Dissecting structural and biochemical features of DNA methyltransferase 1

Dissertation

zur Erlangung des Grades eines
Doktors der Naturwissenschaften

Fachbereich Biologie, Chemie, Pharmazie

Institut für Biologie

Freie Universität Berlin

vorgelegt von

Nina Bailly

Berlin 2022

Erstgutachter:	Prof. Dr. Alexander Meissner Max-Planck-Institute for Molecular Genetics Berlin
Zweitgutachter:	Prof. Dr. Daniel Schubert Freie Universität Berlin Berlin
Disputation:	20.12.2022

Nina Bailly, M.Sc. RWTH

Dissecting structural and biochemical features of DNA methyltransferase 1

Dissertation, 17.10.2022

Max-Planck-Institue for Molecular Genetics
Department of Genome Regulation
Meissner Lab
Ihnestraße 63-73
14195 Berlin

Abstract

DNA methylation is an epigenetic modification found in every branch of life. An essential enzyme for the maintenance of DNA methylation patterns in mammals is DNA methyltransferase 1 (DNMT1). Its recruitment is regulated through its large N-terminus, which contains six annotated domains. Although most of these have been assigned a function, we are still lacking a holistic understanding of the enzyme's spatio-temporal regulation. Interestingly, a large segment of the N-terminus is devoid of any known domain and appears to be disordered in its sequence. Over the past years, such disordered sequences have increasingly gained attention, due to their role in forming biomolecular condensates through liquid-liquid phase separation (LLPS). These liquid compartments offer specific environmental conditions distinct from the surrounding that can enhance protein recruitment and function.

In this work, we explore a potential role for the intrinsically disordered region (IDR) in the recruitment of DNMT1. Taking an evolutionary approach, we uncover that structural features of the region that are key for IDR function are highly conserved. Moreover, we find conserved biochemical signatures compatible with a role in LLPS. Using a reconstitution assay and an opto-genetic approach in cells, we for the first time show that the DNMT1 IDR is capable of undergoing LLPS *in vitro* and *in vivo*. In addition, we define a novel region of interest region of interest (ROI) of about 120 amino acids in the IDR that appears to have been inserted in the ancestor of eutherian mammals. Although the ROI has a distinct biochemical signature, we find no effect on the LLPS behavior of the IDR. Therefore, we discuss other potential roles of the ROI related to DNA methylation, for example, imprinting.

Finally, we lay the foundation for investigating a biological function of the IDR and establish a system for screening DNMT1 mutant phenotypes in mouse embryonic stem cells. Swift depletion of the endogenous protein is enabled by degron-mediated degradation, while our optimized construct design and efficient derivation strategy ensure the robust expression of the large transgenes. In combination with different methods for DNA methylation read-out, this system can now be used to study the role of the IDR and ROI in maintaining the steady-state level of DNA methylation against mechanisms of passive and active demethylation, but also for studying phenotypes affecting the efficiency of DNMT1 recruitment in the future.

Zusammenfassung

DNA-Methylierung ist eine epigenetische Modifikation, die in fast allen Lebewesen vorkommt. Ein wesentliches Enzym für die Aufrechterhaltung der DNA-Methylierung in Säugetieren ist die DNA-Methyltransferase 1 (DNMT1). Ihre Rekrutierung wird durch ihren langen N-Terminus reguliert, der sechs annotierte Domänen enthält. Obwohl den meisten dieser Domänen eine Funktion zugewiesen wurde, fehlt uns immer noch ein ganzheitliches Verständnis der räumlich-zeitlichen Regulierung des Enzyms. Interessanterweise enthält ein großer Abschnitt des N-Terminus keine bekannte Domäne und scheint in seiner Sequenz ungeordnet zu sein. In den letzten Jahren haben solche ungeordneten Sequenzen aufgrund ihrer Rolle bei der Bildung biomolekularer Kondensate durch Flüssig-Flüssig-Phasentrennung (LLPS) zunehmend an Aufmerksamkeit gewonnen. Diese flüssigen Kompartimente bieten spezifische Umgebungsbedingungen, die sich von denen der Umgebung unterscheiden und die Rekrutierung und Funktion von Proteinen verbessern können.

In dieser Arbeit untersuchen wir eine mögliche Rolle der intrinsisch ungeordneten Region (IDR) bei der Rekrutierung der DNMT1. Mit Hilfe eines evolutionären Ansatzes decken wir auf, dass strukturelle Merkmale der Region, die für die Funktion der IDRs entscheidend sind, in hohem Maße konserviert sind. Darüber hinaus finden wir konservierte biochemische Signaturen, die mit einer Rolle in der LLPS vereinbar wären. Mithilfe eines Rekonstitutionsassays und eines optogenetischen Ansatzes in Zellen zeigen wir zum ersten Mal, dass die IDR der DNMT1 sowohl *in vitro* als auch *in vivo* in der Lage ist, eine separate flüssige Phase zu bilden. Darüber hinaus definieren wir eine neue 'region of interest' (ROI) von etwa 120 Aminosäuren in der IDR, die anscheinend in den Vorfahren der eutherischen Säugetiere eingefügt wurde. Obwohl die ROI eine ausgeprägte biochemische Signatur aufweist, finden wir keine Auswirkungen auf das LLPS-Verhalten der IDR. Daher diskutieren wir andere mögliche Rollen der ROI im Zusammenhang mit DNA-Methylierung, z.B. bei der genomischen Prägung.

Schließlich legen wir die Grundlage für die Untersuchung einer biologischen Funktion der IDR und etablieren ein System zum Screening von DNMT1-Mutanten-Phänotypen in embryonalen Stammzellen der Maus. Der rasche Abbau des endogenen Proteins wird durch die Fusion mit einem Degron ermöglicht, während unser optimiertes Konstruktdesign und unsere effiziente Derivationsstrategie die robuste

Expression der großen Transgene sicherstellen. In Kombination mit verschiedenen Methoden zum Auslesen der DNA-Methylierung kann dieses System nun verwendet werden, um die Rolle der IDR und der ROI bei der Aufrechterhaltung der DNA-Methylierung im dynamischen Gleichgewicht mit Mechanismen der passiven und aktiven Demethylierung zu untersuchen, aber auch, um in Zukunft Phänotypen zu untersuchen, die die Effizienz der DNMT1-Rekrutierung beeinflussen.

Bibliographic Information

Section 1.1 of this thesis is partially based on the following publication:

Alexandra Mattei*, **Nina Bailly*** and Alexander Meissner. *DNA methylation: a historical perspective*. **Trends in Genetics**, Volume 38, July 01 2022, pages 676–707.
DOI: 10.1016/j.tig.2022.03.010
The article's materials are reused under the open access CC BY license.

*joint first authorship

The data shown in Figures 4.11 and Figure 4.12 in Section 4.2 has in part been published in the following article:

Chuck Haggerty, Helene Kretzmer, Christina Riemenschneider, Abhishek Sampath Kumar, Alexandra L. Mattei, **Nina Bailly**, Judith Gottfreund, Pay Giesselmann, Raha Weigert, Björn Brändl, Pascal Giehr, René Buschow, Christina Galonska, Ferdinand von Meyenn, Melissa B. Pappalardi, Michael T. McCabe, Lars Wittler, Claudia Giesecke-Thiel, Thorsten Mielke, David Meierhofer, Bernd Timmermann, Franz-Josef Müller, Jörn Walter and Alexander Meissner. *Dnmt1 has de novo activity targeted to transposable elements*. **Nature Structural & Molecular Biology**, Volume 28, June 17 2021, pages 594–603.
DOI: 10.1038/s41594-021-00603-8

Contributions

The experimental work in this study was performed by Nina Bailly and supported by trainees and technicians of the Meissner Lab as well as the core facilities of the Max Planck Institute for Molecular Genetics. The d1 and d2 synthetic gene blocks were designed by Dr. Jocelyn Charlton. All RRBS libraries were prepared by Dr. Raha Weigert. The WGBS libraries for V6.5 wt, KH2 TKO and KH2 TKO + V5-Dnmt1^{co} C37 were prepared by Dr. Alexandra Mattei. The WGBS library for J1 DKO_0 was prepared by Chuck Haggerty. The resulting RRBS and WGBS data were processed and analyzed by Sara Hetzel and Dr. Helene Kretzmer, respectively. Sabine Otto performed all Western blots shown in this study. Tim Cheng acquired the images of the optoDroplet assays involving P15PAF. The bioinformatic analyses presented in Section 2 expand on the work of Pascal Mühlhausen, who carried out the initial analysis for DNMT1 for a smaller subset of species.

Acknowledgements

First, I would like to thank Alexander Meissner for the opportunity to work in his laboratory, starting with my internship at Harvard University. I am grateful for your trust that I would probably know best myself what avenues to pursue next. Working with you has taught me a great deal about keeping attention to detail, while not losing sight of the bigger picture.

I would also like to thank the other two members of my thesis advisory committee, Daniel Schubert and Aydan Bulut-Karslioglu, for their time, ears, and advice. Moreover, I would like to thank the members of the core facilities of the Max Planck Institute for Molecular Genetics, especially Uta Marchfelder, Claudia Gieseke-Thiel, René Buschow and Beatrix Fauler, for their help, support and cookies.

I would like to thank all members of the Meissner Lab, present and past, who created an amazing work environment. Thanks to Maria, Sabine, Petra, Cordula, Jen and Birgit, for always helping and taking care of us! I would also like to thank Sara Hetzel for being such a great partner on all of our projects. A special thanks goes to my Berlin family, Abhi, Christina, Raha, Adri and Sandy, who endured this journey at my side from the beginning to the end. Sandy, I have always admired you for your integrity and strength. You are a great inspiration, role model, and friend. Raha, your hard work and endurance have also pushed me to the next level and I loved working with you in good and bad times. Our friendship was essential for completing this work.

Moreover, I would like to thank my family for their unconditional love and support. I would like to especially thank my parents, who have always been there for me, no matter whether they were pulling me out of the dirt or celebrating my successes with me. Mom and dad, you have always encouraged me to be self-confident and to reach for the stars. You are my biggest heroes.

Most importantly, I would like to thank my loving husband, Anthony Wang, for bearing with me, day after day. Clearly, I would have never made it without you. The past years have been a beautiful chapter in our adventure together and I look forward to the future at your side! You are my soulmate.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

5mC	5-methylcytosine
ADD	ATRX-DNMT3-DNMT3L
BAH	Bromo-Adjacent Homology
BIC	blue-light inhibitor of CRYs
CCE	C-terminal extension domain
Cdc6	Cell Division Cycle 6
Cdt1	Chromatin licensing and DNA replication factor 1
CGI	CpG island
CIB	CRY-interacting bHLHs
COP1	CONSTITUTIVE PHOTOMORPHOGENIC 1
CRY2	CRYPTOCHROME 2 of *Arabidopsis thaliana*
CXXC	Zinc finger binding
D	D-box
DMAP	DNA methyltransferase 1-associated Protein
DMAP1	DNA Methyltransferase 1-associated Protein
DNMT	DNA methyltransferase
DRM	domains re-arranged
GK	glycine lysine linker
H3K36me2	dimethylated histone 3 lysine 36
H3K4	histone 3 lysine 4
H3K9me3	histone 3 lysine 9 trimethylation
H4K20	histone 4 lysine 20
HP1α	Heterochromatin Protein 1
HY5	HYPOCOTYL 5
ICR	imprinting control region
ID	intervening domain
IDR	intrinsically disordered region
In	initiation motif
LCR	low complexity region
LLPS	liquid-liquid phase separation
MBD	methyl-DNA binding domain
MeCP2	Methyl CpG binding Protein 2
MoRF	molecular recognition feature
MTase	methyltransferase
N-WASP	Wiskott-Aldrich syndrome protein
NID	NCoR/SMRT interaction domain
NLS	nuclear localization signal
NSD1	Nuclear Receptor Binding SET Domain Protein 1
ORC	origin recognition complex

Orc1	Origin Recognition Complex subunit 1
P15PAF	PCNA-associated factor 15
PC	principal components
PCA	principal component analysis
PCNA	Proliferating Cell Nuclear Antigen
PHR	photolyase homology region
PIP	PCNA-interacting protein box
PRC1	Polycomb Repressive Complex 1
PTM	post-translational modification
PWWP	Pro-Trp-Trp-Pro
RFTS	replication foci targeting sequence
ROI	region of interest
RRBS	reduced representation bisulfite sequencing
SETD1	SET Domain Containing 1
SH3	SRC Homology 3 Domain
SLiM	small linear motif
SPA	suppressor of PHYTOCRHOME A
TE	transposable element
TET	translocation methylcytosine dioxygenase
TRD	transcriptional repressor domain
UDR	ubiquitin-dependent recruitment region
UHRF1	Ubiquitin Like With PHD And Ring Finger Domains 1
ZFP57	Zinc-Finger Protein 57

1

Introduction

1.1 DNA methylation across the phylogenetic tree of life

1.1.1 Bacterial DNA methylation

DNA methylation was first discovered in the form of 5-methylcytosine (5mC) in nucleic acid extracts of *Mycobacterium tuberculosis* at the beginning of the 20th century [1]. Despite this seemingly relevant discovery, it was only reported again 23 years later by Hotchkiss et al. (1948) who studied the composition of bovine DNA using paper chromatography [2]. The observation of a faint band near that of cytosine lead him to propose the existences of a modified form of cytosine, which he named "epi-cytosine" [2]. He hypothesized this epi-cytosine would be related to cytosine the same way thymine (5-methyluracil) is related to uracil, inferring it could possibly be 5mC.

The dawn of molecular biology set the stage for a more thorough investigation and appreciation of DNA methylation from prokaryotes to eukaryotes. Essential progress was first made by studying the methylation of nucleic acids in bacteria and archaea, where DNA methylation in the form of 5mC, 4-methylcytosine and N6-methyladenine was detected [3, 4]. As tractable and abundant model organisms, bacteria provided two major insights into the biology of DNA methylation: (i) the methyl groups are added to the DNA polymer, mainly to the unmethylated, nascent strand after replication [5–7], and (ii) methylation of nucleic acids is carried out by enzymes called methyltransferases [8]. This suggested that DNA methylation could be a regulated process, and thus provide a path for specific target modification.

Importantly, it was found that different bacteria have strain-specific methyltransferase activity. This raised the possibility of a role for DNA methylation in the defense against phages [9]: Arber proposed the Restriction and Modification system (R-M system) where methylation-sensitive "restriction enzymes" (R) defend the bacterial host against invading viruses by digesting their DNA. Bacterial DNA is protected from these restriction enzymes due to modifications to their DNA in the form of species-specific DNA methylation placed by the cognate methyltransferase (M) [10]. To date, over 5000 R-M systems have been identified in bacteria [11].

Generally, prokaryotic DNA methyltransferases (DNMTs) and restriction endonucleases target specific sequences within DNA. The structural basis underlying the sequence specificity of bacterial DNMTs was uncovered in the 1980s and is determined by a series of non-conserved amino acid patches within the archetypical catalytic domain [12–16]. This sequence specificity restricts the presence of DNA methylation in prokaryotic genomes to defined DNA sequence motifs defined by the cell's genotype. Interestingly, a study published in 2016 identified a wide range of DNMTs that do not match with any cognate restriction enzyme in more than 100 strains of bacteria and archaea [17]. This suggests that DNA methylation has functions besides controlling lateral gene flow in prokaryotes. However, whether DNA methylation also plays a role in regulating DNA replication and gene expression, as hypothesized by the authors, remains to be determined.

Taken together, the first biological roles for DNA methylation were derived from studies on the basics of bacterial immunity. Once it became clear in the 1960's that DNA methylation, despite its relatively low abundance, does have a biological function in bacteria, the possibility that this chemical modification could play a more general regulatory role across species gained credibility and DNA methylation was increasingly studied in eukaryotes.

1.1.2 Distribution of 5-methylcytosine in Eukaryota

Generally, most eukaryotes display a patchy distribution of DNA methylation, where specific functional elements within the genome attract DNA methylation in the form of 5mC (Figure 1.1 A). The main targets for DNA methylation in plants and fungi are repeat elements and actively transcribed genes, where cytosines may be methy-

lated in the contexts of CpG, CpHpH and CpHpG, where H may be A, T, C or G [18–21]. An exception are some yeast species, which entirely lack DNA methylation [22, 23].

Most invertebrate genomes are sparsely methylated. When methylation is found, it occurs in the CpG context and is concentrated to longer stretches of highly methylated DNA, containing the bodies of a subset of expressed genes [24, 25]. Overall, little data on DNA methylation in invertebrates is available, in part because the main model invertebrates *Drosophila melongaster* and *Caenorhabditis elegans* display no methylation at all [26, 27]. In contrast, 5mC within the CpG context appears to be the default state in vertebrates. Most vertebrate methylomes exhibit a bimodal distribution. On the one hand, the majority of CpGs are highly methylated; on the other hand, few CpGs—clustered in CpG-dense regions termed CpG islands (CGIs)—remain unmethylated [28, 29]. A recent preliminary study profiling DNA methylation in 535 vertebrate and 45 invertebrate species using reduced representation bisulfite sequencing (RRBS) confirmed these patterns. Focusing on the evolution of vertebrate DNA methylation, the study found that marsupials and birds have an overall lower methylation level compared to other vertebrates, while fish and amphibia exhibit the highest level of DNA methylation [30]. However, additional comparative high-throughput studies are required to confirm these results and further explore the conservation and evolution of DNA methylation patterns among eukaryotes.

As in prokaryotes, DNA methylation in eukaryotes is catalyzed by DNMTs. Ancestral duplications and neofunctionalization, including the re-arrangement of domains and the acquisition of new domains, has led to a plethora of DNMTs which can be classified into five families (Figure 1.1 B). Importantly, the presence and absence of specific DNMT family members does not correlate with the global DNA methylation patterns observed. For example, DNMT1 and DNMT3 orthologues are deeply conserved in vertebrates and invertebrates, yet the two lineages exhibit dramatically different patterns of DNA methylation. These appear to be rather dictated by lineage-specific cofactors and patterns of histone modifications which recruit the DNMTs to their respective targets [22].

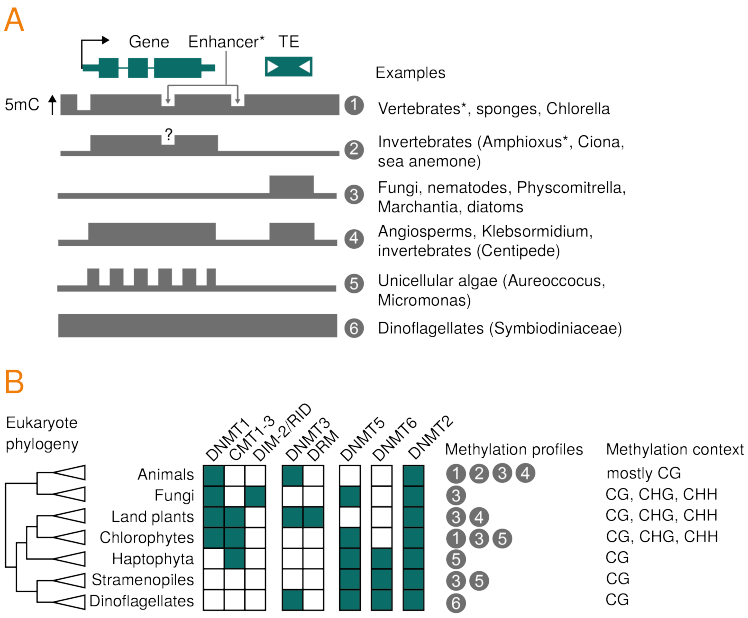

Fig. 1.1.: DNA methylation across Eukaryota. A DNA methylation profiles present in eukaryotic species. The height of the grey bar indicates the average DNA methylation at the features indicated above. Based on the profile, eukaryotic DNA methylation profiles can be categorized into six types. Example species for each profile are given on the right. Asterix indicate species or clades for which enhancer demethylation has been reported. B Presence of DNMTs across the eukaryotic phylogeny. Green boxes indicate the presence of a DNMT family member in at least one species within the clade (adapted from [22]). 5mC: 5-methylcytosine; DNMT: DNA methyltransferase; DRM: domains re-arranged; TE: transposable element.

1.1.3 Recruitment mechanisms of mammalian DNMTs

Most of our knowledge on DNA methylation has been gleaned from studies in mammals, specifically mouse and human. In these organisms, 5mC is mainly catalyzed by three DNMTs: DNMT1, DNMT3A and DNMT3B. Like all eukaryotic DNMTs, they harbor an MTase domain in their C-terminus (Figure 1.2). Using a conserved reaction mechanism, the domain catalyzes the transfer of a methyl group from S-adenosyl methionine to the 5' carbon of cytosine, yielding 5mC. Despite their similar catalytic activity, the three enzymes act on different targets which are in part determined by subtle differences in the structure of the MTase domain. DNMT3A and -B can add 5mC to unmethylated CpGs *de novo*, but can also act on hemimethylated DNA. DNMT1 has a strong preference for hemimethylated DNA and is mainly active during and after DNA replication, when it copies the methylation pattern on the parent strand to the daughter strand in a process referred to as maintenance

methylation. Thus, the DNMTs have been canonically divided into maintenance (DNMT1) and *de novo* methyltransferases (DNMT3A, DNMT3B) [31].

Recent work challenged this strict division of labor, by highlighting that DNMT3A and -B are also required to maintain steady-state 5mC levels in cells with high activities of ten-eleven translocation methylcytosine dioxygenases (TETs), which oxidize 5mC [32]. Conversely, our laboratory has established that DNMT1 is also capable of *de novo* methylation *in vivo*, where it targets certain repetitive elements [33].

Besides their preference for un- or hemimethylated DNA, the targets of the DNMTs are defined by the different regulatory domains in the N-terminus, which recruit the enzymes to specific features and are also involved in their allosteric regulation. The canonical *de novo* methyltransferases have similar N-termini (Figure 1.2 B, C). Both harbor an ATRX-DNMT3-DNMT3L (ADD) domain which occludes the DNA binding pocket of the MTase domain [34]. This autoinhibitory conformation can be released in the presence of unmodified histone 3 lysine 4 (H3K4) tails, which are preferentially bound by the ADD domain. This regulation has important implications for the methylome: CGIs marked by methylated H3K4 remain unmethylated, supporting the bimodal distribution of 5mC [29, 35].

Overall, both DNMT3A and -B preferentially locate to CpG-dense regions, but also have differential targets [36]. Upstream of the ADD domain, both have a PWWP domain, named after its core Pro-Trp-Trp-Pro motif. Due to slight differences in their PWWP domains, DNMT3A is more strongly recruited to dimethylated histone 3 lysine 36 (H3K36me2) in intergenic regions, while DNMT3B is rather recruited to bodies of actively transcribed genes marked by H3K36me3 [36, 37]. The recruitment to H3K36me2 and -me3 is particularly evident in the germ cells, where the histone methyltransferases Nuclear Receptor Binding SET Domain Protein 1 (NSD1) and SET Domain Containing 1 (SETD1) place H3K36me2 and -me3, respectively, in response to the transcriptional activity and thus set the DNA methylation landscape [38, 39].

These landscapes are erased in a wave of demethylation during preimplantation development, a process unique to mammals [22]. However, some remnants of germ cell methylation are maintained by DNMT1 [44]. Differences in sperm and oocyte methylation are propagated and in some cases lead to the differential expres-

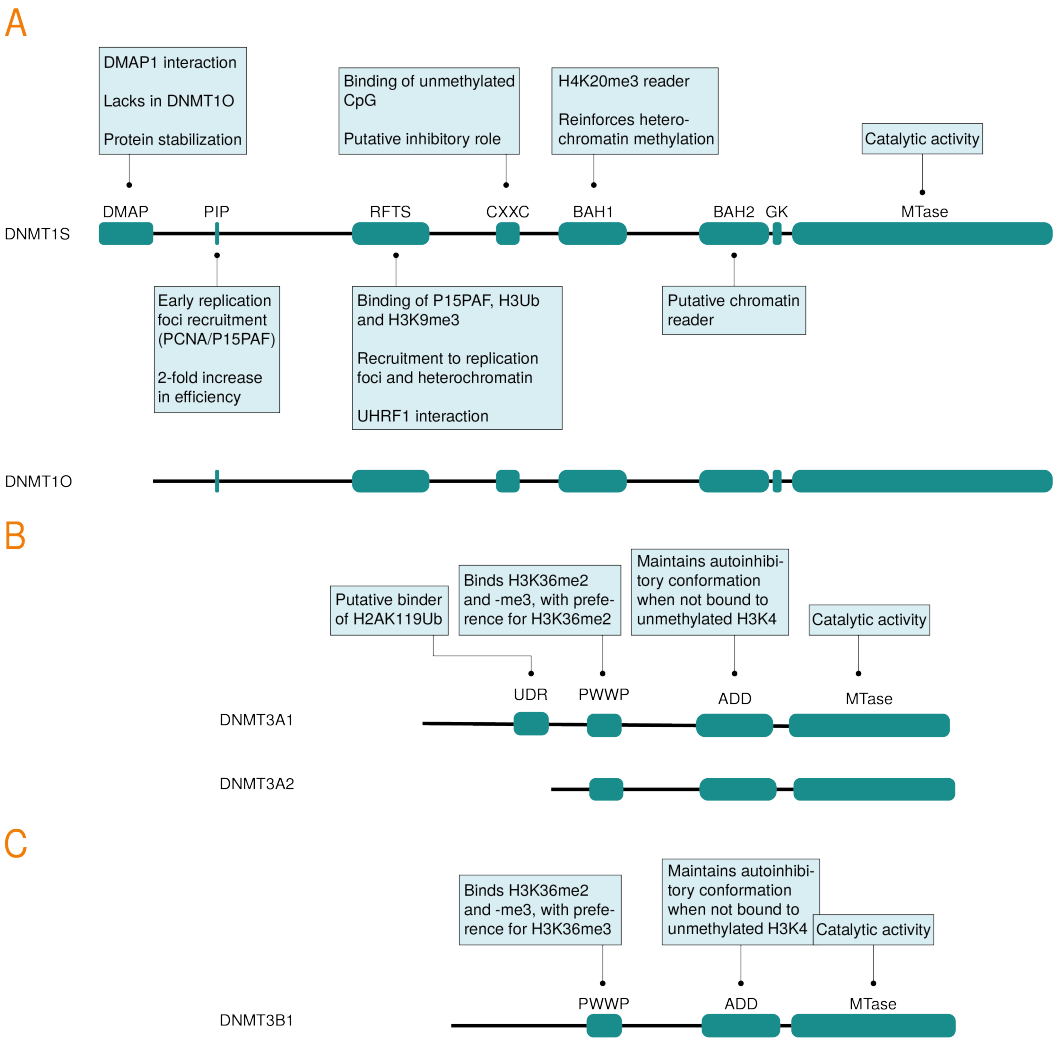

Fig. 1.2.: Domain structure of the DNMTs. All DNMTs have conserved catalytically active MTase domains in the C-terminus and a variety of regulatory domains in the N-terminus. A The somatic isoform DNMT1S of the maintenance methyltransferase is ubiquitously expressed. The shorter isoform DNMT1O is only expressed in oocytes due to alternative promoter usage [40]. B DNMT3A has two major splice isoforms. DNMT3A1 is widely expressed in somatic cells, while expression of DNMT3A2 is only high in the early embryo, in hematopeitic stem cells and the gonads [41, 42]. C DNMT3B has over 30 splice isoforms expressed in various developmental and disease contexts, but the main catalytically active isoform is DNMT3B1 [43]. All domain annotations are illustrated to scale. ADD: ATRX-DNMT3-DNMT3L domain; MTase: methyltransferase domain; BAH: bromo-adjacent homology domain; CXXC: Zinc finger binding domain; DMAP: DNA methyltransferase 1-associated Protein; GK: glycine lysine linker; PIP: PCNA-interacting protein box; PWWP: Pro-Trp-Trp-Pro domain; RFTS: replication foci targeting sequence; UDR: ubiquitin-dependent recruitment region.

sion of the genes associated with these so-called imprinting control regions (ICRs). Imprinted genes have crucial roles during prenatal growth and their disrupted ex-

pression underlies many diseases [45]. Which loci become ICRs is not entirely understood; however, most imprints in eutherian mammals contain DNA motifs derived from repeat sequences that are bound by Zinc-Finger Protein 57 (ZFP57), and ablation of ZFP57 leads to partial or complete loss of these imprints [46, 47]. ZFP57 binds Tripartite motif-containing 28 (TRIM28, also known TIF1β and KAP1) which is also required for the maintenance of methylation imprints [48]. How these factors eventually recruit DNMT1 and why DNMT1 only maintains DNA methylation at these loci during preimplantation development are still open questions.

Once the nadir of DNA methylation levels is reached around the time of implantation, the bimodal methylome characteristic of somatic cells in set up anew. In somatic cells, the effects of many DNMT recruitment mechanisms overlap, making their individual relevance and hierarchy difficult to discern. For example, only recently the recruitment of DNMT3A by histone 2A lysine 118/119 ubiquitination (H2AK118/119Ub) has been uncovered in cells with inactivating mutations in the PWWP domain. The recruitment by H2AK118/119Ub is mediated by the ubiquitin-dependent recruitment region (UDR) of DNMT3A [49]. The enzyme has two main catalytically active isoforms, DNMT3A1 and DNMT3A2, which differ by the presence of the first 219 amino acids in the N-terminus, including the UDR (Figure 1.2 B). Recent work dissecting the roles of DNMT3A1 and DNMT3A2 during mouse development has found that DNMT3A1 is necessary for post-natal survival as well as the DNA methylation and proper expression of developmental genes marked by H2AK118/119Ub [50]. Thus, despite being subordinate to the recruitment to H3K36me2, recruitment to H2AK118/119Ub by the UDR is essential for normal development.

With a size of approximately 190 kDa, DNMT1 is much larger than DNMT3A and -B and has a more complex domain architecture (Figure 1.2 A). To date, six different regulatory domains have been mapped to the DNMT1 N-terminus. Located at the proximal end, the DNA Methyltransferase 1-associated Protein (DMAP1) domain has been implicated in the binding of the transcriptional repressor DMAP1, but little is known about potential targets and mechanisms [51]. The DMAP1 binding domain is missing in the oocyte-specific isoform DNMT1O, which is more stable than the somatic isoform DNMT1S, a property that may be important for its prolonged storage in oocytes and maternal effect [52].

The sequence downstream of the DMAP1 binding domain contains a PIP box, which has been shown to be required for the binding of Proliferating Cell Nuclear Antigen (PCNA), an essential factor of eukaryotic DNA replication [53]. During early S-phase its interaction with PCNA recruits DNMT1 to replication foci, where its substrate—hemimethylated DNA—is produced [54, 55]. While this interaction has been found to be dispensable for the maintenance of steady-state DNA methylation levels, it increases the efficiency of maintenance methylation two-fold [56, 57].

In addition, DNMT1 can be recruited to replication foci via its Replication Foci Targeting Sequence (RFTS). It contains motifs that bind dual mono-ubiquitin modifications on PCNA-associated factor 15 (P15PAF) and H3 tails that depend on the ubiquitin ligase Ubiquitin Like With PHD And Ring Finger Domains 1 (UHRF1). These interactions recruit DNMT1 to replication foci during early and late replication, respectively [58–62]. Of note, UHRF1 can also bind histone 3 lysine 9 trimethylation (H3K9me3) and thus has been hypothesized to lead to the preferential recruitment of DNMT1 to H3K9me3-demarcated regions. However, UHRF1 mutants defective in H3K9 binding still localize to heterochromatin and are able to rescue DNA methylation in *Uhrf1* KO cells. Only when both of UHRF1's abilities to bind H3K9me3 and hemimethylated CpGs are impaired, do steady-state levels of 5mC drop globally [63]. In addition to ubiquitin modifications, the RFTS can also directly recognize H3K9me3 *in vitro* [64]. However, the relevance of this interaction for the recruitment of DNMT1 *in vivo* remains elusive. Besides its role in recruiting DNMT1 to P15PAF and modified H3 tails, biochemical experiments and partial crystal structures of the RFTS suggest it also has the ability to bind the MTase domain, implying an autoinhibitory mechanism [65].

Further downstream of the RFTS, DNMT1 possesses a Zn-binding, cystein-rich domain (CXXC), which has been shown *in vitro* to bind DNA and has been implicated in the allosteric regulation of the enzyme [66]. A mutant lacking the CXXC was shown to locate to replication foci, but did not exhibit any catalytic activity [67]. Yet, a mechanistic understanding of this process is missing and its role in shaping the DNA methylation landscape is unknown.

Finally, the two domains located in the distal N-terminus have low sequence similarity but nonetheless both adopt a Bromo-Adjacent Homology (BAH) fold [68]. Proteins containing BAH domains are common across all eukaryotic clades and are associated with functions in chromatin regulation, *e.g.*, DNA replication, nucleo-

some remodeling, and epigenetic modifications [69]. Ren et al. (2020) recently discovered that DNMT1 can bind histone 4 lysine 20 (H4K20) tails, with a preference for H4K20me3 [64]. Mutations in the BAH1 domain that abolished this interaction did not affect global methylation levels, but led to a slight local hypomethylation of H4K20me3-demarcated regions. About three quarters of these hypomethylated regions overlapped with Long Interspersed Nuclear Elements (LINEs), a dominant class of repetitive elements in mammalian genomes [64]. Thus, the BAH1 domain plays a role in reinforcing DNA methylation at these regions. In contrast, interaction partners of the BAH2 domain have not yet been identified.

Besides the multiple defined regulatory domains, the N-termini of the full-length DNMT isoforms in mouse and human contain a long stretch of disordered amino acid residues [50, 70–72]. In recent years, such intrinsically disordered domains (IDRs) have been increasingly studied in the context of liquid-liquid phase separation (LLPS) (see Section 1.2). However, the function of the DNMT IDRs remains unknown.

All DNMTs methylate DNA *in vitro*; however, establishment and maintenance of methylation of chromatinized DNA *in vivo* requires specific recruitment and regulatory mechanisms mediated by the DNMTs' N-terminal domains. With the exception of the DNMT1 BAH2 domain and the IDRs, each DNMT regulatory region has been assigned a function and has shed some light on the complex regulation of the DNMTs. Yet, a holistic understanding of their spatio-temporal control is obscured in part by the redundancy of these mechanisms. Their hierarchy and individual contributions remain to be untangled.

1.2 Regulation of cellular functions by liquid-liquid phase separation

1.2.1 Biomolecular condensates in cellular organization

Within cells, many complex reactions need to take place to sustain life, and they require precise spatio-temporal control and specific environmental conditions. Thus,

cells—specifically eukaryotic cells—are further compartmentalized by organelles. Canonical organelles are membrane-bound, *e.g.*, mitochondria, chloroplasts, and the nucleus. In addition, organelles without phospho-lipid membranes exist. Despite the lack of a membrane, this latter type of organelles can maintain its shape and size over hours and days [73]. Yet, the principals underlying their organization remained enigmatic until the study of P granules, perinuclear membrane-less organelles in the germ cells of *Caenorhabditis elegans* composed of proteins and RNA. In 2009, Brangwynne et al. were the first to discover their liquid-like behavior. They showed that P granules undergo fusion and fission, after which they relax back into a spherical shape. Moreover, they flow freely in the cytoplasm and deform along the surface of other structures. In accordance with their liquid nature, they freely exchange molecules across their boundary with the surrounding environment [74].

Unblending of liquids into distinct phases has emerged as a key principle for the intracellular organization of cells by membrane-less organelles [75–78]. Importantly, the same mechanisms also apply to the formation of other subcellular structures, *e.g.*, nuclear pore complexes and heterochromatin (see Section 1.2.3). These assemblies are collectively referred to as "biomolecular condensates" and their formation is the result of a process known as liquid-liquid phase separation (LLPS) [73, 79].

Biomolecular condesates are formed biopolymers (proteins, and sometimes RNA or DNA) that follow the same physiochemical laws as other polymers. When biomolecules undergo LLPS, they become concentrated in a dense phase which is surrounded by a dilute phase. The coexistence of dense droplets within the dilute phase represents an thermodynamically favorable state. Whether a solution of biomolecules undergoes LLPS depends on many factors that influence the thermodynamic state of the system, such as temperature, salt concentration, pH and the concentration of the biomolecules themselves.

Conditions favoring a one-phase or two-phase regime can be visualized using phase diagrams and are delineated by the binodal curve (Figure 1.3). At a given set of conditions (*e.g.*, constant temperature or pH, as defined on the y-axis), a two-phase system is established when the total concentration of the biomolecule (x-axis) reaches the solubility limit or saturation concentration c_{sat}, which lies on the binodal. Within the two-phase regime, the concentration of the biomolecule

within the light phase (c_L, which is equivalent to c_{sat}) and dense phase (c_D) can be read using horizontal tie lines within the two-phase regime. With increasing concentration of the biomolecule, c_L and c_D remain fixed, while their volumetric contributions change relative to each other, *i.e.*, more and/or larger droplets form. Once the total concentration surpasses c_D, the system again transitions into a one-phase system [80, 81]. Due to its strict dependence on concentration, the formation of droplets and the volumetric ratio between the light and dense phases can be controlled by the cells through the biogenesis, degradation and localization of the biomolecule [79].

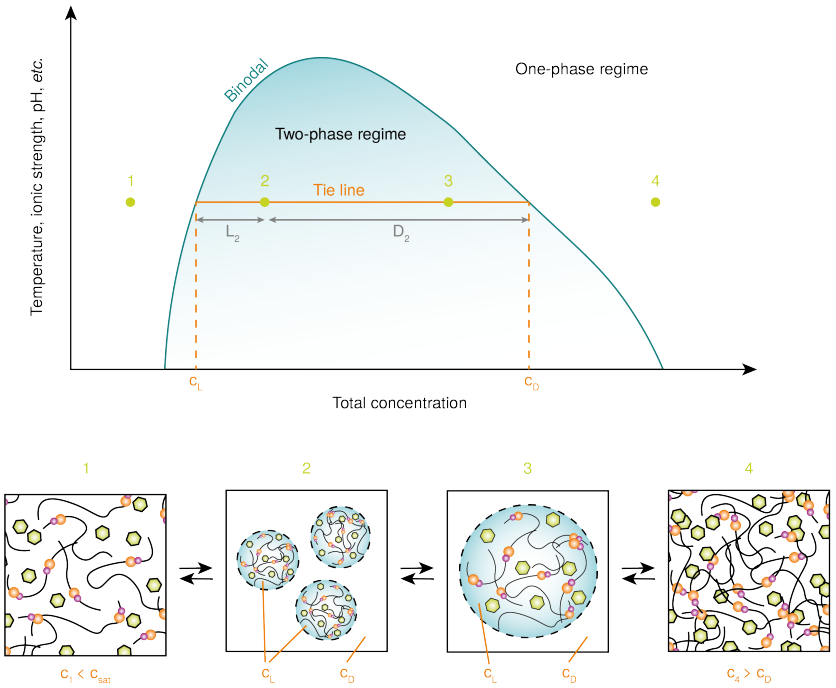

Fig. 1.3.: Schematic phase diagram of LLSP. Depending on the total concentration of the biomolecule along the x-axis, the system exists as one or two phases. At low concentrations (1), the biomolecule is mixed with the diluting liquid. Once the binodal is overstepped (2), the system splits into a dense phase containing the biomolecule at concentration c_D and a light phase where the biomolecule has the concentration c_L. With increasing concentration of the biomolecule (3), the concentrations in the dense and light phase remain constant, while the volume of the dense phase increases relative to the light phase. When the total concentration is larger than c_D, the phases mix again. States 1, 2, 3, and 4 are reversible. Adapted from [80, 81].

Since dense phase droplets are liquid, they assume spherical shapes that can undergo fusion and fission. On surfaces, they can exhibit wetting behavior. Molecules between the light and the dense phase are dynamically exchanged [80]. Usually, the formation of droplets is reversible. However, at conditions within the two-

regime space far away from the binodal, LLPS can also produce gel-like assemblies and even solids, a transformation that is often irreversible at physiological conditions.

1.2.2 Determinants of liquid-liquid phase separation

The essential property of biopolymers driving LLPS is multivalency [79]. With respect to proteins, multivalency is achieved by two archetypical architectures: multiple folded domains and IDRs. In both cases, proteins interact with their partners through multiple interacting motifs or domains. Generally, the higher the valency of a protein, the lower is its c_{sat}, *i.e.*, the stronger is its tendency to undergo LLPS [80].

Multivalent interactions mediated by folded domains function according to the canonical structure-function paradigm established in structural biology. Formation of a defined tertiary structure is mediated by the minimization of the exposure of hydrophobic residues to the surrounding water molecules. The assumed unique fold represents a state of minimal free energy of solvation (Figure 1.4). Possible domain interactions are set by the defined surface topography of the fold [82]. An example for multivalent interactions by folded domains that drive LLPS are the SRC Homology 3 Domain (SH3) domains in the adaptor protein Nck. Nck contains three SH3 domains that each bind to short linear motifs in actin nucleation promoting factor neural Wiskott-Aldrich syndrome protein (N-WASP). Each N-WASP protein contains six of these motifs. The multivalent interactions between Nck and N-WASP concentrate actin nucleation factors and induce the formation of phase-separation droplets in which actin starts to polymerize [76].

In contrast to folded domains, the lack of one defined structural fold is characteristic for IDRs. Instead, they may sample several different structural conformations within a flat energy landscape, or not assume any specific fold at all [82]. Most often, their conformational heterogeneity is the result of the scarcity of aliphatic and aromatic amino acids residues that usually engage in the long-range hydrophobic interactions driving domain folding [80, 84]. The promiscuous interaction behavior of IDRs is often encouraged by short motifs. According to Morris et al. (2021), these may be classified into three categories:

1. Molecular recognition features (MoRFs): about ten to 70 amino acids in length, they adapt a specific secondary structure upon binding that depends on the identity of the interaction partner.
2. Small linear motifs (SLiMs): usually less than ten amino acids long, their primary sequence is sought by specific binding partners and their structure may be ordered or disordered. They may mediate complex formation or signal post-translational modifications or processing.
3. Low complexity regions (LCRs): these regions consist of highly repetitive arrays of select amino acids and may promote binding promiscuity. Of note, regions of low complexity are not necessarily intrinsically disordered [84, 85].

There are two main "flavors" of IDRs, depending on their biochemical composition [85]: (i) IDRs rich in charged amino acid residues often form heterotypic complexes, called coacervates, with nucleic acids of opposite charge (most often negative), and (ii) polar IDRs, which may be punctuated by charged or hydrophobic acids, are usually involved in the formation of homotypic condensates.

Thus, the primary sequence of the IDR dictates the biophysical nature of the emerging condensate [86]. Moreover, variations in the sequence such as IDR length; the

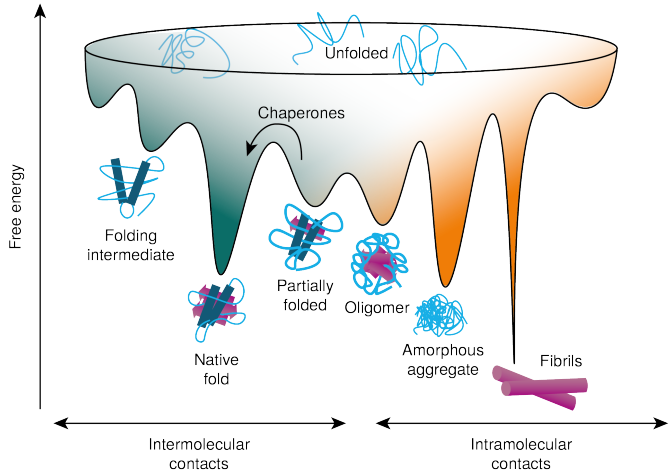

Fig. 1.4.: Schematic energy landscape of protein folding. Protein folding is driven by the requirement to minimize the free energy. The native fold lies in a local free energy minimum, which can be accessed with the help of chaperones. Aberrant intermolecular interactions can lead to trapping in local energy minima, corresponding to the formation of amorphous aggregates or fibrils. The energy landscape for disordered proteins is flat and does not contain deep valleys. Therefore, any specific fold cannot be stabilized within a local energy minimum. Adapted from [83].

number, identity and patterning of interaction motifs; and the frequency and patterning of charged residues can influece the IDR's behavior [80, 87]. Due to this dependence on the biochemical nature of the amino acid residues, the behavior of the IDR and thus the formation, dissolution and composition of condensates can be dynamically regulated by post-translational modifications (PTMs) that change or mask charges of certain residues [79].

We can use algorithms to predict the presence of an IDR from the primary sequence, *e.g.*, using MobiDB or PONDR [73, 80]. Over the past decade, these algorithms have become increasingly accurate. However, the exact mechanisms driving phase separation through the multivalent interactions of IDRs remain still poorly understood. Thus, the phase separation behavior of an IDR within the biochemical milieu of cells can hardly be predicted and requires experimental testing [80].

The RNA helicases DDX4 and LAF-1 are well researched examples of proteins that undergo LLPS driven by their IDRs. DDX4 and LAF-1 play important roles in germ cell development of metazoans and *Caenorhabditis elegans*, respectively [88, 89]. When their IDRs are fused to fluorescent proteins, these fusion proteins form phase separated droplets *in vitro*. As expected, their formation is sensitive to temperature and salt concentrations. In both cases the IDRs contain both negatively and positively charged patches of amino acids [89, 90]. In addition, DDX4 phase separation also depends on numerous repeats containing the aromatic amino acid phenylalanine [79]. However, despite the intense research on these proteins and their IDRs, including atomic simulations and a plethora of *in vitro* and *in vivo* experiments, the mechanisms underlying the phase separation of LAF-1 and DDX4 remain to be understood in full [73, 91, 92].

1.2.3 Liquid-liquid phase separation in chromatin regulation

Nuclear bodies and chromatin are not randomly distributed inside the nucleus. For example, the positioning of interphase chromosomes within their territories is tissue-specific [93]. Many transcriptionally inactive genes are compacted into heterochromatin that is located towards the periphery of the nucleus where it interacts with the lamina [94]. Genes that become activated concomitantly move away from the lamina, towards the center of the nucleus [95]. Active enhancers come

into close spatial proximity to boost the transcription of the associated gene(s) [96, 97]. Spatial organization by LLPS appears to be prevalent in the nucleus. Indeed, proteins with IDRs are enriched in nuclear annotations, *e.g.*, "nucleosome", "spliceosomal complex", and "transcription factor complex" [98]. In addition, chromatin components are particularly rich in multivalent interactions. Histone tails are rich in PTMs and bound by a plethora of chromatin readers that often contain arrays of reader domains as well as IDRs [79, 99]. The following three examples illustrate how chromatin-associated proteins with IDRs influence chromatin organization by undergoing LLPS.

The first example involves the formation of Polycomb bodies. These nuclear structures clustered in pericentromeric heterochromatin were first described in stainings of human cells using antibodies against three components of the Polycomb Repressive Complex 1 (PRC1) [100, 101]. PRCs are transcriptional silencers of developmental genes conserved across metazoans. All PRC1 complexes contain RING1A- or B, which acts as ubiquitin-ligase on H2AK118/119Ub. Based on the selection from a myriad of accessory proteins, numerous variants form which can be subdivided into canonical and non-canonical complexes [102]. One of the canonical PRC1 complexes contains ChromoBox 2 (CBX2), which can undergo LLPS *in vitro* and *in vivo*. Its ability to phase separate has been demonstrated to depend on its positively charged IDR. Upstream of the IDR, CBX2 contains a chromodomain that binds H3K27me3, a histone modification placed by another Polycomb repressive complex, PRC2.

