually only access open chromatin, which is marked by activating histone modifications such as H3K4me and H3K27ac, while low signal chromatin remains inaccessible. Pioneer transcription factors (PFs) can, however, access their binding sites in both open and low signal chromatin, i.e. chromatin with a low amount of histone modifications, and are able to open the latter for NPF binding. Repressive chromatin, which presents repressive histone marks such as H3K9me3 and H3K27me3 and shows dense nucleosome packaging, remains inaccessible to both.

Aside from sliding and ejection of nucleosomes, involving chromatin remodelers, chromatin can also be made accessible by binding of pioneer transcription factors (PFs). Borrowing its

nomenclature from the American Frontier, PFs are the first TFs to bind chromatin, inaccessible to most other TFs, and open it. This allows additional TFs, called “settlers” to access and bind these regions. For illustration purposes, chromatin can be divided into three major groups, regarding its accessibility by different TFs: (1) active or open chromatin, presenting activating histone marks such as H3 methylation on lysine 4 (H3K4me) and histone acetylation, (2) low signal chromatin, without particular histone marks, but with tighter packaging and linker histone H1 occupation, and (3) repressive chromatin, which is very tightly packed and highly enriched in H3K9 and H3K27 methylation (Fig. 1.6). Histone modifications will be explored in detail in chapter 1.1.6. The vast majority of TFs can only access active chromatin, while PFs can access their binding sites in both active and low signal conditions. Repressive conditions, however, are also closed to PFs. While PFs have their trait of gaining access in common, they show vast structural differences and even differ highly in their DBD, implying that their modes of binding to closed chromatin also differ. Members of the FOXA family were among the first to be identified as PFs¹⁷² and it has been shown that their winged helix DBD resembles the histone H1 linker^{173,174}, making them a competitor for this spot on the nucleosome. And it has indeed been shown that FOXA1 kicks out H1 and allows the recruitment of additional TFs¹⁷⁵⁻¹⁷⁸. This process is independent of ATP or other ATPases¹⁷⁸. Other prominent PFs are three of the four Yamanaka pluripotency factors Oct4, Sox2, and Klf4 (OSK)^{179,180}. As already mentioned in chapter 1.1.2, the fourth member, c-Myc, requires coordination with other factors to access its degenerate E-box motif on the nucleosome with its bHLH domain⁷³. O, S, and K themselves possess different DBDs with Pit-Oct-Unc (POU), Sry-related High Mobility Group (HMG), and Zinc Fingers (ZF), respectively. All are able to bind partial motifs, e.g. Oct4 can bind to a hexamer motif representing either half of its canonical octamer motif, Sox2 can access a ‘bent’ motif missing one of its canonical nucleotides, and Klf4 can bind a shorter motif with only two out of its three ZFs⁷³. Several PFs also play crucial roles in differentiation and reprogramming of cells. The already mentioned Yamanaka factors are sufficient to de-

differentiate fibroblasts into induced pluripotent stem cells (iPSCs)¹⁷⁹. Other examples include PU.1, GATA binding protein 4 (Gata4), achaete-scute homolog 1 (Ascl1), and FoxA, which take part in the conversion from fibroblasts to macrophage-like¹⁸¹, cardiomyocyte-like¹⁸², glutaminergic neurons¹⁸³, and hepatocyte-like cells¹⁸⁴, respectively. However, since most PFs do not bind all their possible binding sites in nucleosomal DNA¹⁸⁵, the extent of their independence from other co-factors and from specific genomic contexts is a matter of debate^{130,186}. Some may argue that PFs act in a rather facultative manner¹³⁰.

Taken together, chromatin accessibility is subject to a multitude of factors, such as complex molecular machines acting as remodelers, seemingly independent working PFs, and cooperativity between TFs. Another element, interacting with these factors are histone modifications which will be further explored in the next chapter.

1.1.6 Histone Modifications

One of the most studied epigenetic modifications, are post-translational covalent modifications of histone tails. These modifications occur at different amino acid residues mostly at the N-terminal histone tails that protrude out of the nucleosome. The number of known histone modifications is steadily increasing and already exceeds 100¹⁸⁷. Histone modifications regulate chromatin accessibility and therefore are involved in all cellular processes that use the DNA as a template, including gene expression regulation, DNA replication and DNA repair. They are highly regulated, show crosstalk among each other and with other epigenetic marks, highly affect nucleosome density, and interact with a plethora of TFs and co-factors. The nomenclature for different modifications starts with the name of the modified histone (e.g. H3), followed by single-letter abbreviation of the affected amino acid (e.g. K for Lysine) as well as its position. Finally, the type of modification is stated, e.g. ac for acetylation or me1/me2/me3 for mono-, di- and trimethylation, respectively. For example, H3K27me3 represents trimethylation of

lysine 27 of histone H3. Concerning their role in transcription regulation, some modifications are considered activating (e.g. H3K4 acetylation and methylation), while others fulfill an opposing role and correlate with repression (e.g. H3K27me_{2/3} and H3K9me_{2/3}). Moreover, regardless of their isolated proper function, different modifications could work in combination towards a common outcome (histone code). Among the most common and best understood histone modifications are acetylation and methylation of lysines (Fig. 1.7), which were also the first ones to be discovered¹⁸⁸. Histone acetylation involves the introduction of an acetyl group to a lysine (K) residue of a histone by histone acetyl transferases (HATs). This reaction is reversible through the activity of histone deacetylases (HDACs)¹⁸⁹. While all HATs require acetyl CoA as cofactor, they can be classified into two groups (A and B). HATs of group A (further divided into GNAT, MYST, and CBP/p300 families¹⁹⁰) are multisubunit complexes that predominantly work on chromatin in the nucleus¹⁸⁹, while HATs of type B mostly modify free histones in the cytosol¹⁹¹. Other subunits of HAT complexes with no acetyltransferase activity, play an important role in regulating the ability and specificity to modify certain residues. For example, scGCN5 by itself is only able to modify free histones *in vitro*, whereas *in vivo* it works on nucleosomal histones¹⁹².

Histone acetylation, e.g. on H3K9, H3K14, H3K18, H3K27, H4K5, H4K8 and H4K12, correlates exclusively with open chromatin and transcription activation. One important mechanism to explain the role of this modification in de-condensing the chromatin is the fact that the addition of an acetyl group to a lysine residue neutralizes its positive charge, and thereby weakens the interaction between DNA and histones¹⁹³. Evidently, *loci* with high acetylation are also enriched with DNase I hypersensitive sites¹⁹⁴. Additionally, histone acetylation could prevent repression by competing out other modifications at the same amino acid. For example, H3K27ac found at active and poised enhancers, as well as active promoter regions, competes with H3K27me₃, which is a repressive mark, in a mutually exclusive manner¹⁹³. Interestingly,

activating and repressive modifications affecting different substrates, such as H3K4me2/3 and H3K27me3 can coexist at the same promoter that is set to an inactive state but is poised for activation (bivalent promoters)¹⁹⁵. Finally, aside from neutralizing the charge, acetylated lysine residues are also recognized by bromodomains, which are present in the complexes of chromatin remodelers such as SWI/SNF^{196,197}.

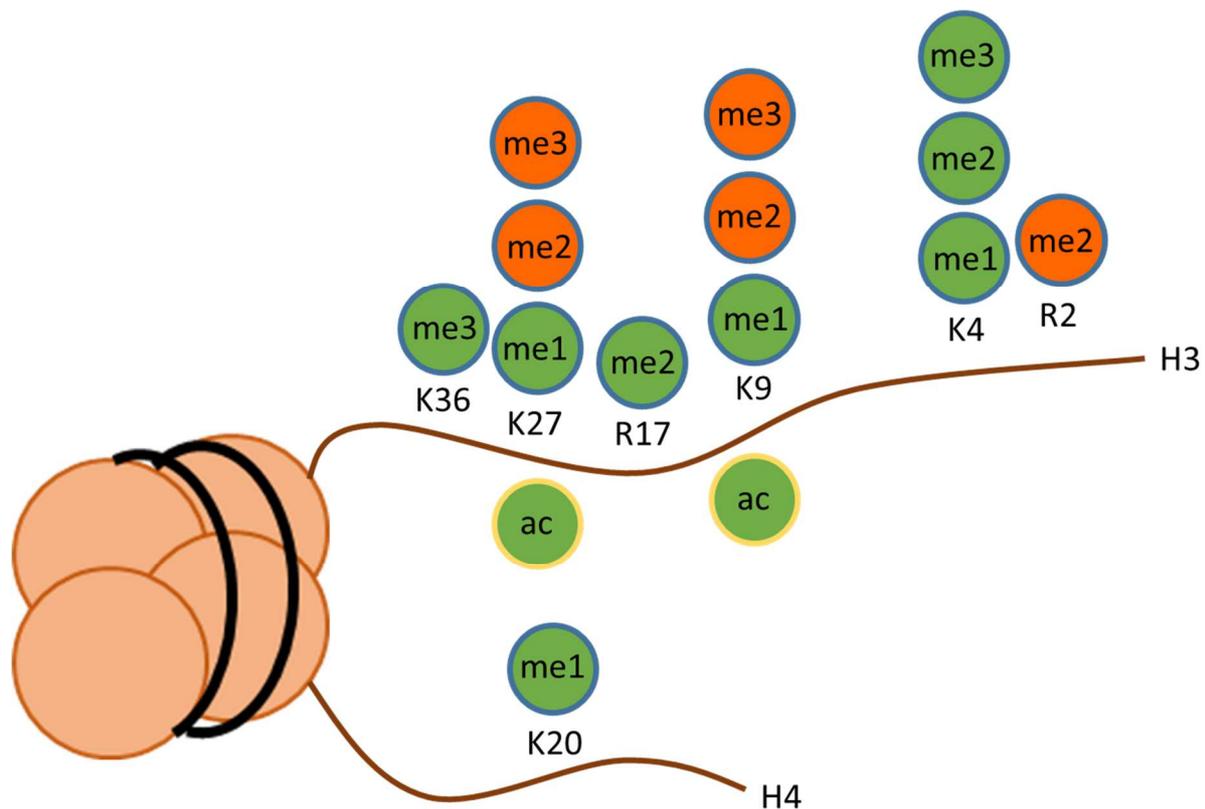


Figure 1.7: Representative example of histone methylation and acetylation and their Association with Gene Expression. The N-terminal tail of histone H3 harbors some of the most significant and well understood histone modifications in regard to gene expression. It must be noted, that different, mutually exclusive modifications can occur on the same residue, which may have antagonistic effects. Modifications indicated in green are associated with active gene expression, orange ones with repression.

In contrast to acetylation, histone methylation does not change the charge of the residue. Both lysine and arginine residues can be methylated and demethylated, each with its own specific set of enzymes. Using S-adenosylmethionine (SAM) as a methyl donor, histone lysine

methyltransferases (HKMTs), such as SUV39H1 and arginine methyltransferases, such as PRMT1, introduce one or multiple methyl groups to a lysine's ϵ -amino group¹⁹⁸ or to the ω -guanidino group of arginine¹⁹⁹, respectively. Histone methyltransferases tend to be highly specific both regarding their target residue as well as regarding the number of transferred methyl groups^{200,201} (mono-, di-, or trimethylation for lysine and mono-, or dimethylation for arginine). In contrast to HDACs, which remove their target modifications with relatively low specificity, histone demethylases are also highly selective regarding their targets and require proper complex composition. For example, lysine-specific demethylase 1 (LSD1) demethylates H3K4me1/2²⁰², but only in presence of the Co-REST repressor complex²⁰³. If androgen receptor is part of the complex, H3K9me becomes the target²⁰³. While LSD1 can only remove mono-, and dimethylation, members of the JMJD2 class demethylate H3K9me3 and H3K36me3²⁰⁴.

Instead of changing the histone's charge, methylation mostly serves for interaction with various proteins, which may vastly differ in their effect, sometimes activating, sometimes repressing. For example, H3K4me3 binds ING family proteins which in turn recruits HATs²⁰⁵, therefore leading to chromatin opening. On the other hand, H3K9me3, a typical mark for repressive chromatin, binds HP1 (Heterochromatin protein 1), an important component of heterochromatin^{206,207}. In contrast to providing a platform for binding, some modifications can also serve as obstruction preventing factor binding. This is the case for H3K4me3 that prohibits the binding of nucleosome remodeling deacetylase (NuRD)^{208,209}, a complex which both serves as remodeler for nucleosome assembly as well as deacetylase.

Throughout the genome, certain histone marks may be predictive of certain chromatin features, not only hetero- and euchromatin, but also specific CREs, like enhancers and promoters. H3K4me1 is present in active and poised enhancers²¹⁰, while H3K27ac is a typical mark for active enhancers and can be used to distinguish the two. On the other hand, H3K4me3 is found at active promoters²¹¹ and H3K36me3 throughout the actively transcribed gene body²¹².

Conversely, H3K27me3 is mark for gene repression, more specifically, facultative heterochromatin²¹³, and H3K9me2/3 for constitutive heterochromatin²¹³.

Other post-translational histone modifications include phosphorylation, ubiquitination, sumoylation, deimination, addition of β -N-acetylglucosamine (O-GlcNAc), and histone tail clipping, however these will not be covered in this thesis. It is, nevertheless, important to note, that these and other, yet to be discovered, modifications play a role in a multitude of known and unknown pathways and one should not exclude the possibility of their interference in some of the observations we made (see results and discussion).

1.1.7 DNA Methylation

The main focus of the present work is the synthetic activation of genes by removal of DNA methylation, a covalent binding of a methyl group (CH₃) to a DNA nucleotide.

DNA Methylation in Prokaryotes

DNA methylation has been observed in the majority of prokaryotes²¹⁴. Initially, it has been functionally linked to the restriction-modification (R-M) system, making methylated DNA sequences unrecognizable to restriction enzymes²¹⁵⁻²¹⁷ and thus protecting eukaryotic genomes against their action. However, the roles of DNA methylation in prokaryotes extend to other processes such as cell cycle control, DNA replication, and transcription regulation²¹⁸. Methylation can occur on both adenine and cytosine and is facilitated by a diverse group of methyltransferases (MTases)²¹⁴, which recognize specific motifs. For example, methylation of the first cytosine of the 5'-CCATGG-3' restriction site in *Citrobacter freundii* is facilitated by the MTase CfrBIM, which prevents cleavage by the restriction enzyme CfrBIR^{219,220}. At the same time, methylation of the site also increases the expression CfrBIR and decreases the expression of CfrBIM^{219,220}. 'Orphan' MTases, which do not have a cognate restriction enzyme, are considered involved in other biological processes. Indeed, cell cycle-regulated DNA MTase (CcrMT), which facilitates adenine methylation at 5'-GANTC-3' in *Caulobacter crescentus*²²¹, is involved in cell cycle regulation by insuring that the replication does not start before complete methylation of the chromosome²²². Furthermore, adenine methylation is involved in directing the mismatch repair mechanism into the unmethylated base located in newly synthesized unmethylated strand. Finally, adenine methylation of 5'-GATC-3' in promoters, which is established by DNA adenine methylase (Dam), is thought to regulate gene expression. Indeed, expression of these genes with high adenosine methylation has been increased after Dam removal²²³⁻²²⁵. These examples, illustrating the diversity of DNA methylation in prokaryotes both in establishment and function, are by no means exhaustive. However, since this work

focusses predominantly on eukaryotic processes, prokaryotic methylation will not be further discussed.

DNA Methylation in Eukaryotes

While in eukaryotes methylation can also be observed on adenine, it is by far most common on cytosine, where it gets attached to the fifth carbon of its pyrimidine ring, forming 5-methylcytosine (5-mC). In contrast to the previously described epigenetic marks, DNA methylation, together with its oxidation products 5-hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-fC), and 5-carboxylcytosine (5-caC), is the only one to directly target the DNA molecule by forming a covalent bond. 5-mC most commonly occurs in the context of a cytosine upstream of a guanine (CG or CpG), although it also exists in different contexts usually termed CHG or CHH, where H is either A, T or C. Non-CpG methylation is rare and only occurs in certain cell types such as embryonic stem cells but also neurons or glia cells, and is thought to be involved in brain development and neurological disease²²⁶. Adenine methylation, on the other hand, is still not very well understood, however, research suggests that it is distinct from cytosine methylation, both in terms of involved proteins as well as function²²⁷. Indeed, adenine methylation is, in contrast to cytosine methylation, associated with gene activation rather than silencing^{228,229}. For the rest of the manuscript, we will focus exclusively on CpG methylation. As with other epigenetic marks, addition and removal of the methyl group is a highly organized affair and is involved in physiological processes such as gene regulation, development and differentiation, imprinting, X-chromosome inactivation, and silencing of transposable elements. Methylation aberrations therefore could lead to defects in any of these processes and subsequently to pathologies, like cancer and obesity. In order to understand the different roles of DNA methylation, it is important to be able to quantitatively analyze its spatiotemporal occurrence in the genome at high resolution.

Methods for Assessment of 5-mC Methylation

Since the discovery of chemical modifications on DNA nucleosides via paper chromatography in the 1940s²³⁰, and the establishment of the role of DNA methylation in the regulation of gene expression in the 70s²³¹ and early 80s²³², great strides have been made in the detection of DNA methylation across the genome. Early methods, however, lacked both resolution and throughput. 5-mC can be quantified by reversed-phase high performance liquid chromatography (RP-HPLC)²³³, which later also incorporated mass spectrometry²³⁴. These methods are not able to give information regarding the location of the methylation within the genome, but can be used to precisely quantify the overall ratio of modified to unmodified cytosine, e.g. between tissues or different species. One of the first methods to distinguish between methylated and unmethylated genomic regions takes advantage of isoschizomer pairs of restriction enzymes with differential methylation sensitivity²³⁵. HpaII and MspI both recognize CCGG, however, HpaII can only cut the unmethylated sequence, while MspI can digest both. The digested DNA was subsequently radioactively labeled and separated by thin-layer chromatography (TLC). The major disadvantage of such approach is that it has a low resolution as it is restricted to CCGG sites.

Another way to determine the enrichment of 5-mC in specific genomic regions, is by immunoprecipitating methylated DNA with 5-mC-specific antibodies. The first step of methylated DNA immunoprecipitation (MeDIP) consists in fragmentation of extracted genomic DNA by sonication, followed by denaturation, and antibody precipitation²³⁶. The enriched DNA can subsequently be identified by qPCR, microarray analysis²³⁷ or high-throughput sequencing²³⁸. While this method does not provide methylation profiles at base pair resolution, methylation levels of regions can be quantified. Furthermore, unlike most base pair resolution methods (discussed below), MeDIP does not require harsh treatment of the DNA and, therefore, requires less sample material.

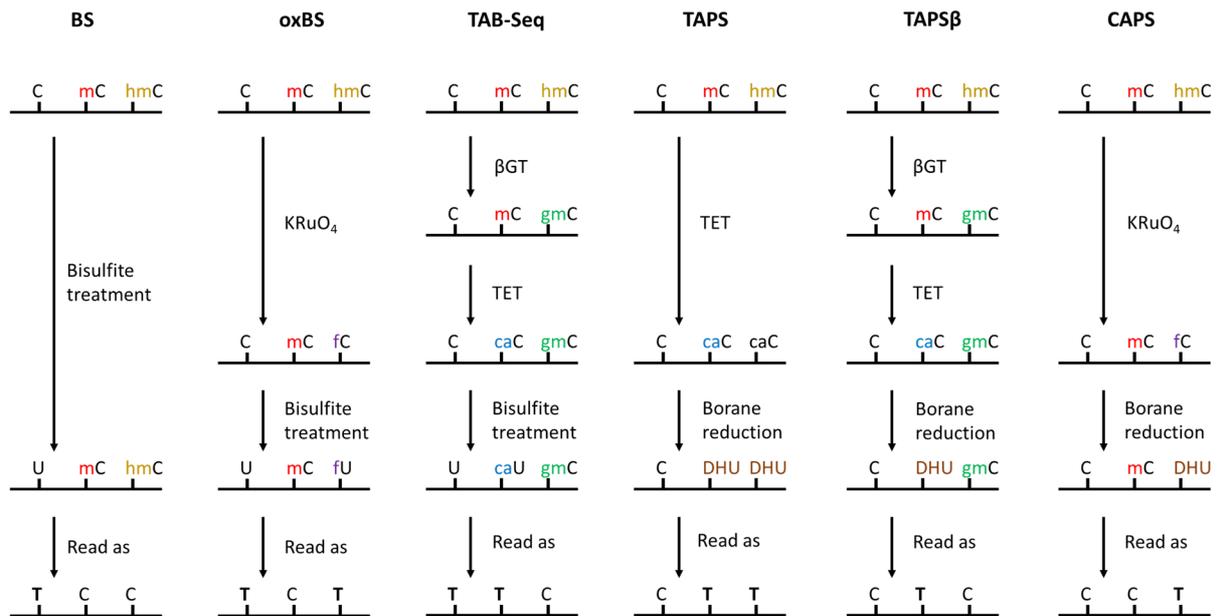


Figure 1.8: Determination of Methylation and Hydroxymethylation by Sequencing. Overview of the various conversion steps and final readouts for traditional BS, oxBS, TAB-Seq, TAPS, TAPSB, and CAPS.

A significant leap forward in detection of DNA methylation at base pair resolution, was the establishment of bisulfite sequencing (BS)²³⁹, which is still the gold standard for the quantification of DNA methylation today (Fig. 1.8; BS, oxBS, and TAB-Seq). Here, DNA is treated with sodium bisulfite that deaminates cytosine residues, but not 5-mC or 5-hmC, into uracil. Following this treatment, methylation levels of individual CpGs at a specific could be detected by PCR and subsequent Sanger sequencing, as uracil will be replaced by a thymine, while 5-mC or 5-hmC will still be read as cytosine. Overlay with the known non-converted sequence of the amplified fragment reveals the levels of CpG methylation at base pair resolution, and, by comparison with other samples, differentially methylated regions (DMRs) can be identified²³⁹. The emergence of next generation sequencing (NGS) techniques made genome-wide approaches for methylation analyses possible. Whole genome bisulfite sequencing (WGBS) represents the sequencing by NGS of a full genome upon bisulfite treatment and represents a powerful tool to cover almost the entire genome^{240,241}. However, roughly 500 million reads of a length of 50-mers are needed to cover about 95% of all CpGs in

a human sample with at least one read, which is in many cases not particularly economical^{242,243}. Moreover, the guidelines suggest a minimum of 15-fold coverage for a given cytosine for the calculation of methylation ratio. Depending on the objective of the project, reduced representation bisulfite sequencing (RRBS) may therefore be more suitable and sufficient²⁴⁴. In contrast to WGBS, where the DNA is usually sonicated, in RRBS, a restriction enzyme, such as MspI, can be utilized to fractionate the DNA. This leads to an enrichment of CpG-rich regions, such as many promoters, but also reduced coverage for CpG-poor regions, including many enhancers²⁴⁵. However, a lower number of total reads is needed (around 10 million), which makes this method more economical. In a given cell, a cytosine can be either methylated or unmethylated on each allele, therefore having a methylation level of 0%, 50%, or 100%. On the population level however, the percentage of methylation on a given cytosine represents the percentage of the methylated form at this cytosine in all sequenced alleles. Since cell populations, especially when harvested from organs with multiple tissues, are often heterogeneous, the obtained methylation readout might therefore be skewed by undesired cell populations in the bulk. This issue can be addressed by single cell bisulfite sequencing (scBS)^{246,247}. However, since the source material is minimal (1 cell), degradation due to the bisulfite treatment becomes an even bigger issue and reduces the coverage even further in such techniques^{246,247}. This can be increased to roughly 18% of all CpGs by switching the adaptor ligation and bisulfite treatment steps in the post-bisulfite adaptor tagging (PBAT) protocol²⁴⁸. Alternatively, one can also consider single-cell locus-specific bisulfite sequencing (SLBS), if coverage of the entire genome is not needed for the study²⁴⁹.

BS is unable to distinguish between 5-mC and 5-hmC, an oxidative intermediate of 5mC observed in TET-dependent demethylation process (described in more detail later in this chapter). To address this issue, variations of the treatment protocol have been developed, such as oxidative bisulfite sequencing (oxBS-Seq)²⁵⁰ and TET-assisted bisulfite sequencing (TAB-

Seq)²⁵¹ (Fig. 1.8). OxBS-Seq involves oxidation of 5-hmC to 5-fC and of 5-mC to 5-hmC with potassium perruthenate (K₂Cr₂O₇). This results in deamination of 5-fC (initially 5-hmC) into uracil upon bisulfite treatment whereas the 5-hmC (initially 5-mC) is not converted. In TAB-Seq, 5-mC is oxidized by a treatment with the TET enzyme, while 5-hmC is protected due to previous treatment with β -glucosyltransferase (β GT). In combination with traditional BS, 5-mC and 5-hmC marks can thus be clearly distinguished. The role of TET enzymes in the removal of methylation, will be further explained later in this chapter.

While still being the ‘gold standard’, bisulfite treatment comes with several caveats. First, the treatment itself is harsh as it produces abasic sites by depyrimidination, therefore leading to strand scission and DNA degradation, which is an issue for samples with low quantity of DNA²⁵². Second, conversion of all unmodified cytosines, not only in the context of a CpG dinucleotide, to thymine leads to severe loss of sequence complexity, which, in turn, may cause poor sequencing quality for certain regions and skewed coverage of the genome²⁵³. The recently developed TET-assisted pyridine borane sequencing (TAPS)²⁵⁴, and the related TAPS β and chemical-assisted pyridine borane sequencing (CAPS)²⁵⁴ (Fig. 1.8), provide a bisulfite-free alternative to BS-based methods. Taking advantage of the oxidizing enzymatic activity of TET on modified cytosine, conversion of only a much smaller portion of the genome is possible, allowing for genome complexity to be maintained. Together with CAPS and TAPS β , which incorporate similar steps as oxBS-Seq and TAB-Seq, respectively, it is also possible to distinguish between 5-mC and 5-hmC in a bisulfite-free manner.

Distribution of DNA Methylation Across the Genome

Methods such as the ones described earlier allow to precisely localize methylated cytosines in the genome. Where does DNA methylation actually occur? And what distinguishes methylated regions from other genomic regions and among one another? Currently, genomic regions can be divided into three major groups, regarding their methylation status: (1) Fully methylated

regions (FMRs) with a methylation frequency of 50% or higher, (2) unmethylated regions (UMRs) with less than 10% methylation frequency, and (3) low methylated regions (LMRs) where the methylation frequency falls in between²⁵⁵ (Fig. 1.9A,C). As a general rule, DNA methylation levels generally seem to anti-correlate with the levels of CpG content in a given region. Indeed, in metazoan genomes, roughly 75% to 90% of CpGs, located mostly at CpG-poor regions, are methylated²⁵⁵⁻²⁵⁷, making FMRs the most common of the three groups. FMRs are mostly found in intergenic regions, repeats, and introns²⁵⁵ (Fig. 1.9F), which is also not surprising since these features make up by far the bulk of the genome. UMRs, on the other hand, are CpG rich, mostly present in the promoter regions of genes, followed by intergenic regions and introns²⁵⁵ (Fig 1.9F). Most of the UMRs correspond to CpG islands (CGIs), which are defined as regions of at least 200bp in length, having a GC content of 50% or more, and an observed:expected ratio of CpG dinucleotides greater than 0.6^{258,259}. CGIs are present in about 50% of gene promoters, and most of them, especially the ones in the promoters of housekeeping genes, remain unmethylated^{255,260,261} (Fig. 1.9E). Conversely, aberrant hypermethylation of CGIs in tumor suppressor gene (TSG) promoter regions is a typical hallmark of many cancers²⁶².

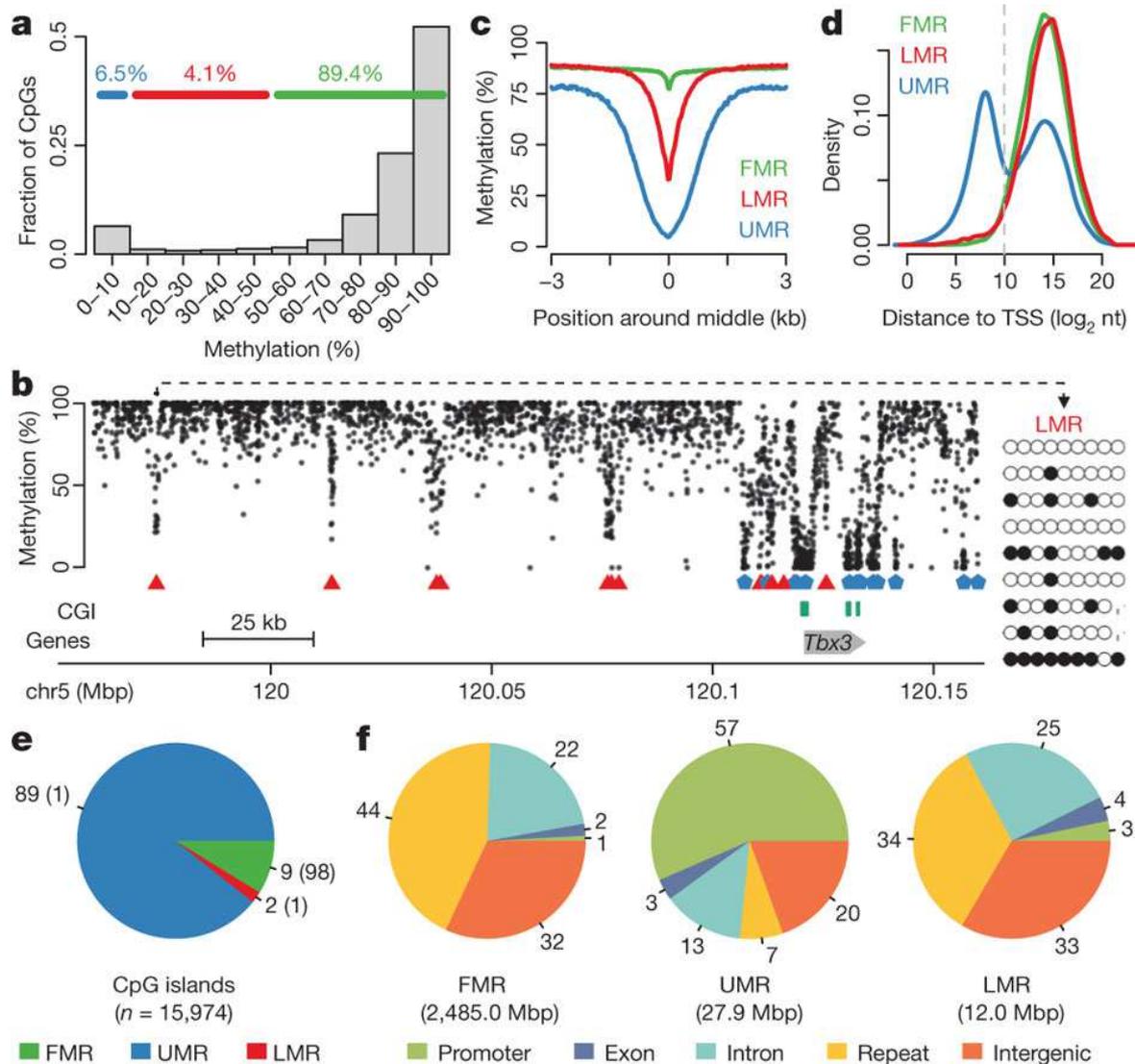


Figure 1.9: Features of the mESC Methylome²⁵⁵. **A)** Frequency of occurrence of CpG methylation percentages across the mESC genome. **B)** LMRs (red triangles) tend to be distal from genes, while UMRs (blue pentagons) tend to be proximal to genes. **C)** Average methylation of FMRs, LMRs, and UMRs. **D)** Distance of each feature from the TSS. **E, F)** Overlap of respective genome features with methylome features.

LMRs, which cover about 4% of the mESC genome, are CpG-poor regions that can be usually found distal from the TSS in intergenic regions, repeats, and introns²⁵⁵ (Fig. 1.9B,D,F). Their sensitivity to DNase I, and enrichment in the enhancer-typical marks H3K4me1 and H3K27ac, support the case for LMRs being distal regulatory regions such as active enhancers²⁵⁵. Furthermore, LMRs are shown to be highly dynamic during differentiation, which may be the result of differential TF binding²⁵⁵. In the end, however, the exact methylation pattern is dependent on the sequence composition, mainly CG content, the cell type, and external *stimuli*

in healthy or diseased cells. Identification of differentially methylated regions (DMRs) between different cell types may, therefore, provide answers regarding underlying physiological or pathological mechanisms active in these cells, and perhaps even identify genes that are aberrantly regulated by epigenetic mechanisms in diseases. These could represent novel medical targets, whose expression could be changed by epigenome editing approaches.

Establishment and Maintenance of DNA Methylation

DNA methylation is established and maintained by DNA methyltransferases (DNMTs)^{263,264}. The catalytically functional DNMTs in humans are DNMT1, DNMT3A, and DNMT3B²⁶⁵. DNMT2 and DNMT3L, which show sequence conservation with the other DNMTs, are also encoded, but have not shown to be active in DNA methylation, although DNMT2 has been shown to facilitate tRNA methylation²⁶⁶ and DNMT3L acts as cofactor for DNMT3A/B²⁶⁷. All metazoan DNMTs, with the exception of DNMT3L, contain a highly conserved C-terminal catalytic domain^{268,269}. Moreover, both DNMT2 and DNMT3L lack an N-terminal regulatory part present in other DNMTs and consisting of several domains for molecular interactions (Fig. 1.10)²⁷⁰. DNMT2's catalytic domain also contains a Cys-Phe-Thr tripeptide (CFT), that is not present in other DNMTs^{271,272}. The N-terminal regulatory domain of DNMT1 consists of multiple subdomains, including 1) DNMT1-associated protein 1 (DMAP1) binding domain that links DNMT1 to histone deacetylation, by interacting with histone deacetylase 2 (HDAC2)²⁷³, 2) replication *foci* targeting sequence (RFTS) for targeting the replication focus through direct binding to ubiquitylated histone H3 and facilitating maintenance of post-replicative methylation^{274,275}, 3) CXXC domain for binding unmethylated DNA²⁷⁶, and 4) two bromo-adjacent homology (BAH) domains. The regulatory domain of both DNMT3A and DNMT3B contains a Pro-Trp-Trp-Pro (PWWP) domain and an ATRX–DNMT3–DNMT3L (ADD) domain, which enable binding to histone H3 harboring trimethylation mark on lysine 36 (H3K36me3) and an unmethylated lysine 4 (H3K4), respectively²⁷⁷.

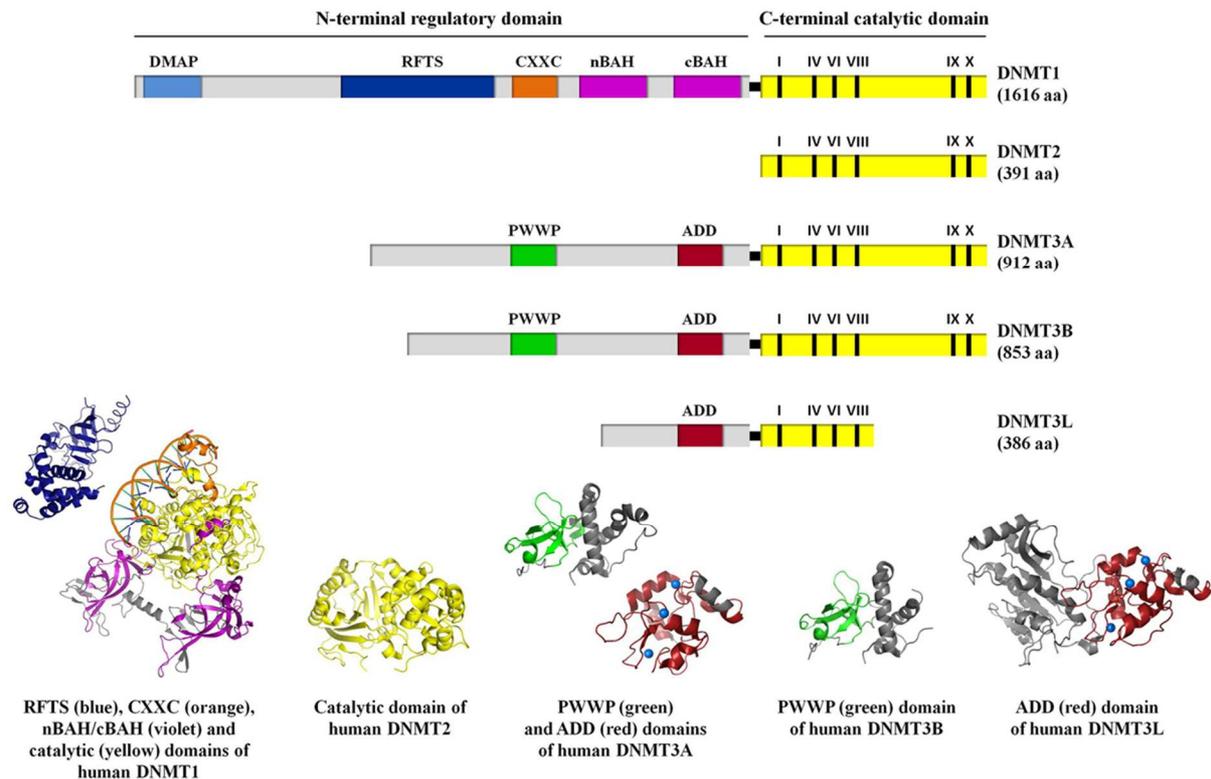


Figure 1.10: Conserved Domain Structures of DNMTs²⁷⁸. All human DNMTs contain a highly conserved C-terminal catalytic domain that is truncated in the case of DNMT3L. With the exception of DNMT2, all DNMTs also contain a regulatory N-terminal domain, consisting of multiple subdomains.

The catalytic reaction of adding a methyl group to the fifth carbon of cytosine (Fig. 1.11A) requires a rotation movement of the target base into close proximity of a cysteine residue within the catalytic motif IV of DNMT1/3A/3B^{279,280}. This cytosine interacts with the sixth carbon of the pyrimidine ring in a nucleophilic manner, opening up the fifth carbon for a methyl transfer from S-adenosylmethionine (SAM) as methyl donor^{279,280}.

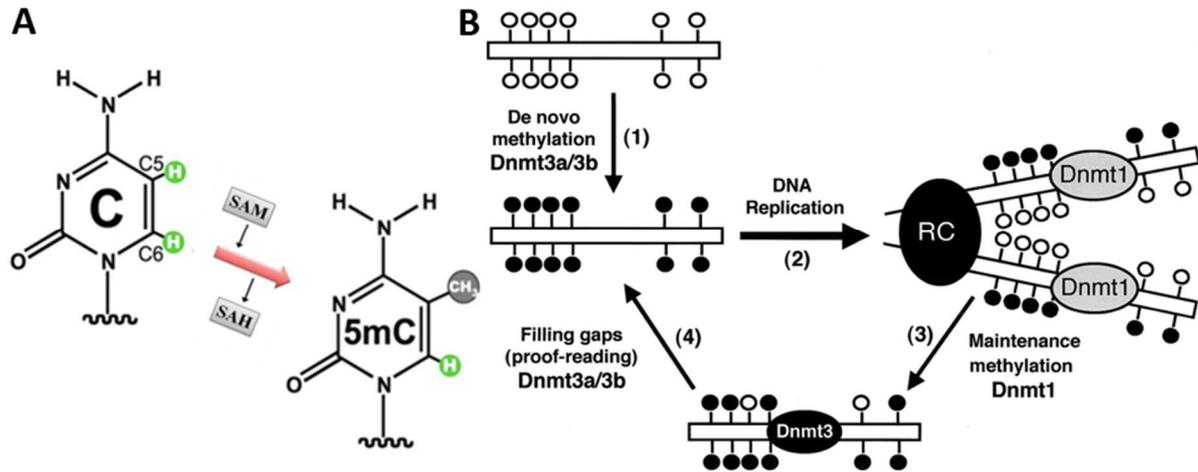


Figure 1.11: The DNA Methylation Reaction (adapted from ²⁷⁸ and ²⁸¹). **A)** Methylation reaction of cytosine by DNMTs using S-adenosylmethionine (SAM) as a methyl group donor, resulting in 5-methylcytosine (5-mC) and S-adenosyl homocysteine (SAH) as cofactor product. **B)** Classical model of *de novo* methylation by DNMT3 and methylation maintenance by DNMT1.

De novo methylation is established by DNMT3A/B^{282,283} (Fig. 1.11B), and, although they do not target specific DNA motifs, their binding is targeted to unmethylated H3K4 residues by their ADD domain²⁸⁴. This leaves regions with enrichment for H3K4me1 and H3K4me3, which are characteristic for enhancers and promoters, respectively, protected from *de novo* methylation. DNMT1, on the other hand, is responsible for methylation maintenance²⁷⁴ (Fig. 1.11B). This is facilitated by its RFTS domain, which targets DNMT1 to replication *foci*, where it subsequently methylates the hemimethylated DNA strands²⁷⁴. The responsibility of DNMT1 maintenance and DNMT3 for establishment of DNA methylation is, however, not strict. Inactivation models of DNMT3 isotypes have shown progressive loss of methylation at certain loci, implying a contributive role of DNMT3 also in maintenance²⁸¹. Similarly, some *de novo* activity of DNMT1 was reported *in vitro* as well as in DNMT3s KO cells *in vivo*²⁸⁵⁻²⁹⁰. Moreover, the affinity of DNMT1 to hemimethylated DNA makes it active at hemimethylated sites outside the replication fork. This led to a model that proposes that DNMT3s add a methyl group on a CpG *de novo* on one strand while DNMT1 adds a methyl group on the corresponding CpG on the other strand (Fig. 1.11B)²⁹¹. This methylation pattern is then maintained during

replication by DNMT1 but could also be changed by the activity of demethylases and transcription factors. However, this model does not imply that DNMT1 is needed for the replication-independent addition of methyl groups on hemimethylated DNA, as DNMT3s can perfectly fulfill this role because they have a similar affinity to unmethylated and hemimethylated DNA.

DNMT activity is regulated on multiple levels, including transcriptionally by alternative splicing, post-translational modifications, and interaction with various cofactors. Splicing variants are most prominently found in DNMT3B and vary in enzymatic activity^{292,293}. In cancer, more than 20 transcripts of DNMT3B have been identified, some of which missing parts of the catalytic domain and some retaining intron sequences²⁹³. These aberrant isoforms, e.g. DNMT3B7, have been found to affect the methylation of certain CGIs and the expression of the corresponding genes²⁹³. As example for post-translational regulation, DNMT1 can be methylated at lysine residues 142 and 1094^{294,295}, thus marking the protein for degradation. However, phosphorylation of serine 143 prevents the neighboring methylation, and, therefore, protects from degradation²⁹⁶. The aforementioned DNMT3L is an important cofactor for the activity of DNMT3s as it interacts with both DNMT3A and 3B^{267,297}. Moreover, it was shown that two copies of each of DNMT3A and DNMT3L form a complex with increased DNA affinity²⁹⁸. Finally, although DNMT1 can, à priori, bind hemimethylated DNA, its interaction with protein E3 ubiquitin-protein ligase UHRF1 (Ubiquitin-like, containing PHD and RING finger domains, 1) is essential for it to perform its maintenance function^{299,300}. UHRF1 shows, via its SET and RING Finger-associated (SRA) Domain, high affinity for hemimethylated DNA and renders the modified cytosine more recognizable for DNMT1^{301,302}. Furthermore, UHRF1 mutants show global hypomethylation²⁹⁹. In addition, DNMT1 can also interact with proliferating cell nuclear antigen (PCNA), which facilitates recruitment to replication *foci*³⁰³. This interaction was first deemed necessary, however, more recent studies have shown that

global methylation levels in post-replicative cells with PCNA-binding domain mutants is still maintained^{304,305}.

Although an increasing number of factors regulating DNMT activity, as well as their interaction with chromatin, is being identified and understood, several biological mechanisms, such as the adaptation to external and internal *stimuli*, remain to be elucidated.

DNA Demethylation

Unlike other epigenetic modifications, DNA methylation is considered to have a long term stability. In fact, until recently, the presence of active DNA demethylation mechanisms was not proven. However, it is now established that DNA methylation is reversible. In general, DNA methylation can be removed by two distinct mechanisms: active and passive (Fig. 1.12A). Passive demethylation implicates the absence or inhibition of the methylation maintenance machinery during replication of DNA. Assuming a total absence of DNA methylation maintenance, each round of replication transforms one methylated strand into a hemimethylated double strand of DNA (50% loss), producing a progressive dilution of methylation over consecutive replication cycles.

Active demethylation involves ten-eleven translocation (TET) proteins that catalyze the sequential oxidation from 5-mC to 5-hmC³⁰⁶⁻³⁰⁸, 5-fC³⁰⁹, and 5-caC^{309,310}. Although the first step during TET-mediated demethylation is active involving the enzymatic activity of TETs, the following step can be either passive, as the oxidation products of 5-mC are not recognized by DNMT1 and therefore not maintained during replication³¹¹, or active, by thymine DNA glycosylase (TDG)-mediated excision of 5fC and 5caC and subsequent base excision repair (BER)^{310,312,313}.

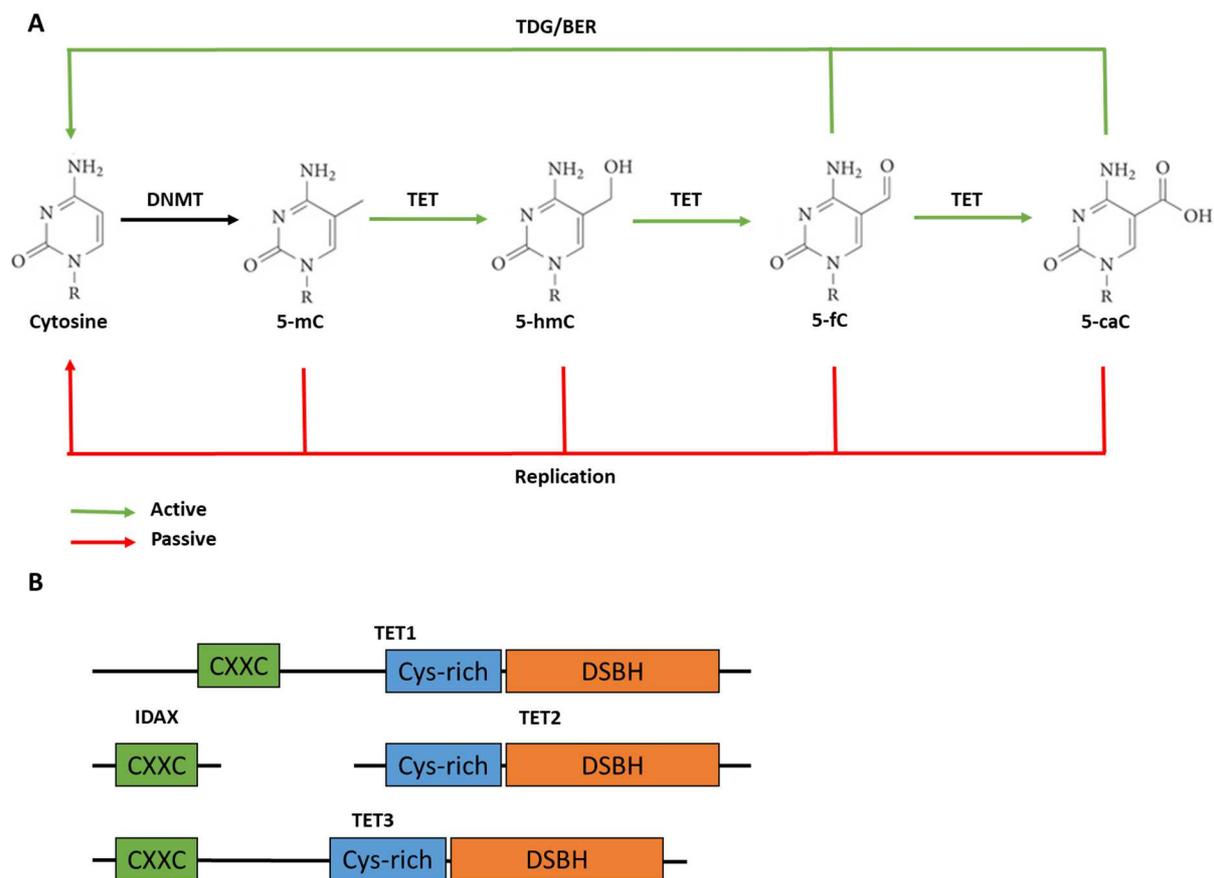


Figure 1.12: Active and Passive DNA Demethylation. A) Sequential oxidation from 5-mC to 5-hmC, 5-fC, and 5-caC by TET proteins, which is followed by either active removal via TDG/BER or passive dilution via replication. If the methylation maintenance machinery is inhibited, dilution may also occur directly from 5-mC. B) The different human TET proteins and their domains.

There are three main TET proteins in humans, TET1, TET2, and TET3, all of which oxidize 5-mC using iron(II) (Fe(II)) as a cofactor and oxygen and α -ketoglutarate as substrates, respectively³¹⁴. All TET proteins contain a double-stranded β -helix (DSBH) domain and a cysteine-rich domain (Cys-rich) (Fig. 1.12B). The former catalyzes the reaction between 5-mC, α -KG, and Fe(II), while the latter stabilizes the complex³¹⁴. Upon interaction of TET proteins with DNA, both domains will come together forming a catalytic pocket. The methyl group is inserted into this pocket, however it does not have direct contact with the protein, thus allowing other oxidized forms of cytosine to be substrates of TET³¹⁵. Two pathways could be used to replace the oxidized forms by an unmodified cytosine. One pathway involves excision of 5-fC

or 5-caC by TDG and repair of the single-strand break by BER³¹³. The other pathway depends on passive dilution during replication, which can be explained for by the lower affinity of the DNMT1 cofactor UHRF1 to hemihydroxymethylated CpG dyads^{316,317}.

5-mC and all its oxidation products are suitable substrates for TET proteins. However, affinity towards the different substrates may vary. For example, human TET1 and TET2 catalyze the reaction from 5-mC to 5-hmC faster than the subsequent conversions from 5-hmC to 5-fC and 5-fC to 5-caC^{309,318}. Also the genomic context may play a role, as TET2 prefers 5-mC substrates in a CpG context rather than CpC or CpA, although these reactions are also catalyzed³¹⁵. At the CpG dyad there are 25 possible substrates depending on the modification status of either strand, all of which, with the exception of completely unmethylated, hemicarboxylated, and fully carboxylated, are substrates for TET^{313,319}. Except for the aforementioned preference of 5-mC, there is no difference in affinity at the dyad in regards to the modification status of complementary strand³¹⁹. Substrate and cofactor availability also influence reaction kinetics of TETs. For instance, overexpression of isocitrate dehydrogenase 1 (IDH1) and IDH2, which produce α -KG, has been shown to increase global 5-hmC levels³²⁰, while downregulation of IDH2, as is the case in certain cancers, leads to a decrease in global 5-hmC levels³²¹. 5-hmC levels also correlate with cellular Fe(II) levels. For example, it has been shown that deprivation of cellular iron with a specific chelator stops 5-hmC production, while addition of iron increases 5-hmC levels³²². In addition, both TET and TDG activity can be regulated on post-transcriptional and post-translational levels by a vast array of microRNAs (miRNAs) as well as ubiquitylation, phosphorylation, and GlcNAcylation³²³⁻³²⁵. TET1 and TET3 contain a CXXC domain³¹⁴, while TET2 directly interacts with IDAX a protein containing CXXC domain^{314,326}. Interestingly, IDAX was once part of ancestral TET2, however, it became separated by chromosomal gene inversion³²⁶. The CXXC domain, through its preference to unmethylated cytosines is probably responsible for the enrichment of TET proteins at these sites leading to

their protection from methylation³²⁷. However, also truncated isoforms of TET1, which lack the CXXC domain and can be found in mice, localize, with lower affinity, to these regions³²⁸. This suggests that CXXC plays only a partial role in TET localization. In fact, evidence suggests that recruitment of TET to specific genomic locations is facilitated by a multitude of different DNA-binding interaction partners. For instance, alongside the process of differentiation of 3T3-L1 pre-adipocytes to mature adipocytes, both TET3 enrichment and DNA demethylation were observed around regions bound by CTCF and peroxisome proliferator-activated receptor- γ (PPAR γ)^{329,330}, which leads to adipogenic transcriptional enhancer hydroxymethylation, suggesting that these factors could be involved in the recruitment of TETs. The RE1-silencing transcription factor (REST) has also been shown to recruit TET3 to its binding sites³³¹, as was TET2 recruitment by the pioneer transcription factor (PF) PU.1³³². Finally, both TET1 and TET2 colocalize with the pluripotency factor NANOG, while depletion of the latter reduces TET binding³³³. Taken together, these findings indicate a strong relationship between TET localization and the binding of certain TFs. The exact extent of TF-TET interactions is still unknown, but evidence suggests that the findings to date are just the tip of the iceberg. Furthermore, cell type-specific interactions between TFs and TET proteins as well as DNMTs may shine a light on the establishment of cell type-specific DNA methylation and chromatin landscapes.