While Polycomb bodies are mostly associated with facultative heterochromatin, LLPS also plays a role in the formation of heterochromatin. A main form of constitutive heterochromatin conserved across eukaryotes is enriched in H3K9me2/3 and Heterochromatin Protein 1 α (HP1α). This type of heterochromatin has the properties of a phase-separated compartment, including nucleation, growth and fusion of spherical foci, and the active diffusion of components within the condensate and across its boundary [103]. In the prevailing model, the heterochromatin phase results from the oligomerization and compaction by HP1α [103–105]. HP1α contains a chromodomain that binds H3K9me2/3, a chromoshadow domain (CSD) for dimerization and ligand binding, and a disordered hinge region in between. The chromobox and CSD are flanked by short IDRs in the N- and C-terminus. Phosphorylation of the proximal N-terminus and DNA binding stimulate LLPS of HP1α *in vitro*.

Besides HP1α, the 5mC reader Methyl CpG binding Protein 2 (MeCP2) is a dynamic component of heterochromatin [106, 107]. The protein possesses two IDRs flanking its methyl-DNA binding domain (MBD), intervening domain (ID), transcriptional repressor domain (TRD) and a NCoR/SMRT interaction domain (NID) [108]. Basic residues arranged in patches in the TRD-ID are required for self-oligomerization, which is necessary and sufficient for MeCP2 LLPS [108, 109]. The basic patches that overlap with the NID appear to be promoting the formation of MeCP2 condensates, probably through electrostatic interactions [107, 108, 110]. Moreover, MeCP2 contains nine MoRFs, which may enable multivalent interactions between MeCP2 and binding partners and further enhance LLPS [111]. The C-terminal IDR, the MBD and the NID are involved in both chromatin compaction and LLPS by MeCP2 [107, 110, 112].

Considering these examples for the formation of condensates in facultative and constitutive heterochromatin, a general model emerges in which high local concentrations of a phase-separating components are achieved by its binding to epigenetic modifications such as 5mC or histone PTMs (H3K27me3, H3K9me2/3). Subsequent oligomerization and phase separation mediated by multivalent interactions leads to the formation of small compartments that recruit more protein. The growing condensates then fuse into larger domains. An increasing fraction of immobile oligomers formed by chromatin-bound protein may lead to the maturation into more gel-like structures. Common to all three examples, the residues promoting LLPS are also required for chromatin compaction. Thus, phase separation and chromatin compaction are intimately linked [103–105, 108, 110, 113].

As a consequence of LLPS, the chromatin encapsulated by the condensates become shielded from certain factors. For example, core transcription factors such as TFIIB cannot enter the HP1α phase, while other components can freely diffuse in and out [104, 114]. Moreover, the condensates exhibit a high resistance to force, contributing to the mechanical stability of heterochromatin compared to euchromatin [114, 115]. Thus, unblending into a separate phase provides both specificity and mechanical stability. Yet, the liquid nature and continuous exchange of the mobile fraction also let this state be reconcilable with the ability of heterochromatin to quickly dissolve in response to intracellular signaling [114].

Besides in the formation of heterochromatin, LLPS also seems to play a role in DNA replication. In 2019, Parker et al. showed that the components of the Cell Divi-

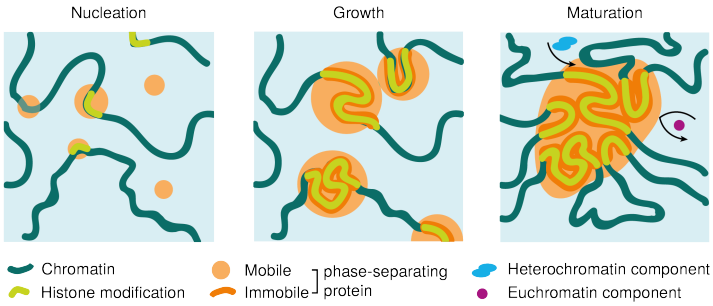

Fig. 1.5.: Model of chromatin compartment formation by LLPS. During nucleation, through the binding of a histone modification, oligomerization, or both, small phase-separated condensates are formed. These condensates subsequently grow and fuse. An increasing fraction of the phase-separating protein becomes immobilized on chromatin, leading to altered properties of the matured phase. This model has been adapted from Strom et al. (2017), who originally devised it for HP1α[103]. However, similar principles have been proposed for MeCP2 and may also be applicable to other chromatin associated condensates [112].

sion Cycle 6 (Cdc6), Chromatin licensing and DNA replication factor 1 (Cdt1) and the Origin Recognition Complex subunit 1 (Orc1) in *Drosophila melongaster* can undergo LLPS on DNA *in vitro* [116]. Together with the multimeric origin recognition complex (ORC), Cdc6 and Ctd1 assemble at replication origins to facilitate the loading of the Mcm2-7 helicase, which co-partitions into ORC-Cdc6-Cdt1 droplets [116, 117]. Cdc6, Cdt1 and Orc1 contain IDRs in their N-terminus, which are required for their LLPS behavior. While the sequence itself is not conserved, their biochemical signature with respect to isoelectric point (pI) and fraction of charged residues is similar across metazoans (with the exception of *Saccharomyces cerevisiae*). In line with this, human CDT1 is also able to phase-separate *in vitro*. Specifically, the IDRs of the three replication initiation factors have approximately equal fractions of charged, hydrophilic and hydrophobic residues, and lack aromatic residues as well as glycine. The large percentage of hydrophobic residues is very unusual for phase-separating IDRs and distinct from the composition of IDRs of other known phase-separating proteins such as LAF-1 (see Section 1.2.2). Their amino acid composition results in a positive pI, which predisposes them to complex coacervation with DNA [116, 118]. The Orc1, Cdc6 and Cdt1 IDRs contain a swath of CDK phosphorylation motifs, and at least ORC phosphorylation inhibits its assembly *in vivo* [119–122]. *In vitro*, phosphorylation inhibits LLPS of all three initiators [116].

Taking this experimental evidence together, a possible model emerges in which the ORC condenses along chromatin whereby its readout of histone modifications, such as H4K20me2 by the Orc1 BAH domain, and other features of chromatin architecture guide them to licensed replication origins [123]. Intramolecular interactions

between the IDR further enrich for ORC protein and attract the initiators Cdc6 and Cdt1. Mcm2-7 diffuses into the phase and is loaded onto chromatin, assembling the pre-replicative complex. After loading, phosphorylation of the initiator IDRs disturbs the weak interactions and the condensed phase disperses. In this model, LLPS would both increase the efficiency of the loading process and present a mode permitting its regulation by phosphorylation [116, 118]. Though this model awaits further experimental validation, it is attractive to speculate that the role of LLPS extends beyond the initiation of DNA replication, since many conserved proteins involved in later steps of DNA replication also possess long IDRs that may mediate LLPS [116, 118]. For example, spreading of a replication phase along chromatin could promote the *cis*-activation of nearby replication origins and *de novo* assembly of replisomes observed *in vivo* [124–126]. Moreover, LLPS may be used for the temporal control of replication foci composition, or simply generally increase the recruiting efficiency of required factors, including proteins essential for the post-replicative maintenance of chromatin such as DNMT1.

1.3 Motivation and aims of this study

While 5mC patterns in prokaryotes are determined by the genetically encoded structure of the DNMT MTase domain, the diversity of methylomes in eukaryotes cannot be explained by the presence of specific DNMTs alone. Instead, an important factor that contributes to the multitude of different 5mC patterns is the spatio-temporal control of DNMT activity. This includes a complex interaction network of the DNMTs with various partners through their regulatory N-terminal domains.

Since *Dnmt1* was first cloned in 1988, many of its regulatory N-terminal domains have been mapped and investigated for their role in maintenance methylation. In total, six different domains have been found and for most of them, their interaction partners have been identified. Yet, a holistic model of the spatio-temporal control of maintenance methylation by DNMT1 is still missing, as the different recruitment mechanisms appear to be partially redundant and/or ordered in a complex hierarchy.

As the majority of past research has focused on the structured parts of the protein, a potential role of the DNMT1 IDR has long been overlooked. However, over the past years a large collection of studies on other proteins has highlighted that these unstructured stretches can play important roles in protein function. This finding extends to many proteins that operate in the same nuclear space as DNMT1, such as heterochromatin and DNA replication foci. Extrapolating from the work on the IDRs of other epigenetic regulators, we hypothesize that engagement of the DNMT1 IDR in multivalent interactions may predispose the enzyme to partition into a separate liquid phase and increase the efficiency of its recruitment to specific chromatin contexts.

Thus, this project aims to understand the contribution of the DNMT1 IDR to the enzyme's behavior. First, we approach the analysis of the IDR from an evolutionary perspective to identify conserved signatures that may point towards critical biochemical features. Second, we characterize the LLPS behavior of the IDR *in vitro* and by opto-genetic analysis in live cells. Finally, we establish and optimize a system for the controlled exchange of the endogenous DNMT1 protein with mutant isoforms in mouse embryonic stem cells, providing a platform to investigate the effect of various IDR mutants on maintenance methylation in the future.

2

Evolutionary conservation of the DNMT1 N-terminus

A handful of prior studies noted the N-terminus of mouse and human DNMT1 is predicted to contain an IDR [70–72]. These unstructured regions commonly provide a platform to regulate protein behavior [84]. For example, embedded linear peptide motifs can tune the interaction specificity, affinity and avidity. Moreover, IDRs provide accessible sites for PTMs that can alter the chemical composition and thereby modulate interaction properties. Lastly, IDRs can connect structured domains in a non-stoichiometric fashion, increasing conformational heterogeneity [82]. Thus, we hypothesize that the DNMT1 IDR may provide a means to dynamically regulate the enzyme's recruitment and function, thereby affecting DNA methylation patterns.

To study the DNMT1 IDR, we first used state-of-the-art methods to define its location and to compare it to the structure predictions of the *de novo* DNMTs, DNMT3A and -B. Specific functions of a protein place constraints on their evolution, which generally results in the conservation of the primary sequence [127]. Therefore, we sought to compare the conservation of the DNMT1 IDR with the other regulatory domains in the N-terminus and to compare it to the disordered regions in DNMT3A and DNMT3B. We then took a closer look at the conservation of the biochemical composition of the DNMT1 IDR to identify features that may predict a function in LLPS.

2.1 Prediction of intrinsically disordered regions in mammalian DNMTs

Over 40 computational tools exist to predict the presence of an IDR from the primary sequence of a protein. A prediction of disorder can be made based on physiochemical properties of the amino acid residues (*e.g.*, hydrophobicity, net charge), machine-learning based methods (classification models and sequence labeling methods), or template-based methods using homology of protein sequences, or a mix of the mentioned methods [128]. Previously, a disordered region in the DNMT1 N-terminus was predicted using DISOPRED2, a classification model using support vector machines and trained on a set of X-ray structures from the Protein Data Bank [70, 129, 130]. However, DISOPRED2 has been found to markedly under-predict disorder [131].

Since the initial prediction, many other predictors have been developed that outperform DISOPRED2. To follow up on the originally reported IDR predictions of DNMT1, we used Metapredict, a state-of-the-art deep-learning-based prediction tool for disordered regions in proteins, and used it to search the mouse DNMT protein sequence for disordered regions (see Section 6.1) [132]. Metapredict is trained on the output of eight different prediction tools and computes a consensus disorder score, representing the confidence in the predicted disordered state of an amino acid residue. In addition, Metapredict is also trained on AlphaFold2 pLDDT scores. AlphaFold2 pLDDT scores represent the local confidence of the AlphaFold2 algorithm in its predicted structure [133]. Metapredict computes the AlphaFold2 pLDDT scores and combines these with its consensus disorder score to provide highly accurate predictions of disordered amino acid residues.

We used Metapredict to compute the consensus disorder score and the AlphaFold2 pLDDT score of mouse DNMT1, DNMT3A and DNMT3B (Figure 2.1 A, 2.2 A, 2.3 A). All described, annotated domains of the DNMTs overlap with high AlphaFold2 pLDDT scores and low Metapredict consensus disorder scores, confirming the presence of defined tertiary structures in these regions. Conversely, all three DNMTs display long stretches characterized by high scores for disorder and simultaneously low AlphaFold2 pLDDT scores in their N-terminus, indicative of a disordered region. We therefore used the built-in Metapredict `predict_disorder_domains()` function

to define disordered domains in each protein. The N-termini of DNMT1, DNMT3A and DNMT3B are predicted to contain contiguous disordered domains of 217, 278 and 223 amino acid residues in length, respectively (Figure 2.1 A, 2.2 A, 2.3 A).

We then confirmed these results by comparing the disorder prediction by Metapredict with the protein structure as predicted by AlphaFold2 [133, 134] (Figure 2.1 B, 2.2 B, 2.3 B). Generally, AlphaFold2 displays a very high confidence (dark blue) in predicting the structure of the known DNMT domains and its prediction corresponds well to the determined partial crystal structures published in literature. However, AlphaFold2 displays low confidence (orange) across large parts of the N-termini of all three DNMTs, resulting in the lasso-like placement of the proximal N-terminal sequences around the predicted core structure. In addition, AlphaFold2 returns high predicted alignment error values (light green) for the N-termini, representing a high uncertainty in the relative orientation of the amino acid residues towards each other, as opposed to the annotated domains which correspond to areas of low predicted alignment errors (dark green) (Figure 2.1 C, 2.2 C, 2.3 C). Exceptions are the medium to high confidence in the predicted α helix formed by L196 to E225 of DNMT3A, roughly corresponding to the UDR reported by Weinberg et al. (2021), and the four α helices between N22 and E93 forming the DMAP1 domain in the proximal N-terminus of DNMT1 [49].

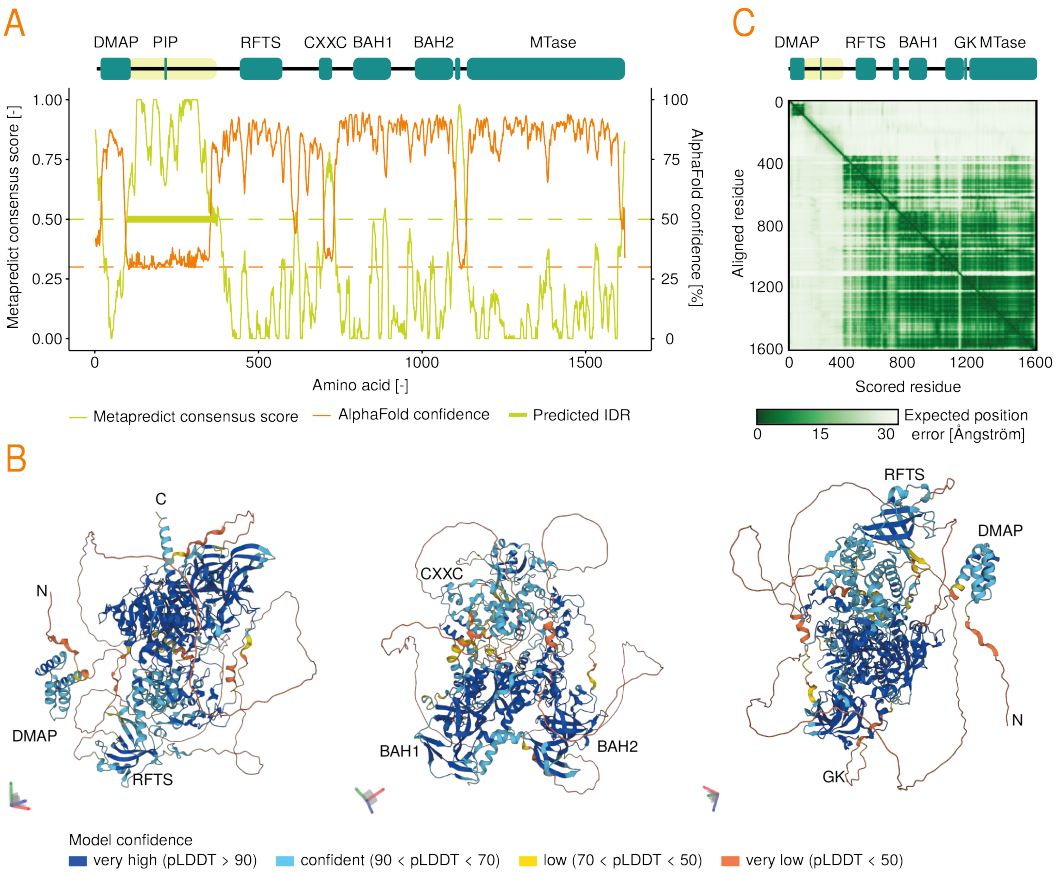

Fig. 2.1.: Predicted structure of mouse DNMT1. A Metapredict scores and AlphaFold2 pLDDT scores for mouse DNMT1. The predicted IDR in the N-terminus (L99-P367) is highlighted. B Structure of DNMT1 as predicted by AlphaFold2 in three different orientations. C Predicted aligned error of the structure generated by AlphaFold2. BAH: bromo-adjacent homology domain; CXXC: Zinc finger binding domain; DMAP: DNA methyltransferase 1-associated Protein; GK: glycine lysine linker; MTase: methyltransferase domain; PIP: PCNA-interacting protein box; RFTS: replication foci targeting sequence.

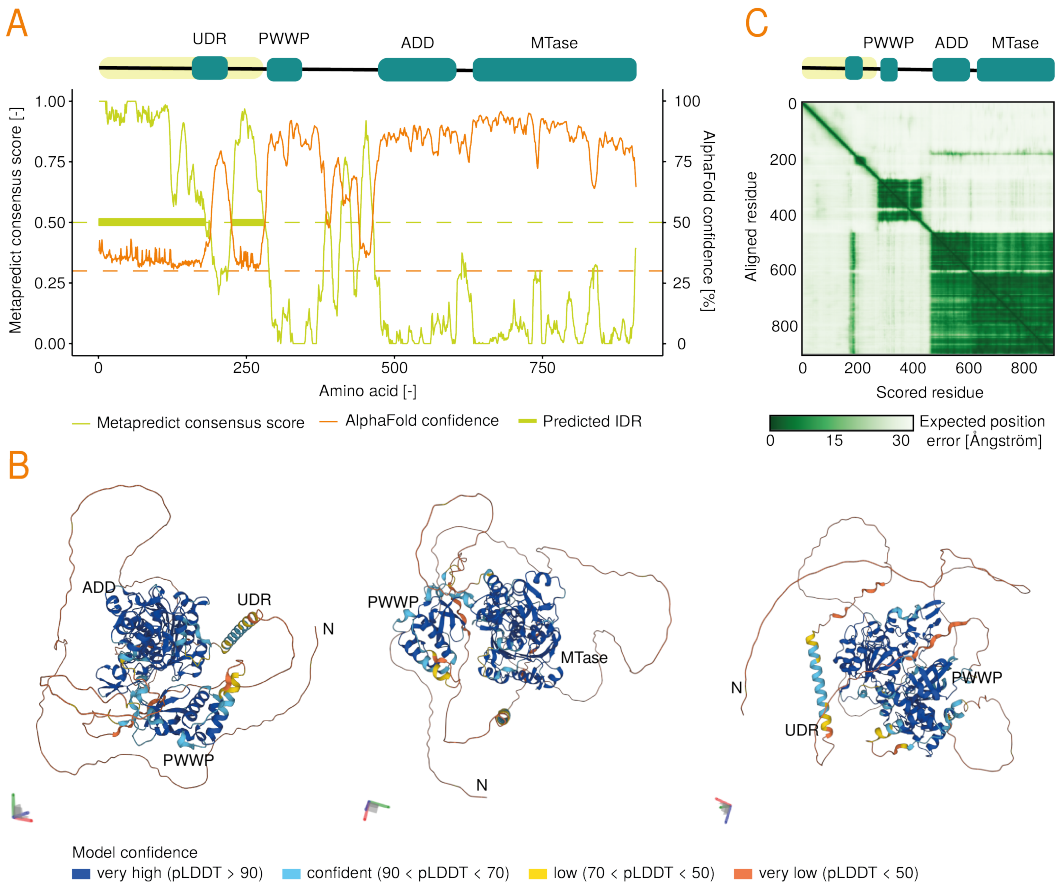

Fig. 2.2.: Predicted structure of mouse DNMT3A. A Metapredict scores and AlphaFold2 pLDDT scores for mouse DNMT3A1. The predicted N-terminal IDR (M1-P279) is highlighted. B Structure of DNMT3A1 as predicted by AlphaFold2 in three different orientations. C Predicted aligned error of the structure generated by AlphaFold2. ADD: ATRX-DNMT3-DNMT3L domain; MTase: methyltransferase domain; PWWP: Pro-Trp-Trp-Pro domain; UDR: ubiquitin-dependent recruitment region.

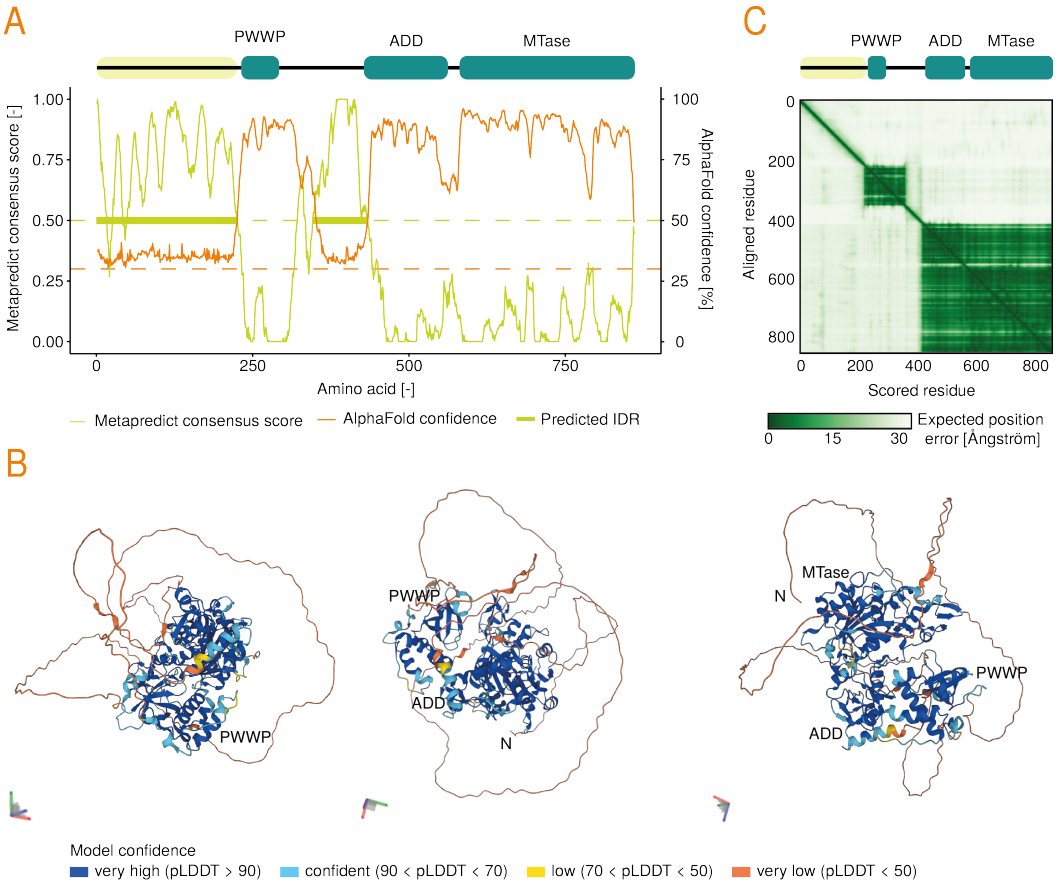

Fig. 2.3.: Predicted structure of mouse DNMT3B. A Metapredict scores and AlphaFold2 pLDDT scores for mouse DNMT3B1. The predicted N-terminal IDR (M1-E224) is highlighted. B Structure of DNMT3A1 as predicted by AlphaFold2 in three different orientations. C Predicted aligned error of the structure generated by AlphaFold2. ADD: ATRX-DNMT3-DNMT3L domain; MTase: methyltransferase domain; PWWP: Pro-Trp-Trp-Pro domain.

2.2 Evolutionary conservation of the DNMT IDR structure

Given that all three main DNMTs in mouse contain strong IDR signatures, we wondered whether this region contributes to the proteins' function. Since conservation across evolution is a good predictor for function [135], we were interested in determining how strong this signature is preserved among closely and more distantly related species. We therefore first determined how the disorder predictions for mouse DNMTs compare to DNMTs of three other mammals, namely human, pig and koala. Using Metapredict, we computed the disorder consensus scores and the AlphaFold scores and plotted them along the protein sequence (Figure 2.4) (see Section 6.1). All four species display a very similar pattern of disorder within the N-termini of the proteins and the predicted IDRs fall nearly within the same regions (Figure 2.4). Of note, the predicted IDR of DNMT1 in the most distant species compared to mouse, the koala, is much shorter compared to the IDR in the other three mammalian DNMT1s. While the IDR in murine DNMT1 is predicted to be 273 amino acids long (L98 to P369), the length of the predicted IDR of koala DNMT1 is only 104 amino acids (T112 to K216). Besides the predicted disordered domain, all DNMTs contain shorter disordered regions which may serve as linker, *e.g.*, between the PWWP and ADD domains with DNMT3A and -B.

To cover more distantly related species, we obtained the protein sequences for DNMT1, DNMT3A and DNMT3B of 47 animal species, including eu- and metatherian mammals, fish, amphibians, reptiles and insects, as well as the corresponding homologs from two plant species (Dnmt1 and Met1 for DNMT1, and Drm2 for DNMT3A/B, respectively) from the NCBI database [136] (see Section 6.1). We then aligned the protein sequences of each enzyme and computed their phylogenetic relationship based on sequence similarity. Figures 2.5, 2.6 and 2.7 visualize the multiple sequence alignments (MSA) of DNMT1, DNMT3A and DNMT3b, respectively. Grey boxes denote aligned sequences, which may not be identical but generally similar. The black lines represent empty positions in the MSA, meaning that for a given species' protein sequence, there are no amino acids matching other sequences present in the alignment. The phylogenetic relationship shown left to the MSAs roughly correspond to the evolutionary relationship between the differ-

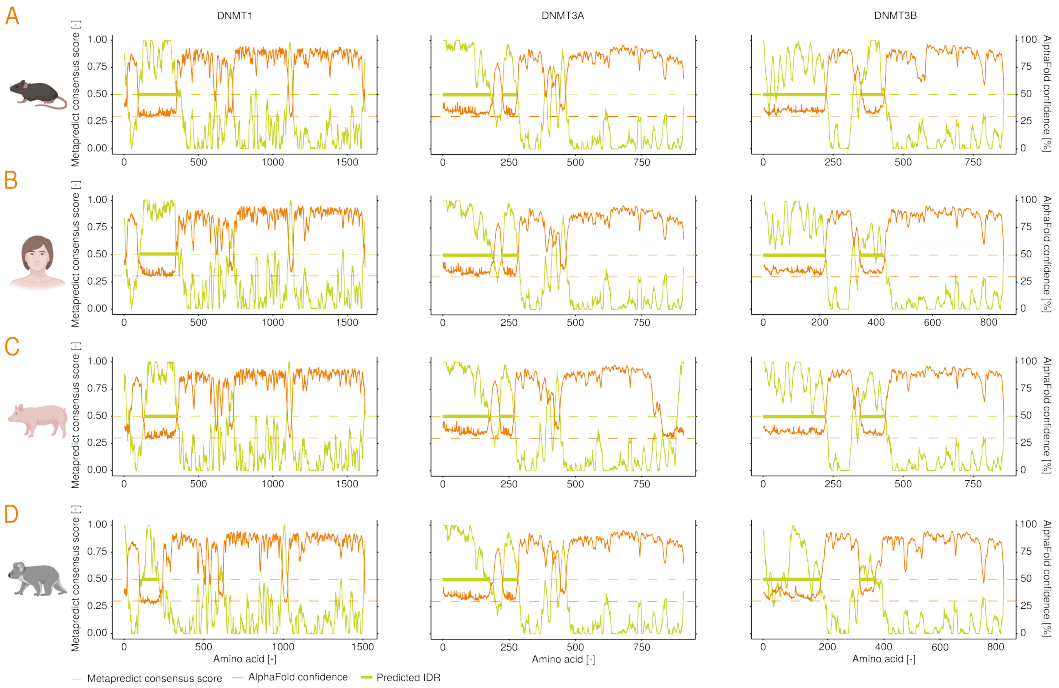

Fig. 2.4.: DNMT1 IDR prediction for mammalian species. Metapredict scores and AlphaFold2 pLDDT scores for DNMT1, DNMT3A and DNMT3B in A mouse, B human, C pig, and D koala. Animal pictograms were created with `BioRender.com`.

ent species. For example, all eutherian mammals cluster together in the DNMT1 phylogenetic tree (Figure 2.5).

The domain structure of the DNMTs is highly conserved among all animal species. With a few exceptions, all annotated DNMT domains are present across the animal kingdom. In addition, the protein sequence corresponding to the IDR in mouse DNMT3A and DNMT3B is also present in all other queried animal species. For DNMT3B, often smaller pieces of the IDR are missing in non-eutherian species. Interestingly, eutherian mammals display an extended N-terminus compared to other animal species. This region of interest (ROI) is located just downstream of the PIP box and corresponds to the amino acids L98 to P369 in mouse DNMT1. Of note, the ROI is located within the IDR of mouse DNMT1. Thus, the longer predicted IDR of mouse, human and pig DNMT1 compared to koala is not specific to this selection of species, but a characteristic that applies to the entire extant eutherian infraclass of Placentalia compared to all other queried animal species. Thus, Placentalia appear to have evolved a longer IDR, extended by the ROI, that sets the DNMT1 structure apart even from the next closes relatives, the Marsupiala. From here on, we focus on studying the IDR within the DNMT1 N-terminus and its extension in eutherian

mammals by the novel ROI. As a side note, both extant metatherian (Marsupiala) and eutherian mammals (Placentalia) develop functional placentas [137]. Thus, we refrain from using the name "placental mammals" to describe members of Placentalia in this study and instead collectively refer to them as eutherian mammals, since Placentalia is the only extant infraclass of Eutheria.

To quantify how conserved the aligned sequences are, we calculated a conservation score based on Jensen-Shannon divergence (see Section 6.1). Shown in Figure 2.8 is the conservation score plotted against the MSA of DNMT1, DNMT3A and DNMT3B across all queried species. Confirming the visualization of the MSA in Figures 2.5, 2.6 and 2.7, the conserved regulatory domains, such as PWWP, ADD and BAHs, as well as the MTase domain have high conservation scores (>60), in line with their structural and functional constraints. On the lower end, sequences present in only one species per MSA return a conservation score near 0, *e.g.*, the extension within the MTase domain of chicken (compare Figure 2.7 and Figure 2.8 C). Between the sequences present in only few species and the ancient DNMT domains, the DMAP1 domain of DNMT1 has a low conservation score of 0.32. Similarly, the N-terminal IDRs of the DNMTs exhibit low sequence conservation, scoring 0.33, 0.41 and 0.36 for DNMT1, DNMT3A and DNMT3B, respectively.

Given the low sequence conservation of the IDRs, we wondered whether the molecular feature of disorder is conserved across the species. We therefore computed the Metapredict consensus score and AlphaFold confidence score for all the species and plotted the mean values along the MSA of each protein (Figure 2.9). The shaded area reflects the standard deviation as an indicator for the variability of the disorder prediction per amino acid residue. For better interpretability, we only show the MSA positions aligning to the murine protein. AlphaFold does not predict any structure formed by the N-termini with little variability in its prediction. In addition, Metapredict clearly considers the N-termini disordered, with score well above 0.50. However, the variability of disorder across all species is higher in DNMT3A and DNMT3B compared to DNMT1, where the IDR has uniformly high disorder scores. Thus, even though the primary sequence is only moderately conserved in the DNMT N-termini, these regions are still disordered in all the species.

Moreover, the IDR extension present only in the eutherian mammals likely confounds the calculation of the conservation score for the DNMT1 IDR. Therefore, we filtered the species list to contain only eutherian mammals and re-calculated

the conservation score for this selection of species only (Figure 2.10). Restricting the analysis of DNMT1 to eutherian mammals increases the scores compared to the analysis of the complete species set. Nonetheless, the conservation score of the IDR is still much lower than that of other domains, such as the RFTS domain, even though in this analysis all species contain the ROI extension within the IDR. Like the surrounding IDR, the amino acid sequence of the ROI itself is also only moderately conserved. Also for the subset of eutherian mammals, the Metapredict consensus score is high across the IDR, including the ROI, and is only slightly lower around the PIP motif (Figure 2.10 B). Accordingly, the AlphaFold confidence score is low throughout the IDR, sharply increasing beyond its predicted boundary. The standard deviation of both consensus and confidence scores is narrow, indicating that despite slightly different primary sequence, the disordered nature is conserved across all sampled species.

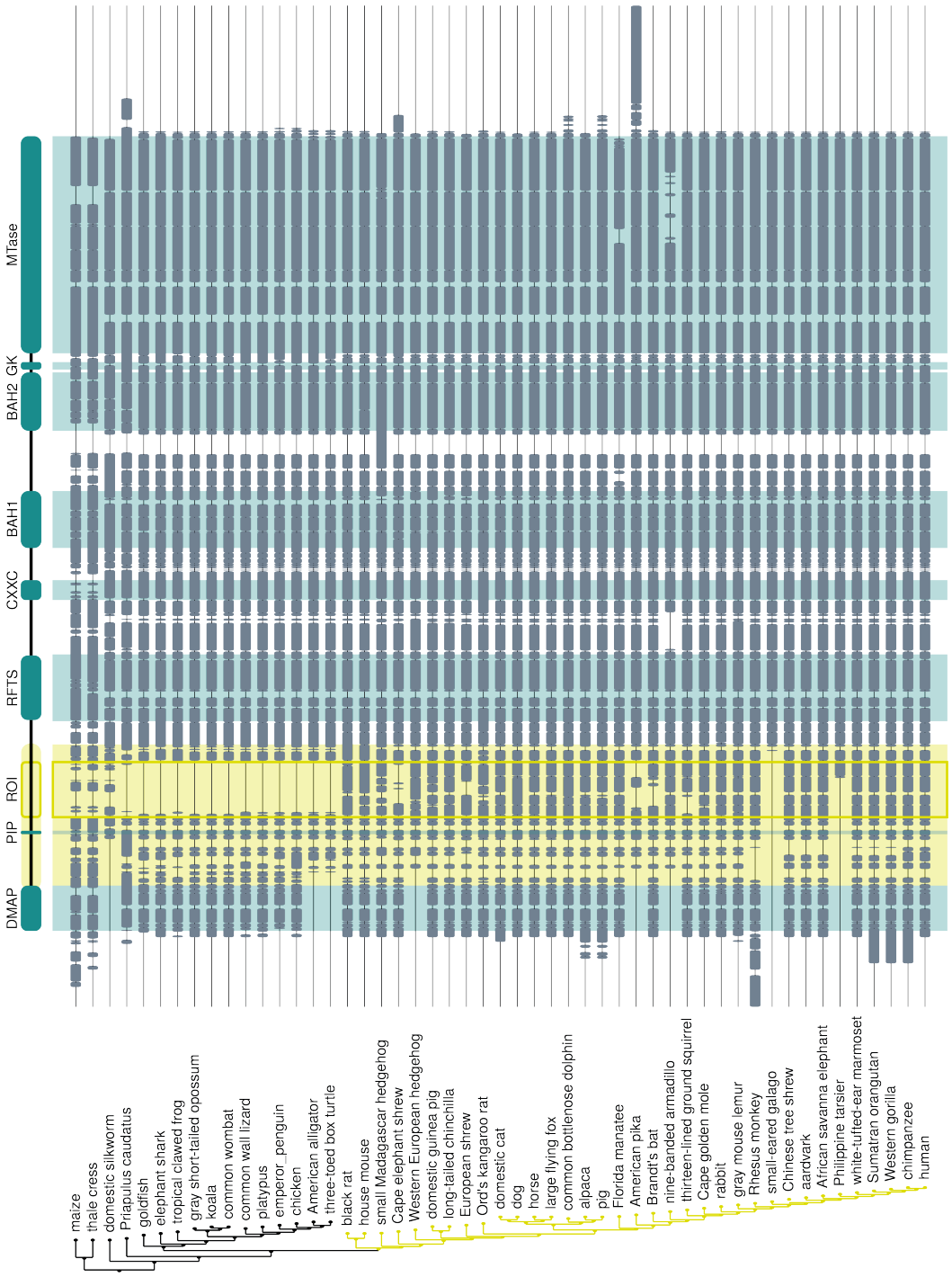

Fig. 2.5.: MSA of eukaryotic DNMT1 across eukaryotes. In total 49 eukaryotic species, including 35 eutherian mammals (highlighted in green), were analyzed. The domain annotation is given for murine DNMT1. Position and length of the domains were obtained from (The UniProt Consortium (2021), accession no. P13864). The region of interest (ROI) is outlined in light green. The IDR identified by Metapredict is shaded in light green. BAH: bromo-adjacent homology domain; CXXC: Zinc finger binding domain; DMAP: DNA methyltransferase 1-associated Protein; GK: glycine lysine linker; MTase: methyltransferase domain; PIP: PCNA-interacting protein box; RFTS: replication foci targeting sequence.

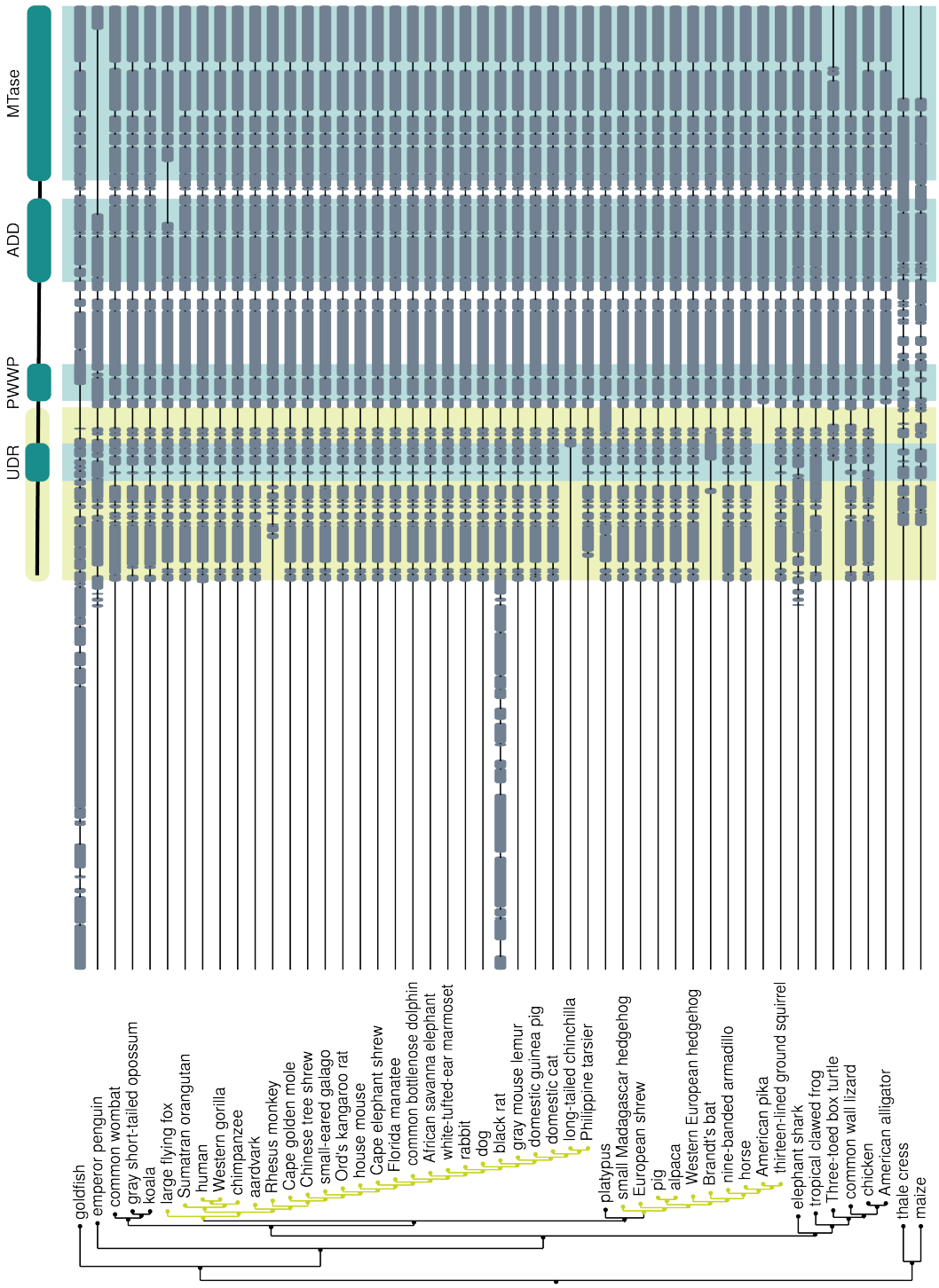

Fig. 2.6.: MSA of DNMT3A across eukaryotes. In total, 49 eukaryotic species, including 35 eutherian mammals (highlighted in green), were analyzed. The domain annotation is given for murine DNMT3A1. Position and length of the domains were obtained from (The UniProt Consortium (2021), accession no. O88508). The UDR was annotated according to [49]. The N-terminal IDR identified by Metapredict is shaded in light green. ADD: ATRX-DNMT3-DNMT3L domain; MTase: methyltransferase domain; PWWP: Pro-Trp-Trp-Pro domain; UDR: ubiquitin-dependent recruitment region.

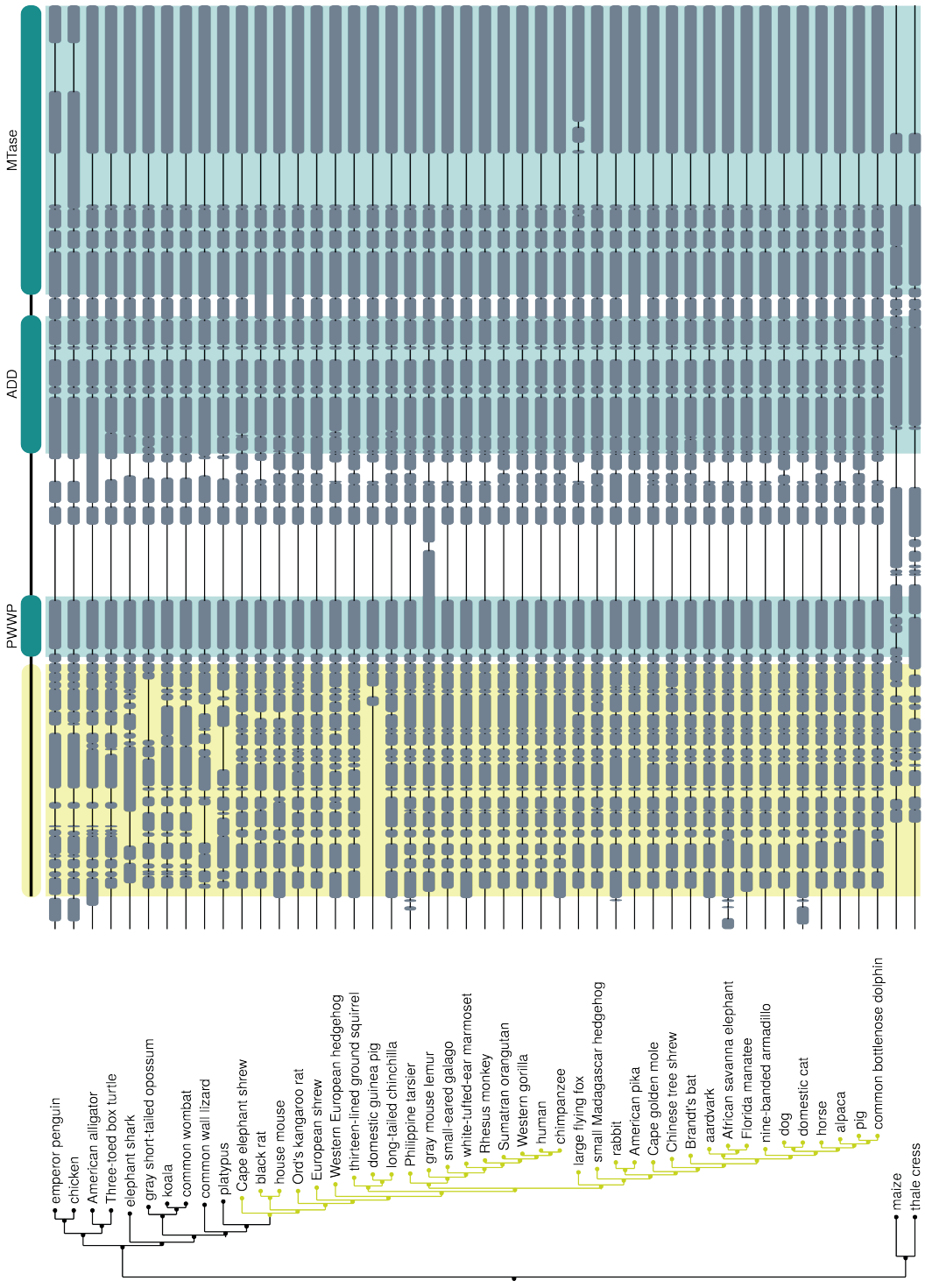

Fig. 2.7.: MSA of eukaryotic DNMT3B across eukaryotes. In total, 47 eukaryotic species, including 35 eutherian mammals (highlighted in green), were analyzed. The domain annotation is given for murine DNMT3B1. Position and length of the domains were obtained from (The UniProt Consortium (2021), accession no. O88509). The N-terminal IDR identified by Metapredict is shaded in light green. ADD: ATRX-DNMT3-DNMT3L domain; MTase: methyltransferase domain; PWWP: Pro-Trp-Trp-Pro domain.

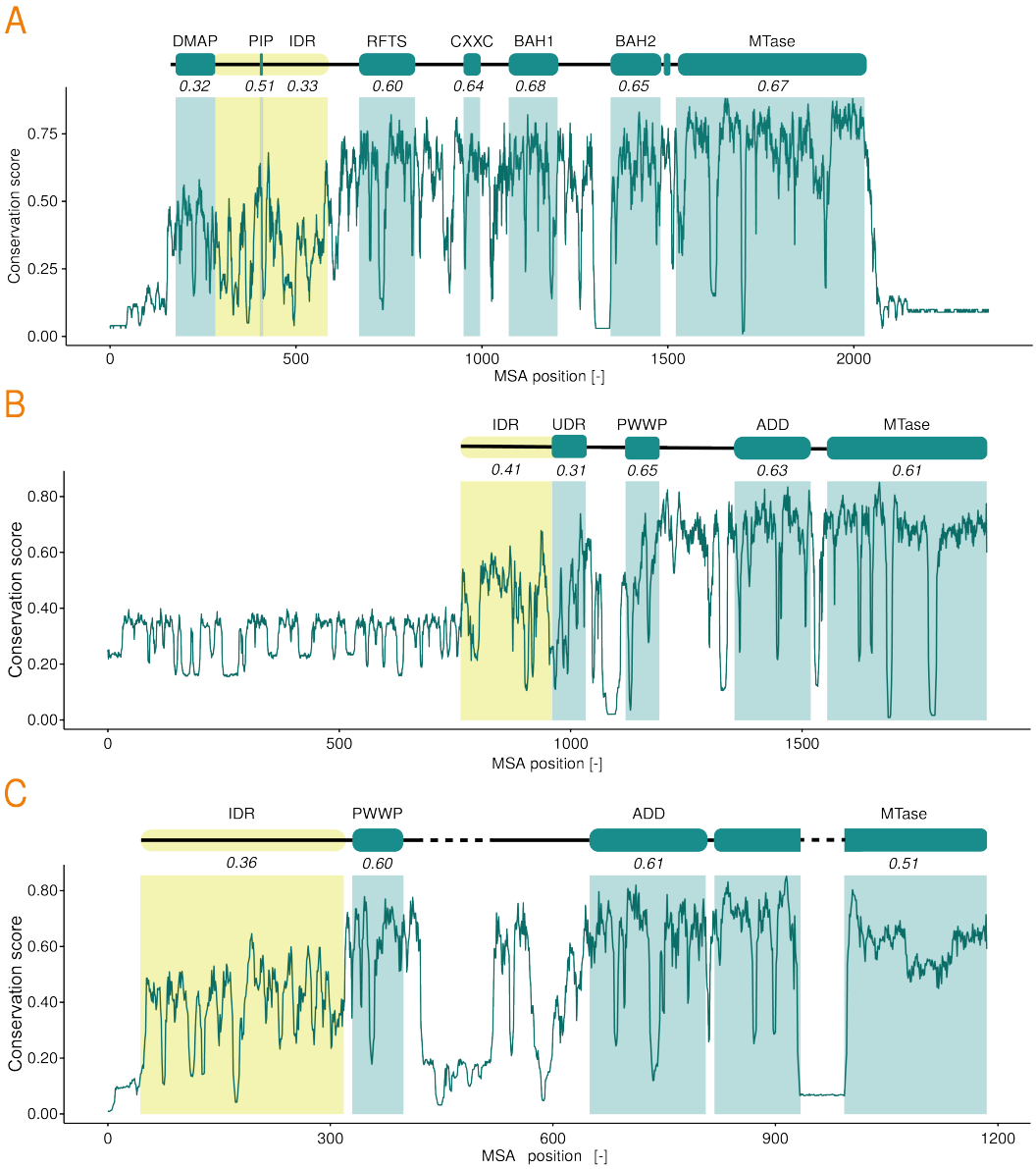

Fig. 2.8.: Conservation of DNMT sequence across eukaryotes. Jensen-Shannon divergence based conservation score for the multiple sequence alignments of A DNMT1, B DNMT3A and C DNMT3B shown in Figures 2.5, 2.6 and 2.7. ADD: ATRX-DNMT3-DNMT3L domain; MTase: methyltransferase domain; BAH: bromo-adjacent homology domain; CXXC: Zinc finger binding domain; DMAP: DNA methyltransferase 1-associated Protein; GK: glycine lysine linker; PIP: PCNA-interacting protein box; PWWP: Pro-Trp-Trp-Pro domain; RFTS: replication foci targeting sequence; UDR: ubiquitin-dependent recruitment region.

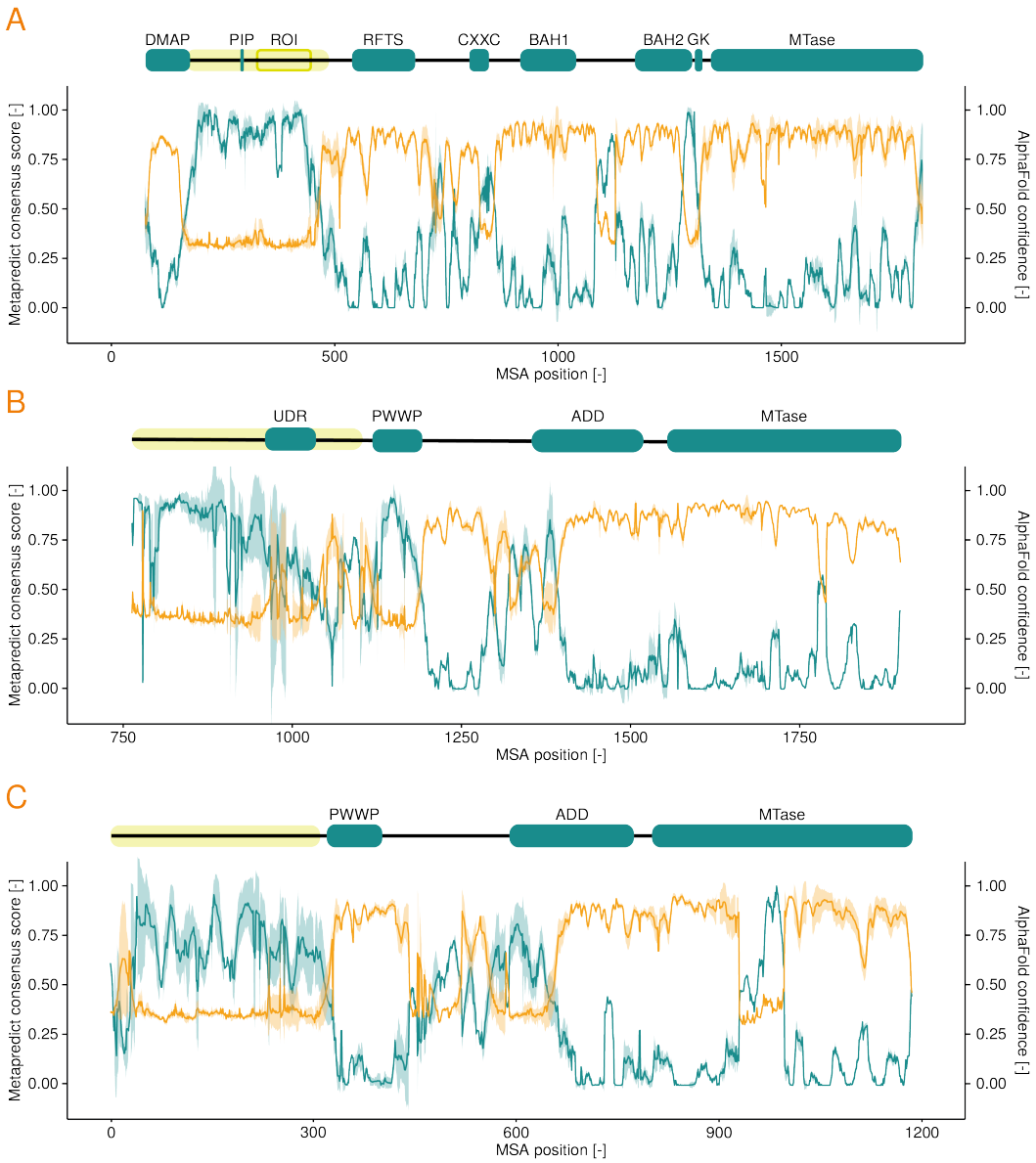

Fig. 2.9.: Disorder prediction for DNMTs across eukaryotes. Metapredict scores and AlphaFold2 pLDDT scores for the mulitple sequence alignment of A DNMT1, B DNMT3A and C DNMT3B. ADD: ATRX-DNMT3-DNMT3L domain; MTase: methyltransferase domain; BAH: bromo-adjacent homology domain; CXXC: Zinc finger binding domain; DMAP: DNA methyltransferase 1-associated Protein; GK: glycine lysine linker; PIP: PCNA-interacting protein box; PWWP: Pro-Trp-Trp-Pro domain; RFTS: replication foci targeting sequence; UDR: ubiquitin-dependent recruitment region.

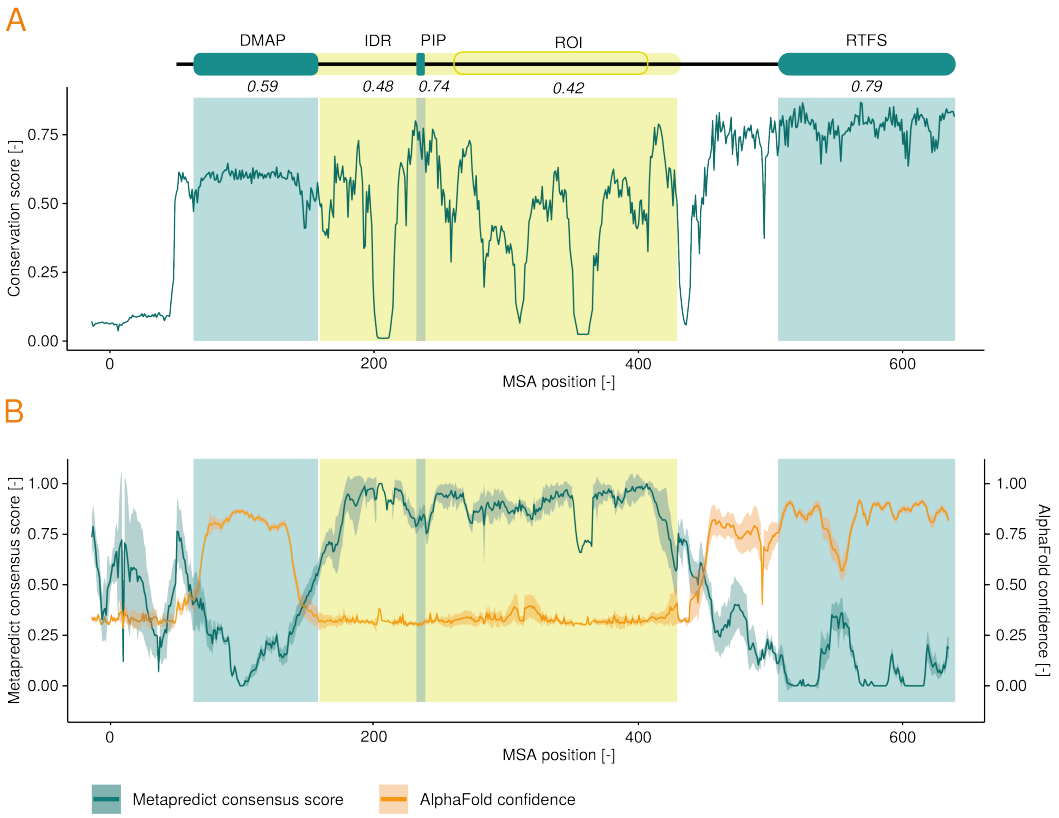

Fig. 2.10.: Conservation of the DNMT1 N-terminal region across Placentalia. A Jensen-Shannon divergence based conservation score and B metapredict consensus score and AlphaFold confidence for the multiple sequence alignment of DNMT1 in Placentalia only. The IDR is shaded in green and the eutherian-specific ROI is outlined in green. DMAP: DNA methyltransferase 1-associated Protein; IDR: intrinsically disordered domain; PIP: PCNA-interacting protein box; RFTS: replication foci targeting sequence; ROI: region of interest.

2.3 Biochemical features of the DNMT1 intrinsically disordered region

Given the relatively low sequence conservation of the DNMT1 IDR across eutherian mammals, we wondered whether other features such as biochemical composition and pI are conserved across this infraclass. Analogous to Parker et al. (2019), we classified the amino acids into five categories of aromatic (F, W, Y), hydrophobic (A, G, I, L, M, P, V), hydrophilic (C, N, Q, S, T), basic (H, K, R) and acidic (D, E) [116]. Aliphatic and aromatic amino acids residues that usually engage in the long-range hydrophobic interactions driving domain folding are commonly scarce in IDRs [80, 84]. The fraction of charged and polar residues may hint at the type of potential condensates formed. An over-representation of charged amino acids could indicate an IDR that tends to form heterotypic condensates. Particularly positively charged IDRs are prone to form coacervates with negatively charged nucleic acids. Conversely, IDRs rich in hydrophilic acids rather form homotypic condensates [85].

We used these five categories to compare the amino acid composition of the IDR, exemplary for mouse, human, pig, and rabbit (Figure 2.11; see Section 6.1). In all four species, the IDR contains about 20 % acidic, 25 % basic, 20 % hydrophilic, 30 % hydrophobic and less than 1 % aromatic amino acids (Figure 2.11 A-D). Compared to the average amino acid composition of non-membrane bound proteins (12 % acidic, 14 % basic, 24 % hydrophilic, 43 % hydrophobic, and 7 % aromatic), hydrophilic, acidic, and basic amino acids are over-represented, while hydrophobic and aromatic amino acids occur less frequently [139].

However, the pI resulting from this amino acid composition is markedly different between the species, ranging from 6.4 in rabbit to 9.5 in pig. Given an approximately neutral nuclear pH, the IDR of mouse, human and pig DNMT1 are positively charged, while the IDR of rabbit may be slightly negatively charged [140]. However, when only considering the novel eutherian-specific ROI (Figure 2.11 E-F), both amino acid composition and a pI around 5 appear to be conserved, meaning the ROI in each case adds mostly negative charges to the IDR.

Nonetheless, when subjected to a principal component analysis (PCA) based on amino acid composition and pI, the DNMT1 IDRs from mouse, human, rabbit, and

Fig. 2.11.: Amino acid composition of the eutherian DNMT1 IDR. A-D Amino acid composition and pI of the mouse, human, rabbit and pig DNMT1 IDR. E-H Amino acid composition and pI of the mouse, human, rabbit and pig DNMT1 ROI. Amino acids were categorized as aromatic (F, W, Y), hydrophobic (A, G, I, L, M, P, V), hydrophilic (C, N, Q, S, T), basic (H, K, R), or acidic (D, E), based on Parker et al. (2019) [116]. Animal pictograms were created with `BioRender.com`.

pig are clustered close to each other and separate from the IDRs of other epigenetic regulators (Figure 2.12). Generally, the IDRs from the same molecule appear to have similar characteristics and form clusters. In this PCA, principal components (PC) 1 and 2 together explain 78.21 % of the variation (51.34 % by PC1 and 26.87 % by PC2). Interestingly, the DNMTs appear most similar to the IDRs of CBX2, presumably a driving component of the phase-separated Polycomb bodies [141], as well as components of the replication machinery, such as POLD3 and CDT1, along PC1. Conversely, the IDRs of the TET proteins gather on the opposing end of the DNMT1 IDRs with respect to PC1 and PC2. The IDRs known of components of heterochromatin such as HP1α and TRIM28 cluster away from the DNMT1 IDRs along PC1 and PC2, but are also diametrically positioned with respect to each other along PC1 and PC2, despite being known interaction partners and co-locating in heterochromatic condensates [142, 143].

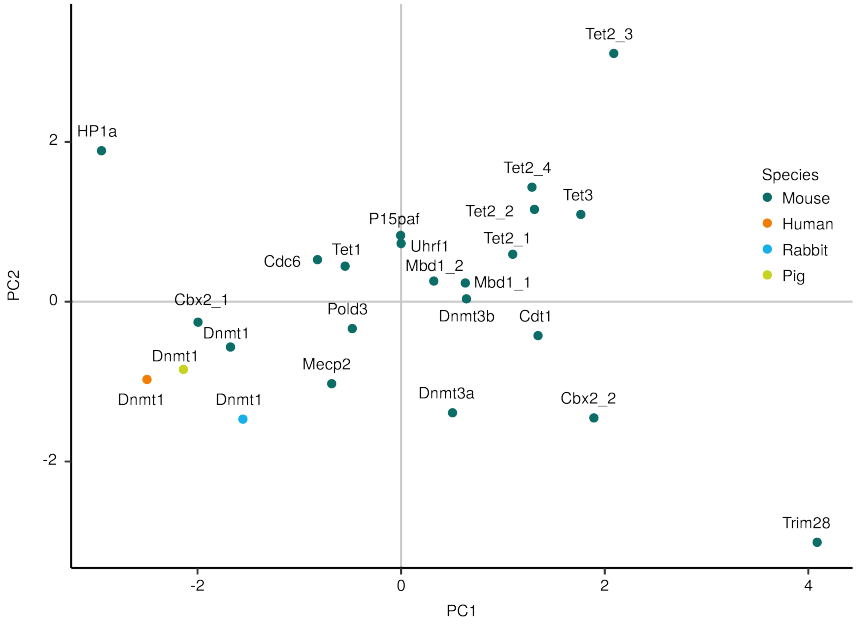

Fig. 2.12.: PCA of the IDR composition of mammalian DNMT1 and its interaction partners. The analysis was carried out on the bases of the amino acid composition and the pI of the IDRs of 23 known interaction partners of DNMT1 and the IDR of DNMT1 from four mammalian species. PC1 and PC2 explain 51.34 % and 26.87 % of the variation, respectively.

Taking a closer look at the amino acid distribution within the DNMT1 IDR, we plotted the presence of all 20 naturally occurring amino acids along its primary sequence. This visualization is particularly useful for identifying amino acid patches characteristic of LCRs. As shown in Figure 2.11, the four selected mammalian IDRs are consistently rich in proline (P), glutamic acid (E), lysine (K) and arginine (R). The latter three frequently occur in patches, particularly within the ROI. Here, especially the repeats of glutamic acid seem to be conserved across the selected species, consistent with the weakly acidic pI of the ROI (Figure 2.11 E-F). The proximal part of IDR conserved in all species is rich in serine (S) and proline (P). Downstream of the serine and proline rich region, in line with its high conservation score, the PIP motif is almost identical in all four species (QTTITAHF for mouse, human and pig; QTTITSHF for rabbit), and is located in a region rich in tyrosines (T). Overall, the biochemical features conserved in the DNMT1 IDR appear to be compatible with a role in LLPS.

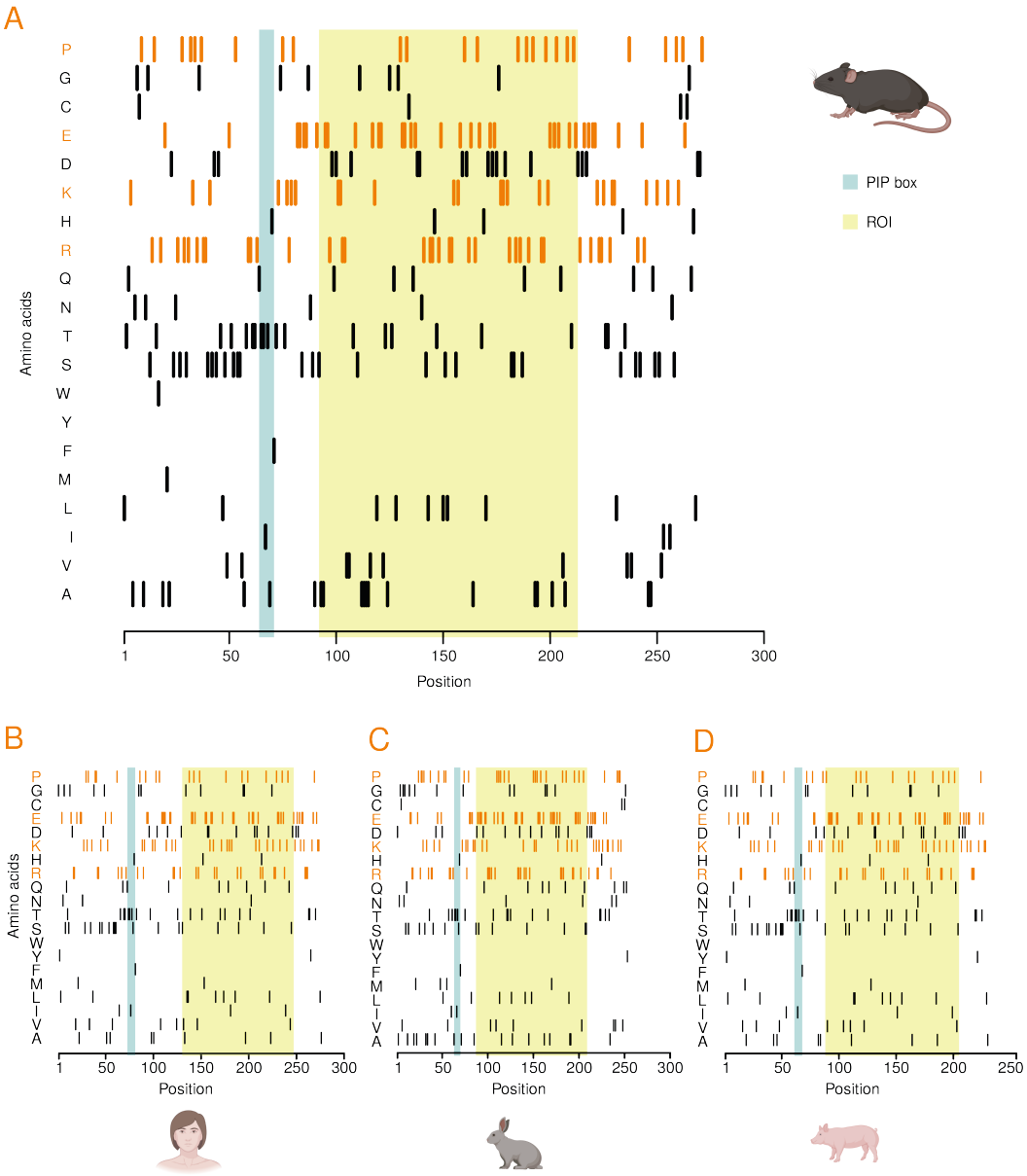

Fig. 2.13.: Barcode plot of mammalian DNMT1 IDRs. Shown is the position of each amino acid in the DNMT1 IDR in A mouse, B human, C rabbit, and D pig. The presence of an amino acid is indicated by a vertical bar along the length of the IDR on the x-axis. The frequent amino acids proline (P), glutamic acid (E), lysine (K), and arginine (R) are highlighted in orange. The PIP box and eutherian-specific ROI are shaded in dark and light green, respectively. Animal pictograms were created with `BioRender.com`.

3

Phase separation properties of the DNMT1 IDR

In recent years, IDRs have been increasingly studied in the context of LLPS [87]. It was found that proteins involved in the formation of biomolecular condensates often contain IDRs [85, 144]. In many cases, the IDRs have been shown to promote this phase-separation behavior [145]. As described in Section 1.2, the IDRs of heterochromatin components such as HP1α and MeCP2 play an important role for their condensation behavior [103, 104, 107]. The replication initiation factors Orc1, Cdc6, and Cdt1 in *Drosophila melongaster* also contain IDRs that facilitate their recruitment and function through LLPS [116].

DNMT1 is mainly active during S-phase at sites of DNA replication, where it interacts with other disordered proteins such as PCNA-associated factor 15 (P15PAF) [62]. After completion of S-phase, DNMT1 continues to be active, particularly in heterochromatin, which represents a phase-separated compartment [103]. Moreover, DNMT1 is recruited to sites of DNA repair damage, which may also represent phase-separated compartments [146–148]. Based on these observations, it is plausible that DNMT1 may undergo LLPS, for example, to facilitate its recruitment and access to these compartments. Thus, we set out to investigate whether the IDR of DNMT1 undergoes LLPS using an *in vitro* reconstitution assay of the IDR, complemented by opto-genetic analysis in live cells. In addition to studying the full-length IDR, we also tested how the presence and absence of the eutherian-specific ROI influences the phase-separation behavior.

3.1 *In vitro* phase separation behavior

LLPS behavior of proteins can be observed through turbidity measurements, since condensates scatter light. However, this assay does not allow to determine the size, shape and other material properties of the condensates. Thus, observation of condensate formation by light microscopy is a more common way to characterize phase separation of proteins or IDRs *in vitro* [80]. In this kind of cell-free reconstitution assay, purified proteins or IDRs dissolved in buffer with physiological salt concentrations are mixed with crowding agents such as polyethylene glycol (PEG) or lipids to simulate the crowded environment of cellular compartments (*e.g.*, cytoplasm, nucleus). The candidate proteins or IDRs are tagged with fluorescent proteins to allow their detection using fluorescent microscopy. The system is reconstituted at different protein concentrations to test for the concentration dependence of condensate formation. This method also allows for the observation of fusion and fission events between condensate as well as surface wetting, two properties characteristic of liquid-like materials [80].

Moreover, using image recognition software, the image pixels can be classified into object (dense condensate phase) and background (light phase), allowing for the quantification of the fluorescent signal intensity within the dense phase compared to the light phase. Using this data, the phase-shifted protein faction (PSF), *i.e.*, the relative amount of the protein present in the condensate phase, can be determined. To calculate the PSF, the intensity of the pixels of all the objects within one image are added and divided by the sum of signal intensity of all objects and the background. We then calculate the average PSF across all images taken for a given protein concentration:

$$\mathrm{PSF} = \frac{\sum \frac{\sum\sum \mathrm{intensity}_{\mathrm{object}}}{\sum\sum \mathrm{intensity}_{\mathrm{object}} + \sum \mathrm{intensity}_{\mathrm{background}}}}{\mathrm{n}}$$

Analogous to the determination of the IC_{50} of dose response curves, the saturation concentration c_{sat} can be determined from the saddle point of a sigmoidal fit curve to the data.

Thus, as a first test to determine whether the DNMT1 IDR has the potential to mediate LLPS, we expressed a recombinant protein consisting of a His-tag for column purification, mCherry as a fluorophore, the DNMT1 IDR as predicted by Metapredict and a solubility tag (Figure 3.1). We also studied the condensate formation behavior of the ROI alone. In order to investigate how the insertion of the ROI during the evolution of eutherian mammals could have influenced the phase-separation behavior of the shorter IDR present in all other animals, we designed a third construct consisting of the IDR lacking the ROI ("ΔROI"). All three constructs were purified from bacterial cultures using His-tag affinity chromatography and tested for their condensate formation capability at different concentrations ranging from 0.01 μM to 10 μM.

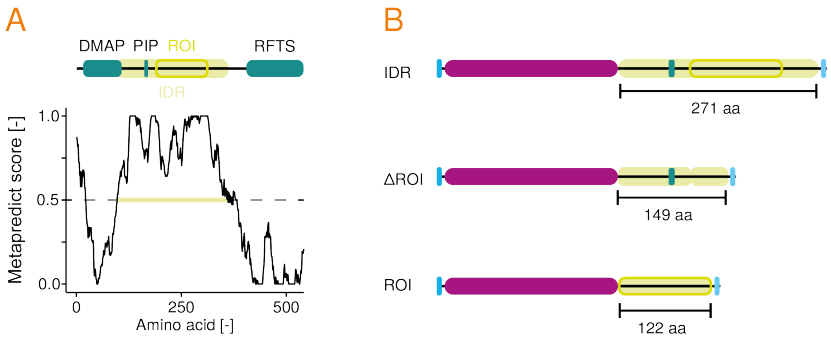

Fig. 3.1.: Construct design for the *in vitro* reconstitution assay. A IDR prediction in the murine DNMT1 N-terminus based on the Metapredict score. B Constructs tested in the *in vitro* reconstitution assay. "IDR" contains the full-length DNMT1 IDR, including PIP box and ROI. "ΔROI" contains the IDR, but lacks the ROI. "ROI" contains only the eutherian-specific ROI. Each construct is fused to mCherry at its N-terminus for visualization in the assay. At its N-terminus, mCherry is fused to a His-tag to enable column purification after recombinant expression in bacteria. At its C-terminus, each construct has a solubility tag to improve solubility during production and isolation of the constructs. DMAP: DNA methyltransferase 1-associated Protein; PIP: PCNA-interacting protein box; RFTS: replication foci targeting sequence; ROI: region of interest.

At 10 μM, we observed robust formation of spherical mCherry+ droplets for all three constructs (Figure 3.2 A). The condensates formed by the "IDR" and "ΔROI" constructs were similar in number, size and signal intensity, while "ROI" condensates were smaller, less numerous and less bright. For "IDR", "ΔROI" and "ROI", the number and size of condensates was dependent on the protein concentration. With decreasing protein concentration, condensates became smaller but more numerous, until no condensates could be detected anymore. "IDR" and "ΔROI" behaved similarly, with small but few condensates still present at 0.05 μM, while for the "ROI", no condensates could be observed below a concentration of 1 μM (Figure 3.2 A). Over time, condensates sank to the bottom, where they fused and wetted the surface of

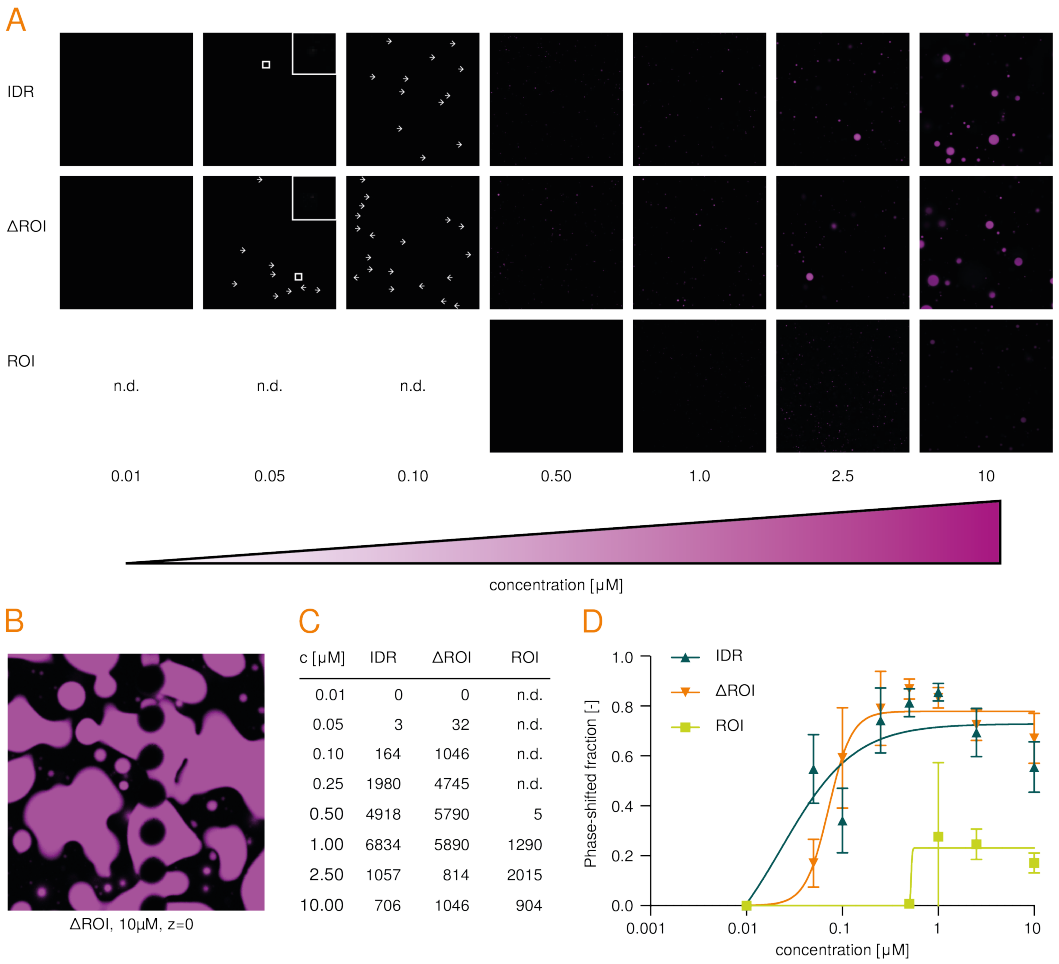

Fig. 3.2.: *In vitro* reconstitution assay of the DNMT1 IDR. A Representative examples of fluorescent microscopy images taken at z = 25 μm at different concentrations of the IDR, ΔROI and ROI constructs. No condensates can be observed in the images shown at the lowest concentration for each construct. At very low intensities, small condensates are marked with an arrow. B Fluorescent microscopy image of the ΔROI construct at 10 μM taken at z = 0 μm. C Number of identified and quantified objects across all images analyzed by construct and concentration. D Phase-shifted fraction of the proteins over protein concentration, calculated from the condensate and background intensities determined from the fluorescent microscopy images.

the object slide, demonstrating the liquid state of the condensates (Figure 3.2 B). Using a machine learning algorithm trained on a small subset of the imaging data, condensate and background pixels were readily classified (Figure 3.2 C).

Figure 3.2 D shows the PSF plotted over the increasing protein concentration of the respective construct. For the construct containing only the ROI, a maximum PSF of 0.28 is reached at 1 μM, while no condensates were observed at the next lower protein concentration. Thus, $c_{sat,ROI}$ likely lies between 1 μM and 0.5 μM. Of note,

the calculation of the PSF at 1 μM is potentially not precise, since the data here is distributed with a large standard deviation. For the "IDR" and "ΔROI" constructs, the maximum PSF is reached in the range of 1 μM to 0.5 μM. Below 2.5 μM, the PSF declines rapidly. For "ΔROI", the half-maximal PSF is reached at $c_{sat,\Delta ROI}$ = 0.014 μM. For the "IDR" construct, only three condensates were identified in the images at 0.05 μM. Thus, the calculation of the PSF at this concentration and the sigmoidal approximation are inaccurate for this construct at low concentrations, making it difficult to accurately determine $c_{sat,IDR}$. Nonetheless, the data indicates that the constructs "IDR" and "ΔROI" behave similarly in the context of this assay, and differences—if any—in their phase-separation behavior are small. Accordingly, we estimate c_{sat} of the DNMT1 IDR to be around 0.015 μM.

For all three constructs, a decline of the PSF can be observed at concentrations exceeding 1 μM. This effect is likely a technical artifact resulting from the limited signal detection range of the microscope. At low protein concentrations, the background signal is at the lower end of the detection limit and likely to be underestimated. At increasing protein concentrations, both background and object signal intensity increase. Relative to the object signal intensity, more of the background signal is detected, resulting in an overall lower PSF.

In summary, both the full-length IDR of DNMT1 as well as the IDR without the eutherian-specific region ("ΔROI") form liquid condensates *in vitro* and show similar phase-separation behavior. In contrast, the ROI alone forms condensates only at high concentrations and compared to the longer IDR constructs, only a small fraction demixes and accumulates in the condensate phase.

3.2 Opto-genetic analysis of LLPS by DNMT1

While *in vitro* reconstitution assays such as the assay described in Section 3.1 serve as a good indicator whether a protein or IDR in principle can undergo LLPS, the limited number of components within the mixture does not reflect the complex conditions present in live cells. Hence, we complemented our investigation with an orthogonal approach in tissue-cultured cells, the optoDroplet assay. This optogenetic tool for mammalian cells increases the multivalency of the tested IDR to

facilitate nucleation of condensates in their natural environment, which can be observed using live cell imaging [145]. In this system, the IDR of interest is fused to the photolyase homology region (PHR) of *Arabidopsis* CRYPTOCHROME (CRY2). CRY2 is a blue light receptor common to plants where it controls floral induction [149]. Upon reception of blue light the PHR drives oligomerization of the protein, leading to the formation of heterotypic nuclear speckles called "photobodies" that regulate the transcription of genes involved in plant morphogenesis through a mechanism that is not yet well understood (Figure 3.3 A) [150–152].

The optoDroplet system takes advantage of this light-inducible oligomerization to stimulate the accumulation of an IDR fused to the PHR with spatio-temporal control. Under the conditions of the optoDroplet assay, the PHR by itself does not form visible condensates. However, if the IDR stimulates LLPS, the PHR-IDR fusion protein undergoes a phase transition and forms liquid condensates once c_{sat} is reached through the initial nucleation. Through fusion with a fluorophore, *e.g.*, mCherry, the condensates can be observed using fluorescence microscopy. In principle, this phase transition is reversible when the cell is no longer exposed to blue light. However, prolonged exposure to blue light can lead to the maturation of gels into irreversible aggregates [145].

To investigate the condensation behavior of the DNMT1 IDR, the ROI alone and the IDR lacking the ROI ("ΔROI"), we transiently transfected HEK293T cells with the respective fusion construct (Figure 3.3 B). The sequences of the disordered regions are the same as used in the *in vitro* reconstitution assay described in Section 3.1. We chose HEK293T cells to carry out this assay since they are (i) easily cultivated, (ii) have large nuclei, (iii) grow in monolayers which makes single cells easy to observe using microscopy, and (iv) are very amiable to transfection.

Figure 3.4 shows HEK293T cells 48 h after transfection with the "opto-IDR", "opto-ΔROI" or "opto-ROI" construct. At $t = 1$, the cells were not yet exposed to blue light. Thus, these images display the native state of the assay. Subsequently, the imaged area was scanned with the 488 nm laser with an intensity of 0.5 in nine cycles, and an image to detect the mCherry signal was taken after each cycle. As expected, the fluorescent signal in cells transfected with "opto-mCherry" was diffuse within the nuclei and no formation of puncta was observed, since the PHR domain alone is not sufficient to form condensate under these settings. Similarly, no changes in the distribution of the nuclear mCherry signal were observed for "opto-ROI",

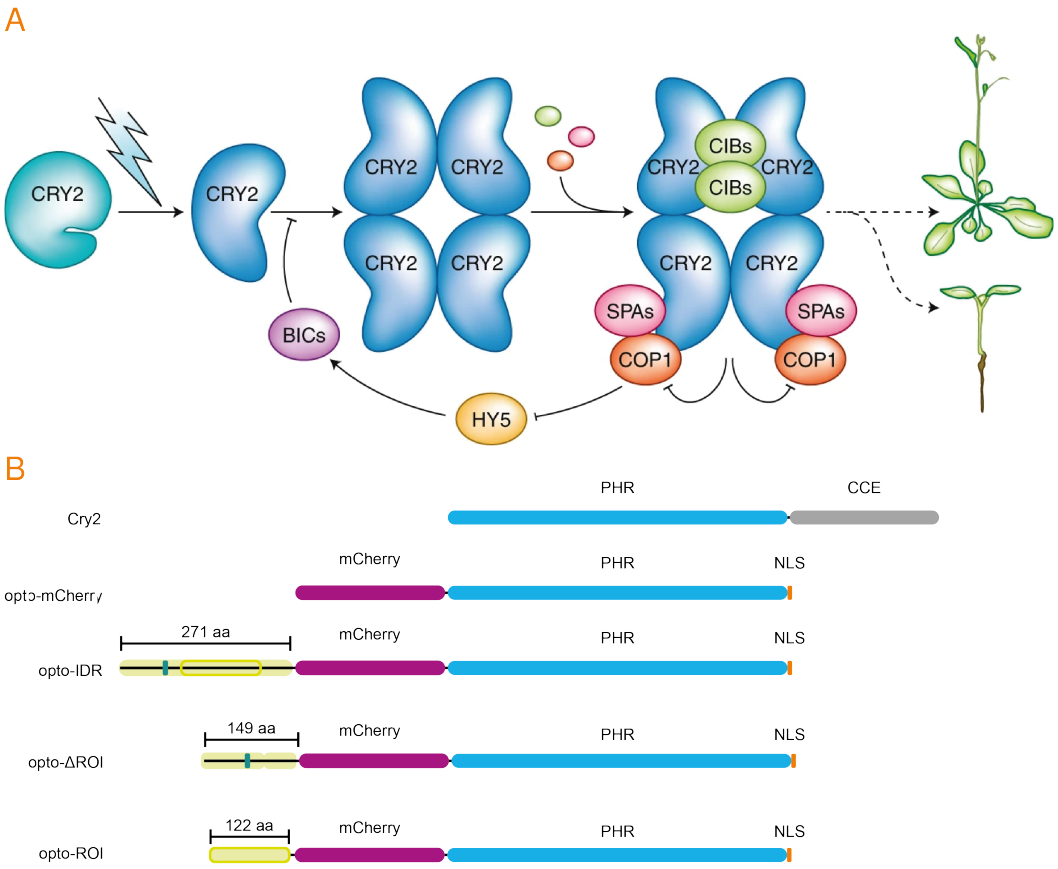

Fig. 3.3.: Adaptation of CRY2 for the optoDroplet assay. A Function of CRY2 in plant transcriptional regulation (from [153]). Monomeric CRY2 undergoes a conformational change upon blue light activation. In this photoactivated state, the protein forms oligomers which interact with several signaling proteins to alter gene expression. CRY2 oligomerization is controlled by a negative feedback loop. B Domain structure of *Arabidopsis thaliana* CRY2 and the derived photoactive fusion constructs between the CRY2 PHR domain (blue) and the IDRs investigate in our optoDroplet assay. The "opto-IDR" construct contains the full-length DNMT1 IDR (shaded in light green), including PIP box (dark green) and ROI (outlined in light green). The "opto-ΔROI" construct contains the DNMT1 IDR but lacks the ROI, while "opto-ROI" contains the eutherian-specific ROI only. Each construct is fused to mCherry (magenta) at its N-terminus for visualization by fluorescence microscopy. The construct "opto-mCherry" does not contain any an IDR sequence and serves as a negative control for condensate formation. Each construct has an NLS (orange) at its C-terminus to ensure localization of the fusion protein into the target compartment of DNMT1, the nucleus. blue-light inhibitor of CRYss (BICs): blue-light inhibitors of CRYs; CCE: C-terminal extension domain; CRY-interacting bHLHss (CIBs): CRY-interacting bHLHs; COP1: CONSTITUTIVE PHOTOMORPHOGENIC 1; CRY2: CRYPTOCHROME 2 of *Arabidopsis thaliana*; HY5: HYPOCOTYL 5; NLS: nuclear localization signal; PHR: photolyase homology region; ROI: region of interest; SPA: suppressors of PHYTOCRHOME A

indicating that the ROI alone is not sufficient to drive LLPS of the construct. In contrast, both "opto-IDR" and "opto-ΔROI" did not display a diffuse localization. Already in the unstimulated cells, both fusion proteins formed structures within the

nuclei. Upon exposure to blue light, the puncta appeared to grow slightly in size and signal intensity while the signal from the nucleoplasm decreased. Most if these changes occurred within one to two cycles, with little additional changes in signal intensity distribution in the remaining cycles. This indicates that protein from the nucleoplasm is quickly recruited into the pre-existing puncta.

The propensity of IDRs to undergo LLPS has been shown in depend on their length. The DNMT1 in its full-length consists of 271 amino acids. The ROI is comprised of 122 of these amino acids. Thus, the full-length IDR as well as the "ΔROI" construct are 2.2-fold and 1.2-fold longer than the ROI, respectively. To exclude the possibility that the length alone drives the formation of the observed structures in the context of the optoDroplet assay, we also tested two other constructs with lengths similar to that of the "opto-IDR" and "ΔROI", respectively: "opto-MTase" consists of the proximal half of the MTase domain (274 amino acids) and "opto-RFTS" contains the full RFTS (136 amino acids) of DNMT1. Both sequences are predicted to be highly ordered and fold into defined structures. Neither "opto-MTase" nor "opto-RFTS" formed nuclear speckles when stimulated with blue light over nine cycles (Figure 3.5). This confirms that the total length of the fusion protein does not underlie the formation of the structures formed by "opto-IDR" and "ΔROI". Interestingly, both "opto-MTase" and "opto-RFTS" fusion proteins were not only diffusely distributed throughout the nuclei of the cells but also within the cytoplasm. At high expression levels, "opto-MTase" formed cytoplasmic aggregates (Figure 3.5 A, inlet).

As shown by Shin et al. (2017), the intensity of the blue light can influence the speed at which the formation of the condensates in the context of the optoDroplet assay occurs [145]. Therefore, we attempted to improve the dynamic range of the assay and repeated the experiment with a ten-fold lower intensity of the 488 nm laser (0.05). Representative images from this second experiment are shown in Figure 3.6.

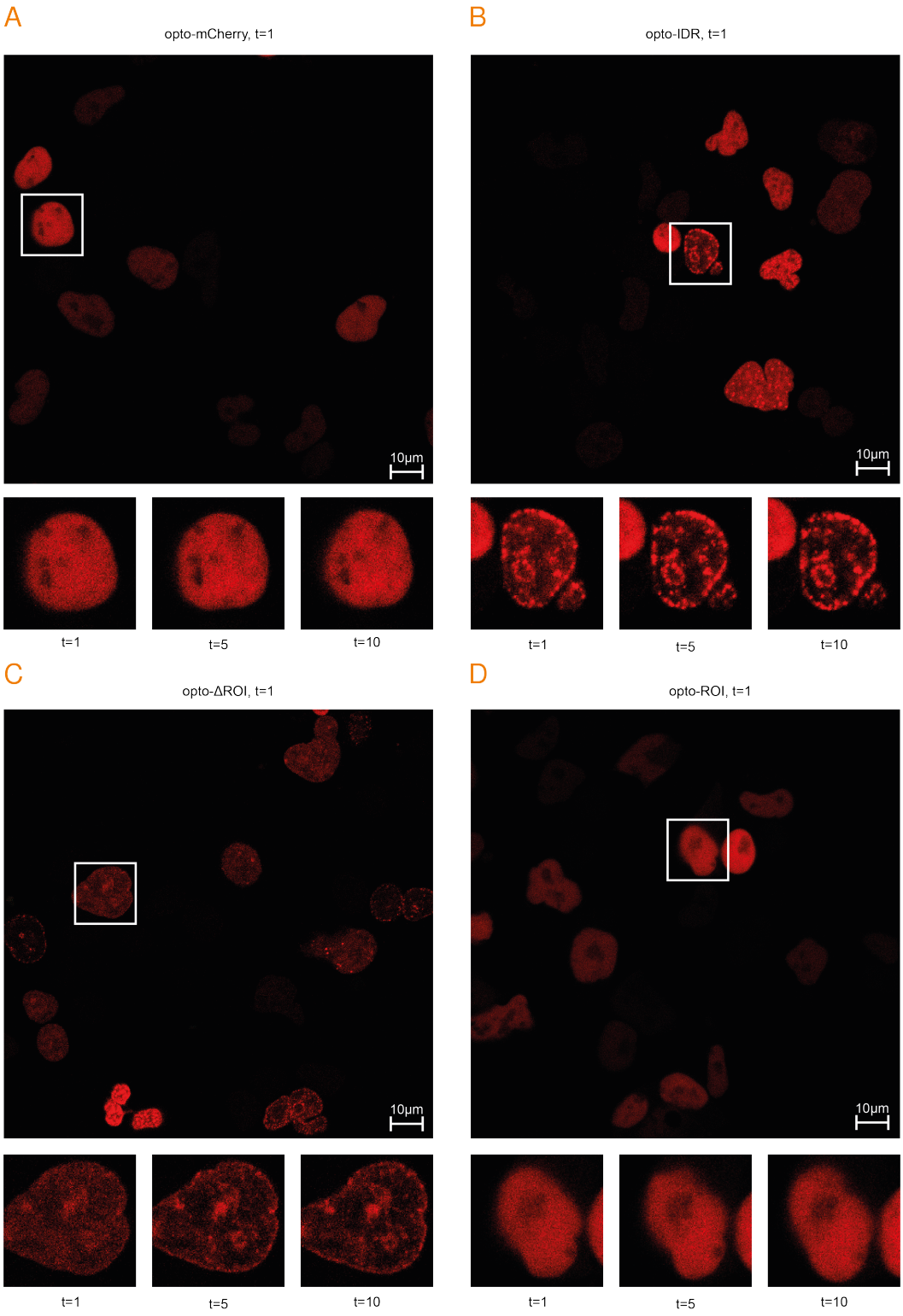

Fig. 3.4.: OptoDroplet assay of the DNMT1 IDR. HEK293T cells imaged with a 488 nm laser intensity of 0.5 48 h post-transfection with A "opto-mCherry", B "opto-IDR", C "opto-ΔROI", and D "opto-ROI", as detailed in Figure 3.3 B. At t = 1, the cells were not yet exposed to blue light. At t = 5 and t = 10, the cells have been scanned with the 488 nm laser for four and nine cycles, respectively.