Roles of DNA Methylation in Cell Physiology

DNA methylation is involved in a multitude of physiological and pathological processes (Fig. 1.13). One of the most prominent functions of DNA methylation is the silencing of genes. Although there is a strong correlation between DNA methylation at regulatory elements and gene silencing²⁶¹, the exact mechanisms by which the two are linked is still not fully understood. Initially, it was thought that methylated DNA is refractory to TF binding and can only recruit methyl-CpG-binding domain (MBD) proteins^{334,335}, which in turn interact with HDACs and

nucleosome remodeling complexes, ultimately leading to less accessible chromatin^{336,337}. Furthermore, interactions between DNMT3s, and H3K9 methyltransferases^{338,339}, HDACs^{340,341}, as well as remodelers such as lymphocyte-specific helicase (LSH)^{338,342}, which form complexes, result in the establishment of heterochromatin. This initial hypothesis seems to be only part of the answer, since recent studies show that only a subset of TFs seem to be methylation sensitive, while others are either not sensitive or have a preference to methylated sites. A recent study using SELEX shows that only around 22% of all tested human TFs actually are methylation sensitive³⁴³. Moreover, certain TFs that can bind to methylated DNA were shown to initiate demethylation of the region around their binding site^{255,344}. A renowned example for genes silenced by CGI promoter methylation are germline-specific genes³⁴⁵, which can be reactivated in cell lines by DNMT mutations^{345,346}, knockdown³⁴⁷ or knockout³⁴⁸.

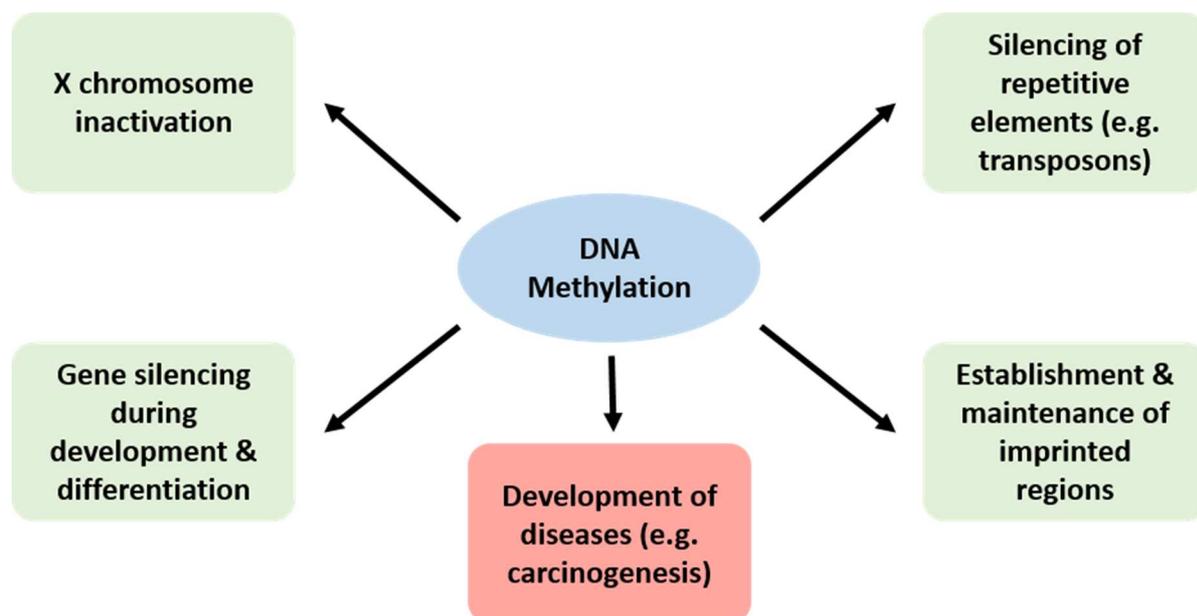


Figure 1.13: The Roles of DNA Methylation. DNA methylation is involved in a series of physiological processes. Therefore, its aberrations could also contribute to the development of certain diseases.

Due to spontaneous deamination from 5-mC to T³⁴⁹ and the potential of DNMTs to cause DNA lesions³⁵⁰, DNA methylation could be considered genome-destabilizing or even cytotoxic. And

indeed, in higher organisms the observed CpG content is only about a fifth of the expected value^{351,352}. Oddly enough though, most CpGs are still methylated despite the mutagenic potential of 5-mC²⁵⁵⁻²⁵⁷. Another way to see it, is that DNA methylation is a driving power of evolution as it accelerates the mutational capacity of the genome. Indeed, when looking at genetic differences between the human and chimpanzee genomes, one can observe a much higher variability in CGs that are methylated in one of the genomes. Moreover, it has been found that DNA methylation actually contributes to genome stability, since not only genes can be silenced by it, but also retrotransposons^{353,354}, self-amplifying genetic elements, which get reinserted into other sites within the genome. Knockout of DNMT1 in mouse embryos, for example, results in reactivation of intracisternal A particle (IAP) retrotransposon expression and lethal developmental defects³⁵⁵. A special case is the newly identified DNMT3C^{354,356}, which protects from IAP retrotransposition in mouse testis. Knockout of DNMT3C, which has so far only been discovered in mice, results in smaller testis and infertility³⁵⁴. Finally, DNA methylation could contribute to genomic stability simply due to its role in chromatin compaction that protects DNA from insults.

Two further physiological examples for the involvement of DNA methylation-mediated silencing in biological functions are gene imprinting and X-chromosome inactivation. Gene imprinting results in stable mono-allelic expression of genes that is already established in the germ lines. In most cases, imprinting depends on the parental origin of the allele and is controlled by differential methylation of regulatory regions called imprinting control regions (ICRs). In other words, the ICRs are methylated specifically either in the maternal (maternal imprinting) or in the paternal (paternal imprinting) germline, thus resulting in differential expression of the genes regulated by these ICRs. 20 ICRs which are able to retain methylation during development, have been identified in the maternal germline, while only three paternally imprinted ICRs, all of which are intergenic, have been identified³⁵⁷.

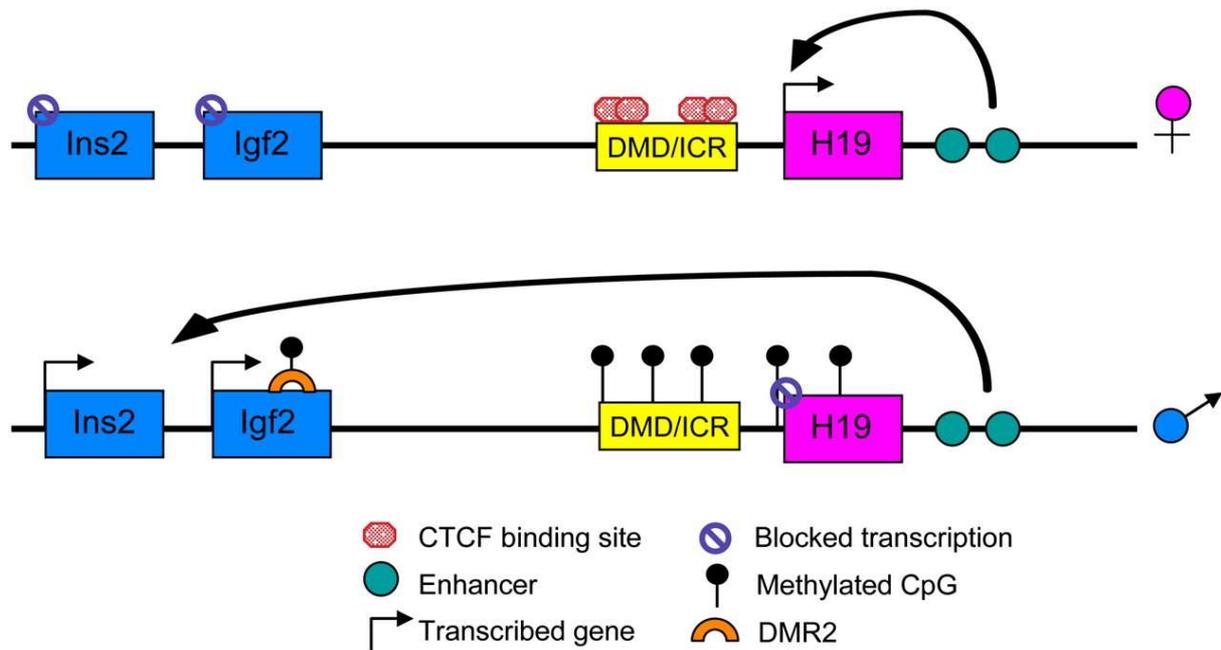


Figure 1.14: Regulation of the *Igf2/H19* Locus via Imprinting³⁵⁸. CTCF is able to bind the differentially methylated domain (DMD)/ICR on the maternal chromosome, setting a boundary which enables *H19*, but not *Igf2*, expression. On the paternal allele, the ICR is methylated, thus prohibiting CTCF binding, which allows loop formation between the enhancers and *Igf2*, ultimately leading to its expression.

This allele specific methylation regulates gene expression in multiple ways. An excellent example of imprinting is the epigenetic control of the *insulin-like growth factor 2 (Igf2)/H19* gene cluster (Fig. 1.14). The genes are reciprocally imprinted, leading to *Igf2* expression exclusively from the paternal allele while *H19* is expressed from the maternal allele. Key to this imprinted control are two regulatory regions, one pair of shared enhancers downstream of *H19* as well as an ICR in between *Igf2* and *H19*. The ICR contains a CTCF-mediated boundary site that is methylated in the paternal allele, thus preventing CTCF binding and disruption of the boundary, ultimately leading to loop formation between the enhancers and *Igf2*, and the expression of the latter. Vice versa, on the maternal allele, the ICR is unmethylated, thus allowing CTCF binding and loop formation with *H19* (summarized in ³⁵⁹ and ³⁶⁰) while insulating *Igf2* and *Ins2*. ICRs are established during oocyte growth, when global methylation is re-established^{267,361-364}. ICRs, unlike other regulatory regions, contain a large number of zinc-

finger protein 57 (ZFP57) motifs^{365,366}. ZFP57 contains a KRAB domain, which in turn recruits DNMT1, UHRF1, and other silencing factors³⁶⁵. This ensures selective DNMT activity and, therefore, repression of the maternally imprinted genes throughout development.

X-chromosome inactivation (XI) is the silencing of one of the two X-chromosomes in females, which allows equivalent expression of X-related genes between males and females. XI is initiated by coating of the X-chromosome with the long non-coding RNA (lncRNA) X-inactive specific transcript (XIST), which is produced by the very same chromosome it is going to silence³⁶⁷. Gene silencing is subsequently established by the XIST-mediated recruitment of various factors involving histone modifiers and nucleosome remodelers³⁶⁸. For example, Drosophila Split ends homolog (SPEN), is recruited to XIST via its multiple RNA binding domains³⁶⁹, and, in turn, recruits HDAC3 and other transcriptional corepressors³⁷⁰. Other complexes recruited are polycomb repressive complex 1 (PRC1) and PRC2^{369,370}, which are responsible for establishing repressive H3K27me3 marks. This is followed by the recruitment of DNMT3B and the acquisition of DNA methylation that is thought as a way to further maintain the repression of genes on the inactivated X chromosome as H3K27me3 is more easily reversible than DNA methylation. How DNMT3B is recruited to the inactivated X remains elusive³⁷¹.

DNA methylation in Development and Differentiation

Subsequent to its role in development and the determination of cell fate, DNA methylation is also involved in the formation and function of tissues, organs and whole organisms. Therefore, both in the germline and embryogenesis, it is extremely important to establish a correct pattern of DNA methylation that will contribute grandly to the determination of gene expression profiles of the gametes and the embryo. In order to set up the correct pattern of methylation, the genome first undergoes genome-wide demethylation at the early stages of both developmental stages. In the fertilized zygote, demethylation is first actively mediated through TET3, and is

accompanied with a rise in 5-hmC levels^{372,373}. The second phase in early embryonic demethylation occurs passively, with 5-mC levels reaching their low point at the pre-implantation blastocyst stage³⁷⁴. This phase occurs due to the absence of DNMT1 and the fact that UHRF1 and DNMT1o, the DNMT1 isoform expressed in oocytes, are largely excluded from the nucleus in the very early embryo^{374,375}. Subsequently, post-implantation, methylation patterns are *de novo* established by DNMT3A, DNMT3B, and DNMT3L³⁷⁶, although in humans the latter is not crucial for this step³⁷⁷. How DNA methylation pattern is reestablished post-implantation depends on several factors including sequence composition and transcription factor availability, and is still a matter of intense research.

Certain cells within the epiblast, which is one of the two layers of the inner cell mass of the post-implantation blastocyst, are reprogrammed to become primordial germ cells (PGCs). At this stage, a second global demethylation event takes place, which, in contrast to the one in the fertilized egg, starts with a passive phase and later switches to active demethylation mediated by TET1 and TET2³⁷⁸⁻³⁸¹. While both events occur globally, they do not cause complete demethylation of the genome. In fact, in both mice and humans about 20% of parental methylation is retained in the inner cells mass of the pre-implantation blastocyst^{382,383}. These regions include among others the aforementioned ICRs^{366,384,385}, where methylation is retained thanks to the activity of ZFP57³⁶⁶. While these ICRs remain stably methylated, the remaining methylated regions, which cover also by far a larger portion of the genome, may get reprogrammed at later stages of development^{383,386,387}, or belong to retrotransposons^{383,388}. On the other hand, in PGCs, the methylation level drops to around 6-8% and occurs mostly at retrotransposons³⁸². At this point, also ICRs get demethylated in a TET-mediated manner³⁸⁹.

Remethylation of PGCs on their way to mature germ cells, results in methylation levels of roughly 80% in sperm, similar in pattern to somatic cells, and about 50% in oocytes^{376,382,390}. The comparatively low methylation levels in oocytes are most likely due to the retention of

DNMT1o and UHRF1 to the cytoplasm by the regulator Stella³⁹¹. When Stella is experimentally depleted, UHRF1 accumulates in the nucleus, which in turn leads to translocation of DNMT1o and an increase of methylation by roughly twofold, accompanied by female infertility³⁹¹. Interestingly, this finding may be an indication of *de novo* methylation capabilities of DNMT1 in the oocyte. Most of the *de novo* methylation in germ cells, however, is mediated by DNMT3A in both humans and mice^{376,377}, as well as DNMT3C in sperm^{354,356} of mice and rats. Furthermore, DNMT3A activity in mice germ cells needs to be mediated by DNMT3L³⁷⁶, while in humans, DNMT3L is not expressed at this stage³⁷⁷.

While methylation at the stage of the epiblast reaches roughly the same levels of adult somatic tissues, extra-embryonic tissue, such as visceral endoderm and the extra-embryonic endoderm, and later also the placenta remain in a hypomethylated state relative to the epiblast³⁹²⁻³⁹⁴. The hypomethylated state is probably a consequence of a lower expression of DNMT3 enzymes³⁹². Interestingly, embryonic cells in the freshly methylated epiblast are still pluripotent, suggesting maintenance of the methylation patterns of pluripotent stem cells (PSCs) throughout life. While the acquired methylation pattern of the epiblast stage is largely maintained, adult tissues show focal but distinct differences among each other, e.g. at enhancers, resulting in tissue-specific DMRs^{386,394-396}. Pluripotency TFs such as OCT4 establish and maintain regions of accessible chromatin, or nucleosome-depleted regions (NDRs)³⁹⁷. NDRs can be found in the distal or proximal regulatory regions of *OCT4* and *NANOG*, another pluripotency gene, respectively, as well as other target genes³⁹⁷. During differentiation, silencing of these genes coincides with hypermethylation and nucleosome enrichment. In case of OCT4, silencing is initiated by a repressor, and subsequent G9A-mediated H3K9 methylation, which in turn recruits heterochromatin protein 1 (HP1)³⁹⁸. Hypermethylation occurs as the last step and can be established by both DNMT3A and DNMT3B^{398,399}.

Role of DNA Methylation in Cell Plasticity

Throughout development, cells acquire a tissue-specific phenotype, usually without changes in genotype. While the established cell fate is relatively stable, cells of a certain type maintain a certain degree of plasticity, which allows them to transdifferentiate into closely-related cell types. An example for that would be the transdifferentiation from glucagon-producing α pancreatic cells to insulin-producing cells following severe β cell ablation⁴⁰⁰. Another example for tissue plasticity following external *stimuli* are thermogenic adipocytes. While white adipose tissue (WAT) represent the vast majority of adipocytes and are responsible for energy storage, brown adipose tissue (BAT) expends energy rather than storing it and thereby contributes to non-shivering thermogenesis^{401,402}. Active BAT is associated with a healthy phenotype, correlating with lower body mass indices (BMI) and a lower risk for diabetes⁴⁰³. Characteristic for BAT is a high enrichment in mitochondria and high expression of uncoupling protein 1 (UCP1) present on the inner mitochondrial membrane. UCP1 is responsible for the uncoupling of oxidative respiration from ATP synthesis, and, therefore, energy expenditure in form of heat. Interestingly, after being exposed to cold temperatures, ‘browning’ occurs within WAT, generating cells phenotypically similar to brown adipocytes and called “beige” or “brite”⁴⁰⁴⁻⁴⁰⁶. Just as BAT, these beige adipocytes also express UCP1⁴⁰⁴. The exact origin of beige adipocytes remains elusive, with both transdifferentiation from mature white adipocytes⁴⁰⁷, or differentiation from precursors^{408,409} being possible explanations. A multitude of TFs and effectors that contribute to the ‘brown’ phenotype, have been identified. One of which is PR domain containing 16 (PRDM16), which gets recruited to the regulatory regions of BAT marker genes, such as the aforementioned *UCP1* and *cell death-inducing DNA fragmentation factor alpha-like effector A (Cidea)*^{410,411}. This recruitment is in turn facilitated by other TFs, including proliferator activated receptor gamma (PPAR γ)⁴¹⁰, and PPAR γ coactivator 1-alpha (PGC1 α)⁴¹¹. While the former is expressed in both WAT and BAT, its binding is cell type specific⁴¹², with

UCP1, *PRDM16*, and *PPAR γ* itself as its targets in BAT. Finally, another striking example of cell type plasticity is the capacity of several cell types to dedifferentiate into embryonic stem cell like cells called induced pluripotent cells (iPS) through the ectopic expression of key TFs^{179,413}. In all the aforementioned examples, the role of epigenetic modifications is key as these cells change phenotype without any change in the DNA sequence. For instance, several papers indicated the importance of DNA methylation changes in iPS generation exemplified by an increase in the reprogramming efficiency due to 5-azacytidine treatment⁴¹⁴. While genome wide methylation analysis establishing DMRs between white, brown, and beige adipocytes have yet to be conducted, several studies have shown a link between the methylation status of BAT marker genes, their expression, and the ultimate phenotype. Indeed, it was shown that cold exposure leads to demethylation and chromatin remodeling at the *UCP1* enhancer inducing its expression⁴¹⁵. It has also been shown that receptor-interacting protein 140 (RIP140) is responsible for recruitment and assembly of DNMTs and HMTs at the *UCP1* enhancer, thus maintaining its repression⁴¹⁶. RIP140-mutant cells, on the other hand, exhibit higher energy expenditure and *UCP1* expression⁴¹⁷. In mice lacking Janus kinase Tyk2, hypermethylation and a decrease H3K4me3 at the promoters of both *UCP1* and *Cidea*, together with decreased expression and thermogenesis was reported⁴¹⁸. Furthermore, elevated non-CpG methylation of the *PGC1 α* promoter, accompanied with lower expression and number of mitochondria, has been observed in skeletal muscle cells of diabetes patients⁴¹⁹. Furthermore, RRBS has identified 31 differentially methylated promoters between WAT and BAT, five of them belonging to members of the Hox gene family, whose expression anti-correlated with the methylation level⁴²⁰. Since there is a bias for CG-rich regions, which are usually present at promoters, RRBS does not cover other regulatory regions, such as enhancers. This issue can be addressed by WGBS. Taken together, these findings indicate a crucial role for DNA methylation in adipogenesis and the robust maintenance of the thermogenic program. However, only locus-specific methylation analysis has been performed, while genome wide approaches could

provide us with tissue specific DMRs, shedding light on possibly affected genes that may have been missed in previous studies.

DNA Methylation in Cancer

As DNA methylation plays important roles in the regulation of physiological processes, it is no surprise that divergent DNA methylation is a hallmark of plethora of pathologies and diseases. For instance, aberrant distribution of DNA methylation is also an important hallmark in many cancers. In particular, two major aberrations of DNA methylation could be observed in cancer: (1) global hypomethylation, and (2) local hypermethylation at CGIs. Cancer-related DNA hypomethylation occurs mostly at repetitive sequences, introns, and gene bodies⁴²¹, resulting in chromosomal instability due to mitotic recombination events, and reactivation of transposable elements, and ultimately leading to deletions, chromosome aberrations and aneuploidy⁴²²⁻⁴²⁴. Also imprinted genes may get reactivated, such as *insulin-like growth factor 2 (Igf2)*, which is a risk factor in colorectal cancer⁴²⁵. Furthermore, the degree of loss of DNA methylation seems to follow the progression of the disease, as a mouse model for skin cancer has shown progressive loss of global 5-mC from initially clonal expansion to a malignant carcinoma⁴²⁶. DNA hypermethylation, on the other hand, can be observed at CGIs including those at the promoters of tumor suppressor genes (TSGs). These genes are responsible for processes such as proliferation regulation, cell cycle control, apoptosis, and angiogenesis, and their inactivation leads to the acquisition of several cancer hallmarks such as uncontrolled proliferation with no contact inhibition, apoptosis evasion and *de novo* angiogenesis^{427,428}. Moreover, these differences in methylation levels and distribution can be used to distinguish healthy tissue from cancer tissue^{429,430}, but also as a biomarker in bodily fluids, such as serum or urine, for clinical diagnostic⁴³¹⁻⁴³³.

An excellent case for the involvement of DNA methylation in cancer is hepatocellular carcinoma (HCC)^{434,435}. Prominent examples for TSGs with hypermethylated promoter regions

in HCC are *p16^{INK4a}*, *p15*, *RASSF1A*, and *GSTP1*, with observed methylation frequencies of 16-83%, 42-47%, 59-75%, and 41-76%, respectively, in comparison to 0-10% in adjacent normal tissue⁴³⁵⁻⁴³⁸. Hepatitis B virus (HBV) infection is a major risk factor for HCC and it might as well be involved in DNA hypermethylation events observed in HCC. Indeed, it has been found that HBV encoded protein X (HBx) that is expressed upon HBV genomic insertion, leads to an increase in the expression of DNMTs via its transactivation domain⁴³⁹⁻⁴⁴⁵. DNA hypomethylation and hypermethylation of specific TSG promoters may also occur at different stages of cancer development and can be either a cause or consequence of cancer progression. For instance, *p16^{INK4a}*, *p15*, *RASSF1A*, and *GSTP1* hypermethylation is usually observed at all stages of HCC, while *SYK* and *CHFR* methylation only occurs at later stages⁴⁴⁶⁻⁴⁴⁸. A recent study in acute myeloid leukemia (AML) has also shown that hypomethylation at the initial stages is due to a mutation in DNMT3A, which occurs in about 25% of AML⁴⁴⁹. Hypermethylation of CGIs, on the other hand, is observed in the later stages of AML with wild-type DNMT3A, most likely as a consequence of cancer progression, since this increase in methylation can also be achieved in hematopoietic stem cells after cytokine-mediated expansion⁴⁴⁹. These stage-related methylation changes could therefore be used as classification biomarkers.

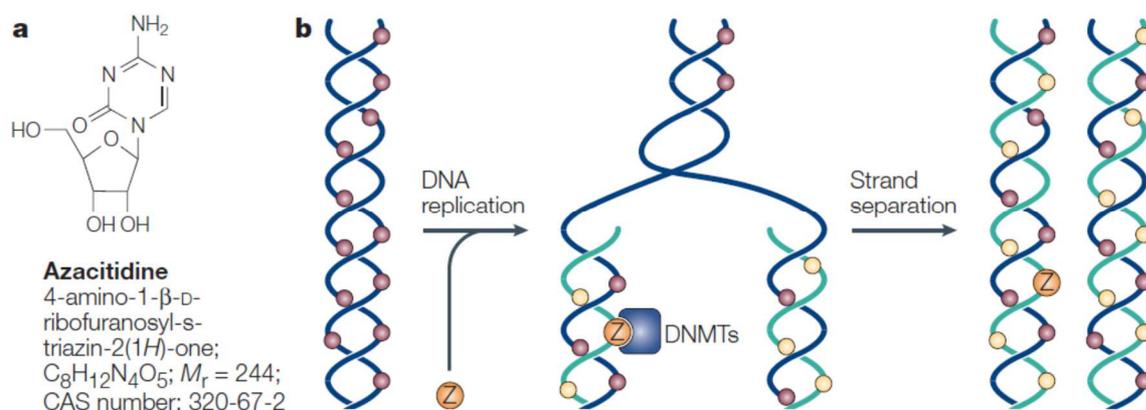


Figure 1.15: Mode of action of 5-Azacytidine⁴⁵⁰. A) Chemical formula of 5-azacytidine (also known as azacytidine or Vidaza) B) 5-Aza integrates newly synthesized DNA leading to inhibition of DNMTs and loss of maintenance of DNA methylation following DNA replication.

Alongside HDAC inhibitors, which can cause cell cycle arrest and apoptosis⁴⁵¹, approved epigenetic therapies against cancer already exist in the form of 5-azacytidine (5-aza) and 5-aza-2'-deoxycytidine, which are DNMT-inhibiting cytidine analogs that get incorporated into the genome upon DNA replication⁴⁵²⁻⁴⁵⁴ (Fig. 1.15). This causes progressive DNA demethylation in proliferating cells, which makes this treatment somewhat selective for cancer cell, although the exact affected cell type cannot be controlled. However, due to the random incorporation of 5-aza into the genome, demethylation occurs indiscriminately of the region, potentially hitting undesired targets. Furthermore, 5-aza has not been shown to be effective in the treatment of solid tumors⁴⁵². Consequently, there is demand for more targeted approaches, ideally only affecting TSGs.

DNA Methylation in Obesity and Diabetes

Indications of the impact of DNA methylation on two related metabolic diseases, obesity and diabetes, have emerged in recent years. As with the aforementioned role of DNA methylation in regulating lineage commitment of adipocytes for thermogenesis or energy storage, abnormal methylation has been observed concordant with altered adipokine levels and expression of key lineage markers. The levels of Leptin, a key adipokine responsible for maintaining energy homeostasis and body weight, are highly increased in patients with obesity⁴⁵⁵. The expression of leptin negatively correlates with the methylation levels at its promoter⁴⁵⁶. High expression in adipocytes correlates with hypomethylation of the promoter while low expression in non-adipocytes coincides with its hypermethylation. A more direct evidence of the importance of DNA methylation in the regulation of Leptin, came from a study showing that its expression can be increased in non-adipocytes by inhibiting DNMTs⁴⁵⁶. Interestingly, high-fat diets in rodents, increase, on a long-term scale, DNMT recruitment and methylation levels at the *Lep* promoter, after initial decrease in methylation^{457,458}. This may suggest that methylation the *Lep* promoter could represent a feedback response to curb the initially increasing leptin levels.

Another adipokine with altered levels in obesity is adiponectin, a major regulator of glucose and lipid metabolism⁴⁵⁹. In both mice and humans, adiponectin gene promoter hypermethylation is observed in concordance with lower adiponectin levels^{460,461}, which in turn correlates with obesity, increased insulin resistance and diabetes⁴⁶². In humans, methylation levels of the *ADIPOQ* promoter in subcutaneous adipose tissue (SAT) correlate with body-mass-index (BMI) and plasma levels of LDL cholesterol⁴⁶¹. Recent studies have even suggested a transgenerational relationship between epigenetics and obesity. Maternal obesity in rodents has been shown to induce hypomethylation of the CGI in the *Zfp423* promoter of the fetal adipose tissue⁴⁶³. *Zfp423* is a key regulator of adipogenesis and its overexpression leads to increased adipose expansion⁴⁶³. Its high expression is therefore an important factor in the predisposition to obesity observed in the offspring of obese female mice. On the other hand, humans conceived in cold months tend to have a lower BMI and higher respiratory activity⁴⁶⁴. A similar relationship has been observed between the date of conception and BMI and BAT activity in humans. Moreover, cold exposure of male mice before mating increases BAT activity and ameliorates obesity related effect in the offspring⁴⁶⁴. Here, *Adrb3*, a gene that responsible for β -adrenergic stimulation in BAT, has been found hypomethylated in sperm⁴⁶⁴. While the connection between DNA methylation and metabolic diseases has been made over the last decade, growing evidence already suggests its importance for understanding obesity and diabetes. However, it is important to keep in mind that most of these studies remain observational or correlational. It is therefore important to address the causality of such epigenetic events in diseases by epigenetic editing.

It has to be noted at this point that DNA methylation, while involved in all the previously described physiological and pathological processes, is not the only player, but rather a piece in the puzzle of our understanding of the machinery involved in gene expression regulation and its aberrations. Moreover, the causative role of epigenetic modifications, including DNA

methylation, in both normal and aberrant transcription regulation is a matter of intense debate. Indeed, a lot of studies tend to imply causality from correlations, which is wrong no matter how strong the correlations are. This being said, not many would argue against the important causative implication of DNA methylation in the mechanisms described above. The important next step is to try to understand when are changes in DNA methylation causative and when, on the contrary, these changes are mere effects of other events. This also applies on changes in other epigenetic marks such as histone modifications and nucleosome remodeling that usually go hand in hand with DNA methylation changes.

Finally, it is important to note that genetic predispositions, mutations and aneuploidy have the strongest impact on both cancer and metabolic diseases. So why is it important to study the less obvious epigenetic component of these diseases? A part of the answer resides in the important fact that, unlike genetic aberrations, epigenetic modifications are easily reversible. Therefore, finding differences in DNA methylation between healthy and diseased individuals/cells may provide us with novel medical targets.

DNA Methylation and its Relationship with TFs – Identification of Super Pioneer Transcription Factors (SPFs)

While the exact relationship between TF activity and DNA methylation remains elusive, a more comprehensive picture of possible binding modes is emerging (Fig. 1.16). Traditionally, DNA methylation has been regarded as an obstacle for TF binding, and only factors with a 5-mC-binding domain (MBD) are able to bind methylated DNA^{335,465-467}. While for some TFs, the methylation itself might be the obstacle, in other cases occupation by an MBD protein and resulting formation of heterochromatin may be the reason for obstruction. However, in recent years other TF-5-mC relationships have emerged. One already mentioned example for a TF which selectively targets methylated DNA sequences is ZFP57 that binds at ICRs where it retains methylation for imprinting purposes^{366,468}. Another example is Kaiso, a repressive TF

with multiple DBDs, that can bind methylated DNA with its C2H2 ZF domain⁴⁶⁹. Interestingly, it has been shown that amongst its targets is the methylated promoter of the TSG p16^{470,471}. Since it does not bind its methylated motif under all circumstances,⁴⁷² however, the relationship between Kaiso and its methylated binding site may be cell type specific. The list of TFs able to bind 5-mC is continuously expanding and includes many representatives with ZF DBDs, and also known pioneer factors such as KLF4, GATA4, and FOXA1⁴⁷³. Findings suggest that hydrophobic interactions, due to amino acid residues such as arginine, facilitate 5-mC binding, and indeed, structural similarities can be found between KLF4, and methyl-CpG-binding protein 2 (MeCP2)^{474,475}. In general, however, proteins with a variety of DBDs including ZF, homeobox, bHLH and E2F have shown 5-mC-binding potential⁴⁷⁵. Furthermore, the sequence context between methylated and unmethylated motifs of the same TF may change. Indeed, KLF4 recognizes the TCCmCGCCC motif only if the CpG is methylated, while its canonical motif TTTACGCC is only bound when unmethylated⁴⁷⁵. Also the affinity to motifs, which are recognized in both methylated and unmethylated context, may differ. Keeping KLF4 as an example, *in vitro* pulldown-coupled MS/MS experiments have shown higher affinity for the methylated GGGCGTG motif, rather than the unmethylated one⁴⁷⁶.

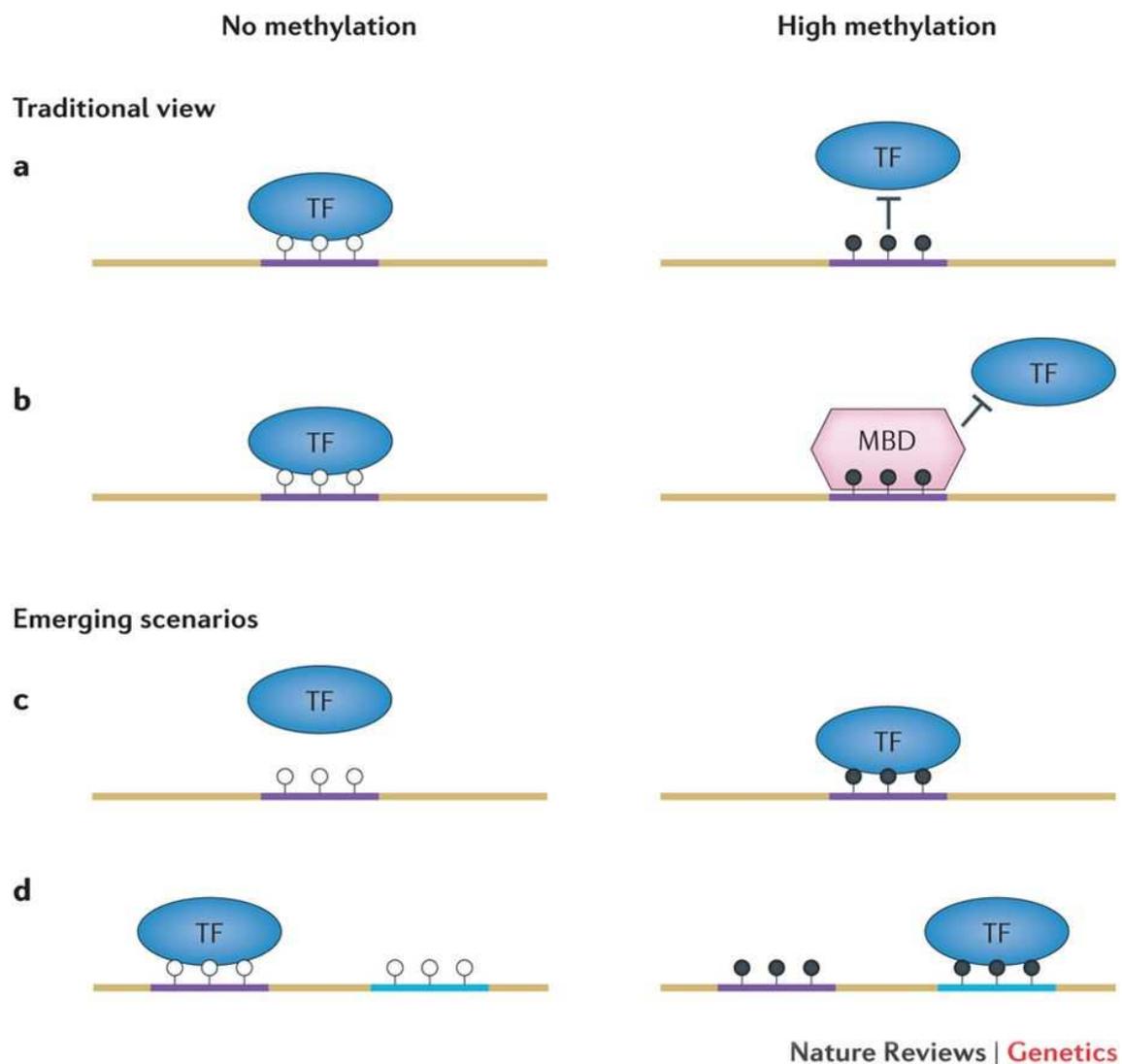


Figure 1.16: Interaction Modes between TFs and DNA⁴⁷³. While it was traditionally considered that DNA methylation prevents TF binding, either due to inability to bind to methylated DNA (**A**) or due to MBD-facilitated obstruction (**B**), new binding scenarios emerge, where methylated motifs are preferred (**C**), or differential binding between methylated and unmethylated motifs occurs (**D**).

An important question regarding the relationship between DNA methylation and TFs is that of cause or consequence. As mentioned before, historically it was believed that DNA methylation dictates TF binding or absence thereof. And while this is certainly true for some TFs, other TFs have been shown to influence the methylation status upon binding in a certain location (Fig. 1.17A). Examples for local demethylation after binding are the architect of chromatin CTCF and RE1-silencing transcription factor (REST)²⁵⁵. In a study by Stadler et al., a reporter region was inserted into an epigenetically neutral locus within the mouse genome. The region

was either *in vitro* methylated prior to insertion or not, and either contained a wild-type or scrambled CTCF motif. It has been shown that in case of the wild-type motif, CTCF is getting recruited and, as a result, the region around the motif became demethylated or stayed unmethylated, depending of the prior methylation status. These results revealed the ability of CTCF to be involved in the demethylation process²⁵⁵. Similarly, a KO of REST has shown an increase in methylation around its endogenous binding sites in ES cells²⁵⁵. In contrast to CTCF and REST, which are associated with demethylation, proteins like nuclear receptor subfamily 6 group A member 1 (NR6A1) interact with DNMT3A/B, which leads to methylation, e.g. at the OCT4 promoter⁴⁷⁷.

Vice versa, DNA methylation can also dictate TF binding (Fig. 1.17B). While several studies have shown this for various TFs *in vitro*⁴⁷⁸⁻⁴⁸¹, this was also confirmed by *in vivo* studies, e.g. on NRF1. Here, a triple KO of DNMT1/3A/3B in mouse ES cells (EStko), created new, albeit non-conserved, binding sites for NRF1⁴⁸², indicating that the presence of methylation in WT cells blocks NRF1 binding at these sites. Interestingly, CTCF, which can bind its motif under most circumstances regardless of its methylation status, has also been shown to be methylation sensitive, i.e. its binding is dictated by the methylation status, in some contexts. One example for such a context is the imprinted *Igf2/H19* locus, where it binds the ICR on the unmethylated maternal allele, but not on the methylated maternal allele^{483,484}. Since CTCF binding, however, is not globally changed in EStko cells, its methylation sensitivity represents the exception rather than the rule²⁵⁵.

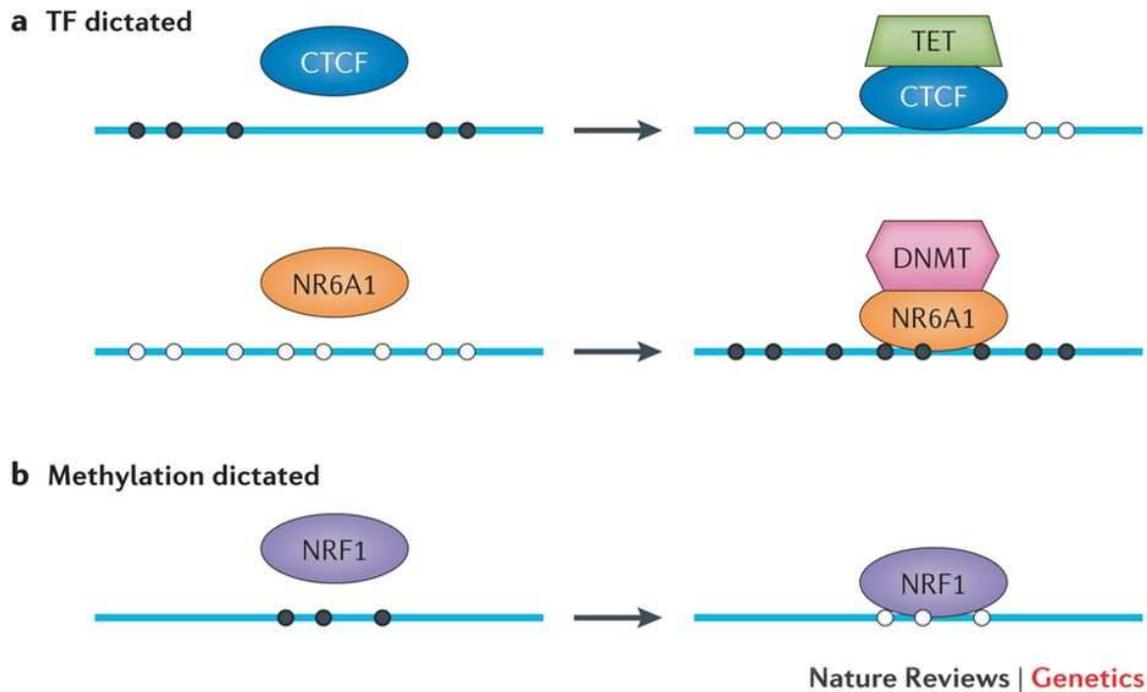


Figure 1.17: TF Binding and DNA Methylation: which is the cause and which is the consequence⁴⁷³? A) TFs such as CTCF or NR6A1 may shape the epigenetic landscape upon binding, by interacting with enzymes such as TETs and DNMTs B) In other cases, methylation dictates the binding behavior of TFs, e.g. NRF1.

To better understand the dynamics between TFs and DNA methylation, further TFs, whose binding is the cause rather than the consequence for demethylation, need to be identified. Promising candidates would be PFs, which can already access repressive chromatin and open it for other TFs. Such demethylating PFs have been identified in a study recently published by our laboratory³⁴⁴, where PFs were categorized by their ability to protect from or to remove methylation. CTCF, REST, SOX2, KLF4, FOXD3, CREB, FOXA1, and SOX17 were identified as Super Pioneer Transcription Factors (SPFs), as they can bind closed methylated chromatin and, in addition to their ability to induce nucleosome remodeling, have the ability to induce DNA demethylation around their binding sites. Additionally, it was shown that, with the exception of SOX2 that inhibits DNMT1 during replication, these SPFs seem to induce TET-mediated, active demethylation.

The relationship between DNA methylation and TF binding is not as straightforward as initially expected. However, new studies focusing on individual TFs and taking genomic contexts into consideration, have revealed a more complete picture. Perhaps these newly identified ‘writers/erasers’ of DNA methylation could be harnessed for applications, such as targeted activation or repression of gene expression. The project presented here stem from this idea and aims at recruiting SPFs and other epigenetic editors, via CRISPR-Cas9 technologies, to activate epigenetically repressed genes that could be beneficial in diseases like cancer or obesity.

1.2 CRISPR

Some discoveries may seem insignificant at first, and some do not receive any attention for decades, because their impact is not well understood or underestimated. An example for such a long-neglected finding was the discovery of clustered regularly interspaced short palindromic repeats (CRISPR) in *E. coli* in 1987⁴⁸⁵. However, since the turn of the millennium much has changed in our understanding of the function of these sequences, and the resulting genome editing techniques developed since 2012/2013 can beyond doubt be regarded as one of the greatest breakthroughs in life science during the last decade. Nowadays, CRISPR-based techniques have found a home in laboratories all over the world, first clinical applications are being developed, and, also outside the realm of sciences, a discussion on the impact and ethics of genome editing is in full swing. The following chapter discusses CRISPR, from its finding and its biological function, to the techniques and its impact on human life.

1.2.1 Exploits from an Ancient War

When mentioned nowadays, CRISPR is usually discussed by both the public and the scientific community in the context of genome editing to which it became synonymous. However, the bacterial defense mechanism, which it originally stands for, is not less extraordinary. Bacteria are, in fact, at constant war with viruses such as bacteriophages, which are with around ten billion particles per liter of seawater the most abundant biological agent on Earth⁴⁸⁶⁻⁴⁸⁸. To cope with this biological pressure, archaea and bacteria have developed innate immune defenses, such as restriction enzymes as well as adaptive immunity through CRISPR, which involves the incorporation of invading viral sequences (spacers) into a ‘genomic archive’. These spacers can be utilized in the form of guide RNA (gRNA) during future infections for targeted guidance of a DNA-cutting enzyme, reminiscent of a “WANTED” poster. The characteristic spacer-repeat-spacer sequence was first discovered in *E. coli* by Ishino et al. in 1987⁴⁸⁵, although further findings were a long time in coming. After the turn of the century it has been revealed that

CRISPR *loci* are quite abundant, being present in around 50% of bacteria and 90% of archaea⁴⁸⁹⁻⁴⁹¹. The repeats vary to great extent in both length and sequence between species, with some coding CRISPR RNA (crRNA) for specific structures such as hairpins and others not⁴⁹². Nevertheless, some generally common features are present, like a GAAA(C/G) motif at the 3' end, or an AT-rich sequence at the flanks of the CRISPR locus, called leader sequence^{489,492,493}. It has also been shown that within these leader sequences, promoter elements are present⁴⁹⁴⁻⁴⁹⁶. Furthermore, not only do most prokaryotes display CRISPR mediated immunity, some even have more than one CRISPR locus⁴⁹⁰.

In 2002 an additional cluster in close proximity of the CRISPR locus has been discovered, expressing genes for CRISPR-associated (Cas) proteins⁴⁹⁷. Over 45 different Cas protein families have been identified so far⁴⁹⁸, leading to various classifications of CRISPR-associated systems^{492,498-500}. Until recently, the most commonly used classification system was divided into three types with multiple subtypes, based on phylogenies of the CRISPR repeats and *Cas* genes⁴⁹⁹. Type I has six subtypes (A-F), while both Type II and Type III have two subtypes each (A/B)⁴⁹⁹. However, the discovery of Types IV-VI further categorizes these types in two classes⁵⁰¹. Class 1 contains types I, III, and IV, which all employ a large complex of Cas proteins involved in interference (Fig. 1.18A), while Class 2 includes types II, V, VI, which employ a single large Cas protein (Fig. 1.18B)⁵⁰¹. For sake of simplicity, only types I-III will be covered in this chapter. All types and subtypes have Cas1 and Cas2 in common, which may serve as a hallmark for the entire system^{498,500}. The hallmark protein for Type I is Cas3 that possesses helicase and cleavage domains for degradation of dsDNA^{499,500}. Cas3 by itself, however, is not sufficient to target specific DNA sequences and requires association with crRNA-guided surveillance complexes^{502,503}. The first targeted antiviral defense has been described in *E. coli* in 2008, showing the association between crRNA and a complex of Cas proteins, forming a ribonucleoprotein complex recognizing specific sequences⁵⁰². This CRISPR-associated

complex for antiviral defense (Cascade) consists of eleven subunits from five Cas proteins, with Cas6 responsible for maturation of the crRNA transcript⁵⁰²⁻⁵⁰⁵. Type II is only present in bacteria⁴⁹⁹ and consists of fewer Cas genes than Type I, with Cas9 being the hallmark gene for this type. Cas9 is a large protein, responsible for both crRNA maturation and cleavage of dsDNA, with a HNH nuclease domain cutting the strand complementary to crRNA2 and a RuvC-like domain, cutting the opposite strand⁵⁰⁶⁻⁵⁰⁸.

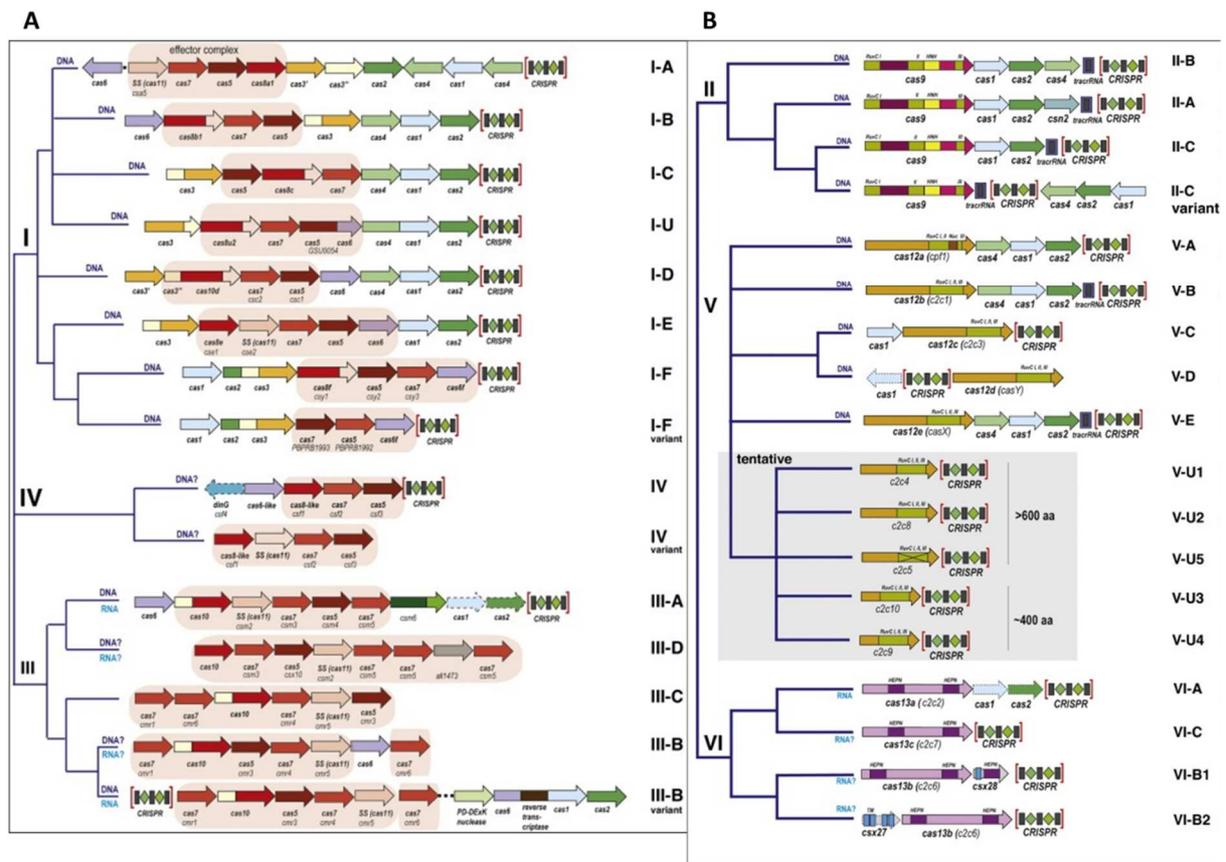


Figure 1.18: Classification of CRISPR-Cas Systems⁵⁰¹. Phylogenetic studies have revealed several types and subtypes of CRISPR *loci* in prokaryotes. The types can be categorized into two classes, with Class 1 (A) employing a complex of multiple Cas proteins as effector, and Class 2 (B) using a single Cas effector protein, e.g. Cas9.

In contrast to the other two types, crRNA maturation in Type II requires a trans-activating crRNA (tracrRNA), which is encoded upstream of the CRISPR locus in *Streptococcus pyogenes* (Sp)⁵⁰⁶. The hallmark genes of Type III, which is mostly present in archaea, are Cas6 and Cas10 that show endoribonuclease and target interference, respectively⁴⁹⁹. However, Type

III-A and III-B show distinct target difference, with III-A cleaving DNA⁵⁰⁹ and III-B cleaving RNA⁵¹⁰.

CRISPR-mediated adaptive immunity can be divided into three phases (Fig. 1.19): CRISPR adaptation, crRNA biogenesis, and crRNA-guided interference⁵¹¹.

CRISPR adaptation

In 2005, it has been discovered that the spacer sequences actually match with phage or plasmid sequences⁵¹²⁻⁵¹⁴. This extraordinary discovery not only enabled us to make a connection between CRISPR and a possible immune response⁵¹⁴, but also revealed active and controlled insertion of foreign sequences into the prokaryotic genome, which then serve as a memory system that allows a future efficient response against invading phages. This process is referred to as spacer acquisition or CRISPR adaptation, while the foreign DNA fragments being selected for insertion are called protospacers. Interestingly, a conserved sequence immediately upstream or downstream of the protospacer, called protospacer-adjacent motif (PAM) has also been identified⁵¹³. While being fairly constant within a Cas subtype, this sequence shows high variability in its composition or location between subtypes⁵¹⁵. Although the exact mechanism of protospacer acquisition remains elusive, the conserved presence of the PAM suggests that it is recognized by the integration machinery⁵¹⁶. Mutational analysis has also identified proteins involved in the process. Early studies focused on Type II and identified Csn2 and Cas4 for II-A and II-B, respectively^{498-500,507}. Since Csn2 and Cas4 are not present in the other types, this suggests that other enzymes should be responsible there. Indeed, Cas1 and Cas2, which are contained in all clusters have been identified as nucleases for integration⁵¹⁶ and their overexpression leads to an increase of spacer elements⁵¹⁶.

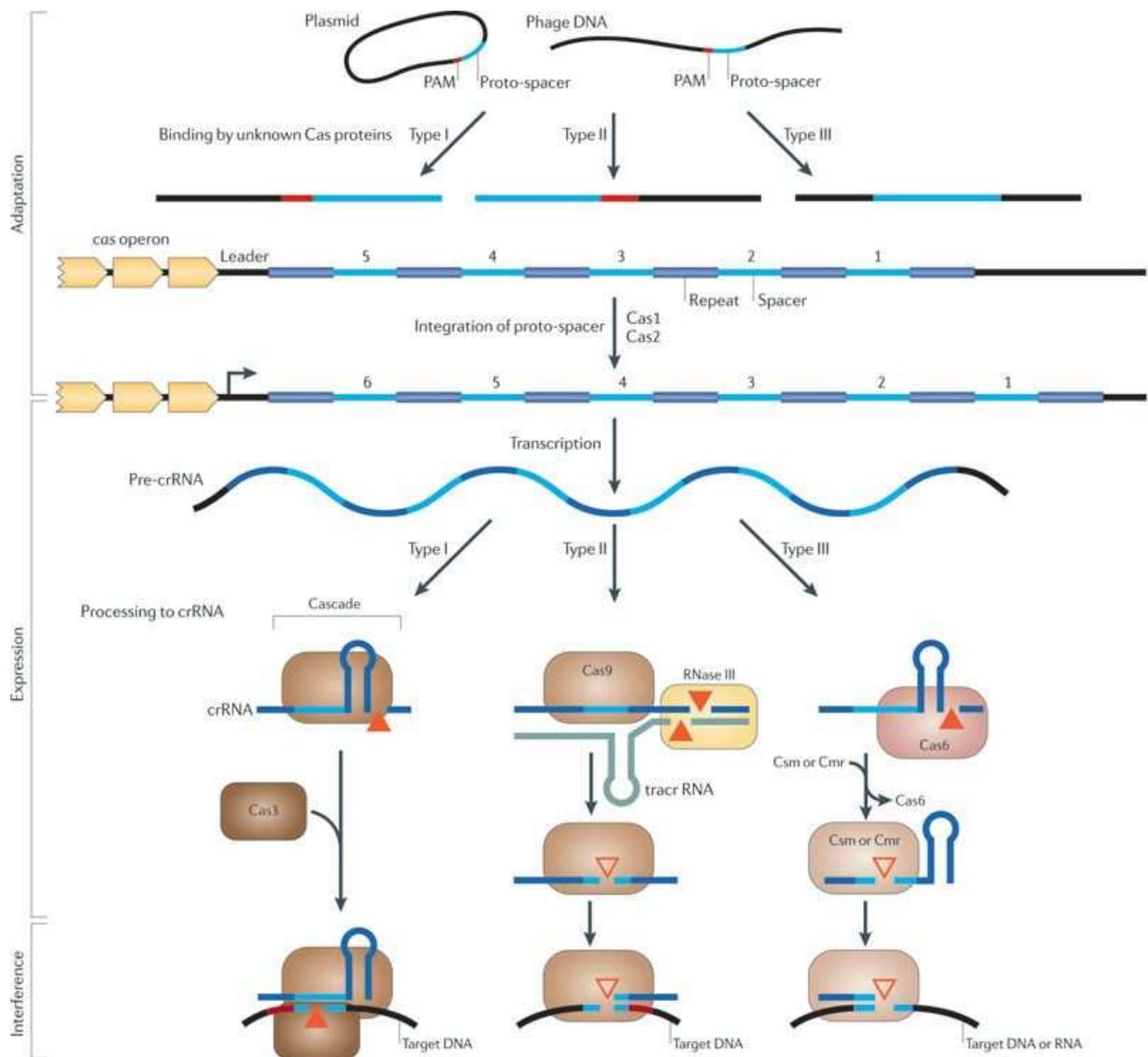


Figure 1.19: The Three Stages of CRISPR-mediated Immunity⁴⁹⁹. During adaptation, pieces of the invading genome are getting integrated into the CRISPR locus in a process called spacer acquisition, which is mediated by Cas1 and Cas2. The resulting pre-crRNA is expressed in a long transcript, covering most of the CRISPR locus. During its biogenesis, it gets trimmed into individual, mature crRNAs, a process which differs between the types. Finally, the foreign DNA or RNA is intercepted by the newly formed ribonucleoprotein complex and destroyed.

crRNA biogenesis

Individual, mature crRNAs are derived from a single long transcript that is expressed from the CRISPR locus⁵¹⁷. crRNA biogenesis in Type I is facilitated by Cas6 that recognizes the stem-loops formed by the repeats within the premature crRNA^{499,502,505}. One exception is I-C that

lacks Cas6, and therefore the maturation is processed by Cas5d⁵¹⁸. In Type III, Cas6 is also responsible for crRNA maturation, however, the mechanism differs since no loops are present^{519,520}. Here, the pre-crRNA is cleaved at the 5'-end of the single-stranded RNA repeat, which is followed by 3' trimming in III-A⁵²⁰. In Type II, the mechanism diverges drastically, with Cas6 not present at all. Instead, tracrRNA forms a complex with Cas9 and the complementary sequence present in the repeat sequence of the pre-crRNA. The resulting dsRNA is processed by RNase III, leaving a cleaved tracrRNA-crRNA duplex behind in the Cas9^{506,508}.

crRNA-guided interference

Interference of invading viruses generally involves recognition of the foreign sequences by the complementary crRNA and subsequent cleavage by a Cas protein. In Type I, hybridization of crRNA within Cascade leads to the recruitment of Cas3, which in turn degrades the DNA with its nuclease domain^{502,503,521,522}. Similar to Type I, in Type III, a complex of several Cas proteins is involved in the destruction of either invading DNA or RNA, depending on the subtype^{510,523}. The interference mechanism in Type II on the other hand is exclusively carried out by one protein, Cas9, which cleaves the dsDNA with its HNH and RuvC-like domains, one for each strand, leaving behind a blunt double strand break^{507,508}. In addition to being involved during spacer acquisition, the PAM is also essential for binding during interference in both Type I and Type II^{507,524}, however, no PAM sequences have been discovered for Type II as of yet.

Taken together, our understanding of CRISPR has evolved from an obscure sequence pattern to an intricate system of adaptive immune response against phages in prokaryotes. While certain questions remain to be addressed, such as the mechanism behind spacer selection and integration, opportunities for its application in genome editing and beyond have opened up. These and more will be discussed in the following chapters.

1.2.2 Harnessing CRISPR for Genome Editing

Since our understanding of the composition and function of the genetic code, there has been a desire to edit it in a spatio-temporal controlled and non-random manner. One key aspect of genome editing at a desired location, is the introduction of a DNA double strand break (DSB), which dramatically increases the probability of targeted insertion^{525,526}. Following a DSB, the two main pathways of repair are homology-directed repair (HDR) and non-homologous end joining (NHEJ)^{527,528}. HDR involves the presence of a repair template, which can be other DNA alleles in diploid organisms or artificial exogenous templates⁵²⁷. This mechanism is desired when the introduction of a sequence, ranging from a single base pair to entire genes, is the goal. NHEJ, on the other hand, does not require a template and often leads to the occurrence of random insertions or deletions (indels)⁵²⁸. In certain applications NHEJ can be desired, e.g. in the establishment of gene knockouts (KOs), due to the introduction of a missense in the amino acid code.

Enzymes with endonuclease activity recognizing specific sequences have been used for cutting DNA since the 1970s⁵²⁹. In fact, the already mentioned restriction enzymes are a crucial component of the innate defense system of bacteria^{530,531}. However, the sequences recognized by these enzymes tend to be too short and therefore too abundant for specific genome editing purposes. Meganucleases, on the other hand, are enzymes that recognize longer sequences of 14 to 40 bp, and thus circumvent the problem of multiple cut sites due to their increased sequence specificity⁵²⁶. These enzymes can be found in a broad variety of organisms like yeast, archaea, bacteria, and plants, but although the catalog of available meganucleases has been expanding, they still remain a rigid tool in regards to target selection. Also the preferred repair mechanism for meganuclease-induced DSBs is NHEJ over HDR⁵³². Furthermore, their activity can be influenced by DNA methylation and the chromatin status of their target site^{533,534}. The first modular enzymes developed for targeted genome editing were zinc finger nucleases

(ZFNs). These artificial, multi-modular proteins consist of the catalytic domain of a FokI endonuclease fused to a series of ZFs, each of which recognizing a particular 3bp sequence, leading to target flexibility^{535,536}. In order to fulfill its catalytic function, FokI needs to form a homodimer, i.e. two ZFNs have to be recruited to the same location, further increasing specificity. ZFNs have proven to be successful genome editing tools, greatly improving HDR at the desired target sites across multiple organisms^{537,538}. About a decade later, transcription activator-like effector (TALE) proteins, which recognize single bases were utilized for genome editing, also by fusing them to FokI⁵³⁹⁻⁵⁴². This takes advantage of the highly conserved DBD of TALEs which consists of 33-34 amino acids, with only the ones at positions 12 and 13 being variable for specific nucleotide recognition^{539,540}. While both techniques greatly improved genome editing, the requirement of protein engineering made them quite costly and time consuming, limiting their every-day application.

Enter CRISPR: In this system, the endonuclease module, in form of a Cas protein, remains constant, disregarding special applications for now, while the targeting is flexible due to an easily programmable guide RNA (gRNA). Special attention has been given to Type II CRISPR enzyme Cas9 that can perform its nuclease activity on its own and does not require a complex (Fig. 1.20). In 2012, it was demonstrated that Cas9 together with a gRNA could target bacterial sequences^{508,543}. Notably, while Gasiunas et al. used both crRNA and tracrRNA for guidance, Jinek et al. fused the two together into a single guide RNA (sgRNA), which remains the most commonly used system today. One year later, the application of CRISPR/Cas9 for genome editing has been shown in human cells⁵⁴⁴⁻⁵⁴⁶ and the age of CRISPR had begun.

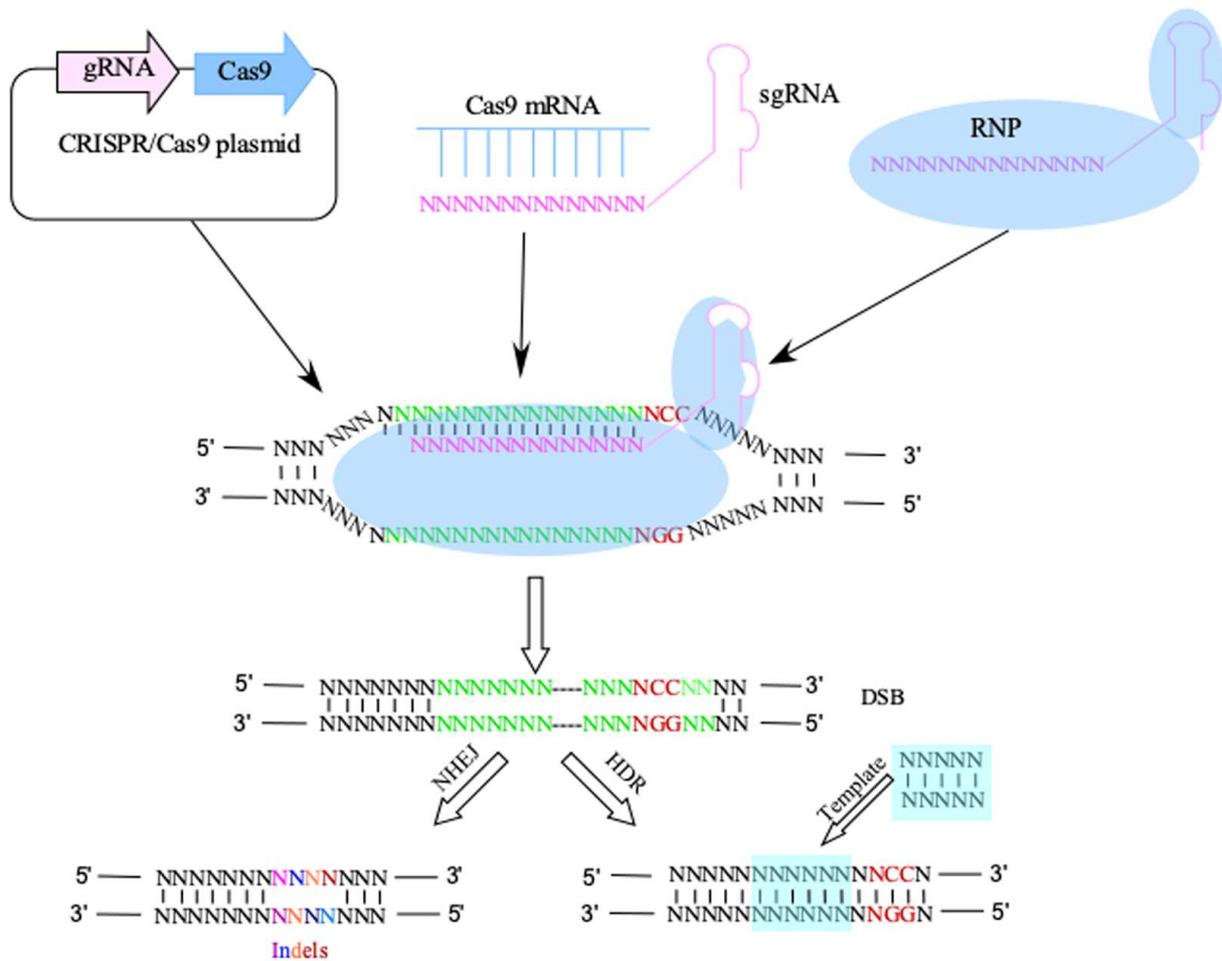


Figure 1.20 : Genome Editing via CRISPR/Cas9⁵⁴⁷. Subsequent to delivery via plasmid, RNA, or ribonucleoprotein (RNP) transfection, or viral transduction (not depicted), the RNP complex binds to its target site within the genome and produces a double strand break (DSB). The DSB may be repaired via non-homologous end joining (NHEJ), which may lead to the formation of desired or undesired indels, or homology-directed repair (HDR), which requires a repair template.

While the Cas9 of *Streptococcus pyogenes* (spCas9) remains the most commonly used CRISPR enzyme, its large size of 1368 aa puts limitations on its delivery via viruses with small packaging capacity, such as adeno-associated viruses (AAVs). Also its PAM sequence (NGG), while being relatively short, limits its flexibility in regards to target choice. Therefore, efforts are being made in finding or engineering Cas proteins with smaller size or alternative PAMs. An example for a smaller variant would be the Cas9 from *Campylobacter jejuni* (CjCas9) with 984 aa, but requiring a fairly complex PAM (NNNNACAC)⁵⁴⁸. On the other hand, Cas9 from

Francisella novicida (FnCas9) has been engineered to recognize a YG (Y = C or T) PAM sequence instead of NGG. However, FnCas9 is, even larger than spCas9⁵⁴⁹.

Engineering of Cas9 also goes beyond size and flexibility. Early on after the first applications, concerns regarding off-target effects of Cas9⁵⁵⁰⁻⁵⁵³ were reported. These concerns have been addressed by changing the delivery method from plasmids to direct transfection of the ribonucleoprotein (RNP) complexes, which ultimately leads to a shorter retention time of the Cas9 within the cell^{554,555}. Regarding the actual protein, engineering approaches have been performed exchanging specific amino acid residues, which led to the creation of more specific high fidelity spCas9 variants^{556,557}. Another approach is the creation of a Cas9 nickase that induces a single strand break, by mutating either one of its catalytic HNH and RuvC-like domains. The induction of a DSB would then be achieved by the presence of two nickases at the same location, therefore increasing overall specificity^{551,558}. Going one step further, mutating both domains (D10A and H840A for SpCas9) leads to the creation of a catalytically dead Cas9 (dCas9)⁵⁵⁹ that maintains its ability to be recruited to target sequences but does not induce breaks at these sequences. dCas9 could be fused to FokI to increase the specificity of the cut^{560,561}, but more interestingly, it could be fused to a multitude of different proteins/effectors that you would wish to recruit to a specific sequence of the genome. The many applications which opened up due to dCas9-effectors will be explored in the following chapter.

1.2.3 No Cutting Needed: dCas9-Effectors

Over the last years, the flexibility and easy programmability of dCas9 has led to applications of CRISPR that go far beyond genome editing. The first application of dCas9 was interference with gene expression by steric hindrance of RNA polymerase binding or transcription elongation, simply due to its recruitment⁵⁵⁹. This was the first example for CRISPR interference (CRISPRi) and it utilized a 'naked' dCas9 (Fig. 1.21A) that by simply binding to regulatory

regions would block the binding of transcription factors, thus leading to transcription repression. However, when fused to effector proteins, a whole new tool box opens for researchers. For example, another CRISPRi method, utilizing the Krüppel-associated Box (KRAB) repressor complex as effector (Fig. 1.18A), results in more robust gene repression than just a naked dCas9⁵⁶². Among the effectors are also fluorescent proteins, which enable visualization of genomic loci in the nucleus⁵⁶³, or the aforementioned FokI endonucleases to improve target specificity^{560,561}. Here, however, we will focus on applications useful for the regulation of gene expression and epigenetics.

First approaches of CRISPR activation (CRISPRa) fused VP64 (Fig. 1.21B), i.e. four copies of the 16-amino-acid transactivation domain (VP16) of the Herpes simplex virus, to dCas9^{551,564,565}. Later approaches involved the fusion of several different transactivation domains, like VP64, mammalian P65, and Rta from Epstein-Barr virus (collectively VPR), resulting in more stable induction of gene expression⁵⁶⁶. Instead of fusing the effector directly to the dCas9, multiple copies of the effector can be either recruited to a combinatorial version of the sgRNA, or to an adapter protein which in turn is fused to the dCas9. In case of the former method, termed synergistic activation mediator (SAM), MS2 bacteriophage coat protein (MCP) fused to the effector is getting recruited to the MS2 RNA aptamer extensions of the combinatorial sgRNA^{567,568}. The SunTag method, on the other hand, employs a protein adapter fused to dCas9, which recruits several copies of an antibody-tagged effector⁵⁶⁹. Each of these methods has been shown to successfully activate genes, with VPR, SAM, and SunTag, i.e. the recruitment of multiple transactivators, outperforming simple VP64⁵⁷⁰.

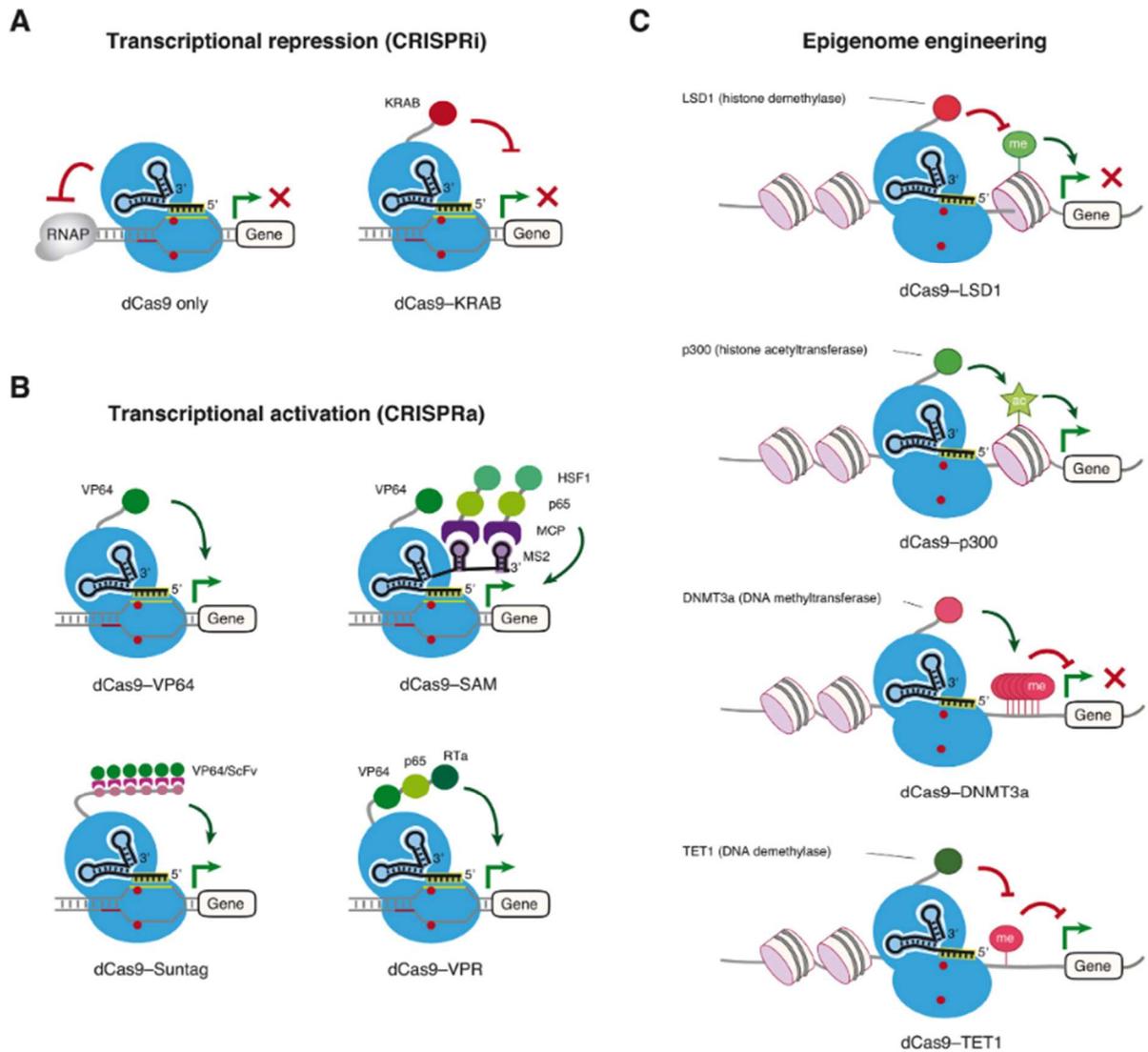


Figure 1.21: CRISPR-based Systems Involving Regulation of Gene Expression and Epigenetics⁵⁷¹.

A) CRISPRi systems interfere with gene expression by either steric hindrance or through repressors such as KRAB, which facilitate the formation of heterochromatin. **B)** CRISPRa systems employ transactivators from various species, which can be either directly fused to dCas9 or recruited in multiple copies through combinatorial RNA or protein adapters. The latter approaches also work for effectors such as TET1. **C)** Various enzymes regulating histone modifications (e.g. LSD1 or p300) or DNA methylation (e.g. TET1 or DNMT3A) can be fused to dCas9 for epigenome engineering.

While already recruitment of dCas9-KRAB leads to the formation of heterochromatin, other epigenetic modulators have been fused to dCas9 to directly alter the epigenetic landscape (Fig. 1.21C). So does recruitment of dCas9-DNMT3A methylate its target region, which leads to stable repression of gene expression⁵⁷²⁻⁵⁷⁴. Employing TET1 as effector does, as expected,

lead to demethylation and has been used to activate gene expression^{572,574}. Instead of direct fusion to the dCas9, also here recruitment of multiple effectors is possible, e.g. via SunTag adapters⁵⁷⁴. Aside from DNA methylation, histone modifications can also be altered by recruiting the corresponding enzymes to their target site, such as LSD1 for histone demethylation⁵⁷⁵ or p300 for acetylation⁵⁷⁶. The former has shown to greatly reduce H3K4me2 and H3K27ac marks at distal enhancers⁵⁷⁵, while H3K27ac levels have been increased with dCas9-p300⁵⁷⁶ at regulatory regions leading to increased expression of the target genes.

Finally, the expression or presence of the CRISPR system highly depends on the mode of delivery. Thus delivery via plasmid or ribonucleoprotein complex is only transient, while transduction via lentivirus (LV) leads to permanent expression of CRISPR/Cas9 due to its insertion into the host genome. For more temporal control, inducible CRISPR (iCRISPR) systems have been developed. Such systems can be light inducible^{577,578} and take advantage of light-sensitive heterodimerizing proteins like CRY2 and CIB1, one of which is fused to the dCas9, the other is fused to the effector⁵⁷⁸. While the dCas9 remains stably present at its target site, the effector only gets recruited after light induction, which allows temporal control of gene activation⁵⁷⁸. Other inducible systems may employ chemicals, like abscisic acid (ABA), gibberellin (GA), or rapamycin (RAP)^{579,580}, all of which trigger protein dimerization⁵⁸¹⁻⁵⁸³. One particularly elaborate system, Fkbp/Frb-based inducible recruitment for epigenome editing by Cas9 (FIRE-Cas9), first recruits multiple MSPs fused to FK506 binding protein (Fkbp) via M2 RNA loops, which is followed by RAP-induced Fkbp/Frb dimerization with FKBP12-Rapamycin Binding domain (Frb)⁵⁸⁰. The Frb itself is fused to a chromatin modulator, of which multiple copies are present at the target site in the presence of RAP⁵⁸⁰. On the other hand, auxin-inducible degron methods have been developed to induce rapid degradation of the targeting complex⁵⁸⁴.

Taken together, the toolbox for CRISPR-based genome editing and beyond is growing and enables researchers and clinicians alike to harness its potential. Some possible applications and the caveats that come with them, will be discussed in the following chapter.

1.2.4 Opportunities and Challenges

CRISPR has proven to be a powerful tool for a broad range of applications, which have the potential to greatly impact human life and our environment. However, it also faces challenges; not only on a technical level, but also on ethical and ecologic levels.

One of the first things that come to mind regarding CRISPR usage are clinical applications. There are high hopes for novel treatment strategies, particularly for monogenic diseases. In case of sickle cell disease (SCD), genome engineering of human hematopoietic stem cells (HSCs) by disruption of the GATA1 motif within the *BCL11A* erythroid enhancer, results in expression of therapeutic levels of fetal hemoglobin (HbF) from engrafted SCD HSC progeny, which also do not exhibit the ‘sickling’ phenotype⁵⁸⁵. A clinical trial for an *ex vivo* CRISPR gene therapy to restore HbF production is currently under way (NCT03745287 on clinicaltrials.gov). Another example for potential CRISPR gene therapy in monogenic diseases, would be the restoration of the *dystrophin* gene in Duchenne muscular dystrophy (DMD). CRISPR-based restoration of dystrophin in DMD mice, results in amelioration of the DMD phenotype, which ultimately leads to enhancement of muscle force and muscle biochemistry^{586,587}. Long-term studies in mice have shown that the levels of dystrophin are sustained over a period of at least one year⁵⁸⁸. Specific cancers also present targets for CRISPR, such as chronic myeloid leukemia (CML), which often exhibits a Philadelphia chromosome containing a fusion gene of *BCR-ABL1*, originally from chromosome 22 and 9, respectively. The unique sequence resulting from the fusion makes targeted ablation and reversal of the tumorigenic process in CML possible^{589,590}.

An obvious matter of concern for clinical applications is safety. CRISPR has been shown to not only being susceptible to off-target effects, but also unexpected on-target effects, which may result in large deletions sometimes undetectable by conventional sequencing methods and rearrangements⁵⁹¹. Advanced, long-range PCR sequencing methods, however are able to detect these lesions⁵⁹¹. Off-target effects can furthermore be reduced by high-fidelity Cas9⁵⁵⁶, or the very recent Prime Editing⁵⁹². The latter employs a Cas9 nickase fused to a reverse transcriptase, while prime editing guide RNA (pegRNA) serves both as guide and template for editing⁵⁹².

Another CRISPR-related issue is the manipulation of human embryos or the human germline. In November 2018, He Jiankui announced the birth of twin girls, who are the first genetically modified human beings (<http://www.chictr.org.cn/showprojen.aspx?proj=32758>). The C-C chemokine receptor type 5 (CCR5) has been mutated in these girls. This mutation, when homozygous, might result in resistance to HIV-1 infection by its carrier⁵⁹³. The girls, known under the pseudonyms Lulu and Nana, carry only one copy of the mutated gene and are supposed to be in a healthy state. The study, however, has received world-wide criticism from outside and within the scientific community⁵⁹⁴. The exact impact of the CRISPR treatment on the girls has not yet been reported and its long-term ramifications are still up for speculation. The prenatal application of CRISPR on humans is, however, most likely feasible and its use will be more an ethical question rather than a technical one. The consequences of such procedures, especially when it goes beyond the prevention of disease, are still difficult to predict, but the debate on designer babies and the ‘new normal’ is in full swing.

Leaving the clinic and CRISPR applications on humans for the moment, genome engineering not only of individuals, but also of entire species is one of the most recent breakthroughs. An approach called ‘gene drive’ enables the spread of the desired modification to both alleles, and, therefore, all offspring (Fig. 1.22). Targeted disruption of the splice variant *dsx-female* (*AgdsxF*) of the *doublesex* (*Agdsx*) gene in mosquitos combined with a gene drive, results in

female, infertility and ultimately to population control in a lab⁵⁹⁵. Other methods target genes which enable infection of mosquitos by the malaria parasite Plasmodium. Disruption of these genes paired with a gene drive results in a malaria-resistant mosquito population^{596,597}. Both approaches have as ultimate consequence that the malaria parasite will lose its main vector, and there are considerations releasing the modified mosquitos into the wild to battle the malaria epidemic. However, such measures are serious and most importantly, irreversible. The consequences of population-wide genome manipulation, particularly when it comes to population control, are unforeseeable, perhaps even resulting in a disruption of the food chain.

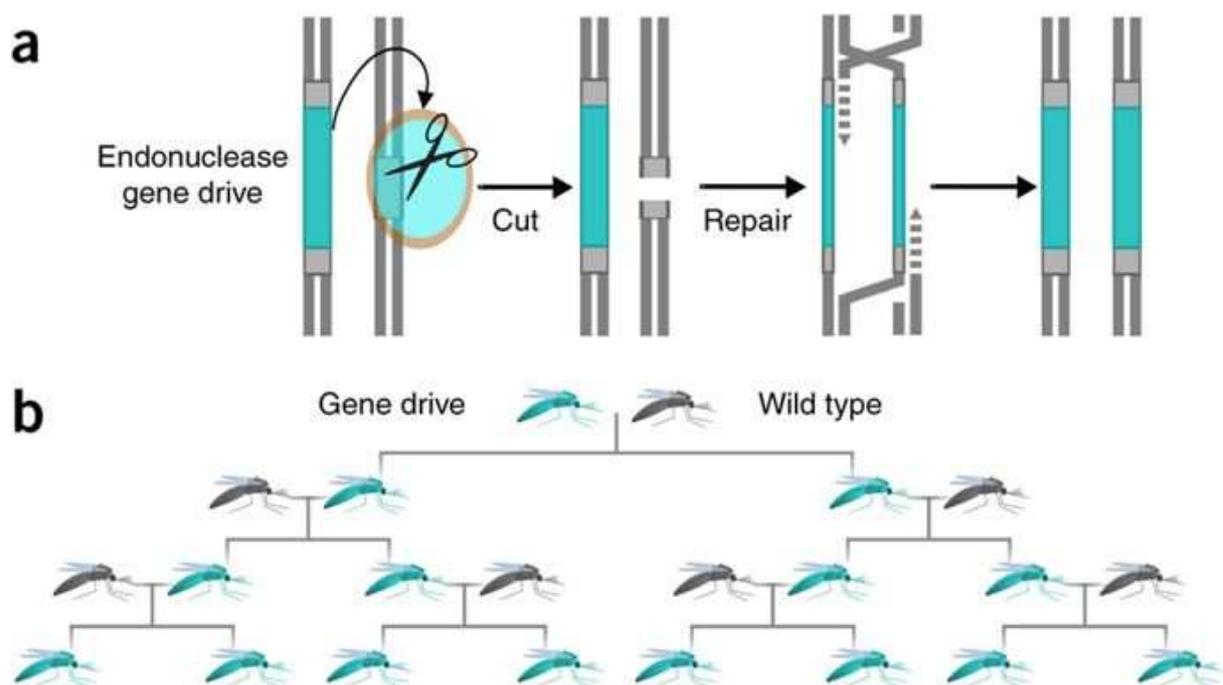


Figure 1.22: Effect of a Gene Drive on the Population Level⁵⁹⁸. **A)** A gene drive is a cassette consisting of the desired modification and an expression system for both Cas9 and gRNA targeting the competing allele. Subsequent to the cut, HDR enables the spread of the modified region to the other allele. **B)** Since both alleles carry the modification, it will spread easily throughout the population.

All in all, CRISPR remains a tool, and, while being a powerful one, its applications are ultimately in our hands. Due to its accessibility and low economic cost it is of outmost

importance that its use is regulated and remains within safe, ethical, and ecologically sound boundaries. Finally, the limitations of CRISPR are also not known yet, and whether it turns out to be the silver bullet for all diseases, the opening of Pandora's Box, or something in between, is unforeseeable from our current position.

2 Aim of the Thesis

As explored during the introduction, epigenetic mechanisms contribute to the acquisition of different cell types within an organism exhibiting distinct traits and phenotypes despite having the same genotype. The different epigenetic patterns and characteristic gene expression profiles in various cell types arise through controlled physiological processes during development. However, through a plethora of mechanisms, some of which are unknown, aberrant epigenetic modifications and expression profiles contribute to the etiology of certain pathologies. DNA methylation, is a major epigenetic player in transcription control and, like all epigenetic marks, is reversible. Due to this reversible nature, aberrations in DNA methylation in diseases represent desirable therapeutic targets. Ultimately, the aim of the thesis is the **synthetic activation of epigenetically silenced genes in disease**, in a **targeted** and **stable** manner. Specifically, this issue has been addressed in two projects targeting two major diseases:

(1) Reactivation of Epigenetically Silenced Tumor Suppressor Genes with Super Pioneer Transcription Factors

In many cancers, such as hepatocellular carcinoma (HCC), tumor suppressor genes (TSGs) are silenced by hypermethylation of their promoter region. In this project, we aim at synthetically recruiting super pioneer transcription factors (SPFs) to methylated TSG promoter regions to induce local demethylation. The hypothesis is that a decrease in methylation will lead to reactivation of the gene and subsequently to cell death or cycle arrest dictated by the function of the targeted TSGs. Recruitment of the SPFs will be facilitated through CRISPR-directed insertion of their corresponding consensus binding motifs into the promoter region close to the transcription start site (TSS). Subsequently, gene expression, proliferation and cell death, will be assessed. In contrast to other approaches targeting cancer epigenetics, like 5-azacytidine, the reversal in methylation will only affect one locus and, due to insertion, should be permanent. In addition to the

establishment of a novel application for gene activation, this project should give insight into our understanding of how DNA methylation affects gene expression, and if its removal is sufficient to reactivate target genes.

(2) Epigenetic Plasticity and Cell Fate of Adipose Tissues: New Targets for Epigenome Editing

During development, cells differentiate into distinct tissues each with its own characteristic epigenetic profile. Certain tissues exhibit plasticity and may give rise to different cell types upon external *stimuli*. The occurrence of beige adipocytes that expend rather than store energy, within white adipose tissue, a process called browning, after cold exposure or calorie restriction of mice is a striking example of this phenomenon. Mice in which these beige adipocytes arise, have a leaner phenotype, higher glucose tolerance, lower insulin resistance, and an overall amelioration of the Diabetes phenotype. Since DNA methylation plays an important role in differentiation and cell fate determination, it is important to identify genes regulated by DNA methylation and that contribute to the browning. Towards this aim, browning of white adipose tissue in mice will be stimulated by cold exposure and microbiota depletion after treatment with a cocktail of antibiotics. Subsequently, the adipose tissue will be harvested from treated and untreated mice, and the transcriptomes and methylomes at single basepair resolution will be profiled. Comparison between beige and white adipocytes will lead to the identification of Differentially Methylated Regions (DMRs) that contribute to the differential expression of genes involved in browning and therefore represent new targets for epigenome editing.

3 Results

3.1 Reactivation of Epigenetically Silenced Tumor Suppressor Genes with Super Pioneer Transcription Factors

Transcription is regulated by differential binding of transcription factors (TFs) to specific sequence motifs. As genomic sequence composition is the same in all cells of an organism, epigenetic “sequence-independent” mechanisms are needed to control correct spatiotemporal recruitment of TFs. However, it is becoming evident that epigenetic mechanisms can, in turn, be influenced by the sequence-composition of their target sites^{255,344,599}. Traditionally, repressive epigenetic marks are thought to inhibit transcription by preventing the binding of TFs. However, recent studies began to shed light on a new class of TFs called pioneer factors (PFs) that are the first to engage target sites containing their binding-motifs in nucleosomal DNA. Such initial binding was shown to induce nucleosome remodeling and chromatin decondensation, allowing additional TFs that are unable to bind to closed chromatin, called “settlers”, to bind the DNA, thus generating a transcriptionally competent chromatin around the binding site^{255,600-603}. Studies in our laboratory additionally identified a subclass of PFs that can bind methylated closed chromatin, initiating DNA demethylation in addition to their ability to remodel the nucleosomes³⁴⁴. These were called Super Pioneer Transcription Factors (SPFs) as they overcome virtually all repressive chromatin marks. These findings suggest that the epigenetic and transcriptional status of a locus could be changed by modifying its sequence composition in a way that permits recruitment of SPFs. Thus, our overarching aim is to activate methylated endogenous genes by SPF-dependent epigenetic editing. As a proof of concept, we focus on activating tumor suppressor genes (TSGs).

TSG inhibition by DNA hypermethylation is an important cancer hallmark^{427,604}. As DNA methylation is reversible, reactivation of methylated TSGs represents an attractive strategy for cancer therapy. Current approaches for TSG demethylation with 5-azacytidin (5-aza) are

promising, but lack specificity and stability⁶⁰⁵. We aim at taking advantage of the unique SPF properties to reactivate methylated TSG promoters, by inserting SPF motifs within these promoters using CRISPR/Cas9 technology. The hypothesis is, that these motifs would recruit SPFs, which leads to demethylation, transcription activation, and ultimately cancer cell proliferation arrest or death.

Abnormal hypermethylation of ordinarily unmethylated TSG promoters is involved in cancer formation^{427,604}. In particular, promoter methylation of TSGs *p16* and *Rassf1a* was shown to play a role in the formation of Hepatocellular Carcinoma (HCC)^{606,607}. *p16* acts as inhibitor of cyclin dependent kinases (CDKs). By binding to CDK4/6, it prevents the formation of the CDK4/6-cyclin D complex, and, further downstream, the phosphorylation of retinoblastoma protein (Rb), which ultimately inhibits S-phase entry^{608,609}. *Rassf1a*, on the other hand, does not exhibit enzymatic activity, but is suspected to be an effector of the Ras oncoprotein and is involved in multiple pathways regulating cell cycle and apoptosis⁶¹⁰. Importantly, it was shown that induced demethylation in HCC cells by treatment with 5-azacytidin reactivates *p16* and *Rassf1a*, resulting in cell cycle arrest and apoptosis^{611,612}. We, therefore, propose to investigate the ability of SPFs to specifically and stably demethylate and reactivate the promoters of TSGs *Rassf1a* and *p16* in HCC cells.

The SPFs tested in this project are CCCTC-binding factor (CTCF) and SRY (sex determining region Y)-box 2 (SOX2) that were recently confirmed as SPFs^{255,344}.

While issues of tumor penetration and *in vivo* targeting in humans remain to be addressed, results could provide novel strategies for epigenetic therapeutics. The study should also contribute to our understanding of the complex relationship between epigenetic and genetic-driven events in transcription regulation.

3.1.1 Validation of the Experimental Approach by Targeting TSGs with dCas9-TET1

To validate our hypothesis and the feasibility of our experimental approach, we first assessed whether targeted demethylation of TSG promoters by established dCas9-effector proteins is not only possible, but also sufficient to induce reactivation of gene expression. Suitable effectors for active demethylation are Ten-eleven translocation methylcytosine dioxygenases (TETs) that oxidize 5-mC into 5-hmC and further oxidation products, an essential step in the process of active DNA demethylation. Previous studies have shown that recruitment of TET1 through dCas9-dependent approaches facilitates both local demethylation and gene activation^{572,574}. To edit the methylation status of the p16 and Rassf1a promoters, we first used transient transfection of HCC cell line Huh7 with a plasmid expressing dCas9-SunTag and antibody-fused TET1CD that allows the recruitment of multiple copies of TET1 to target sites⁵⁷⁴. TET1 fusion was guided to four locations in proximity of the transcription start sites (TSS) of each of the genes by four transiently expressed single guide RNAs (sgRNAs). Unfortunately, despite trying several conditions, the transfection efficacy of Huh7 cells with the rather large dCas9-SunTag-TETCD plasmid proved to be too low for further experiments. Furthermore, the delivery of the CRISPR machinery via plasmid is only transient, and, as the demethylation process and the expected gene reactivation may require more time, stable expression of the components and later measurements were anticipated for the following experiments. Therefore, we decided to apply a more efficient way of CRISPR delivery via lentiviral (LV) transduction. The LV-based delivery system of dCas9-TET1 and sgRNAs has previously been tested on a DNA demethylation-dependent reporter^{572,613}, endogenous *brain-derived neurotrophic factor (BDNF)*, *myoblast determination protein (MyoD)*⁵⁷², and the *fragile X mental retardation 1 (FMR-1)* gene⁶¹⁴. In all those cases, this resulted in demethylation and gene activation. In these studies, transduction of all viruses was performed at the same time, without prior establishment of a dCas9-TET1-expressing cell line. However, we opted of establishing stable cell lines expressing the fusion proteins, which should improve the efficiency of TET1 recruitment to

target sites. Therefore, Huh7 cells were transduced with LV for stable expression of dCas9-TET1 or a catalytically inactive dCas9-dTET1 as a control. Transduced cells were allowed to form colonies that were picked and tested for dCas9-effector protein expression via quantitative reverse transcription polymerase chain reaction (qRT-PCR) two weeks post-transduction. The acquired stable cell lines were subsequently transduced with a second round of LV for the simultaneous delivery of all four sgRNAs as well as a constitutively expressed mCherry, which gets also delivered with the sgRNA LV (Fig. 3.1A). Seven days post transduction, mCherry-positive cells were sorted and kept at regular culture conditions. On day 13 and day 26 after transduction, cells were harvested and RNA, DNA, and proteins (only for day 26) were extracted.

When we used this strategy to target the *Rassfla* promoter, no significant changes in mRNA expression could be observed under these conditions (Fig. 3.1B-C). Therefore, we focused in further experiments on the effects of p16 reactivation. qRT-PCR analysis shows significantly higher p16 mRNA expression at both time points when dCas9-TET1 was targeted to *p16* in comparison to its controls, which either recruit a catalytically inactive dCas9-dTET1 to *p16*, or express dCas9-TET1 that is not targeted to *p16*, or both (Fig. 3.1D-E). The expression, however, remains relatively low in comparison to the p16 expression level in HeLa cells. Despite being tumor cells themselves, p16 expression does not affect the cell cycle in HeLa cells. As many cervical carcinomas, HeLa cells express high-risk human papillomavirus (HPV) E7 protein, which in turn binds hypophosphorylated Rb, leading to destabilization of the Rb/E2F repressor complex^{615,616}. Thus, the downstream target of p16 and CDK4/6 is virtually not present in HeLa cells, diminishing its tumor suppression capabilities.

In order to address whether the increase in p16 mRNA levels upon dCas9-TET1 targeting to the *p16* promoter results in higher protein production, we performed Western blots using a specific p16 antibody. Surprisingly, either transductions did not lead to higher p16 protein

levels (Fig. 3.1F). Whether the absence of p16 protein is due to the fact that the RNA levels are still too low in targeted cells despite the significant increase, or to other post-transcriptional events, cannot be concluded at this point. Moreover, we cannot exclude that the antibody's sensitivity does not allow detection of low levels of proteins. Supporting the latter explanation, the band corresponding to p16 in HeLa cells is quite faint, although p16 mRNA levels were multiple magnitudes higher than the ones observed in Huh7 cells upon dCas9-TET1 targeting. If there is any correlation between the level of RNA expression and protein expression, it would not be surprising that the amount of protein produced in the targeted cells is simply not detectable using this antibody. We therefore sought of testing the status of p16 downstream targets as a proxy of its activity. Since p16 is ultimately responsible for inhibiting the phosphorylation of Rb, we investigated the levels of pRb and Rb on day 26 post transduction. Correlating with mRNA expression levels, dCas9-TET1 recruitment to the p16 promoter reduces the amount of pRb relative to the other transductions, while total Rb levels are not affected, or even slightly higher after dCa9-TET1 recruitment (Fig. 3.1F). These results suggest that targeting dCas9-TET1 to the p16 promoter in Huh7 cells, leads to a small but significant increase in p16 mRNA expression resulting in lower phosphorylation of its downstream target Rb.

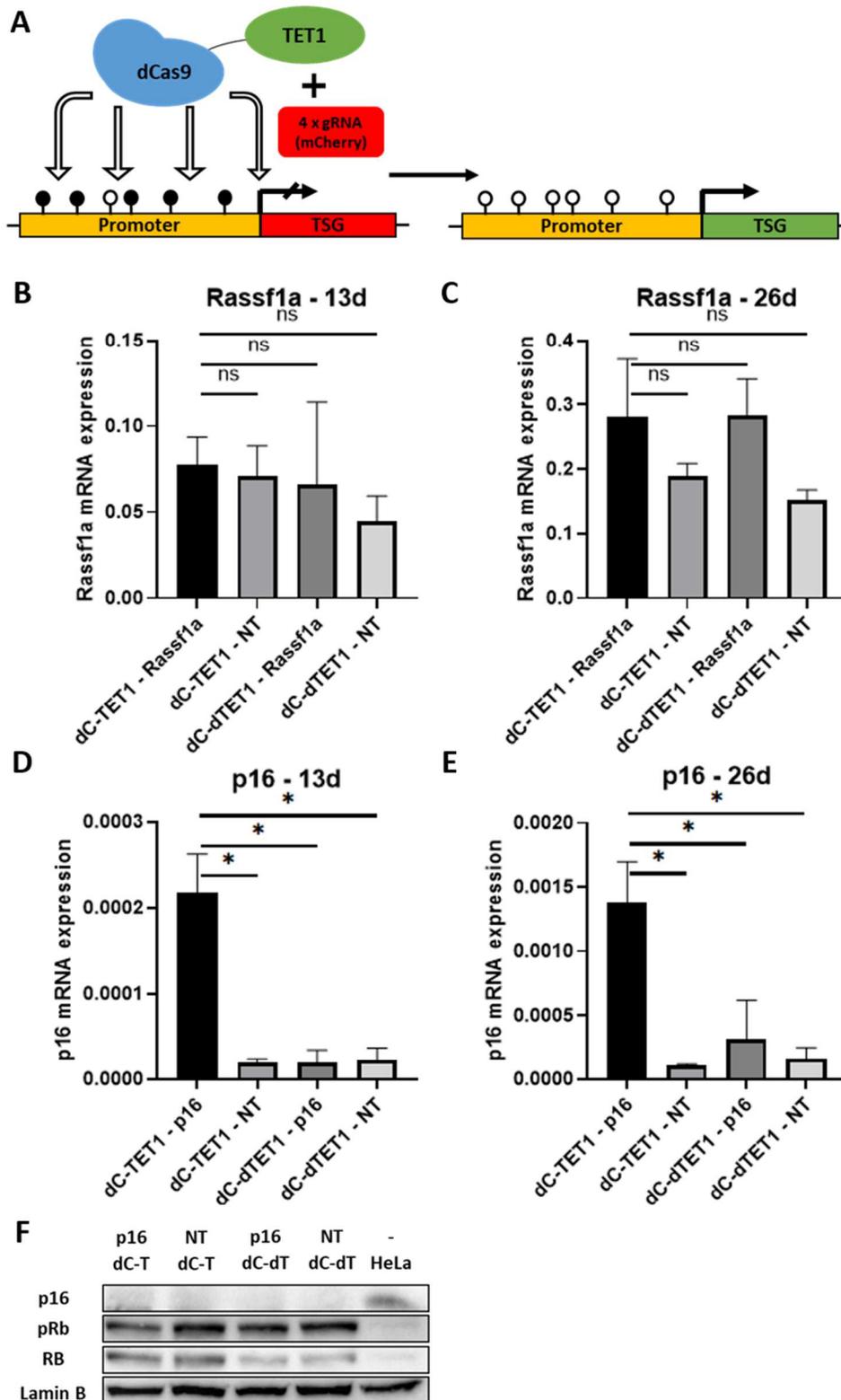


Figure 3.1: Activation of TSG Expression after Epigenome Editing. **A)** Schematic representation of the strategy used for epigenome editing on TSG promoters with dCas9-TET1. **B-E)** mRNA expression of Rassf1a or p16 relative to HeLa control 13 and 26 days post transduction. qPCR was performed with SNRPD3 as housekeeping gene. **F)** Protein levels of p16, pRb, Rb, and Lamin B. dC: catalytically dead Cas9; dT/dTET1: catalytically dead Tet1; NT: non-targeting guide RNA; pRb: phosphorylated Rb.

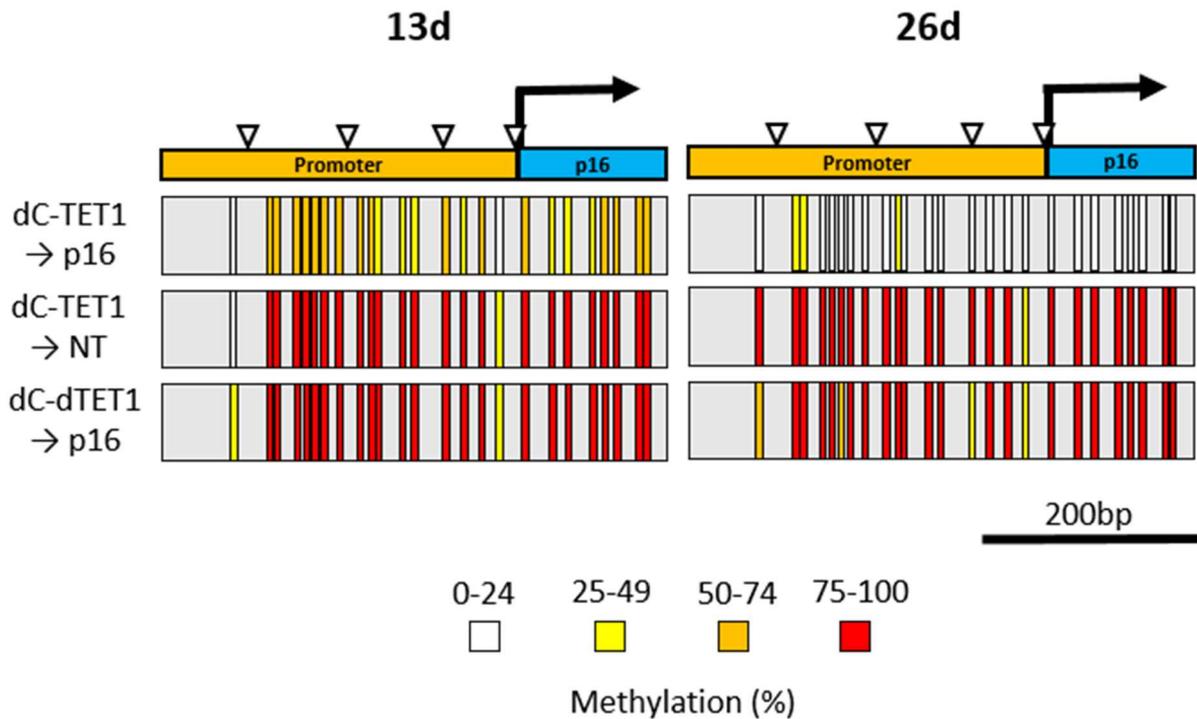


Figure 3.2: Methylation Landscape after Epigenome Editing of the p16 promoter. Bisulfite profile of the p16 promoter in Huh7 cells (stably expressing either dCas9-TET1 or dCas9-dTET1) 13 days and 26 days post transduction. Each vertical bar represents a CpG. CpG methylation percentages based on at least ten reads of bisulfite sequencing are indicated by a color code. The white triangles depict the locations of sgRNA targeting within the p16 promoter. The non-targeting (NT) sgRNA has no specific target in the human genome. dC-TET1-p16: TET1 targeting p16, dC-dTET1-p16: catalytically inactive TET1 targeting p16, dC-TET1-NT: non-targeted TET1, dC-dTET1-NT: non-targeted catalytically inactive TET1

In order to address whether changes in p16 gene expression in upon TET1 recruitment are driven by DNA demethylation, we assessed the methylation landscape of the p16 promoter by bisulfite sequencing on day 13 and day 26 post transduction. Sequencing results show clear demethylation patterns when p16 is targeted with dCas9-TET1, while methylation is maintained in the controls (Fig. 3.2). Furthermore, the methylation level decreases over time when compared to the results from day 13 and day 26. These results confirm the ability of dCas9-TET1 to demethylate and activate a novel target (p16), while methylation is not affected in case a catalytically inactive dCas9-dTET1 is recruited.