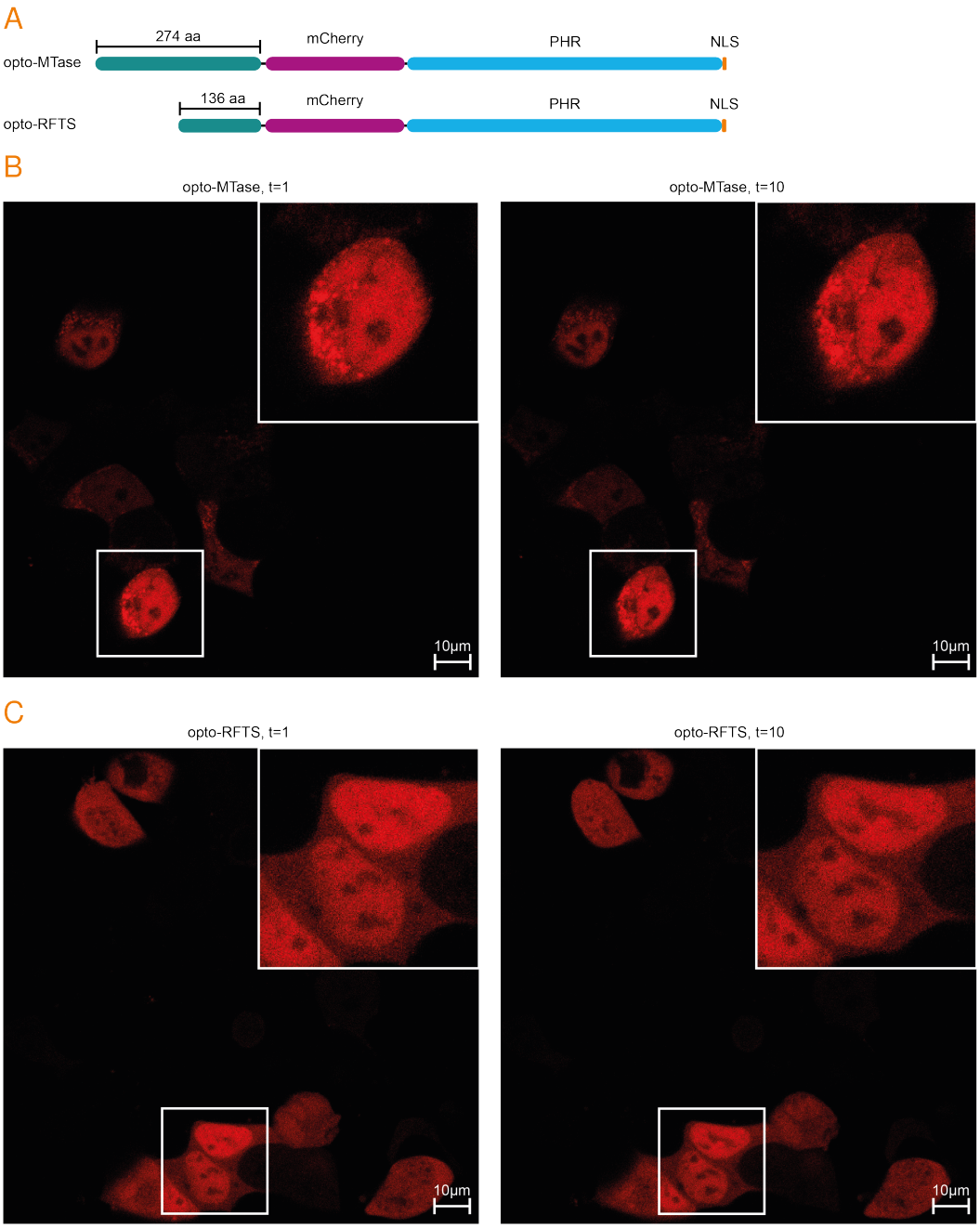

Fig. 3.5.: OptoDroplet assay with "opto-RFTS" and "opto-MTase". A The optoDroplet assay constructs "opto-MTase" and "opto-RFTS" consist of the proximal 247 amino acids of the catalytic MTase domain of murine DNMT1 and the 136 amino acid long RFTS domain, fused to mCherry-PHR-NLS. A HEK293T cells stimulated with a 488 nm laser intensity of 0.5 after t = 1 for nine cycles, 48 h post-transfection with "'opto-RFTS", or B "opto-MTase". MTase: methyltransferase domain; NLS: nuclear localization signal; PHR: photolyase homology region; RFTS: replication foci targeting sequence.

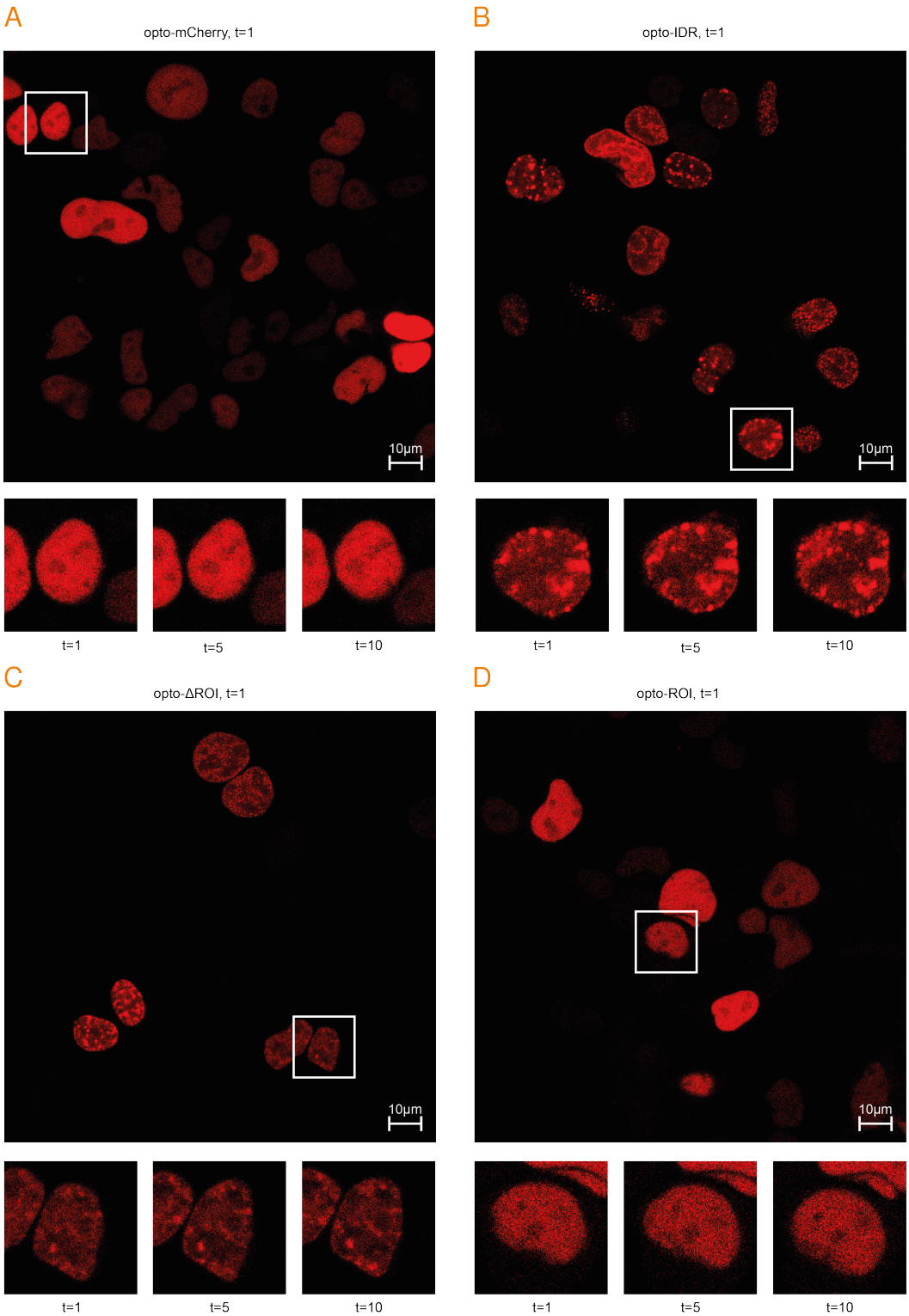

Fig. 3.6.: OptoDroplet assay of the DNMT1 IDR with reduced laser intensity. HEK293T cells imaged with a 488 nm laser intensity of 0.05 48 h post-transfection with A "opto-mCherry", B "opto-IDR", C "opto-ΔROI", and D "opto-ROI", as detailed in Figure 3.3 B. At t = 1, the cells were not yet exposed to blue light. At t = 5 and t = 10, the cells have been scanned with the 488 nm laser for four and nine cycles, respectively.

Using this very low laser intensity, the nucleoplasmic fraction of the mCherry signal from “opto-IDR” and “opto-ΔROI” decreased within two to three cycles while the pre-existing puncta slightly gained in signal intensity. Thus, we proceeded to use this laser setting to acquire data from a large number of cells for the quantification of our observations.

The behavior of “opto-IDR” and “opto-ΔROI” are very similar and there does not seem to be an observable difference by eye. To detect more subtle differences, we quantified the changes in condensate and background signal intensity. Using machine-learning based image recognition software, we detected each nucleus in every image. We then measured the mean signal intensity across all pixels within one nucleus as a measure for overall signal intensity and the standard deviation as a measure for the “granularity” of the cell (*i.e.*, the signal intensity of the condensates relative to the background nucleoplasm). After the detection of the cells, we first checked whether the numbers of detected nuclei and their signal intensity distribution for each construct were comparable. Shown in Figure 3.7 A is the mean intensity per cell by construct. Overall, we quantified between 211 and 298 cells per construct. The spread of the mean signal intensity is slightly larger for “opto-IDR” and “opto-mCherry”. However, overall the distribution of mean signal intensity and its median are similar between the four conditions.

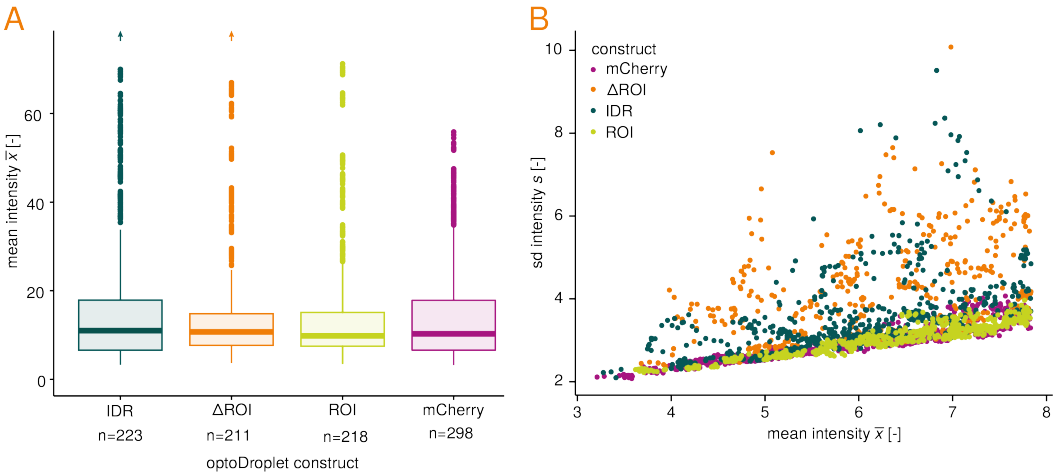

Fig. 3.7.: Signal intensity distribution of DNMT1 IDR optoDroplet constructs. A Distribution of mean signal intensity per cell quantified from images of the optoDroplet assay with “opto-IDR”, “opto-ΔROI”, “opto-ROI” and “opto-mCherry” carried out with a 488 nm laser intensity of 0.05, across all time points. The box includes the 2nd and 3rd quartile, the line shows the median. Some outliers for “opto-IDR” and “opto-ΔROI” were cropped, as indicated by the arrows. The number of quantified cells is shown below the y-axis. B Standard deviation of the signal intensity over mean signal intensity per cell for all time points by construct.

Cells with an evenly diffuse signal are expected to have a low standard deviation of signal intensity, while cells with bright condensates but overall darker nucleoplasm have a high standard deviation. However, the absolute value of the standard deviation alone is not sufficient to describe the difference between diffuse and structured signaling patterns, since cells with an overall high signal intensity also tend to have a larger signal intensity standard deviation (Figure 3.7 B). Thus, we normalized the standard deviation of the signal intensity $s_{intensity}$ by the mean signal intensity $\bar{x}_{intensity}$ to derive the mean normalized standard deviation $sd_{mean\ norm}$ as a measure to compare the "granularity" of the nuclei:

$$sd_{mean\ norm} = \frac{s_{intensity}}{\bar{x}_{intensity}}$$

The $sd_{mean\ norm}$ is calculated on a per cell basis. We then pooled all images by construct and time point and plotted the median $sd_{mean\ norm}$ over the ten cycles of the optoDroplet assay. Figure 3.8 A shows the development of the median $sd_{mean\ norm}$ over time for all cells, grouped by optoDroplet construct. As expected, cells with a diffuse signal (*e.g.*, cells transfected with "opto-mCherry") start with a low $sd_{mean\ norm}$. Since no condensates are formed in these cells, the $sd_{mean\ norm}$ stays constant over time. In contrast, cells transfected with either "opto-IDR" or "opto-ΔROI" already form condensates prior to the exposure to the 488 nm laser after $t = 1$. This is reflected by the high $sd_{mean\ norm}$ at $t = 1$, as compared to "opto-mCherry" and "opto-ROI". As we observed by eye, the signal intensity in the nucleoplasm decreases while the signal intensity stemming from the condensates increases, leading to an increase in the $sd_{mean\ norm}$ over time Figure 3.8 A.

However, our data also includes very bright cells where almost all of the fusion protein is located within condensates from the beginning on, so that the signal distribution does not appear to change over time. To confirm this using our data, we split the cells into four groups according to the quartiles for the mean signal intensity distribution and plotted the median $sd_{mean\ norm}$ over time for each of the quartiles separately (Figure 3.8 B-E). For "opto-ROI" and "opto-mCherry", the signal distribution did not change over time, independent of the mean signal intensity. As expected, very bright cells (4^{th} quartile, Figure 3.8 E) transfected with "opto-IDR" and "opto-ΔROI" displayed a constantly high $sd_{mean\ norm}$. In contrast, and as reflected by the trend in the overall population, the median $sd_{mean\ norm}$ for

"opto-IDR" and "opto-ΔROI" in the lower three quartiles increases over time as recruitment of fusion protein into pre-formed condensates is stimulated (Figure 3.8 B-D).

Interestingly, in the lower three quartiles "opto-ΔROI" cells show an overall higher median $sd_{mean\ norm}$ than "opto-IDR" cells for all time points. This trend is particularly evident for dim cells in the 1st quartile. This could imply that "opto-ΔROI" forms condensates more readily, and attracts free fusion protein from the nucleoplasm into condensates at low protein concentrations (which are reflected by an overall low mean signal intensity in the 1st quartile) more efficiently than "opto-IDR". However, the interquartile range (shaded area) of "opto-IDR" and "opto-ΔROI" well overlap, indicating that the two data series are very similar.

The structures formed by "opto-IDR" and "opto-ΔROI" are not randomly distributed throughout the nuclei, but manifest in a variety of distinct patterns (Figure 3.9). In few cells, relatively large puncta are evenly distributed throughout the nucleus. More commonly, smaller condensates are located within the nucleoplasm. In addition, when smaller condensates are formed, these are often additionally located along the nuclear lamina, and sometimes surround other spherical intranuclear structures. Of note, cells adjacent to each other often display similar distributions of puncta.

To test whether this behavior is specific to optoDroplet constructs of mouse DNMT1, we cloned further constructs with the IDRs of human DNMT1, murine DNMT3A and DNMT3B as well as murine P15PAF, a highly disordered protein that associates with the replication fork where it interacts with and recruits DNMT1 in early S-phase [62] (Figures 3.10-3.12). The human versions of "opto-IDR" ("opto-hs_IDR"), "opto-ΔROI" ("opto-hs_ΔROI") and "opto-ROI" ("opto-hs_ROI") behaved analogous to their murine counterparts (Figure 3.10), with condensates formed by "opto-hs_IDR" and "opto-hs_ΔROI", but not "opto-hs_ROI". Overall, the same phenotypes were observed; however, fewer cells exhibited only small droplets that were not associated with any nuclear structure. Of note, for this experiment, the transfection was generally less efficient than with the murine sequences. Thus, there were too few cells with a high enough signal intensity for quantification.

For "opto-Dnmt3a" and "opto-Dnmt3b", only few cells expressing high levels of the respective fusion protein had pre-formed speckles (Figure 3.11 B-C, left panels). Ex-

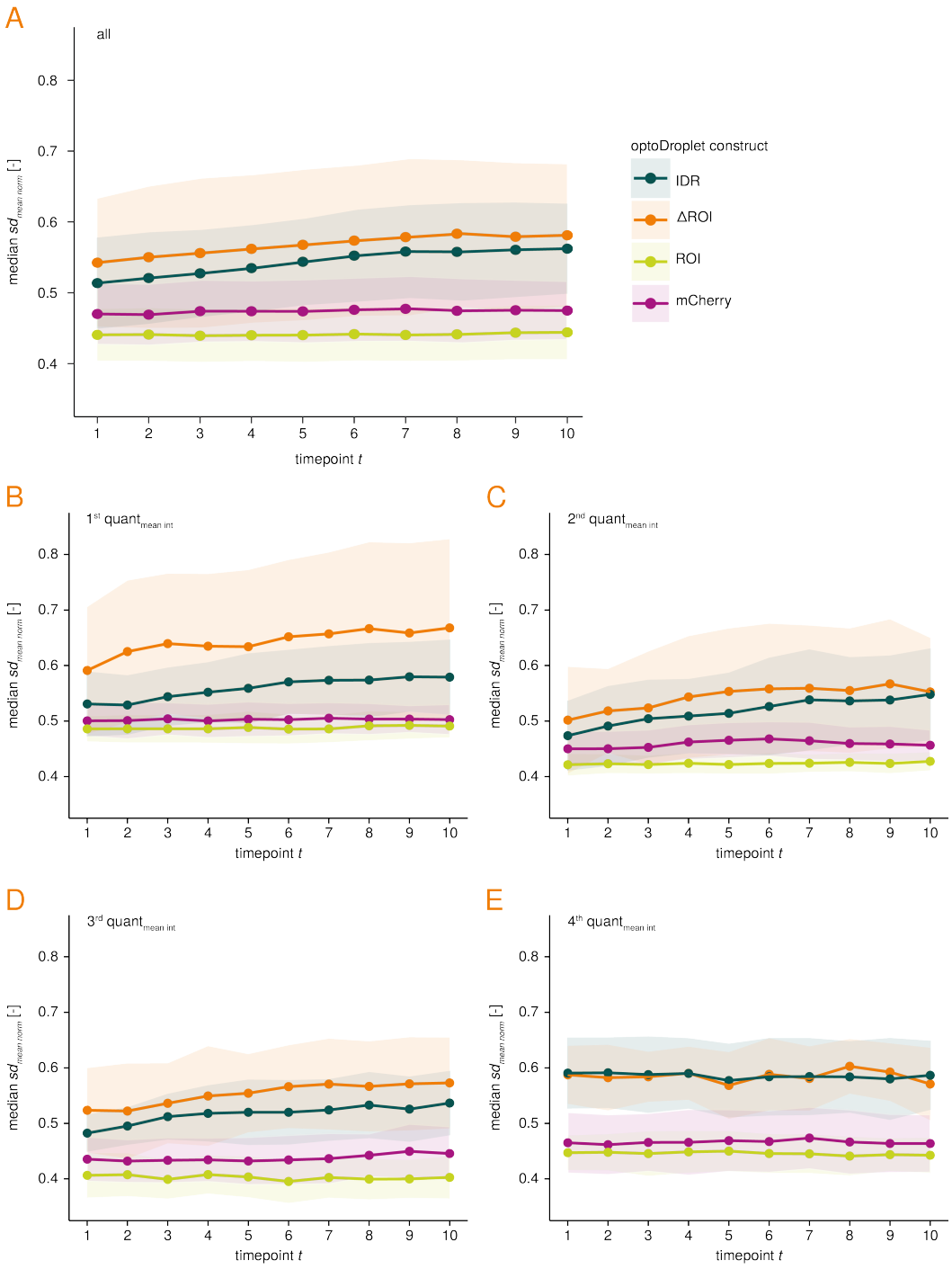

Fig. 3.8.: Signal quantification in the DNMT1 IDR optoDroplet assay. A Median of the signal intensity standard deviation, normalized by the mean signal intensity, over time by optoDroplet construct. Shown is the data for all detected cells. The shaded area indicates the interquartile range by construct. B-E Same data as shown A, split by quartiles of the mean signal intensity. For example, B shows data for the dimmest 25 % of the cells, while the brightest 25 % cells are shown in E.

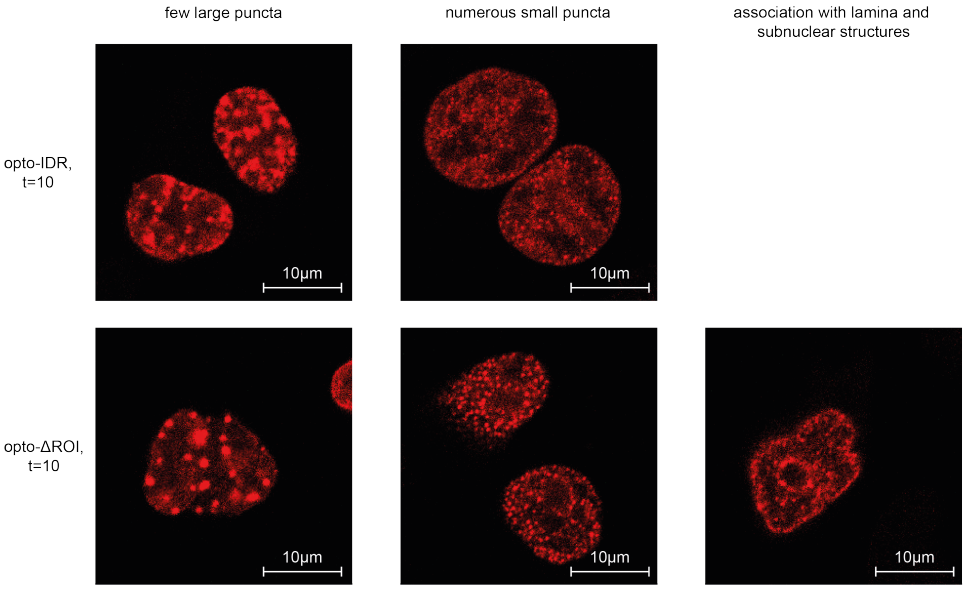

Fig. 3.9.: Nuclear patterns formed by "opto-IDR" and "opto-Δ ROI". HEK293T cells were stimulated with a 488 nm laser intensity of 0.05 over nine cycles and imaged 48 h after transfection with either "opto-IDR" or "opto-ΔROI". With either construct, several patterns emerge. Some cells form fewer large puncta distributed throughout the nucleus. In most cells, the fusion protein accumulates in numerous smaller speckles. These are either even distributed throughout the nucleus or predominantly associate with the lamina and subnuclear structures, resulting in ring shaped patterns.

posure to the 488 nm laser at an intensity of 0.05 stimulated droplet formation in "opto-Dnmt3a" and "opto-Dnmt3b" cells (Figure 3.11 B-C, right panels). However, the extent of droplet formation and the intranuclear distribution of either construct was distinct from the patterns observed with the DNMT1 "opto-IDR" or "opto-ΔROI" (Figure 3.4-3.6). For "opto-Dnmt3a", two different phenotypes were observed. In some cells, the fusion protein was diffusely distributed throughout the nucleus, and one or two large foci appeared upon blue light exposure (Figure 3.11, arrows in left panels). In most highly expressing cells, condensates appeared throughout the nucleus and stayed clear of the nucleoli (Figure 3.9 B, arrows in right panels). For "opto-Dnmt3b", condensates formed throughout the nucleus, with some accumulation around the nucleoli which themselves remained free of mCherry signal (Figure 3.9 C, arrows in right panels). In contrast to DNMT1 "opto-IDR" or "opto-ΔROI", a significant portion of "opto-Dnmt3a" and "opto-Dnmt3b" remained diffusely distributed throughout the nucleus, even after nine cycles of blue light exposure.

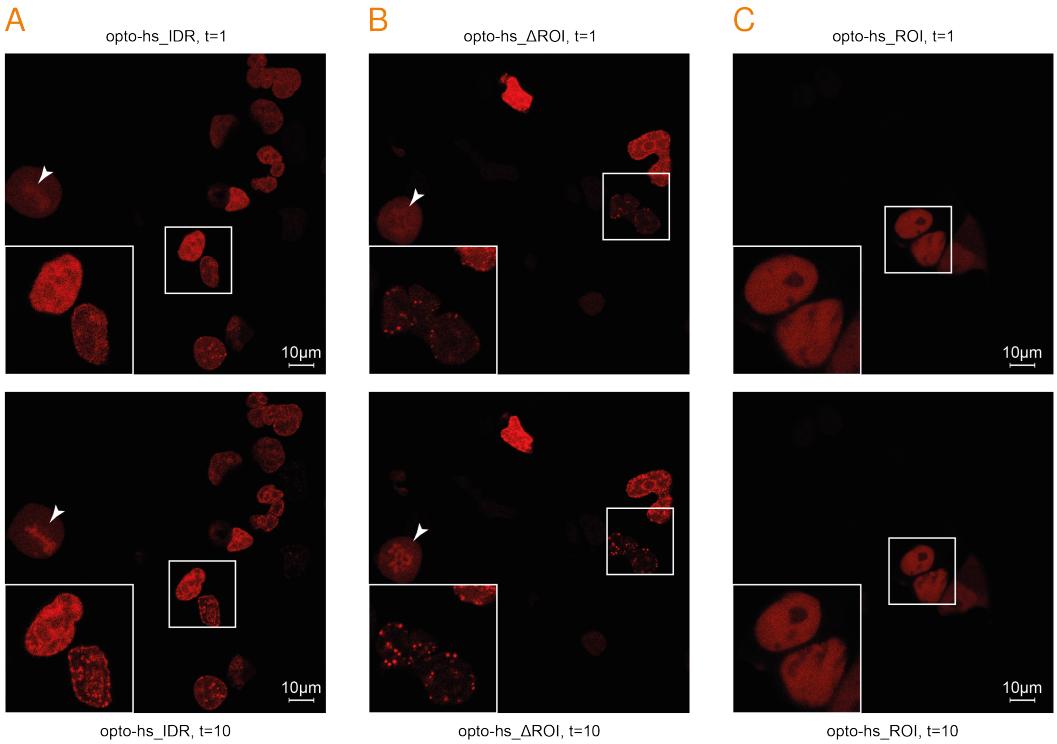

Fig. 3.10.: OptoDroplet assay with human "opto-IDR", "opto-ΔROI" and "opto-ROI". HEK293T cells imaged with a 488 nm laser intensity of 0.5 48 h post-transfection with A "opto-hs_IDR", B "opto-hs_ΔROI", and C "opto-hs_ROI". At t = 1, the cells were not yet exposed to blue light. At t = 10, the cells have been scanned with the 488 nm laser for nine cycles.

Cells transfected with "opto-p15paf" also exhibited two different phenotypes (Figure 3.12). In few cells, the mCherry signal was diffuse in the nucleus (Figure 3.12 B-C). Already after the first cycle of blue light exposure, many small droplets and few large patches appeared. Most other cells displayed pre-formed droplets throughout the nucleus, with a tendency to associate with the nuclear lamina (Figure 3.12 B, right panels). Unlike all other tested optoDroplet constructs, "opto-p15paf" also entered the nucleoli (Figure 3.12 B, right panels, arrowheads). Except for the occupancy of nucleoli, the distribution of "opto-p15paf" was similar to that of DNMT1 "opto-IDR" or "opto-ΔROI".

If DNMT1 and P15PAF undergo LLPS at replication foci, both IDRs should be able to localize to the same condensates, since DNMT1 and P15PAF were shown to interact in early S-phase [62]. We therefore asked whether "opto-IDR" or "opto-ΔROI" would co-localize with "opto-p15paf" when co-transfected. In order to observe two different IDRs within one optoDroplet assay, we exchanged mCherry with GFP

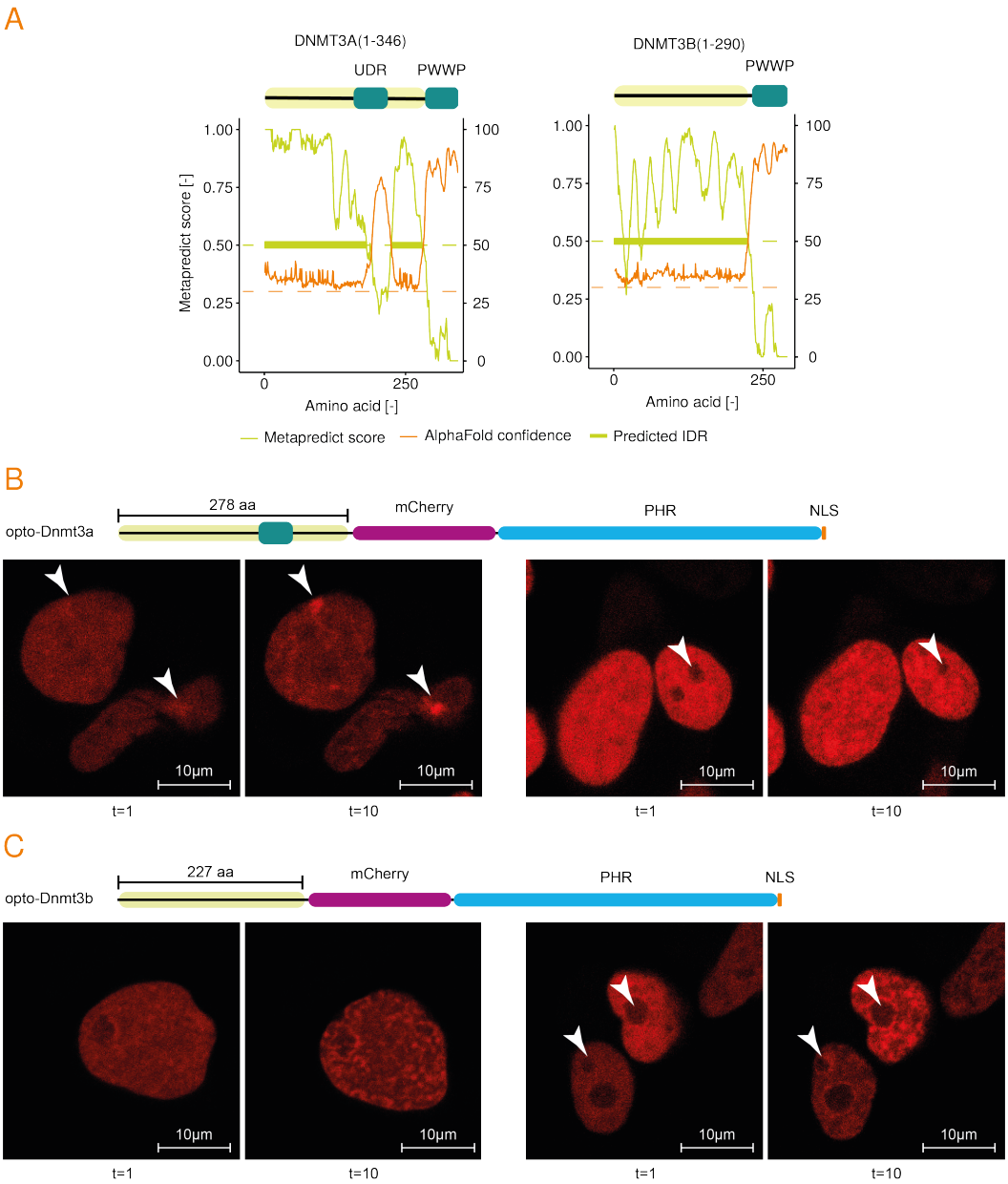

Fig. 3.11.: OptoDroplet assay with "opto-Dnmt3a" and "opto-Dnmt3b". A Metapredict scores, AlphaFold2 pLDDT confidence scores and predicted IDRs within the N-termini of mouse DNMT3A1 and DNMT3B. B-C Nuclear patterns formed in HEK293T cells 48 h after transfection with "opto-Dnmt3a" and "opto-Dnmt3b", respectively. After the first image (t = 1), the cells were stimulated with a 488 nm laser intensity of 0.05 over nine cycles. PWWP: Pro-Trp-Trp-Pro domain; UDR: ubiquitin-dependent recruitment region.

in "opto-p15paf". We then co-transfected the cells with GFP-tagged "opto-p15paf" and mCherry-tagged "opto-p15paf", "opto-IDR" or "opto-ΔROI" at a molar ratio of 1:1 (Figures 3.12 C, 3.13). Exchanging mCherry for GFP in the optoDroplet construct did not influence the intranuclear distribution of "opto-p15paf", and GFP-

tagged “opto-p15paf” retained its ability to accumulate in puncta upon blue light stimulation (Figure 3.12 C). When co-transfected with mCherry-tagged “opto-IDR” or “opto-ΔROI”, each construct behaved as observed when transfected alone (Figure 3.13). In most cases where a cell was transfected with both constructs, GFP+ and mCherry+ puncta overlapped. As seen with “opto-IDR” or “opto-ΔROI”, and with some tendency also with “opto-p15paf”, the condensates had a preference to localize to the nuclear lamina. An exception are “opto-ΔROI” condensates on the surface of subnuclear structures (ring shaped formations), which contain almost no GFP-tagged “opto-p15paf” in contrast to other speckles within the same nucleus (*e.g.*, Figure 3.13). This is consistent with the observation that the formation of ring shaped patterns is not a predominant phenotype of “opto-p15paf” (Figure 3.12 B-C).

Interestingly, while studying the behavior of the different IDRs in the context of the optoDroplet assay, we noticed distinct patterns in cells undergoing metaphase. Depending on the construct, the fusion protein either localized to the condensed chromosome, was repelled by them, or was not influenced by their presence (Figure 3.14). The constructs “opto-mCherry” and “opto-Dnmt3b” were repelled by the condensed chromatin (Figure 3.14 A, D), while “opto-IDR” and “opto-ΔROI” stained metaphase chromosomes (Figure 3.14 B, C, also highlighted in Figure 3.10 A, B). The localization of “opto-p15paf” was not influenced by the presence of metaphase chromosomes, which were stained by “opto-IDR” in the same cell (Figure 3.14 B). No cells in metaphase were observed for the assays with “opto-ROI” and “opto-Dnmt3a”.

Given the similar behavior of “opto-IDR” and “opto-ΔROI” constructs, which part of the DNMT1 IDR other than the ROI is responsible for the strong drive of the DNMT1 IDR to form condensates, and what dictates their distribution? In order to address this question, we divided the DNMT1 IDR conserved in all animal species into a proximal part (cIDR1, upstream of the PIP box) and a distal part (cIDR2, downstream of the ROI), and studied the performance of these parts separately using the optoDroplet assay. We also considered whether the intranuclear distribution of the “opto-IDR” and “opto-ΔROI” may result from binding of the constructs to endogenous PCNA, since both constructs harbor the PIP box. To test whether this is the case, we designed the ΔPIP mutant where the mutation of Q in the PIP box motif (QTTITSHF) to E abolishes the ability to bind PCNA (Figure 3.15) [57]. Compared to “opto-IDR” and “opto-ΔROI”, “opto-cIDR1” and “opto-cIDR2”

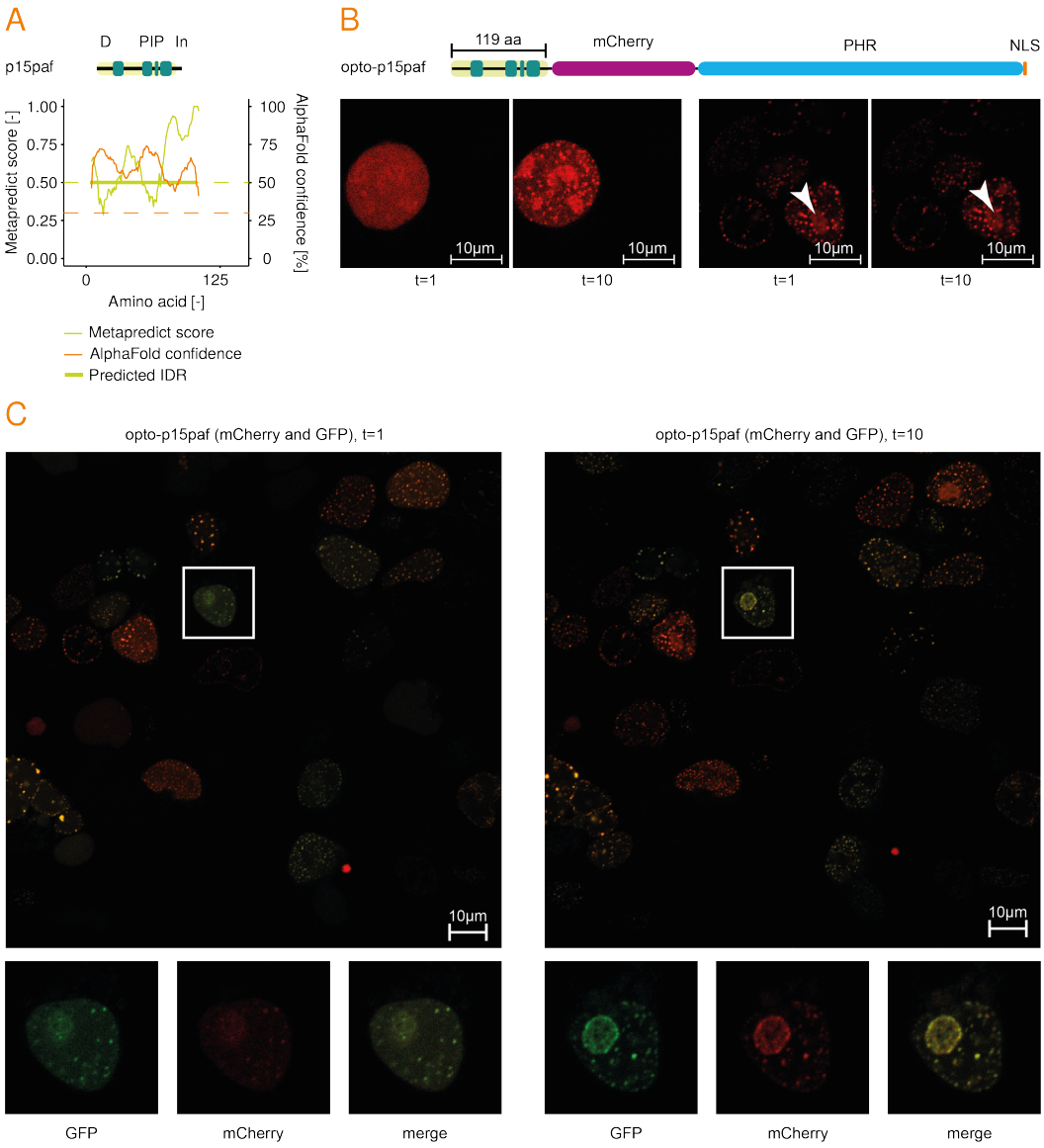

Fig. 3.12.: Dual optoDroplet assay with "opto-p15paf". A Metapredict scores, AlphaFold2 pLDDT confidence scores and predicted IDR of mouse P15PAF. B-D Nuclear patterns formed in HEK293T cells 48 h after transfection with "opto-p15paf", either fused to mCherry or to GFP. After the first image (t = 1), the cells were stimulated with a 488 nm laser intensity of 0.05 over nine cycles. D: D-box; In: initiation motif; PIP: PCNA-interacting protein box.

displayed a much weaker tendency to form condensates under the same conditions (Figure 3.15 B-C). Without blue light stimulation (t = 1), the mCherry signal was diffusely distributed throughout the nuclei (Figure 3.15 B-C, top panels). In most weakly expressing cells, no condensate formation was observed upon exposure to blue light (Figure 3.15 B-C, bottom panels). Only strongly fluorescent cells reacted and some condensation was observed. For both "opto-cIDR1" and "opto-

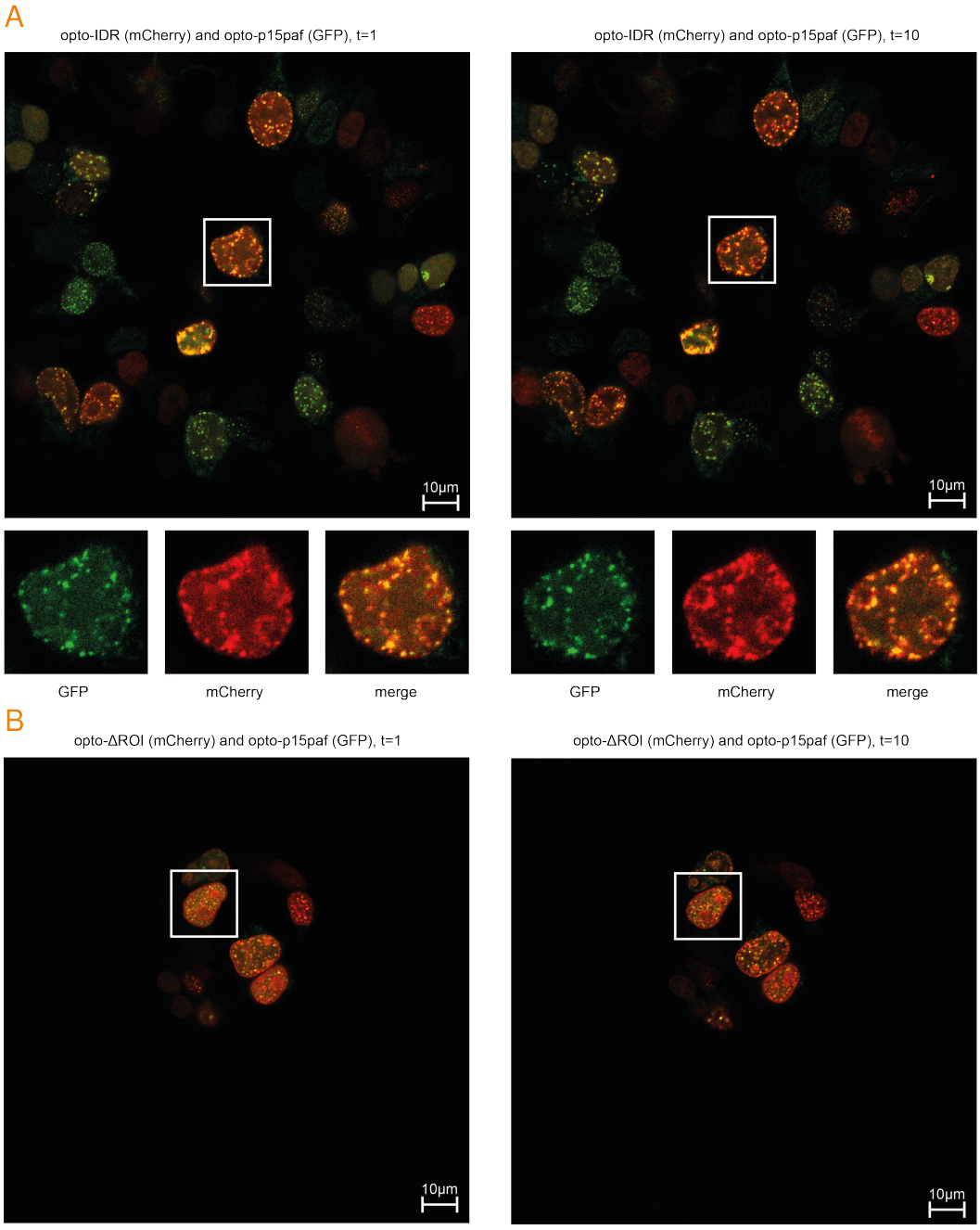

Fig. 3.13.: Dual optoDroplet assay with "opto-p15paf" and "'opto-IDR" or "opto-ΔROI". Nuclear patterns formed in HEK293T cells 48 h after transfection with "opto-p15paf" (GFP) and A "opto-IDR" (mCherry) or B "opto-ΔROI" (mCherry). After the first image (t = 1), the cells were stimulated with a 488 nm laser intensity of 0.05 over nine cycles.

cIDR2", subnuclear patterning was observed to some extent when condensates formed; however, most of the signal remained diffuse in the nucleus. Few ring shaped features were observed in cells transfected with "opto-cIDR2" (Figure 3.15

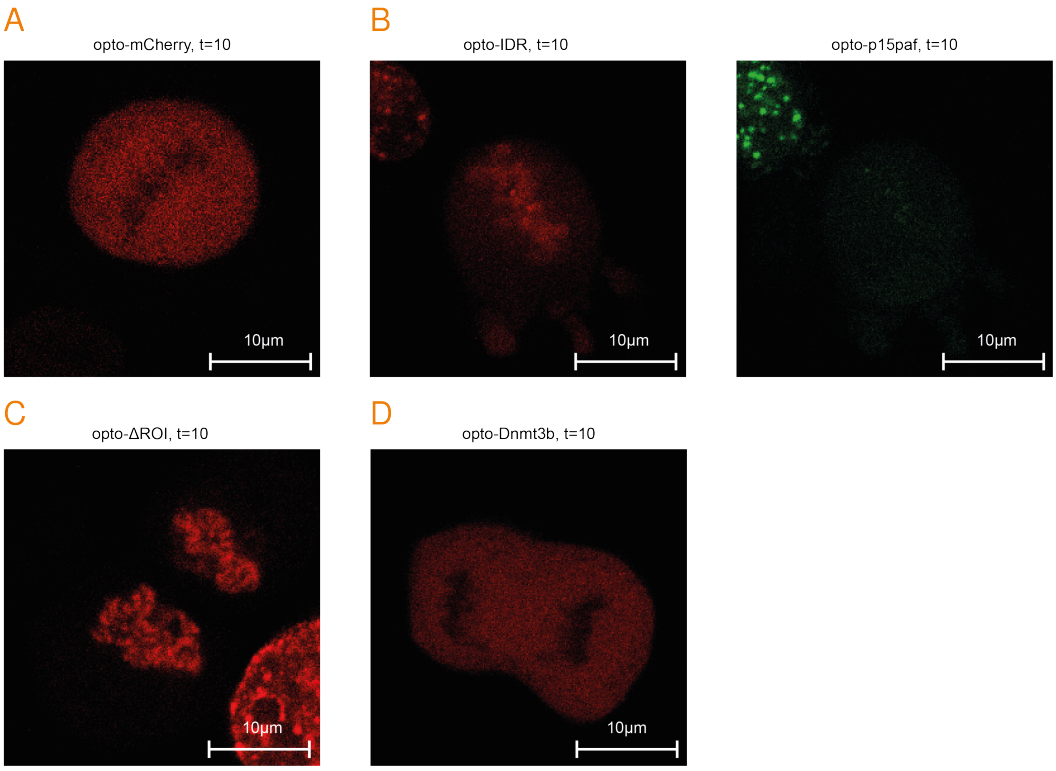

Fig. 3.14.: OptoDroplet construct behavior in metaphase. Nuclear patterns formed in metaphase HEK293T cells after nine cycles of exposure to the 488 nm laser with an intensity of 0.05. The cells were imaged 48 h after transfection with A "opto-mCherry", B "opto-IDR" (mCherry) and "opto-p15paf" (GFP), C "opto-ΔROI", and D "opto-Dnmt3b".