3.1.2 Phenotypic Changes Subsequent to Targeted p16 Reactivation

Since p16 is involved in cell cycle control, we started assessing phenotypic changes of the transduced cells by first analyzing their proliferation rate. 10^5 transduced Huh7 cells were plated and kept under regular cell culture conditions. On day 1, day 3, and day 5 post plating cells were counted, revealing a significant difference in cell number between cells in which dCas9-TET1 was targeted to p16 and the controls, indicative of reduced proliferation upon activation of p16 (Fig. 3.3A).

To further confirm this observation, we assessed proliferation by live cell imaging (LCI). Cells were plated into an 8-well chamber and stained with SiR-Hoechst, a cell-permeable DNA probe for fluorogenic visualization of the nucleus. Subsequently, we used live imaging with pictures taken every three minutes over the course of 20 hours with 2 μm z-stacks, followed by the counting of mitotic cells within this time course (mitotic timing). Overall, Huh7 cells clearly performed fewer mitoses within the given time frame, if dCas9-TET1 has been recruited to p16 (Fig. 3.3B). While the difference in proliferation compared to the controls is not statistically significant due to variability between biological replicates, the result could still be considered as biologically relevant.

As the cells seem to have a decreased proliferation rate, we next investigated the cell cycle profile, with the aim of identifying possible cell cycle arrest. The state of the cell cycle was assessed by staining dsDNA with Hoechst 33342. Since cells should contain twice as much DNA in G2 phase than in G0/G1, the cell populations in different phases can clearly be distinguished by signal strength. p16 acts at the G1/S checkpoint. One would therefore expect an accumulation of cells where p16 is activated in G1 and a reduction in the S and G2 phases. Surprisingly, the result shows a similar cell cycle profile between all conditions with about 70% of the cells in G1 phase and 15-20% of cells in G2 phase (Fig. 3.3C-D). On the other hand, using this experimental approach, it is impossible to distinguish between G1 and G0 indicative

for a quiescent cell state. To do so one could investigate the expression of M1 subunit of ribonucleotide reductase (M1-RR), a marker for cycling cells⁶¹⁷.

Cellular senescence, i.e. irreversible cell cycle arrest⁶¹⁸, is associated to a number of hallmarks, including p16 expression, but also β -galactosidase activity⁶¹⁹⁻⁶²¹, which is induced by expression of *GLB1*, encoding lysosomal β -D-galactosidase⁶²². Senescence-associated (SA) β -galactosidase activity has been assessed by X-gal staining of the transduced cells. Microscopic images after overnight staining reveal dye accumulation and, therefore, SA- β -galactosidase activity, only in Huh7 cells where p16 has been targeted with dCas9-TET1 (Fig. 3.4).

Taken together, the results validate our approach and suggest that targeted local demethylation of a part of the p16 promoter, is not only possible, but also leads to p16 expression and changes of the cancer cell phenotype, such as slower cell growth and senescence. The extent of the effects, however, still show room for improvement, which could potentially be accomplished by the recruitment of different or additional effectors, or by the insertion of SPF motifs.

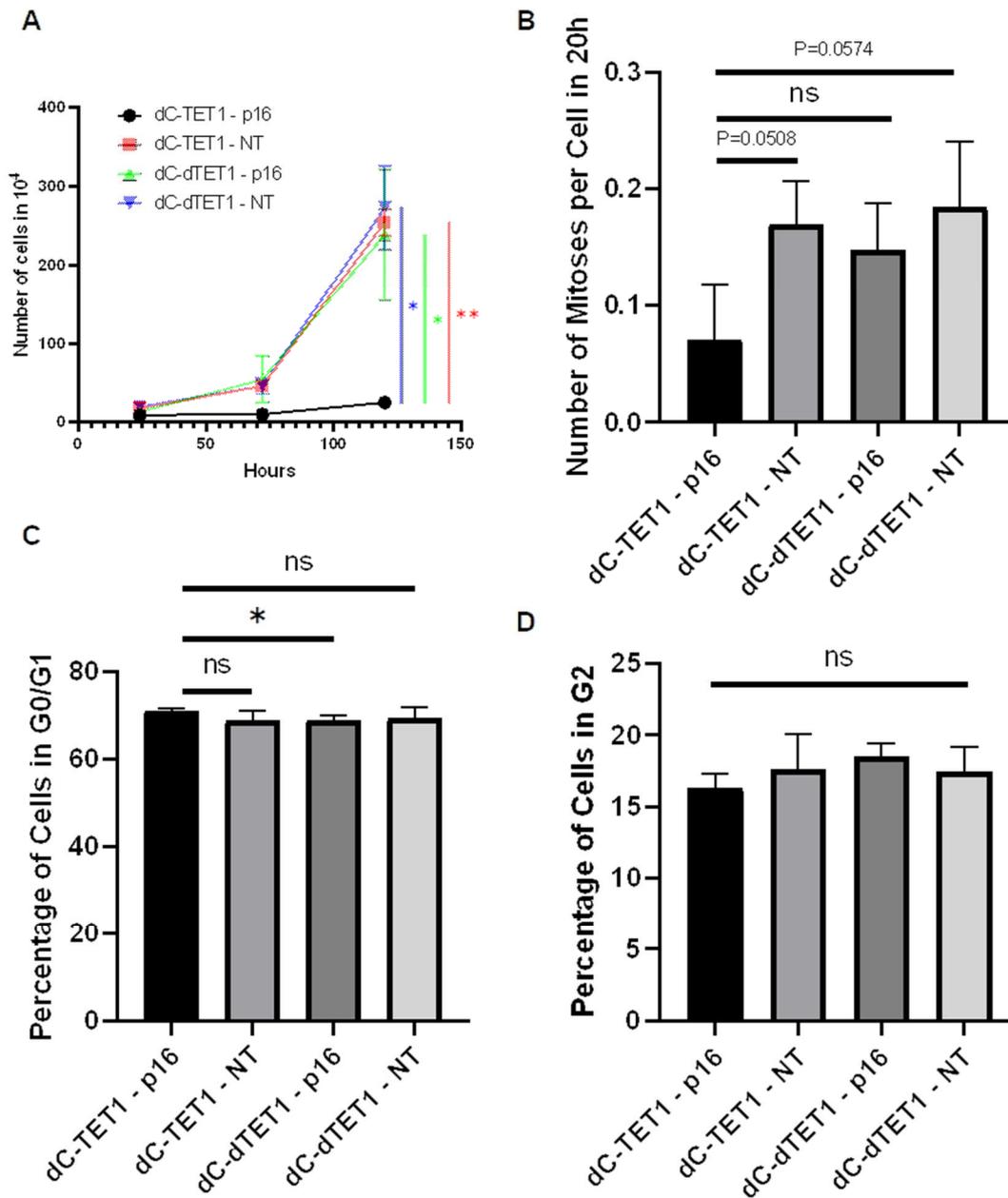


Figure 3.3: Cell Cycle and Proliferation after Epigenome Editing. **A)** Cell count at day one, day three, and day five after plating 10^5 Huh7 cells. dC-TET1-p16: TET1 targeting p16, dC-dTET1-p16: catalytically inactive TET1 targeting p16, dC-TET1-NT: non-targeted TET1, dC-dTET1-NT: non-targeted catalytically inactive TET1. **B)** Cell proliferation assay via life cell imaging. Cells were plated into an 8-well dish and dyed with SiR-Hoechst. Subsequently, the cells were monitored for 20 hours and the mitoses were counted. The number of observed mitoses was divided by the cells initially visible on the plate. **C-D)** FACS analysis of modified Huh7 cells after staining with Hoechst 33342 shows percentages of cells with fluorescent intensities corresponding to G0/G1 (**C**) or G2 phase (**D**).

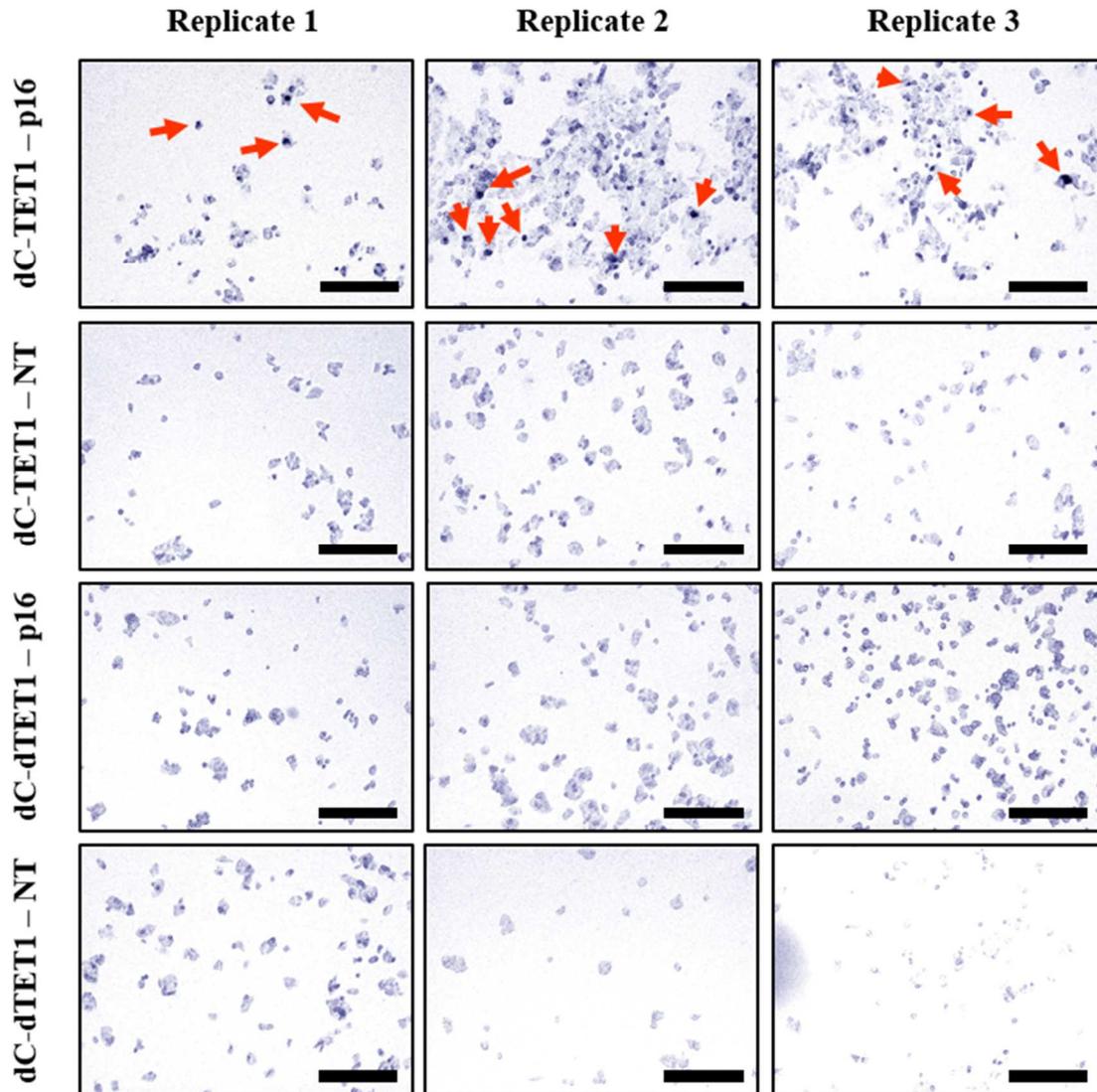


Figure 3.4: Senescence after Epigenome Editing. Fixed epigenome-modified Huh7 cells (dC-TET1 – p16) and their controls were stained with X-gal overnight and pictures from each biological replicate were taken 12 hours later at 200x magnification. Increased β -galactosidase activity is associated with cellular senescence and can be visualized by staining with X-gal (red arrows). Only dC-TET1 – p16 cells show increased senescence. dC-TET1-p16: TET1 targeting p16, dC-dTET1-p16: catalytically inactive TET1 targeting p16, dC-TET1-NT: non-targeted TET1, dC-dTET1-NT: non-targeted catalytically inactive TET1. Scale bar: 200 μ m.

3.1.3 Insertion of the CTCF motif into TSG Promoter Regions

Previous studies have shown that the presence of a binding motif of CTCF, an established SPF, leads to its recruitment and demethylation of an otherwise methylated region^{255,344}. However, this observation has been made in exogenous DNA sequences containing the CTCF motif. Can recruitment of CTCF and subsequent local demethylation be directed by motif insertion into a methylated endogenous regulatory region of the genome? If yes, does this lead to activation of the gene regulated by the targeted regulatory region?

To answer these questions, we chose to target the epigenetically silenced promoters of TSGs *p16* and *Rassf1a* in HCC cell lines, as their activation by targeted demethylation is possible and may be of therapeutic interest as shown above. Specifically, we inserted either wild-type “WT” (TGGCCACCAGGGGGCGCTA) or scrambled “SC” (AATGGCTGGCCACCCGGGG) CTCF consensus binding motifs into the methylated promoter regions in close proximity of the TSS of *p16* and *Rassf1a* in HCC cell line Huh7. Motifs of CTCF and SOX2, which was inserted in a later experiment, were previously tested experimentally for their ability to specifically and efficiently bind to their corresponding PFs, either by ChIP experiments, or DNA/protein microarray and EMSA. Transgenic cell lines with stable insertion of each motif were established. The first insertion was located 141bp upstream of the *p16* TSS. Bisulfite analysis upon this insertion revealed local demethylation close to the insertion site of the CTCF WT motif, which was not the case for SC motif insertion (Fig. 3.5A). In contrast to previous studies that showed large demethylation extent around the insertion site on exogenous fragments (400 bp), the demethylation observed here is confined around the insertion site and did not reach the TSS³⁴⁴. Consistently, this limited demethylation did not yield an increase in mRNA or protein levels of *p16* relative to mock transfected cells for either WT or SC motifs (Fig. 3.5B, C). The reason for the lack of expression could be, in contrast to dCas9TET1 targeting, insufficient demethylation, which did not extend to the TSS. We, therefore, inserted the motifs

closer to the TSS in a distance of 35bp. Bisulfite sequencing shows an even lesser extent of demethylation with only two of the endogenous CpGs demethylated (Fig. 3.5D). Again, this

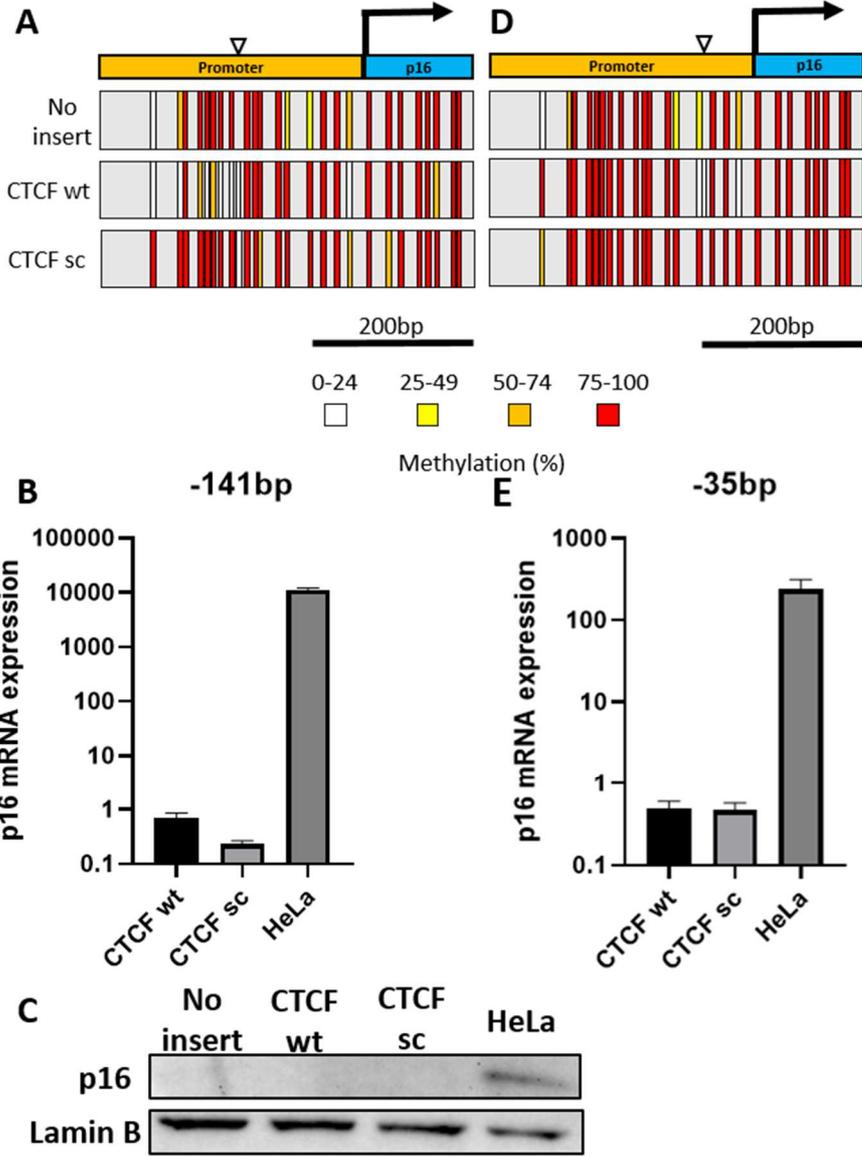


Figure 3.5: CTCF Motif Insertion into the p16 Promoter Region. **A** and **D**) CpG methylation percentages as analyzed by bisulfite sequencing of the p16 promoter after different CTCF motif insertions either -141bp (**A**) or -35bp (**D**) from the p16 TSS. Each vertical bar represents a CpG and the color code indicates the methylation percentage based on at least ten reads. The white triangle depicts the motif insertion site. **B** and **E**) p16 mRNA expression in Huh7 cells after CTCF motif insertion and in HeLa cells (as positive control), relative to mock transfected Huh7 cells. Measurements were made on mRNA from biological triplicates (n = 3) and normalized to the housekeeping gene *SNRPD3*. **C**) Protein levels of p16 and Lamin B (as loading control) after motif insertion -141bp from the TSS.

insertion did not result in an increase in p16 mRNA expression (Fig. 3.5E). This data suggests that insertion of a CTCF motif into the p16 promoter region only leads to limited demethylation that does not seem to be sufficient for activation of gene expression. This, however, does not rule out that other insertion sites or other SPFs could lead to p16 activation (see discussion).

Using the same approach, we also inserted the CTCF motif into the *Rassf1a* promoter. Here, bisulfite analysis shows clear demethylation around the insertion site of the CTCF WT motif in both directions, spanning at least 200 bp (Fig. 3.6A). The CpGs around the TSS however were not demethylated. In contrast, the region around the inserted SC remained methylated. This larger demethylation zone around the WT motif led to 2.4-fold increase of expression on the mRNA level relative to mock transfected Huh7 cells. Nevertheless, the *Rassf1a* expression level was still low in comparison to its expression in HeLa cells that were used as a positive control (17.9 fold higher than Huh7) (Fig. 3.6B). Finally, this increase could not be confirmed at the protein level (Fig. 3.6C). Whether this lack of protein is due to the expression level at the RNA level being too low, or other post-transcriptional events affecting protein production or stability, cannot be concluded at this point.

As with the insertions into *p16*, several options could be used to improve the activation of the gene, such as choice of SPF or the exact motif location and direction. While candidates for the former, which also have to be expressed in Huh7 cells, could be Krüppel-like factor 4 (KLF4) or Forkhead box protein A1 (FOXA1), the possible options for the latter are more numerous. In order to test the possible locations and directions, a streamlined system with a faster, more convenient readout is needed. One such option could be the establishment of reporter cell lines, in which a reporter would be driven by the endogenous TSG promoter. Another issue, which could be circumvented with such a reporter, is that it could prevent the reactivation of the endogenous TSG and thus would not inflict a selective disadvantage such as cell cycle arrest or senescence on the cells. Indeed, it cannot be ruled out that clones with successful *p16* or *Rassf1a*

activation were generated by CTCF motif insertion, but were unable to proliferate and form colonies that could be expanded for the establishment of a cell line.

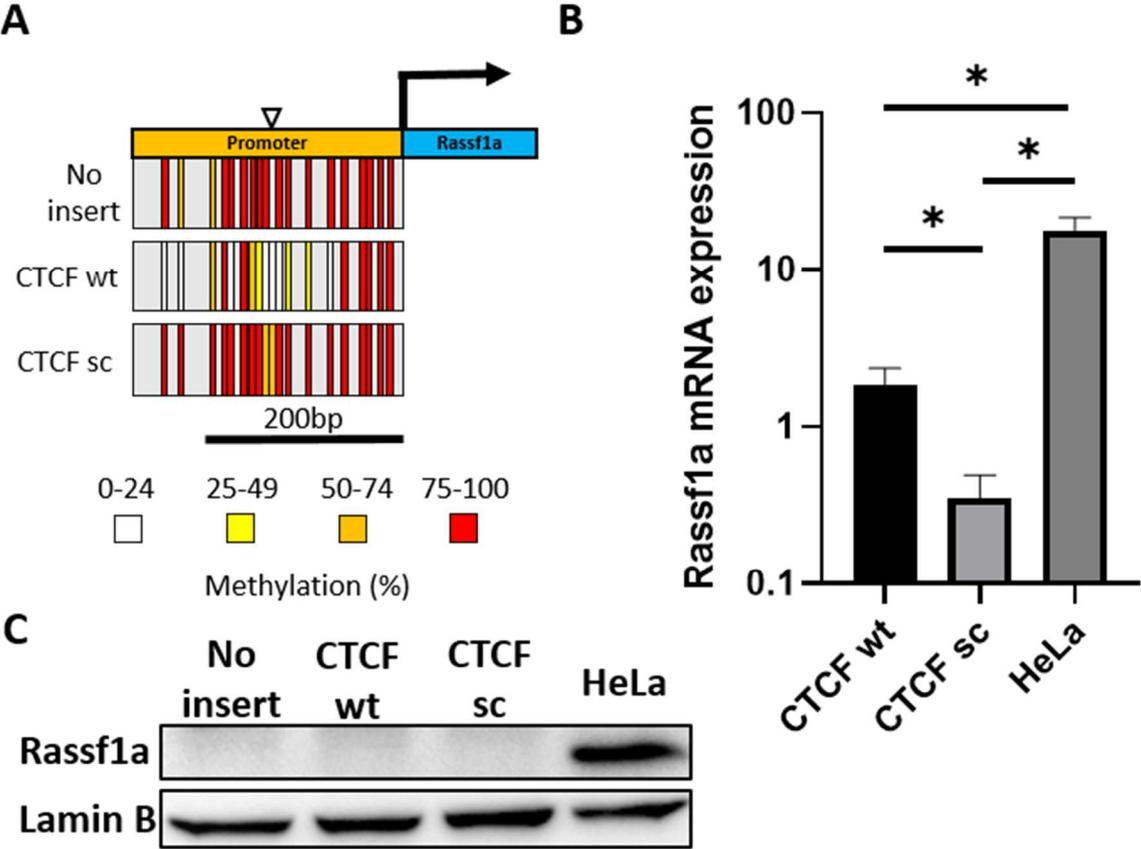


Figure 3.6: CTCF Motif Insertion into the Rassf1a Promoter Region. **A)** Bisulfite profile of the Rassf1a promoter after CTCF motif insertion close to the Rassf1a TSS. Each vertical bar represents a CpG and the color code indicates the methylation percentage based on at least ten reads. The white triangle depicts the motif insertion site. **B)** Rassf1a mRNA expression in Huh7 cells after CTCF motif insertion and in HeLa cells (as positive control), relative to mock transfected Huh7 cells. Measurements were made on mRNA from biological triplicates (n = 3). The housekeeping gene for calculations was SNRPD3. **C)** Protein levels of Rassf1a and Lamin B (as loading control) after motif insertion.

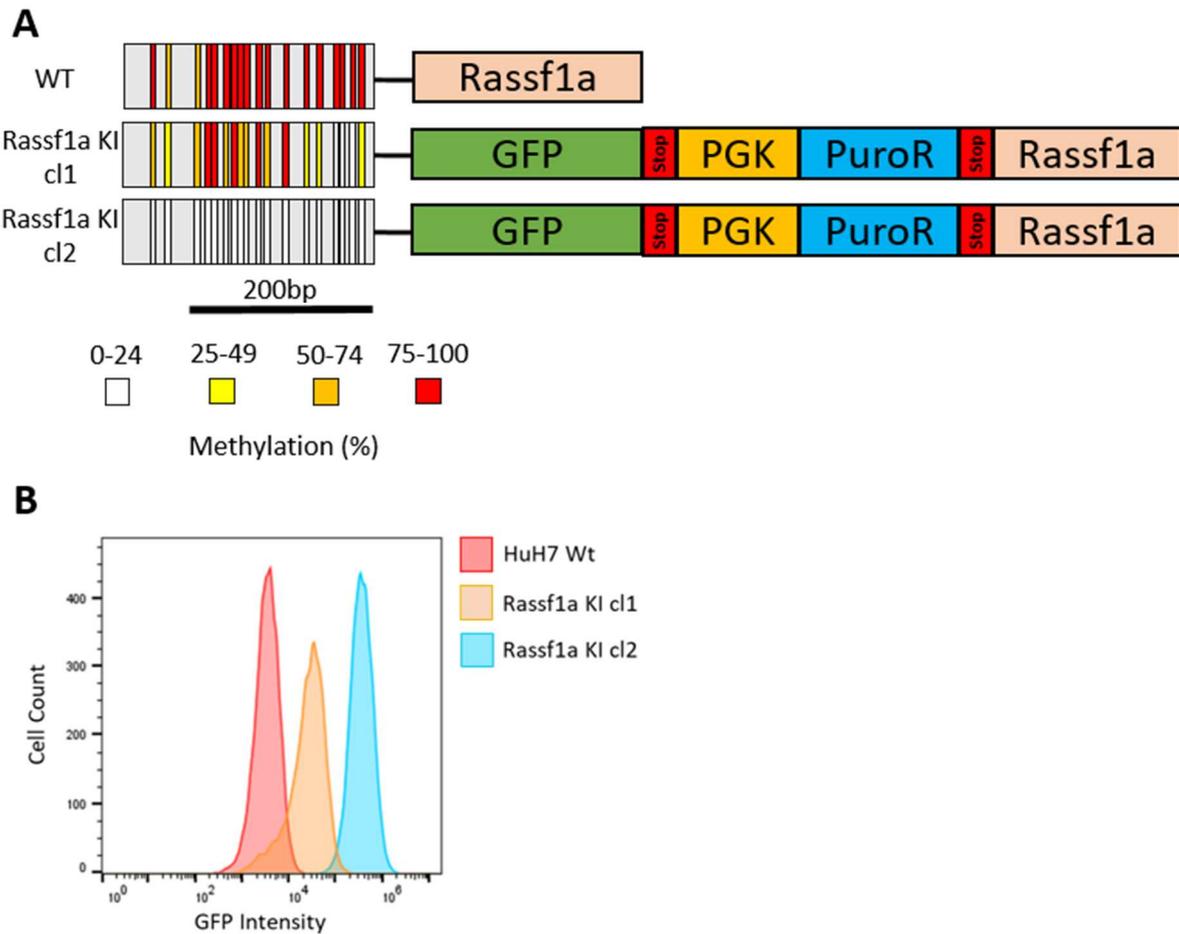


Figure 3.7: Establishment of a Methylation Sensitive Reporter to assess SPF-dependent activation of TSGs. A) Bisulfite profile of the *Rassf1a* promoter before (WT) or after knock in (KI) of a GFP cassette downstream of *Rassf1a* ATG. Note that the 5' homology region for the KI covers the 600bp in front of the Start codon including the entire region analyzed by bisulfite sequencing. This region does not contain an SPF motif, however, it has been inserted unmethylated. Two clones with successful KI were established (*Rassf1a* KI c1 and c2). Each vertical bar represents a CpG and the color code indicates the methylation percentage based on at least ten reads. **B)** Histogram showing the differences in GFP intensity between the two established clones and Huh7 WT cells.

To create such a reporter, we performed a knock-in of a cassette containing the GFP gene and a PGK promoter-driven puromycin resistance gene into the *Rassf1a* locus. The start codon of the GFP replaced the original *Rassf1a* start codon. Furthermore, the GFP contained a stop codon to prevent possible expression of the largely intact *Rassf1a* further downstream. Upon genotyping, two clones with successful knock-ins were identified and expanded. In these resulting cell lines, GFP expression is now under the control of the endogenous promoter of

Rassf1a and its activation could therefore be used as a proxy for the promoter activity. We then checked the methylation status of the promoter in these two cell lines. Bisulfite sequencing revealed that Rassf1a promoter regions are differentially methylated with Rassf1a KI c12 being completely unmethylated and Rassf1a KI c11 partially methylated with lower methylation towards the TSS (Fig. 3.7A). This is most likely due to the fact that the plasmid-based repair template for the knock-in got inserted unmethylated and contained a 5'-region for homology-directed repair, which covers a large portion of the Rassf1a promoter region. Nevertheless, it is puzzling that this region did not get methylated upon insertion, as previous studies reported that insertion of exogenous DNA fragments with the same sequence as endogenous fragments, result in the same methylation status as the endogenous counterpart^{255,599}. Since the region for the knock-in cannot be changed, perhaps *in vitro* methylation of the repair plasmid can ensure that the promoter stays methylated. Regardless of this discrepancy, FACS analysis of the two cell lines indicated GFP intensities that correlate with the methylation status of the Rassf1a promoter (Fig. 3.7B). This result further validates the original hypothesis that demethylation on a local level could indeed lead to expression of Rassf1a. Moreover, as the methylation of the promoter in Rassf1a KI c11 is partially reduced and the GFP expression only slightly increased, it could represent a sensitized form in which a small further decrease in methylation could have a higher impact on expression. This clone could therefore, in principle, be used for our purpose of screening for the best SPF insertion strategy that could lead to activation of Rassf1a.

Taken together, CTCF motif insertion into defined locations of endogenous promoter regions of *p16* and *Rassf1a* has resulted in local DNA demethylation, but no robust reactivation of gene expression. The use of a cell line with a reporter system that allows to assess best insertion strategy for the reactivation of Rassf1a promoter will be considered in the future.

3.1.4 Insertion of SPF motifs into a Methylation Reporter

While biologically interesting, reactivating epigenetically silenced TSGs has a technical downside. Indeed, it could be technically challenging to detect the cells with active *p16* or *Rassf1a* as this could present a selective disadvantage compared to other cells, and, therefore, would not form colonies. We may have therefore excluded potentially ‘successful’ SPF insertions with subsequent gene activation from our results. Furthermore, a more convenient and direct readout for gene reactivation alternatively to qPCR and Western blot is desired. While the fluorescence reporter for *Rassf1a* in Huh7 cells was under development, we targeted the already established *Dazl-Snrpn* methylation reporter developed by the Jaenisch lab⁶¹³. The reporter consists of a small promoter of *nuclear ribonucleoprotein polypeptide N (Snrpn)* and a downstream GFP, both of which were inserted into the *deleted in azoospermia-like (Dazl)* locus of mouse embryonic stem cells (mESCs), retaining part of the *Dazl* promoter upstream of the reporter cassette. *Snrpn* is an imprinted gene whose methylation status was shown to be sensitive to the methylation status of its surrounding locus. Moreover, it is expressed in most tissues making it a good candidate for a methylation sensitive reporter system^{623,624}. *Dazl*, on the other hand, is a germ-line specific gene that is methylated in mESCs⁶²⁵. Importantly, the reporter was shown to be stably methylated in the *Dazl* locus and successfully activated by recruitment of dCas9-TET1⁵⁷².

To confirm activation of the reporter following demethylation, *Dazl-Snrpn-GFP* V6.5 mESCs were treated with 1 μ M 5-aza for 7 and 14 days. FACS analysis shows a clear shift of the treated cell population towards higher GFP intensity (Fig. 3.8A-B). However, despite the shift, there is still a substantial overlap in intensity shared by both populations, making it difficult to determine the absolute amount of GFP-positive cells. Alternatively, one can calculate the median or mean fluorescence intensities (MFI) and their ratio between treated and untreated cells. In case the distributions are symmetrical, median and mean should be almost identical.

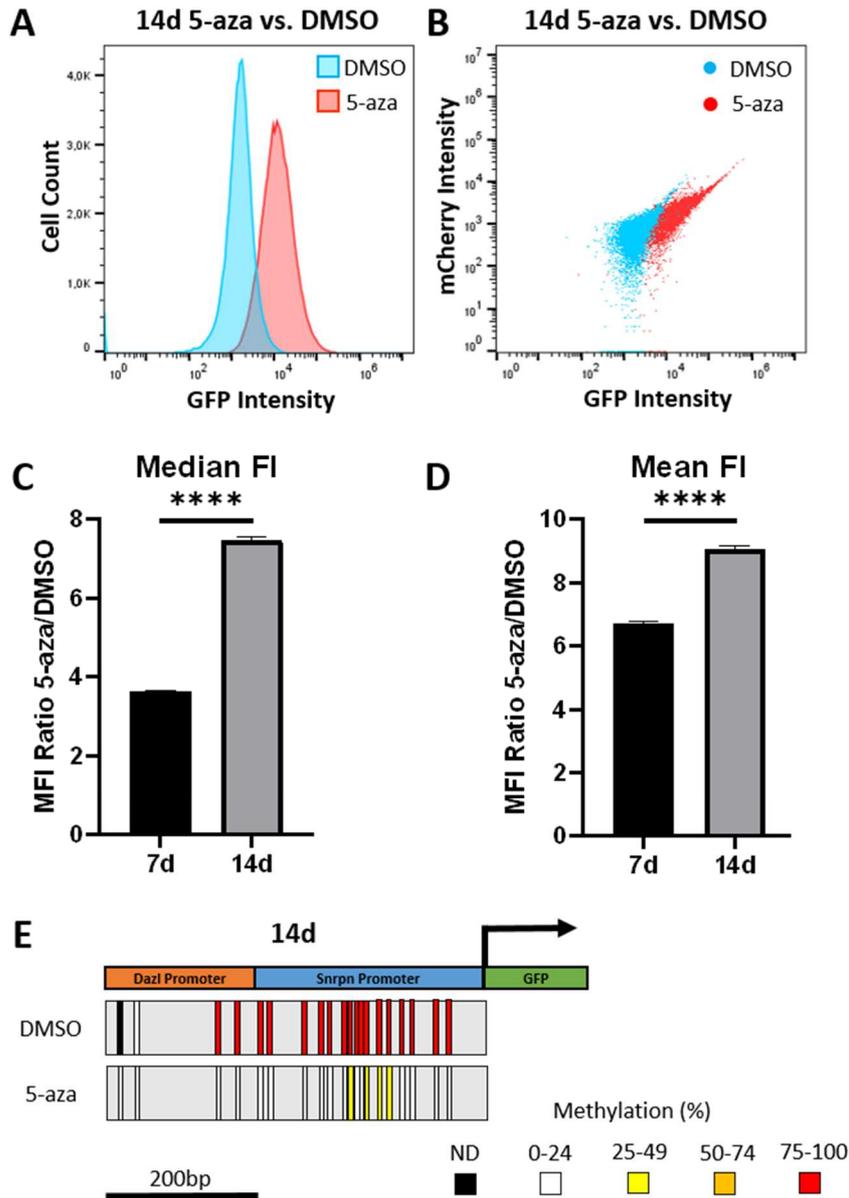


Figure 3.8: Activation of the Dazl-Snrpn Methylation Reporter with 5-aza. **A)** Representative histogram showing the difference in GFP intensity between Dazl-Snrpn-GFP V6.5 mESCs either treated with DMSO or $1\mu\text{M}$ 5-aza for 14 days. **B)** Representative scatter plot showing the population shift towards higher GFP intensity in Dazl-Snrpn-GFP V6.5 mESCs either treated with DMSO or $1\mu\text{M}$ 5-aza for 14 days. **C)** Median fluorescence intensity ratios of GFP between cells treated with 5-aza and DMSO was measured by FACS after seven days and 14 days of treatment ($n = 3$). **D)** Mean fluorescence intensity ratios of GFP between cells treated with 5-aza and DMSO was measured by FACS after seven days and 14 days of treatment ($n = 3$). **E)** Bisulfite profile of the Dazl-Snrpn promoter after 14 days of 5-aza treatment. Each vertical bar represents a CpG and the color code indicates the methylation percentage based on at least ten reads. The white triangle depicts the motif insertion site.

However, in case a distribution is skewed either positively or negatively, the median will be closer to the mode, representing more ‘typical’ events. For the following experiments, the ratios for both median and mean fluorescence intensity were calculated. The MFI ratios are increased after 5-aza treatment with further increase between 7 and 14 days (Fig. 3.8C-D), suggesting that longer exposure to demethylating agents increases gene activation. Importantly, these expression differences correlate well with changes in the methylation landscape of the reporter that were confirmed by bisulfite sequencing (Fig. 3.8E). This set of experiments confirm the utility of using the *Dazl-Snrpn* as methylation-dependent reporter for gene expression.

Next, we inserted SPF motifs in either the *Dazl* part or the *Snrpn* part of the promoter region by CRISPR/Cas9 (Fig. 3.9A). The first SPF we tested was CTCF. Several clones with CTCF motif insertions either as WT or SC as well as in forward (fwd) or reverse (rev) were established. Of all insertions, only one clone with the CTCF WT rev motif inserted into the *Dazl* part showed a significant shift in GFP intensity (Fig. 3.9B-D). While being significant, this shift is relatively small in comparison with the one achieved with 5-aza treatment (Fig. 3.8A-D). This insertion also caused local demethylation that was restricted to only a few CpGs around the target site (Fig. 3.9E). Surprisingly, demethylation, albeit to a bit lesser extent, was also observed around CTCF SC motif at the same insertion site. As for insertions into the *Snrpn* part, none of which seem to induce GFP expression, the bisulfite profile between two clones containing the CTCF WT fwd motif, i.e. both having the same genotype, differs drastically (Fig. 3.9F). While WT fwd clone 1 shows demethylation around the insertion site, WT fwd clone 2 seems to retain full methylation. These results indicate that CTCF motif insertion can lead to reactivation of silenced genes under certain circumstances. However, both location and direction may play a role. Also the extent on how much DNA demethylation, is involved in the reactivation is uncertain, since one insertion into the *Snrpn* part leads to widespread demethylation, but no change in GFP intensity. Perhaps the demethylation of

certain key CpGs is more important than demethylation over a larger region. Furthermore, since there are two clones with the exact same insertion, but different methylation profiles, it is unsure whether the demethylation is due to differential CTCF recruitment, or other unknown factors potentially involving the repair process of the region. ChIP experiments investigating differential CTCF enrichment in these clones may in part answer this question.

Another recently identified SPF is SOX2³⁴⁴, a prominent pluripotency factor expressed in mESCs. Therefore, SOX2 represents an additional suitable candidate for the assessment of SPF-dependent transcription activation. As in the earlier experiment, we inserted as WT (fwd or rev) or SC SOX2 motifs into the *Dazl* or *Snrpn* part of the reporter. Of the established clones only two, WT fwd 2 in *Dazl* and WT fwd in *Snrpn*, show a shift in fluorescence intensity (Fig. 3.10A-C). While the former does not show a significant shift, the latter appears to contain a population, that is statistically different from the original population in mock transfected cells (Fig. 3.10C). This pattern results in a significant shift for mean fluorescence intensity, but not for median fluorescence intensity (Fig. 3.10A-B). Insertion of the SOX2 WT fwd motif into the *Dazl* part also leads to similar demethylation in both established clones (Fig. 3.10D). Although SOX2 WT fwd cl2 seems to have higher GFP expression, the difference is not significant. Surprisingly, the promoter seems to maintain complete methylation when the SOX2 WT motif is inserted into the *Snrpn* part despite the increase in GFP expression, while moderate demethylation was observed when the SC motif was inserted in the same location (Fig. 3.10E). Sequencing of the region did not reveal genomic heterogeneity, with all 36 analyzed sequences containing the SOX2 WT motif and no further abnormalities. These results raise the questions whether the insertion of the SOX2 motif leads to demethylation, and whether demethylation is necessary for reactivation of gene expression.

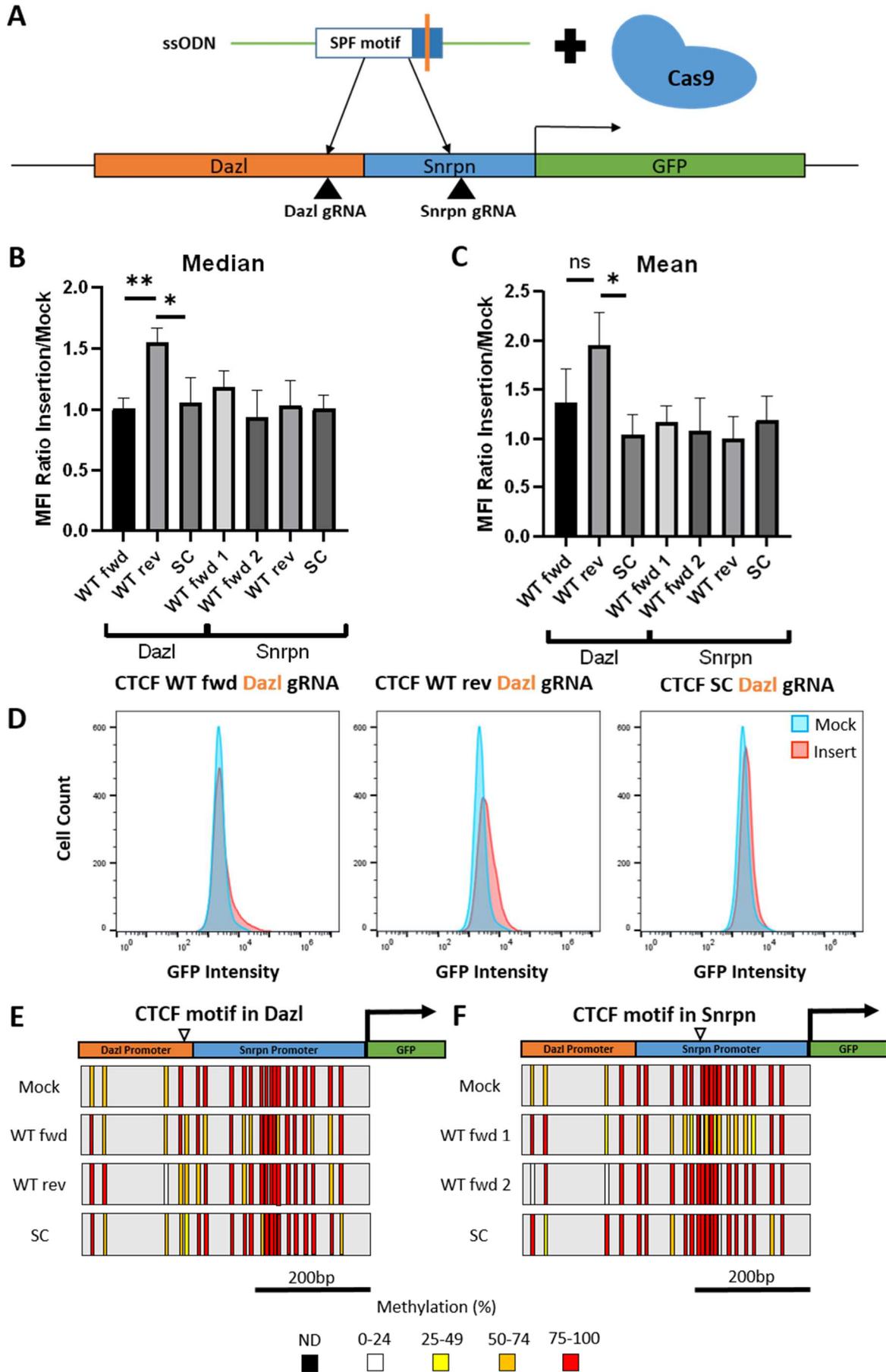


Figure 3.9: Targeted Insertion of CTCF Motifs into the Dazl-Snrpn Promoter. **A)** Schematic representation of CTCF motif insertion into the Dazl-Snrpn-GFP reporter. **B)** Median fluorescence intensity ratios of GFP between clones with insertions and mock transfected Dazl-Snrpn-GFP V6.5 mESCs. GFP intensity was measured by FACS for biological triplicates of each clone (n = 3) **C)** Mean fluorescence intensity ratios of GFP between clones with insertions and mock transfected Dazl-Snrpn-GFP V6.5 mESCs. Snrpn WT fwd 1 and 2 are two distinct clones with the same insertion. **D)** Representative histograms of clones with the highest MFI ratio difference and their SC motif insertion counterparts. **E-F)** Bisulfite profile of the Dazl-Snrpn promoter after different CTCF motif insertions into either the Dazl region (**E**) or the Snrpn region (**F**). Each vertical bar represents a CpG and the color code indicates the methylation percentage based on at least ten reads. The white triangle depicts the motif insertion site.

Finally, we also assessed GFP expression of the clones showing the strongest intensity shifts on the mRNA level. All tested clones with a WT SPF motif show higher GFP expression (Fig. 3.11), which corresponds to the results observed during FACS.

Taken together, SPF motif insertion into the Dazl-Snrpn reporter induces gene expression in some cases. However, having activation without demethylation and differential activation or demethylation between clones with the exact same insertion, suggest that other mechanisms, possibly CRISPR- or DNA repair-related, are involved, making SPF insertion not a robust technique for gene reactivation as of yet.

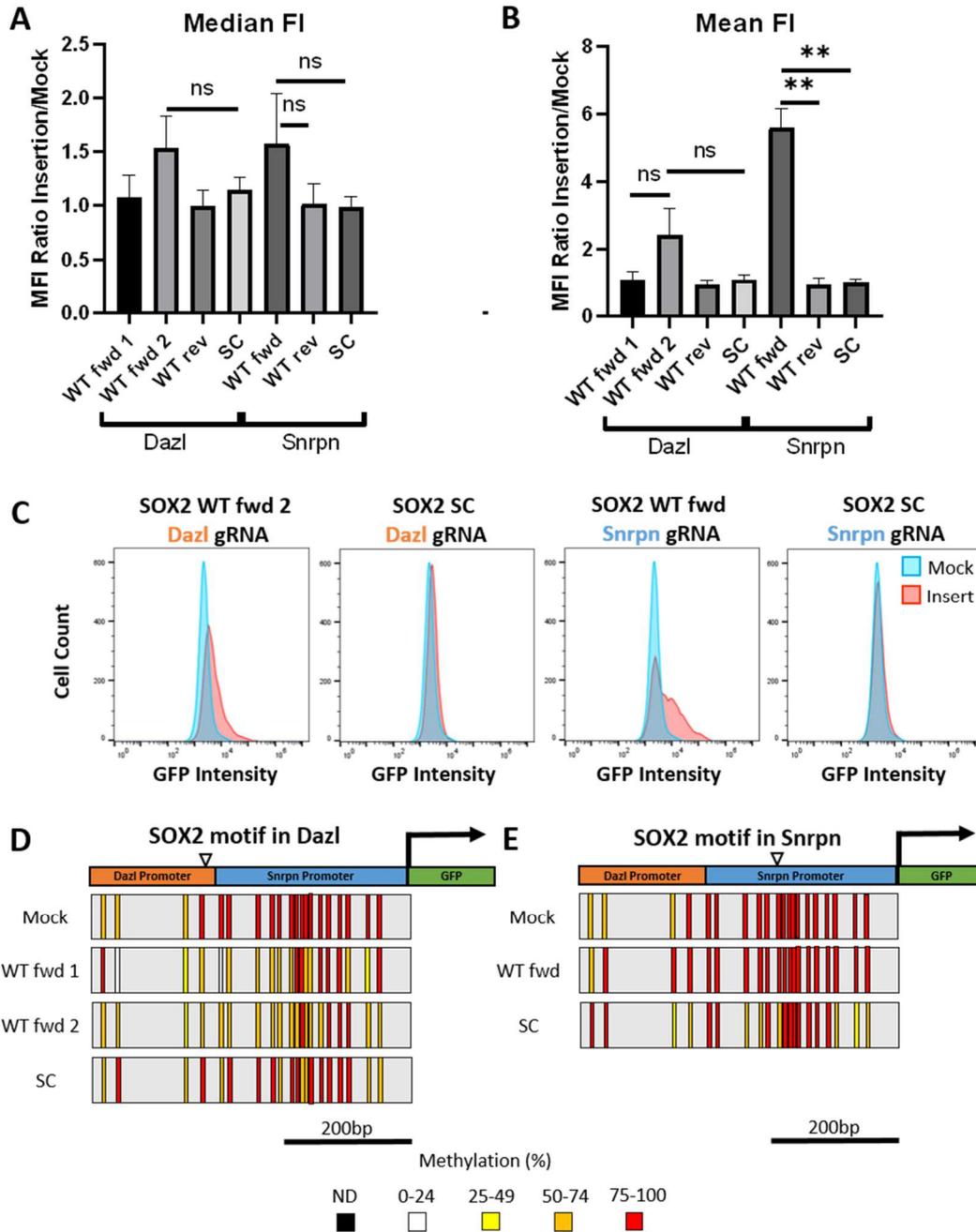


Figure 3.10: Targeted Insertion of a SOX2 Motif into the Dazl-Snrpn Promoter. **A)** Median fluorescence intensity ratios of GFP between clones with insertions and mock transfected Dazl-Snrpn-GFP V6.5 mESCs. GFP intensity was measured by FACS for three passages of the same clone ($n = 3$) **B)** Mean fluorescence intensity ratios of GFP between clones with insertions and mock transfected Dazl-Snrpn-GFP V6.5 mESCs. Dazl WT fwd 1 and 2 are two distinct clones with the same insertion. **C)** Representative histograms of clones with the highest MFI ratio difference and their SC motif insertion counterparts. **D-E)** Bisulfite profile of the Dazl-Snrpn promoter after different SOX2 motif insertions into either the Dazl region (**D**) or the Snrpn region (**E**). Each vertical bar represents a CpG and the color code indicates the methylation percentage based on at least ten reads. The white triangle depicts the motif insertion site.

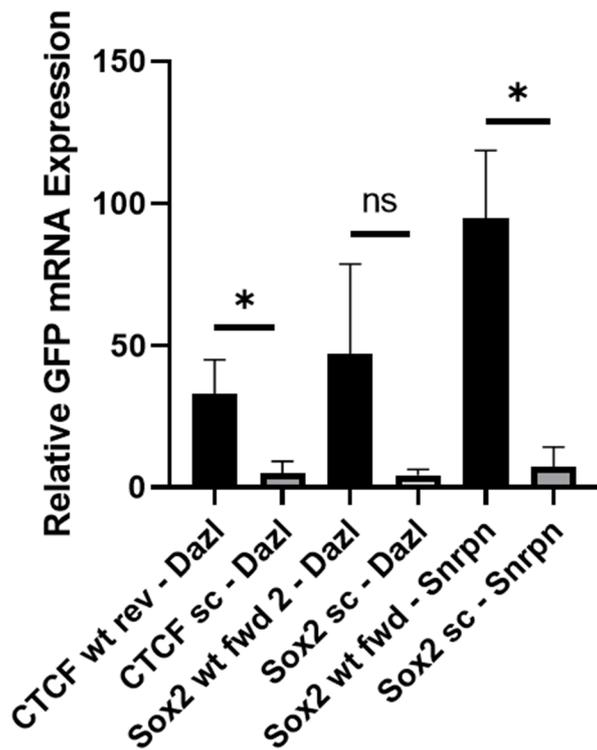


Figure 3.11: GFP Expression on the mRNA Level after SPF Motif Insertion. GFP mRNA expression relative to mock-transfected Dazl-Snrpn-GFP V6.5 mESCs. Clones with WT motif insertions most likely to induce GFP expression as well as their SC motif counterparts, were assessed. Measurements were performed on RNA from three different passages of the same clone (n = 3), which has been extracted on the same days as the previously described FACS experiments (error bars, s.d.). SOX2 wt fwd 2 – Snrpn refers to the clone, which previously showed higher increase in GFP intensity.

3.2 Cell Fate of Adipose Tissues: New Targets for Epigenome Editing

- In collaboration with the Trajkovski laboratory, University of Geneva, Switzerland.
- Contributions: Nicolas Hafner performed genomic DNA extraction, WGBS, 5-aza treatment and differentiation of 3T3-L1 pre-adipocytes, as well as interpretation of the results; Nicolas Suarez-Zamorano performed the *in vivo* experiments, tissue extraction, RNAseq, and qRT-PCR; Victor Ythier performed WGBS data analysis; Silas Kieser performed RNAseq data analysis.