C, bottom panel inlet). In contrast, "opto-ΔPIP" behaved similar to "opto-IDR" and "opto-ΔROI". Cells already pre-formed condensates without exposure to blue light, and these condensates gained in intensity, while the signal within the nucleoplasm decreased (Figure 3.15 D). For the "opto-ΔPIP" construct, we also observed a tendency to localize to the nuclear lamina and subnuclear structures, or to accumulate in large foci. Compared to transfections with "opto-IDR" and "opto-ΔROI", cells with many small puncta were rare.

The formation of condensates by "opto-IDR", "opto-ΔROI" and "opto-ΔPIP" even without exposure to blue light indicates that the DNMT1 IDR—both in eutherian-mammals with the ROI and in other animals lacking the ROI—possesses a strong tendency to undergo LLPS. In order to confirm that this characteristic depends on the IDR alone, we removed the light-responsive PHR domain from "opto-IDR" and "opto-ΔROI", resulting in the new constructs "opto-IDR-no Cry2" and "opto-ΔROI-no Cry2", respectively. We then transfected HEK293T cells with either constructs and imaged the cells 48 h post-transfection. In this case, we did not apply the

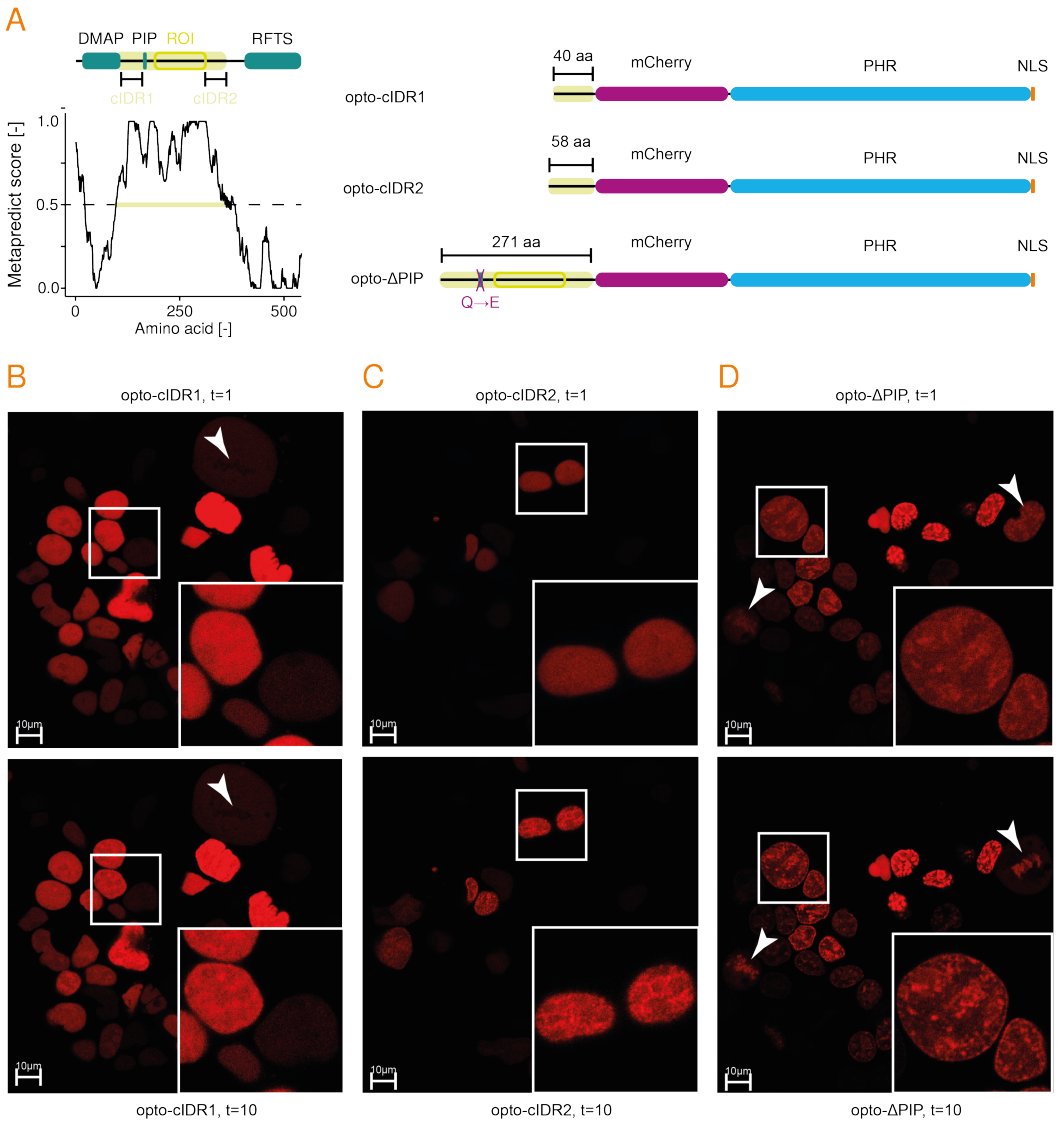

Fig. 3.15.: Dissection of the DNMT1 IDR using the optoDroplet assay. A Metapredict score for the DNMT1 N-terminus. The proximal and distal conserved IDR (cIDR1 and -2, respectively) are indicated. The right side shows the three new constructs based on the division of the IDR as well as the ΔPIP mutant. B Nuclear patterns formed in HEK293T cells before and after nine cycles of exposure to the 488 nm laser with an intensity of 0.05. The cells were imaged 48 h after transfection with "opto-cIDR1". B In a second experiment, cells were transfected with the "opto-cIDR2.". C In this experiment, cells were transfected with the ΔPIP mutant version of "opto-IDR", "opto-ΔPIP", which cannot bind PCNA.

488 nm laser to the cells, since the construct lacks the light-sensitive domain and cannot respond to blue light. On the one hand, as shown in Figure 3.16, both "opto-IDR-no Cry2" and "opto-ΔROI-no Cry2" formed puncta within the nuclei of the cells that preferentially associate with the nuclear lamina. As we also observed with "opto-IDR" and "opto-ΔROI", some cells displayed large speckles dis-

tributed throughout the nucleus (*e.g.*, Figure 3.16 B, bottom left panel). As their PHR-containing counterparts, "opto-IDR-no Cry2" and "opto-ΔROI-no Cry2" also accumulated on metaphase chromosomes (data not shown). On the other hand, "opto-IDR-no Cry2" and "opto-ΔROI-no Cry2" differed from "opto-IDR" and "opto-ΔROI" in that they also occupied nucleoli and very few, if any, cells had numerous small condensates distributed throughout the nucleoplasm, a phenotype commonly observed in "opto-IDR" and "opto-ΔROI" cells.

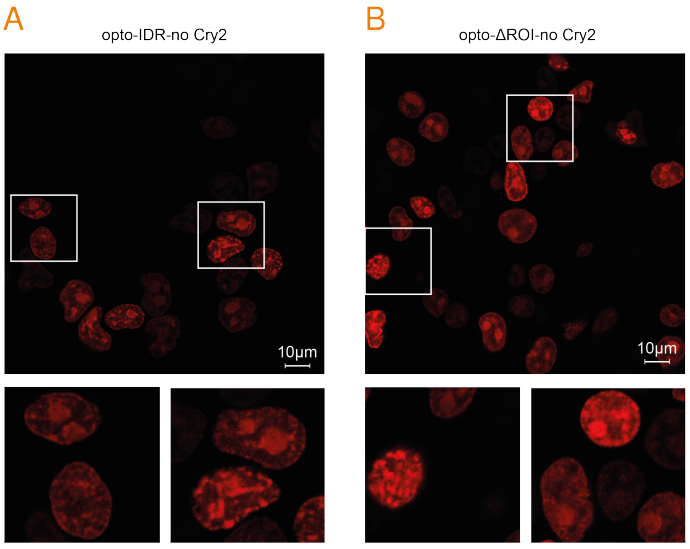

Fig. 3.16.: PHR-less "opto-IDR" and "opto-ΔROI". HEK293T cells were transfected with A "opto-IDR-no Cry2" or B "opto-ΔROI-no Cry2", lacking the PHR domain. Cells were imaged 48 h post-transfection.

To summarize, all tested optoDroplet constructs exhibit construct-specific behavior and formed condensates to different extents and at different locations within the nucleus. The behavior of mouse and human DNMT1 IDR and derivative constructs was comparable. Interestingly, "opto-IDR" or "opto-ΔROI" behaved similarly: they exhibited the same phenotypes with respect to localization and both had the ability to co-occupy condensates with "opto-p15paf". However, condensates localized around nuclear substructures were predominantly formed by "opto-IDR" or "opto-ΔROI" alone. Because adjacent cells often displayed similar phenotypes, we hypothesize that cell cycle dependent presence of endogenous proteins influences the distribution of DNMT1 and P15PAF condensates.

Further splitting the "opto-ΔROI" into a proximal conserved part ("opto-cIDR1") and a distal conserved part ("opto-cIDR2") effectively decreased the ability of the DNMT1 IDR to respond to the blue light and to form condensates. As a whole,

the DNMT1 IDR has a much stronger tendency to form condensates than the IDRs of DNMT3A and DNMT3B, and does so even without blue light stimulation or the presence of the CRY2 PHR domain within the context of the optoDroplet assay.

4

A platform to investigate methylation phenotypes of DNMT1 mutants

After having tested the ability of the IDR to undergo LLPS by itself *in vitro* and *in vivo*, we sought to study the role of the IDR in the context of the full-length protein and its primary function, namely maintenance methylation. To determine the relevance of the IDR for the maintenance activity of DNMT1, we envisioned a platform with which we can study loss-of-function phenotypes of various DNMT1 mutants in cells starting with a fully methylated genome.

Our platform is based on a mouse embryonic stem cell (mESC) line in which the endogenous DNMT1 wt protein can be swiftly depleted. Here, we make use of a degron-tag that leads to the depletion of DNMT1 at the protein level. After the derivation and thorough characterization of the degron cell line (Section 4.1), we set out to find a strategy to efficiently and reproducibly express *Dnmt1* wt and mutant transgenes at wt protein levels. However, the large size of DNMT1 proved extremely challenging for over-expression. Thus, we first tested several approaches to ectopically express *Dnmt1* transgenes (Section 4.2) and finally optimized expression through an improved derivation strategy and enhanced construct design (Section 4.3).

4.1 Inducible DNMT1 degradation

There are generally two options for the order in which endogenous DNMT1 can be replaced by the mutant protein using a classic genetic KO approach:

1. The endogenous protein can be depleted first, followed by the introduction of the mutant. Generating a *Dnmt1* KO cell line prior to the introduction of the mutant has the advantage that the KO only needs to be performed and confirmed once, after which many different mutant cell lines can be established in this line [57, 71]. However, upon *Dnmt1* KO, the cells will lose most DNA methylation. Thus, the starting point for evaluating the mutant phenotype with respect to DNA methylation will not be a fully methylated genome, which may confound the analysis. In addition, this approach is limited to blastocyst-derived stem cells, since *Dnmt1* KO and the concomitant loss of DNA methylation is lethal for further differentiated cell types. Lastly, a transient lack of DNMT1 leads to the loss of imprints, which can only be established through germ line passage [154]. Thus, the maintenance of imprints cannot be studied in a cell culture system once DNMT1 and subsequently DNA methylation have been lost.

2. Alternatively, the mutant protein can be introduced first, followed by the removal of the endogenous protein. In this approach both endogenous and mutant *Dnmt1* are simultaneously expressed at first, and no phenotypes of excess DNMT1 have been reported so far. However, the drawback of this approach is that the subsequent *Dnmt1* KO needs to be performed and confirmed in every mutant cell line.

Thus, neither of these classic approaches was ideal for our envisioned platform. However, various methods have been developed to deplete DNMT1 without having to conduct and confirm a gene KO after the introduction of the mutant transgene to be studied. For example, replacement of the endogenous *Dnmt1* promoter with a TET^{off} promoter silences the gene in the absence of the drug tetracycline. Expression of the gene requires continuous addition of tetracycline to the cell culture medium, while removal of the drug leads to the silencing of *Dnmt1* [70]. Alternatively, parts of the *Dnmt1* gene can be floxed, enabling an inducible knockout upon transient transfection with Cre [155]. However, the efficient excision of the floxed region needs to be subsequently confirmed.

Here, we use an alternative approach taking advantage of the recent developments in the field of target-specific protein degradation. Specifically, the dTAG system enables the quick and reversible degradation of any protein that is fused to the $FKBP12^{F36V}$ degron (Figure 4.1 A). Addition of the heterobifunctional degrader

dTAG-13 to the cell culture medium recruits the endogenous E3 ubiquitin ligase Cereblon to the degron fusion protein, leading to its polyubiquitination and subsequent proteasomal degradation [156].

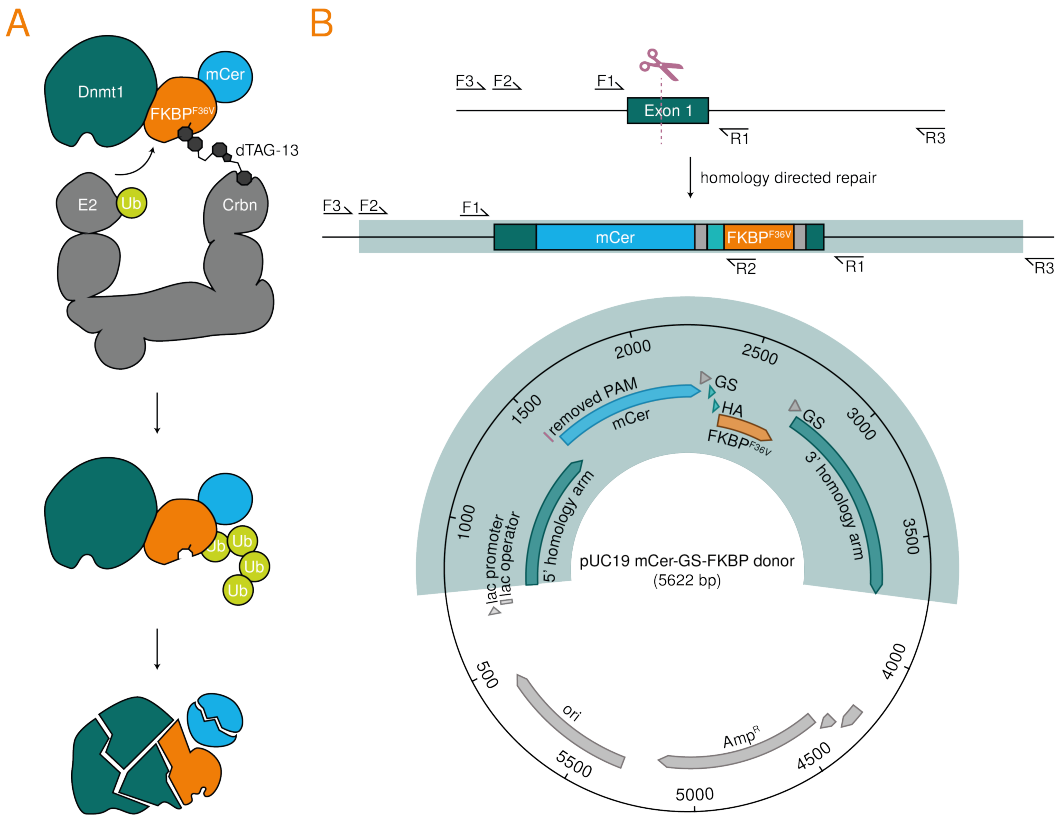

Fig. 4.1.: Degron-tagging of the endogenous *Dnmt1* locus. A The heterobifunctional dTAG-13 molecule dimerizes the FKBP12^{F36V} degron, fused to DNMT1, and Cereblon. Subsequent polyubiquitination targets the fusion protein for proteasomal degradation (adapted from [156]). B During Cas9-mediated knock-in, the guide targets exon 1 in the *Dnmt1* locus, 44 bp upstream of the translation start site. The cut site is repaired through homology-directed repair, using the sequence provided by the pCU19 mCer-GS-FKBP donor plasmid (shaded green). Thereby, the mCerulean and FKBP12^{F36V} sequences are inserted in-frame downstream of the start codon in exon 1. C Genotyping by PCR with the indicated primers. D Expression levels of *Dnmt1* determined by qPCR, normalized to wt. E Brightfield images at 4X magnification. Crbn: Cereblon; GS: glycine-serine linker; HA: hemagglutinin tag; mCer: mCerulean; Ub: ubiquitination.

We inserted an mCerulean fluorescent reporter and the FKBP12^{F36V} degron using a Cas9-mediated knock-in guided by the donor template shown in Figure 4.1 B. Among 48 isolated monoclonal cell lines, clone 39 (C39) carries the insertion on both alleles, displays normal cell morphology and continues to express *Dnmt1* slightly below wt levels (Figure 4.2 A-C).

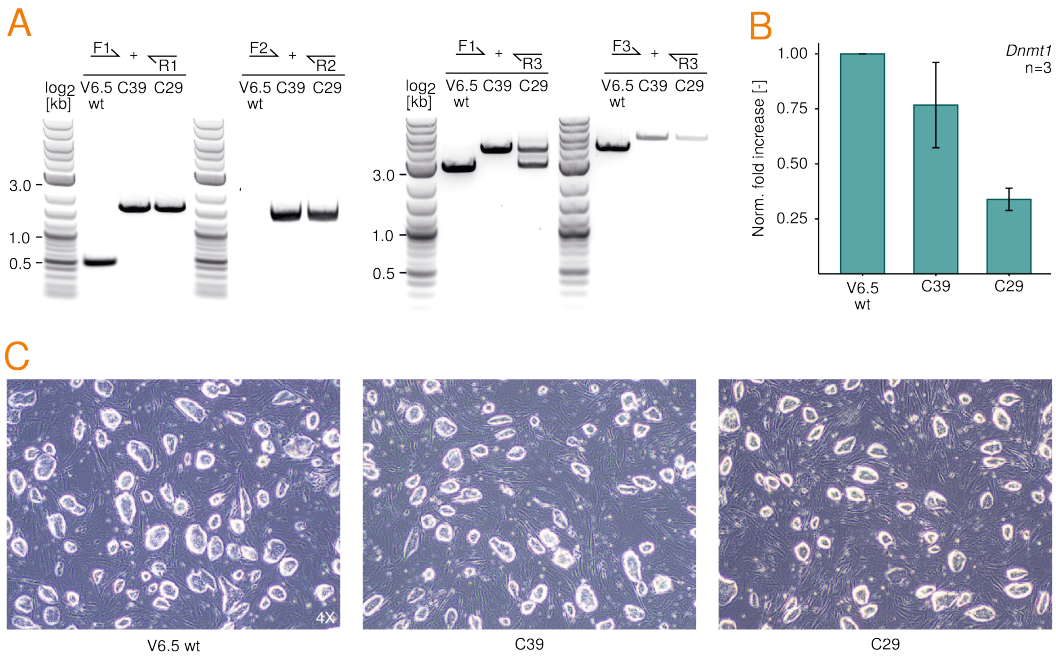

Fig. 4.2.: Validation of two degron-tagged clones. A Genotyping by PCR around the insertion site. The location of the indicated primers' binding sites are shown in 4.1 B. B Expression levels of *Dnmt1* determined by qPCR, normalized to V6.5 wt. C Brightfield images at 4X magnification.

Addition of 250 nM dTAG-13 to the cell culture medium resulted in the swift depletion of DNMT1 below levels detectable by Western blot within 4 h (Figure 4.3 A). However, residual fluorescent signal was detectable using FACS, reaching a minimal level by 24 h and remaining even after 8 d of treatment (Figure 4.3 B-C).

To determine whether this degree of degradation is sufficient to deplete the cells of maintenance methylation activity, we used RRBS to profile the methylation levels of C39 over 7 d of dTAG-13 treatment, followed either by treatment for additional 7 d or by a recovery period of 6 d (Figure 4.3 D-F). Within 7 d, methylation dropped from 73.9 % to 19.7 % over 1 kb tiles and hardly decreased further by 14 d of treatment. Upon removal of dTAG-13 after 7 d, methylation levels almost recovered to wt levels within 6 d (Figure 4.3 F), with the exception of a subset of CGIs covering imprinted regions (Figure 4.4 A-B).

These results are in line with data obtained from *Dnmt1* KO embryos, which maintain a residual level of methylation of about 20 % due to the activity of the *de novo* methyltransferases DNMT3A and DNMT3B [157, 158]. However, the continued presence of DNMT3A and -B in our cell line confounds the detection of potentially

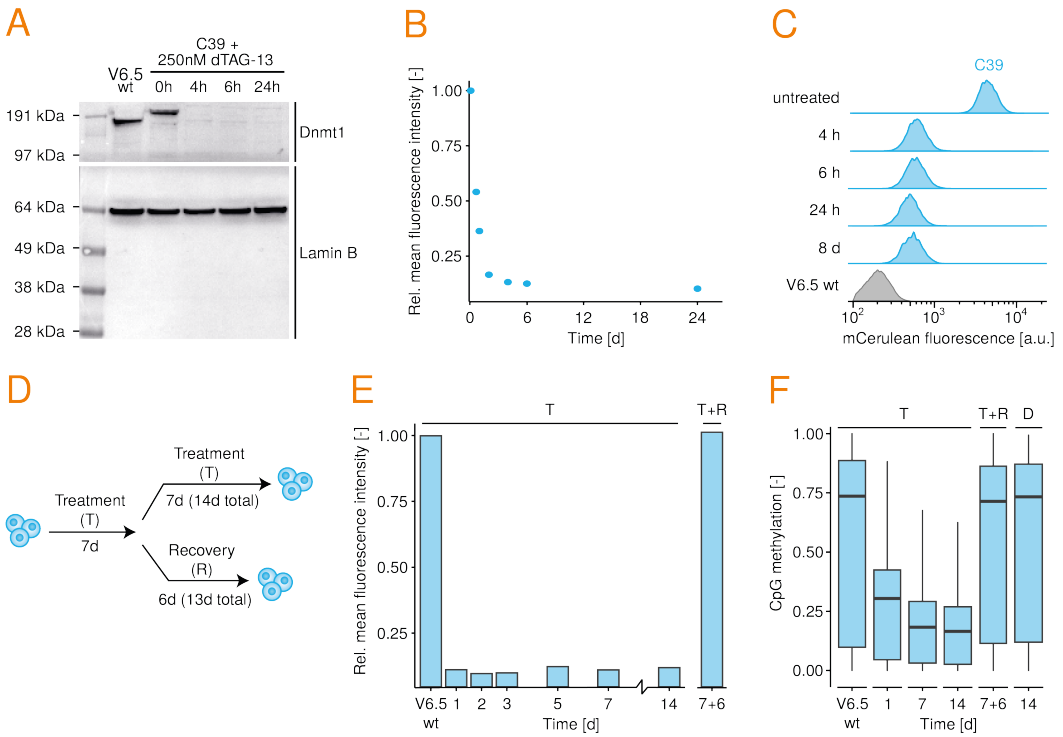

Fig. 4.3.: Effect of dTAG-13 treatment on the Dnmt1 degron cell line. A Western blot detecting DNMT1 and Lamin B in C39 treated with 250 nM dTAG-13. B, C Fluorescence intensity of mCerulean in C39 treated with 250 nM dTAG-13. D Schematic of the treatment time course with 250 nM dTAG-13 profiled by RRBS. E Mean fluorescence intensity of mCerulean over the treatment time course. F CpG methylation over 1 kb tiles measured by RRBS over time. D: DMSO only; R: recovery in DMSO; T: treatment with 250 nM dTAG-13.

low maintenance methylation activity by residual DNMT1. Therefore, we removed *Dnmt3a* and *Dnmt3b* in C39 via Cas9-mediated KO (Figure 4.5 A-B). The monoclonal cell line C46 carries deletions on both alleles of *Dnmt3a* (2 and 87 bp in exon 19 on the respective alleles) and *Dnmt3b* (15 bp deletion in exon 20 on one allele, 120 bp in intron 19 and exon 20 on the other allele) and does not express either protein (Figure 4.5 C-D).

We then treated C46 with DMSO or 250 nM dTAG-13 over the course of 7 d, followed by a recovery period of 6 d under the addition of DMSO only or another 7 d treatment (Figure 4.6 A). By day 1 of the treatment time course, the methylation dropped to 11.6 % measured over 1 kb tiles (compared to 34.5 % for C39 treated for 1 d under the same conditions), underscoring that the sole presence of DNMT1 is not sufficient to maintain methylation in mouse ESCs (Figure 4.6 B).

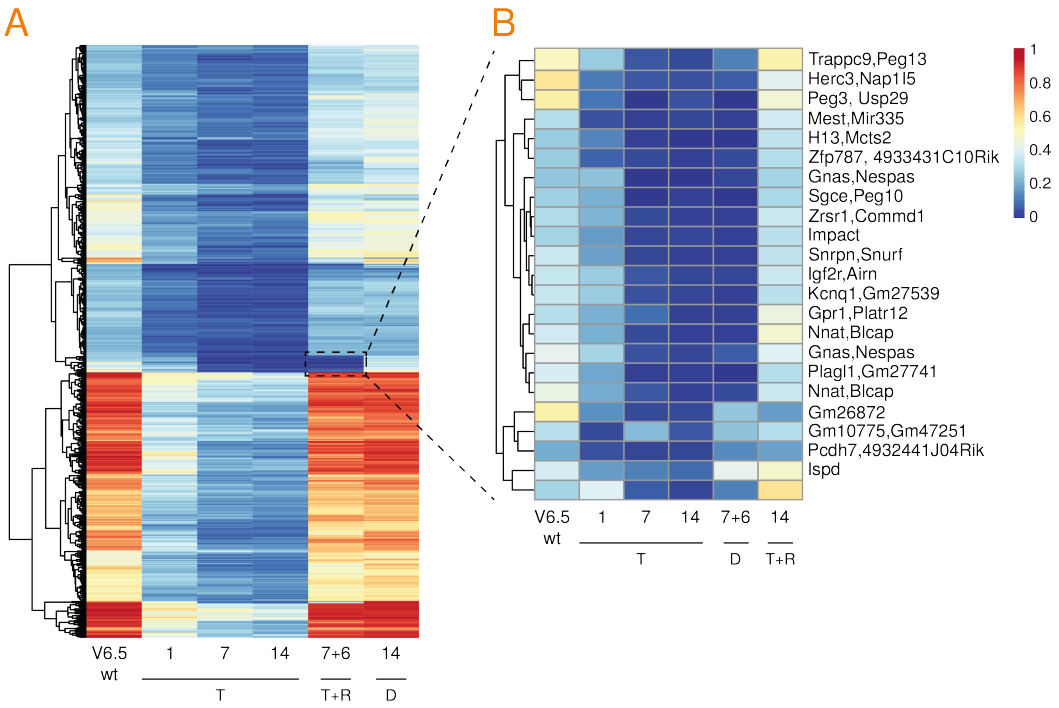

Fig. 4.4.: CGI methylation in dTAG-13 treated cells. A Mean CpG methylation at CpG islands throughout the treatment time course illustrated in Figure 4.3 D. B Mean CpG methylation at imprinted control regions. The same color scale applies to both panels. D: DMSO only; R: recovery in DMSO; T: treatment with 250 nM dTAG-13.

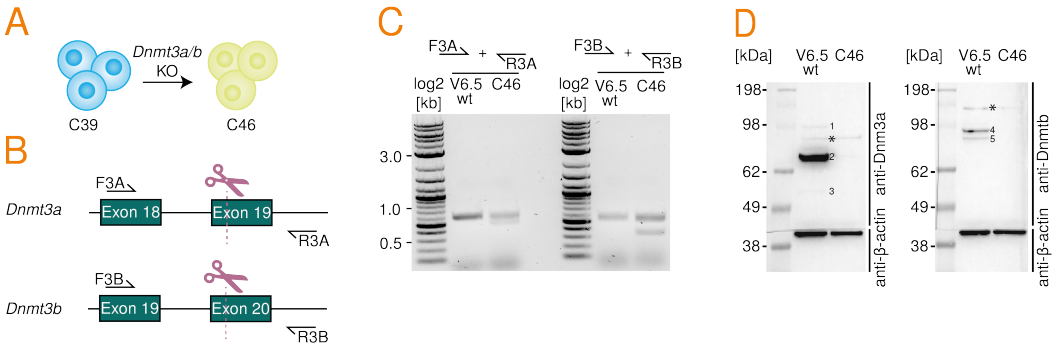

Fig. 4.5.: Derivation and characterization of a *Dnmt3a/b* DKO in the Dnmt1-degron cell line. A The monoclonal cell line C46 was derived from C39 by simultaneous Cas9-mediated KO of *Dnmt3a* and *Dnmt3b*. B Genotyping of C46 by PCR. C Expression levels of *Dnmt3a* and *Dnmt3b* determined by qPCR. D Western blot against DNMT3A and -B in V6.5 wt mESCs and C46, with β-actin as loading control.

Upon treatment with dTAG-13 for 7 d, C46 lost the majority of DNA methylation (Figure 4.6 B). However, certain genomic features, particularly repeat elements such as LINE1 and long terminal repeats (LTRs), retained relatively high levels of DNA methylation (11.5 % and 13.4 %, respectively Figure 4.6 C). Interestingly, upon removal of dTAG-13, few regions gained methylation *de novo* (minimal gain of

10 %; Figure 4.6 D) despite the absence of the canonical *de novo* methyltransferases DNMT3A and DNMT3B. Feature analysis revealed these regions to cover repeat elements, mainly Endogenous Retrovirus (ERV) K and LINE1 elements (Figure 4.6 D). In total, 907 repeat elements gained methylation over the recovery period of 6 d. In parallel to this work, our laboratory demonstrated that DNMT1 possesses *de novo* activity at specific regions of the genome, resulting in *de novo* differentially methylated regions (DMRs). These regions exhibit a great overlap with LTRs, in particular intracisternal A-particle repeats (IAPs), a sub-class of ERVK) [33]. The recovering regions identified in this work partially overlap with the *de novo* DMRs described by Haggerty et al. (2021), particularly at ERVK elements (Figure 4.6 E) [33].

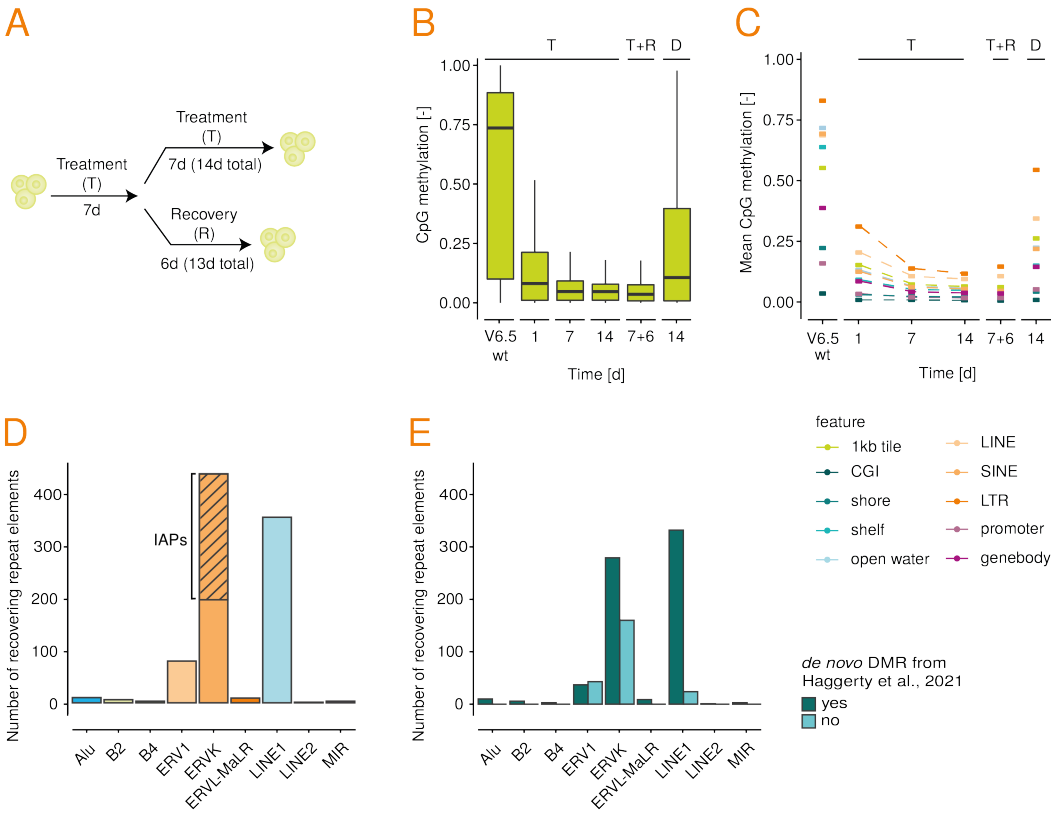

Fig. 4.6.: Methylation analysis of dTAG-13 treated *Dnmt3a/b* DKO in the Dnmt1-degron cell line. A Schematic of the treatment time course with 250 nM dTAG-13 profiled by RRBS. B CpG methylation over 1 kb tiles measured by RRBS over time. C Mean CpG methylation by genomic features measured by RRBS over time. D Distribution of repeat elements that gain at least 10 % methylation between day 7 of the treatment and day 6 of the recovery period. E Overlap of recovering repeat elements with *de novo* DMRs identified by Haggerty et al. (2021) [33]. CGI: CpG island; D: DMSO only; DMR: differentially methylated region; ERV: endogenous retrovirus, IAPs: intracisternal A-particle repeats; LINE: long interspersed nuclear element; LTR: long terminal repeat; MIR: mammalian-wide interspersed repeats; R: recovery in DMSO; T: treatment with 250 nM dTAG-13.

4.2 Over-expression of Dnmt1 as transposon

For the introduction of the Dnmt1 mutant transgene, we utilized the transposon-based PiggyBac system which results in multiple random integrations of the transgene in the host genome via a "cut and paste" mechanism [159]. On the transposon vector, the transgene—including the promoter and poly-adenylation signal—is flanked by inverted terminal repeat sequences. The host cells are transfected with the transposon vector, alongside a second vector coding for the PiggyBac transposase. Upon successful transfection, the host cell transiently expresses the transposase which recognizes the transgene transposon by its inverted terminal repeat sequences and mobilizes it into the host genome.

To facilitate the generation of various mutant plasmids and to preclude secondary mutations that may occur during the amplification of long cDNAs such as Dnmt1, we used Gateway cloning to generate the transposon vectors (Figure 4.7). We constructed a Gateway destination vector including an EF1α promoter, mCherry and a triple FLAG epitope sequence (3XFLAG) upstream of the ccdB cassette (Figure 4.7 A). These sequences were flanked by insulator sequences and the inverted terminal repeat sequences on both sides. A codon-optimized cDNA of Dnmt1 tagged with a V5 epitope (V5-Dnmt1^{co}) was synthesized in two pieces (d1 and d2) and cloned into the pENTR4 entry vector (Figure 4.7 B). Upon recombination, the V5-Dnmt1^{co} sequence was inserted in-frame downstream of the mCherry-3XFLAG open reading frame (Figure 4.7 C). For efficient screening of transfected cells, the *Dnmt1* transgene was tagged with mCherry, while the transposase was expressed as a BFP fusion with a self-cleaving T2A linker. Double positive cells were isolated by FACS 2 d post-transfection (Figure 4.8 B). Single clones were then isolated and screened for mCherry fluorescence signal by the transgene. The clones with the strongest mCherry fluorescence signal were expanded and transgene expression subsequently characterized by qPCR and Western blot.

In order to assess the performance of this construct in a cell line lacking endogenous DNMT1, we generated monoclonal cell lines expressing the transgene from KH2 triple KO (TKO; *Dnmt1-/-, Dnmt3a-/-, Dnmt3b-/-*) mouse ESCs. Of the eleven clones with the strongest mCherry fluorescent signal, nine had detectable levels of mCherry mRNA, correlating with the determined number of integrations (Figure 4.8 C-E). However, hardly any levels of DNMT1 protein were detectable by West-

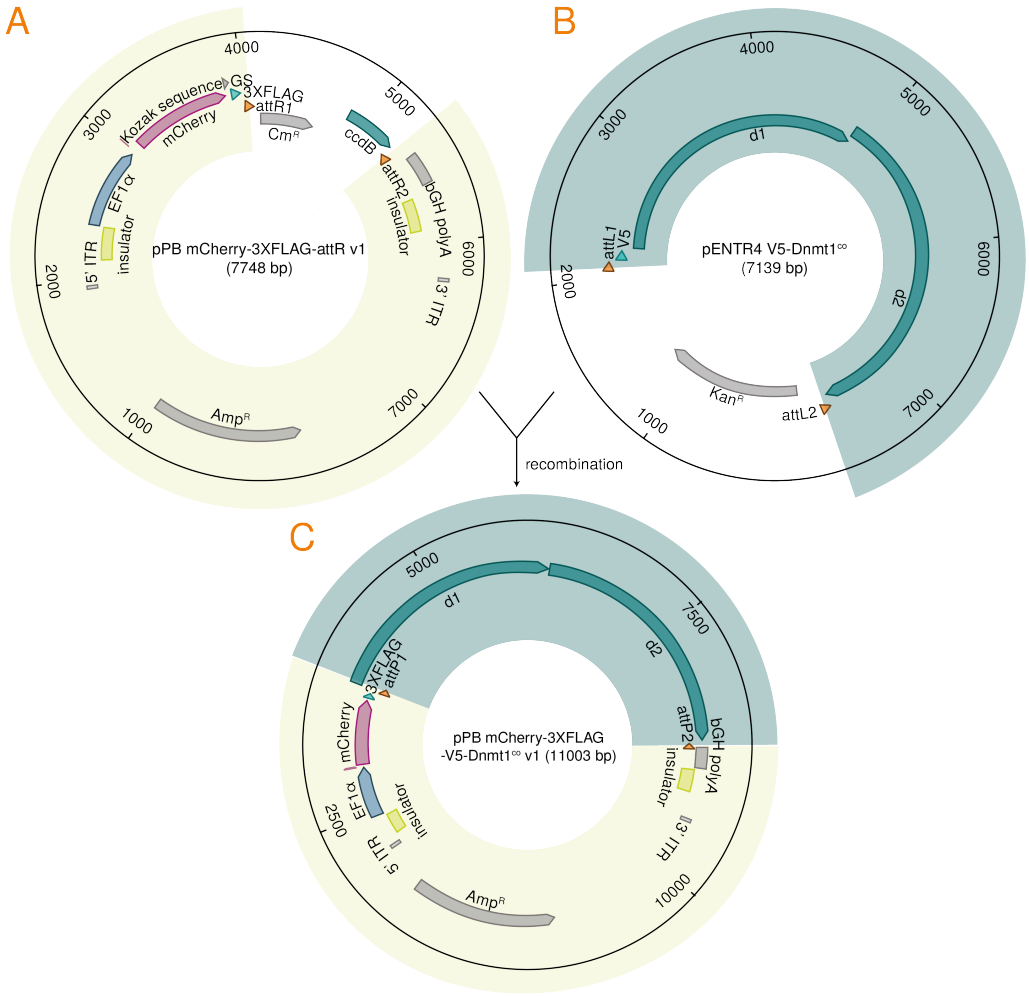

Fig. 4.7.: Schematics of Gateway vectors. A The destination vector contains the EF1α promoter, an mCherry-3XFLAG open reading frame, followed by the ccdB cassette flanked by attR sites and a bGH polyA signal. This whole sequences in flanked by insulator sequences and the ITRs. B The entry vector contains the codon-optimized V5-Dnmt1 (V5-Dnmt1^{co}) sequences flanked by attL sites. C Upon recombination by LR clonase, the ccdb cassette is swapped with the V5-Dnmt1^{co} sequence, resulting in an open reading frame containing mCherry-3XFLAG-attP-V5-Dnmt1^{co} driven by the EF1α promoter on the expression vector. The attP1 sequence results in the linker sequence TSLYKKAGL. 3XFLAG: triple FLAG epitode; co: codon-optimized; ITR: inverted terminal repeat; bGH polyA: bovine growth hormone poly adenylation signal

ern blot (Figure 4.8 F). We therefore aimed to improve the stability of the protein by introducing a self-cleaving linker (P2A) between mCherry and the 3XFLAG epitope. Using this vector also increased the number of cells falling with the mCherry+ gate by about three-fold.

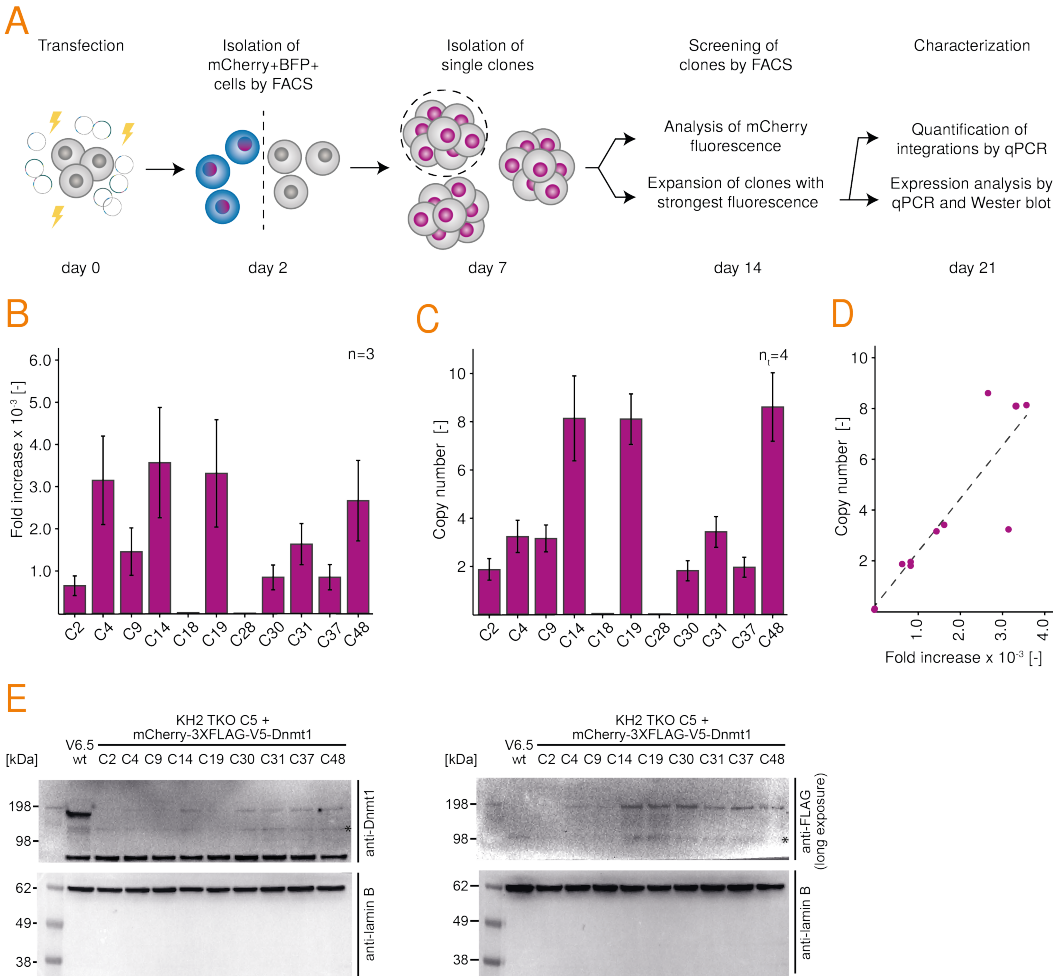

Fig. 4.8.: Ectopic expression of V5-Dnmt1^{co} in KH2 TKO cells. A Scheme illustrating the derivation and characterization of clones with stable ectopic expression of mCherry-3XFLAG-V5-Dnmt1^{co} after the transfection with pPB mCherry-3XFLAG-V5-Dnmt1^{co} and a vector expressing the BFP-tagged PiggyBac transposase. B Quantification of V5-Dnmt1^{co} expression by qPCR. Error bars represent the standard deviation. C Quantification of mCherry copy numbers integrated into the genome by qPCR. Error bars represent the standard deviation. D Correlation between V5-Dnmt1^{co} mRNA levels and copy number. E Western blot with anti-Dnmt1 antibody (left blot) and anti-FLAG antibody (right blot), with Lamin B as loading control. Unspecific bands are denoted with an *.

In an alternative approach, we aimed to improve the expression of the transgene by increasing the distance between the EF1α promoter and the Kozak sequence, which had been shortened compared to the original backbone (pPB mCherry/GFP loxP) during the construction of the destination vector (Figure 4.9 A). In HEK 293T cells transiently expressing the transgene, the second generation destination vector (pPB mCherry-3xFLAG-attR v2) generated a robust mCherry signal (Figure 4.9 B, C). Transfection with the second generation expression vector (pPB mCherry-3xFLAG-V5-Dnmt1^{co} v2) resulted in an approximately three-fold increase of mCherry fluo-

rescent signal within the mCherry+ gate compared to the original expression vector (pPB mCherry-3xFLAG-V5-Dnmt1^{co} v1; Figure 4.9 B). Using pPB mCherry-3xFLAG-V5-Dnmt1^{co} v2, the signal intensity within the mCherry+ fraction reached 140 % when compared to using the v1 backbone in combination with the self-cleaving linker P2A.

Using the second generation expression vector to generate new monoclonal cell lines in the KH2 TKO background resulted in an approximately ten-fold increase of mCherry mRNA levels (Figure 4.10 A) compared to the previously established clone 19 (v1). Albeit not reaching wt levels, a band for DNMT1 protein, shifted due to the fusion with mCherry and the epitopes, was clearly detectable in all four clones (Figure 4.10 C). Clone 37 (C37) displayed the highest DNMT1 protein levels and the highest number of integrations (Figure 4.10 B, C).

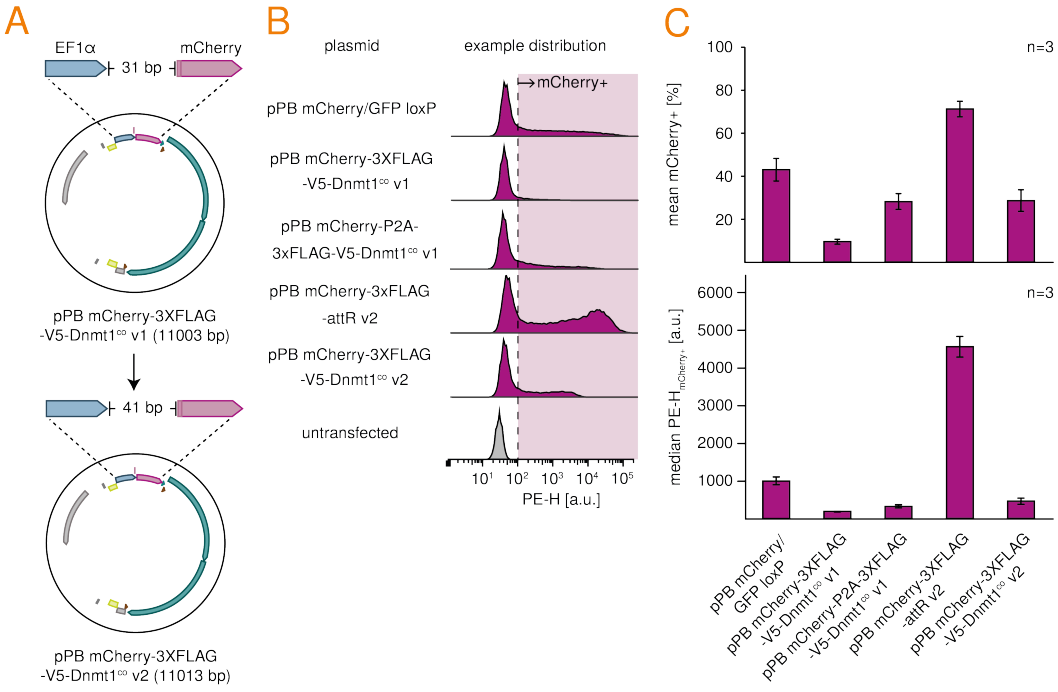

Fig. 4.9.: Test of 2nd generation expression vectors in HEK293T. A From v1 to v2, the promoter downstream sequence was extended by 10 bp, reconstructing the original sequence present in pPB mCh/GFP loxP. B Fluorescence intensity corresponding to mCherry in HEK293T cells transiently expressing the indicated plasmid. The experiment was carried out in triplicates. Shown is a representative example of the fluorescence signal distribution. The dashed line indicates the start of the mCherry+ gate. C The percentage of cells falling into the mCherry+ gate (top) and the median fluorescence intensity within this gate (bottom) are given as mean calculated over all three triplicates. Error bars represent the standard deviation.