Adipose tissues come in different forms and functions⁶²⁶. While white adipose tissue (WAT) constitutes the majority of adipocytes in adult mammals and is responsible for energy storage, brown adipose tissue (BAT) regulates non-shivering thermogenesis^{401,402}, i.e. it expends energy rather than storing it. Active BAT is associated with a healthy phenotype, correlating with a lower body mass index (BMI) and a lower risk for diabetes⁴⁰³. Interestingly, it has been shown that after cold exposure or caloric restriction, a third type of adipose tissue called “beige” emerges within WAT⁴⁰⁴⁻⁴⁰⁶. Beige adipocytes have similar properties to brown adipocytes. This process is therefore called “browning”. Browning has been consistently shown to occur in subcutaneous white adipose tissue (SAT), which is located underneath the skin, but not visceral white adipose tissue (VAT), which is located between the inner organs. Recently however, it has been observed that cold exposure in combination with gut microbiota depletion also leads to browning of VAT, which leads to a further increase in glucose tolerance and insulin sensitivity⁶²⁷. These ‘beige’ adipocytes emerging from VAT, while sharing traits with SAT-related beige and brown adipocytes, show also differences in gene expression. For example, *Uncoupling Protein 1 (UCPI)*, a marker gene for the brown phenotype, responsible for heat generation, does not show higher expression in this emerging cell type (unpublished data, Suarez-Zamorano et al). Targeting browning in VAT over SAT might also be clinically more

relevant since VAT is associated with chronic inflammation negatively impacting metabolic diseases.

The precise mechanisms and origin underlying the browning phenomenon are still not fully understood. However, because of the positive contribution of active brown and beige adipose tissue towards glucose uptake and insulin sensitivity, controlled transdifferentiation of white adipocytes into beige adipocytes may present a new therapeutic strategy against diabetes and obesity. Since cell fate is largely determined by epigenetic mechanisms, such as DNA methylation, the identification of differentially methylated regions (DMRs) related to regulatory elements of differentially expressed genes (DEGs) before and after adipocyte browning in VAT, may present targets for epigenome editing. We therefore aim at identifying these elements *in vivo* by using mouse models. Furthermore, we aim to develop a cell culture system for fast and convenient assessment of the newly identified targets and their possible activation.

3.2.1 Identification of Differentially Expressed Genes and Differentially Methylated Regions Upon Browning in VAT Adipocytes

To induce browning of VAT, mice were kept for 30 days at 6°C for cold exposure and received an antibiotic cocktail, administered in the drinking water (CAx). Both CAx treated mice and mice kept at room temperature (RT) without antibiotics administration were sacrificed and RNA and DNA were extracted from perigonadal VAT (Fig. 3.12A). RNA sequencing (RNAseq) analysis revealed 152 DEGs with fairly even distribution between upregulated and downregulated genes upon browning (Fig. 3.12B). Whole genome bisulfite sequencing (WGBS) was then performed and resulted in a mean read coverage of 18.87 and 20.20 per CG for CAx-treated and RT samples, respectively. Using a sliding window approach, we identified 419 DMRs between CAx-treated and RT-control VAT, most of which (356) were hypomethylated after CAx treatment (Fig. 3.12C). Importantly, none of the DMRs were in the promoter region or in close proximity to the TSS of the DEGs (Tab. 1). This is not surprising, since changes in gene expression during cellular differentiation is mostly driven by differential activation of distal regulatory elements. These could be located at different distances from TSS, as well as up- or downstream of their target gene(s). Therefore, an overlap of these regions with our identified DMRs cannot be ruled out. For this new, emerging cell type, however, tissue-specific enhancers are still unknown. We, therefore, looked at previously published enhancer-specific histone marks in WAT and BAT, which were identified by chromatin immunoprecipitation (ChIP)⁶²⁸. Peaks for H3K27ac and H3K4me1 enrichment in these cells in different distances to the TSS of the DEGs were identified (Tab. 2-3). The results identified six DEGs with both DMRs and enhancer-specific histone marks in proximity to their TSS (Tab. 4).

Table 1: DMRs close to DEGs (CAx vs RT). The first column depicts the maximal distance from TSS of DEGs used for the identification of related DMRs. The second column depicts the number of DEGs that have a neighboring DMR within the defined distance and their names are depicted in the third column. The color code refers to whether the genes were upregulated (red) or downregulated (blue) after CAx.

Distance from TSS (kb)	Number of DEGs	DEG names
1	0	-
2	1	Thbs1
3	2	Thbs1, Itgax
5	4	Thbs1, Itgax, Acly, Acs11
10	8	Thbs1, Itgax, Acly, Acs11, Scd2, Pdha1, Serpine1, Serpina3k

Two of the genes have enhancer-specific histone marks in WAT, two in BAT, and two in both (Tab. 4; blue: lower expression, red: higher expression after CAx). The identified genes showing nearby enhancer marks in both WAT and BAT are *Acyl-CoA synthetase-1 (Acs11)*, which is responsible for the direction of fatty acids (FAs) towards β -oxidation and cold thermogenesis⁶²⁹, and *pyruvate dehydrogenase E1 subunit alpha 1 (Pdha1)*, which converts pyruvate into acetyl-CoA⁶³⁰. The genes with nearby enhancer marks only in BAT are *ATP citrate lyase (Acly)*, which is involved in FA synthesis^{631,632}, and *Serpin Family E Member 1 (Serpine1)*, which is responsible for blood clot degradation by inhibiting plasminogen activator (PLAT)⁶³³. The genes with enhancer marks only in WAT are *Stearoyl-CoA desaturase-2 (Scd2)*, which is involved in lipid synthesis⁶³⁴, and *Thrombospondin 1 (Thbs1)*, which is a component of the extracellular matrix and cell adhesion machinery⁶³⁵. Interestingly, while none of the DMRs directly overlap with the putative enhancers as defined by the related histone modification (Fig. 3.13 and Fig. 3.14), all DMRs associated with the above-mentioned genes and distal regulatory elements are hypomethylated after CAx treatment and located at the 3' end of the gene.

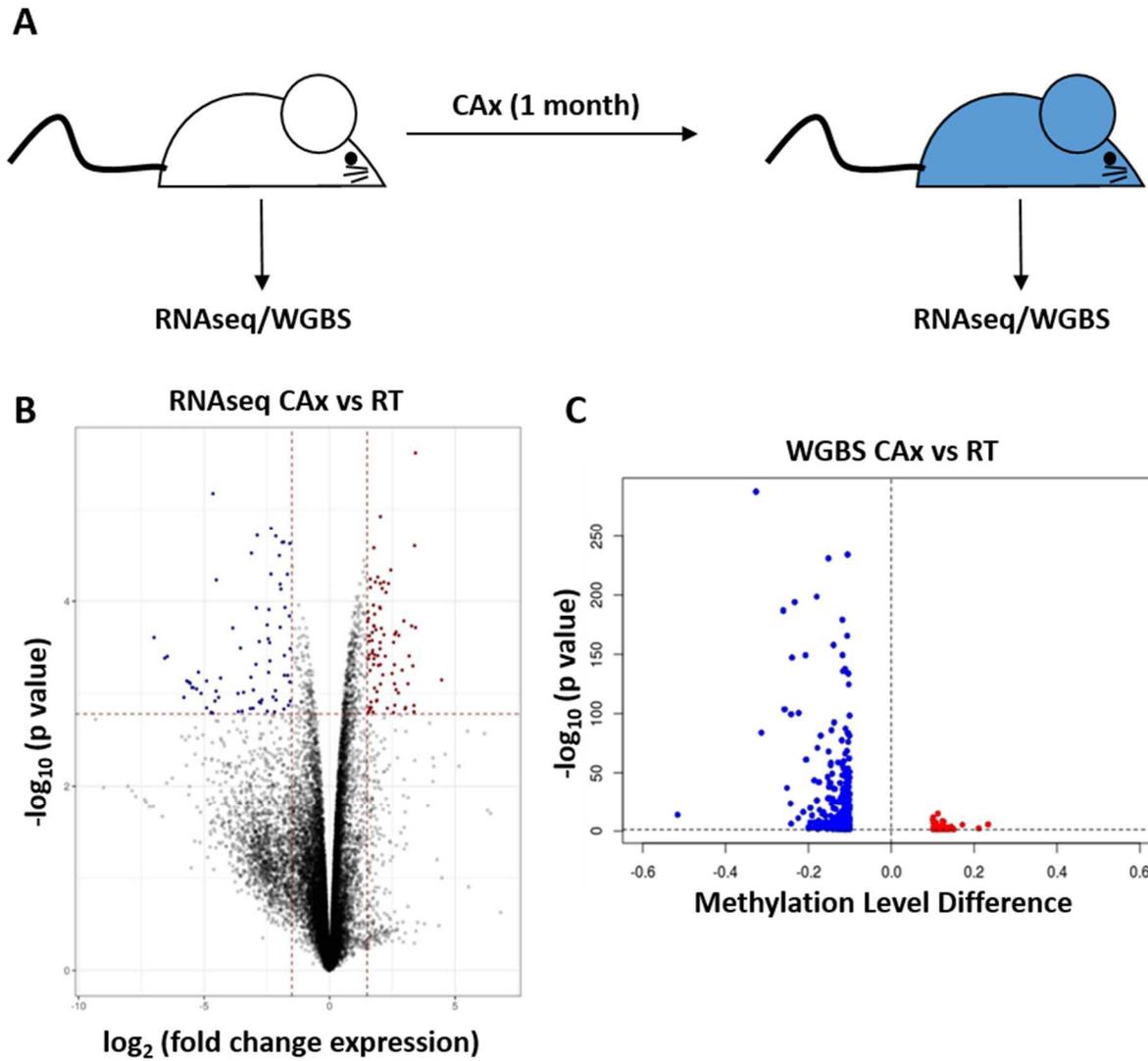


Figure 3.12: Identification of DMRs and DEGs after CAx Treatment. **A)** Schematic representation of cold and antibiotics (CAx) treatment of mice. **B)** Volcano blot of RNA-seq showing DEGs after CAx treatment with a fold change in expression of at least 1.5. downregulated genes after CAx treatment are depicted in blue, while upregulated genes are depicted in red. **C)** Volcano blot of WGBS showing DMRs after CAx treatment with a methylation level (ML) difference of at least 10%. Hypomethylated DMRs are depicted in blue, hypermethylated regions are depicted in red.

Table 2: Enhancer-specific Histone Marks near DEGs (CAx vs RT) in Adipose Tissues

Distance from TSS (kb)	Number of peaks	Number of DEGs (152 max)
H3K27ac peaks near DEGs (CAx vs RT) in BAT		
1	330	82
2	398	85
3	448	89
5	569	93
10	941	105
1000	34275	151
H3K27ac peaks near DEGs (CAx vs RT) in WAT		
1	135	39
2	168	41
3	183	45
5	215	51
10	287	59
1000	8364	147
H3K4me1 peaks near DEGs (CAx vs RT) in BAT		
1	87	38
2	145	63
3	200	76
5	314	90
10	627	113
1000	33632	151
H3K4me1 peaks near DEGs (CAx vs RT) in WAT		
1	235	81
2	347	96
3	412	104
5	563	113
10	965	125
1000	50777	152

Table 3: DEGs with both H3K27ac and H3K4me1 peaks close to their TSS. The first column depicts the maximal distance from TSS of DEGs where both enhancer-specific histone marks can occur. The second column depicts the number of DEGs that have both histone marks within the defined distance in BAT. The second column depicts the number of DEGs that have both histone marks within the defined distance in WAT.

Distance from TSS (kb)	Number of DEGs in BAT	Number of DEGs in WAT
1	26	26
2	50	36
3	60	42
5	73	50
10	94	58
1000	151	147

Table 4: DEGs with both DMRs and Enhancer-Specific Histone Marks in Proximity to their TSS. The first column depicts the maximal distance from TSS of DEGs where both enhancer-specific histone marks as well as DMRs can occur. The second and third columns depict the names of the DEGs fulfilling these requirements within the given distance. The color code refers to whether the genes were upregulated (red) or downregulated (blue) after CAx.

Distance from TSS (kb)	DEGs with enhancer-specific histone marks in BAT	DEGs with enhancer-specific histone marks in WAT
1	-	-
2	-	Thbs1
3	-	Thbs1
5	Acs11 , Acly	Acs11 , Thbs1
10	Acs11 , Acly , Pdha1 , Serpine1	Acs11 , Thbs1 , Scd2 , Pdha1



Figure 3.13: DEGs related to DMRs and Enhancer-Specific Histone Marks in BAT. Genomic regions covered by the DEGs are highlighted in yellow. Below are shown, in order, the ChIP-seq tracks of H3K4me1 in BAT, H3K4me1 in WAT, H3K27ac in BAT and H3K27ac in WAT. The lowest tracks depict the ratio of CpG methylation measured by WGBS in RT VAT and CAX VAT, in order. The DMRs are highlighted at the very bottom, while ChIP-Seq peaks are highlighted below the corresponding ChIP-Seq experiment. The genes displayed are (A) *Acly*, (B) *Acs11*, (C) *Pdha1*, (D) *Serpine1*.

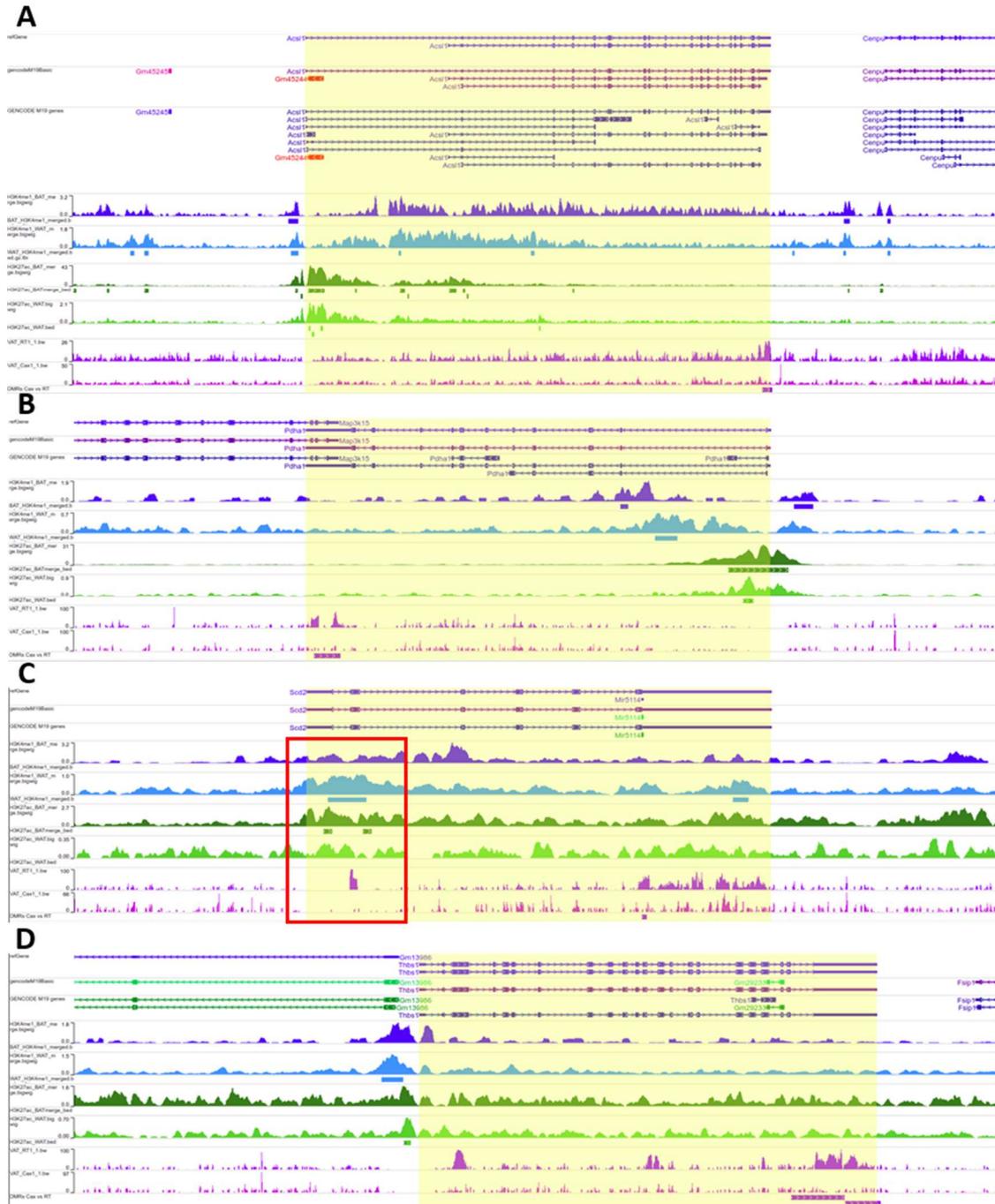


Figure 3.14: DEGs with both DMRs and Enhancer-Specific Histone Marks in WAT. Genomic regions covered by the DEGs are highlighted in yellow. Below are shown, in order, the ChIP-Seq tracks of H3K4me1 in BAT, H3K4me1 in WAT, H3K27ac in BAT and H3K27ac in WAT. The lowest tracks depict the ratio of CpG methylation measured by WGBS in RT VAT and CAx VAT, in order. The DMRs are highlighted at the very bottom, while ChIP-Seq peaks are highlighted below the corresponding ChIP-Seq experiment. The genes displayed are (A) *Acs11*, (B) *Pdha1*, (C) *Scd2*, (D) *Thbs1*. The red box in (C) highlights a potential DMR, which may have escaped our parameters for DMR identification.

Therefore, one would expect that these DMRs should be correlating with upregulation of genes upon CAX treatment. Indeed, of the identified related-DEGs, only *Serpine1* and *Thbs1* are downregulated in VAT from CAX-treated mice, while the other genes are upregulated. Interestingly, *Serpine 1* and *Thbs1* are also the only genes that are not directly involved in FA metabolism, but rather platelet biology and cells adhesion.

While these findings could be interesting, their biological significance, as so far, remains elusive. Indeed, in all the cases above, DMRs do not seem to overlap directly with enhancer marks. Therefore, it is difficult to conclude whether these DMRs have a direct effect on the neighboring regulator elements. Furthermore, a quick look on the WGBS tracks reveals several apparent DMRs that were not identified as such. For example, a region overlapping with enhancer-specific histone marks in *Scd2* and loosing methylation after CAX treatment seems not having been recognized by the set criteria (Fig. 3.14C, red box). In fact, the standard criteria that we applied for DMR identification seem to be too stringent and probably not adapted for a heterogeneous population, such as the VAT (see discussion). Changing the criteria for DMR identification may identify additional DMRs overlapping with enhancer marks.

Taken together, the results from WGBS identified DMRs between CAX-treated and RT VAT, with a strong shift towards hypomethylation. While most of the genes showing both enhancer-specific marks and DMRs are involved in FA metabolism, no apparent targets for epigenome editing could be identified, due to a lack of overlap between the two. Other epigenetic mechanisms beyond DNA methylation, such as histone modifications, may also be considered for target identification. Furthermore, heterogeneity of cell types within the sample, and pooling of multiple mice to obtain a sufficient amount of DNA for WGBS, may have prevented the identification of DMRs with lower methylation level difference (see discussion).

3.2.3 Differentiation of 3T3L1-preadipocytes under the Influence of 5-aza

While, ultimately, the assessment of potential epigenome editing targets for browning of WAT, which includes testing of glucose uptake and insulin sensitivity, needs to take place in mice, prior investigation of the targets on development and transdifferentiation in a more convenient cell culture system is desired. Here, we take advantage of the differentiation potential of 3T3-L1 pre-adipocytes, which initially show fibroblast-like morphology, but have the known ability to differentiate into white adipocyte-like cells, synthesizing and accumulating triglycerides as well as adopting a round morphology including lipid droplets⁶³⁶. Aside from this ‘whitening’ protocol, we also applied a browning protocol, adopted from the differentiation protocol of primary immortalized brown adipocytes (PIBA)⁶³⁷, which requires indomethacin, isobutylmethylxanthine, dexamethasone, insulin, triiodothyronine or rosiglitazone, and isoproterenol⁶³⁸. Since 3T3-L1 pre-adipocytes treated with either protocol are morphologically indistinguishable, but show different gene expression profiles, especially in regards to FA metabolism and some browning markers, the idea of this experiment is to differentiate the cells with either protocol, and test the ability of favoring browning by epigenome editing. Since no apparent targets for epigenome editing have been identified as of yet, and because during WGBS a predominant shift towards hypomethylation after CAx-treatment has been observed (Fig. 3.12B), we hypothesized that 5-aza treatment, a general demethylating agent, may enhance the browning potential of these cells. Therefore, we first tested our differentiation protocol in combination with 5-aza treatment. The treatment was applied at two different time points during the differentiation protocol, which lasts seven days (Fig. 3.15A). The first treatment, called START, was applied during the first two days of differentiation, while END treatment was applied during the last two days. Two concentrations of 5-aza together with a DMSO control were tested. Subsequently, RNA was extracted and the expression of adipogenesis and browning markers, as well as mitochondrial genes and genes involved in FA metabolism was measured by qPCR.

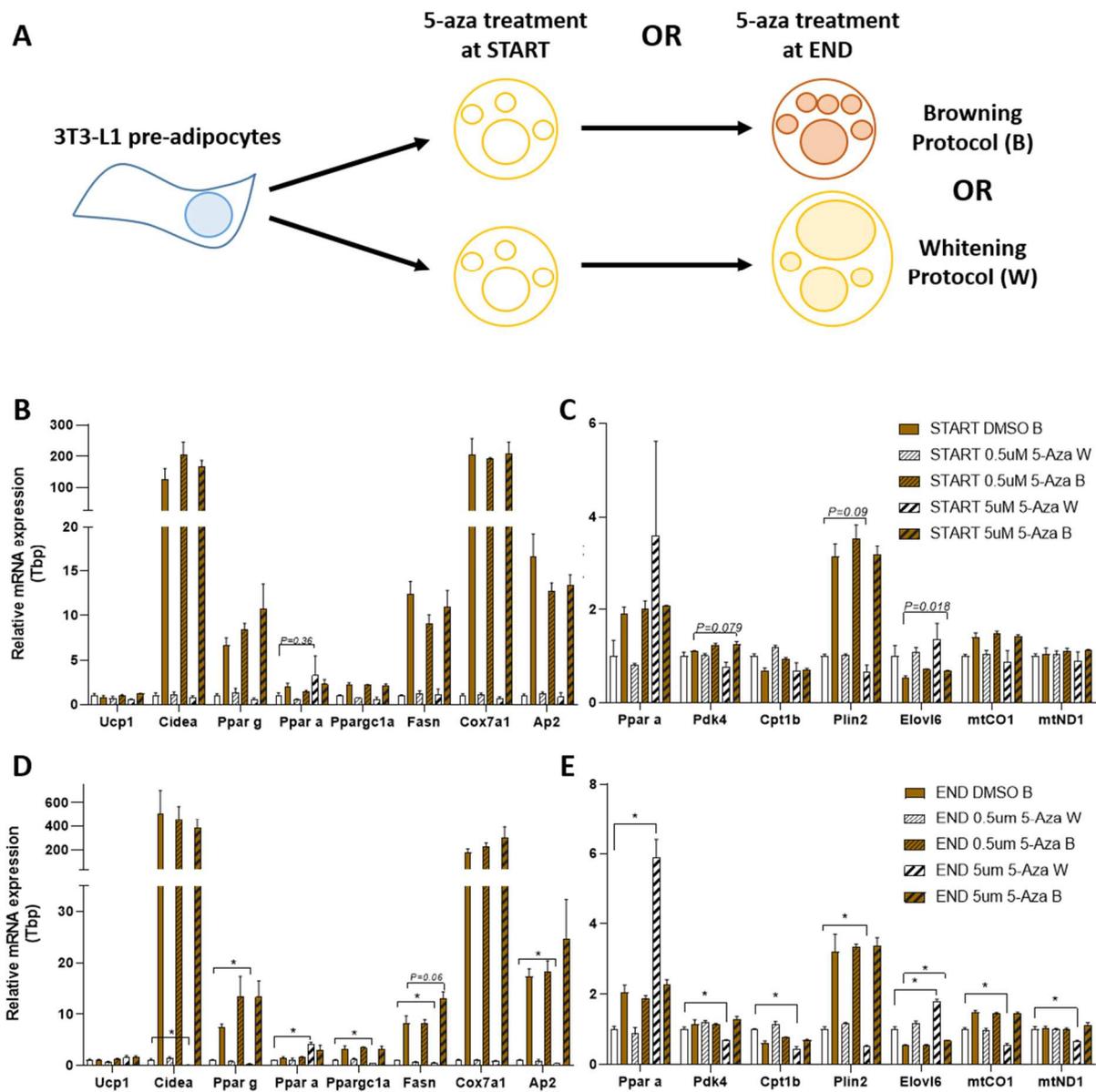


Figure 3.15: Differentiation of 3T3L1-preadipocytes under the Influence of 5-aza. A) Schematic representation of 3T3-L1 differentiation with either browning or whitening protocol under the influence of different 5-aza concentrations at the start or end of the differentiation. B) mRNA expression of adipose tissue and browning marker genes under the influence of 5-aza at the start of differentiation. C) mRNA expression of downstream genes of Ppar- α under the influence of 5-aza at the start of differentiation. D) mRNA expression of adipose and browning marker genes under the influence of 5-aza at the end of differentiation. E) mRNA expression of downstream genes of Ppar- α under the influence of 5-aza at the end of differentiation. Data were normalized to START DMSO W (B,C) or END DMSO W (D,E). $n = 3$, Housekeeping gene used for calculations is Tbh (B-E).

In both START and END treated cells, two of the browning markers, *cell death-inducing DNA fragmentation factor alpha-like effector A (Cidea)* and *peroxisome proliferator-activated*

receptor gamma coactivator 1-alpha (Pparg1 α), but not *UCP1*, were higher expressed when the browning protocol was applied (Fig. 3.15B and D). As expected, adipocyte marker *adipocyte protein 2 (Ap2)*, as well as other genes involved in lipid metabolism, such as *fatty acid synthase (Fasn)*, *peroxisome proliferator-activated receptor alpha (Ppar α)*, and *peroxisome proliferator-activated receptor gamma (Ppar γ)*, or *cytochrome c oxidase polypeptide 7A1 (Cox7a1)*, which is a component of the mitochondrial respiratory chain, were also more expressed after browning, most likely due to higher metabolic rate.

Low concentrations of 5-aza do not seem to affect gene expression significantly (Fig. 3.15B and D). However, high concentrations of 5-aza during whitening decreased the expression of *Cidea*, *Ppar γ* , *Pparg1 α* , *Fasn*, and *Ap2*, especially after END treatment. The only gene upregulated after 5-aza treatment to roughly the same levels as after browning was *Ppar α* (Fig. 3.15B and D). We, therefore, looked at the gene expression of several downstream genes of *Ppar α* . Surprisingly, three of the downstream targets, pyruvate dehydrogenase kinase 4 (*Pdk4*), carnitine palmitoyltransferase 1B (*Cpt1b*), and perilipin 2 (*Plin2*), were downregulated after high concentrations of 5-aza treatment (Fig. 3.15C and E). Only elongation of long-chain fatty acids family member 6 (*Elovl6*), which is also a browning marker and the gene with the highest expression fold change after our *in vivo* CAx treatment, was upregulated.

Taken together, the browning protocol provides us with several differentially expressed marker genes which could be used for the assessment of browning in cells treated with both whitening protocol and targeted epigenome editing. However, 5-aza treatment, which should support a hypomethylated state, in reference to our previous findings during WGBS, did only partially elevate marker gene expression to browning level and in some cases even worked against our expectations. On the other hand, 5-aza treatment is not targeted and may not be comparable to precise epigenome editing. In any case, the system requires further testing and potentially other strategies for readout, such as reporter genes or testing of respiratory rate.

4 Discussion

While interesting results have been produced for both projects, neither the insertion of super pioneer factor (SPF) motifs could be established as a robust method for gene reactivation, nor have new apparent targets for epigenome editing in adipocytes been identified. Therefore, before interpreting the results, the shortcomings of our experimental approach, as well as alternative routes to accomplish our goals, will be discussed. Since there were also promising findings in regards to *p16* reactivation, an outlook into follow-up experiments and potential therapeutic strategies will be given.

4.1 Experimental Approach

While targeting epigenetically silenced tumor suppressor genes (TSGs) for reactivation via epigenome editing, be it with dCas9-effectors or SPF motif insertion, is both biologically interesting and clinically relevant, it comes at the same time with an inherent flaw: TSG reactivation initiates various mechanisms ranging from cell cycle arrest to apoptosis, thus ultimately leading to selective proliferative disadvantage of the affected cells in comparison with unaffected cells. Although we start with a homogenous cell population (Huh7), our targeting approach invariably produces heterogeneous cell populations, be it because of varying lentivirus (LV) infection rates as well as insertion locations, infidelity in NHEJ repair, or the limited efficiency of motif insertion through homology-directed repair (HDR) after CRISPR. Which means that different activation rates might occur and the more ‘successful’ ones will be selected out. This generates a lot of background and the candidates with the most interesting or desirable phenotype might have been lost at early stages of the respective experiment due to proliferative disadvantage.

In order to get a proof-of-concept for the gene activating capabilities of SPF motifs with a possible therapeutic effect, one could also target other endogenous genes involved in other

diseases and whose activation does not cause a selective disadvantage. However, one could first use target genes that do not have immediate therapeutic benefits, but could provide an easier readout. These should be epigenetically silenced by DNA methylation, and, ideally, proven to be reactivated by demethylation. One such example is the promoter of *Rhox homeobox family member 2 B (RHOXF2B)*, which has been previously reactivated by targeted demethylation via TET1-recruiting dCas9–SunTag system in human embryonic kidney 293 (HEK293) cells⁵⁷⁴. We tried to recreate this activation by CTCF motif insertion into the same targeted location in HEK293 cells, and we observed *RHOXF2B* gene activation for several of the resulting clones (data not shown). However, since there are eight copies of the gene present in this cell type, one motif insertion via HDR co-occurs with multiple random insertions and deletions (indels) through non-homologous end joining (NHEJ) in the other copies. Therefore, one cannot exclude the possibility that reactivation might have been facilitated through a mechanism involving the latter. Indeed, we observed increased *RHOXF2B* expression also for clones, treated with Cas9 and sgRNA, but without repair template, i.e. without the presence of a CTCF motif, but of various indels (data not shown). These results suggest that *RHOXF2B* in HEK293 cells is not a suitable target to establish a proof-of-concept. Another candidate gene is *UCP1*, whose enhancer is demethylated in correlation with its activation after 5-azacytidine treatment in 3T3-L1 cells or cold-stress treatment in brown adipose tissue (BAT)⁴¹⁵. Finally, *myoblast determination protein (MyoD)*, which has been activated with dCas9-TET1 in C3H10T1/2 mouse embryonic fibroblast cells⁵⁷² could also serve as a candidate target. Since CRISPR-based insertions are hard to accomplish in these cell lines, we assessed whether we could activate these genes in mouse embryonic stem cells (mESCs), in which motif insertion works very well. Targeting dCas9-TET1 to these specific *loci* did, indeed, lead to demethylation, but no change in gene expression could be detected (data not shown), perhaps due to a lack of cell type-specific transcription factors (TFs). We, therefore, did not attempt SPF-motif insertions for these loci in mESCs. However, the task of finding a suitable target for epigenome editing has led to the start

of the second project in this thesis: the cell fate of adipose tissues and its linked epigenomic marks.

Another way to for testing the effect of SPF-motif insertion in gene expression with a more convenient readout is the employment of methylation sensitive reporters. One such previously published reporter is the *Dazl-Snrpn-GFP* introduced above⁶¹³. It has been shown that recruitment of dCas9-TET1 leads to demethylation of the *Snrpn* part and subsequent GFP expression⁵⁷². However, the amount of GFP-positive cells after lentivirus transduction remains quite modest and data in regards to changes in mean or median fluorescence intensity has not been shown⁵⁷². Here, we achieved unambiguous increase in GFP intensity after 5-aza treatment, which was also accompanied by demethylation of the promoter region (Fig. 3.8). The activation by SPF-motif insertion, on the other hand, was less obvious (Fig. 3.9 and Fig. 3.10). However, in comparison with the other target genes that were tested, the *Dazl-Snrpn-GFP* reporter still provided the most reliable readout and represents the most convenient testing system for SPF-dependent activation so far. In addition, activation of the *Dazl-Snrpn-GFP* reporter does not lead to an apparent selective disadvantage. Even so, aside from proof-of-concept testing, this system does not provide assessment capabilities in regards to potential medical targets, such as TSGs. To achieve that, a knock-in of GFP downstream of *p16* or *Rassfla* promoters, thus acting as a reporter for the activity of the TSG promoter, was assessed. We were successful in establishing two cell lines containing a GFP gene directly driven by the *Rassfla* promoter. However, the downsides of this approach became evident after bisulfite sequencing. Neither of the clones shared the methylation profile with the endogenous region, with one clone being completely demethylated, while the other is partially demethylated (Fig. 3.7). The homologous region of the repair template, which also covers a large portion of the promoter was inserted unmethylated. While GFP intensity correlated with the methylation status, their use for assessing SPF motif insertion, or other demethylation approaches, is questionable. Even so,

SPF motif insertions can still be performed on the established knock-in clone with partial methylation, to see if comparable GFP expression to the unmethylated clone can be achieved. Furthermore, one could attempt to remethylate the promoter by targeting it with dCas9-DNMT3A⁵⁷². However, this does not necessarily achieve complete methylation, e.g. the *Snrpn* reporter was remethylated to about 70% 5-mC⁵⁷². How well this translates to the *Rassf1a* promoter, which shows close to 100% 5-mC in Huh7, has not been tested. Another way to retain promoter methylation could be *in vitro* methylation of the donor plasmid for the repair template. While this can be achieved by treatment of the plasmid with bacterial methylase SssI⁶³⁹, it would also result in methylation of the entire construct including the reporter gene body and potential antibiotic resistance genes for clonal selection. Finally, one could try to insert a cassette formed of the GFP gene driven by the TSG promoter, in another location of the genome that could help maintaining the methylation status of the insert. In general, any knock-in may alter the epigenomic landscape of its target region in unforeseeable ways and needs to be investigated and compared to the endogenous state, not only in regards to DNA methylation, but also histone modifications, before serious assessment of demethylation approaches can be conducted.

In order to identify new medical targets for epigenome editing, we took a closer look at the DNA methylation profile of an emerging cell type in visceral adipose tissue (VAT) after a combination treatment with cold temperature and an antibiotics cocktail (CAx). Whole genome bisulfite sequencing (WGBS) on these beige adipocytes is challenging though, since they emerge from within the white adipose tissue (WAT), resulting in a heterogeneous sample. Hematoxylin and eosin (H&E) staining on perigonadal VAT and quantification of multilocular cells has revealed that beige adipocytes make up roughly 40% of the tissue after CAx treatment, making them a minority in the analyzed sample. Furthermore, an unknown amount of other cell types, especially from the immune system, may be present. This heterogeneity needs to be taken

into account when looking at the WGBS data, as well as the gene expression profile. Indeed, for the identification of DMRs and DEGs, only the largest changes in methylation and expression may be detected, while others may escape due to the background. For example, changes of 20% in the methylation level of a CpG in beige cells may appear as 8% in total methylation differences. Indeed, several regions with methylation differences could be detected by visual inspection of the methylomes of VAT and VAT+CAx. However, many of these regions did not pass the thresholds for the degree or extent of DNA methylation differences that we have set. For example, for one of the identified DEGs, *Stearoyl-CoA desaturase-2 (Scd2)*, which also contains a DMR and enhancer-specific marks in proximity to the TSS, another potential DMR overlapping with the enhancer marks is visible, but that did not pass our parameters for DMRs (Fig. 3.14C, red box). Our parameters were the default ones of swDMR software version 1.6.2⁶⁴⁰, which is a sliding window approach with a window size of 1000 bp and a step size of 100 bp. 1000 bp-regions with at least 10 CpGs and a methylation level (ML) difference of at least 10% (p value < 0.05 using a Fisher test) were considered as DMRs. These are very stringent conditions that are suitable for the comparison of two homogeneously constituted cell lines, but maybe less in our case. Indeed, to detect a DMS and assuming that the difference occurs only in beige cells, the minimum difference in methylation has to be >40%. Therefore, we are currently considering changing these parameters or using different software, e.g. DMRfinder⁶⁴¹ or DSS⁶⁴², which could confirm this region, and potentially others, as DMRs, and, eventually, as potential targets for epigenome editing. On the contrary, detected changes might be driven by non-beige cells and therefore constitute false positives.

To address the problem of heterogeneity in our samples, one could attempt cell sorting of the beige adipocytes. Sorting could be performed based on the differences in the number of mitochondria, which can be stained with agents such as MitoTracker® Deep Red⁶⁴³. Alternatively, a mouse line with a tissue-specific reporter for VAT-emerging beige adipocytes

can be created. Similar lines already exist for UCP1 expression, which is a marker in brown adipose tissue (BAT) and beige adipocytes emerging from subcutaneous adipose tissue (SAT)^{407,644}. Since changes in UCP1 are not as apparent in VAT after CAX, another marker should be used. For example, *Elongation of long chain fatty acids family member 6 (Elovl6)*, is the most significant differentially expressed gene (DEG) in VAT after CAX. Furthermore, it is an already known browning marker⁶⁴⁵, and its higher expression is, therefore, unlikely to be associated with an undesired cell type, although this cannot be ruled out categorically. An exemplary reporter cassette could exist of the endogenous *Elovl6* gene and a downstream GFP linked through a 2A self-cleaving peptide, which does not affect *Elovl6* function, reminiscent of the already existing UCP1-2A-GFP reporter⁶⁴⁴. In general, FACS on adipocytes has proven to be difficult, as the pressure leads to cell bursting, and, therefore, sample loss. Strategies to minimize sample loss involve lowering sheath pressure (e.g. to 6 psi), increasing nozzle diameter (e.g. 150 μm), and in-tube stirring⁶⁴⁶. In any case, the final number of sorted cells and, therefore extracted DNA, will be lower than from unsorted tissue. TET-assisted pyridine borane sequencing (TAPS), a method recently developed as alternative to WGBS, which retains higher sequence complexity and involves less harsh treatments, could be performed to counteract smaller samples²⁵⁴. Finally, single-cell locus-specific bisulfite sequencing (SLBS)²⁴⁹ could be performed on sorted cells to confirm candidate DMRs.

In addition to the genome-wide methylation profile of beige adipocytes emerging from VAT, the pattern of other epigenetic marks also remains elusive. ChIP experiments on WAT, BAT and beige adipocytes emerging from SAT, have provided us with tissue-specific histone modification profiles, especially for enhancer-related marks, like H3K4me1 and H3K27ac^{628,647,648}. These studies have given us insights into various switches involved in white tissue browning, e.g. ablation of histone deacetylase 3 (HDAC3) leads to acetylation of UCP1 and *Ppargc1a* enhancers as well as their respective gene activation⁶⁴⁷. Since similar control

mechanisms could also exist for VAT and since epigenome editors targeting these modifications, e.g. dCas9-p300⁵⁷⁶, have already been developed, conducting CHIP experiments on enhancer-specific marks seems a necessary step in our understanding of VAT browning. The heterogeneity could be circumvented with above-mentioned reporters and FACS techniques, and the resulting limited cell number for CHIP, could be counterbalanced with recently developed techniques such as CHIPmentation⁶⁴⁹ or combinatorial barcoding and targeted chromatin release (CoBATCH)⁶⁵⁰.

Finally, one might ask the question of why are we focusing on beige adipocytes emerging from VAT rather than SAT. It is indeed true that sorting techniques for SAT already exist, which would make methylome and expression profiling easier⁴⁰⁷. Likewise, browning of SAT also leads to higher glucose tolerance and insulin resistance⁶²⁷. VAT, however, is a clinically more relevant tissue as it is considered more dangerous than SAT^{651,652}. Aside from fat storage, VAT is also involved in immune responses, and can be considered as intraabdominal adipo-immune organ⁶⁵³. It is strongly associated with chronic inflammation, negatively impacting cardiovascular diseases and Type II Diabetes^{654,655}. Since these aberrant immune phenomena do not occur in subcutaneous obesity, SAT is sometimes considered as healthy fat^{656,657}. Increased immune response and inflammation, manifesting in an increase in macrophages and cytokines, are, on the other hand, characteristic for VAT obesity^{653,658,659}. At the same time, VAT retains insulin sensitivity, resulting in a positive feedback loop consisting of increased VAT leading to increased inflammation that leads to increased peripheral insulin resistance, leading in turn to VAT prioritizing for energy storage, finally resulting in further expansion of VAT⁶⁶⁰. Shifting its function from energy storage towards energy expenditure could, therefore, be a potential strategy in breaking this feedback loop and ameliorating the negative effects of VAT obesity. Furthermore, this new emerging cell type is still very little understood and

assessment of its epigenomic as well as expression profile will contribute to our insight into its origin and function.

4.2 Interpretation of the Results

On a technical level, insertion of SPF motifs into the desired target regions using CRISPR/Cas9 has been successful. In most cases, this insertion has led to local demethylation and in some cases, it has also led to an increase in target gene expression. Since overall, both demethylation and activation of gene expression were met with heterogeneous outcomes and neither has been achieved in a robust manner, it is clear that our efforts cannot be regarded as a robust proof-of-concept for this approach.

It is important to note that the motifs of CTCF and SOX2 were previously tested experimentally for their ability to specifically and efficiently bind to their corresponding PFs, either by ChIP experiments, or DNA/protein microarray and EMSA. Therefore, we made the assumption that the TFs bind to the WT motif in our setting. However, one could confirm that this event is indeed taking place by performing chromatin immunoprecipitation (ChIP) experiments, assessing CTCF or SOX2 enrichment at the target regions.

In case of TSG promoter targeting, insertion of the wild-type CTCF motif has led to, albeit sometimes modest, local demethylation, which could not be observed for the scrambled motif (Fig. 3.5 and Fig. 3.6). Only for *Rassfla*, which also experienced the largest extent of demethylation, we could observe an increase in mRNA expression, that could not be confirmed on the protein level. It is unknown whether this is due to regulatory events at the post-transcriptional level, or simply because the amount of protein is below the detection limit. For p16 we could not detect any increase in gene expression. Some possible explanations could be insufficient extent or bad positioning of demethylation at the target region, steric hindrance due to CTCF occupancy, TF motif disruption due to motif insertion, or absence of other key TFs

that are needed for activation. However, the fact that we obtained a moderate p16 reactivation with dCas9-TET1 in the same cell line indicates that the conditions to obtain a detectable level of activation, including the presence of TFs necessary for activation, are present (Fig. 3.1). In comparison to CTCF motif insertion, targeting with dCas9-TET1 resulted in more widespread demethylation, which also included the transcription start site (TSS). This is probably due to the fact that multiple copies of the effector protein were spread out over the promoter region, while the insertion just occurred in one spot. This activation, while significant, was relatively low and not detectable on the protein level, so one should be careful when using it as a reference point. CTCF occupancy has yet to be confirmed by ChIP, however, steric hindrance is probably lower in comparison to the recruitment of up to four larger proteins in form of dCas9-TET1. On the other hand, CTCF recruitment could also lead to the disruption or the establishment of new DNA loops, which may influence gene expression. The importance of a CTCF-mediated chromatin boundary site about -1 kb from the TSS has previously been shown in regards to regulating p16 expression, with its loss resulting in gene silencing in various cancer cell lines⁶⁶¹. Similar results have been observed for a boundary site around 1.8 kb upstream of the *Rassfla* gene⁶⁶¹. Changes in chromatin architecture, such as new loop formations upon CTCF motif insertion, could be assessed via circular chromosome conformation capture (4C) or related techniques.

Using CTCF, only insertion of the reverse wild-type motif into the Dazl part of the Dazl-Snrpn reporter lead to an increase in GFP expression (Fig. 3.9 and Fig. 3.11). While this was accompanied by some degree of demethylation, moderate rates of demethylation have also been observed in clones with scrambled or other wild-type motif insertion that did not show an increase in GFP expression. One extreme example was the insertion of the wild-type CTCF motif in forward direction, which resulted in two clones with identical sequences, but two vastly different bisulfite profiles (Fig. 3.9F) and no apparent change in expression. For SOX2 motif

insertion into the *Snrpn* part, the clone with the wild-type motif in forward direction showed the strongest increase in GFP expression, while retaining complete methylation. At the same time, the clone with the scrambled motif experienced a slight decrease in methylation but no changes in expression. These findings open up several questions: (1) Is the observed demethylation caused by the SPFs? (2) Is demethylation necessary for the activation of the *Snrpn-Dazl* reporter? (3) Is the observed increase in gene expression facilitated through other functions of the SPFs? – It has been shown that the insertion of unmethylated fragments can lead to demethylation of neighboring sequences, which may be an alternative explanation for the first question. Moreover, a part of the answer for the second question is provided by the fact that we have shown that 5-aza treatment results in demethylation and activation of *Snrpn-Dazl* reporter, which indicates that demethylation is sufficient, but not whether it is necessary, for the activation of the reporter. Indeed, other 5-aza-dependent demethylation events occurring at different locations might be necessary for this activation. Both CTCF and SOX2 are known to interact with a variety of other proteins influencing gene expression and might act through these mechanisms. As mentioned above, CTCF is an architect of chromatin topology, and may act through loop formation. SOX2, has been shown to recruit Krüppel-like factor 4 (KLF4), another identified SPF, as well as to interact with some other 143 proteins, several of which are transcriptional activators and coactivators^{662,663}. Taken together, the results at their current state are not sufficient to support a proof-of-concept for SPF motif insertion-mediated reactivation of gene expression via demethylation. Further ChIP and 4C experiments, on the other hand, should elucidate the mechanisms at play.

Interestingly, WGBS in VAT either from RT or CAX treated mice, revealed many DMRs, which were hypomethylated after CAX treatment. This result is in concordance with previous studies using reduced representation bisulfite sequencing (RRBS) and RNAseq in BAT and WAT, and that identified mostly hypomethylated DMRs in BAT versus WAT at the identified DEG

promoters⁴²⁰. In addition, injection of mice with 5-aza has resulted in an increase in UCP1 and Ppargc1 α expression in SAT, as well as in a decrease in inflammation factors IL1 β and TNF α ⁶⁶⁴. None of our identified DMRs, however, overlapped with a DEG promoter region, nor, after looking at H3K4me1 and H3K27ac marks in WAT or BAT, with any apparent enhancer. As of yet, no candidate genes for epigenome editing in VAT have been revealed. This may change by changing the parameters for DMR recognition or by reducing the heterogeneity of the sample with above-mentioned FACS methods.

4.3 Outlook and Perspectives

While a robust proof-of-concept for SPF motif insertion-mediated reactivation of gene expression via demethylation is still missing, there are still other SFP candidates that should be tested. These could use different mechanisms to activate genes. Two possible candidates are the established SPFs KLF4 and FOXA1. KLF4 is a pluripotency factor that is expressed in mouse embryonic stem cells (mESC) and that has shown to have strong demethylating capabilities in mESCs³⁴⁴. Unlike SOX2, which induces demethylation passively by inhibiting DNMT1, KLF4 interacts with TET2⁶⁶⁵, and its recruitment, results in TET-dependent active DNA demethylation. KLF4 is expressed in Huh7, although at low levels, and could, therefore, serve as an activating SPF. FOXA1, on the other hand, is highly expressed in Huh7 cells, which makes it a better candidate. Furthermore, FOXA1 has been shown to interact with TET1 in prostate cancer and was associated with DNA demethylation at DNA repair sites^{602,666,667}. Another option is to insert SOX2 motif alongside an OCT4 motif, since the two occur often adjacent to each other and were shown to collaborate to achieve pluripotency^{668,669}. Moreover, the presence of the two motifs was shown to increase SOX2 ability to demethylate its binding sites³⁴⁴.

In case robust gene activation using SPF motif insertion is achieved, one may think of applying this method for therapeutic purposes. However, one apparent downside of insertion-based

activation methods has not yet been addressed: indels. Using Cas9 to induce double strand breaks (DSB), results predominantly in NHEJ and the creation of indels. This is particularly problematic when you have to target virtually the whole cell population, e.g. in cancer. The occurrence of random indels, could not only lead to unpredictable effects on gene expression, but also masks the target site from further recognition by the sgRNA, making repeated treatment, in case the initial motif insertion failed, impossible. A possible solution could be the employment of double-nickases, which reduces both off-target effects and indels caused by DSBs⁶⁷⁰.

To avoid DSBs and indels altogether, one can also recruit dCas9-effectors to the target locus. Here, we successfully induced DNA demethylation by recruiting dCas9-TET1 to the *p16* promoter, resulting in activation of gene expression (Fig. 3.1 and Fig. 3.2). While the expression of p16 mRNA remained relatively low in comparison to its expression in HeLa cells (used as positive control) and could not be detected on the protein level, significant phenotypic changes in regards to senescence and proliferation could be observed (Fig. 3.3 and Fig. 3.4). This was achieved using one of the least elaborate dCas9-effectors. Employing dCas9-SunTags recruiting multiple TET1 proteins⁵⁷⁴ or a combination of various transactivators and epigenome editors may further increase reactivation of p16 expression. To have a stronger effect on the cancer phenotype, one can also target multiple TSGs. Polycistronic tRNA and CRISPR guide-RNA (PTG) enables targeting of multiple sites with just one construct^{671,672}. Finally, dCas9-SPF fusions were not tested in our study. These could prove more efficient, as the recruitment of multiple proteins to multiple sites is more feasible when compared to motif insertion. The downside of such fusion proteins is that they are not stably expressed and therefore do not allow stable recruitment of the SPFs.

To assess these strategies on TSGs, one has to address the issue of creating a selective disadvantage, as discussed at the beginning of this chapter. The presence of all required

components (dCas9, effector, and guide RNA) in the cell, as well as temporal control of their assembly at the target site is key to properly investigate their effect without generating large heterogeneity within a cell population and to enable short term analysis avoiding sorting or clonal selection. This also provides the option to perform a xenograft of the established cell lines prior to induction. Several inducible CRISPR systems already exist⁵⁷⁷⁻⁵⁸⁰, but care must be taken when choosing the means of induction. Rapamycin, for example, may influence cell growth, although these effects have not been observed at such low doses⁵⁸⁰. Light-inducible systems might not have effects on cell physiology, but may prove to be useless in a xenograft model.

Finally, once a robust method for TSG activation has been developed, which also leads to significant phenotypic changes in cancer physiology both in cell culture and xenograft models, preclinical evaluation of the epigenome therapy can be conducted in an inducible hepatocellular carcinoma (HCC) mouse model⁶⁷³. The CRISPR machinery could, for example, be delivered via adeno-associated virus (AAV), which shows high infection rate of HCC cell lines and can also be directed to the liver in mouse models, using the right AAV serotype⁶⁷⁴⁻⁶⁷⁶. One issue with AAV is the limited packaging capacity. Since spCas9 protein is of considerable size (1368 aa), smaller Cas proteins, such as the recently discovered Cas14 that is 400 to 700 aa in length, could prove useful⁶⁷⁷. Whether the project will advance to such a stage cannot be predicted at this stage, however, the possibilities for using dCas9-effectors to achieve TSG activation for potential cancer therapy are both numerous and promising, and, since the tools are available, these should be assessed as soon as possible.

As concluding remark, it can be said that the interaction between epigenetic marks and transcription factors, and the resulting control of gene expression remains as complex and fascinating as ever, and, more often than not, unpredictable. New “plug and play” emerging techniques allow us to edit these marks and influence both cell fate and the progression of

diseases. However, to master such control over ancient physiological processes, and to prevent random tinkering, a thorough and deep understanding of the underlying biology is needed. And here, still a lot can be learned.

“Science is not about building a body of known ‘facts’. It is a method for asking awkward questions and subjecting them to a reality-check, thus avoiding the human tendency to believe whatever makes us feel good.”

— Terry Pratchett, *The Science of Discworld*

5 Materials and Methods

5.1 Cell Culture

Dazl-Snrpn-GFP V6.5 mESCs⁶¹³ were cultivated on dishes coated with 0.2% porcine skin gelatin (Sigma, cat. No. G1890) in high glucose DMEM medium (Thermo Fisher, USA) supplemented with 15% heat-inactivated (HI) fetal bovine serum (FBS), 1% NAA (Thermo Fisher, USA), 1:1000 homemade LIF and 1:100000 beta mercaptoethanol. Huh7 cells were cultivated in DMEM/F-12, GlutaMAX (Thermo Fisher, USA) supplemented with 10% HI FBS. HT1080 cells, HEK293T cells, and 3T3L1-preadipocytes were cultivated in high glucose DMEM medium (Thermo Fisher, USA) supplemented with 10% HI FBS. All cells were kept at 37°C. Dazl-Snrpn-GFP V6.5 mESCs, Huh7 cells, HT1080 cells, and HEK293T cells were cultured at 5% CO₂ and 3T3L1-preadipocytes were both cultivated and differentiated at 7.5% CO₂.

5.2 Treatment of Dazl-Snrpn-GFP V6.5 mESCs with 5-azacytidine

2*10⁵ Dazl-Snrpn-GFP V6.5 mESCs were plated per well in a 6-well plate and cultivated under regular culture conditions (see 6.1). 24 hours after plating, medium with 1 μM 5-azacytidine (5-aza) (Sigma-Aldrich, USA) was added. Medium exchange with fresh 5-aza medium was performed every 48 hours. Cells were harvested for FACS 7 days and 14 days after start of treatment, and after 14 days for bisulfite analysis.