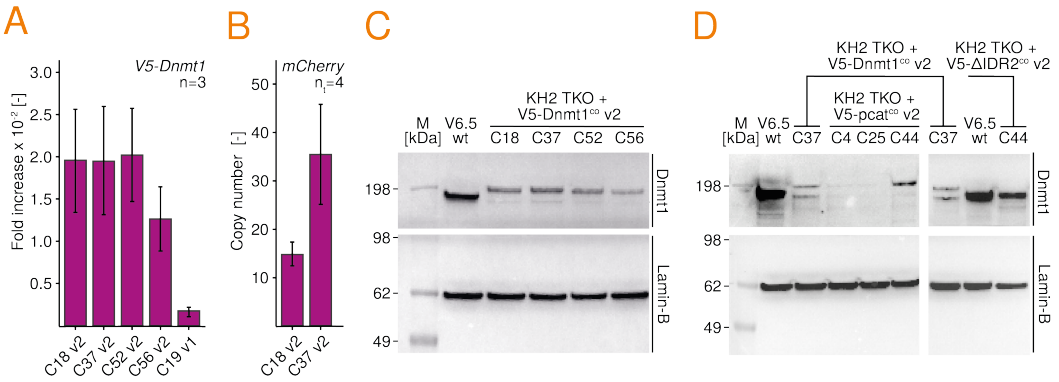

Fig. 4.10.: Expression from 2nd generation vectors in KH2 TKO. A Expression levels of codon-optimized *Dnmt1* expressed from pPB mCherry-3XFLAG-V5-Dnmt1^{co} v1 or v2 in KH2 TKO, determined by qPCR. Errors bars represent the standard deviation among replicates. B Number of integrated construct copies of selected clones measured by qPCR. Errors bars represent the error among technical replicates determined using Gaussian error propagation. C Western blot against DNMT1 in clones expressing pPB mCherry-3XFLAG-V5-Dnmt1^{co} v2. C Western blot against DNMT1 in clone 37 expressing pPB mCherry-3XFLAG-V5-Dnmt1^{co} v2, and three clones expressing pPB mCherry-3XFLAG-V5-pcat^{co} v2 and clone 44 expressing pPB mCherry-3XFLAG-V5-ΔIDR2^{co} v2.

In order to test the functionality of the ectopically expressed DNMT1, we analyzed the DNA methylation of C37 cultivated for 17 d and 60 d using whole-genome bisulfite sequencing (WGBS). Due to the lack of all three DNMTs, KH2 TKO cells lack DNA methylation (Figure 4.11). Ectopic expression of V5-Dnmt1^{co} in TKO cells resulted in a minor genome-wide gain in methylation, but particularly in the *de novo* methylation of distinct DMRs (Figure 4.11 A-C). After the introduction of the Dnmt1 rescue construct, methylation at DMRs and in the background increased over time, with the gain being most pronounced at DMRs and flanking regions (Figure 4.11 A-D). After 60 d in culture, methylation within DMRs of V5-Dnmt1^{co} reached 18 % on average (Figure 4.11 C).

The appearance of *de novo* DMRs supports our laboratory's findings that DNMT1 possess *de novo* activity mainly targeted to repeat elements [33]. Despite not expressing DNMT1 at wt protein levels, total methylation gain and rate at *de novo* DMRs in C37 and DKO$_0$ cells were comparable . The latter are derived from TKO-like cells which are *Dnmta-/-; Dnmt3b-/-* with an shRNA knockdown of *Dnmt1*. Like true TKO cells, TKO-L thus lack all main DNMTs and have lost all DNA methylation. DKO$_0$ cells are generated by reversing the shRNA knockdown of *Dnmt1* and express DNMT1 at wt levels (reference [33], Figure 2B). DKO$_0$ gain 5mC over time, where

five and fifteen passages are approximately equivalent to 15 d and 45 d in culture (Figure 4.11 A-D).

De novo DMRs show a strong overlap with repeat elements, particularly IAPs (Figure 4.11 A). However, also other repeat elements are targeted. For example, compared to transcription start sites (TSSs), which are also highly methylated in the wt, 5mC gain in both C37 and DKO_0 was pronounced at LTRs and LINEs (Figure 4.12).

Generally, these results show that the randomly integrated V5-Dnmt1^{co} transgene produces a functional, enzymatically active protein capable of *de novo* methylation. To confirm that the *de novo* methylation depends on the catalytic activitiy of DNMT1, we mutated the conserved PC motif within the MTase domain using QuikChange mutagenesis (P1228G, C1229S). Supporting our findings, clone 44 expressing this catalytically inactive mutant V5-pcat^{co} did not display any changes in methylation, neither globally nor at DMRs (Figures 4.10 D, 4.11,4.12).

Next, we wondered whether the IDR of DNMT1 is required for its recruitment to the *de novo* DMRs and its *de novo* activity. Thus, analogous to V5-Dnmt1^{co} and V5-pcat^{co}, we derived a cell line V5-ΔIDR2^{co} which lacks 151 amino acids in the IDR downstream of the PIP box, including most of the eutherian-specific ROI as well as the distal part conserved in other animals. The remaining IDR is approximately equivalent to the sequence in the "opto-cIDR1" construct which was hardly able to form condensates in the context of the optoDroplet assay (Section 3.1, Figure 3.15). V5-ΔIDR2^{co} C44 expressed mutant DNMT1 near wt levels, and after 17 d in culture, exhibited 5mC at DMRs and the background similar to V5-Dnmt1^{co} C37 at day 17 and DKO_0 cells at passage 5 (roughly 15 d in culture) (Figures 4.10 D, 4.11 A). Likewise, V5-ΔIDR2^{co} C44 gained methylation at the same genomic features as V5-Dnmt1^{co} C37 after 17 d and DKO_0 p5 (Figure 4.12). Hence, neither the ROI nor the ability of the IDR to undergo LLPS are required to recruit DNMT1 to *de novo* DMRs and to catalyze the addition of methylgroups to unmethylated CpGs.

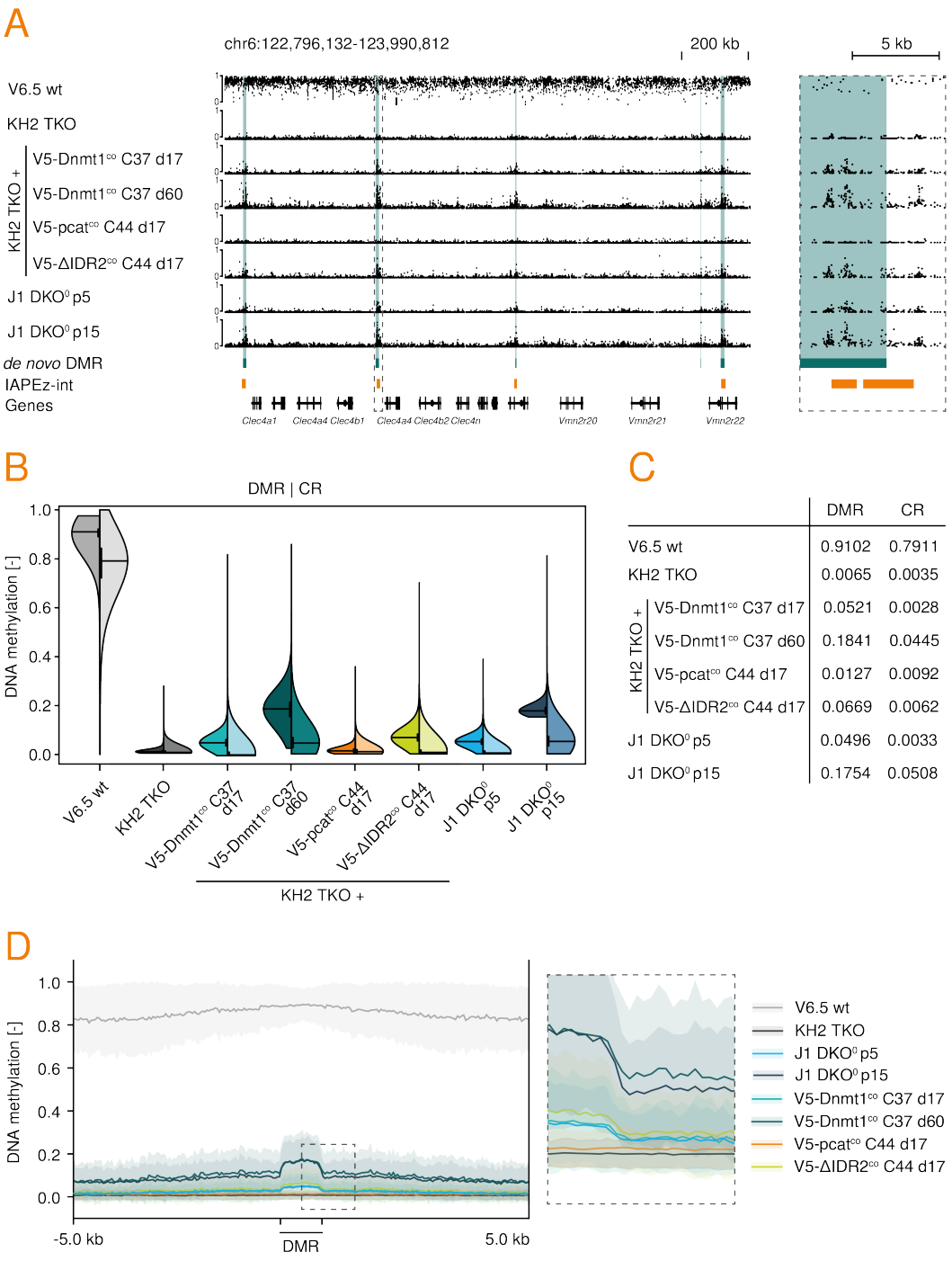

Fig. 4.11.: Rescue of DNMT1 activity in 5mC deficient cells. A Browser tracks of a 1 Mbp region on the q arm of chromosome 6 containing five DMRs. The right panel shows a zoom of the DMR upstream of the *Clec4a4* gene. DMRs are shaded in green and IAPEz-int elements are marked in orange. Each dot represents the mean 5mC level of one CpG covered. B 5mC methylation levels at DMRs (left violin half) and control regions (right violin half). Horizontal bars indicates the median and vertical bars the interquartile range. C 5mC mean methylation across DMRs and control regions. D Composite plots of 5mC levels across DMRs and 5 kbp flanking regions. The right panel shows a zoom of the outlined windows in the left panel. CGI: CpG island; CR: control region; DMR: differentially methylated region; LINE: long interspersed nuclear element; LTR: long terminal repeat.

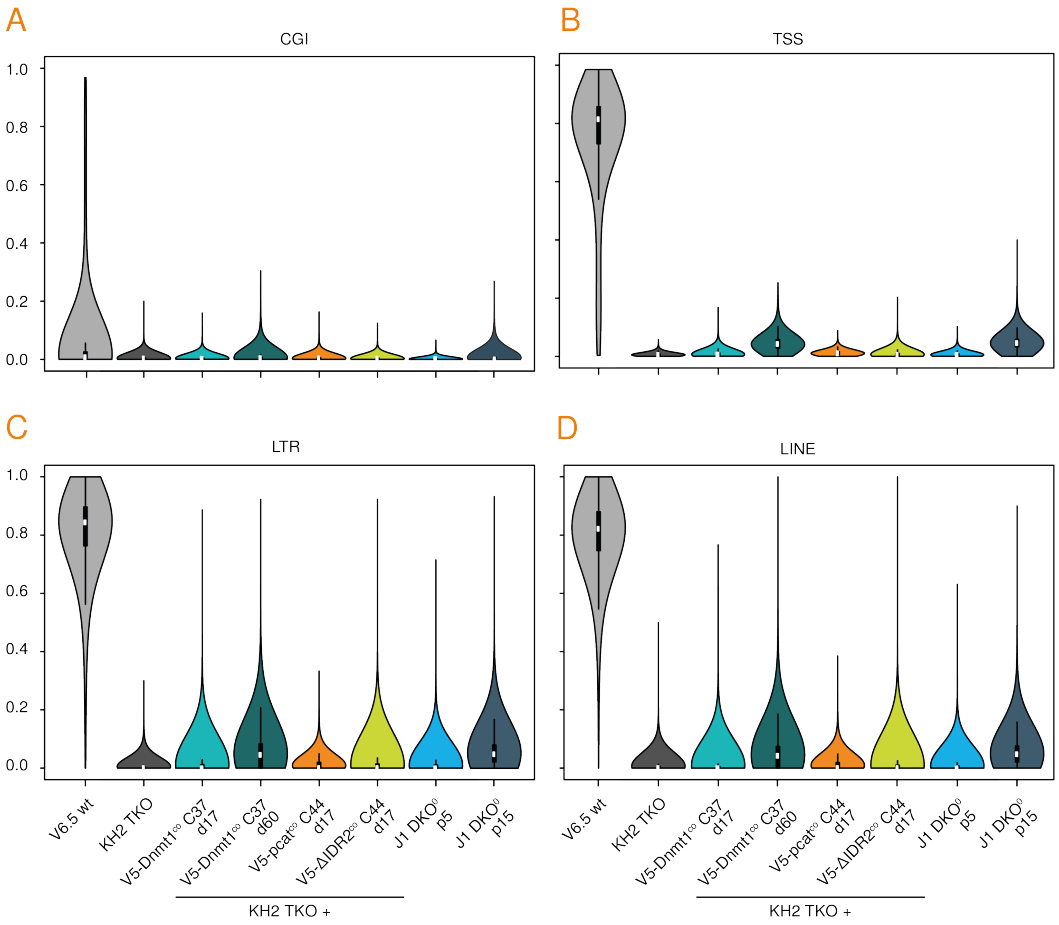

Fig. 4.12.: 5mC at genomic features in rescue cell lines. A 5mC levels in wt, TKO and various rescue cell lines at CGIs, B at TSSs, C at LTRs, and D at LINEs. CGI: CpG island; LINE: long interspersed nuclear element; LTR: long terminal repeat; TSS: transcriptional start site.

4.3 Optimization of ectopic Dnmt1 expression using introns

Despite many integrations, DNMT1 levels achieved by the random integration of pPB mCherry-3XFLAG-V5-Dnmt1^{co} v2 were below the ones of wt cells, which express the protein from two alleles only (Figure 4.10 B, C). Thus, we strove to further optimize DNMT1 expression to wt levels. First, we wondered whether the codon-optimized cDNA is recognized as foreign by the cells, resulting in the silencing of most integrations. We therefore devised a new entry vector by cloning the Dnmt1 cDNA from V6.5 ESCs and created a new expression vector coding for an mCherry-3XFLAG-Dnmt1 fusion protein (pPB mCherry-3XFLAG-Dnmt1 v2). In transiently transfected HEK293T cells, this vector resulted in an almost two-fold increase of the mCherry signal intensity within the mCherry+ gate compared to the second generation expression vector containing the codon-optimized cDNA (Figure 4.14).

While this work was in preparation, a new study showed that intronless cDNAs are prone to silencing by the human silencing hub (HUSH) complex [160]. HUSH initiates the silencing of invading DNA species, such as LINE1 elements, and is conserved across vertebrates [161]. Seczynska et al. (2022) show that HUSH distinguishes such invading retrogenes from host genes by their intronless cDNAs, which are usually a product of reverse transcription [160]. Hence, HUSH is not specific to repetitive elements but may also recognize other long, intronless cDNAs such as ectopically introduced transgenes. Introduction of an intron, *e.g.*, the second intron from the human β-globin gene (HBB), into the transgene can alleviate the repression by HUSH [161]. In addition, Lacy-Hulbert et al. (2001) demonstrated that the introduction of introns into transgenes can boost their expression [162]. Against this background, we therefore wondered whether the introduction of an intron into our rescue constructs improves the ectopic expression of Dnmt1.

To approach this question, we devised three different constructs containing introns:

1. “Hbb” contains the mouse *Hbb* intron 2 between *Dnmt1* exon 3 and 4, analogous to the construct used by Seczynska et al. (2022) (Figure 4.13 A) [160].
2. “End” contains the endogenous Dnmt1 introns 2 and 3 (Figure 4.13 B).

3. “Igh” contains the *Ighm* intron 1 between *Dnmt1* exon 18 and 19, and the *Ighe* intron 3 between *Dnmt1* exon 28 and 29, based on the report by Lacy-Hulbert et al. (2001) (Figure 4.13 C) [162].

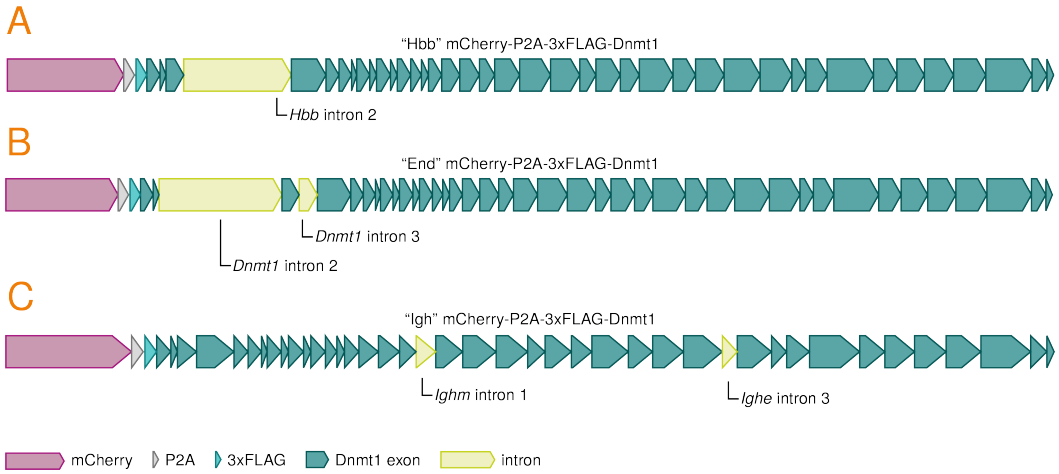

Fig. 4.13.: Linear maps of three Dnmt1 constructs for ectopic expression containing introns. A “Hbb” contains intron 2 from mouse *Hbb* between *Dnmt1* exon 3 and exon 4. B “End” contains the endogenous *Dnmt1* introns 2 and 3 at their original location. C “Igh” contains *Ighm* intron 1 and *Ighe* intron3 from mouse, located between *Dnmt1* exons 18 and 19, and exons 28 and 29, respectively.

Taking into consideration our previous results, all three intron-containing constructs were based on the second generation expression vector (v2) and include the endogenous, non-codon-optimized Dnmt1 cDNA. To minimize any effect on DNMT1 stability, the novel constructs also encode a P2A linker between mCherry and the 3XFLAG tag (Figure 4.13). We then compared the expression from these three intron-containing constructs to our previous rescue constructs in transiently transfected HEK293T cells in three replicate experiments (Figure 4.14). Compared to the constructs without introns, the geometric mean of the fluorescence signal of transfected, mCherry+ cells doubled using the intron-containing constructs. All three intron-containing constructs performed well with respect to transfection efficiency and fluorescence signal.

Due to the simpler cloning approach and smaller plasmid size compared to the intron containing the endogenous *Dnmt1* introns, we selected the “Hbb” Dnmt1 construct to derive over-expressing monoclonal cell lines in the *Dnmt1* degron background C39. In parallel, we derived three additional cell lines in order to investigate the role of the IDR and the ROI in maintenance methylation. “Hbb” ΔcIDR2 lacks S190 to P369, which includes the IDR conserved across animals downstream of the

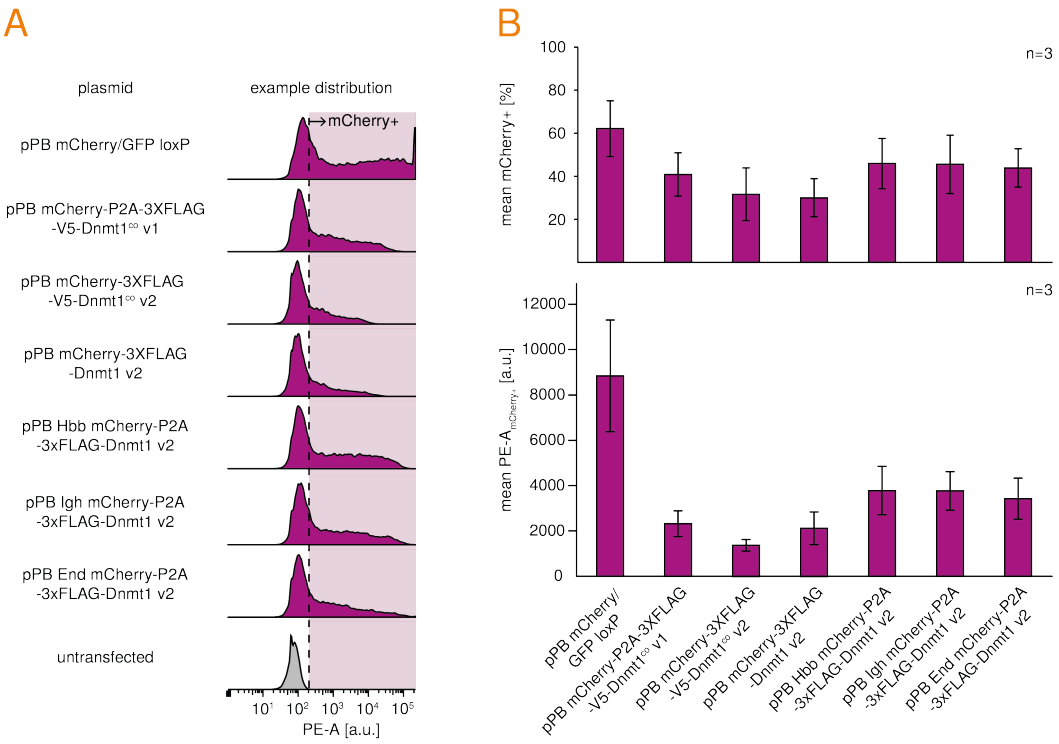

Fig. 4.14.: Test of intron-containing Dnmt1 constructs in HEK293T. A Fluorescence intensity corresponding to mCherry signal in HEK293T cells transiently expressing the indicated plasmid. The experiment was carried out in triplicates. Shown is a representative example of the fluorescence signal distribution. The dashed line indicates the start of the mCherry+ gate. B The percentage of cells falling into the mCherry+ gate (top) and the geometric mean fluorescence intensity within this gate (bottom) are given as mean calculated over all three triplicates. Error bars represent the standard deviation.

NLS, including the ROI (Figure 4.15 A). Thus, "Hbb" ΔcIDR2 only contains the conserved IDR limited to the sequence upstream of the PIP box (cIDR1), which hardly promoted condensate formation in the OptoDroplet assay (Figure 3.15 C). To investigate the role of the ROI alone, "Hbb" ΔROI lacks the ROI, comprised of S190 to D311 (Figure 4.15 B). Finally, "Hbb" ΔPIP carries the Q162E mutation within the PIP box, compromising its ability to bind PCNA (Figure 4.15 C) [57]. This construct can be used to compare the effect of the ΔcIDR2 and ΔROI on recruitment efficiency, since ΔPIP has been described to halve the efficiency of DNMT1 recruitment without reducing steady-state levels [57].

As previously described, mCherry+BFP+ double positive cells were isolated by FACS 2 d post-transfection (Figure 4.16 A, B). Figure 4.16 C shows the mCherry fluorescence signal of the mCherry+BFP+ population isolated in Figure 4.16 B 9 d post-transfection. As expected, the cells did not retain the BFP signal of the tran-

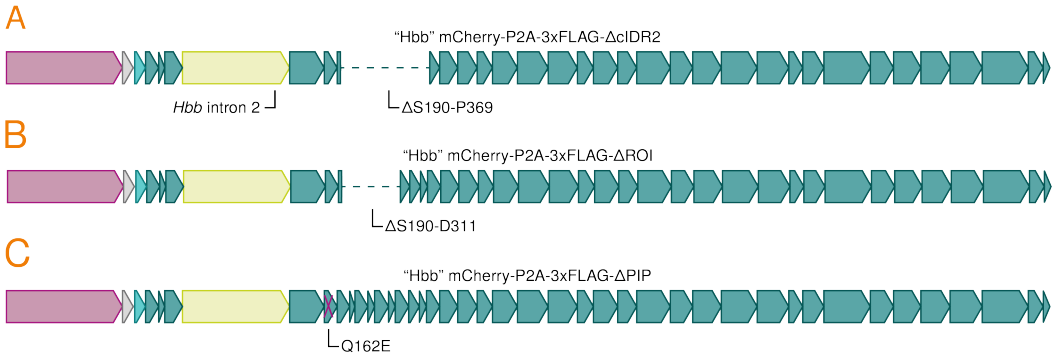

Fig. 4.15.: Linear maps of three Dnmt1 mutant constructs for ectopic expression containing the Hbb intron. A Analogous to "Hbb" Dnmt1, "Hbb" ΔcIDR2 contains intron 2 from mouse *Hbb* between *Dnmt1* exon 3 and exon 4, and partially lacks exons 6 and 15 and all exons in between, translating into ΔS190-P369. B "Hbb" ΔROI also contains the *Hbb* intron 2 and lacks the ROI between S190 and D311. C "Hbb" ΔPIP carries the Q162E mutation within the PIP box.

siently expressed BFP-transposase. Within the cell population, many cells retained expression of mCherry, indicating the stable integration of the "Hbb" Dnmt1 construct. In order to enrich for strong expression of the transgene, we isolated the top 10 % mCherry+ cells. After cultivation for an additional 7 d, we again enriched for the top 10 % mCherry+ cells (Figure 4.16 D). While most cells had retained the mCherry signal 16 d post-transfection, some cells exhibited reduced expression of mCherry, indicating occasional silencing of the transgene. However, overall a clear shift in the signal intensity of the population is evident between day 9 and day 16. On day 16, only 10 % of the cells fall within the negative gate, which also include feeder cells contained within the sample.

After expanding the double-enriched cell population for 7 d, we isolated single clones and screened for mCherry fluorescence signal by the transgene using FACS (Figure 4.17 A, B). Compared to the clone C37 producing DNMT1 from a second generation expression vector using the codon-optimized cDNA, the mCherry fluorescence among the top six clones from each construct was on average ten times higher, and in some clones of "Hbb" ΔcIDR2 more than twenty times higher. These top six clones were expanded and subsequently treated with 500 nM dTAG-13 for 2 d to deplete the endogenous DNMT1 protein. The expression and efficient splicing of the transgene was characterized by qPCR and PCR, respectively (Figure 4.17 C, D). Among the selected clones, expression of *Dnmt1* and *mCherry* was within similar range. Of note, the fold increase (FI) for *Dnmt1* and *mCherry* is not directly comparable due to different primer efficiencies. Both endogenous and ectopic tran-

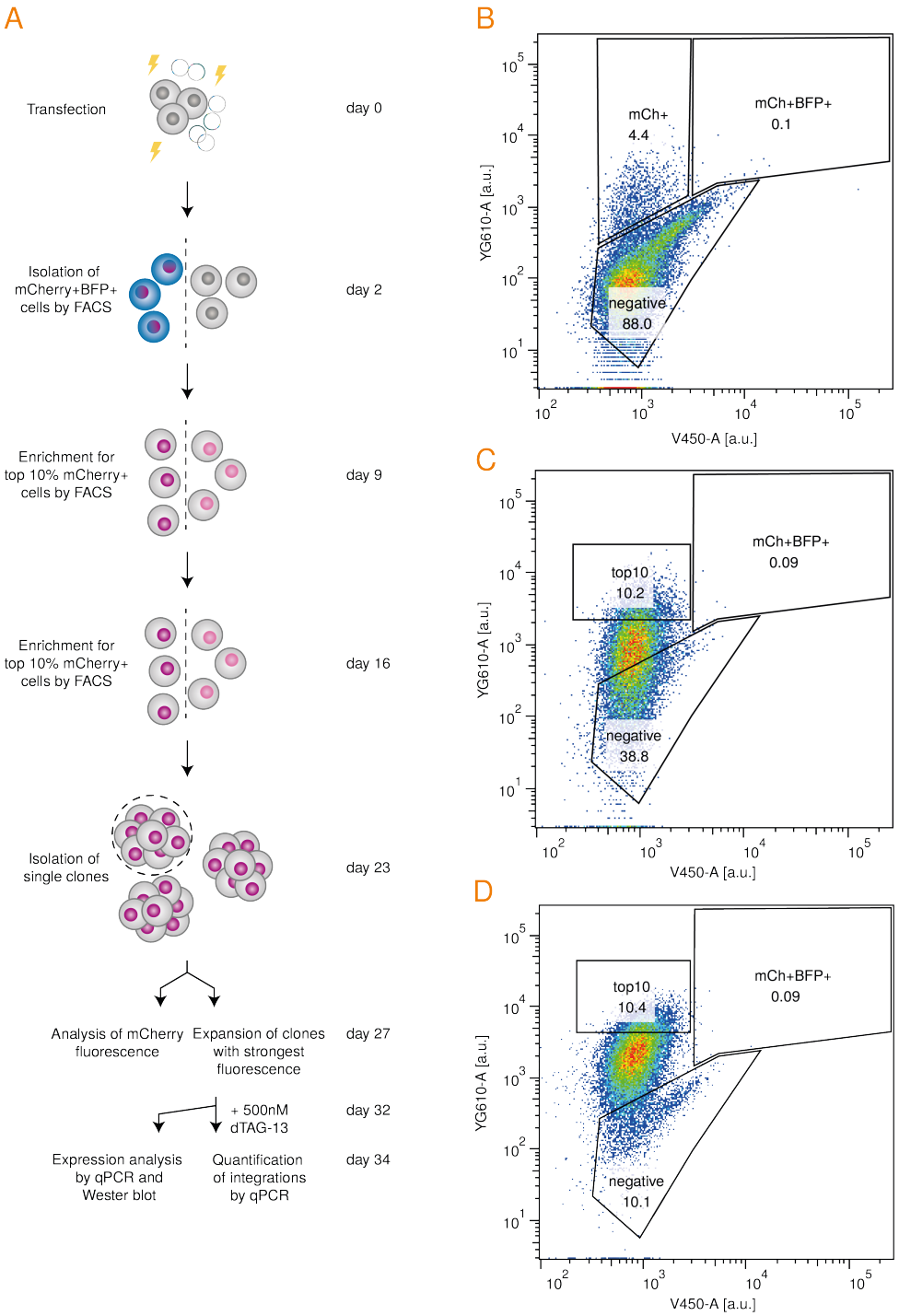

Fig. 4.16.: Derivation of "Hbb" Dnmt1 cells in the *Dnmt1* degron cell line C39. A Scheme illustrating the derivation and characterization of clones stably expressing "Hbb" Dnmt1. B Signal intensity of mCherry (YG610-A) over signal intensity of BFP (V450-A) 2 d post-transfection. Cells falling within the mCherry+BFP+ gate were propagated and enriched for the top 10 % mCherry+ cells (top10) C 9 d and again D 16 d post-transfection.

scripts of *Dnmt1* are measured. With the exception of "Hbb" ΔcIDR2 B2 and "Hbb" ΔROI G10, all clones expressed *Dnmt1* at least two-fold compared to the parent cell line C39. (Figure 4.17 C). We also counted the number of integrated copies of the transgene by qPCR. The clones have 15 integrations on average, ranging from four to 43 (Figure 4.17 D). Efficient splicing was determined by PCR on cDNA isolated from each clone (Figure 4.17 E). The primers were positioned in exon 3 and exon 4, just up- and downstream of the *Hbb* intron, respectively. Removal of the intron by splicing is reflected by a 188 bp long amplicon, while the amplicon would be 842 bp long if the intron was not spliced. With the exception of "Hbb" ΔcIDR2 C9 and "Hbb" ΔROI D12, all clones produce 188 bp amplicons, indicating complete removal of the intron from the transgene mRNA. For "Hbb" ΔcIDR2 C9 and "Hbb" ΔROI D12, there is neither an amplicon of 188 bp nor 842 bp, indicating that the cDNA concentration may have been too low in these two reactions.

We next determined the transgene protein levels by Western blot against DNMT1 (Figure 4.18 A). "Hbb" Dnmt1 and "Hbb" ΔPIP expressing clones show robust expression of the transgene at protein level. Unlike previous attempts to over-express DNMT1, the constructs containing the Hbb intron yielded near wt levels (compare Figures 4.8 C and 4.10 E). DNMT1 bands for "Hbb" ΔROI clones were slightly thinner than for "Hbb" Dnmt1 and "Hbb" ΔPIP clones; however, the loading control's signal was also slightly weaker. Hardly any signal was visible in "Hbb" ΔcIDR2 clones. When over-exposing the blot, bands of the correct size appeared; however, the protein levels were much lower than in the parental cell line C39 and the other mutant cell lines.

Since mCherry fluorescence signal and transgene expression at the mRNA level were similar between the "Hbb" Dnmt1, ΔcIDR2, ΔROI and ΔPIP clones, we wondered whether the signal using the anti-DNMT1 antibody against the ΔcIDR2 and ΔROI mutants could be affected by a full or partial lack of the antibody epitope. According to the manufacturer, the antibody is directed against a fairly long N-terminal sequence, approximately up to R260. ΔcIDR2 and ΔROI lack S190 to P369 and S190 to D311, respectively. Thus, up to 70 amino acids of the epitope may be missing in either construct. We therefore carried out a second Western blot, this time using an antibody against the N-terminal FLAG-tag of the ectopically expressed Dnmt1 constructs (Figure 4.18 B). Using this epitope, the fusion protein level in the ΔROI clones was more similar to that of ΔPIP clones, if not higher. Thus, the signal of the anti-DNMT1 antibody indeed seems to be diminished by the

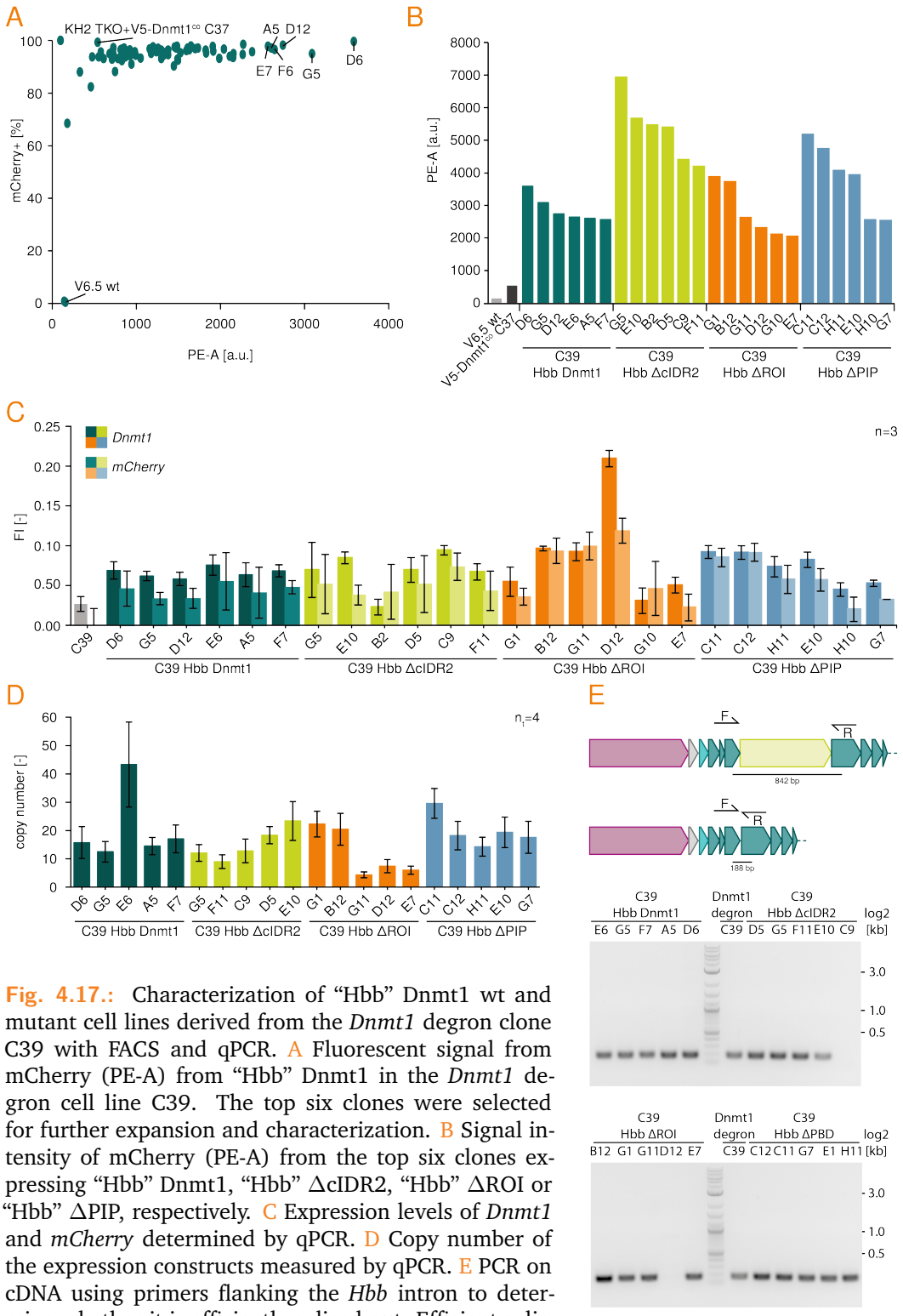

Fig. 4.17.: Characterization of "Hbb" Dnmt1 wt and mutant cell lines derived from the *Dnmt1* degron clone C39 with FACS and qPCR. A Fluorescent signal from mCherry (PE-A) from "Hbb" Dnmt1 in the *Dnmt1* degron cell line C39. The top six clones were selected for further expansion and characterization. B Signal intensity of mCherry (PE-A) from the top six clones expressing "Hbb" Dnmt1, "Hbb" ΔcIDR2, "Hbb" ΔROI or "Hbb" ΔPIP, respectively. C Expression levels of *Dnmt1* and *mCherry* determined by qPCR. D Copy number of the expression constructs measured by qPCR. E PCR on cDNA using primers flanking the *Hbb* intron to determine whether it is efficiently spliced out. Efficient splicing results in an amplicon of 188 bp, while the amplicon including the intron is 842 bp long.

deletion in the IDR. Nonetheless, the signal for the fusion protein in the ΔcIDR2 clones was much lower compared to clones expressing the full-length version of DNMT1. Thus, in addition to the mutant protein likely not being recognized efficiently by the anti-DNMT1 antibody, the level of the ΔcIDR2 fusion protein appears to be much lower despite the mCherry fluorescence signal and transcription level being similar to all the other over-expression lines.

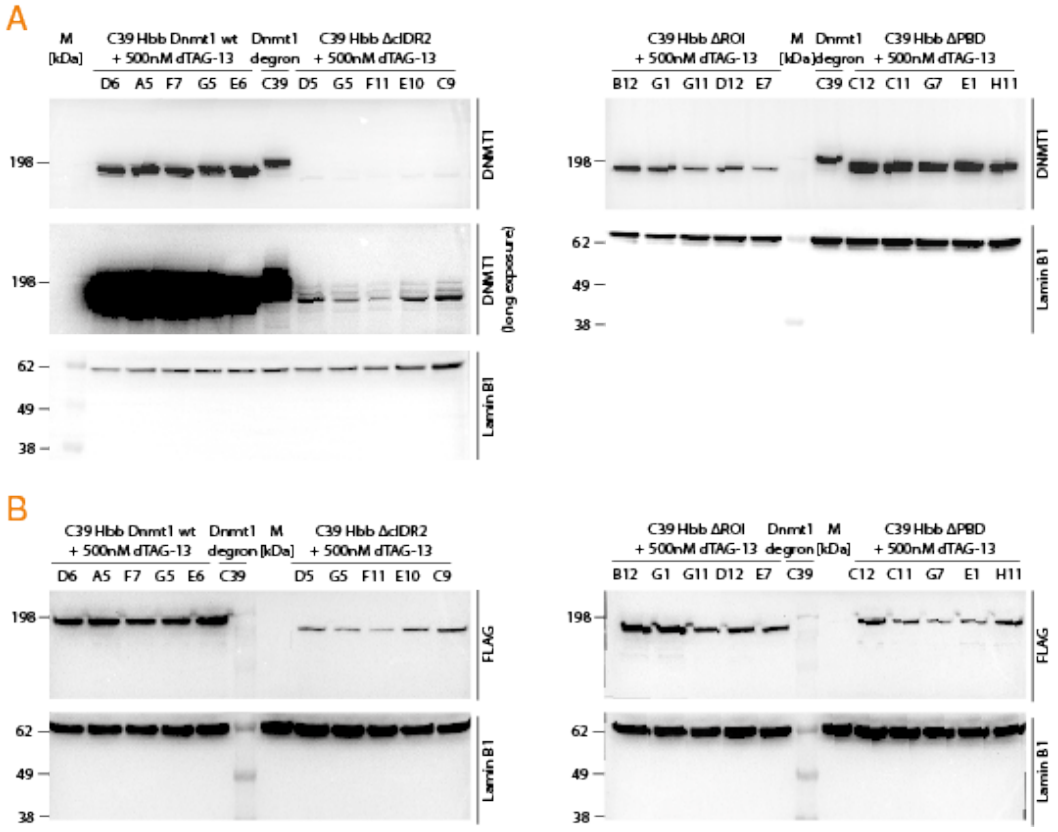

Fig. 4.18.: Western blot of DNMT1 and the FLAG epitope in "Hbb" Dnmt1 wt and mutant cell lines. A Western blot using an anti-DNMT1 antibody directed against the DNMT1 N-terminus. The middle panel on the left shows the blot in the upper panel additionally exposed for 14 s. B Western blot against the FLAG epitope at the N-terminus of the expression construct.

In summary, we were able to successfully establish a platform for the swift exchange of endogenous DNMT1 with mutant versions ectopically expressed at wt levels. Beyond optimizing the expression construct itself with respect to promoter-TSS distance, fusion protein length (Section 4.2), and native codon usage, the introduction of an intron and the sequential enrichment of highly expressing clones were key to reproducibly achieving robust transgene expression resulting in wt protein levels.

5

Discussion, conclusion and future directions

5.1 Overview and summary of results

In eukaryotes, complex patterns of DNA methylation result from the activity of DNMTs which is modulated by complex interaction networks. Studies of the spatio-temporal control of the DNMTs have been focused on the structured domains in the DNMT N-termini. These regulatory domains read post-translational histone tail modifications or interact with other nuclear proteins. Their interactions lead to preferential recruitment of the DNMTs to certain nuclear environments. For example, DNMT1 is brought to replication foci through its interaction with the replication processivity factor PCNA.

In addition to interactions following the "lock-and-key" fit of two molecules with defined topographies, recruitment of proteins driven by LLPS has emerged as an important mechanism for the spatial organization within the nucleus and the regulation of nuclear processes. LLPS is driven by multivalent interactions of proteins with other proteins, DNA or RNA. Proteins that undergo LLPS often form these multivalent interactions via unstructured IDRs. This includes examples of proteins involved in the read-out of 5mC, heterochromatin compaction and DNA replication. The N-terminus of DNMT1 also contains such an unstructured region; however, its function remains unknown.

This study explores a potential role of the DNMT1 IDR in shaping the enzyme's recruitment through LLPS. We find the length, degree of disorder and biochemical composition of the IDR to be conserved throughout evolution, substantiating

a possible functional role. In addition, we define a novel ROI within the DNMT1 IDR exclusive to extant eutherian mammals. For the first time, we demonstrate that the DNMT1 IDR by itself is capable of LLPS and IDR condensates are recruited to specific nuclear locations independent of the ROI and the PIP box. Lastly, we devise a platform to screen DNMT1 mutants for their phenotype in the context of maintenance methylation, to study the role of the IDR and the ROI in the future.

5.2 Conserved signatures of the DNMT1 IDR

As a basis for our study, we used the deep-learning-based prediction tool Metapredict to define the IDR within the DNMT1 N-terminus. Previous studies using the predictor DISOPRED2 located the IDR roughly between T100 and P400 [70, 129]. In comparison, Metapredict detects an IDR between L99 and P368 (Figure 2.1). Thus, the previous prediction extended too far into the N-terminus, which was unexpected since it has been found that DISOPRED2 tends to under-predict disorder [131]. Besides, Qin et al. (2015) identified a ubiquitin interaction motif (UIM) between A380 and T399 with 14 amino acids being highly conserved in species representing various classes of Metazoa (*i.a., Xenopus laevis, Danio rerio, Gallus gallus, Mus musculus*) [59]. Crystal structures show that the UIM assumes a helical shape, which is consistent with the structure prediction made by AlphaFold [60, 134]. This confirms our own prediction that the IDR of DNMT1 does not extend up to P400, but ends around P368.

Similar to DNMT1, Metapredict detects N-terminal IDRs also in the canonical *de novo* methyltransferases DNMT3A and -B. The IDRs have a predicted length of 278 and 223 amino acids, respectively, and are located in the proximal N-terminus just upstream of the PWWP domain (Figures 2.2, 2.3). Corroborating the IDR predictions, no crystal structure of DNMT1, DNMT3A and DNMT3B including amino acids that are part of these disordered regions has been solved. Due to their inherent structural flexibility, IDRs cannot be visualized in protein crystals used for conformational studies. All structures of DNMT1, DNMT3A and DNMT3B deposited in the Protein Data Bank and Uniprot have been carried out with isolated domains or protein fragments not including the IDRs [130, 138]. Thus, both the lack of crystal

structures and the computational predictions support the absence of a defined fold in the DNMT N-termini.

Importantly, we find the N-terminal IDRs to be present in all animal species queried in this study. However, the primary sequence shows much lower conservation compared to the folded domains. This is not surprising, given the lower evolutionary constraints on IDRs compared to domains with a defined fold. The function of structured domains is directly dependent on their correct surface topography that mediates interactions with binding partners. This requirement limits sequence variations within the primary sequence—and particularly of functional residues—of folded domains throughout evolution and has led to substantial sequence homology between conserved folds [163]. In contrast, the function of IDRs can by definition not depend on a fixed conformation. Accordingly, these evolutionary constraints do not apply to IDRs. Indeed, on a global scale, proteomes with a higher disorder content are overall less conserved [98]. Instead, the function of IDRs is defined by other molecular features such as length, net charge, hydrophobicity, and complexity [164]. While all of these features are encoded by the IDR's primary sequence, the exact order and identity of the amino acid residues play only a minor role in defining these properties, allowing for greater plasticity with respect to the addition, removal or substitution of amino acids. However, this does not imply that the maintenance of disordered protein regions is trivial, since excessive disorder may result in protein aggregation or degradation [165].

Importantly, our evolutionary analysis reveals the position and length of the DNMT IDRs to be overall conserved (Figures 2.4-2.7). Long IDRs as the ones identified here in the DNMTs are frequently found in N- and C-termini of proteins, likely because disorder at these positions can only interfere with ordered structures from one end, and not both as would be the case if the IDR was located in the center of the protein. IDR chain length is an important parameter that influences its potential LLPS behavior. At a fixed sequence composition, longer chains favor more multivalent interactions, therefore increasing the tendency to unblend from a homogeneous solution [166, 167]. Thus, it is interesting to note that the length of the DNMT IDRs is overall roughly conserved. This may indicate that this IDR length results in properties that are under selective pressure. An obvious exception is the DNMT1 IDR in eutherian mammals which is separately discussed in Section 5.4. In addition, the DNMT3B IDR also appears to be more variable in species other

than eutherian mammals, which may point towards a convergence of IDR length in eutherian mammals or simply reflect their closer evolutionary distance.

Besides directly influencing the physical conditions under which an IDR or protein will undergo LLPS, the IDR length between interaction motifs is also important. Interaction motifs and their spacing have co-evolved and may serve as a base for the functional selection of IDRs. This conformational buffering of interaction motifs allows sequence plasticity as long as the sequence remains disordered and the functional length is preserved [168]. Interaction motifs may be MoRFs, SLiMs or LCRs [84] (see Section 1.2.2). In the case of DNMT3A, the IDR harbors the UDR which can bind H2AK118/K119Ub. According to its length (69 amino acids) the UDR could be classified as a MoRF. These motifs adopt a secondary structure only upon binding their interaction partner. Yet, the primary sequence of the UDR is not disordered and AlphaFold predicts with low to medium confidence a helical conformation of the UDR. Thus, the UDR should be viewed as a short folded domain placed within an otherwise disordered structure.

The DNMT1 IDR is known to contain a SLiM, an interaction motif often found in IDRs that undergo LLPS driven by multivalent interactions [84]. The SLiM within the DNMT1 IDR is a PIP box motif, an archetypical SLiM mostly found in IDRs [169]. While AlphaFold does not predict any specific conformation for the PIP box, the structure of the DNMT1 PIP box bound to PCNA has been solved [170]. Thus, at least when bound to its interaction partner, the PIP box assumes a defined secondary structure that is similar to other PCNA-bound PIP motifs [169, 170]. PCNA assembles into a homotrimer that encircles the DNA and tethers DNA polymerase, an essential step in DNA replication. However, the function of PCNA is not limited to DNA replication as it has been shown to interact with more than 200 proteins involved in over 22 different vital activities, from chromatin remodeling to DNA repair and epigenetic modification. Many of these proteins contain PIP boxes and since PCNA is a homotrimer, in theory, up to three interaction partners can bind PCNA at the same time.

PCNA is a member of the DNA sliding clamps, a protein family which is functionally and structurally conserved across all branches of life. It has well conserved homologues in all eukaryotes and archaea [171–173]. Thus, it is not surprising that the PIP box is conserved in the DNMT1 N-terminus in all queried animal species (Figure 2.5). The interaction between PCNA and DNMT1 is not essential to maintain

DNA methylation, but it increases the efficiency of DNMT1 association with replication foci two-fold, particularly in early S-phase [56, 57]. Hence, the conserved DNMT1 IDR length could also serve to position the PIP box at a specific distance to surrounding motifs or domains. Following up on this, it would be interesting to test whether the recruitment via the PIP box is affected by varying the IDR length. Moreover, in this study, we did not systematically search the DNMT IDRs for novel interaction motifs. However, such an analysis would be helpful to further contextualize the conserved IDR length. In addition, a recent study has shown that the net charge of immediate PIP box flanking regions modulate its affinity to PCNA. Even though PIP box motifs are strongly conserved across evolution, the binding site does not appear to be overly specific on its own. Instead, positively charged residues in the surrounding region can increase the affinity to PCNA, by improving charge complementarity between the PIP box and PCNA [169]. Accordingly, traces of a selective pressure to modulate PCNA affinity could be identified by investigating if the selection for positive charges is reflected in the sequence variations of DNMT1 IDRs across evolution. Such a selection could result in a higher binding affinity to PCNA and thus a potentially more efficient recruitment of DNMT1 in early S-phase.

In addition to overall length and position, the disorder content of the IDRs is high irrespective of the exact primary sequences. As we show in Figure 2.8, the Metapredict consensus score and Alphafold confidence scores are uniformly high and low, respectively, along the entire length of the IDR when averaging the prediction for all species in our query. Of note, the disorder score for the DNMT3A and -B IDRs are, on average, slightly lower than those of the DNMT1 IDR and also exhibit greater variability across the species. This could indicate that, in contrast to DNMT1, the IDRs of DNMT3A and -B have continued to evolve while a very high degree disorder, that may be required for a certain function, has been fixed by natural selection in DNMT1 over longer evolutionary distances.

Focusing on the DNMT1 IDR in eutherian mammals, we furthermore analyzed the conservation of its amino acid composition and pI. Aromatic amino acids are scarce and make up less than 1 % of the residues within the IDR. In fact, the only consistently conserved aromatic amino acid is the phenylalanine that is part of the conserved PIP box motif (Figure 2.13). Moreover, hydrophobic residues are less frequent than on average in non-membrane bound proteins. This is typical for IDRs, since the drive of aromatic and hydrophobic residues to avoid contact with H_2O

molecules is an underlying force of domain folding [80, 84]. In contrast, acidic, basic and other non-charged hydrophilic amino acids are strongly over-represented (Figure 2.11). The overall positive charge of the region could promote the association with negatively charged molecules such as the nucleic acids. In the context of LLPS, charged IDRs typically form heterotypic condensates with RNA and DNA [80].

We also compared the amino acid composition of the murine DNMT1 IDR to that of the IDRs of other epigenetic regulators that are known to be active in similar compartments as DNMT1. Importantly, the DNMT1 IDRs of our four eutherian example species (mouse, human, rabbit, and pig) cluster more closely among themselves than with the IDRs of other proteins. This trend can also be observed for IDRs from the same protein, *e.g.*, the TETs or MBD1. Thus, intramolecular and interspecies differences in the amino acid composition of the IDRs are less pronounced than differences between proteins, at least in our selection of epigenetic regulators. However, since we hardly understand a possible "sorting code" that may depend on the properties of the IDRs, we cannot yet predict which IDRs may interact based on their amino acid composition [174].

5.3 Liquid-liquid phase separation by the DNMT1 IDR

Based on our computational analysis, we thus conclude that the structural and biochemical features of the DNMT1 IDR are compatible with a role in LLPS. To test this experimentally, we subjected the IDR to two orthogonal methods that can be used to determine whether the IDR can indeed undergo LLPS. The first assay uses the purified IDR to reconstitute a minimal assembly of critical components to test whether the IDR can unblend from the solution. Cues indicating that this condensation is indeed driven by LLPS are the formation of droplets with (i) spherical shape that (ii) fuse and wet surfaces, (iii) contain the IDR at higher concentrations than the light phase ($c_D > c_L$; see Section 1.2.1), and (iii) whose appearance depends on the protein concentration. Two additional strong indicators for LLPS are (iv) a constant c_{sat} and (v) the growth of the dense phase volume with increasing protein concentrations, which is not expected for other association processes not involving LLPS, such as dimer- and oligomerization [80].

In this study, we could demonstrate that the DNMT1 IDR forms condensates starting at 0.05 μM, which increase in size with increasing protein concentration. The droplets formed in our assay are spherical and fuse into larger droplets that eventually wet the surface of the cover slip, forming "lakes" (Figure 3.2 B). As we did not obtain a sufficient number of quantifiable objects at lower concentrations, we could not calculate the c_{sat}, but estimate it around 0.015 μM. The studies investigating LLPS of MeCP2, the DNA replication initiators or HP1α do not quantify the droplets at such low concentrations. However, at 2 μM, all these proteins robustly form droplets [103, 104, 107, 112, 116, 175]. Testing MeCP2, the studies of Wang et al. (2022) and Zhang et al. (2022) used a setup similar to ours, and the resulting MeCP2 droplets appear to be similar in size and volume fraction compared to our DNMT1 IDR construct at this concentration [112, 175]. Importantly, the results of these *in vitro* reconstitution assays do not lend themselves for quantitative comparisons, since the separation into light and dense phase as well as dense phase properties are highly sensitive to a plethora of factors, such as salt concentration, the type and concentration of crowders and temperature [80]. Even small inaccuracies in protein concentration measurements may impact the interpretation of these assays. For example, we determined the protein concentration of our samples using the Nanodrop spectrophotometer with an mCherry standard curve, which is only moderately accurate. A systematic error in the standard curve could easily result in precise but inaccurate measurements of our protein concentration. Thus, within the same experiment, the behavior of the reconstituted systems is comparable but should not be directly compared with other publications even if the conditions were somewhat similar. Nonetheless, in our hands, the droplet formation of the DNMT1 IDR fulfills all five criteria characterizing condensate formation by LLPS. On this basis, we conclude that the DNMT1 IDR phase separates *in vitro*.

We could corroborate this finding in live cells using the optoDroplet assay. Expressed in HEK293T cells, the murine DNMT1 IDR readily forms condensates, even without Cry2, which normally serves to nucleate condensates upon blue light stimulation (Figures 3.4, 3.6, 3.16). When exposed to the 488 nm laser, the "opto-IDR" fraction that is initially diffusely distributed within the nucleoplasm is quickly recruited into the pre-existing speckles. Our quantification shows that this behavior depends on the overall construct concentration within the nucleus. In cells expressing "opto-IDR" at weak to medium levels (1^{st} to 3^{rd} quartile), nucleoplasmic fusion protein is concentrated into droplets within the first six cycles, whereas no change in mean normalized standard deviation is observed in cells that express the most

"opto-IDR" (4^{th} quartile). This could indicate that the higher the protein concentration, the stronger the tendency for assembly within a dense phase, which is in general agreement with the fundamental properties of phase separated systems.

Contrasting these results to the behavior of "opto-MTase" containing an ordered sequence of similar length, we can demonstrate that the behavior of the DNMT1 IDR does not simply depend on the construct length but also requires the presence of a disordered sequence (Figure 3.5). Moreover, the IDRs of DNMT3A and -B are also disordered, but each exhibit distinct condensation behavior (Figure 3.11). Of note, Gu et al. (2022) demonstrated that in their hands, the DNMT3A IDR also forms condensates in *in vitro* reconstituted systems as well as the OptoDroplet assay [50]. Eventually, the authors concluded that DNMT3A does not form liquid condensates *in vivo*, since the fluorescent signal within condensates formed in DNMT3A1-over-expressing cells did not recover after bleaching. However, as discussed below, this conclusion might be premature since at high levels in over-expressing cells even condensates of proteins that at physiological concentrations form liquid-like condensates may be driven into a less dynamic gel-like state (see below). Nonetheless, the phenotypes of the optoDroplet constructs are clearly dependent on the identity of IDR.

Conversely, we show that part of the IDR sequence that is unique to eutherian DNMT1 (ROI) does not form foci when expressed as optoDroplet construct, despite having approximately the same length as the more ancient part of the IDR, which does accumulate in condensates (see also Section 5.4). In addition, we show that the proximal and distal part of the IDR conserved in all animal species (cIDR1 and cIDR2, respectively) also do not form condensates alone and require all ten cycles of blue light stimulation to weakly nucleate (Figure 3.15). Combined, these results show that disorder alone is also not sufficient to form condensates. Using the same setup with the human sequence yielded similar results to the ones obtained using the murine sequence (Figure 3.10). This further validates that not the exact primary sequence, but the structural and biochemical features which are conserved across the species, determine the LLPS behavior of the IDRs.

Intriguingly, the "opto-IDR" condensates are not randomly located within the nucleus but reproducibly form specific patterns, which are similar in adjacent cells (Figure 3.9). Since adjacent cells are likely to be daughter cells and be in the same cell cycle, this observation leads us to hypothesize that the localization of

"opto-IDR" condensates is cell-cycle dependent. Indeed, the observed patterns of "opto-IDR" and "opto-ΔROI" are reminiscent of but distinct from the DNMT1 distribution reported during different stages of S-phase, where DNMT1 accumulates in many small, disperse puncta in early S-phase and is detected within few ring-shaped structures or large patches in late S-phase [176, 177]. To test the cell-cycle dependency of condensate pattern, one could track transfected cells over several hours to observe one cell cycle, but HEK293T cells are quite mobile, making this technically challenging. Alternatively, the condensate pattern could be analyzed in fixed DAPI-stained cells in a high-throughput manner. Using the DAPI signal as a proxy for DNA content, the cells could be classified according to cell cycle phase, similar to cell cycle analyses using flow cytometry. In a second step, the condensate patterns could be classified and overlaid with the cell cycle of each individual cell. However, due to the size of the condensates, classification of their pattern requires high-resolution (confocal) microscopy, which is challenging to set up as a high-throughput system, especially because HEK293T cells produce vast amounts of extra-cellular matrix and usually do not grow in one focal plane. Thus, either approach will likely benefit from a change in cell type that is more amiable to high-throughput imaging and/or live cell tracking. As a third alternative—or perhaps complementary—approach, we could use a PCNA-reporter cell line to identify cells in S-phase, since PCNA forms specific nuclear patterns in early, middle and late S-phase that can be used for cell-cycle classification [178].

Of note, if some of the condensate locations indeed represent sites of replication, this would mean that the DNMT1 IDR can associate with these sites independently of the PIP box and its association with PCNA, since condensates of "opto-ΔPIP" also form the patterns observed with "opto-IDR". Excitingly, this would imply that the IDR may be recruited to replication foci by virtue of its phase separation behavior only and may have evolved as an additional or redundant mean to efficiently recruit DNMT1 to the sites of its main activity, namely maintenance methylation. However, prior to experimental confirmation of a PIP box-independent association of the IDR with replication foci, *e.g.*, by the above mentioned approaches, this mechanism is entirely speculative.

In addition to dissecting the role of the PIP box in the localization of the condensates in more detail, it would be informative to stain "opto-IDR" cells against various binding partners to determine whether other proteins are localized to the optoDroplet condensates. Lastly, a shortcoming of this study is that we did not test the liquid

nature of condensates in live cells, *e.g.*, using Fluorescence Recovery After Bleaching (FRAP). If the condensates are liquid, protein is continuously exchanged with the mobile fraction in the nucleoplasm. Thus, after the fluorescent signal of a condensate has been bleached, the signal recovers within a few minutes, depending on the rate of diffusion, since unbleached "opto-IDR" from other sites is swapped against the bleached molecules over time. This analysis can be considered as a first test of a liquid-like state, but it is important to keep in mind that "opto-IDR" is present within the cells at supranatural concentrations at which the state of the system could be driven into more gel-like states. Thus, the absence of recovery after bleaching within the context of the optoDroplet setup, or any type of overexpression for that matter, does not necessarily mean that the IDR forms gel-like or even solid condensates also at physiological concentrations. FRAP experiments on naturally occurring condensates would be therefore more meaningful.

Its conserved amino acid composition furthermore predicts that the IDR would form heterotypic coacervates and bind negatively charged DNA. Indeed, we find the DNMT1 IDR to associate with DNA, in particular in metaphase cells where the "optoIDR" fusion protein entirely stains the mitotic chromosomes, while "opto-mCherry", "opto-p15paf" and "opto-Dnmt3b" do not (Figure 3.14). *In vitro*, this could be further experimentally validated by reconstitution of the IDR with DNA. Using the same experimental setup, it could be tested whether the presence of DNA promotes LLPS by the DNMT1 IDR (*i.e.* by lowering c_{sat}). Using a DNA stain such as DAPI, it could be further determined whether DNA partitions into DNMT1 IDR droplets. LLPS by the DNA methylation reader MeCP2 and the DNA replication initiators Orc1, Cdc6 and Cdt1 is promoted in the presence of DNA [107, 110, 112, 113, 116, 118].

Interestingly, the growth of MeCP2 droplets *in vitro* is influenced by DNA methylation, too. There are conflicting data whether DNA methylation promotes or inhibits the growth of MeCP2 droplets *in vitro*, but growth of MeCP2 condensates *in vivo* appears to be restricted by the presence of DNA methylation [107, 112]. In contrast to DNMT1, MeCP2 has a dedicated domain for binding methylated DNA. Hence, the association of DNMT1 with methylated DNA is likely to be more transient. Nonetheless, methylation of DNA could modulate the interaction between the DNMT1 IDR and DNA, thus influencing coacervation. To reconstitute their system with methylated DNA, the above-mentioned studies on MeCP2 used PCR amplicons from plasmids or DNA oligomers methylated *in vitro* with M.SssI, a bacterial restric-

tion enzyme that methylates CpGs. This methylation reaction, however, is very inefficient and hardly leads to complete methylation. Above all, the CpG frequency of bacterial plasmid DNA does not reflect the CpG density in mammalian genomes. In addition, care should be taken to use long DNA molecules that accurately reflect the conditions *in vivo*. Yet, these caveats could be overcome by shearing genomic DNA from wt V6.5 mESCs and KH2 TKO to a uniform size. This would result in DNA fragments with clearly defined levels of 5mC and an appropriate distribution.

To start investigating whether the composition of the IDR droplets observed in our assays would be compatible with the interaction profile of DNMT1, we tested whether IDR condensates can partition P15PAF, a known interaction partner at the replication fork, in live cells using the optoDroplet assay (Figure 3.14 A). When co-transfected, the signal of "opto-IDR" and "opto-p15paf" clearly overlapped, demonstrating that the two proteins cannot only partition into the same condensates but also co-localize. To conclude, we could show that the known interaction partner P15PAF can diffuse into DNMT1 IDR condensates, demonstrating that the dense phase allows other proteins to enter. This finding should be further corroborated by testing the ability of the DNMT1 IDR to also partition other known DNMT1 interaction partners.

With respect to both the *in vitro* reconstitution and the optoDroplet assay, an important caveat is that we do not know the physiological concentration of DNMT1 in the nucleus. Thus, we cannot assert whether c_{sat} is reached under normal conditions. As a first step, this would require the determination of the DNMT1 concentration in the nucleus which is complicated by the redundant mechanisms that recruit DNMT1 to the replication foci during S-phase. Since these processes locally concentrate DNMT1, the measurement would need to be made at these sites. Zhang et al. (2022) used a combination of Western blot, FACS analysis and 3D-fluorescent microscopy of cells over-expressing GFP-MeCP2 to estimate the nuclear local concentrations [112]. A similar approach would be feasible for measuring the physiological, local concentration of DNMT1 during S-phase.

Nonetheless, we conclude that our results *in vitro* and *in vivo* form a strong basis indicating that LLPS is a relevant mechanism for DNMT1 behavior. However, we only tested the propensity of the IDR alone to form a separate phase. To reach a more definite conclusion, it would need to be demonstrated that also the entire protein becomes part of a separate dense phase. This could be attempted with the *in*

vitro reconstitution assay. However, DNMT1 is a 190 kDa protein whose production requires the use of insect cells and an optimized purification setup to obtain the required amounts [60, 62, 66]. Likewise, the optoDroplet assay could be adapted to constructs with the full-length cDNA, but transfection and expression of such a large protein are non-trivial as demonstrated and discussed in Sections 4.2-4.3 and 5.5, respectively. Overall, techniques to identify and characterize liquid-like assemblies *in vivo* are still limited and can often not determine whether a structure has formed via LLPS or another recruitment mechanism such as canonical direct domain interactions. Although more advanced non-invasive techniques like super-resolution or Brillouin microscopy are being developed to better probe biomolecular condensates and their properties *in vivo* [80], ultimately, loss- and gain-of-function studies are required to demonstrate that the phase separation ability is relevant for protein function. These studies require careful design as to dissect the influence of LLPS without disturbing other properties. While residues required for LLPS can be identified using *in vitro* assays, the most definitive proof of an effect of LLPS on protein function is achieved *in vivo* under physiological conditions by replacing the endogenous protein with mutants.

Which processes involving DNMT1 could benefit from the enzyme undergoing LLPS? First of all, several *in vitro* biochemical studies have proposed that DNMT1 moves along the DNA processively by linear diffusion [179–182]. However, it remains elusive how this processive movement comes about. One possible mechanism could be the spreading of a dense phase along the chromatin fiber which would require LLPS.

Second, in its natural function DNMT1 requires to enter many phase separated compartments, for example, heterochromatin [103, 104]. Although LLPS has so far only been demonstrated to play a role in the initiation of DNA replication, it is reasonable to consider LLPS also as a potential process to mediate other steps of DNA replication, making the replisome a candidate phase separated environment [116, 118]. Moreover, a separated phase is thought to condense at sites of DNA damage, which are also visited by DNMT1 [146, 148]. While DNMT1 likely does not act as an initiation scaffold for these compartments, it may require to undergo LLPS to gain access to these phases as a client [79].

Lastly, local concentration of DNMT1 could improve the efficiency of maintenance methylation. By interacting with PCNA, P15PAF, and H3K18/K23Ub, DNMT1 is

recruited to the replication fork, where DNA polymerase synthesizes the unmethylated daughter strand that is the main target of DNMT1 [56–58, 62]. DNA polymerase moves at about $0.02\,\mathrm{s\,bp^{-1}}$, whereas DNMT1 requires 70 s to 450 s per methyl group transfer [183, 184]. Thus, the speed of DNA replication and maintenance methylation differ by three orders of magnitude. In support of this, we and others have shown that maintenance methylation starts to lag behind DNA replication early on and requires additional time in G2 phase to fully methylated the nascent DNA [185, 186]. Hence, we can envision that maintenance methylation would greatly benefit from a process such as LLPS that promotes high concentrations of DNMT1 at the replication fork and ensures a continuous supply of DNMT1 to the emerging nascent strand. On the one hand, this is clearly a requirement since at least three other partially redundant mechanisms have evolved to recruit DNMT1 to replication foci [56–58, 62]. On the other hand, these could compensate a defect in LLPS and complicate the identification of a functional role of LLPS. However, if this is not the case, steady-state methylation levels in the absence of LLPS may decline over time which would be detectable by methylation analysis of cell populations. Hence, we set out to develop a system which can be used to study the function if DNMT1 mutants in the context of maintenance methylation, which we discuss in Section 5.5.

5.4 Extension of the DNMT1 IDR in Placentalia and possible functional implications

Highlighting the greater evolutionary plasticity of disordered regions, we identified a segment of approximately 120 amino acids within the DNMT1 IDR that is exclusively present in species belonging to the infraclass Placentalia. This novel ROI inserted downstream of the PIP box and extends the IDR conserved in the other animal species by almost 200 %. A previous report noted that a region encompassing about 160 amino acids is present in mouse, human and cow, but not chicken and zebra fish, and thus concluded that it may have a mammal-specific role, such as imprinting [70]. However, our analysis uncovers that the novel extension is only present in Placentalia and has thus likely evolved after the divergence of Metatheria (including Marsupiala, such as koala and wombat) and Eutheria (including Placentalia, such as mouse, pig, and human) 148 Myr ago [187].

We compared the biochemical composition of the ROI among mouse, human, pig and rabbit and found that in all four species the ROI is particularly rich in proline and charged residues (glutamic acid, lysine and arginine), which appear in clusters (Figure 2.11, 2.13). Compared to the full-length IDR, the ROI is richer in acidic residues, lowering the pI of the ROI (between 4.6 and 5.3) relative to that of the full-length IDR (between 6.4 and 9.6). Thus, the insertion of the ROI may have significantly lowered the net charge of the IDR. To that end, it would be informative to compare the amino acid composition of the IDR conserved in all species to the extended IDR in eutherian mammals.

Net charge, acidic residue content and pI have been identified among several molecular features that appear to be under selection in IDRs [164]. Due to the change in IDR length and additional charges, we hypothesized that the ROI may influence the tendency of the IDR to undergo LLPS. To this end, we compared the phenotype of the full-length IDR with the isolated ROI as well as the IDR missing the ROI (ΔROI). Keeping the limitations of the two employed assays in mind (Section 5.3), neither our *in vitro* reconstitution assay nor our opto-genetic approach revealed any differences between the behavior of the full-length IDR and the ΔROI constructs. Both the complete IDR and ΔROI formed condensates *in vitro* within the same range of concentrations. At all concentrations, IDR and ΔROI condensates were similarly sized and numerous (Figure 3.2). In live cells, IDR and ΔROI alike preformed speckles that exhibited indistinguishable nuclear localization in interphase and metaphase cells (Figures 3.4, 3.6, 3.9, 3.10, 3.14), recruited mobile nucleoplasmic mCherry fusion constructs upon blue light stimulation with similar dynamics (Figures 3.7, 3.8) and overlapped with P15PAF condensates (Figures 3.12). The only difference in behavior was noted in co-transfection of "opto-ΔROI" with "opto-p15paf", where "opto-ΔROI" condensates associated with the nuclear lamina and presumably the surface of the nucleoli remained free of "opto-p15paf", while "opto-IDR" and "opto-p15paf" also overlapped at these locations (Figures 3.12). Hence, it remains a possibility that the extension by ROI fine-tunes the LLPS behavior with respect to the composition of condensates in specific contexts. To further investigate the interaction between the DNMT1 IDR and the ROI with P15PAF, co-partitioning could be also tested using the *in vitro* reconstitution assay. Lastly, the ROI formed condensates at only the highest concentrations *in vitro* and not at all in the context of the optoDroplet assay. Thus, we conclude that the ROI—despite being disordered and of similar length compared to the conserved IDR—cannot form a condensed phase by itself. This in turn also validates our findings with respect to the

full-length IDR and ΔROI sequence, since it demonstrates that not simply every disordered sequence undergoes LLPS but must have additional features that support this process.

Irrespective of a potential influence on the LLPS behavior of DNMT1, it is attractive to speculate that the emergence of the ROI is in some way linked to the evolution of a feature unique to Placentalia. Interestingly, marsupials appear to have generally lower levels of DNA methylation compared to eutherian species, with methylation on autosomes mostly concentrated at telo- and centromeres [30, 188]. Thus, we can hypothesize that one function of the IDR extended by the ROI in eutherian mammals could be an increase in the efficiency of maintenance methylation to retain the high steady-state levels of eutherian methylomes. Methylation analysis of the ΔROI mutant in our DNMT1 degron cell line will shed light on the role of the ROI in maintaining the bimodal somatic methylation landscape. Of note, other metazoan classes such as reptiles, fish and amphibians also have high levels of methylation in somatic cells [30]. However, it is possible that these species evolved distinct mechanisms for efficient maintenance that did not involve innovations in the DNMT1 structure.

Since detailed methylomes of marsupials have not been published, we cannot search them for more subtle differences between meta- and eutherian methylation landscapes that may relate to the IDR extension. However, when looking beyond the steady-state somatic methylome, we can consider alternative or additional functions for the ROI, *e.g.* during embryonic development. A first feature that may come to mind when thinking of differences that distinguish the embryonic development of Marsupalia and Placentalia is the placenta. Yet, as briefly mentioned in Section 2.2, placentation is not exclusive to the infraclass Placentalia despite their name. Indeed, when broadly defining the placenta as a fusion of fetal membranes with the uterine mucosa to enable physiological exchange, placentation evolved multiple times as an adaptation to the transition from ovipatry to vivipatry in all vertebrate branches but birds. Alternatively, the definition of the placenta can be made more narrow as a structure at the maternal-fetal interface derived from the trophectoderm. This cell type is formed as the result of the first lineage decision made during embryonic development in Placentalia, when the trophectoderm segregates from the other pluripotent cells that will develop into the embryo proper. However, the trophectoderm evolved in the therian ancestors prior to the split between Metatheria and Eutheria. In both clades it develops into a fully functional

placenta that is essential for embryonic development [137]. Hence, the name "placental mammals" to refer to the eutherian infraclass Placentalia is misleading and is not used in this study.

With respect to DNA methylation, the placental precursors derived from the trophectoderm are particularly interesting since they adopt a methylation pattern distinct from the canonical bimodal landscape established in the embryo around the time of implantation [31]. Numerous studies have noted that placental tissues in mouse and human are globally less methylated than the embryonic counterparts [189–191]. Specifically, partially methylated domains (PMDs) exhibit markedly lower levels of 5mC in the extra-embryonic tissues than in the embryonic tissues, where these megabase-scale domains are hardly detectable [192, 193]. PMDs are best correlated with lamina-associated domains and late DNA replication. Since PMDs are also prominent in most cancer types and aged cells, the prevailing hypothesis for the formation of PMDs is that they result from a lack of efficient maintenance methylation by DNMT1 towards the end of the cell cycle, leading to the passive erosion of 5mC over subsequent mitoses [193].

However, like the development of the placenta from the trophectoderm, its global undermethylation is not unique to extant eutherian mammals—as it has been found in several primates, dog, horse, and cow—but also in the extra-embryonic tissues of opossum, a marsupial. This finding makes a role of the DNMT1 ROI in the bifurcation of the methylome into clear PMDs and HMDs unlikely [194]. Nonetheless, due to a lack of high-resolution data, our picture of 5mC distribution in marsupial extra-embryonic tissues remains incomplete at best. Thus, more subtle difference between the placental methylome of meta- and eutherian mammals may exist that have so far escaped our attention.

Besides the development of the trophectoderm, the evolution of mammalian placentation has been accompanied by the emergence of genomic imprinting. Imprinted genes are specifically expressed from either the maternal or paternal allele, resulting in functional haploidy. Imprinted genes have functions in placentation, metabolism, fetal and post-partum growth as well as the weaning and adult social behavior [195]. Thus, all theories to explain the selective advantage of imprinting are centered around the maternal-offspring relationship [196]. In many cases, imprinted genes expressed from the paternal allele promote growth, while genes expressed only from the maternal allele restrict growth [197]. This observation forms

the basis of the most popular theory of imprinting evolution, the conflict/kinship theory, which proposes imprinting as a means of the parental genomes to influence the allocation of resources between the mother and the offspring [196].

Imprinting evolved after the monotreme-therian split about 166 Myr and was likely an adoption of epigenetic defense mechanisms against foreign DNA insertions. In the common eutherian model organisms, over 150 imprinted genes have been identified to date. Of those 150 genes, 30 orthologues have been investigated in marsupials, but only 19 were found to be present and of those, only six are imprinted. Thus, it appears that imprinting initially evolved in the therian ancestor but diversified quickly prior to the eutherian radiation. This may indicate an intensification of the resource allocation conflict during the prolonged gestation which evolved at the same time. Concurrently, retroviral elements further expanded in the eutherian branch and may have provided ample opportunities for the adoption of beneficial silencing events as imprints [196].

In marsupials, imprinting of two loci (*Peg10* and *Igf2/H19*) relies on parent-of-origin specific DNA methylation. In extant eutherian mammals, the paternal and maternal allele at all known canonical imprints are distinguished by differential DNA methylation at a defined regulatory region, the ICR. Most imprinted genes are located in clusters, each controlled by one ICR [196]. So far, 25 ICRs have been identified, 22 maternal ones and three paternal ones [198]. The difference in methylation at the ICR is a consequence of the distinct methylation patterns established in ova and sperm, where the region is methylated in maternal germline but not the paternal, or vice versa (see Section 1.1.3). While the global differences between the maternal and paternal methylome are erased during a wave of epigenetic reprogramming during preimplantation development, ICR methylation is maintained by DNMT1 and the allele-specific differences persist into adulthood [44]. How ICRs are selected for protection against the global demethylation is hardly understood. The majority of ICRs contain motifs recognized by ZFP57 that binds to the methylated alleles and maternal-zygotic depletion of ZFP57 leads to loss of methylation at these ICRs [46, 199]. ZFP57 recruits TRIM28, an important regulator of repressive chromatin environments which is also required for the maintenance of imprints [48].

Interestingly, some imprints in embryos lacking ZFP57 retain methylation either partially or even completely. Of further note, all methylation imprints found in mar-

supials fall into this category. This not only points towards at least one alternative mechanism for the retention of methylation at ICRs that evolved prior to the divergence of Marsupalia and Placentalia. It also indicates that imprint maintenance by ZFP57 is exclusive to Placentalia. Indeed, ZFP57 is not present in marsupial genomes and must have evolved in the eutherian ancestors. Thus, despite imprinting *per se* being used as a dosage-compensation mechanism in both Marsupalia and Placentalia, most DNA methylation imprints are only found in Placentalia and rely on a mechanism exclusive to this infraclass [196]. Intriguingly, it remains elusive how ZFP57 and TRIM28 are involved in the recognition of ICRs by DNMT1. Thus, it would be worthwhile investigating whether the insertion of the ROI within the DNMT1 IDR was coincidental or co-evolved with ZFP57-mediated imprinting.

One way to test a role for the ROI in the maintenance of these imprints would be to derive mice from mESCs expressing a DNMT1 mutant lacking the ROI and profiling the methylation status of the ICRs after methylation has been reset during embryonic development, *e.g.*, using hybrid-capture bisulfite sequencing. Alternatively, a more time efficient setup could be the assessment of ICR methylation status in late blastocysts derived from DNMT1 mutant mESCs using somatic cell nuclear transfer (SCNT). Like embryos derived from fertilized oocytes, embryos cloned using SCNT undergo epigenetic reprogramming during pre-implantation development. By the time the blastocyst would implant, ICRs would have been lost in DNMT1 mutant embryos if the ROI was essential for their maintenance. However, this second method has two caveats. (i) DNMT1 is a maternal-effect protein and oocyte-derived DNMT1O is known to maintain imprints at least during the first cleavages [44]. Therefore, a potential phenotype of ROI-less DNMT1 may be partially rescued by the presence of DNMT1O from the oocyte. Because DNMT1O is present in the cytoplasm of oocytes, enucleation would not deplete the maternal stores. Hence, a potential solution could be to use donor oocytes devoid of DNMT1O [200]. However, some studies have shown that imprints are partially lost in embryos lacking a maternal supply of DNMT1O, perhaps because DNMT1O associated with the maternal chromosomes is essential [44, 201, 202]. (ii) SCNT is an inefficient method and only few cloned embryos develop. This limitation has been mainly attributed to the delayed and incomplete epigenetic reprogramming in SCNT embryos [203–205]. Conversely, imprints are often not properly maintained in cloned embryos, but the underlying reasons are not clear [206]. Thus, both incomplete erasure and error-prone maintenance of imprints could confound the analysis of DNMT1 mutant phenotypes using this method.

Nonetheless, using either approach, it would not only be possible to study a potential function of the ROI on imprinting. Other mutants could be tested to investigate a more general role of the IDR in embryonic development. Yet, the mutant cell lines that can be derived using the platform optimized and characterized in this study (see Section 4) would not be suitable to study the role of DNMT1 mutants during embryonic development. This is because the depletion of the endogenous protein requires continuous treatment with dTAG-13, which may not reach all cells within larger embryos at sufficient concentrations and it would be challenging to tell whether the endogenous protein is sufficiently depleted. This could be circumvented by irreversibly depleting the endogenous enzyme using Cas9-mediate KO, if it turns out that the mutants can maintain a fully methylated genome in the steady state.

Besides, the constitutive over-expression of DNMT1 mutants could pose an issue because naturally nuclear DNMT1 levels in the early cleavage blastocyst are extremely low. In fact, hardly any protein can be detected using conventional methods, which has even sparked a discussion spanning more than two decades whether maternal DNMT1O, zygotic DNMT1S, both or neither maintain imprints in 2-cell to 8-cell embryos [40, 44, 52, 202, 207–210]. On the one hand, particularly during this stage of low protein abundance, DNMT1 could benefit from specialized recruiting mechanisms through the IDR or the ROI to find the few targets to be maintained throughout the genome. On the other hand, constitutive over-expression at wt levels from our transposon constructs could rescue any mutant phenotype from less efficient recruitment since the protein is present at much higher levels than would naturally occur in the early embryo.

Hence, the investigation of the ROI function—or the function of the IDR more generally—during preimplantation development would require the establishment of mESCs with endogenous mutants. These can be generated by removing genic regions encoding the ROI or an extended sequence. In the case of the ROI, this would require the removal of exons 5 to 12. This region spans about 11.5 kb, including introns, which is a deletion that can be achieved with reasonable efficiency using Cas9-mediated genome editing. In-frame fusion of the first and last amino acids encoded by exons 5 and 12 could be guided by supplying a donor template, a strategy analogous to Cas9-mediated knock-ins as described in Section 4.1.

5.5 Development of a screening platform for DNMT1 mutants

So far, we have concluded that the continued evolutionary conservation of the DNMT1 IDR's molecular features is a strong indication for a conserved function. The structural and biochemical signatures as well as our *in vitro* and *in vivo* assays confirm a possible role in LLPS, a process that may serve to locally concentrate DNMT1 to enhance the efficiency of maintenance methylation and/or mediate its access to other phase separated compartments as a client. Alternatively, the IDR and specifically the ROI in eutherian mammals could also have LLPS-independent functions regulating the recruitment of DNMT1 to its targets. Either way, the study of IDR and ROI mutants could reveal a loss-of-function phenotype as a cue towards the biological function of these disordered sequences.

To start, we thus set out to investigate potential loss-of-function phenotypes in the context of maintenance methylation in mESCs. To exchange the endogenous enzyme with a mutant version, we derived the monoclonal cell line C39 in which DNMT1 is fused to the fluorophore mCerulean and an $FKBP12^{F36V}$ degron at its N-terminus (Figures 4.1, 4.2). We could demonstrate that the fusion protein is stable and present at wt levels, but quickly and efficiently degraded upon addition of the small molecule drug dTAG-13 (Figure 4.3 A). Following the degradation of the fusion protein over longer periods of time, we noticed weak but persistent residual mCerulean fluorescent signal (Figure 4.3 C). This remaining signal could have two sources: (i) during the proteasomal degradation process, some mCerulean could become separated from the $FKBP12^{F36V}$-DNMT1 and remain undegraded in the cell, and (ii) not all fusion protein can be degraded due to the Hook effect. This effect applies to the docking of ternary complexes, in this case the mCerulean-$FKBP12^{F36V}$-DNMT1 fusion protein, dTAG-13 and Cereblon. Both the fusion protein and Cereblon can bind free dTAG-13 so that at times a number of proteins cannot undergo complex formation since the protein binding pockets are already occupied. This is especially the case when the heterobifunctional linker is present at high concentrations, but statistically even at low concentrations, some protein will remain which we may detect using FACS [211, 212]. Since we also detect residual DNMT1 activity (discussed in Section 5.6), this second explanation likely applies but both options are not mutually exclusive.

Nonetheless, our results show that these residual levels are not sufficient to maintain DNA methylation. Within 1 d of dTAG-13 treatment CpG methylation measured by RRBS is halved, corresponding with the passive mode of demethylation resulting from DNA replication in the absence of maintenance methylation (Figure 4.3 F). After 7 d, methylation levels stabilize at around 20 %, since DNMT3A and -B are still active and continuously methylate the DNA *de novo* [157, 158]. Hence, keeping the final application in mind, we will be able to detect methylation loss as a consequence of DNMT1 loss-of-function phenotypes down to 20 %. Below this, any residual activity will be masked by the *de novo* activity of DNMT3A and -B. Upon recovery, wt methylation levels are restored, which was expected since DNMT1 fusion protein also returns to wt levels (Figure 4.3 E, F). An exception are CGIs within ICRs which do not recover their intermediate methylation (since only one allele is methylated) upon restoration of DNMT1 wt levels (Figure 4.4). This is because the methylation at ICRs is set up following the transcriptional program and H3K36 methylation patterns that are unique to the germ line [38, 39]. Thus, we will be able to detect loss-of-function phenotypes in the context of imprint maintenance despite DNMT3A and -B activity.

Besides being an essential part of our mutant screening platform, the Dnmt1 degron cell line C39 could also prove useful to answer a few unanswered questions about DNA methylation. For example, micro-dosing the cell line with dTAG-13 to lower DNMT1 levels in a controlled manner could be used to determine the lower limit of DNMT1 necessary to maintain a highly methylated genome. Furthermore, the fast and reversible degradation of DNMT1 could be used for examining the question why and at what point differentiated cells die from a lack of DNA methylation and what minimum level is required for their survival.

To complement the activity of endogenous DNMT1, we chose to constitutively express (mutant) DNMT1 from transgenes randomly integrated into the genome using PiggyBac transposase. We designed three DNMT1 mutants: analogous to our *in vitro* reconstitution and OptoDroplet assay constructs, (i) ΔROI is missing the eutherian-specific ROI, (ii) ΔcIDR2 lacks the ROI as well as the distal half of the conserved IDR, and (iii) ΔPIP has a Q→E mutation in the PIP box motif (QTTITAHF) that abolishes the ability to bind PCNA [57]. Although we expect ΔPIP to be able to maintain DNA methylation at wt levels, it will be useful as a benchmark for effects on recruiting efficiency of the other two mutants in the future.

In our initial attempt using a codon-optimized cDNA fused to mCherry, we only achieved very low protein levels despite counting up to ten integrations. Since we moved the start of the genes closer to the promoter in the process of cloning, we suspected that this hampered efficient transcription from the construct. One reason could be that the closer distance is too short for the preinitiation complex and transcriptional machinery to assemble and efficiently initiate transcription upstream of the start codon. While prokaryotic promoters are most often located just –10 bp to –35 bp upstream of the TSS, eukaryotic transcription required more spacing and core promoters are often found –50 bp to –150 bp away from the TSS [213]. Indeed, restoring the original promoter-gene distance doubled the mCherry fluorescent signal in transiently transfected cells (Figure 4.9).

Using this second generation construct, we integrated a wt version of DNMT1 in methylation-deficient KH2 TKO cells and obtained clones with ten-fold higher transcript levels with only double the number of integrations compared to our first generation construct (Figure 4.10 A, B). We also obtained a clear signal for the transgenic protein in our Western blot that; however, it still clearly fell below the wt level (Figure 4.10 C). Since our laboratory unequivocally demonstrated that DNMT1 possesses *de novo* activity, we analyzed methylation genome-wide and confirmed that rescuing DNMT1 results in *de novo* methylation of specific loci, which we discuss separately in Section 5.6. Hence, our over-expression constructs clearly produce enzymatically active protein. Because KH2 TKO cells also lack DNMT3A and -B, we could not test whether this amount of DNMT1 would be able to maintain methylation at high wt levels. In anticipation that this amount would not be sufficient, we first optimized the protein expression further.

As our qPCR results showed, transcription from the second generation construct was robust (Figure 4.10 A). Therefore, we asked ourselves whether the bottle neck was related to translation or protein stability. To address these factors, we tested whether the exchange of the codon-optimized cDNA with the endogenous sequence or the insertion of a self-cleaving P2A linker improved the fraction and signal intensity of fluorescent cells. Either change improved the fluorescent signal by about 30 %, but the signal increase upon introduction of the P2A linker may also be the result of a longer half-life of free mCherry compared to the fusion protein. However, we achieved the biggest improvement in combination with expression constructs that contain one or more introns. This observation could be explained by the recent finding that long cDNAs tend to be silenced by the HUSH complex. This

mechanism is part of the cell's innate "immune" response against invading or expanding transposable elements. During their transposition, these species undergo reverse transcription, producing a long intron-less cDNA. The lack of introns is used by the cell to distinguish this type of foreign DNA from its own genes and to repress the former through the activity of the HUSH complex [160]. Comparing the number of integrations and transcription levels of the clones C18 and C37 transfected with the intron-less second generation construct, we can deduce that some copies in C37 were probably silenced, since this clone has a more than two-fold higher copy number than C18, but both clones produce the same amount of mRNA (Figure 4.10 A, B). Yet, in both clones transcription is high and we would therefore expect a higher protein level as well.

Thus, we hypothesize that an additional effect of the introns could be an improvement of transcript processing, resulting in larger amounts of mRNA available for translation. Several groups have shown that the export of spliced transcripts can be up to ten-fold faster than that of intron-less cDNA [214–220]. This is due to the coupling of transcription, mRNA splicing, processing, and export, where several routes exist through which the machinery required for one process is recruited by interactions with RNA-binding proteins from previous maturation steps [221, 222]. To test this hypothesis, fluorescence *in situ* hybridization with probes against mCherry (to distinguish the ectopic from the endogenous transcript) could be used to compare the fraction of nuclear and cytoplasmic transcript in intron-less *versus* intron-containing expression constructs. Alternatively, qPCR could be used to measure mRNA levels in cytoplasmic and nuclear fractions separately.

Of note, some previous studies performing rescue experiments by ectopic *Dnmt1* expression used the so-called *Dnmt1* "minigene" cloned by the Jaenisch laboratory [223–227]. The first study involving rescue experiments of *Dnmt1* used an intron-less cDNA consisting of exon 4 to 39, since the ORF of *Dnmt1* was initially thought to start with the ATG codon in exon 4 [154]. However, a follow-up study identified three additional exons upstream of exon 4 [223]. These exons—alongside a 12 kb region containing the *Dnmt1* promoter—were then added to the initial construct, creating the so-called *Dnmt1* "minigene". Due to this cloning history, the "minigene" contains three introns from the endogenous locus.

Finally, we used the optimized version of our expression vector to obtain monoclonal derivatives of our Dnmt1 degron cell line C39 ectopically expressing mutant

DNMT1 or the wt protein as a control. In addition to using the novel constructs, we also included two sequential enrichment steps in our derivation process to isolate the most highly expressing clones (Figure 4.16). Our optimization translated into robust protein levels comparable to wt, with the exception of the "Hbb" ΔcIDR2 clones. This was surprising, given that the transcript levels of these clones were in the same range of those of the other lines and because "Hbb" ΔcIDR2 clones had the strongest mCherry fluorescence signal (Figure 4.17 B, C). This could point towards the deletion of a large part of the IDR leading to decreased protein stability. On the one hand, expressing ΔIDR2 as codon-optimized, intron-less cDNA, we achieved high protein levels of the mutant. On the other hand, we counted approximately 140 transgene integrations for this clone (data not shown), that could compensate for the little protein produced per integration. Interestingly, Ortega-Alarcon et al. (2021) recently showed that the IDRs of MeCP2 contribute to its stability and proposed that this effect may generally apply to multidomain proteins where IDRs promote intermediate partially folded states and limit unfolding cooperativity [228]. Hence, it is possible that the IDR evolved to assist the folding of the large multi-domain protein DNMT1 and/or increase its structural flexibility for conformational changes, *e.g.*, upon DNA binding [181]. In light of this, the effect of the IDR on DNMT1 stability poses an attractive avenue to pursue further.

All in all, performing several rounds of optimization of ectopic DNMT1 expression, we eventually derived a system for the efficient screening of DNMT1 mutant phenotypes in mESCs. Our optimized constructs enable the robust transcription and translation of Dnmt1 from cDNA sequences with endogenous codon usage and containing an efficiently spliced intron. Separation of the mCherry fusion by a self-cleaving linker may confer additional protein stability. Combining this with our improved derivation strategy consistently enriching for highly expressing clones, we could reproducibly derive monoclonal cell lines expressing wt or mutant DNMT1 at wt protein levels. This system now paves the way for investigating the role of the IDR and ROI in maintenance methylation. Moreover, our Gateway-based set of entry and destination vectors also allows a swift adaptation to the investigation of other mutants of DNMT1 as well as other proteins.