5.3 Cloning of guide RNAs

Guide RNAs (gRNAs) for lentiviral transduction were cloned into pgRNA-modified (Addgene plasmid: 84477)⁵⁷² containing an AarI restriction site. gRNAs for plasmid-based insertion of SPF motifs were cloned into pSPgRNA (Addgene plasmid: 47108)⁵⁶⁵ containing a BbsI restriction site.

Table 5: Primer Sequences to Construct Guide RNAs

Name	Distance from TSS (bp)	Sequence
Targeting p16		
SgRNA1 fwd	+6	ttggCCCGGGGGAGACCCAACCTG
SgRNA1 rev	+6	aaacCAGGTTGGGTCTCCCCGGG
SgRNA2 fwd	-35	ttggGCCAACGCTGGCTCTGGCGA
SgRNA2 rev	-35	aaacTCGCCAGAGCCAGCGTTGGC
SgRNA3 fwd	-131	ttggACGCCTTTGCTGGCAGGCGG
SgRNA3 rev	-131	aaacCCGCCTGCCAGCAAAGGCGT
SgRNA4 fwd	-207	ttggTTAGGAAGGTGTATCGCGG
SgRNA4 rev	-207	aaacCCGCGATAACAACCTTCCTAA
SgRNA2 fwd	-35	caccGCCAACGCTGGCTCTGGCGA
SgRNA2 rev	-35	aaacTCGCCAGAGCCAGCGTTGGC
SgRNA5 fwd	-141	caccGGCACTCAAACACGCCTTTGC
SgRNA5 rev	-141	aaacGCAAAGGCGTGTTTGAGTGCC
Targeting Rassf1a		
SgRNA6 fwd	+28	ttggCCTCCCCCAGGATCCAGACT
SgRNA6 rev	+28	aaacAGTCTGGATCCTGGGGGAGG
SgRNA7 fwd	-21	ttggGCACCCAGGTTTCCATTGCG
SgRNA7 rev	-21	aaacCGCAATGGAAACCTGGGTGC
SgRNA8 fwd	-30	ttggAACCTGGGTGCAGGGACTGT
SgRNA8 rev	-30	aaacACAGTCCCTGCACCCAGGTT
SgRNA9 fwd	-137	ttggACCCCGGACGGCCACAACGA
SgRNA9 rev	-137	aaacTCGTTGTGGCCGTCCGGGGT
SgRNA9 fwd	-137	caccACCCCGGACGGCCACAACGA
SgRNA9 rev	-137	aaacTCGTTGTGGCCGTCCGGGGT
Targeting Dazl-Snrpn		
Dazl SgRNA fwd	-260	caccGAGCCGAGCTGTAGGGTGCT
Dazl SgRNA rev	-260	aaacAGCACCTACAGCTCGGCTC
Snrpn SgRNA fwd	-122	caccCGCATGTGCAGCCATTGCCT
Snrpn SgRNA rev	-122	aaacAGGCAATGGCTGCACATGCG

5.4 Lentivirus Production and Transduction

Lentiviruses expressing dCas9-TET1, dCas9-dTET1, and gRNAs, as well as GFP were produced by transfecting HEK293T cells with Fuw-dCas9-Tet1CD (Addgene plasmid: 84475)⁵⁷², Fuw-dCas9-Tet1CD_IM (Addgene plasmid: 84479)⁵⁷², pgRNA constructs, or pCLX-UBI-GFP (Addgene plasmid: 27245), respectively, together with standard packaging vectors pCAG-VSVG (Addgene plasmid: 35616)⁶⁷⁸ and psPAX2 (Addgene plasmid: 12260). GFP-expressing virus was produced in parallel to the other viruses, since it could be used for FACS-based virus titration, while the others may not possess a fluorescent marker. One day before transfection, 2×10^6 HEK293T were plated onto a 10 cm cell culture dish to reach 50-70% confluence at the time of transfection. 4 μ g of pCAG-VSVG, 8 μ g of psPAX2, and 8 μ g of Fuw, gRNA, or GFP plasmid were cotransfected with calcium phosphate. The plasmid mix was adjusted to 250 μ L with H₂O (Stock: 25 mL H₂O, 250 μ L 1M HEPES pH 7.3). To the adjusted plasmid mix 500 μ L of 2x HBS solution (280 mM NaCl, 50 mM HEPES, 1.9 mM Na₂HPO₄, in H₂O) was added. The resulting mixture was dropwise added to 250 μ L 4x CaCl₂ solution (500 mM CaCl₂ in 250 mL H₂O with 2.5 mL 1M HEPES pH 7.3) under vigorous vortexing. After allowing precipitation for roughly 5 minutes, the final mixture was dropwise added onto the cell monolayer. Medium was exchanged after 12 hours. 24 hours post transfection, the medium containing the virus was harvested and either stored at -80°C or used for titer determination.

To test the titer of the harvested virus medium, 5×10^4 HT1080 cells were plated per well in a 6-well plate 24 hours before transduction. 5 μ L, 50 μ L, or 500 μ L virus containing medium was added. On day 5 post-transduction, cells were harvested for either FACS- or qPCR-based lentivirus titration. FACS-based titration was performed on GFP-expressing viruses, which were produced in parallel to the other viruses. Flow cytometry was performed on a BD Accuri C6 (Becton Dickinson, USA) and analyzed using FlowJo V10 (Becton Dickinson, USA).

Titer (HT1080-transducing units/mL) = 100,000 (estimated target HT1080 cells) * (% of GFP-positive cells/100) / volume of virus containing medium (in mL)

Subsequently, qPCR on the presence of the HIV-expressed gag gene (GAG) inside the host genome was performed, with human beta-actin (HB2) as housekeeping gene. Genomic DNA (gDNA) was extracted using the GenElute Mammalian Genomic DNA Miniprep Kit (Sigma-Aldrich, USA) according to manufacturer's instructions. qPCR was performed as described under 5.8. A standard curve was drawn, plotting the previously obtained copy number (of the GFP virus) against the Δ Ct (cycle threshold; Ct GAG minus Ct HB2). By applying the resulting formula to the Δ Ct values of the other samples, the copy number for each sample was calculated, which in turn could be used to calculate the titer and the desired multiplicity of infection (MOI).

Titer (HT1080-transducing units/mL) = 100,000 (estimated target HT1080 cells) * (copy number per cell) / volume of virus containing medium (in mL)

The volume for the wanted MOI, i.e. the desired average virus copy number per cell after infection, was calculated according to the following formula:

Volume (μ L) = X / (copy number per cell) * volume of virus containing medium (in μ L)

X represents the desired MOI. For the establishment of dCas9-effector protein expressing Huh7 cells, 5×10^4 Huh7 cells were plated per well in a 6-well plate 24 hours before transduction. 500 μ L of virus containing medium was added without prior titer testing. Medium was exchanged after 12 hours and clones were picked and tested for dCas9-effector protein expression two weeks post-transduction. The resulting cell lines were infected with gRNA expressing virus, by plating 5×10^4 Huh7 dCas9-TET1/dCas9-dTET1 cells per well in a 6-well plate 24 hours before transduction, and adding the according volume of virus containing medium to achieve an MOI of 10 per gRNA-expressing virus (4 gRNAs at the same time).

5.5 SPF Motif Insertion with CRISPR/Cas9

Both Huh7 cells and Dazl-Snrpn-GFP V6.5 mESCs were split 24 hours before nucleofection. 2 hours prior to nucleofection, the medium for Dazl-Snrpn-GFP V6.5 mESCs was exchanged for standard medium containing 20% FBS. For SPF motif insertion, 1 μ L of 100 μ M single-stranded oligo DNA nucleotides (ssODNs), serving as repair template containing the motif, 5 μ g pSpCas9(BB)-2A-Puro (PX459) V2.0 (Addgene plasmid: 62988)⁵⁵⁸, expressing the Cas9, and 2.5 μ g of the respective pSPgRNA plasmid, were transfected using Nucleofector 2b (Lonza, Switzerland). For nucleofection of Huh7 cells the Cell Line Nucleofector Kit T (Lonza, Switzerland) was used for 10^6 cells for each transfection. For nucleofection of Dazl-Snrpn-GFP V6.5 mESCs the Mouse Embryonic Stem Cell Nucleofector Kit (Lonza, Switzerland) was used for 10^6 cells for each transfection. For both cell types, medium was exchanged 12 hours post-transfection. For the first two days after transfection 2 μ g/mL puromycin (InvivoGen, USA) was added to the medium to enrich for Cas9-transfected cells. After colony formation, clones were picked and genotyped.

List of ssODN Donors for CRISPR-directed SPF Motif Insertion:

CTCT WT fwd (-141 bp from p16 TSS)

gcagttaggaaggttgatcgcggaggaaggaacggggcgggggcggatttcttttaacagagtgaacgcactcaaacacgcctttggccaccagggggcgcta aagctttgctggcaggcgggggagcgcggctgggagcagggaggccggagggcggtgtggggggcaggtggggaggagcccagtcctcctcctcct

CTCT SC (-141 bp from p16 TSS)

gcagttaggaaggttgatcgcggaggaaggaacggggcgggggcggatttcttttaacagagtgaacgcactcaaacacgcctttatggctggccaccgggggaagctttgctggcaggcgggggagcgcggctgggagcagggaggccggagggcggtgtggggggcaggtggggaggagcccagtcctcctcctcctcct

CTCT WT fwd (-35 bp from p16 TSS)

gagcgcggctgggagcagggaggccggagggcggtgtggggggcaggtggggaggagcccagtcctcctcctcctgccaacgctgctctggtggccaccagggggcgctacgagggctgctccggctggtgccccgggggagac

CTCT SC (-35 bp from p16 TSS)

gagcgcggctgggagcagggaggccggagggcggtgtggggggcaggtggggaggagcccagtcctcctcctcctgccaacgctgctctggaatggctggccaccggggcgagggctgctccggctggtgccccgggggagac

CTCT WT fwd (-137 bp from Rassf1a TSS)

ggggctgggcgcgctctcgcagagcccccccgcttgccttccctccttcgtcccctcctcacacccccccccggacggcca
caatggccaccagggggcgctacgacggcgaccgaaagcaccacgaggagatacccg

CTCT SC (-137 bp from Rassf1a TSS)

agggctgggcgcgctctcgcagagcccccccgcttgccttccctccttcgtcccctcctcacacccccccccggacggcca
caaatggctggccaccggggcgacggcgaccgaaagcaccacgaggagatacccg

CTCT WT fwd (-260 bp from GFP TSS in Dazl-Snrpn Reporter)

aacaagttaggccagctgagagaattctagaacattctcaagccagagaaacggggcctacctacctacagcagagccgagctgta
gggtggccaccagggggcgctagcttggcaattgacgctcaaatttccgagtaggaa

CTCT WT rev (-260 bp from GFP TSS in Dazl-Snrpn Reporter)

aacaagttaggccagctgagagaattctagaacattctcaagccagagaaacggggcctacctacctacagcagagccgagctgta
gggttagcggcccctgggtggccaagcttggcaattgacgctcaaatttccgagtaggaa

CTCT SC (-260 bp from GFP TSS in Dazl-Snrpn Reporter)

aacaagttaggccagctgagagaattctagaacattctcaagccagagaaacggggcctacctacctacagcagagccgagctgta
gggtaatggctggccaccggggcgcttggcaattgacgctcaaatttccgagtaggaa

CTCT WT fwd (-122 bp from GFP TSS in Dazl-Snrpn Reporter)

ccttttgtagctgccttttggcaggacattccggtcagagggacagagaccctgcattgaggcaaaaatgtgcgcatgtgcagccatt
gtggccaccagggggcgctacctgggacgcatgctagggagccgcgcgacaaaacc

CTCT WT rev (-122 bp from GFP TSS in Dazl-Snrpn Reporter)

ccttttgtagctgccttttggcaggacattccggtcagagggacagagaccctgcattgaggcaaaaatgtgcgcatgtgcagccatt
gtagcggcccctgggtggccaagctgggacgcatgctagggagccgcgcgacaaaacc

CTCT SC (-122 bp from GFP TSS in Dazl-Snrpn Reporter)

ccttttgtagctgccttttggcaggacattccggtcagagggacagagaccctgcattgaggcaaaaatgtgcgcatgtgcagccatt
gaatggctggccaccggggcgctgggacgcatgctagggagccgcgcgacaaaacc

SOX2 WT fwd (-260 bp from GFP TSS in Dazl-Snrpn Reporter)

aacaagctaggccagctgagagaattctagaacattctcaagccagagaaacggggcctacctacctacagcagagccgagctgta
gggtccttgtgcttggcaattgacgctcaaatttccgagtaggaa

SOX2 WT rev (-260 bp from GFP TSS in Dazl-Snrpn Reporter)

aacaagctaggccagctgagagaattctagaacattctcaagccagagaaacggggcctacctacctacagcagagccgagctgta
gggtaacaaaggcttggcaattgacgctcaaatttccgagtaggaa

SOX2 SC (-260 bp from GFP TSS in Dazl-Snrpn Reporter)

aacaagctaggccagctgagagaattctagaacattctcaagccagagaaacggggcctacctacctacagcagagccgagctgta
gggtctgtcttggcttggcaattgacgctcaaatttccgagtaggaa

SOX2 WT fwd (-122 bp from GFP TSS in Dazl-Snrpn Reporter)

ccttttgtagctgccttttggcaggacattccggtcagagggacagagaccctgcattgaggcaaaaatgtgcgcatgtgcagccatt
gccttggcttgggacgcatgctagggagccgcgcgacaaaacc

SOX2 WT rev (-122 bp from GFP TSS in Dazl-Snrpn Reporter)

ccttttgtagctgccttttggcaggacattccggtcagagggacagagaccctgcattgaggcaaaaatgtgcgcatgtgcagccatt
gaacaaaggcctgggacgcatgcgtagggagccgcgcgacaaaacc

SOX2 SC (-122 bp from GFP TSS in Dazl-Snrpn Reporter)

Ccttttgtagctgccttttggcaggacattccggtcagagggacagagaccctgcattgaggcaaaaatgtgcgcatgtgcagccat
tgctgtcttcctgggacgcatgcgtagggagccgcgcgacaaaacc

5.6 Genotyping

gDNA from the picked SPF motif insertion clones was extracted using the GenElute Mammalian Genomic DNA Miniprep Kit (Sigma-Aldrich, USA) according to manufacturer's instructions. Genotyping PCR was performed using GoTaq G2 Hot Start Green Master Mix (Promega, USA) according to the following protocol: 95°C for 2 minutes; 95°C for 30 seconds; 64°C for 30 seconds; 73°C for 1 minute; Repeat steps 2-4 20 X; 73°C for 5 minutes; Hold at 4°C. The resulting PCR fragments were run on a 1% agarose gel and the bands were extracted using the GenElute Gel Extraction Kit (Sigma-Aldrich, USA). The samples were sequenced at Microsynth, Switzerland, and the sequences were analyzed using CLC Workbench (QIAGEN, Germany).

Table 6: Genotyping Primers

Name	Sequence
p16 Genotype fwd	AGAATTCTCCCCCGTCCGTA
p16 Genotype rev	CGACCCTGTCCCTCAAATCC
Rassf1a Genotype fwd	CAGCTCCCGCAGCTCAAT
Rassf1a Genotype rev	ACCTCAAGATCACGGTCCAG
Dazl-Snrpn Genotype fwd	CGACTAGAGAGCAGGCCTTG
Dazl-Snrpn Genotype rev	CAGAACCAAGCGTCTGGCAT

5.7 Bisulfite Conversion, PCR, and Sequencing

gDNA was extracted using the GenElute Mammalian Genomic DNA Miniprep Kit (Sigma-Aldrich, USA) according to manufacturer's instructions. Bisulfite conversion of 800 ng gDNA was performed using the EZ DNA Methylation-Gold Kit (Zymo Research, USA) following the manufacturer's instructions. 3 μ L of the resulting modified DNA was amplified by PCR using AmpliTaq Gold DNA Polymerase (Thermo Fisher, USA) according to the following program: 95°C for 15 minutes; 95°C for 30 seconds; 61°C for 30 seconds; 72°C for 1 minute; Repeat steps 2-4 20 X, with a decrease of 0.5°C each cycle for step 3; 95°C for 30 seconds; 53°C for 30 seconds; 72°C for 1 minute; Repeat steps 5-7 40 X; 72°C for 15 minutes; Hold at 4°C. The resulting PCR fragment was run on a 1% agarose gel and the band was extracted using the GenElute Gel Extraction Kit (Sigma-Aldrich, USA). The fragment was cloned into One Shot TOP10 Chemically Competent E. coli (Thermo Fisher, USA) using the TOPO TA Cloning Kit (Thermo Fisher, USA) according to manufacturer's instructions. The samples were sequenced at Microsynth, Switzerland, and the sequences were analyzed using BISMAs software⁶⁷⁹.

Table 7: Primers for Bisulfite PCR

Name	Sequence
p16 BisSeq fwd	GTGGGGTTTTTATAATTAGGAAAGAATA
p16 BisSeq rev	CTATCCCTCAAATCCTCTAAAAAAC
Rassf1a BisSeq fwd	TTTATTTAGTGGGTAGGTTAAGTGTGTT
Rassf1a BisSeq rev	AAACCTAAATACAAAACTATAAAACCC
Dazl-Snrpn BisSeq fwd	TGTTTATTATGTTAGTAGAATTTATAAGTTTAG
Dazl-Snrpn BisSeq rev	AAAAAACACAACAATAACCAAACCAC

5.8 RNA Extraction and qPCR

RNA was extracted with the RNeasy Plus Mini Kit (QIAGEN, Germany) according to manufacturer's instructions. RNA content was measured with the Qubit RNA HS Assay Kit

(Thermo Fisher, USA) according to manufacturer's instructions. 200 µg of RNA were converted to cDNA with PrimeScript 1st Strand cDNA Synthesis Kit (Takara Bio, Japan) using oligo-dT primers. qPCR was performed using a StepOnePlus Real-Time PCR System (Thermo Fisher, USA) and SYBR Green PCR Master Mix (Thermo Fisher, USA) according to the following protocol: 95°C for 10 minutes, followed by 40 cycles at 95°C for 15 seconds and 60°C for 1 minute. Expression levels were calculated according to the Livak method⁶⁸⁰. Statistical analysis was performed with Prism 8 (GraphPad, USA). The p-value was calculated via an unpaired parametric t-test with Welsh's correction.

Table 8: qRT-PCR Primers

Name	Sequence
p16 qPCR fwd	CTTCGGCTGACTGGCTGG
p16 qPCR rev	TCATCATGACCTGGATCGGC
Rassf1a qPCR fwd	GTTACCTGCCACTACCGC
Rassf1a qPCR rev	CACAGGCTCGTCCACGTTC
SNRPD3 qPCR fwd	CAGCGGACCGAAGAGAAGAA
SNRPD3 qPCR rev	TGTGATGTTGGACATCTGGCA
GFP qPCR fwd	CTCGATGTTGTGGCGGATCT
GFP qPCR rev	GGGCACAAGCTGGAGTACAA
Hprt qPCR fwd	GTTGGGCTTACCTCACTGCT
Hprt qPCR rev	TCATCGCTAATCACGACGCT
GAG qPCR fwd	GGAGCTAGAACGATTCGCAGTTA
GAG qPCR rev	GGTTGTAGCTGTCCCAGTATTTGTC
H2B qPCR fwd	TCCGTGTGGATCGGCGGCTCCA
H2B qPCR rev	CTGCTTGCTGATCCACATCTG

5.9 Protein Extraction and Western Blot

For protein extraction, 200 – 500 μ L RIPA-like lysis buffer (50 mM Tris pH 7.4, 250 mM NaCl, 0.1% SDS, 2 mM DTT, 0.5% NP40, 1x protease inhibitors ()) were added to previously harvested cell pellets and thoroughly mixed. Subsequently the suspension was incubated for 30 minutes on ice and centrifuged at 10,000 g at 4°C for 15 minutes. Protein concentration of the supernatant was determined with the Pierce BCA Protein Assay Kit (Thermo Fisher, USA) according to manufacturer’s instructions. 20 μ g of protein were denatured by addition of Laemmli buffer (4x: 20% glycerol, 2% SDS, 25mM Tris pH 6.8, 3 mM bromophenol blue, 100 mM DTT) and incubation at 95°C for 5 minutes. Proteins were separated via sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) at 70 V for 20 minutes followed by 120 V for 2 hours. The percentage of the gel was chosen based on the size of the protein of interest (Tab. 9).

Table 9: Gel Recipes for SDS-PAGE

Reagent	Separating Gel					Stacking Gel
	5%	6%	8%	10%	12%	
Distilled water	8.4 mL	7.9 mL	6.9 mL	5.9 mL	4.9 mL	6.8 mL
30% Acrylamide solution	2.5 mL	3.0 mL	4 mL	5 mL	6 mL	1.7 mL
1.5M Tris pH 8.8	3.8 mL	3.8 mL	3.8 mL	3.8 mL	3.8 mL	-
1M Tris pH 6.8	-	-	-	-	-	1.25 mL
SDS 10%	150 μ L	150 μ L	150 μ L	150 μ L	150 μ L	100 μ L
10% Ammonium persulfate (APS)	100 μ L	100 μ L	100 μ L	100 μ L	100 μ L	100 μ L
TEMED	10 μ L	10 μ L	10 μ L	10 μ L	10 μ L	20 μ L

Proteins were transferred from the gel to an Immuno Blot polyvinylidene difluoride (PVDF) membrane (Bio-Rad Laboratories, USA) via Western blot at 400 mA/300 V for 2 hours. After blocking with 5% bovine serum albumin (BSA) in TBST (20 mM Tris pH 7.4, 150 mM NaCl,

0.1% Tween 20) for 1 hour, the membrane was incubated over night at 4°C with the primary antibody. The next morning, after three washing steps with TBST for 5 minutes each, the membrane was incubated with secondary HRP-conjugated antibody for 3 hours at 4°C. The membrane was treated with Pierce ECL Western Blotting Substrate (Thermo Fisher, USA) according to manufacturer's instructions, pictures were taken with ChemiDoc XRS+ (Bio-Rad Laboratories, USA), and images were processed with Image Lab (Bio-Rad Laboratories, USA).

Table 10: Antibodies

Name	Source	Identifier
Primary Antibodies		
Rabbit Anti-Lamin B1 antibody	Abcam United Kingdom	[EPR8985(B)] (ab133741)
Rabbit Anti-p16 antibody	Abcam United Kingdom	[EP4353Y(3)] (ab81278)
Rabbit Anti-pRb antibody	Cell Signaling Technology, USA	#9307
Mouse Anti-Rassf1a antibody	Abcam United Kingdom	[3F3] (ab23950)
Rabbit Anti-Rb antibody	Abcam United Kingdom	[EPR17512] (ab181616)
Secondary Antibodies		
Goat Anti-Rabbit HRP Conjugate	Bio-Rad Laboratories, USA	170-6515
Goat Anti-Mouse HRP Conjugate	Bio-Rad Laboratories, USA	170-6516

5.10 Cell Counting Assay

10⁵ LV transduced Huh7 cells were plated per well onto a 6-well cell culture dish. 24 hours, 72 hours, and 120 hours after plating cells were harvested with Trypsin/EDTA and subsequent resuspension of the pellet in 1 mL PBS. The concentration of cells was determined with a TC 20 Automated Cell Counter (Bio-Rad Laboratories, USA). Statistical analysis was performed with Prism 8 (GraphPad, USA). The p-value was calculated via an unpaired parametric t-test with Welsh's correction.

5.11 Live Cell Imaging

8 hours prior to monitoring, 5×10^5 LV transduced Huh7 cells were plated per well onto a chambered μ -Slide 8 Well dish (Ibidi, Germany), which has been previously coated overnight at 37°C with poly-L-lysine solution (Sigma-Aldrich, USA). After attachment, cells were coincubated for 4 hours with 25 nM SiR-Hoechst (Spirochrome AG, Switzerland). Cells were monitored using a Nikon Eclipse Ti-E wide-field microscope (Nikon, Switzerland) equipped with a DAPI/eGFP/ TRITC/Cy5 filter set (Chroma, USA) and a 40× NA 1.3 objective (mitotic timing) and recorded with an Orca Flash 4.0 CMOS camera (Hamamatsu, Japan) and NIS software. To measure mitotic timing of cells, images were taken every 3 minutes for 20 hours with 2 μ m z-stacks. To calculate the number of mitoses per cell, the total number of cells at the beginning was determined and subsequently the mitoses occurring during the 20h were counted. Finally, the number of mitoses was divided by the total number of cells at the beginning of the experiment. Statistical analysis was performed with Prism 8 (GraphPad, USA). The p-value was calculated via an unpaired parametric t-test with Welch's correction.

5.12 Cell Cycle Analysis of Live Cells

LV transduced Huh7 cells were harvested and counted. 1×10^6 cells were resuspended in 1 mL cell culture medium containing 5 μ g/mL Hoechst 33342 dye (Thermo Fisher, USA). The cells were incubated in a water bath at 37°C for 1 hour, while being briefly vortexed every 15 minutes. The cells were transferred to FACS without further washing steps. Flow cytometry was performed using a BD LSRFortessa (Becton Dickinson, USA) and analysis was performed with FlowJo V10 (Becton Dickinson, USA).

5.13 Senescence β -Galactosidase Staining

One day prior to staining, 5×10^4 LV transduced Huh7 cells were plated per well onto a 12-well cell culture dish. Staining was performed overnight with Senescence β -Galactosidase Staining

Kit (Cell Signaling Technology, USA) according to manufacturer's instructions. Pictures were taken at 200x magnification with EVOS M5000 Imaging System (Thermo Fisher, USA).

5.14 Flow Cytometry of Dazl-Snrpn Reporter Cells

After harvesting, cell pellets were resuspended in 1 mL FACS buffer (PBS with 2% BSA and 5 mM EDTA). Flow cytometry was performed on a BD Accuri C6 (Becton Dickinson, USA). Mean and median fluorescence intensity (MFI) of GFP was determined using FlowJo V10 (Becton Dickinson, USA). MFI ratios were calculated by dividing the MFI of the target sample through the MFI of the negative control (mock transfected Dazl-Snrpn mESCs): MFI (sample)/MFI (control). Statistical analysis was performed with Prism 8 (GraphPad, USA). The p-value was calculated via an unpaired parametric t-test with Welsh's correction.

5.15 Animal Housing and CAx Treatment

C57Bl/6J mice were kept at 12-hour day and night cycles in a specific pathogen-free facility (SPF), and they were fed standard chow. Treatment with cold temperature and antibiotics was performed over 30 days at 6°C and by *ad libitum* administration of the antibiotics cocktail (100 µg per mL Neomycin, 50 µg per mL Streptomycin, 100 U per mL Penicillin, 50 µg per mL Vancomycin, 100 µg per mL Metronidazole, 1 mg per mL Bacitracin, 125 µg per mL Ciprofloxacin, 100 µg per mL Ceftazidime and 170 µg per mL Gentamycin⁶⁸¹) in the drinking water, which was freshly replaced every second day.

5.16 Whole Genome Bisulfite Sequencing

DNA was isolated from perigonadal visceral adipocyte tissue (pgVAT) from one replicate of pooled CAx treated mice or RT control mice using DNeasy Blood & Tissue Kit (QIAGEN, Germany). cDNA libraries were constructed by the Genomic platform of Novogene, China, adding 26 ng lambda DNA. Libraries were sequenced using paired-end sequencing (150 nt-long) on Illumina HiSeq4000. FastQ reads were mapped to the UCSC reference genome (mm10

last update) using Bismark version 0.12.5⁶⁸² and bowtie2 version⁶⁸³ with standard settings, except that any reads mapping to more than one location in the genome (ambiguous reads) were discarded (m= 1).

DMRs were identified with a sliding-window approach using swDMR software version 1.6.2⁶⁴⁰. The window was set to 1000 bp, sliding 100 bp per step. Regions with at least 10 CpGs and a methylation level (ML) difference of at least 10% (p value < 0.05 using a Fisher test) were considered DMRs.

5.17 RNA Sequencing

RNA was isolated from pgVAT. cDNA libraries were constructed by the Genomic platform of the University of Geneva (iGe3) using the Illumina TruSeq RNA Sample Preparation Kit according to the manufacturer's protocol. Libraries were sequenced using single-end sequencing (50 nt-long) on Illumina HiSeq4000. FastQ reads were mapped to the UCSC reference genome (mm10, last update) using STAR version 2.6.1b⁶⁸⁴ with standard settings, except that any reads mapping to more than one location in the genome (ambiguous reads) were discarded (m = 1).

A unique gene model was used to quantify reads per gene. Briefly, the model considers all annotated exons of all annotated protein coding isoforms of a gene to create a unique gene where the genomic region of all exons are considered coming from the same RNA molecule and merged together.

All reads overlapping the exons of each unique gene model were reported using featureCounts version 1.6.2⁶⁸⁵. Gene expressions were reported as raw counts and in parallel normalized in RPKM in order to filter out genes with low expression value (0.5 RPKM) before calling differentially expressed genes. Library size normalization and differential gene expression calculation was performed using the package edgeR⁶⁸⁶ designed for the R software. As

biological coefficient of variation (bcv) 0.1 and performance of an exactTest were chosen. Only genes having a significant fold-change (Benjamini-Hochberg corrected p-value < 0.05) were considered for the rest of the RNAseq analysis.

5.18 Differentiation of 3T3-L1 Pre-Adipocytes

3T3-L1 pre-adipocytes were grown until complete confluency prior to differentiation. For ‘whitening’, 500 μ M 3-isobutyl-1-methylxanthine (IMBx), 1 μ M dexamethasone, and 850 nM insulin bovine were added to the culture medium on day 1. On day 3, the medium was exchanged and regular medium with 850 nM insulin bovine was added. On day 5 and on day 7 medium was exchanged without further additions. Cells were harvested on day 9. The ‘browning’ followed the same steps, except additional 1 nM T3 antigen, 125 μ M indomethacin, and 1 μ M rosiglitazone were added on day 1, as well as additional 1 nM T3 antigen and 1 μ M rosiglitazone on day 3 and on day 5.

6 References

- 1 Koonin, E. V. & Novozhilov, A. S. Origin and evolution of the genetic code: the universal enigma. *IUBMB Life* **61**, 99-111, doi:10.1002/iub.146 (2009).
- 2 Crick, F. Ideas on protein synthesis. *Unpublished note. Wellcome Library, PPCRI/H/2/6*. <https://wellcomelibrary.org/item/b18174139> (1956).
- 3 Crick, F. Central Dogma of Molecular Biology. *Nature* **227**, 561-563, doi:10.1038/227561a0 (1970).
- 4 Schuler, G. *et al.* A gene map of the human genome. *Science (New York, N.Y.)* **274**, 540-546 (1996).
- 5 Antequera, F. & Bird, A. Predicting the total number of human genes. *Nature genetics* **8**, 114 (1994).
- 6 Fields, C., Adams, M. D., White, O. & Venter, J. C. How many genes in the human genome? *Nature genetics* **7**, 345 (1994).
- 7 Consortium, I. H. G. S. Initial sequencing and analysis of the human genome. *Nature* **409**, 860 (2001).
- 8 Consortium, I. H. G. S. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931 (2004).
- 9 Pertea, M. *et al.* Thousands of large-scale RNA sequencing experiments yield a comprehensive new human gene list and reveal extensive transcriptional noise. *bioRxiv*, 332825, doi:10.1101/332825 (2018).
- 10 Berget, S. M., Moore, C. & Sharp, P. A. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 3171-3175, doi:10.1073/pnas.74.8.3171 (1977).
- 11 Chow, L. T., Gelinas, R. E., Broker, T. R. & Roberts, R. J. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* **12**, 1-8, doi:10.1016/0092-8674(77)90180-5 (1977).
- 12 Jacob, F., Perrin, D., Sanchez, C. & Monod, J. [Operon: a group of genes with the expression coordinated by an operator]. *C R Hebd Seances Acad Sci* **250**, 1727-1729 (1960).
- 13 Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* **3**, 318-356, doi:10.1016/s0022-2836(61)80072-7 (1961).
- 14 Spieth, J., Brooke, G., Kuersten, S., Lea, K. & Blumenthal, T. Operons in *C. elegans*: polycistronic mRNA precursors are processed by trans-splicing of SL2 to downstream coding regions. *Cell* **73**, 521-532, doi:10.1016/0092-8674(93)90139-h (1993).
- 15 Brogna, S. & Ashburner, M. The Adh-related gene of *Drosophila melanogaster* is expressed as a functional dicistronic messenger RNA: multigenic transcription in higher organisms. *The EMBO journal* **16**, 2023-2031, doi:10.1093/emboj/16.8.2023 (1997).

- 16 Hampsey, M. Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol Mol Biol Rev* **62**, 465-503 (1998).
- 17 Haberle, V. & Stark, A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nature Reviews Molecular Cell Biology* **19**, 621-637, doi:10.1038/s41580-018-0028-8 (2018).
- 18 Ponjavic, J. *et al.* Transcriptional and structural impact of TATA-initiation site spacing in mammalian core promoters. *Genome biology* **7**, R78, doi:10.1186/gb-2006-7-8-R78 (2006).
- 19 Patikoglou, G. A. *et al.* TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes & development* **13**, 3217-3230, doi:10.1101/gad.13.24.3217 (1999).
- 20 Lifton, R. P., Goldberg, M. L., Karp, R. W. & Hogness, D. S. The organization of the histone genes in *Drosophila melanogaster*: functional and evolutionary implications. *Cold Spring Harbor symposia on quantitative biology* **42 Pt 2**, 1047-1051, doi:10.1101/sqb.1978.042.01.105 (1978).
- 21 FitzGerald, P. C., Sturgill, D., Shyakhtenko, A., Oliver, B. & Vinson, C. Comparative genomics of *Drosophila* and human core promoters. *Genome biology* **7**, R53, doi:10.1186/gb-2006-7-7-r53 (2006).
- 22 Smale, S. T. & Baltimore, D. The "initiator" as a transcription control element. *Cell* **57**, 103-113, doi:10.1016/0092-8674(89)90176-1 (1989).
- 23 Chalkley, G. E. & Verrijzer, C. P. DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250-TAF(II)150 complex recognizes the initiator. *The EMBO journal* **18**, 4835-4845, doi:10.1093/emboj/18.17.4835 (1999).
- 24 Vo Ngoc, L., Cassidy, C. J., Huang, C. Y., Duttke, S. H. & Kadonaga, J. T. The human initiator is a distinct and abundant element that is precisely positioned in focused core promoters. *Genes & development* **31**, 6-11, doi:10.1101/gad.293837.116 (2017).
- 25 Burke, T. W. & Kadonaga, J. T. *Drosophila* TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes & development* **10**, 711-724, doi:10.1101/gad.10.6.711 (1996).
- 26 Burke, T. W. & Kadonaga, J. T. The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAFII60 of *Drosophila*. *Genes & development* **11**, 3020-3031, doi:10.1101/gad.11.22.3020 (1997).
- 27 Louder, R. K. *et al.* Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. *Nature* **531**, 604-609, doi:10.1038/nature17394 (2016).
- 28 Nogales, E., Louder, R. K. & He, Y. Structural Insights into the Eukaryotic Transcription Initiation Machinery. *Annual Review of Biophysics* **46**, 59-83, doi:10.1146/annurev-biophys-070816-033751 (2017).
- 29 Orphanides, G., Lagrange, T. & Reinberg, D. The general transcription factors of RNA polymerase II. *Genes & development* **10**, 2657-2683, doi:10.1101/gad.10.21.2657 (1996).

- 30 Sainsbury, S., Bernecky, C. & Cramer, P. Structural basis of transcription initiation by RNA polymerase II. *Nat Rev Mol Cell Biol* **16**, 129-143, doi:10.1038/nrm3952 (2015).
- 31 He, Y. *et al.* Near-atomic resolution visualization of human transcription promoter opening. *Nature* **533**, 359-365, doi:10.1038/nature17970 (2016).
- 32 Kadonaga, J. T. Perspectives on the RNA polymerase II core promoter. *Wiley Interdiscip Rev Dev Biol* **1**, 40-51, doi:10.1002/wdev.21 (2012).
- 33 Allen, B. L. & Taatjes, D. J. The Mediator complex: a central integrator of transcription. *Nat Rev Mol Cell Biol* **16**, 155-166, doi:10.1038/nrm3951 (2015).
- 34 Banerji, J., Rusconi, S. & Schaffner, W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299-308, doi:10.1016/0092-8674(81)90413-x (1981).
- 35 Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430-435, doi:10.1038/nature09380 (2010).
- 36 Weintraub, A. S. *et al.* YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell* **171**, 1573-1588.e1528, doi:10.1016/j.cell.2017.11.008 (2017).
- 37 Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319, doi:10.1016/j.cell.2013.03.035 (2013).
- 38 Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-947, doi:10.1016/j.cell.2013.09.053 (2013).
- 39 Rao, S. S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680 (2014).
- 40 Proudhon, C. *et al.* Active and inactive enhancers cooperate to exert localized and long-range control of gene regulation. *Cell reports* **15**, 2159-2169 (2016).
- 41 Song, W., Sharan, R. & Ovcharenko, I. The first enhancer in an enhancer chain safeguards subsequent enhancer-promoter contacts from a distance. *Genome biology* **20**, 197, doi:10.1186/s13059-019-1808-y (2019).
- 42 Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science (New York, N.Y.)* **316**, 1497-1502 (2007).
- 43 Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science (New York, N.Y.)* **295**, 1306-1311, doi:10.1126/science.1067799 (2002).
- 44 Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (New York, N.Y.)* **326**, 289-293, doi:10.1126/science.1181369 (2009).
- 45 Zhao, Z. *et al.* Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nature genetics* **38**, 1341-1347, doi:10.1038/ng1891 (2006).

- 46 Dostie, J. *et al.* Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome research* **16**, 1299-1309, doi:10.1101/gr.5571506 (2006).
- 47 Arnold, C. D. *et al.* Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science (New York, N.Y.)* **339**, 1074-1077, doi:10.1126/science.1232542 (2013).
- 48 Catarino, R. R. & Stark, A. Assessing sufficiency and necessity of enhancer activities for gene expression and the mechanisms of transcription activation. *Genes & development* **32**, 202-223, doi:10.1101/gad.310367.117 (2018).
- 49 Fong, A. P. & Tapscott, S. J. Skeletal muscle programming and re-programming. *Current Opinion in Genetics & Development* **23**, 568-573, doi:<https://doi.org/10.1016/j.gde.2013.05.002> (2013).
- 50 Lee, T. I. & Young, R. A. Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237-1251 (2013).
- 51 Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic acids research* **44**, D279-D285 (2015).
- 52 Letunic, I., Doerks, T. & Bork, P. SMART: recent updates, new developments and status in 2015. *Nucleic acids research* **43**, D257-D260 (2014).
- 53 Johnson, P. F. & McKnight, S. L. Eukaryotic transcriptional regulatory proteins. *Annual review of biochemistry* **58**, 799-839 (1989).
- 54 Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650-665, doi:<https://doi.org/10.1016/j.cell.2018.01.029> (2018).
- 55 Han, B. Y., Foo, C.-S., Wu, S. & Cyster, J. G. The C2H2-ZF transcription factor Zfp335 recognizes two consensus motifs using separate zinc finger arrays. *Genes & development* **30**, 1509-1514, doi:10.1101/gad.279406.116 (2016).
- 56 Mesika, A., Ben-Dor, S., Laviad, E. L. & Futerman, A. H. A new functional motif in Hox domain-containing ceramide synthases: identification of a novel region flanking the Hox and TLC domains essential for activity. *The Journal of biological chemistry* **282**, 27366-27373, doi:10.1074/jbc.M703487200 (2007).
- 57 Ecco, G., Imbeault, M. & Trono, D. KRAB zinc finger proteins. *Development (Cambridge, England)* **144**, 2719-2729 (2017).
- 58 Panne, D. The enhanceosome. *Current opinion in structural biology* **18**, 236-242 (2008).
- 59 Frietze, S. & Farnham, P. J. in *A handbook of transcription factors* 261-277 (Springer, 2011).
- 60 Rosenfeld, M. G., Lunyak, V. V. & Glass, C. K. Sensors and signals: a coactivator/corepressor/epigenetic code for integrating signal-dependent programs of transcriptional response. *Genes & development* **20**, 1405-1428 (2006).

- 61 Meijnsing, S. H. *et al.* DNA binding site sequence directs glucocorticoid receptor structure and activity. *Science (New York, N.Y.)* **324**, 407-410 (2009).
- 62 Wong, K. H. & Struhl, K. The Cyc8–Tup1 complex inhibits transcription primarily by masking the activation domain of the recruiting protein. *Genes & development* **25**, 2525-2539 (2011).
- 63 Jolma, A. *et al.* DNA-binding specificities of human transcription factors. *Cell* **152**, 327-339 (2013).
- 64 Jolma, A. *et al.* DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature* **527**, 384 (2015).
- 65 Rhee, H. S. & Pugh, B. F. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* **147**, 1408-1419 (2011).
- 66 Hunt, R. W. & Wasserman, W. W. Non-targeted transcription factors motifs are a systemic component of ChIP-seq datasets. *Genome biology* **15**, 412 (2014).
- 67 Stormo, G. D. & Zhao, Y. Determining the specificity of protein–DNA interactions. *Nature Reviews Genetics* **11**, 751 (2010).
- 68 Cusanovich, D. A., Pavlovic, B., Pritchard, J. K. & Gilad, Y. The functional consequences of variation in transcription factor binding. *PLoS genetics* **10**, e1004226 (2014).
- 69 Fu, Y., Sinha, M., Peterson, C. L. & Weng, Z. The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS genetics* **4**, e1000138 (2008).
- 70 Kim, T. H. *et al.* Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128**, 1231-1245 (2007).
- 71 Morgunova, E. & Taipale, J. Structural perspective of cooperative transcription factor binding. *Current opinion in structural biology* **47**, 1-8 (2017).
- 72 Stampfel, G. *et al.* Transcriptional regulators form diverse groups with context-dependent regulatory functions. *Nature* **528**, 147, doi:10.1038/nature15545
<https://www.nature.com/articles/nature15545#supplementary-information> (2015).
- 73 Soufi, A. *et al.* Pioneer Transcription Factors Target Partial DNA Motifs on Nucleosomes to Initiate Reprogramming. *Cell* **161**, 555-568, doi:10.1016/j.cell.2015.03.017 (2015).
- 74 Heitz, E. *Das heterochromatin der moose.* (Bornträger, 1928).
- 75 Waddington, C. H. The epigenotype. *Endeavour* **1**, 18-20 (1942).
- 76 Waddington, C. H. Canalization of development and the inheritance of acquired characters. *Nature* **150**, 563 (1942).
- 77 Waddington, C. H. Towards a theoretical biology. *Nature* **218**, 525 (1968).

- 78 Wu, C. & Morris, J. R. Genes, genetics, and epigenetics: a correspondence. *Science (New York, N.Y.)* **293**, 1103-1105, doi:10.1126/science.293.5532.1103 (2001).
- 79 Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nature Reviews Genetics* **17**, 487, doi:10.1038/nrg.2016.59 (2016).
- 80 Daskalos, A. *et al.* Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer. *International Journal of Cancer* **124**, 81-87 (2009).
- 81 Zaratiegui, M., Irvine, D. V. & Martienssen, R. A. Noncoding RNAs and gene silencing. *Cell* **128**, 763-776 (2007).
- 82 Dodge, J. E. *et al.* Inactivation of Dnmt3b in mouse embryonic fibroblasts results in DNA hypomethylation, chromosomal instability, and spontaneous immortalization. *Journal of Biological Chemistry* **280**, 17986-17991 (2005).
- 83 Perkel, J. M. Visiting "noncodarnia". *Biotechniques* **54**, 301, 303-304, doi:10.2144/000114037 (2013).
- 84 Mercer, T. R., Dinger, M. E. & Mattick, J. S. Long non-coding RNAs: insights into functions. *Nature reviews. Genetics* **10**, 155-159, doi:10.1038/nrg2521 (2009).
- 85 Dinger, M. E., Amaral, P. P., Mercer, T. R. & Mattick, J. S. Pervasive transcription of the eukaryotic genome: functional indices and conceptual implications. *Brief Funct Genomic Proteomic* **8**, 407-423, doi:10.1093/bfpg/elp038 (2009).
- 86 Goodrich, J. A. & Kugel, J. F. Non-coding-RNA regulators of RNA polymerase II transcription. *Nat Rev Mol Cell Biol* **7**, 612-616, doi:10.1038/nrm1946 (2006).
- 87 Wutz, A. & Gribnau, J. X inactivation Xplained. *Curr Opin Genet Dev* **17**, 387-393, doi:10.1016/j.gde.2007.08.001 (2007).
- 88 Umlauf, D. *et al.* Imprinting along the Kcnq1 domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes. *Nature genetics* **36**, 1296-1300, doi:10.1038/ng1467 (2004).
- 89 Bolzer, A. *et al.* Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol* **3**, e157, doi:10.1371/journal.pbio.0030157 (2005).
- 90 Nagano, T. *et al.* Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* **547**, 61-67, doi:10.1038/nature23001 (2017).
- 91 Cremer, T. & Cremer, M. Chromosome territories. *Cold Spring Harb Perspect Biol* **2**, a003889, doi:10.1101/cshperspect.a003889 (2010).
- 92 Bickmore, W. A. & van Steensel, B. Genome architecture: domain organization of interphase chromosomes. *Cell* **152**, 1270-1284, doi:10.1016/j.cell.2013.02.001 (2013).
- 93 Osborne, C. S. *et al.* Active genes dynamically colocalize to shared sites of ongoing transcription. *Nature genetics* **36**, 1065-1071, doi:10.1038/ng1423 (2004).

- 94 Shah, S. *et al.* Dynamics and Spatial Genomics of the Nascent Transcriptome by Intron seqFISH. *Cell* **174**, 363-376.e316, doi:10.1016/j.cell.2018.05.035 (2018).
- 95 van Steensel, B. & Belmont, A. S. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* **169**, 780-791, doi:10.1016/j.cell.2017.04.022 (2017).
- 96 Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380, doi:10.1038/nature11082 (2012).
- 97 Symmons, O. *et al.* Functional and topological characteristics of mammalian regulatory domains. *Genome research* **24**, 390-400, doi:10.1101/gr.163519.113 (2014).
- 98 Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**, 458-472, doi:10.1016/j.cell.2012.01.010 (2012).
- 99 Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-385, doi:10.1038/nature11049 (2012).
- 100 Dali, R. & Blanchette, M. A critical assessment of topologically associating domain prediction tools. *Nucleic acids research* **45**, 2994-3005, doi:10.1093/nar/gkx145 (2017).
- 101 Dixon, J. R., Gorkin, D. U. & Ren, B. Chromatin Domains: The Unit of Chromosome Organization. *Molecular cell* **62**, 668-680, doi:10.1016/j.molcel.2016.05.018 (2016).
- 102 Zhan, Y. *et al.* Reciprocal insulation analysis of Hi-C data shows that TADs represent a functionally but not structurally privileged scale in the hierarchical folding of chromosomes. *Genome research* **27**, 479-490 (2017).
- 103 Bonev, B. *et al.* Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**, 557-572.e524, doi:10.1016/j.cell.2017.09.043 (2017).
- 104 Zuin, J. *et al.* Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 996-1001, doi:10.1073/pnas.1317788111 (2014).
- 105 Vietri Rudan, M. *et al.* Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell reports* **10**, 1297-1309, doi:10.1016/j.celrep.2015.02.004 (2015).
- 106 Fudenberg, G. *et al.* Formation of Chromosomal Domains by Loop Extrusion. *Cell reports* **15**, 2038-2049, doi:10.1016/j.celrep.2016.04.085 (2016).
- 107 Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E6456-6465, doi:10.1073/pnas.1518552112 (2015).
- 108 Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R. & Darzacq, X. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *eLife* **6**, doi:10.7554/eLife.25776 (2017).

- 109 Guo, Y. *et al.* CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* **162**, 900-910, doi:10.1016/j.cell.2015.07.038 (2015).
- 110 Flavahan, W. A. *et al.* Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**, 110-114, doi:10.1038/nature16490 (2016).
- 111 Franke, M. *et al.* Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* **538**, 265-269, doi:10.1038/nature19800 (2016).
- 112 Lupianez, D. G. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012-1025, doi:10.1016/j.cell.2015.04.004 (2015).
- 113 Lupianez, D. G., Spielmann, M. & Mundlos, S. Breaking TADs: How Alterations of Chromatin Domains Result in Disease. *Trends in genetics : TIG* **32**, 225-237, doi:10.1016/j.tig.2016.01.003 (2016).
- 114 Dixon, J. R. *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331-336, doi:10.1038/nature14222 (2015).
- 115 Acemel, R. D., Maeso, I. & Gómez-Skarmeta, J. L. Topologically associated domains: a successful scaffold for the evolution of gene regulation in animals. *Wiley Interdisciplinary Reviews: Developmental Biology* **6**, e265 (2017).
- 116 Harmston, N. *et al.* Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation. *Nature communications* **8**, 441 (2017).
- 117 Kaaij, L. J. T., van der Weide, R. H., Ketting, R. F. & de Wit, E. Systemic Loss and Gain of Chromatin Architecture throughout Zebrafish Development. *Cell reports* **24**, 1-10.e14, doi:10.1016/j.celrep.2018.06.003 (2018).
- 118 Ahmed, K. *et al.* Global chromatin architecture reflects pluripotency and lineage commitment in the early mouse embryo. *PLoS One* **5**, e10531, doi:10.1371/journal.pone.0010531 (2010).
- 119 Flyamer, I. M. *et al.* Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* **544**, 110-114, doi:10.1038/nature21711 (2017).
- 120 Du, Z. *et al.* Allelic reprogramming of 3D chromatin architecture during early mammalian development. *Nature* **547**, 232-235, doi:10.1038/nature23263 (2017).
- 121 Ke, Y. *et al.* 3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis. *Cell* **170**, 367-381.e320, doi:10.1016/j.cell.2017.06.029 (2017).
- 122 Naumova, N. *et al.* Organization of the mitotic chromosome. *Science (New York, N.Y.)* **342**, 948-953, doi:10.1126/science.1236083 (2013).
- 123 Zheng, H. & Xie, W. The role of 3D genome organization in development and cell differentiation. *Nature Reviews Molecular Cell Biology* **20**, 535-550, doi:10.1038/s41580-019-0132-4 (2019).

- 124 Fraser, J. *et al.* Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. *Mol Syst Biol* **11**, 852, doi:10.15252/msb.20156492 (2015).
- 125 Gorkin, D. U., Leung, D. & Ren, B. The 3D genome in transcriptional regulation and pluripotency. *Cell stem cell* **14**, 762-775, doi:10.1016/j.stem.2014.05.017 (2014).
- 126 Smallwood, A. & Ren, B. Genome organization and long-range regulation of gene expression by enhancers. *Curr Opin Cell Biol* **25**, 387-394, doi:10.1016/j.ceb.2013.02.005 (2013).
- 127 Rubin, A. J. *et al.* Lineage-specific dynamic and pre-established enhancer–promoter contacts cooperate in terminal differentiation. *Nature genetics* **49**, 1522-1528, doi:10.1038/ng.3935 (2017).
- 128 Zhou, Y. B., Gerchman, S. E., Ramakrishnan, V., Travers, A. & Muyltermans, S. Position and orientation of the globular domain of linker histone H5 on the nucleosome. *Nature* **395**, 402-405, doi:10.1038/26521 (1998).
- 129 Papamichos-Chronakis, M. & Peterson, C. L. Chromatin and the genome integrity network. *Nature Reviews Genetics* **14**, 62-75, doi:10.1038/nrg3345 (2013).
- 130 Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nature Reviews Genetics* **20**, 207-220, doi:10.1038/s41576-018-0089-8 (2019).
- 131 Boyle, A. P. *et al.* High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**, 311-322, doi:10.1016/j.cell.2007.12.014 (2008).
- 132 Hesselberth, J. R. *et al.* Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nature methods* **6**, 283-289, doi:10.1038/nmeth.1313 (2009).
- 133 Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75-82, doi:10.1038/nature11232 (2012).
- 134 Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature methods* **10**, 1213-1218, doi:10.1038/nmeth.2688 (2013).
- 135 Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nature methods* **14**, 959-962, doi:10.1038/nmeth.4396 (2017).
- 136 Mieczkowski, J. *et al.* MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. *Nat Commun* **7**, 11485, doi:10.1038/ncomms11485 (2016).
- 137 Mueller, B. *et al.* Widespread changes in nucleosome accessibility without changes in nucleosome occupancy during a rapid transcriptional induction. *Genes & development* **31**, 451-462, doi:10.1101/gad.293118.116 (2017).

- 138 Krebs, A. R. *et al.* Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. *Molecular cell* **67**, 411-422.e414, doi:10.1016/j.molcel.2017.06.027 (2017).
- 139 Kelly, T. K. *et al.* Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome research* **22**, 2497-2506, doi:10.1101/gr.143008.112 (2012).
- 140 Clapier, C. R., Iwasa, J., Cairns, B. R. & Peterson, C. L. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nature Reviews Molecular Cell Biology* **18**, 407, doi:10.1038/nrm.2017.26
<https://www.nature.com/articles/nrm.2017.26#supplementary-information> (2017).
- 141 Sirinakis, G. *et al.* The RSC chromatin remodelling ATPase translocates DNA with high force and small step size. *The EMBO journal* **30**, 2364-2372, doi:10.1038/emboj.2011.141 (2011).
- 142 Harada, B. T. *et al.* Stepwise nucleosome translocation by RSC remodeling complexes. *eLife* **5**, doi:10.7554/eLife.10051 (2016).
- 143 Singleton, M. R., Dillingham, M. S. & Wigley, D. B. Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev Biochem* **76**, 23-50, doi:10.1146/annurev.biochem.76.052305.115300 (2007).
- 144 Deindl, S. *et al.* ISWI remodelers slide nucleosomes with coordinated multi-base-pair entry steps and single-base-pair exit steps. *Cell* **152**, 442-452, doi:10.1016/j.cell.2012.12.040 (2013).
- 145 Velankar, S. S., Soultanas, P., Dillingham, M. S., Subramanya, H. S. & Wigley, D. B. Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell* **97**, 75-84, doi:10.1016/s0092-8674(00)80716-3 (1999).
- 146 Sen, P. *et al.* The SnAC domain of SWI/SNF is a histone anchor required for remodeling. *Molecular and cellular biology* **33**, 360-370, doi:10.1128/mcb.00922-12 (2013).
- 147 Corona, D. F. & Tamkun, J. W. Multiple roles for ISWI in transcription, chromosome organization and DNA replication. *Biochim Biophys Acta* **1677**, 113-119, doi:10.1016/j.bbaexp.2003.09.018 (2004).
- 148 Clapier, C. R. & Cairns, B. R. Regulation of ISWI involves inhibitory modules antagonized by nucleosomal epitopes. *Nature* **492**, 280-284, doi:10.1038/nature11625 (2012).
- 149 Grune, T. *et al.* Crystal structure and functional analysis of a nucleosome recognition module of the remodeling factor ISWI. *Molecular cell* **12**, 449-460, doi:10.1016/s1097-2765(03)00273-9 (2003).
- 150 Hauk, G., McKnight, J. N., Nodelman, I. M. & Bowman, G. D. The chromodomains of the Chd1 chromatin remodeler regulate DNA access to the ATPase motor. *Molecular cell* **39**, 711-723, doi:10.1016/j.molcel.2010.08.012 (2010).

- 151 Ryan, D. P., Sundaramoorthy, R., Martin, D., Singh, V. & Owen-Hughes, T. The DNA-binding domain of the Chd1 chromatin-remodelling enzyme contains SANT and SLIDE domains. *The EMBO journal* **30**, 2596-2609, doi:10.1038/emboj.2011.166 (2011).
- 152 Tran, H. G., Steger, D. J., Iyer, V. R. & Johnson, A. D. The chromo domain protein chd1p from budding yeast is an ATP-dependent chromatin-modifying factor. *The EMBO journal* **19**, 2323-2331, doi:10.1093/emboj/19.10.2323 (2000).
- 153 Kunert, N. & Brehm, A. Novel Mi-2 related ATP-dependent chromatin remodelers. *Epigenetics* **4**, 209-211, doi:10.4161/epi.8933 (2009).
- 154 McKnight, J. N., Jenkins, K. R., Nodelman, I. M., Escobar, T. & Bowman, G. D. Extranucleosomal DNA binding directs nucleosome sliding by Chd1. *Molecular and cellular biology* **31**, 4746-4759, doi:10.1128/mcb.05735-11 (2011).
- 155 Bargaje, R. *et al.* Proximity of H2A.Z containing nucleosome to the transcription start site influences gene expression levels in the mammalian liver and brain. *Nucleic acids research* **40**, 8965-8978, doi:10.1093/nar/gks665 (2012).
- 156 Rege, M. *et al.* Chromatin Dynamics and the RNA Exosome Function in Concert to Regulate Transcriptional Homeostasis. *Cell reports* **13**, 1610-1622, doi:10.1016/j.celrep.2015.10.030 (2015).
- 157 Morrison, A. J. & Shen, X. Chromatin remodelling beyond transcription: the INO80 and SWR1 complexes. *Nat Rev Mol Cell Biol* **10**, 373-384, doi:10.1038/nrm2693 (2009).
- 158 Bao, Y. & Shen, X. SnapShot: Chromatin remodeling: INO80 and SWR1. *Cell* **144**, 158-158.e152, doi:10.1016/j.cell.2010.12.024 (2011).
- 159 Watanabe, S. *et al.* Structural analyses of the chromatin remodelling enzymes INO80-C and SWR-C. *Nat Commun* **6**, 7108, doi:10.1038/ncomms8108 (2015).
- 160 Jha, S. & Dutta, A. RVB1/RVB2: running rings around molecular biology. *Molecular cell* **34**, 521-533, doi:10.1016/j.molcel.2009.05.016 (2009).
- 161 Szerlong, H. *et al.* The HSA domain binds nuclear actin-related proteins to regulate chromatin-remodeling ATPases. *Nature structural & molecular biology* **15**, 469-476, doi:10.1038/nsmb.1403 (2008).
- 162 Mizuguchi, G. *et al.* ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science (New York, N.Y.)* **303**, 343-348, doi:10.1126/science.1090701 (2004).
- 163 Papamichos-Chronakis, M., Watanabe, S., Rando, O. J. & Peterson, C. L. Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. *Cell* **144**, 200-213, doi:10.1016/j.cell.2010.12.021 (2011).
- 164 Ranjan, A. *et al.* H2A histone-fold and DNA elements in nucleosome activate SWR1-mediated H2A.Z replacement in budding yeast. *eLife* **4**, e06845, doi:10.7554/eLife.06845 (2015).

- 165 Mohrmann, L. & Verrijzer, C. P. Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes. *Biochim Biophys Acta* **1681**, 59-73, doi:10.1016/j.bbaexp.2004.10.005 (2005).
- 166 Schubert, H. L. *et al.* Structure of an actin-related subcomplex of the SWI/SNF chromatin remodeler. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 3345-3350, doi:10.1073/pnas.1215379110 (2013).
- 167 Ho, L. & Crabtree, G. R. Chromatin remodelling during development. *Nature* **463**, 474-484, doi:10.1038/nature08911 (2010).
- 168 Ho, L. *et al.* An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proceedings of the National Academy of Sciences* **106**, 5181-5186 (2009).
- 169 Kadoch, C. *et al.* Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nature genetics* **45**, 592 (2013).
- 170 Mashtalir, N. *et al.* Modular Organization and Assembly of SWI/SNF Family Chromatin Remodeling Complexes. *Cell* **175**, 1272-1288.e1220, doi:<https://doi.org/10.1016/j.cell.2018.09.032> (2018).
- 171 Kadoch, C. Diverse compositions and functions of chromatin remodeling machines in cancer. *Science Translational Medicine* **11**, eaay1018, doi:10.1126/scitranslmed.aay1018 (2019).
- 172 Gualdi, R. *et al.* Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control. *Genes & development* **10**, 1670-1682 (1996).
- 173 Ramakrishnan, V., Finch, J., Graziano, V., Lee, P. & Sweet, R. Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature* **362**, 219 (1993).
- 174 Clark, K. L., Halay, E. D., Lai, E. & Burley, S. K. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* **364**, 412 (1993).
- 175 Iwafuchi-Doi, M. *et al.* The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. *Molecular cell* **62**, 79-91, doi:10.1016/j.molcel.2016.03.001 (2016).
- 176 Cirillo, L. A. *et al.* Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. *The EMBO journal* **17**, 244-254 (1998).
- 177 Cirillo, L. A. & Zaret, K. S. An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. *Molecular cell* **4**, 961-969 (1999).
- 178 Cirillo, L. A. *et al.* Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Molecular cell* **9**, 279-289 (2002).
- 179 Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676 (2006).

- 180 Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **151**, 994-1004 (2012).
- 181 Feng, R. *et al.* PU. 1 and C/EBP α / β convert fibroblasts into macrophage-like cells. *Proceedings of the National Academy of Sciences* **105**, 6057-6062 (2008).
- 182 Ieda, M. *et al.* Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* **142**, 375-386 (2010).
- 183 Vierbuchen, T. *et al.* Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**, 1035 (2010).
- 184 Huang, P. *et al.* Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. *Nature* **475**, 386 (2011).
- 185 Barozzi, I. *et al.* Coregulation of transcription factor binding and nucleosome occupancy through DNA features of mammalian enhancers. *Molecular cell* **54**, 844-857, doi:10.1016/j.molcel.2014.04.006 (2014).
- 186 Zaret, K. S. Pioneering the chromatin landscape. *Nature genetics* **50**, 167-169, doi:10.1038/s41588-017-0038-z (2018).
- 187 Zhao, Y. & Garcia, B. A. Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harb Perspect Biol* **7**, a025064, doi:10.1101/cshperspect.a025064 (2015).
- 188 Allfrey, V. G., Faulkner, R. & Mirsky, A. E. ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS. *Proceedings of the National Academy of Sciences of the United States of America* **51**, 786-794, doi:10.1073/pnas.51.5.786 (1964).
- 189 Yang, X. J. & Seto, E. HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention. *Oncogene* **26**, 5310-5318, doi:10.1038/sj.onc.1210599 (2007).
- 190 Hodawadekar, S. C. & Marmorstein, R. Chemistry of acetyl transfer by histone modifying enzymes: structure, mechanism and implications for effector design. *Oncogene* **26**, 5528-5540, doi:10.1038/sj.onc.1210619 (2007).
- 191 Parthun, M. R. Hat1: the emerging cellular roles of a type B histone acetyltransferase. *Oncogene* **26**, 5319-5328, doi:10.1038/sj.onc.1210602 (2007).
- 192 Grant, P. A. *et al.* Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. *Genes & development* **11**, 1640-1650, doi:10.1101/gad.11.13.1640 (1997).
- 193 Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693-705 (2007).
- 194 Kiefer, C. M., Hou, C., Little, J. A. & Dean, A. Epigenetics of beta-globin gene regulation. *Mutat Res* **647**, 68-76, doi:10.1016/j.mrfmmm.2008.07.014 (2008).
- 195 Voigt, P., Tee, W. W. & Reinberg, D. A double take on bivalent promoters. *Genes & development* **27**, 1318-1338, doi:10.1101/gad.219626.113 (2013).

- 196 Mujtaba, S., Zeng, L. & Zhou, M. M. Structure and acetyl-lysine recognition of the bromodomain. *Oncogene* **26**, 5521-5527, doi:10.1038/sj.onc.1210618 (2007).
- 197 Hassan, A. H. *et al.* Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. *Cell* **111**, 369-379, doi:10.1016/s0092-8674(02)01005-x (2002).
- 198 Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593 (2000).
- 199 Bedford, M. T. & Clarke, S. G. Protein arginine methylation in mammals: who, what, and why. *Molecular cell* **33**, 1-13, doi:10.1016/j.molcel.2008.12.013 (2009).
- 200 Tamaru, H. *et al.* Trimethylated lysine 9 of histone H3 is a mark for DNA methylation in *Neurospora crassa*. *Nature genetics* **34**, 75-79, doi:10.1038/ng1143 (2003).
- 201 Xiao, B. *et al.* Structure and catalytic mechanism of the human histone methyltransferase SET7/9. *Nature* **421**, 652 (2003).
- 202 Shi, Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941-953 (2004).
- 203 Klose, R. J. & Zhang, Y. Regulation of histone methylation by demethylimination and demethylation. *Nat Rev Mol Cell Biol* **8**, 307-318, doi:10.1038/nrm2143 (2007).
- 204 Whetstine, J. R. *et al.* Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. *Cell* **125**, 467-481 (2006).
- 205 Champagne, K. S. & Kutateladze, T. G. Structural insight into histone recognition by the ING PHD fingers. *Curr Drug Targets* **10**, 432-441, doi:10.2174/138945009788185040 (2009).
- 206 Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120 (2001).
- 207 Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116-120, doi:10.1038/35065132 (2001).
- 208 Nishioka, K. *et al.* Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. *Genes & development* **16**, 479-489, doi:10.1101/gad.967202 (2002).
- 209 Zegerman, P., Canas, B., Pappin, D. & Kouzarides, T. Histone H3 lysine 4 methylation disrupts binding of nucleosome remodeling and deacetylase (NuRD) repressor complex. *The Journal of biological chemistry* **277**, 11621-11624, doi:10.1074/jbc.C200045200 (2002).
- 210 Hon, G. C., Hawkins, R. D. & Ren, B. Predictive chromatin signatures in the mammalian genome. *Human molecular genetics* **18**, R195-201, doi:10.1093/hmg/ddp409 (2009).

- 211 Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823-837, doi:10.1016/j.cell.2007.05.009 (2007).
- 212 Bannister, A. J. *et al.* Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *The Journal of biological chemistry* **280**, 17732-17736, doi:10.1074/jbc.M500796200 (2005).
- 213 Trojer, P. & Reinberg, D. Facultative heterochromatin: is there a distinctive molecular signature? *Molecular cell* **28**, 1-13 (2007).
- 214 Blow, M. J. *et al.* The Epigenomic Landscape of Prokaryotes. *PLoS genetics* **12**, e1005854, doi:10.1371/journal.pgen.1005854 (2016).
- 215 Arber, W. DNA modification and restriction. *Progress in nucleic acid research and molecular biology* **14**, 1-37, doi:10.1016/s0079-6603(08)60204-4 (1974).
- 216 Bickle, T. A. & Kruger, D. H. Biology of DNA restriction. *Microbiological reviews* **57**, 434-450 (1993).
- 217 Boyer, H. GENETIC CONTROL OF RESTRICTION AND MODIFICATION IN ESCHERICHIA COLI. *Journal of bacteriology* **88**, 1652-1660 (1964).
- 218 Casadesús, J. & Low, D. Epigenetic gene regulation in the bacterial world. *Microbiol Mol Biol Rev* **70**, 830-856, doi:10.1128/MMBR.00016-06 (2006).
- 219 Beletskaya, I. V., Zakharova, M. V., Shlyapnikov, M. G., Semenova, L. M. & Solonin, A. S. DNA methylation at the CfrBI site is involved in expression control in the CfrBI restriction-modification system. *Nucleic acids research* **28**, 3817-3822, doi:10.1093/nar/28.19.3817 (2000).
- 220 Zakharova, M., Minakhin, L., Solonin, A. & Severinov, K. Regulation of RNA polymerase promoter selectivity by covalent modification of DNA. *J Mol Biol* **335**, 103-111, doi:10.1016/j.jmb.2003.09.081 (2004).
- 221 Malone, T., Blumenthal, R. M. & Cheng, X. Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *J Mol Biol* **253**, 618-632, doi:10.1006/jmbi.1995.0577 (1995).
- 222 Berdis, A. J. *et al.* A cell cycle-regulated adenine DNA methyltransferase from *Caulobacter crescentus* processively methylates GANTC sites on hemimethylated DNA. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 2874-2879, doi:10.1073/pnas.95.6.2874 (1998).
- 223 Barras, F. & Marinus, M. G. The great GATC: DNA methylation in *E. coli*. *Trends in genetics : TIG* **5**, 139-143, doi:10.1016/0168-9525(89)90054-1 (1989).
- 224 Sternberg, N., Sauer, B., Hoess, R. & Abremski, K. Bacteriophage P1 cre gene and its regulatory region. Evidence for multiple promoters and for regulation by DNA methylation. *J Mol Biol* **187**, 197-212, doi:10.1016/0022-2836(86)90228-7 (1986).
- 225 Plumbridge, J. & Soll, D. The effect of dam methylation on the expression of *glnS* in *E. coli*. *Biochimie* **69**, 539-541, doi:10.1016/0300-9084(87)90091-5 (1987).

- 226 Jang, H. S., Shin, W. J., Lee, J. E. & Do, J. T. CpG and Non-CpG Methylation in Epigenetic Gene Regulation and Brain Function. *Genes (Basel)* **8**, 148, doi:10.3390/genes8060148 (2017).
- 227 Iyer, L. M., Zhang, D. & Aravind, L. Adenine methylation in eukaryotes: Apprehending the complex evolutionary history and functional potential of an epigenetic modification. *BioEssays : news and reviews in molecular, cellular and developmental biology* **38**, 27-40, doi:10.1002/bies.201500104 (2016).
- 228 Greer, E. L. *et al.* DNA Methylation on N6-Adenine in *C. elegans*. *Cell* **161**, 868-878, doi:10.1016/j.cell.2015.04.005 (2015).
- 229 Fu, Y. *et al.* N6-methyldeoxyadenosine marks active transcription start sites in *Chlamydomonas*. *Cell* **161**, 879-892, doi:10.1016/j.cell.2015.04.010 (2015).
- 230 Hotchkiss, R. D. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *The Journal of biological chemistry* **175**, 315-332 (1948).
- 231 Holliday, R. & Pugh, J. E. DNA modification mechanisms and gene activity during development. *Science (New York, N.Y.)* **187**, 226-232 (1975).
- 232 Razin, A. & Riggs, A. D. DNA methylation and gene function. *Science (New York, N.Y.)* **210**, 604-610, doi:10.1126/science.6254144 (1980).
- 233 Kuo, K. C., McCune, R. A., Gehrke, C. W., Midgett, R. & Ehrlich, M. Quantitative reversed-phase high performance liquid chromatographic determination of major and modified deoxyribonucleosides in DNA. *Nucleic acids research* **8**, 4763-4776, doi:10.1093/nar/8.20.4763 (1980).
- 234 Annan, R. S., Kresbach, G. M., Giese, R. W. & Vouros, P. Trace detection of modified DNA bases via moving-belt liquid chromatography-mass spectrometry using electrophoric derivatization and negative chemical ionization. *J Chromatogr* **465**, 285-296, doi:10.1016/s0021-9673(01)92666-4 (1989).
- 235 Cedar, H., Solage, A., Glaser, G. & Razin, A. Direct detection of methylated cytosine in DNA by use of the restriction enzyme MspI. *Nucleic acids research* **6**, 2125-2132, doi:10.1093/nar/6.6.2125 (1979).
- 236 Weber, M. *et al.* Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nature genetics* **37**, 853-862, doi:10.1038/ng1598 (2005).
- 237 Zhang, X. *et al.* Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell* **126**, 1189-1201, doi:10.1016/j.cell.2006.08.003 (2006).
- 238 Down, T. A. *et al.* A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nature biotechnology* **26**, 779-785, doi:10.1038/nbt1414 (2008).
- 239 Frommer, M. *et al.* A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National*

- Academy of Sciences of the United States of America* **89**, 1827-1831, doi:10.1073/pnas.89.5.1827 (1992).
- 240 Cokus, S. J. *et al.* Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* **452**, 215-219, doi:10.1038/nature06745 (2008).
- 241 Lister, R. *et al.* Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* **133**, 523-536, doi:10.1016/j.cell.2008.03.029 (2008).
- 242 Harris, R. A. *et al.* Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nature biotechnology* **28**, 1097-1105, doi:10.1038/nbt.1682 (2010).
- 243 Beck, S. Taking the measure of the methylome. *Nature biotechnology* **28**, 1026-1028, doi:10.1038/nbt1010-1026 (2010).
- 244 Meissner, A. *et al.* Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic acids research* **33**, 5868-5877, doi:10.1093/nar/gki901 (2005).
- 245 Gu, H. *et al.* Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nature protocols* **6**, 468-481, doi:10.1038/nprot.2010.190 (2011).
- 246 Guo, H. *et al.* Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. *Genome research* **23**, 2126-2135, doi:10.1101/gr.161679.113 (2013).
- 247 Guo, H. *et al.* Profiling DNA methylome landscapes of mammalian cells with single-cell reduced-representation bisulfite sequencing. *Nature protocols* **10**, 645-659, doi:10.1038/nprot.2015.039 (2015).
- 248 Smallwood, S. A. *et al.* Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nature methods* **11**, 817, doi:10.1038/nmeth.3035
<https://www.nature.com/articles/nmeth.3035#supplementary-information> (2014).
- 249 Gravina, S., Ganapathi, S. & Vijg, J. Single-cell, locus-specific bisulfite sequencing (SLBS) for direct detection of epimutations in DNA methylation patterns. *Nucleic acids research* **43**, e93, doi:10.1093/nar/gkv366 (2015).
- 250 Booth, M. J. *et al.* Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution. *Science (New York, N.Y.)* **336**, 934-937, doi:10.1126/science.1220671 (2012).
- 251 Yu, M. *et al.* Base-Resolution Analysis of 5-Hydroxymethylcytosine in the Mammalian Genome. *Cell* **149**, 1368-1380, doi:<https://doi.org/10.1016/j.cell.2012.04.027> (2012).
- 252 Tanaka, K. & Okamoto, A. Degradation of DNA by bisulfite treatment. *Bioorganic & medicinal chemistry letters* **17**, 1912-1915 (2007).

- 253 Olova, N. *et al.* Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data. *Genome biology* **19**, 33 (2018).
- 254 Liu, Y. *et al.* Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. *Nature biotechnology* **37**, 424-429, doi:10.1038/s41587-019-0041-2 (2019).
- 255 Stadler, M. B. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**, 490-495, doi:10.1038/nature10716 (2011).
- 256 Tost, J. DNA methylation: an introduction to the biology and the disease-associated changes of a promising biomarker. *Mol Biotechnol* **44**, 71-81, doi:10.1007/s12033-009-9216-2 (2010).
- 257 Li, E. & Zhang, Y. DNA methylation in mammals. *Cold Spring Harb Perspect Biol* **6**, a019133, doi:10.1101/cshperspect.a019133 (2014).
- 258 Antequera, F. & Bird, A. Number of CpG islands and genes in human and mouse. *Proceedings of the National Academy of Sciences* **90**, 11995-11999 (1993).
- 259 Ioshikhes, I. P. & Zhang, M. Q. Large-scale human promoter mapping using CpG islands. *Nature genetics* **26**, 61-63, doi:10.1038/79189 (2000).
- 260 Mohn, F. *et al.* Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Molecular cell* **30**, 755-766, doi:10.1016/j.molcel.2008.05.007 (2008).
- 261 Weber, M. *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature genetics* **39**, 457-466, doi:10.1038/ng1990 (2007).
- 262 Esteller, M. CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. *Oncogene* **21**, 5427-5440, doi:10.1038/sj.onc.1205600 (2002).
- 263 Wu, J. C. & Santi, D. V. On the mechanism and inhibition of DNA cytosine methyltransferases. *Prog Clin Biol Res* **198**, 119-129 (1985).
- 264 Wu, J. C. & Santi, D. V. Kinetic and catalytic mechanism of HhaI methyltransferase. *The Journal of biological chemistry* **262**, 4778-4786 (1987).
- 265 Ponger, L. & Li, W. H. Evolutionary diversification of DNA methyltransferases in eukaryotic genomes. *Mol Biol Evol* **22**, 1119-1128, doi:10.1093/molbev/msi098 (2005).
- 266 Jurkowski, T. P. *et al.* Human DNMT2 methylates tRNA(Asp) molecules using a DNA methyltransferase-like catalytic mechanism. *Rna* **14**, 1663-1670, doi:10.1261/rna.970408 (2008).
- 267 Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B. & Bestor, T. H. Dnmt3L and the establishment of maternal genomic imprints. *Science (New York, N.Y.)* **294**, 2536-2539, doi:10.1126/science.1065848 (2001).

- 268 Posfai, J., Bhagwat, A. S., Posfai, G. & Roberts, R. J. Predictive motifs derived from cytosine methyltransferases. *Nucleic acids research* **17**, 2421-2435, doi:10.1093/nar/17.7.2421 (1989).
- 269 Bestor, T., Laudano, A., Mattaliano, R. & Ingram, V. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J Mol Biol* **203**, 971-983, doi:10.1016/0022-2836(88)90122-2 (1988).
- 270 Aapola, U. *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* **65**, 293-298, doi:10.1006/geno.2000.6168 (2000).
- 271 Dong, A. *et al.* Structure of human DNMT2, an enigmatic DNA methyltransferase homolog that displays denaturant-resistant binding to DNA. *Nucleic acids research* **29**, 439-448, doi:10.1093/nar/29.2.439 (2001).
- 272 Jurkowski, T. P. & Jeltsch, A. On the evolutionary origin of eukaryotic DNA methyltransferases and Dnmt2. *PLoS One* **6**, e28104, doi:10.1371/journal.pone.0028104 (2011).
- 273 Rountree, M. R., Bachman, K. E. & Baylin, S. B. DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. *Nature genetics* **25**, 269-277, doi:10.1038/77023 (2000).
- 274 Leonhardt, H., Page, A. W., Weier, H. U. & Bestor, T. H. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* **71**, 865-873, doi:10.1016/0092-8674(92)90561-p (1992).
- 275 Misaki, T. *et al.* The replication foci targeting sequence (RFTS) of DNMT1 functions as a potent histone H3 binding domain regulated by autoinhibition. *Biochemical and biophysical research communications* **470**, 741-747, doi:10.1016/j.bbrc.2016.01.029 (2016).
- 276 Song, J., Rechkoblit, O., Bestor, T. H. & Patel, D. J. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science (New York, N.Y.)* **331**, 1036-1040, doi:10.1126/science.1195380 (2011).
- 277 Du, J., Johnson, L. M., Jacobsen, S. E. & Patel, D. J. DNA methylation pathways and their crosstalk with histone methylation. *Nat Rev Mol Cell Biol* **16**, 519-532, doi:10.1038/nrm4043 (2015).
- 278 Maresca, A., Zaffagnini, M., Caporali, L., Carelli, V. & Zanna, C. DNA methyltransferase 1 mutations and mitochondrial pathology: is mtDNA methylated? *Frontiers in Genetics* **6**, doi:10.3389/fgene.2015.00090 (2015).
- 279 Du, Q., Wang, Z. & Schramm, V. L. Human DNMT1 transition state structure. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 2916-2921, doi:10.1073/pnas.1522491113 (2016).

- 280 Klimasauskas, S., Kumar, S., Roberts, R. J. & Cheng, X. HhaI methyltransferase flips its target base out of the DNA helix. *Cell* **76**, 357-369, doi:10.1016/0092-8674(94)90342-5 (1994).
- 281 Chen, T., Ueda, Y., Dodge, J. E., Wang, Z. & Li, E. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Molecular and cellular biology* **23**, 5594-5605, doi:10.1128/mcb.23.16.5594-5605.2003 (2003).
- 282 Okano, M., Xie, S. & Li, E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nature genetics* **19**, 219-220, doi:10.1038/890 (1998).
- 283 Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-257, doi:10.1016/s0092-8674(00)81656-6 (1999).
- 284 Ooi, S. K. *et al.* DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**, 714-717, doi:10.1038/nature05987 (2007).
- 285 Fatemi, M., Hermann, A., Pradhan, S. & Jeltsch, A. The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. *Journal of molecular biology* **309**, 1189-1199 (2001).
- 286 Vilkaitis, G., Suetake, I., Klimašauskas, S. & Tajima, S. Processive methylation of hemimethylated CpG sites by mouse Dnmt1 DNA methyltransferase. *Journal of Biological Chemistry* **280**, 64-72 (2005).
- 287 Goyal, R., Reinhardt, R. & Jeltsch, A. Accuracy of DNA methylation pattern preservation by the Dnmt1 methyltransferase. *Nucleic acids research* **34**, 1182-1188 (2006).
- 288 Kim, G. D., Ni, J., Kelesoglu, N., Roberts, R. J. & Pradhan, S. Co-operation and communication between the human maintenance and de novo DNA (cytosine-5) methyltransferases. *The EMBO journal* **21**, 4183-4195 (2002).
- 289 Lorincz, M. C., Schübeler, D., Hutchinson, S. R., Dickerson, D. R. & Groudine, M. DNA methylation density influences the stability of an epigenetic imprint and Dnmt3a/b-independent de novo methylation. *Molecular and cellular biology* **22**, 7572-7580 (2002).
- 290 Arand, J. *et al.* In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. *PLoS genetics* **8**, e1002750 (2012).
- 291 Jeltsch, A. & Jurkowska, R. Z. New concepts in DNA methylation. *Trends in Biochemical Sciences* **39**, 310-318, doi:<https://doi.org/10.1016/j.tibs.2014.05.002> (2014).
- 292 Robertson, K. D. *et al.* The human DNA methyltransferases (DNMTs) 1, 3a and 3b: coordinate mRNA expression in normal tissues and overexpression in tumors. *Nucleic acids research* **27**, 2291-2298, doi:10.1093/nar/27.11.2291 (1999).

- 293 Ostler, K. R. *et al.* Cancer cells express aberrant DNMT3B transcripts encoding truncated proteins. *Oncogene* **26**, 5553-5563, doi:10.1038/sj.onc.1210351 (2007).
- 294 Wang, J. *et al.* The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. *Nature genetics* **41**, 125-129, doi:10.1038/ng.268 (2009).
- 295 Esteve, P. O. *et al.* Regulation of DNMT1 stability through SET7-mediated lysine methylation in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 5076-5081, doi:10.1073/pnas.0810362106 (2009).
- 296 Esteve, P. O. *et al.* A methylation and phosphorylation switch between an adjacent lysine and serine determines human DNMT1 stability. *Nature structural & molecular biology* **18**, 42-48, doi:10.1038/nsmb.1939 (2011).
- 297 Hata, K., Okano, M., Lei, H. & Li, E. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development (Cambridge, England)* **129**, 1983-1993 (2002).
- 298 Jia, D., Jurkowska, R. Z., Zhang, X., Jeltsch, A. & Cheng, X. Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature* **449**, 248-251, doi:10.1038/nature06146 (2007).
- 299 Bostick, M. *et al.* UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science (New York, N.Y.)* **317**, 1760-1764, doi:10.1126/science.1147939 (2007).
- 300 Sharif, J. *et al.* The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* **450**, 908-912, doi:10.1038/nature06397 (2007).
- 301 Hashimoto, H. *et al.* The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature* **455**, 826-829, doi:10.1038/nature07280 (2008).
- 302 Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y. & Shirakawa, M. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature* **455**, 818-821, doi:10.1038/nature07249 (2008).
- 303 Chuang, L. S. *et al.* Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science (New York, N.Y.)* **277**, 1996-2000, doi:10.1126/science.277.5334.1996 (1997).
- 304 Schermelleh, L. *et al.* Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. *Nucleic acids research* **35**, 4301-4312, doi:10.1093/nar/gkm432 (2007).
- 305 Spada, F. *et al.* DNMT1 but not its interaction with the replication machinery is required for maintenance of DNA methylation in human cells. *The Journal of cell biology* **176**, 565-571, doi:10.1083/jcb.200610062 (2007).
- 306 Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science (New York, N.Y.)* **324**, 929-930, doi:10.1126/science.1169786 (2009).

- 307 Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science (New York, N.Y.)* **324**, 930-935, doi:10.1126/science.1170116 (2009).
- 308 Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129-1133, doi:10.1038/nature09303 (2010).
- 309 Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science (New York, N.Y.)* **333**, 1300-1303, doi:10.1126/science.1210597 (2011).
- 310 He, Y. F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science (New York, N.Y.)* **333**, 1303-1307, doi:10.1126/science.1210944 (2011).
- 311 Kohli, R. M. & Zhang, Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **502**, 472-479, doi:10.1038/nature12750 (2013).
- 312 Maiti, A. & Drohat, A. C. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *The Journal of biological chemistry* **286**, 35334-35338, doi:10.1074/jbc.C111.284620 (2011).
- 313 Weber, A. R. *et al.* Biochemical reconstitution of TET1-TDG-BER-dependent active DNA demethylation reveals a highly coordinated mechanism. *Nat Commun* **7**, 10806, doi:10.1038/ncomms10806 (2016).
- 314 Pastor, W. A., Aravind, L. & Rao, A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat Rev Mol Cell Biol* **14**, 341-356, doi:10.1038/nrm3589 (2013).
- 315 Hu, L. *et al.* Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. *Cell* **155**, 1545-1555, doi:10.1016/j.cell.2013.11.020 (2013).
- 316 Hashimoto, H. *et al.* Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic acids research* **40**, 4841-4849, doi:10.1093/nar/gks155 (2012).
- 317 Otani, J. *et al.* Cell cycle-dependent turnover of 5-hydroxymethyl cytosine in mouse embryonic stem cells. *PLoS One* **8**, e82961, doi:10.1371/journal.pone.0082961 (2013).
- 318 Hu, L. *et al.* Structural insight into substrate preference for TET-mediated oxidation. *Nature* **527**, 118-122, doi:10.1038/nature15713 (2015).
- 319 Crawford, D. J. *et al.* Tet2 Catalyzes Stepwise 5-Methylcytosine Oxidation by an Iterative and de novo Mechanism. *J Am Chem Soc* **138**, 730-733, doi:10.1021/jacs.5b10554 (2016).
- 320 Xu, W. *et al.* Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. *Cancer Cell* **19**, 17-30, doi:10.1016/j.ccr.2010.12.014 (2011).

- 321 Lian, C. G. *et al.* Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. *Cell* **150**, 1135-1146, doi:10.1016/j.cell.2012.07.033 (2012).
- 322 Zhao, B. *et al.* Redox-active quinones induces genome-wide DNA methylation changes by an iron-mediated and Tet-dependent mechanism. *Nucleic acids research* **42**, 1593-1605, doi:10.1093/nar/gkt1090 (2014).
- 323 Wu, X. & Zhang, Y. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nature Reviews Genetics* **18**, 517, doi:10.1038/nrg.2017.33 (2017).
- 324 Lv, X., Jiang, H., Liu, Y., Lei, X. & Jiao, J. MicroRNA-15b promotes neurogenesis and inhibits neural progenitor proliferation by directly repressing TET3 during early neocortical development. *EMBO Rep* **15**, 1305-1314, doi:10.15252/embr.201438923 (2014).
- 325 Nakagawa, T. *et al.* CRL4(VprBP) E3 ligase promotes monoubiquitylation and chromatin binding of TET dioxygenases. *Molecular cell* **57**, 247-260, doi:10.1016/j.molcel.2014.12.002 (2015).
- 326 Ko, M. *et al.* Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature* **497**, 122-126, doi:10.1038/nature12052 (2013).
- 327 Wu, H. *et al.* Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* **473**, 389-393, doi:10.1038/nature09934 (2011).
- 328 Zhang, W. *et al.* Isoform Switch of TET1 Regulates DNA Demethylation and Mouse Development. *Molecular cell* **64**, 1062-1073, doi:10.1016/j.molcel.2016.10.030 (2016).
- 329 Dubois-Chevalier, J. *et al.* A dynamic CTCF chromatin binding landscape promotes DNA hydroxymethylation and transcriptional induction of adipocyte differentiation. *Nucleic acids research* **42**, 10943-10959, doi:10.1093/nar/gku780 (2014).
- 330 Fujiki, K. *et al.* PPARgamma-induced PARylation promotes local DNA demethylation by production of 5-hydroxymethylcytosine. *Nat Commun* **4**, 2262, doi:10.1038/ncomms3262 (2013).
- 331 Perera, A. *et al.* TET3 is recruited by REST for context-specific hydroxymethylation and induction of gene expression. *Cell reports* **11**, 283-294, doi:10.1016/j.celrep.2015.03.020 (2015).
- 332 de la Rica, L. *et al.* PU.1 target genes undergo Tet2-coupled demethylation and DNMT3b-mediated methylation in monocyte-to-osteoclast differentiation. *Genome biology* **14**, R99, doi:10.1186/gb-2013-14-9-r99 (2013).
- 333 Costa, Y. *et al.* NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature* **495**, 370-374, doi:10.1038/nature11925 (2013).
- 334 Meehan, R. R., Lewis, J. D., McKay, S., Kleiner, E. L. & Bird, A. P. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* **58**, 499-507, doi:10.1016/0092-8674(89)90430-3 (1989).

- 335 Hendrich, B. & Bird, A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Molecular and cellular biology* **18**, 6538-6547, doi:10.1128/mcb.18.11.6538 (1998).
- 336 Ng, H. H. *et al.* MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nature genetics* **23**, 58-61, doi:10.1038/12659 (1999).
- 337 Nan, X. *et al.* Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**, 386-389, doi:10.1038/30764 (1998).
- 338 Myant, K. *et al.* LSH and G9a/GLP complex are required for developmentally programmed DNA methylation. *Genome research* **21**, 83-94, doi:10.1101/gr.108498.110 (2011).
- 339 Esteve, P. O. *et al.* Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. *Genes & development* **20**, 3089-3103, doi:10.1101/gad.1463706 (2006).
- 340 Fuks, F., Burgers, W. A., Godin, N., Kasai, M. & Kouzarides, T. Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. *The EMBO journal* **20**, 2536-2544, doi:10.1093/emboj/20.10.2536 (2001).
- 341 Deplus, R. *et al.* Dnmt3L is a transcriptional repressor that recruits histone deacetylase. *Nucleic acids research* **30**, 3831-3838, doi:10.1093/nar/gkf509 (2002).
- 342 Dennis, K., Fan, T., Geiman, T., Yan, Q. & Muegge, K. Lsh, a member of the SNF2 family, is required for genome-wide methylation. *Genes & development* **15**, 2940-2944, doi:10.1101/gad.929101 (2001).
- 343 Yin, Y. *et al.* Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science (New York, N.Y.)* **356**, doi:10.1126/science.aaj2239 (2017).
- 344 Vanzan, L., Soldati, H., Ythier, V., Anand, S., Francis, N., Murr, R. *SOX2 Acts as A "Super Pioneer Transcription Factor" by Inducing Replication-Dependent DNA Demethylation at its Binding Sites.*
- 345 Borgel, J. *et al.* Targets and dynamics of promoter DNA methylation during early mouse development. *Nature genetics* **42**, 1093-1100, doi:10.1038/ng.708 (2010).
- 346 Auclair, G. *et al.* EHMT2 directs DNA methylation for efficient gene silencing in mouse embryos. *Genome research* **26**, 192-202, doi:10.1101/gr.198291.115 (2016).
- 347 O'Neill, K. M. *et al.* Depletion of DNMT1 in differentiated human cells highlights key classes of sensitive genes and an interplay with polycomb repression. *Epigenetics & chromatin* **11**, 12, doi:10.1186/s13072-018-0182-4 (2018).
- 348 Karimi, M. M. *et al.* DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. *Cell stem cell* **8**, 676-687, doi:10.1016/j.stem.2011.04.004 (2011).

- 349 Holliday, R. & Grigg, G. W. DNA methylation and mutation. *Mutat Res* **285**, 61-67, doi:10.1016/0027-5107(93)90052-h (1993).
- 350 Rosic, S. *et al.* Evolutionary analysis indicates that DNA alkylation damage is a byproduct of cytosine DNA methyltransferase activity. *Nature genetics* **50**, 452-459, doi:10.1038/s41588-018-0061-8 (2018).
- 351 Bird, A. P. & Taggart, M. H. Variable patterns of total DNA and rDNA methylation in animals. *Nucleic acids research* **8**, 1485-1497, doi:10.1093/nar/8.7.1485 (1980).
- 352 Cooper, D. N. & Krawczak, M. Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes. *Human genetics* **83**, 181-188, doi:10.1007/bf00286715 (1989).
- 353 Yoder, J. A., Walsh, C. P. & Bestor, T. H. Cytosine methylation and the ecology of intragenomic parasites. *Trends in genetics : TIG* **13**, 335-340, doi:10.1016/s0168-9525(97)01181-5 (1997).
- 354 Barau, J. *et al.* The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science (New York, N.Y.)* **354**, 909-912, doi:10.1126/science.aah5143 (2016).
- 355 Walsh, C. P., Chaillet, J. R. & Bestor, T. H. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nature genetics* **20**, 116-117, doi:10.1038/2413 (1998).
- 356 Jain, D. *et al.* rahu is a mutant allele of Dnmt3c, encoding a DNA methyltransferase homolog required for meiosis and transposon repression in the mouse male germline. *PLoS genetics* **13**, e1006964, doi:10.1371/journal.pgen.1006964 (2017).
- 357 Bartolomei, M. S. & Ferguson-Smith, A. C. Mammalian genomic imprinting. *Cold Spring Harbor perspectives in biology* **3**, a002592 (2011).
- 358 Bartolomei, M. S. Genomic imprinting: employing and avoiding epigenetic processes. *Genes & development* **23**, 2124-2133, doi:10.1101/gad.1841409 (2009).
- 359 Sasaki, H., Ishihara, K. & Kato, R. Mechanisms of Igf2/H19 imprinting: DNA methylation, chromatin and long-distance gene regulation. *Journal of biochemistry* **127**, 711-715, doi:10.1093/oxfordjournals.jbchem.a022661 (2000).
- 360 Nordin, M., Bergman, D., Halje, M., Engstrom, W. & Ward, A. Epigenetic regulation of the Igf2/H19 gene cluster. *Cell proliferation* **47**, 189-199, doi:10.1111/cpr.12106 (2014).
- 361 Kaneda, M. *et al.* Genetic evidence for Dnmt3a-dependent imprinting during oocyte growth obtained by conditional knockout with Zp3-Cre and complete exclusion of Dnmt3b by chimera formation. *Genes Cells* **15**, 169-179, doi:10.1111/j.1365-2443.2009.01374.x (2010).
- 362 Chotalia, M. *et al.* Transcription is required for establishment of germline methylation marks at imprinted genes. *Genes & development* **23**, 105-117, doi:10.1101/gad.495809 (2009).

- 363 Veselovska, L. *et al.* Deep sequencing and de novo assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape. *Genome biology* **16**, 209, doi:10.1186/s13059-015-0769-z (2015).
- 364 Peaston, A. E. *et al.* Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev Cell* **7**, 597-606, doi:10.1016/j.devcel.2004.09.004 (2004).
- 365 Li, X. *et al.* A maternal-zygotic effect gene, *Zfp57*, maintains both maternal and paternal imprints. *Dev Cell* **15**, 547-557, doi:10.1016/j.devcel.2008.08.014 (2008).
- 366 Quenneville, S. *et al.* In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Molecular cell* **44**, 361-372, doi:10.1016/j.molcel.2011.08.032 (2011).
- 367 Ng, K., Pullirsch, D., Leeb, M. & Wutz, A. Xist and the order of silencing. *EMBO Rep* **8**, 34-39, doi:10.1038/sj.embor.7400871 (2007).
- 368 da Rocha, S. T. & Heard, E. Novel players in X inactivation: insights into Xist-mediated gene silencing and chromosome conformation. *Nature structural & molecular biology* **24**, 197-204, doi:10.1038/nsmb.3370 (2017).
- 369 Chu, C. *et al.* Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404-416, doi:10.1016/j.cell.2015.03.025 (2015).
- 370 McHugh, C. A. *et al.* The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232-236, doi:10.1038/nature14443 (2015).
- 371 Gendrel, A. V. *et al.* Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. *Dev Cell* **23**, 265-279, doi:10.1016/j.devcel.2012.06.011 (2012).
- 372 Gu, T. P. *et al.* The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**, 606-610, doi:10.1038/nature10443 (2011).
- 373 Wossidlo, M. *et al.* 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat Commun* **2**, 241, doi:10.1038/ncomms1240 (2011).
- 374 Howell, C. Y. *et al.* Genomic imprinting disrupted by a maternal effect mutation in the *Dnmt1* gene. *Cell* **104**, 829-838, doi:10.1016/s0092-8674(01)00280-x (2001).
- 375 Maenohara, S. *et al.* Role of UHRF1 in de novo DNA methylation in oocytes and maintenance methylation in preimplantation embryos. *PLoS genetics* **13**, e1007042, doi:10.1371/journal.pgen.1007042 (2017).
- 376 Kobayashi, H. *et al.* Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS genetics* **8**, e1002440, doi:10.1371/journal.pgen.1002440 (2012).
- 377 Okae, H. *et al.* Genome-wide analysis of DNA methylation dynamics during early human development. *PLoS genetics* **10**, e1004868, doi:10.1371/journal.pgen.1004868 (2014).

- 378 Yamaguchi, S. *et al.* Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming. *Cell Res* **23**, 329-339, doi:10.1038/cr.2013.22 (2013).
- 379 Yamaguchi, S. *et al.* Tet1 controls meiosis by regulating meiotic gene expression. *Nature* **492**, 443-447, doi:10.1038/nature11709 (2012).
- 380 Yamaguchi, S., Shen, L., Liu, Y., Sendler, D. & Zhang, Y. Role of Tet1 in erasure of genomic imprinting. *Nature* **504**, 460-464, doi:10.1038/nature12805 (2013).
- 381 Vincent, J. J. *et al.* Stage-specific roles for tet1 and tet2 in DNA demethylation in primordial germ cells. *Cell stem cell* **12**, 470-478, doi:10.1016/j.stem.2013.01.016 (2013).
- 382 Wang, L. *et al.* Programming and inheritance of parental DNA methylomes in mammals. *Cell* **157**, 979-991, doi:10.1016/j.cell.2014.04.017 (2014).
- 383 Zhu, P. *et al.* Single-cell DNA methylome sequencing of human preimplantation embryos. *Nature genetics* **50**, 12-19, doi:10.1038/s41588-017-0007-6 (2018).
- 384 Galonska, C., Ziller, M. J., Karnik, R. & Meissner, A. Ground State Conditions Induce Rapid Reorganization of Core Pluripotency Factor Binding before Global Epigenetic Reprogramming. *Cell stem cell* **17**, 462-470, doi:10.1016/j.stem.2015.07.005 (2015).
- 385 Messerschmidt, D. M. *et al.* Trim28 is required for epigenetic stability during mouse oocyte to embryo transition. *Science (New York, N.Y.)* **335**, 1499-1502, doi:10.1126/science.1216154 (2012).
- 386 Proudhon, C. *et al.* Protection against de novo methylation is instrumental in maintaining parent-of-origin methylation inherited from the gametes. *Molecular cell* **47**, 909-920, doi:10.1016/j.molcel.2012.07.010 (2012).
- 387 Smith, Z. D. *et al.* DNA methylation dynamics of the human preimplantation embryo. *Nature* **511**, 611-615, doi:10.1038/nature13581 (2014).
- 388 Guo, H. *et al.* The DNA methylation landscape of human early embryos. *Nature* **511**, 606-610, doi:10.1038/nature13544 (2014).
- 389 Hargan-Calvopina, J. *et al.* Stage-Specific Demethylation in Primordial Germ Cells Safeguards against Precocious Differentiation. *Developmental cell* **39**, 75-86, doi:10.1016/j.devcel.2016.07.019 (2016).
- 390 Smallwood, S. A. *et al.* Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nature genetics* **43**, 811-814, doi:10.1038/ng.864 (2011).
- 391 Li, Y. *et al.* Stella safeguards the oocyte methylome by preventing de novo methylation mediated by DNMT1. *Nature* **564**, 136-140, doi:10.1038/s41586-018-0751-5 (2018).
- 392 Smith, Z. D. *et al.* Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer. *Nature* **549**, 543-547, doi:10.1038/nature23891 (2017).
- 393 Zhang, Y. *et al.* Dynamic epigenomic landscapes during early lineage specification in mouse embryos. *Nature genetics* **50**, 96-105, doi:10.1038/s41588-017-0003-x (2018).

- 394 Hon, G. C. *et al.* Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nature genetics* **45**, 1198-1206, doi:10.1038/ng.2746 (2013).
- 395 Schultz, M. D. *et al.* Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* **523**, 212-216, doi:10.1038/nature14465 (2015).
- 396 Ziller, M. J. *et al.* Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500**, 477-481, doi:10.1038/nature12433 (2013).
- 397 You, J. S. *et al.* OCT4 establishes and maintains nucleosome-depleted regions that provide additional layers of epigenetic regulation of its target genes. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 14497-14502, doi:10.1073/pnas.1111309108 (2011).
- 398 Feldman, N. *et al.* G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nat Cell Biol* **8**, 188-194, doi:10.1038/ncb1353 (2006).
- 399 Athanasiadou, R. *et al.* Targeting of de novo DNA methylation throughout the Oct-4 gene regulatory region in differentiating embryonic stem cells. *PLoS One* **5**, e9937, doi:10.1371/journal.pone.0009937 (2010).
- 400 Thorel, F. *et al.* Conversion of adult pancreatic alpha-cells to beta-cells after extreme beta-cell loss. *Nature* **464**, 1149-1154, doi:10.1038/nature08894 (2010).
- 401 Bartelt, A. & Heeren, J. Adipose tissue browning and metabolic health. *Nature reviews. Endocrinology* **10**, 24-36, doi:10.1038/nrendo.2013.204 (2014).
- 402 Rosen, E. D. & Spiegelman, B. M. What we talk about when we talk about fat. *Cell* **156**, 20-44, doi:10.1016/j.cell.2013.12.012 (2014).
- 403 Stanford, K. I. *et al.* Brown adipose tissue regulates glucose homeostasis and insulin sensitivity. *The Journal of clinical investigation* **123**, 215-223, doi:10.1172/jci62308 (2013).
- 404 Wu, J. *et al.* Beige adipocytes are a distinct type of thermogenic fat cell in mouse and human. *Cell* **150**, 366-376 (2012).
- 405 Wu, J., Cohen, P. & Spiegelman, B. M. Adaptive thermogenesis in adipocytes: is beige the new brown? *Genes & development* **27**, 234-250, doi:10.1101/gad.211649.112 (2013).
- 406 Peirce, V., Carobbio, S. & Vidal-Puig, A. The different shades of fat. *Nature* **510**, 76-83, doi:10.1038/nature13477 (2014).
- 407 Rosenwald, M., Perdikari, A., Rüllicke, T. & Wolfrum, C. Bi-directional interconversion of brite and white adipocytes. *Nature cell biology* **15**, 659 (2013).
- 408 Wang, Q. A., Tao, C., Gupta, R. K. & Scherer, P. E. Tracking adipogenesis during white adipose tissue development, expansion and regeneration. *Nature medicine* **19**, 1338 (2013).

- 409 Lee, M.-W. *et al.* Activated type 2 innate lymphoid cells regulate beige fat biogenesis. *Cell* **160**, 74-87 (2015).
- 410 Seale, P. *et al.* PRDM16 controls a brown fat/skeletal muscle switch. *Nature* **454**, 961 (2008).
- 411 Seale, P. *et al.* Transcriptional control of brown fat determination by PRDM16. *Cell Metab* **6**, 38-54 (2007).
- 412 Rajakumari, S. *et al.* EBF2 determines and maintains brown adipocyte identity. *Cell Metab* **17**, 562-574 (2013).
- 413 Takahashi, K. & Yamanaka, S. A decade of transcription factor-mediated reprogramming to pluripotency. *Nature reviews Molecular cell biology* **17**, 183 (2016).
- 414 De Carvalho, D. D., You, J. S. & Jones, P. A. DNA methylation and cellular reprogramming. *Trends Cell Biol* **20**, 609-617, doi:10.1016/j.tcb.2010.08.003 (2010).
- 415 Shore, A., Karamitri, A., Kemp, P., Speakman, J. & Lomax, M. Role of Ucp1 enhancer methylation and chromatin remodelling in the control of Ucp1 expression in murine adipose tissue. *Diabetologia* **53**, 1164-1173 (2010).
- 416 Kiskinis, E. *et al.* RIP140 directs histone and DNA methylation to silence Ucp1 expression in white adipocytes. *The EMBO journal* **26**, 4831-4840 (2007).
- 417 Christian, M. *et al.* RIP140-targeted repression of gene expression in adipocytes. *Molecular and cellular biology* **25**, 9383-9391 (2005).
- 418 Derecka, M. *et al.* Tyk2 and Stat3 regulate brown adipose tissue differentiation and obesity. *Cell Metab* **16**, 814-824 (2012).
- 419 Barres, R. *et al.* Non-CpG methylation of the PGC-1alpha promoter through DNMT3B controls mitochondrial density. *Cell Metab* **10**, 189-198, doi:10.1016/j.cmet.2009.07.011 (2009).
- 420 Lim, Y. C. *et al.* Dynamic DNA methylation landscape defines brown and white cell specificity during adipogenesis. *Mol Metab* **5**, 1033-1041, doi:10.1016/j.molmet.2016.08.006 (2016).
- 421 Feinberg, A. P. & Tycko, B. The history of cancer epigenetics. *Nature Reviews Cancer* **4**, 143 (2004).
- 422 Eden, A., Gaudet, F., Waghmare, A. & Jaenisch, R. Chromosomal instability and tumors promoted by DNA hypomethylation. *Science (New York, N.Y.)* **300**, 455-455 (2003).
- 423 Karpf, A. R. & Matsui, S.-i. Genetic disruption of cytosine DNA methyltransferase enzymes induces chromosomal instability in human cancer cells. *Cancer research* **65**, 8635-8639 (2005).
- 424 Bestor, T. H. Transposons reanimated in mice. *Cell* **122**, 322-325 (2005).
- 425 Cui, H. *et al.* Loss of IGF2 imprinting: a potential marker of colorectal cancer risk. *Science (New York, N.Y.)* **299**, 1753-1755 (2003).

- 426 Fraga, M. F. *et al.* A mouse skin multistage carcinogenesis model reflects the aberrant DNA methylation patterns of human tumors. *Cancer research* **64**, 5527-5534 (2004).
- 427 Esteller, M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nature reviews genetics* **8**, 286 (2007).
- 428 Herman, J. G. & Baylin, S. B. Gene silencing in cancer in association with promoter hypermethylation. *N Engl J Med* **349**, 2042-2054, doi:10.1056/NEJMra023075 (2003).
- 429 Song, M. A. *et al.* Elucidating the landscape of aberrant DNA methylation in hepatocellular carcinoma. *PLoS One* **8**, e55761, doi:10.1371/journal.pone.0055761 (2013).
- 430 Jeronimo, C. *et al.* Quantitation of GSTP1 methylation in non-neoplastic prostatic tissue and organ-confined prostate adenocarcinoma. *J Natl Cancer Inst* **93**, 1747-1752, doi:10.1093/jnci/93.22.1747 (2001).
- 431 Esteller, M. *et al.* Detection of aberrant promoter hypermethylation of tumor suppressor genes in serum DNA from non-small cell lung cancer patients. *Cancer research* **59**, 67-70 (1999).
- 432 Cairns, P. *et al.* Molecular detection of prostate cancer in urine by GSTP1 hypermethylation. *Clin Cancer Res* **7**, 2727-2730 (2001).
- 433 Hoque, M. O. *et al.* Quantitative methylation-specific polymerase chain reaction gene patterns in urine sediment distinguish prostate cancer patients from control subjects. *J Clin Oncol* **23**, 6569-6575, doi:10.1200/jco.2005.07.009 (2005).
- 434 Herceg, Z. & Paliwal, A. Epigenetic mechanisms in hepatocellular carcinoma: How environmental factors influence the epigenome. *Mutation Research/Reviews in Mutation Research* **727**, 55-61, doi:<https://doi.org/10.1016/j.mrrev.2011.04.001> (2011).
- 435 Wahid, B., Ali, A., Rafique, S. & Idrees, M. New Insights into the Epigenetics of Hepatocellular Carcinoma. *BioMed research international* **2017**, 1609575-1609575, doi:10.1155/2017/1609575 (2017).
- 436 Tischoff, I. *et al.* Allele loss and epigenetic inactivation of 3p21.3 in malignant liver tumors. *Int J Cancer* **115**, 684-689, doi:10.1002/ijc.20944 (2005).
- 437 Tomasi, M. L., Li, T. W., Li, M., Mato, J. M. & Lu, S. C. Inhibition of human methionine adenosyltransferase 1A transcription by coding region methylation. *J Cell Physiol* **227**, 1583-1591, doi:10.1002/jcp.22875 (2012).
- 438 Zhang, C. *et al.* CpG island methylator phenotype association with upregulated telomerase activity in hepatocellular carcinoma. *Int J Cancer* **123**, 998-1004, doi:10.1002/ijc.23650 (2008).
- 439 Lambert, M.-P. *et al.* Aberrant DNA methylation distinguishes hepatocellular carcinoma associated with HBV and HCV infection and alcohol intake. *Journal of hepatology* **54**, 705-715 (2011).

- 440 Huang, J., Wang, Y., Guo, Y. & Sun, S. Down-regulated microRNA-152 induces aberrant DNA methylation in hepatitis B virus-related hepatocellular carcinoma by targeting DNA methyltransferase 1. *Hepatology* **52**, 60-70 (2010).
- 441 Park, S.-H., Jung, J. K., Lim, J. S., Tiwari, I. & Jang, K. L. Hepatitis B virus X protein overcomes all-trans retinoic acid-induced cellular senescence by downregulating levels of p16 and p21 via DNA methylation. *Journal of General Virology* **92**, 1309-1317 (2011).
- 442 Um, T.-H. *et al.* Aberrant CpG island hypermethylation in dysplastic nodules and early HCC of hepatitis B virus-related human multistep hepatocarcinogenesis. *Journal of hepatology* **54**, 939-947 (2011).
- 443 Herceg, Z. & Paliwal, A. HBV protein as a double-barrel shot-gun targets epigenetic landscape in liver cancer. *Journal of hepatology* **50**, 252-255 (2009).
- 444 Zhu, Y. Z. *et al.* Hepatitis B virus X protein promotes hypermethylation of p16(INK4A) promoter through upregulation of DNA methyltransferases in hepatocarcinogenesis. *Experimental and molecular pathology* **89**, 268-275, doi:10.1016/j.yexmp.2010.06.013 (2010).
- 445 Fu, X., Song, X., Li, Y., Tan, D. & Liu, G. Hepatitis B virus X protein upregulates DNA methyltransferase 3A/3B and enhances SOCS-1CpG island methylation. *Mol Med Rep* **13**, 301-308, doi:10.3892/mmr.2015.4545 (2016).
- 446 Herceg, Z. Epigenetics and cancer: towards an evaluation of the impact of environmental and dietary factors. *Mutagenesis* **22**, 91-103, doi:10.1093/mutage/gel068 (2007).
- 447 Sawan, C., Vaissiere, T., Murr, R. & Herceg, Z. Epigenetic drivers and genetic passengers on the road to cancer. *Mutat Res* **642**, 1-13, doi:10.1016/j.mrfmmm.2008.03.002 (2008).
- 448 Baylin, S. B. DNA methylation and gene silencing in cancer. *Nat Clin Pract Oncol* **2** **Suppl 1**, S4-11, doi:10.1038/nconpc0354 (2005).
- 449 Spencer, D. H. *et al.* CpG Island Hypermethylation Mediated by DNMT3A Is a Consequence of AML Progression. *Cell* **168**, 801-816.e813, doi:<https://doi.org/10.1016/j.cell.2017.01.021> (2017).
- 450 Issa, J.-P. J., Kantarjian, H. M. & Kirkpatrick, P. Azacitidine. *Nature Reviews Drug Discovery* **4**, 275-276, doi:10.1038/nrd1698 (2005).
- 451 Bolden, J. E., Peart, M. J. & Johnstone, R. W. Anticancer activities of histone deacetylase inhibitors. *Nature reviews. Drug discovery* **5**, 769-784, doi:10.1038/nrd2133 (2006).
- 452 Mack, G. S. Epigenetic cancer therapy makes headway. *J Natl Cancer Inst* **98**, 1443-1444, doi:10.1093/jnci/djj447 (2006).
- 453 Muller, C. I., Ruter, B., Koeffler, H. P. & Lubbert, M. DNA hypermethylation of myeloid cells, a novel therapeutic target in MDS and AML. *Curr Pharm Biotechnol* **7**, 315-321, doi:10.2174/138920106778521523 (2006).

- 454 Oki, Y., Aoki, E. & Issa, J. P. Decitabine--bedside to bench. *Crit Rev Oncol Hematol* **61**, 140-152, doi:10.1016/j.critrevonc.2006.07.010 (2007).
- 455 Myers, M. G., Jr., Leibel, R. L., Seeley, R. J. & Schwartz, M. W. Obesity and leptin resistance: distinguishing cause from effect. *Trends Endocrinol Metab* **21**, 643-651, doi:10.1016/j.tem.2010.08.002 (2010).
- 456 Marchi, M. *et al.* Human leptin tissue distribution, but not weight loss-dependent change in expression, is associated with methylation of its promoter. *Epigenetics* **6**, 1198-1206, doi:10.4161/epi.6.10.16600 (2011).
- 457 Zwamborn, R. A. *et al.* Prolonged high-fat diet induces gradual and fat depot-specific DNA methylation changes in adult mice. *Sci Rep* **7**, 43261 (2017).
- 458 Shen, W. *et al.* Epigenetic modification of the leptin promoter in diet-induced obese mice and the effects of N-3 polyunsaturated fatty acids. *Sci Rep* **4**, 5282, doi:10.1038/srep05282 (2014).
- 459 Wang, Z. V. & Scherer, P. E. Adiponectin, the past two decades. *J Mol Cell Biol* **8**, 93-100, doi:10.1093/jmcb/mjw011 (2016).
- 460 Kim, A. Y. *et al.* Obesity-induced DNA hypermethylation of the adiponectin gene mediates insulin resistance. *Nat Commun* **6**, 7585, doi:10.1038/ncomms8585 (2015).
- 461 Houde, A.-A. *et al.* Leptin and adiponectin DNA methylation levels in adipose tissues and blood cells are associated with BMI, waist girth and LDL-cholesterol levels in severely obese men and women. *BMC medical genetics* **16**, 29 (2015).
- 462 Maeda, N. *et al.* Diet-induced insulin resistance in mice lacking adiponectin/ACRP30. *Nature medicine* **8**, 731-737, doi:10.1038/nm724 (2002).
- 463 Yang, Q.-Y. *et al.* Maternal obesity induces epigenetic modifications to facilitate Zfp423 expression and enhance adipogenic differentiation in fetal mice. *Diabetes* **62**, 3727-3735 (2013).
- 464 Sun, W. *et al.* Cold-induced epigenetic programming of the sperm enhances brown adipose tissue activity in the offspring. *Nature medicine* **24**, 1372 (2018).
- 465 Zhang, X. Y. *et al.* Binding sites in mammalian genes and viral gene regulatory regions recognized by methylated DNA-binding protein. *Nucleic acids research* **18**, 6253-6260, doi:10.1093/nar/18.21.6253 (1990).
- 466 Wade, P. A. Methyl CpG binding proteins: coupling chromatin architecture to gene regulation. *Oncogene* **20**, 3166-3173, doi:10.1038/sj.onc.1204340 (2001).
- 467 Hendrich, B. & Tweedie, S. The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends in genetics : TIG* **19**, 269-277, doi:10.1016/s0168-9525(03)00080-5 (2003).
- 468 Liu, Y., Toh, H., Sasaki, H., Zhang, X. & Cheng, X. An atomic model of Zfp57 recognition of CpG methylation within a specific DNA sequence. *Genes & development* **26**, 2374-2379, doi:10.1101/gad.202200.112 (2012).

- 469 Prokhortchouk, A. *et al.* The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes & development* **15**, 1613-1618, doi:10.1101/gad.198501 (2001).
- 470 Lopes, E. C. *et al.* Kaiso contributes to DNA methylation-dependent silencing of tumor suppressor genes in colon cancer cell lines. *Cancer research* **68**, 7258-7263, doi:10.1158/0008-5472.can-08-0344 (2008).
- 471 Qin, S. *et al.* Kaiso mainly locates in the nucleus in vivo and binds to methylated, but not hydroxymethylated DNA. *Chinese journal of cancer research = Chung-kuo yen cheng yen chiu* **27**, 148-155, doi:10.3978/j.issn.1000-9604.2015.04.03 (2015).
- 472 Blattler, A. *et al.* ZBTB33 binds unmethylated regions of the genome associated with actively expressed genes. *Epigenetics & chromatin* **6**, 13, doi:10.1186/1756-8935-6-13 (2013).
- 473 Zhu, H., Wang, G. & Qian, J. Transcription factors as readers and effectors of DNA methylation. *Nature Reviews Genetics* **17**, 551, doi:10.1038/nrg.2016.83
<https://www.nature.com/articles/nrg.2016.83#supplementary-information> (2016).
- 474 Liu, Y. *et al.* Structural basis for Klf4 recognition of methylated DNA. *Nucleic acids research* **42**, 4859-4867, doi:10.1093/nar/gku134 (2014).
- 475 Hu, S. *et al.* DNA methylation presents distinct binding sites for human transcription factors. *eLife* **2**, e00726, doi:10.7554/eLife.00726 (2013).
- 476 Spruijt, C. G. *et al.* Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**, 1146-1159, doi:10.1016/j.cell.2013.02.004 (2013).
- 477 Sato, N., Kondo, M. & Arai, K. The orphan nuclear receptor GCNF recruits DNA methyltransferase for Oct-3/4 silencing. *Biochemical and biophysical research communications* **344**, 845-851, doi:10.1016/j.bbrc.2006.04.007 (2006).
- 478 Bednarik, D. P. *et al.* DNA CpG methylation inhibits binding of NF-kappa B proteins to the HIV-1 long terminal repeat cognate DNA motifs. *The New biologist* **3**, 969-976 (1991).
- 479 Comb, M. & Goodman, H. M. CpG methylation inhibits proenkephalin gene expression and binding of the transcription factor AP-2. *Nucleic acids research* **18**, 3975-3982, doi:10.1093/nar/18.13.3975 (1990).
- 480 Ehrlich, K. C., Cary, J. W. & Ehrlich, M. A broad bean cDNA clone encoding a DNA-binding protein resembling mammalian CREB in its sequence specificity and DNA methylation sensitivity. *Gene* **117**, 169-178, doi:10.1016/0378-1119(92)90726-6 (1992).
- 481 Falzon, M. & Kuff, E. L. Binding of the transcription factor EBP-80 mediates the methylation response of an intracisternal A-particle long terminal repeat promoter. *Molecular and cellular biology* **11**, 117-125, doi:10.1128/mcb.11.1.117 (1991).
- 482 Domcke, S. *et al.* Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature* **528**, 575-579, doi:10.1038/nature16462 (2015).

- 483 Bell, A. C. & Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* **405**, 482-485, doi:10.1038/35013100 (2000).
- 484 Schoenherr, C. J., Levorse, J. M. & Tilghman, S. M. CTCF maintains differential methylation at the *Igf2/H19* locus. *Nature genetics* **33**, 66-69, doi:10.1038/ng1057 (2003).
- 485 Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of bacteriology* **169**, 5429-5433, doi:10.1128/jb.169.12.5429-5433.1987 (1987).
- 486 Bergh, O., Borsheim, K. Y., Bratbak, G. & Heldal, M. High abundance of viruses found in aquatic environments. *Nature* **340**, 467-468, doi:10.1038/340467a0 (1989).
- 487 Suttle, C. A. Viruses in the sea. *Nature* **437**, 356-361, doi:10.1038/nature04160 (2005).
- 488 Rodriguez-Valera, F. *et al.* Explaining microbial population genomics through phage predation. *Nature reviews. Microbiology* **7**, 828-836, doi:10.1038/nrmicro2235 (2009).
- 489 Mojica, F. J., Diez-Villasenor, C., Soria, E. & Juez, G. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular microbiology* **36**, 244-246, doi:10.1046/j.1365-2958.2000.01838.x (2000).
- 490 Grissa, I., Vergnaud, G. & Pourcel, C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic acids research* **35**, W52-57, doi:10.1093/nar/gkm360 (2007).
- 491 Rousseau, C., Gonnet, M., Le Romancer, M. & Nicolas, J. CRISPI: a CRISPR interactive database. *Bioinformatics (Oxford, England)* **25**, 3317-3318, doi:10.1093/bioinformatics/btp586 (2009).
- 492 Kunin, V., Sorek, R. & Hugenholtz, P. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome biology* **8**, R61, doi:10.1186/gb-2007-8-4-r61 (2007).
- 493 Godde, J. S. & Bickerton, A. The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *Journal of molecular evolution* **62**, 718-729, doi:10.1007/s00239-005-0223-z (2006).
- 494 Lillestol, R. K. *et al.* CRISPR families of the crenarchaeal genus *Sulfolobus*: bidirectional transcription and dynamic properties. *Molecular microbiology* **72**, 259-272, doi:10.1111/j.1365-2958.2009.06641.x (2009).
- 495 Pougach, K. *et al.* Transcription, processing and function of CRISPR cassettes in *Escherichia coli*. *Molecular microbiology* **77**, 1367-1379, doi:10.1111/j.1365-2958.2010.07265.x (2010).
- 496 Pul, U. *et al.* Identification and characterization of *E. coli* CRISPR-cas promoters and their silencing by H-NS. *Molecular microbiology* **75**, 1495-1512, doi:10.1111/j.1365-2958.2010.07073.x (2010).

- 497 Jansen, R., Embden, J. D., Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular microbiology* **43**, 1565-1575, doi:10.1046/j.1365-2958.2002.02839.x (2002).
- 498 Haft, D. H., Selengut, J., Mongodin, E. F. & Nelson, K. E. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS computational biology* **1**, e60, doi:10.1371/journal.pcbi.0010060 (2005).
- 499 Makarova, K. S. *et al.* Evolution and classification of the CRISPR-Cas systems. *Nature reviews. Microbiology* **9**, 467-477, doi:10.1038/nrmicro2577 (2011).
- 500 Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology direct* **1**, 7, doi:10.1186/1745-6150-1-7 (2006).
- 501 Koonin, E. V., Makarova, K. S. & Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. *Current Opinion in Microbiology* **37**, 67-78, doi:<https://doi.org/10.1016/j.mib.2017.05.008> (2017).
- 502 Brouns, S. J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science (New York, N.Y.)* **321**, 960-964, doi:10.1126/science.1159689 (2008).
- 503 Jore, M. M. *et al.* Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nature structural & molecular biology* **18**, 529-536, doi:10.1038/nsmb.2019 (2011).
- 504 Gesner, E. M., Schellenberg, M. J., Garside, E. L., George, M. M. & Macmillan, A. M. Recognition and maturation of effector RNAs in a CRISPR interference pathway. *Nature structural & molecular biology* **18**, 688-692, doi:10.1038/nsmb.2042 (2011).
- 505 Sashital, D. G., Jinek, M. & Doudna, J. A. An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. *Nature structural & molecular biology* **18**, 680-687, doi:10.1038/nsmb.2043 (2011).
- 506 Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602-607, doi:10.1038/nature09886 (2011).
- 507 Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67-71, doi:10.1038/nature09523 (2010).
- 508 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)* **337**, 816-821, doi:10.1126/science.1225829 (2012).
- 509 Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science (New York, N.Y.)* **322**, 1843-1845, doi:10.1126/science.1165771 (2008).
- 510 Hale, C. R. *et al.* RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **139**, 945-956, doi:10.1016/j.cell.2009.07.040 (2009).

- 511 Sorek, R., Lawrence, C. M. & Wiedenheft, B. CRISPR-Mediated Adaptive Immune Systems in Bacteria and Archaea. *Annual Review of Biochemistry* **82**, 237-266, doi:10.1146/annurev-biochem-072911-172315 (2013).
- 512 Pourcel, C., Salvignol, G. & Vergnaud, G. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology (Reading, England)* **151**, 653-663, doi:10.1099/mic.0.27437-0 (2005).
- 513 Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology (Reading, England)* **151**, 2551-2561, doi:10.1099/mic.0.28048-0 (2005).
- 514 Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of molecular evolution* **60**, 174-182, doi:10.1007/s00239-004-0046-3 (2005).
- 515 Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology (Reading, England)* **155**, 733-740, doi:10.1099/mic.0.023960-0 (2009).
- 516 Yosef, I., Goren, M. G. & Qimron, U. Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic acids research* **40**, 5569-5576, doi:10.1093/nar/gks216 (2012).
- 517 Marraffini, L. A. & Sontheimer, E. J. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nature reviews. Genetics* **11**, 181-190, doi:10.1038/nrg2749 (2010).
- 518 Nam, K. H. *et al.* Cas5d protein processes pre-crRNA and assembles into a cascade-like interference complex in subtype I-C/Dvulg CRISPR-Cas system. *Structure (London, England : 1993)* **20**, 1574-1584, doi:10.1016/j.str.2012.06.016 (2012).
- 519 Wang, R., Preamplume, G., Terns, M. P., Terns, R. M. & Li, H. Interaction of the Cas6 ribonuclease with CRISPR RNAs: recognition and cleavage. *Structure (London, England : 1993)* **19**, 257-264, doi:10.1016/j.str.2010.11.014 (2011).
- 520 Hatoum-Aslan, A., Maniv, I. & Marraffini, L. A. Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 21218-21222, doi:10.1073/pnas.1112832108 (2011).
- 521 Westra, E. R. *et al.* CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Molecular cell* **46**, 595-605, doi:10.1016/j.molcel.2012.03.018 (2012).
- 522 Beloglazova, N. *et al.* Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference. *The EMBO journal* **30**, 4616-4627, doi:10.1038/emboj.2011.377 (2011).

- 523 Estrella, M. A., Kuo, F. T. & Bailey, S. RNA-activated DNA cleavage by the Type III-B CRISPR-Cas effector complex. *Genes & development* **30**, 460-470, doi:10.1101/gad.273722.115 (2016).
- 524 Semenova, E. *et al.* Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 10098-10103, doi:10.1073/pnas.1104144108 (2011).
- 525 Rudin, N., Sugarman, E. & Haber, J. E. Genetic and physical analysis of double-strand break repair and recombination in *Saccharomyces cerevisiae*. *Genetics* **122**, 519-534 (1989).
- 526 Rouet, P., Smih, F. & Jasin, M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Molecular and cellular biology* **14**, 8096-8106, doi:10.1128/mcb.14.12.8096 (1994).
- 527 San Filippo, J., Sung, P. & Klein, H. Mechanism of eukaryotic homologous recombination. *Annu Rev Biochem* **77**, 229-257, doi:10.1146/annurev.biochem.77.061306.125255 (2008).
- 528 Lieber, M. R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual review of biochemistry* **79**, 181-211, doi:10.1146/annurev.biochem.052308.093131 (2010).
- 529 Loenen, W. A., Dryden, D. T., Raleigh, E. A., Wilson, G. G. & Murray, N. E. Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic acids research* **42**, 3-19, doi:10.1093/nar/gkt990 (2014).
- 530 Arber, W. & Linn, S. DNA modification and restriction. *Annu Rev Biochem* **38**, 467-500, doi:10.1146/annurev.bi.38.070169.002343 (1969).
- 531 Kruger, D. H. & Bickle, T. A. Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. *Microbiological reviews* **47**, 345-360 (1983).
- 532 Jeggo, P. A. DNA breakage and repair. *Advances in genetics* **38**, 185-218, doi:10.1016/s0065-2660(08)60144-3 (1998).
- 533 Valton, J. *et al.* 5'-Cytosine-phosphoguanine (CpG) methylation impacts the activity of natural and engineered meganucleases. *The Journal of biological chemistry* **287**, 30139-30150, doi:10.1074/jbc.M112.379966 (2012).
- 534 Daboussi, F. *et al.* Chromosomal context and epigenetic mechanisms control the efficacy of genome editing by rare-cutting designer endonucleases. *Nucleic acids research* **40**, 6367-6379, doi:10.1093/nar/gks268 (2012).
- 535 Klug, A. & Rhodes, D. Zinc fingers: a novel protein fold for nucleic acid recognition. *Cold Spring Harbor symposia on quantitative biology* **52**, 473-482, doi:10.1101/sqb.1987.052.01.054 (1987).

- 536 Kim, Y. G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 1156-1160, doi:10.1073/pnas.93.3.1156 (1996).
- 537 Bibikova, M. *et al.* Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Molecular and cellular biology* **21**, 289-297, doi:10.1128/mcb.21.1.289-297.2001 (2001).
- 538 Porteus, M. H. & Baltimore, D. Chimeric nucleases stimulate gene targeting in human cells. *Science (New York, N.Y.)* **300**, 763, doi:10.1126/science.1078395 (2003).
- 539 Boch, J. *et al.* Breaking the code of DNA binding specificity of TAL-type III effectors. *Science (New York, N.Y.)* **326**, 1509-1512, doi:10.1126/science.1178811 (2009).
- 540 Moscou, M. J. & Bogdanove, A. J. A simple cipher governs DNA recognition by TAL effectors. *Science (New York, N.Y.)* **326**, 1501, doi:10.1126/science.1178817 (2009).
- 541 Christian, M. *et al.* Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **186**, 757-761, doi:10.1534/genetics.110.120717 (2010).
- 542 Zhang, F. *et al.* Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nature biotechnology* **29**, 149-153, doi:10.1038/nbt.1775 (2011).
- 543 Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. 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* **109**, E2579-2586, doi:10.1073/pnas.1208507109 (2012).
- 544 Jinek, M. *et al.* RNA-programmed genome editing in human cells. *eLife* **2**, e00471, doi:10.7554/eLife.00471 (2013).
- 545 Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science (New York, N.Y.)* **339**, 819-823, doi:10.1126/science.1231143 (2013).
- 546 Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science (New York, N.Y.)* **339**, 823-826, doi:10.1126/science.1232033 (2013).
- 547 Tian, X. *et al.* CRISPR/Cas9 – An evolving biological tool kit for cancer biology and oncology. *npj Precision Oncology* **3**, 8, doi:10.1038/s41698-019-0080-7 (2019).
- 548 Kim, E. *et al.* In vivo genome editing with a small Cas9 orthologue derived from *Campylobacter jejuni*. *Nature communications* **8**, 14500 (2017).
- 549 Hirano, H. *et al.* Structure and Engineering of *Francisella novicida* Cas9. *Cell* **164**, 950-961, doi:10.1016/j.cell.2016.01.039 (2016).
- 550 Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology* **31**, 827-832, doi:10.1038/nbt.2647 (2013).
- 551 Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature biotechnology* **31**, 833-838, doi:10.1038/nbt.2675 (2013).

- 552 Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature biotechnology* **31**, 839-843, doi:10.1038/nbt.2673 (2013).
- 553 Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature biotechnology* **31**, 822-826, doi:10.1038/nbt.2623 (2013).
- 554 Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* **3**, e04766, doi:10.7554/eLife.04766 (2014).
- 555 Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J. S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome research* **24**, 1012-1019, doi:10.1101/gr.171322.113 (2014).
- 556 Kleinstiver, B. P. *et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490-495, doi:10.1038/nature16526 (2016).
- 557 Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science (New York, N.Y.)* **351**, 84-88, doi:10.1126/science.aad5227 (2016).
- 558 Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nature protocols* **8**, 2281-2308, doi:10.1038/nprot.2013.143 (2013).
- 559 Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173-1183, doi:10.1016/j.cell.2013.02.022 (2013).
- 560 Guilinger, J. P., Thompson, D. B. & Liu, D. R. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nature biotechnology* **32**, 577-582, doi:10.1038/nbt.2909 (2014).
- 561 Tsai, S. Q. *et al.* Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nature biotechnology* **32**, 569-576, doi:10.1038/nbt.2908 (2014).
- 562 Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442-451, doi:10.1016/j.cell.2013.06.044 (2013).
- 563 Chen, B. *et al.* Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479-1491, doi:10.1016/j.cell.2013.12.001 (2013).
- 564 Maeder, M. L. *et al.* CRISPR RNA-guided activation of endogenous human genes. *Nature methods* **10**, 977-979, doi:10.1038/nmeth.2598 (2013).
- 565 Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature methods* **10**, 973-976, doi:10.1038/nmeth.2600 (2013).
- 566 Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nature methods* **12**, 326-328, doi:10.1038/nmeth.3312 (2015).
- 567 Zalatan, J. G. *et al.* Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160**, 339-350, doi:10.1016/j.cell.2014.11.052 (2015).

- 568 Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583-588, doi:10.1038/nature14136 (2015).
- 569 Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**, 635-646, doi:10.1016/j.cell.2014.09.039 (2014).
- 570 Chavez, A. *et al.* Comparison of Cas9 activators in multiple species. *Nature methods* **13**, 563-567, doi:10.1038/nmeth.3871 (2016).
- 571 Lo, A. & Qi, L. Genetic and epigenetic control of gene expression by CRISPR-Cas systems. *F1000Research* **6**, doi:10.12688/f1000research.11113.1 (2017).
- 572 Liu, X. S. *et al.* Editing DNA Methylation in the Mammalian Genome. *Cell* **167**, 233-247.e217, doi:10.1016/j.cell.2016.08.056 (2016).
- 573 Amabile, A. *et al.* Inheritable Silencing of Endogenous Genes by Hit-and-Run Targeted Epigenetic Editing. *Cell* **167**, 219-232.e214, doi:10.1016/j.cell.2016.09.006 (2016).
- 574 Morita, S. *et al.* Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. *Nature biotechnology* **34**, 1060-1065, doi:10.1038/nbt.3658 (2016).
- 575 Kearns, N. A. *et al.* Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nature methods* **12**, 401-403, doi:10.1038/nmeth.3325 (2015).
- 576 Hilton, I. B. *et al.* Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature biotechnology* **33**, 510-517, doi:10.1038/nbt.3199 (2015).
- 577 Nihongaki, Y., Yamamoto, S., Kawano, F., Suzuki, H. & Sato, M. CRISPR-Cas9-based photoactivatable transcription system. *Chemistry & biology* **22**, 169-174, doi:10.1016/j.chembiol.2014.12.011 (2015).
- 578 Polstein, L. R. & Gersbach, C. A. A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. *Nature chemical biology* **11**, 198-200, doi:10.1038/nchembio.1753 (2015).
- 579 Gao, Y. *et al.* Complex transcriptional modulation with orthogonal and inducible dCas9 regulators. *Nature methods* **13**, 1043-1049, doi:10.1038/nmeth.4042 (2016).
- 580 Braun, S. M. G. *et al.* Rapid and reversible epigenome editing by endogenous chromatin regulators. *Nat Commun* **8**, 560, doi:10.1038/s41467-017-00644-y (2017).
- 581 Liang, F. S., Ho, W. Q. & Crabtree, G. R. Engineering the ABA plant stress pathway for regulation of induced proximity. *Science signaling* **4**, rs2, doi:10.1126/scisignal.2001449 (2011).
- 582 Miyamoto, T. *et al.* Rapid and orthogonal logic gating with a gibberellin-induced dimerization system. *Nature chemical biology* **8**, 465-470, doi:10.1038/nchembio.922 (2012).

- 583 Inoue, T., Heo, W. D., Grimley, J. S., Wandless, T. J. & Meyer, T. An inducible translocation strategy to rapidly activate and inhibit small GTPase signaling pathways. *Nature methods* **2**, 415-418, doi:10.1038/nmeth763 (2005).
- 584 Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T. & Kanemaki, M. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nature methods* **6**, 917-922, doi:10.1038/nmeth.1401 (2009).
- 585 Wu, Y. *et al.* Highly efficient therapeutic gene editing of human hematopoietic stem cells. *Nature medicine* **25**, 776-783, doi:10.1038/s41591-019-0401-y (2019).
- 586 Nelson, C. E. *et al.* In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. *Science (New York, N.Y.)* **351**, 403-407, doi:10.1126/science.aad5143 (2016).
- 587 Tabebordbar, M. *et al.* In vivo gene editing in dystrophic mouse muscle and muscle stem cells. *Science (New York, N.Y.)* **351**, 407-411, doi:10.1126/science.aad5177 (2016).
- 588 Nelson, C. E. *et al.* Long-term evaluation of AAV-CRISPR genome editing for Duchenne muscular dystrophy. *Nature medicine* **25**, 427-432, doi:10.1038/s41591-019-0344-3 (2019).
- 589 García-Tuñón, I. *et al.* The CRISPR/Cas9 system efficiently reverts the tumorigenic ability of BCR/ABL in vitro and in a xenograft model of chronic myeloid leukemia. *Oncotarget* **8**, 26027-26040, doi:10.18632/oncotarget.15215 (2017).
- 590 Luo, Z. *et al.* Efficient disruption of bcr-abl gene by CRISPR RNA-guided FokI nucleases depresses the oncogenesis of chronic myeloid leukemia cells. *J Exp Clin Cancer Res* **38**, 224-224, doi:10.1186/s13046-019-1229-5 (2019).
- 591 Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nature biotechnology* **36**, 765, doi:10.1038/nbt.4192
<https://www.nature.com/articles/nbt.4192#supplementary-information> (2018).
- 592 Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, doi:10.1038/s41586-019-1711-4 (2019).
- 593 de Silva, E. & Stumpf, M. P. HIV and the CCR5-Delta32 resistance allele. *FEMS microbiology letters* **241**, 1-12, doi:10.1016/j.femsle.2004.09.040 (2004).
- 594 Li, J.-R., Walker, S., Nie, J.-B. & Zhang, X.-Q. Experiments that led to the first gene-edited babies: the ethical failings and the urgent need for better governance. *J Zhejiang Univ Sci B* **20**, 32-38, doi:10.1631/jzus.B1800624 (2019).
- 595 Kyrou, K. *et al.* A CRISPR–Cas9 gene drive targeting doublesex causes complete population suppression in caged *Anopheles gambiae* mosquitoes. *Nature biotechnology* **36**, 1062, doi:10.1038/nbt.4245
<https://www.nature.com/articles/nbt.4245#supplementary-information> (2018).

- 596 Gantz, V. M. *et al.* Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito *Anopheles stephensi*. *Proceedings of the National Academy of Sciences* **112**, E6736-E6743, doi:10.1073/pnas.1521077112 (2015).
- 597 Dong, Y., Simões, M. L., Marois, E. & Dimopoulos, G. CRISPR/Cas9 -mediated gene knockout of *Anopheles gambiae* FREP1 suppresses malaria parasite infection. *PLOS Pathogens* **14**, e1006898, doi:10.1371/journal.ppat.1006898 (2018).
- 598 DiCarlo, J. E., Chavez, A., Dietz, S. L., Esvelt, K. M. & Church, G. M. Safeguarding CRISPR-Cas9 gene drives in yeast. *Nature biotechnology* **33**, 1250, doi:10.1038/nbt.3412
<https://www.nature.com/articles/nbt.3412#supplementary-information> (2015).
- 599 Lienert, F. *et al.* Identification of genetic elements that autonomously determine DNA methylation states. *Nature genetics* **43**, 1091-1097, doi:10.1038/ng.946 (2011).
- 600 Sherwood, R. I. *et al.* Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nature biotechnology* **32**, 171-178, doi:10.1038/nbt.2798 (2014).
- 601 Feldmann, A. *et al.* Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. *PLoS genetics* **9**, e1003994, doi:10.1371/journal.pgen.1003994 (2013).
- 602 Serandour, A. A. *et al.* Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome research* **21**, 555-565, doi:10.1101/gr.111534.110 (2011).
- 603 Taube, J. H., Allton, K., Duncan, S. A., Shen, L. & Barton, M. C. Foxa1 functions as a pioneer transcription factor at transposable elements to activate Afp during differentiation of embryonic stem cells. *The Journal of biological chemistry* **285**, 16135-16144, doi:10.1074/jbc.M109.088096 (2010).
- 604 Jones, P. A. & Baylin, S. B. The epigenomics of cancer. *Cell* **128**, 683-692, doi:10.1016/j.cell.2007.01.029 (2007).
- 605 Howell, P. M., Liu, Z. & Khong, H. T. Demethylating Agents in the Treatment of Cancer. *Pharmaceuticals (Basel)* **3**, 2022-2044, doi:10.3390/ph3072022 (2010).
- 606 Maeta, Y., Shiota, G., Okano, J. & Murawaki, Y. Effect of promoter methylation of the p16 gene on phosphorylation of retinoblastoma gene product and growth of hepatocellular carcinoma cells. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* **26**, 300-305, doi:10.1159/000089288 (2005).
- 607 Schagdarsurengin, U. *et al.* Frequent epigenetic inactivation of the RASSF1A gene in hepatocellular carcinoma. *Oncogene* **22**, 1866-1871, doi:10.1038/sj.onc.1206338 (2003).
- 608 Serrano, M. The tumor suppressor protein p16INK4a. *Experimental cell research* **237**, 7-13, doi:10.1006/excr.1997.3824 (1997).

- 609 Romagosa, C. *et al.* p16Ink4a overexpression in cancer: a tumor suppressor gene associated with senescence and high-grade tumors. *Oncogene* **30**, 2087-2097, doi:10.1038/onc.2010.614 (2011).
- 610 Donninger, H., Vos, M. D. & Clark, G. J. The RASSF1A tumor suppressor. *Journal of Cell Science* **120**, 3163-3172, doi:10.1242/jcs.010389 (2007).
- 611 Hu, H. *et al.* P16 reactivation induces anoikis and exhibits antitumour potency by downregulating Akt/survivin signalling in hepatocellular carcinoma cells. *Gut* **60**, 710-721, doi:10.1136/gut.2010.220020 (2011).
- 612 Kurita, S. *et al.* DNMT1 and DNMT3b silencing sensitizes human hepatoma cells to TRAIL-mediated apoptosis via up-regulation of TRAIL-R2/DR5 and caspase-8. *Cancer science* **101**, 1431-1439, doi:10.1111/j.1349-7006.2010.01565.x (2010).
- 613 Stelzer, Y., Shivalila, C. S., Soldner, F., Markoulaki, S. & Jaenisch, R. Tracing dynamic changes of DNA methylation at single-cell resolution. *Cell* **163**, 218-229, doi:10.1016/j.cell.2015.08.046 (2015).
- 614 Liu, X. S. *et al.* Rescue of Fragile X Syndrome Neurons by DNA Methylation Editing of the FMR1 Gene. *Cell* **172**, 979-992.e976, doi:10.1016/j.cell.2018.01.012 (2018).
- 615 DeFilippis, R. A., Goodwin, E. C., Wu, L. & DiMaio, D. Endogenous human papillomavirus E6 and E7 proteins differentially regulate proliferation, senescence, and apoptosis in HeLa cervical carcinoma cells. *J Virol* **77**, 1551-1563, doi:10.1128/jvi.77.2.1551-1563.2003 (2003).
- 616 Goodwin, E. C. & DiMaio, D. Repression of human papillomavirus oncogenes in HeLa cervical carcinoma cells causes the orderly reactivation of dormant tumor suppressor pathways. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 12513-12518, doi:10.1073/pnas.97.23.12513 (2000).
- 617 Tay, D. L., Bhathal, P. S. & Fox, R. M. Quantitation of G0 and G1 phase cells in primary carcinomas. Antibody to M1 subunit of ribonucleotide reductase shows G1 phase restriction point block. *The Journal of clinical investigation* **87**, 519-527, doi:10.1172/jci115026 (1991).
- 618 Florido, R., Tchkonja, T. & Kirkland, J. L. in *Handbook of the Biology of Aging (Seventh Edition)* (eds Edward J. Masoro & Steven N. Austad) 119-139 (Academic Press, 2011).
- 619 Dimri, G. P. *et al.* A biomarker that identifies senescent human cells in culture and in aging skin in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 9363-9367, doi:10.1073/pnas.92.20.9363 (1995).
- 620 Hall, B. M. *et al.* Aging of mice is associated with p16(Ink4a)- and beta-galactosidase-positive macrophage accumulation that can be induced in young mice by senescent cells. *Aging* **8**, 1294-1315, doi:10.18632/aging.100991 (2016).
- 621 Hall, B. M. *et al.* p16(Ink4a) and senescence-associated β -galactosidase can be induced in macrophages as part of a reversible response to physiological stimuli. *Aging* **9**, 1867-1884, doi:10.18632/aging.101268 (2017).

- 622 Lee, B. Y. *et al.* Senescence-associated β -galactosidase is lysosomal β -galactosidase. *Aging Cell* **5**, 187-195, doi:10.1111/j.1474-9726.2006.00199.x (2006).
- 623 Buiting, K. *et al.* Inherited microdeletions in the Angelman and Prader-Willi syndromes define an imprinting centre on human chromosome 15. *Nature genetics* **9**, 395-400, doi:10.1038/ng0495-395 (1995).
- 624 Kantor, B., Kaufman, Y., Makedonski, K., Razin, A. & Shemer, R. Establishing the epigenetic status of the Prader-Willi/Angelman imprinting center in the gametes and embryo. *Human molecular genetics* **13**, 2767-2779, doi:10.1093/hmg/ddh290 (2004).
- 625 Hackett, J. A. *et al.* Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science (New York, N.Y.)* **339**, 448-452, doi:10.1126/science.1229277 (2013).
- 626 Frontini, A. & Cinti, S. Distribution and development of brown adipocytes in the murine and human adipose organ. *Cell Metab* **11**, 253-256, doi:10.1016/j.cmet.2010.03.004 (2010).
- 627 Suárez-Zamorano, N. *et al.* Microbiota depletion promotes browning of white adipose tissue and reduces obesity. *Nature medicine* **21**, 1497-1501, doi:10.1038/nm.3994 (2015).
- 628 Roh, H. C. *et al.* Simultaneous Transcriptional and Epigenomic Profiling from Specific Cell Types within Heterogeneous Tissues In Vivo. *Cell reports* **18**, 1048-1061, doi:10.1016/j.celrep.2016.12.087 (2017).
- 629 Ellis, J. M. *et al.* Adipose acyl-CoA synthetase-1 directs fatty acids toward beta-oxidation and is required for cold thermogenesis. *Cell Metab* **12**, 53-64, doi:10.1016/j.cmet.2010.05.012 (2010).
- 630 Borglum, A. D., Flint, T., Hansen, L. L. & Kruse, T. A. Refined localization of the pyruvate dehydrogenase E1 alpha gene (PDHA1) by linkage analysis. *Human genetics* **99**, 80-82, doi:10.1007/s004390050315 (1997).
- 631 Elshourbagy, N. A. *et al.* Cloning and expression of a human ATP-citrate lyase cDNA. *European journal of biochemistry* **204**, 491-499, doi:10.1111/j.1432-1033.1992.tb16659.x (1992).
- 632 Sun, T., Hayakawa, K., Bateman, K. S. & Fraser, M. E. Identification of the citrate-binding site of human ATP-citrate lyase using X-ray crystallography. *The Journal of biological chemistry* **285**, 27418-27428, doi:10.1074/jbc.M109.078667 (2010).
- 633 Lee, M. H., Vosburgh, E., Anderson, K. & McDonagh, J. Deficiency of plasma plasminogen activator inhibitor 1 results in hyperfibrinolytic bleeding. *Blood* **81**, 2357-2362 (1993).
- 634 Miyazaki, M., Dobrzyn, A., Elias, P. M. & Ntambi, J. M. Stearoyl-CoA desaturase-2 gene expression is required for lipid synthesis during early skin and liver development. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 12501-12506, doi:10.1073/pnas.0503132102 (2005).

- 635 Adams, J. C. Thrombospondin-1. *The International Journal of Biochemistry & Cell Biology* **29**, 861-865, doi:[https://doi.org/10.1016/S1357-2725\(96\)00171-9](https://doi.org/10.1016/S1357-2725(96)00171-9) (1997).
- 636 Green, H. & Kehinde, O. An established preadipose cell line and its differentiation in culture. II. Factors affecting the adipose conversion. *Cell* **5**, 19-27, doi:10.1016/0092-8674(75)90087-2 (1975).
- 637 Klein, J., Fasshauer, M., Klein, H. H., Benito, M. & Kahn, C. R. Novel adipocyte lines from brown fat: a model system for the study of differentiation, energy metabolism, and insulin action. *Bioessays* **24**, 382-388 (2002).
- 638 Trajkovski, M., Ahmed, K., Esau, C. C. & Stoffel, M. MyomiR-133 regulates brown fat differentiation through Prdm16. *Nature Cell Biology* **14**, 1330-1335, doi:10.1038/ncb2612 (2012).
- 639 Schubeler, D., Lorincz, M. C. & Groudine, M. Targeting silence: the use of site-specific recombination to introduce in vitro methylated DNA into the genome. *Science's STKE : signal transduction knowledge environment* **2001**, p11, doi:10.1126/stke.2001.83.p11 (2001).
- 640 Wang, Z. *et al.* swDMR: A Sliding Window Approach to Identify Differentially Methylated Regions Based on Whole Genome Bisulfite Sequencing. *PLOS ONE* **10**, e0132866, doi:10.1371/journal.pone.0132866 (2015).
- 641 Gaspar, J. M. & Hart, R. P. DMRfinder: efficiently identifying differentially methylated regions from MethylC-seq data. *BMC Bioinformatics* **18**, 528-528, doi:10.1186/s12859-017-1909-0 (2017).
- 642 Wu, H. *et al.* Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicates. *Nucleic acids research* **43**, e141, doi:10.1093/nar/gkv715 (2015).
- 643 Boumelhem, B. B., Assinder, S. J., Bell-Anderson, K. S. & Fraser, S. T. Flow cytometric single cell analysis reveals heterogeneity between adipose depots. *Adipocyte* **6**, 112-123, doi:10.1080/21623945.2017.1319536 (2017).
- 644 Qiu, Y. *et al.* Screening of FDA-approved drugs identifies sutent as a modulator of UCP1 expression in brown adipose tissue. *EBioMedicine* **37**, 344-355, doi:<https://doi.org/10.1016/j.ebiom.2018.10.019> (2018).
- 645 Tan, Chong Y. *et al.* Brown Adipose Tissue Thermogenic Capacity Is Regulated by Elovl6. *Cell reports* **13**, 2039-2047, doi:<https://doi.org/10.1016/j.celrep.2015.11.004> (2015).
- 646 Hagberg, C. E. *et al.* Flow Cytometry of Mouse and Human Adipocytes for the Analysis of Browning and Cellular Heterogeneity. *Cell reports* **24**, 2746-2756.e2745, doi:10.1016/j.celrep.2018.08.006 (2018).
- 647 Ferrari, A. *et al.* HDAC3 is a molecular brake of the metabolic switch supporting white adipose tissue browning. *Nature communications* **8**, 93-93, doi:10.1038/s41467-017-00182-7 (2017).

- 648 Peng, X., Zhang, Q., Liao, C., Han, W. & Xu, F. Epigenomic Control of Thermogenic Adipocyte Differentiation and Function. *Int J Mol Sci* **19**, 1793, doi:10.3390/ijms19061793 (2018).
- 649 Schmidl, C., Rendeiro, A. F., Sheffield, N. C. & Bock, C. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nature methods* **12**, 963-965, doi:10.1038/nmeth.3542 (2015).
- 650 Wang, Q. *et al.* CoBATCH for High-Throughput Single-Cell Epigenomic Profiling. *Molecular cell* **76**, 206-216.e207, doi:<https://doi.org/10.1016/j.molcel.2019.07.015> (2019).
- 651 Fox, C. S. *et al.* Abdominal visceral and subcutaneous adipose tissue compartments: association with metabolic risk factors in the Framingham Heart Study. *Circulation* **116**, 39-48, doi:10.1161/circulationaha.106.675355 (2007).
- 652 West-Eberhard, M. J. Nutrition, the visceral immune system, and the evolutionary origins of pathogenic obesity. *Proceedings of the National Academy of Sciences* **116**, 723-731, doi:10.1073/pnas.1809046116 (2019).
- 653 Mathis, D. Immunological goings-on in visceral adipose tissue. *Cell Metab* **17**, 851-859, doi:10.1016/j.cmet.2013.05.008 (2013).
- 654 Reaven, G. M. Role of insulin resistance in human disease. *Diabetes* **37**, 1595-1607 (1988).
- 655 Tchernof, A. & Despres, J. P. Pathophysiology of human visceral obesity: an update. *Physiological reviews* **93**, 359-404, doi:10.1152/physrev.00033.2011 (2013).
- 656 Bjorndal, B., Burri, L., Staalesen, V., Skorve, J. & Berge, R. K. Different adipose depots: their role in the development of metabolic syndrome and mitochondrial response to hypolipidemic agents. *Journal of obesity* **2011**, 490650, doi:10.1155/2011/490650 (2011).
- 657 Lee, M. J., Wu, Y. & Fried, S. K. Adipose tissue heterogeneity: implication of depot differences in adipose tissue for obesity complications. *Molecular aspects of medicine* **34**, 1-11, doi:10.1016/j.mam.2012.10.001 (2013).
- 658 Sell, H., Habich, C. & Eckel, J. Adaptive immunity in obesity and insulin resistance. *Nature reviews. Endocrinology* **8**, 709-716, doi:10.1038/nrendo.2012.114 (2012).
- 659 Tchkonja, T. *et al.* Abundance of two human preadipocyte subtypes with distinct capacities for replication, adipogenesis, and apoptosis varies among fat depots. *American journal of physiology. Endocrinology and metabolism* **288**, E267-277, doi:10.1152/ajpendo.00265.2004 (2005).
- 660 Fernandez-Real, J. M. & Pickup, J. C. Innate immunity, insulin resistance and type 2 diabetes. *Diabetologia* **55**, 273-278, doi:10.1007/s00125-011-2387-y (2012).
- 661 Witcher, M. & Emerson, B. M. Epigenetic Silencing of the p16INK4a Tumor Suppressor Is Associated with Loss of CTCF Binding and a Chromatin Boundary. *Molecular cell* **34**, 271-284, doi:<https://doi.org/10.1016/j.molcel.2009.04.001> (2009).

- 662 Fang, X. *et al.* Landscape of the SOX2 protein–protein interactome. *PROTEOMICS* **11**, 921-934, doi:10.1002/pmic.201000419 (2011).
- 663 Tang, J., Zhong, G., Wu, J., Chen, H. & Jia, Y. SOX2 recruits KLF4 to regulate nasopharyngeal carcinoma proliferation via PI3K/AKT signaling. *Oncogenesis* **7**, 61, doi:10.1038/s41389-018-0074-2 (2018).
- 664 Cao, W. *et al.* Homeobox a5 Promotes White Adipose Tissue Browning Through Inhibition of the Tenascin C/Toll-Like Receptor 4/Nuclear Factor Kappa B Inflammatory Signaling in Mice. *Frontiers in Immunology* **9**, doi:10.3389/fimmu.2018.00647 (2018).
- 665 Sardina, J. L. *et al.* Transcription Factors Drive Tet2-Mediated Enhancer Demethylation to Reprogram Cell Fate. *Cell stem cell* **23**, 727-741.e729, doi:10.1016/j.stem.2018.08.016 (2018).
- 666 Yang, Y. A. *et al.* FOXA1 potentiates lineage-specific enhancer activation through modulating TET1 expression and function. *Nucleic acids research* **44**, 8153-8164, doi:10.1093/nar/gkw498 (2016).
- 667 Zhang, Y. *et al.* Nucleation of DNA repair factors by FOXA1 links DNA demethylation to transcriptional pioneering. *Nature genetics* **48**, 1003-1013, doi:10.1038/ng.3635 (2016).
- 668 Rizzino, A. Concise review: The Sox2-Oct4 connection: critical players in a much larger interdependent network integrated at multiple levels. *Stem Cells* **31**, 1033-1039, doi:10.1002/stem.1352 (2013).
- 669 Tapia, N. *et al.* Dissecting the role of distinct OCT4-SOX2 heterodimer configurations in pluripotency. *Sci Rep* **5**, 13533-13533, doi:10.1038/srep13533 (2015).
- 670 Rees, H. A., Yeh, W.-H. & Liu, D. R. Development of hRad51–Cas9 nickase fusions that mediate HDR without double-stranded breaks. *Nature Communications* **10**, 2212, doi:10.1038/s41467-019-09983-4 (2019).
- 671 Xie, K., Minkenberg, B. & Yang, Y. Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proceedings of the National Academy of Sciences* **112**, 3570-3575, doi:10.1073/pnas.1420294112 (2015).
- 672 Dong, F., Xie, K., Chen, Y., Yang, Y. & Mao, Y. Polycistronic tRNA and CRISPR guide-RNA enables highly efficient multiplexed genome engineering in human cells. *Biochemical and biophysical research communications* **482**, 889-895, doi:10.1016/j.bbrc.2016.11.129 (2017).
- 673 Runge, A. *et al.* An Inducible Hepatocellular Carcinoma Model for Preclinical Evaluation of Antiangiogenic Therapy in Adult Mice. *Cancer research* **74**, 4157-4169, doi:10.1158/0008-5472.can-13-2311 (2014).
- 674 Li, R. *et al.* The anti-tumor effect and increased tregs infiltration mediated by rAAV-SLC vector. *Molecular biology reports* **40**, 5615-5623 (2013).

- 675 Qin, T.-J. *et al.* Secretory expression of Par-4 SAC-HA2TAT following adeno-associated virus-mediated gene transfer induces apoptosis in HepG2 cells. *Molecular medicine reports* **3**, 749-757 (2010).
- 676 Sands, M. S. AAV-mediated liver-directed gene therapy. *Methods in molecular biology (Clifton, N.J.)* **807**, 141-157, doi:10.1007/978-1-61779-370-7_6 (2011).
- 677 Harrington, L. B. *et al.* Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science (New York, N.Y.)* **362**, 839-842, doi:10.1126/science.aav4294 (2018).
- 678 Hanawa, H. *et al.* Comparison of various envelope proteins for their ability to pseudotype lentiviral vectors and transduce primitive hematopoietic cells from human blood. *Molecular therapy : the journal of the American Society of Gene Therapy* **5**, 242-251, doi:10.1006/mthe.2002.0549 (2002).
- 679 Rohde, C., Zhang, Y., Reinhardt, R. & Jeltsch, A. BISMA - Fast and accurate bisulfite sequencing data analysis of individual clones from unique and repetitive sequences. *BMC Bioinformatics* **11**, 230, doi:10.1186/1471-2105-11-230 (2010).
- 680 Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻($\Delta\Delta C_T$) Method. *Methods (San Diego, Calif.)* **25**, 402-408, doi:10.1006/meth.2001.1262 (2001).
- 681 Grivennikov, S. I. *et al.* Adenoma-linked barrier defects and microbial products drive IL-23/IL-17-mediated tumour growth. *Nature* **491**, 254 (2012).
- 682 Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics (Oxford, England)* **27**, 1571-1572, doi:10.1093/bioinformatics/btr167 (2011).
- 683 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
- 684 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
- 685 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)* **30**, 923-930, doi:10.1093/bioinformatics/btt656 (2014).
- 686 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).