5.6 *De novo* activity of DNMT1 and role of the IDR

While developing our mutant screening platform, we made several observations that further confirm the *de novo* activity of DNMT1. First, we uncovered increases in DNA methylation in our *Dnmt3a-/-;Dnmt3b-/-* Dnmt1 degron cell line (C46) upon washout of dTAG-13 (Figure 4.6 C, D). As evident from our genotyping, both DNMT3A and DNMT3B are clearly absent (Figure 4.5). The low level CpG methylation in our day 1 sample of the dTAG-13 treatment time course further confirms the absence of the canonical *de novo* methyltransferases (Figure 4.6 A, B, C). If DNMT3A and -B were still active, we would have expected a loss of methylation around 50 % within the first 24 h of dTAG-13 treatment, since within one cell cycle half of the DNA would be passively demethylated during replication, and we confirmed that this is the case during the characterization of our Dnmt1 degron cell line C39 (Figure 4.3 F). On the one hand, in many cell types the high fidelity of DNMT1 would suffice to keep methylation loss in the absence of DNMT3A/B at a minimum. In mESCs, on the other hand, DNMT3A and -B are in constant competition against the demethylating TET enzymes, which are highly active in this cell type. This results in a high methylation turnover in mESCs, while the overall methylation landscape is steady [32]. Hence, upon the KO of DNMT3A and -B, TET activity prevails and our dTAG-13 treated cell line is subject to both passive and active demethylation.

Taken together, the only remaining catalytically active DNMT in dTAG-13 treated C46 cells and possible source of the observed *de novo* methylation is DNMT1, which persists at low levels (Figure 4.3). Setting a minimal threshold at 10 % gain, we identified several hundred repeat elements as primary targets for the *de novo* activity. With over 400 elements, about half of the recovering regions were ERVKs, which are a type of LTR and include the murine-specific IAPs. The second most common repeat to regain methylation were LINE1s, even though LINEs make up a larger fraction of the mouse genome compared to ERVs [229].

Second, when over-expressing DNMT1, but not the catalytically inactive mutant pcat, in KH2 TKO, we clearly found interspersed peaks of 5mC all over the genome. These overlapped with the DMRs our laboratory previously identified in DKO_0 cells, which are derived from reversing the shRNA-knockdown of *Dnmt1* in TKO-like cells [33]. *Dnmt3a* and *Dnmt3b* are genetically ablated in TKO-like cells. Therefore,

TKO-like cells lack the activity of all three DNMTs and are devoid of DNA methylation. Five passages after reversal of the *Dnmt1* knockdown, DKO_0 cells exhibit some focal increases in methylation, which become more obvious after fifteen passages in culture. Despite not producing DNMT1 at wt levels, our KH2 TKO cells rescued with V5-Dnmt1^{co} gained methylation at the same regions at the same rate (Figure 4.11). After 60 d, CpG methylation in DMRs reached on average 18.4 %, which is just a little bit higher than the DMR methylation of DKO_0 cells after 15 passages (being approximately equivalent to 45 d in culture). Haggerty et al. (2021) identified the majority of these *de novo* DMRs as repeat elements, mainly ERVKs which include IAPs, but also LINEs [33]. Accordingly, we find a strong increase in methylation in KH2 TKO + V5-Dnmt1^{co} C37 at LTRs (which include ERVs) and LINEs (Figure 4.11 C, D). After 60 d in culture, we also observed a slight increase in methylation at TSSs, which can be attributed to the overall methylation gain in DMR flanking regions (Figures 4.11 D, 4.12 B). Using this orthogonal approach to rescue the DNMT1 protein by expressing a catalytically inactive mutant, we additionally demonstrated that the recovery of DNA methylation not only depends on the presence of DNMT1 but also its catalytic activity (Figure 4.11).

Importantly, both the recovering elements in our *Dnmt3a-/-;Dnmt3b-/-* Dnmt1 degron cell line (C46) and the *de novo* DMRs in KH2 TKO + V5-Dnmt1^{co} C37 largely overlap. Almost all of the LINE1 elements gaining methylation in C46 recovering from dTAG-13 treatment overlap with *de novo* DMRs in DKO_0 and therefore also in KH2 TKO + V5-Dnmt1^{co} C37. More than half of the recovering ERVKs also lie within *de novo* DMRs; however, we identified about 180 additional hypermethylated ERVKs in C46. This could be due to the lower threshold set for identifying recovering regions upon dTAG-13 washout (minimum 10 % gain) *versus* the definition of a *de novo* DMR (minimum 15 % gain).

Since the *de novo* activity is directed at a narrow set of regions, DNMT1 must be specifically recruited to them. This recruitment has to be highly efficient, since even trace amounts of DNMT1 are sufficient to observe this activity, *e.g.*, in dTAG-13 treated Dnmt1 degron cells. Further experiments by Haggerty et al. (2021) showed that DNMT1's co-factor UHRF1 is required for its activity and is enriched at DMRs compared to flanking regions. Moreover, the DNMT1 *de novo* targets are rich in H3K9me3 and recruit TRIM28, presumably through the binding of ZFPs, also in the absence of DNMT1 [33]. Here we can draw parallels to imprint maintenance throughout the wave of epigenetic reprogramming in preimplantation em-

bryos, where ZFP57 and TRIM28 are key players (see Section 5.4). Indeed, the embryo also avoids complete demethylation of repeat elements during this phase, particularly at LINEs and IAPs, presumably to continue their repression and avoid the harmful effects of transcriptional upregulation of thousands of promoters [33, 230, 231]. However, just as in the case of imprinting, which regulatory domain of DNMT1 is responsible for its recruitment to these sites and what DNMT1's direct interaction partner is in this case remains to be determined.

Wondering about the role of the IDR and the ROI in recruiting DNMT1 to these specific regions, we repeated the rescue with our V5-ΔIDR2^{co} mutant, which does not contain the ROI and additionally lacks half of the remaining IDR. This cell line gained methylation at DMRs to the same extent and with the same kinetics as our wt rescue (Figures 4.11, 4.12). Based on these data, we can conclude that (i) the IDR is neither required for DNMT1 *de novo* activity, (ii) nor to recruit DNMT1 to the *de novo* DMRs. To further investigate how DNMT1 finds its *de novo* targets, we could use our established screening platform to express other DNMT1 mutants, *e.g.*, defective in binding H3K9me3 or H3K18/K23Ub through mutations in the RFTS. The interpretation of the results could be complicated by the residual endogenous DNMT1 protein. However, comparison of the activity of wt *versus* mutant rescues in dTAG-13 treated cells over prolonged periods of time ($\geq$ 60 d) will lead to clear methylation gains in DMRs in case of wt activity.

5.7 Summary

Based on state-of-the-art methods for the prediction of the IDR boundaries, we uncover that the IDR is conserved in the DNMT1 N-terminus in 49 investigated eukaryotic species, spanning from silkworm to humans. While the primary sequence of the IDR is only moderately conserved compared to the folded domains, we find that the degree of predicted disorder of the entire region is high across all investigated species. Moreover, we define a novel ROI exclusive to extant eutherian mammals that extends the IDR by 122 amino acids. Focusing on species of this subset, we find that the amino acid composition of the IDR is very similar, with a high content of charged and hydrophilic residues. Together, our analyses show that structural and biochemical features that are known to define the function of IDRs appear to be

under evolutionary pressure, implying a strong positive selection. These conserved features are generally compatible with a role of the IDR in LLPS.

Using an *in vitro* reconstitution assay and an opto-genetic approach in live cells, we for the first time demonstrate that the DNMT1 IDR by itself is capable of forming condensates at physiological conditions. We establish that the IDR conserved across animals is sufficient for this behavior and that the eutherian-specific ROI does not seem to influence its performance in the context of these assays. Instead, we discuss other potential functions of the ROI in mammalian DNA methylation, such as imprinting. In live cells, IDR condensates' localization is reminiscent of the replication foci patterns and they can partition P15PAF, a known DNMT1 interaction partner known to be present at replication forks. Hence, we hypothesize that IDR condensates may distribute in a cell-cycle dependent manner and suggest a collection of experimental approaches to test this. Overall, our experimental evidence points towards a role of LLPS in the recruitment of DNMT1 to replication foci. Such a mechanism could ensure high local concentrations of DNMT1 at these sites, supporting the efficient post-replicative maintenance of DNA methylation. However, further experiments are required to validate this process and to proof its biological relevance.

To enable the investigation of a potential role for LLPS in DNMT1 activity *in vivo*, we establish a mutant screening platform in mESCs to enable the swift exchange of the endogenous DNMT1 with a mutant version. Using an FKBP-based degron, we show that maintenance methylation can be reversibly inhibited by efficient degradation of DNMT1. Moreover, we demonstrate that even at protein levels below wt, DNMT1 is capable of *de novo* methylation targeted to transposable elements, and that this activity neither requires the ROI nor the remaining conserved IDR in its full length. Hence, even if the IDR is involved in LLPS, this mechanism is not necessary for DNMT1 to find these special targets and to methylate them *de novo*. Overcoming initially weak transcription and potentially inefficient mRNA processing, we achieve robust expression of ectopic DNMT1 using intron-containing transgenes and a sequential enrichment strategy of clones with high transgene expression, paving the way to study the behavior of DNMT1 IDR mutants in the context of maintenance methylation.

5.8 Future directions

In this study we have established that the DNMT1 IDR is capable of undergoing LLPS. To further understand this behavior, it will be helpful to identify the amino acid residues and patterns in the IDR that mediate its separation. For this investigation we can carry out the *in vitro* reconstitution assay as well as the OptoDroplet assay with various IDR mutants. Our thorough analysis of amino acid residues and their distribution can serve as a convenient starting point for the design of such mutant constructs. A persisting challenge in the field of biomolecular condensates is the separating of protein behaviors mediated by LLPS and other mechanisms, *e.g.*, protein-protein interactions, *in vivo*. Against this background, not only will a dissection of the IDR at the amino acid level reveal which exact properties are required for LLPS, but will also generate mutants with defective LLPS and otherwise minimal changes to the primary sequence. These will come in handy when further studying the relevance of LLPS for DNMT1 *in vivo* behavior.

In addition to answering these fundamental questions, it could be interesting to investigate the effect of post-translational modifications on IDR behavior and function. The composition of biomolecular condensates is often controlled by altering the biochemical properties of the contained proteins, such as charge. Besides mediating multivalent interactions that lead to LLPS, a main function of IDRs is to serve as a primary site for PTMs, as they are easily accessible for writers and erasers. Phosphorylation is a common modification to control the assembly and disassembly of biomolecular condensates in a reversible manner [232]. For example, during heterochromatin formation, phosphorylation of the HP1α N-terminus may provide a cue for the formation of condensates by enabling interactions between the N-terminus on one HP1α molecule with the hinge regions of another HP1α molecule. As a result, higher-order oligomers are formed that would then associate with and spread along the DNA [104]. Another example for the control of LLPS by phosphorylation is the formation of DNA replication initiator condensates. Parker et al. (2019) showed that the putative phosphosites in all three investigated proteins (Orc1, Cdt1, Cdc6) are concentrated to their IDRs, and that CDK/Cyclin-dependent phosphorylation inhibits the assembly of their condensates [116]. Control of DNMT stability and activity by PTMs is an understudied area. There have been many occasional reports of PTMs, particularly phosphorylation, in the DNMT1 N-terminus and especially in the IDR, but only less than a handful have been investigated in

more detail [233–235] Hence, it will be worthwhile to figure out how the DNMT1 N-terminus, and particularly the IDR, become a target for PTMs and how these integrate the protein into the cell signaling networks within the cell.

The DNMT1 IDR provides a platform for a wide range of possible multivalent interactions. These could be mediated through the low complexity nature of the amino acid sequence but also through interaction motifs such as SLiMs and MoRFs. So far, only one SLiM, the PIP box, has been mapped to the IDR. However, a first step in the investigation of protein-protein interactions mediated by the IDR, which may or may not involve LLPS, should be the search for other canonical SLiM or MoRF motifs within the IDR. Furthermore, the IDR and ΔROI sequences could be overexpressed in cells and serve as a bait for identifying specific interaction partners. While native co-immunoprecipitation of tagged sequences could serve as a first screen, especially weak or transient interactions are better captured with Rapid Immunoprecipitation Mass Spectrometry of Endogenous proteins (RIME) which involves cross-linking [236]. Beyond the isolated IDR or ROI, one could also expand this to the full-length protein or mutants expressed in our Dnmt1 degron cell line.

In this study, we have utilized this Dnmt1 degron cell line as the basis for our mutant screening platform for the swift exchange of the endogenous wt protein with a mutant. Profiling DNA methylation genome-wide, e.g. using RRBS or WGBS, we can now start to investigate the effect of the IDR and ROI deletions on the maintenance of methylation, which is high and subject to constant turn-over in mESCs. However, beyond probing the steady-state in a cell population, we can also apply other methods to study the efficiency of maintenance methylation by measuring how quickly DNA methylation is restored in the nascent strand using Replication Bisulfite Sequencing (Repli-BS-seq) which combines Bromodeoxyuridine (BrdU) labeling and immunoprecipitation with WGBS to profile 5mC separately on the template and nascent strand [185]. As a complementary approach, we could explore the recruitment behavior of DNMT1 to replication foci using the trapping assay developed by Schermelleh et al. (2005) [237]. This assay probes the association kinetics of fluorescently labeled DNMT1 with DNA that contains cytosine analogues, immobilizing the enzyme on DNA. This readout would require the replacement of the self-cleaving P2A linker in our constructs with a non-cleavable linker so that DNMT1 location can be determined by tracking mCherry. Moreover, this method will likely benefit from the use of a stronger, more photostable fluorophore such as StayGold or mGreenLantern [238, 239]. However, since our expression vectors are

based on a Gateway cloning system, the coding sequence can be easily transferred into new destination vectors.

In addition to studying DNMT1 mutant phenotypes in the context of steady-state maintenance and recruitment kinetics, it may be possible to use our system to investigate the phenotype of N-terminal mutants on the processivity of DNMT1 *in vivo*. So far, processivity has only been studied in *in vitro* biochemical assays [179–182]. However, emerging long-read sequencing technologies open avenues to trace the progression of advancing replication forks along the genome. Methods such as DNAscent have already successfully applied to yeast, but should be transferable to the analysis of other eukaryotic species as well [240]. Optimizing the signal deconvolution of the data, it may be possible to not only identify nascent strands through the incorporated BrdU but also distinguish C from 5mC to specifically read methylation on the nascent strand and determine its distance from the nearest replication fork.

Using this work as a foundation exploring role of intrinsic protein disorder in DNMT1, it would be interesting to expand this research to other DNMTs. In this respect, studying the emergence and evolution of molecular IDR features from the prokaryotic origin to extant eukaryotic species may inform when positive selection of the IDRs began. Concomitant changes in DNA methylation patterns may hint at the functional implications for this evolutionary event. Beyond this, the approaches presented in this work offer a framework to determine how widespread conserved IDRs are in other proteins related to DNA methylation and to pioneer their exploration.

6

Methods

All scripts have been deposited on Github, unless indicated otherwise: https://github.com/ninabailly/ninas_phd_code.

Analysis in R was carried out with R (version 4.1.2) using RStudio 2022.07.1 Build 554 run on Windows 10. Analysis in Python was carried out with Python (version 3.9.5) run through Ubuntu 20.04.3 LTS on Windows 10.

Detailed Standard Operation Procedures (SOPs) for all experimental work are included the in Appendices A-C.

6.1 Evolutionary conservation analysis

Metapredict

The disorder of protein sequences was evaluated using the Metapredict software tool (version 1.51) [132]. To calculate the consensus score and AlphaFold confidence score for the MSA of placental Dnmt1 proteins, the hyphens were removed before the computation and restored at the correct indices afterwards, using the script `metapredict_msa.py`. The average consensus score and AlphaFold confidence were plotted using `script_metapredict_plot_msa.Rmd`. The domain annotation of mouse DNMT1 was obtained from UniProt and added manually to the plots [138]. Consensus score and AlphaFold confidence score for single proteins were computed and plotted with the scripts `metapredict_single_fasta.py` and `script_ metapredict_plot_dual.Rmd`. Putative disordered regions predicted with

the script metapredict_disorder_domains.py were highlighted manually in the plots.

Multiple Sequence Alignment and Conservation score

The protein sequences for DNMT1, DNMT3A and DNMT3B across animal species and for the closely related plant proteins were downloaded from the NCBI database [136] using the NCBI Datasets command line tool [241]. The respective input lists are given in data_table_dnmt1.tsv, data_table_dnmt3a.tsv, and data_table_dnmt3b.tsv, respectively. The data were restructured using dnmt1_R_script_meta_data_analysis.Rmd and filtered for the longest protein isoform using filter_isoforms.Rmd, generating a FASTA file with each specie's longest isoform and the specie's common name (adapted from [242]). The MSA and corresponding guide tree were generated using the ClustalO algorithm (version 1.2.4), selecting FASTA as output file format for the MSA [243] (see clustalO_nina.sh). The MSA and guide tree were visualized using the Python based Environment Tree Exploration (ETE) toolkit (version 3.0; see nina_ete3.py) [244]. The domain annotations of mouse DNMT1, DNMT3A and DNMT3B were obtained from UniProt [138] and added manually to the respective MSA visualization.

The conservation score was computed using the Python based score conservation program by Capra and Singh [245] (version updated as of 09.03.2011; see score_conservation.py), using the Jensen-Shannon divergence algorithm (calculate_conservation.sh). The domain annotation of mouse DNMT1 was obtained from UniProt [138] and added manually to the plots.

Amino acid composition analysis

The amino acid composition of IDRs was evaluated and plotted using script_aa_composition_plots.Rmd, using a FASTA file containing the IDR sequences as input. The PCA of IDR sequences was generated and plotted using script_aa_pca.Rmd. Bar code plots were generated using the script by Basu et al. [246].

6.2 *In vitro* reconstitution assay

Vector construction

The pET-45b-mCherry protein overexpression vector (Addgene #145279) digested with AscI (NEB R0558S) and HindIII-HF (NEB R3104S) according to SOP B.5. The IDR sequences as predicted by Metapredict (see Section 6.1) were amplified from V6.5 wt mESC cDNA using the Gibson primers in Table 6.1 with SOP B.4 and cloned into the digested pET-45b-mCherry vector according to SOP B.7. Subsequently, 50 μL competent BL21(DE3) bacteria (NEB C2527H) were transformed with 5 μL assembly mix using SOP B.10 on LB plates containing ampicillin. Six 5 mL LB cultures supplemented with ampicillin were inoculated with colonies from the plates and incubated at 37 °C overnight, shaking at 200 rpm. A glycerol stock of each culture was prepared by 500 μL of the culture with 50 % glycerol in a 2 mL microcentrifuge tube and stored at –80 °C. The plasmids were isolated from the remaining culture with the Qiagen Mini Prep Kit (Qiagen 27106X4) according to the manufacturer's instructions. The plasmids were eluted in 20 μL EB buffer and the correct assembly of the vector was verified with Sanger sequencing using the primers pET45(+)_backbone_mCherry_F and pET45(+)_backbone_R1.

Tab. 6.1.: Primers for IDR purification constructs. All primers were obtained from SigmaAldrich. This table contains primers for Gibson assembly (GA) and Sanger sequencing (SEQ). All sequences are given in 5' to 3' orientation.

Construct	Primer name	Type	Sequence
IDR	mCh_dnmt1_idr_F	GA	catggatgaattgtacaagtacacg...
			...CTCACTCAAAAAGCCAACGGTTGTC
	mCh_dnmt1_idr_R	GA	ctttaccagactcgagtgcggccgc...
			...AGGGTCGTCTAGGTGCTGGCCACAC
ΔROI	mCh_dnmt1_idr_F	GA	catggatgaattgtacaagtacacg...
			...CTCACTCAAAAAGCCAACGGTTGTC
	dnmt1_cIDR_1_R	GA	TCTCCTCCCTCTCATCCTCGTCTCT...
			...CTCAGCCGAGTTCCCCTCTTCCGAC
	dnmt1_cIDR_2_F	GA	GTCGGAAGAGGGGAACTCGGCTGAG...
			...AGAGACGAGGATGAGAGGGAGGAGA
	mCh_dnmt1_idr_R	GA	ctttaccagactcgagtgcggccgc...
			...AGGGTCGTCTAGGTGCTGGCCACAC
ROI	mCh_dnmt1_roi_F	GA	catggatgaattgtacaagtacacg...
			...TCGGCTGCAGAGGAGAGAGACCAGG
	dnmt1_roi_stag_R	GA	ctttaccagactcgagtgcggccgc...
			...GTCCTCGGGAGTCTCTGGAGCTACC
all	pET45(+)_backbone_mCherry_F	SEQ	gagggcagacatagcacagg
all	pET45(+)_backbone_R1	SEQ	agctgcgctagtagacgagt

Protein isolation

After confirmation of the correct sequence, LB plates containing ampicillin were inoculated from the corresponding glycerol stock using the four quadrant method to obtain single colonies. The plates were incubated overnight at 37 °C. Auto-induction cultures were inoculated with a single colony and purified according to the SOPs in Appendix C. The protein concentration was determined using our laboratoy's mCherry standard curve using the Nanodrop-2000 (Thermo Fisher).

Image acquisition and analysis

The reconstitution assay was carried out by mixing the isolated protein with storage buffer not containing glycerol and 20 % polyethylene glycol to a final protein concentration of 0.01 mM to 10 mM in a final volume of 10 μL. The suspension was pipetted on a chambered cover slip (μ-Slide 8 Well Grid-500, ibidi, 80826-G500) and imaged with an inverted confocal laser scanning microscope (Zeiss LSM880) using the 63x 1.4NA oil objective. The focus plane was set to the grid plane ($z = 0\,\mu m$). Images of the droplets were taken within 3 min to 5 min after reconstitution in the center of the droplet at $z = 25\,\mu m \pm 2.5\,\mu m$.

Image segmentation was carried out using the ZEISS ZEN Intelessis tool and plotted using `plot_psf.rmd`. Sigmoidal fit curves were added using Prism (GraphPad).

6.3 OptoDroplet assay

Vector construction

The vector pHR-mCherry-CRY2-NLS (Addgene #145259) was digested with CspCI (NEB R0645S) and dephosphorylated according to SOP B.5. The IDR sequences as predicted by Metapredict (see Section 6.1) were amplified from V6.5 wt mESC cDNA for mouse sequences and HEK293T cDNA for human sequences using the Gibson primers in Table 6.2 with SOP B.4 and cloned into the digested pHR-mCherry-CRY2-NLS vector according to SOP B.7. To clone the pHR-GFP-CRY2-NLS plasmid, pHR-mCherry-CRY2-NLS was digested with BamHI-HF (NEB R3136S) and SbfI (NEB R3642S) according to SOP B.5. The GFP and Cry2-NLS sequences were amplified from PB_lox_mChe_lox_GFP (in-house) and pHR-mCherry-CRY2-NLS, respectively, and cloned into the digested backbone using SOP B.7. To insert the p15paf cDNA, pHR-GFP-CRY2-NLS was digested with SpeI-HF (NEB R3133S) according to SOP B.5 and assembled with the cDNA according to SOP B.7.

Subsequently, 50 μL competent BL21(DE3) bacteria (NEB C2527H) were transformed with 5 μL assembly mix using SOP B.10 on LB plates containing ampicillin. Six 5 mL LB cultures supplemented with ampicillin were inoculated with colonies from the plates and incubated at 30 °C overnight, shaking at 200 rpm. A glycerol stock of each culture was prepared by 500 μL of the culture with 50 % glycerol in a 2 mL microcentrifuge tube and stored at –80 °C. The plasmids were isolated from the remaining culture with the Qiagen Mini Prep Kit (Qiagen 27106X4) according to the manufacturer's instructions. The plasmids were eluted in 20 μL EB buffer and the correct assembly of the vector was verified with Sanger sequencing using the primers pHR-mCh-Cry2_re4 (for mCherry containing constructs only) and pHR_SFFV_F. A culture of 50 mL of LB medium supplemented with ampicillin was inoculated from the glycerol stock and incubated overnight at 30 °C overnight, shaking at 200 rpm. The plasmids were isolated according to SOP B.11.

Tab. 6.2.: Primers for OptoDroplet constructs. All primers were obtained from SigmaAldrich. This table contains primers for Gibson assembly (GA) and Sanger sequencing (SEQ). All sequences are given in 5' to 3' orientation.

Construct	Primer name	GA	Sequence
opto-IDR	pHR_dnmt1_IDR_F	GA	ggggatctggagctctcgagaattctcacgccacc...
			...ATGCTCACTCAAAAAGCCAACGGTTGTC
	dnmt1_IDR_SGSG_mCh_R	GA	cctcctcgcctttagacaccatggtagctccggatccact...
			...AGGGTCGTCTAGGTGCTGGCCACAC
opto-ΔROI	pHR_dnmt1_IDR_F	GA	ggggatctggagctctcgagaattctcacgccacc...
			...ATGCTCACTCAAAAAGCCAACGGTTGTC
	dnmt1_cIDR_1_R	GA	TCTCCTCCCTCTCATCCTCGTCTCT...
			...CTCAGCCGAGTTCCCCTCTTCCGAC
	dnmt1_cIDR_2_F	GA	GTCGGAAGAGGGGAACTCGGCTGAG...
			...AGAGACGAGGATGAGAGGGAGGAGA
	dnmt1_IDR_SGSG_mCh_R	GA	cctcctcgcctttagacaccatggtagctccggatccact...
			...AGGGTCGTCTAGGTGCTGGCCACAC
opto-ROI	pHR_dnmt1_roi_F	GA	ggggatctggagctctcgagaattctcacgccacc...
			...ATGTCGGCTGCAGAGGAGAGAGACCAGG
	dnmt1_roi_SGSG_mCh_R	GA	cctcctcgcctttagacaccatggtagctccggatccact...
			...GTCCTCGGGAGTCTCTGGAGCTACC
opto-MTase	pHR_dnmt1_mtase-n_F	GA	ggggatctggagctctcgagaattctcacgccacc...
			...ATGCTCCGGACCCTGGATGTGTTTTCCG
	dnmt1_mtase-n_SGSG_mCh_R	GA	cctcctcgcctttagacaccatggtagctccggatccact...

Tab. 6.2.: Primers for OptoDroplet constructs (continued). All primers were obtained from SigmaAldrich. This table contains primers for Gibson assembly (GA) and Sanger sequencing (SEQ). All sequences are given in 5' to 3' orientation.

Construct	Primer name	GA	Sequence
			...ATGGTCCCTGAGGATGGGCTGGTAG
opto-RFTS	pHR_dnmt1_rfts_F	GA	ggggatctggagctctcgagaattctcacgccacc...
			...ATGTATGAAGATTCTCCCATGCATAGGT
	dnmt1_rfts_SGSG_mCh_R	GA	cctcctcgcctttagacaccatggtagctccggatccact...
			...AGTGGTCTCAATCTTATTGATCAGG
opto-cIDR1	pHR_IDR_F2	GA	ggggatctggagctctcgagaattctcacgccacc...
		GA	...ATGGCTGACTCTAATAGATCCCCAC
	mCherry_SGSG_IDR1_R	GA	cctcctcgcctttagacaccatggtagctccggatccact...
			...TCGCCTTGTAGCCACGGATGATGGG
opto-cIDR2	pHR_IDR2_F	GA	ggggatctggagctctcgagaattctcacgccacc...
			...ACTGCACACTTCACTAAAGGTCCTA
	mCherry_SGSG_IDR2_R2	GA	cctcctcgcctttagacaccatggtagctccggatccact...
			...CTTAGGTATAACGCTCTTTGACTGT
opto-ΔPIP	IDR_PIP_mut_F	GA	GTGGCTACGAGGAGAACCACCAGGG...
			...AGACCACCATCACGGCTCACTTCAC
	IDR_PIP_mut_R	GA	GTGAAGTGAGCCGTGATGGTGGTCT...
			...CCCTGGTGGTTCTCCTCGTAGCCAC
opto-Dnmt3a	pHR_dnmt3a_IDR_F	GA	ggggatctggagctctcgagaattctcacgccacc...
			...ATGCCCTCCAGCGGCCCCGGGGACA

Tab. 6.2.: Primers for OptoDroplet constructs (continued). All primers were obtained from SigmaAldrich. This table contains primers for Gibson assembly (GA) and Sanger sequencing (SEQ). All sequences are given in 5' to 3' orientation.

Construct	Primer name	GA	Sequence
	dnmt3a_idr_SGSG_mCh_R	GA	cctcctcgcctttagacaccatggtagctccggatccact...
			...AGGCTCATCGTCGGCTGCTTTGGTA
opto-Dnmt3b	pHR_dnmt3b_IDR_F	GA	ctggagctctcgagaattctcacgccacc...
			...ATGAAGGGAGACAGCAGACATCTGA
	dnmt3b_idr_SGSG_mCh_R	GA	gcctttagacaccatggtagctccggatccact...
			...ATCATCCTGATACTCTGTGCTGTCT
opto-hs_IDR	pHR_hs_dnmt1_IDR_F	GA	ggggatctggagctctcgagaattctcacgccacc...
			...ATGAATGGACGTCTAGAAAACGGGAACC
	hs_dnmt1_IDR_SGSG_mCh_R	GA	cctcctcgcctttagacaccatggtagctccggatccact...
			...CTGAATGCACTTGGGAGGGTGGGTC
opto-hs_ΔROI	pHR_hs_dnmt1_IDR_F	GA	ggggatctggagctctcgagaattctcacgccacc...
			...ATGAATGGACGTCTAGAAAACGGGAACC
	hs_dnmt1_cIDR_1_R	GA	TCTCCTCCTTTTCATCCTCGTCTTT...
			...ATCCGATTTGGCTCTTTCAGACTCT
	hs_dnmt1_cIDR_2_F	GA	AGAGTCTGAAAGAGCCAAATCGGAT...
		GA	...AAAGACGAGGATGAAAAGGAGGAGA
	hs_dnmt1_IDR_SGSG_mCh_R	GA	cctcctcgcctttagacaccatggtagctccggatccact...
			...CTGAATGCACTTGGGAGGGTGGGTC
opto-hs_ROI	pHR_hs_dnmt1_roi_F	GA	ggggatctggagctctcgagaattctcacgccacc...

Tab. 6.2.: Primers for OptoDroplet constructs (continued). All primers were obtained from SigmaAldrich. This table contains primers for Gibson assembly (GA) and Sanger sequencing (SEQ). All sequences are given in 5' to 3' orientation.

Construct	Primer name	GA	Sequence
			...ATGGAGTCCATCAAGGAAGAAGACAAAG
	hs_dnmt1_roi_SGSG_mCh_R	GA	cctcctcgcctttagacaccatggtagctccggatccact...
			...TTCATCAGAAATCTGTGGATTTACT
opto-P15PAF	pHR_p15paf_F	GA	ggggatctggagctctcgagaattctcacgccacc...
			...ATGGTGCGGACCAAAGCAAACTACG
	p15paf_SGSG_mCh_R	GA	cctcctcgcctttagacaccatggtagctccggatccact...
			...TTCGTTTTCATCATCTCTGTGATCA
	p15paf_SGSG_GFP_R	GA	TTGCTggtagctccggatccact...
			...TTCGTTTTCATCATCTCTGTGATCA
pHR-GFP-CRY2-NLS	pHR_Kz_SpeI_GFP_F	GA	atctggagctctcgagaattctcacgccacca...
			...ctagtagtggatccggagctaccAGCAAGG
	Cry2_eGFP_R	GA	tgtccatcttcatacgcgtggccgc...
			...CTTGTACAGCTCGTCCATGCCGAGA
	eGFP_Cry2_F	GA	TCTCGGCATGGACGAGCTGTACAAG...
			...gcggccacgcgtatgaagatggaca
	pHR_Cry2_R	GA	gcttgatatcaagcttgcatgcctg...
			...caggtcgactctagagtcgcggcct
all	pHR_SFFV_F	SEQ	aagagctcacaacccctcactc
	pHR-mCh-Cry2_re4	SEQ	gggtaactgtgactacaccgc

Transfection

HEK293T (ATCC, passage unknown) cells were cultured according to SOP A.6. On the day prior to transfection, cells were seeded on a glass-bottom chambered coverslip (ibidi 80827-90) at a density of 50000 cells/well. For each transfection, 100 ng of plasmid and 0.6 μL LipoD29 In Vitro DNA Transfection Reagent (SignaGen tebu-bio SL100668-1) were *each* diluted in 10 μL KnockOut DMEM (Gibco 10829018). Then the diluted transfection reagent was added to the diluted plasmid, mixed by pipetting up and down three to four times and incubated at RT for 15 min. In the meantime, the medium was replaced by 250 μL fresh KnockOut DMEM (Gibco 10829018) supplemented with 10 % FBS (Pan Biotech, P30-2602) and Penicillin/Streptomycin (Gibco 15140148). The reconstituted transfection mix was added dropwise to the well. The medium was refreshed after 24 h.

Image acquisition and analysis

The OptoDroplet assay was carried out 48 h post-transfection with a confocal laser-scanning microscope (Zeiss LSM880) using the 63x 1.4NA oil objective. The field of view was first scanned with the 561 nm laser (intensity 0.5, gain 700, pinhole 1 AU) to acquire an image and then with the 488 nm (intensity 0.5 or 0.05, gain 700, pinhole 1 AU) for blue-light stimulation over ten cycles. Image pre-processing, segmentation and tracking of individual nuclei was carried out in collaboration with Dr. René Buschow (Microscopy Facility at the Max Planck Institute for Molecular Genetics) using scripts `RB_220720_nina_track.czias`, `220909_chanel_fusion.czmac`, and `220801_track.R` in ZEISS ZEN blue 3.5 and R version 3.6, respectively. These scripts are found online in `https://github.com/ReneBuschow/Fuse_Spot_Track`. Briefly, the channel was re-named for convenience and a Gaussian smooth was applied to all images. Individual nuclei were called as foreground objects by setting the background area zero with rolling ball background subtractions. Binarization was carried out using fixed intensity thresholds and close objects were split by water shedding. Signal intensity and standard deviation were measured by applying the resulting mask to the raw images. Individual foreground objects (nuclei) were filtered by size (75 μm^2 to 500 μm^2) and by circularity (0.6-1), and then tracked using their center of mass using `220801_track.R`. Tracks were output as `.txt` file.

Finally, signal intensity and standard deviation of each track were processed and plotted using `optodroplet_quantification.Rmd`.

6.4 Derivation of cells with inducible degradation of DNMT1

Donor template construction

Using a Cas9-mediated knock-in, we inserted a cDNA coding for the fluorescent reporter protein mCerulean linked to a hemagglutinin (HA) tag and the FKBP12^{F36V} degron downstream of the start codon in *Dnmt1* exon 1 into V6.5 ESCs (Figure 4.1 B). The sequence for the sgRNA (Table 6.3) was obtained as Gibson primers and inserted into the pX458 plasmid (Addgene #48138) according to SOP B.8.

Six 5 mL LB cultures supplemented with ampicillin were inoculated with colonies from the plates and incubated at 30 °C overnight, shaking at 200 rpm. A glycerol stock of each culture was prepared by 500 µL of the culture with 50 % glycerol in a 2 mL microcentrifuge tube and stored at –80 °C. The plasmids were isolated from the remaining culture with the Qiagen Mini Prep Kit (Qiagen 27106X4) according to the manufacturer's instructions and eluted in 20 µL EB buffer. The correct assembly was verified by Sanger sequencing using the primer U6_F.

To construct the donor template pUC19 mCer-GS-FKBP, the cDNA of the mCerulean fluorescent reporter and the HA-tagged FKBP12^{F36V} degron were amplified from pCU19-T (in-house) and pCRIS-PITChv2-Puro-dTAG (BRD4) (Addgene #91793), respectively. The GS linker sequence connecting the fluorescent reporter with the degron was obtained as Gibson primer. The 779 kb long 5' and 963 kb long 3' homology arms were amplified from V6.5 mESC genomic DNA using SOP B.4. The PAM of the guide RNA was scrambled in the reverse primer to amplify the 5' homology arm. A new HpaI site was inserted into the forward primer of the 3'homology arm. All primers used for cloning the template are listed in Table 6.3. The plasmid pUC19 (Addgene #50005) was digested with BamHI-HF (NEB R3136S) and dephosphorylated (SOP B.5). Gel-purified homology arm amplicons and backbone (SOP B.6) were assembled using Gibson assembly (SOPs B.7 and B.10) into pUC19 HA intermediate.

The assembly was verified to be correct by Sanger sequencing using the primers AB_pUC19_colonyPCR_F and -R. The pUC19 HA intermediate plasmid was digested

with HpaI (NEB R0105S), dephosphorylated (SOPs B.5), purified (SOP B.6) and assembled with the mCerulean and FKBP12^{F36V} degron sequence (SOPs B.7 and B.10). Mini preps were formed as described above and the correct assembly was verified by Sanger sequencing using the primers mCer_seq_F1 and FKBP-V_R.

Tab. 6.3.: Primers for construction of the Dnmt1-degron cell line. All primers were obtained from SigmaAldrich. This table contains primers for Gibson assembly (GA) and Sanger sequencing (SEQ), and sgRNA sequences (sgRNA). All sequences are given in 5' to 3' orientation.

Construct	Primer name	Type	Sequence
pX458-Dnmt1 NT2	Dnmt1_NT2	sgRNA	CCGCGCGCGCGAAAAAGCCG
	U6_F	SEQ	ATGGACTATCATATGCTTAC
pUC19 HA intermediate	pUC_mDnmt1_5HA_F	GA	gcatgcctgcaggtcgactctagag...
			...GCTTCGAGTGGCTGGATGTGGGTGG
	mDnmt1_5HA_scramble_R	GA	GCAGGTTGCAGACGACAGAACAGCTCTGAAC...
			...GAGAGGCCGGCTTTTTCGCGCGCGC
	mDnmt1_5HA_scramble_R	GA	GCAGGTTGCAGACGACAGAACAGCTCTGAAC...
			...GAGAGGCCGGCTTTTTCGCGCGCGC
	pUC_mDnmt1_3HA_R	GA	aattcgagctcggtacccggggatc...
			...ATGCACTGAAATGGGGGACTTTACC
	AB_pUC19_colonyPCR_F	SEQ	atggaaaaacgccagcaacg
	AB_pUC19_colonyPCR_R	SEQ	aataggcgtatcacgaggc
pUC19 mCer-GS-FKBP	mDnmt1_5HA_mCer_F	GA	TGTTCTGTCGTCTGCAACCTGCAA...
			...GATGGTGAGCAAGGGCGAGGAGCT
	mCer_GS_R	GA	caccgccaccagatccgccgccacc...
			...CTTGTACAGCTCGTCCATGCCGAGA
	GS linker	GA	ggtggcggcggatctggtgg...
			...cggtggctcgggcggtggtgggtcg
	GS_FKBP_F	GA	cggtggctcgggcggtggtgggtcg...

Tab. 6.3.: Primers for construction of the Dnmt1-degron cell line (continued). All primers were obtained from SigmaAldrich. This table contains primers for Gibson assembly (GA) and Sanger sequencing (SEQ), and sgRNA sequences (sgRNA). All sequences are given in 5' to 3' orientation.

Construct	Primer name	Type	Sequence
			. . . taccctacgacgtgcccgactacg
	mDnmt1_3HA_GS_R	GA	CGCAGGCACTCGGGCTGGAGCTGTTCGCGCTGG. . .
			. . . agatccgccgccacccgacccacca
	mCer_seq_F1	SEQ	GTGAACCGCATCGAGCTGAAG
	FKBP-V_R	SEQ	atggtttccacctgcactcctc

Transfection and screening

V6.5 mESC were cultured according to SOPs A.1-A.4. One million single cells were transfected with 8 μg of pX458-Dnmt1 NT2 and pUC19 mCer-GS-FKBP at a molar ratio of 1:2 using nucleofection (SOP A.7). After 48 h, single GFP+mCerulean+ cells were sorted from the population using FACS (BD FACSAriaII, operated by the FACS & Flow Cytometry Core of the Max Planck Institute for Molecular Genetics). Cells were placed back on 6 cm dishes covered with feeders at a density of about 8000 cell/plate. After 5 d, single colonies were picked according to SOP A.8. Genomic DNA was isolation and screened for the insertion of the reporter-degron construct using SOP B.2 and B.3, respectively. The primers F1 and R1 were used for the initial screen.

Tab. 6.4.: Primers to characterize the Dnmt1-degron cell line. All primers were obtained from SigmaAldrich. This table contains primers for genotyping (GENO). All sequences are given in 5' to 3' orientation.

Primer name	Type	Sequence
F1	GENO	AGCACGGACGAGCCCACTATA
R1	GENO	TCCCTCAAGCTCCCAGTCAATG
F2	GENO	GCTTCGAGTGGCTGGATGTGG
R2	GENO	ATGCACTGAAATGGGGGACTTTACC
F3	GENO	CAGAAGGGGAGGGGGATAGAGT
R3	GENO	ACAGGTAAACCAACCAACCCGA

Genotyping

Two clones producing amplicons of the correct size, C29 and C39, were further expanded. The cells were separated the from the feeders using SOP A.5. Aliquots of 1 million cells were washed once with PBS (Gibco 10010023) and pelleted in 1.5 mL microcentrifuge tubes at 4 °C and 3000 rpm for 3 min. The supernatant was aspirated and the dry pellets was flash frozen on dry ice and stored at –80 °C. Genomic DNA was extracted using SOP B.1 and further characterized using the primer combinations indicated in Figure 4.2 according to SOP B.4. The amplicon produced by the primers F3 and R3 from clone 39 was purified using SOP B.6, cloned into the pJET vector (SOPB.9) and the correct in-frame insertion of the reporter-degron sequence was confirmed using Sanger sequencing the primers mCer_seq_F1 and FKBP-V_R (Table 6.3).

6.5 Derivation of over-expression lines

Transposon vector construction

The codon-optimized cDNA, including the V5-tag, was obtained as two synthetic genes (d1 and d2) from IDT. The synthetic genes were designed by Dr. Jocelyn Charlton (Meissner Lab, Max Planck Institute for Molecular Genetics). The synthetic genes were amplified and cloned into pENTR4-V5-2 (w234-1) entry vector (Addgene #17426) linearized with NcoI-HF (NEB R3193S) and XbaI (NEB R0145S) (SOPs B.7, B.5, B.10). The pcat mutant was derived from pENTR4 V5-Dnmt1^{co} using site-directed mutagenesis using the QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent 210518) according to the manufacturer's recommendation with the mutagenesis primers given in Table 6.5. The ΔIDR2 mutant was derived by amplifying the cDNA up- and downstream of the deletion from pENTR4 V5-Dnmt1^{co} and cloning it back into the NcoI/XbaI-linearized pENTR4-V5-2 (w234-1) entry vector using Gibson assembly (SOPs B.7, B.5, B.10). For the entry vector containing the non-codon optimized (endogenous) Dnmt1 cDNA, a fragment containing the Dnmt1 open reading from was amplified from V6.5 mESC cDNA. The resulting purified amplicon was then re-amplified in two pieces using Gibson primers. The final vector was then constructed in a three-part Gibson assembly. Twelve 5 mL LB cultures supplemented with kanamycin were inoculated with colonies from the plates and incubated at 37 °C overnight, shaking at 200 rpm. A glycerol stock of each culture was prepared by 500 μL of the culture with 50 % glycerol in a 2 mL microcentrifuge tube and stored at –80 °C. The plasmids were isolated from the remaining culture with the Qiagen Mini Prep Kit (Qiagen 27106X4) according to the manufacturer's instructions. The plasmids were eluted in 20 μL EB buffer, and the correct assembly of all vectors and absence of secondary mutations were verified with Sanger sequencing using the primers d1CheckR to d2CheckF5 for the codon-optimized cDNA and gB1checkR4 to mDnmt1_E36_R for the endogenous cDNA, respectively, given in Table 6.5.

Tab. 6.5.: Primers to construct the entry vectors. All primers were obtained from SigmaAldrich. This table contains primers for Gibson assembly (GA), Sanger sequencing (SEQ) and site-directed mutagenesis (MUT). All sequences are given in 5' to 3' orientation.

Construct	Primer name	Type	Sequence
pENTR4 V5-Dnmt1^{co}	attL1_d1-M1L_F	GA	ccaactttgtacaaaaaagcaggct...
			...TGGGCAAACCTATCCCTAAC
	d2_d1_R	GA	GTCCGTTTTTATCTTCCCACAAAGC...
			...TGTTACGCGAGCCAGATACAATGGC
	d1_d2_F	GA	GCCATTGTATCTGGCTCGCGTAACA...
			...GCTTTGTGGGAAGATAAAAACGGAC
	attL2_d2_R	GA	ccaactttgtacaagaaagctgggt...
			...CTAATCCTTAGTGGCAGCTTC
pENTR4 V5-pcat^{co}	pcat_mu_F	MUT	ATGTTGTGCGGAGGTCCCGGC...
			...AGCCAGGGGTTCAGCGGGATG
	pcat_mut_R	MUT	CATCCCGCTGAACCCCTGGCT...
			...GCCGGGACCTCCGCACAACAT
pENTR4 V5-ΔIDR2^{co}	attL1_d1-M1L_F	GA	ccaactttgtacaaaaaagcaggct...
			...TGGGCAAACCTATCCCTAAC
	NLS_d1_R	GA	GGTGCTGCCCACATTCAGGACACTTAGGTGAGTTTAT ...
			...CCTCCTCTTCTTGTCTTGGTCCCGC
	d1_NLS_F	GA	GGTGCTGCCCACATTCAGGACACTTAGGTGAGTTTAT...
			...CCTCCTCTTCTTGTCTTGGTCCCGC
	attL2_d2_R	GA	ccaactttgtacaagaaagctgggt...

Tab. 6.5.: Primers to construct the entry vectors (continued). All primers were obtained from SigmaAldrich. This table contains primers for Gibson assembly (GA), Sanger sequencing (SEQ) and site-directed mutagenesis (MUT). All sequences are given in 5' to 3' orientation.

Construct	Primer name	Type	Sequence
			...CTAATCCTTAGTGGCAGCTTC
all codon-optimized cDNAs	d1CheckR	SEQ	TGGTCAGGCAGGCTTCCG
	d1CheckF	SEQ	CGGAAGCCTGCCTGACCA
	d1CheckF2	SEQ	AAGAAGGTAATTCCGCCGAG
	d1CheckF3	SEQ	CGCACAGTCAAAGAGCGTTA
	d1CheckF4	SEQ	CGTACTCCTTTGATGGCTCC
	d1CheckF5	SEQ	TTTGTCAGCAGCCTGAGTGT
	d2CheckF	SEQ	ATGCAGCTGTCTTACATACACTCAA
	d2CheckF2	SEQ	GTAACCTCGACGCACCTGA
	d2CheckF3	SEQ	AGCGAAGGATTTCATCAAGCA
	d2CheckF4	SEQ	GCAGCTCCAGGTGAGAAACT
	d2CheckF5	SEQ	CTCGGGGACGGAGTGATAG
pENTR4 Dnmt1	Dnmt1_cDNA_F1		TCCCCCACTCTCTTGCCCTGTG
	Dnmt1_cDNA_R1		CATCAGTGCACGTGGGGCTCTG
	attL1_TG_Dnmt1_F	GA	ccaactttgtacaaaaaagcaggct...
			...TGATGCCAGCGCGAACAGCTCCAGC
	Dnmt1_e25_GibR	GA	CCCAGCGCAGAACCAGTGCGCATGG...
			...AACATCATCTGACCATTTTTGTCTT
	Dnmt1_e25_GibF	GA	AAGACAAAAATGGTCAGATGATGTT...

Tab. 6.5.: Primers to construct the entry vectors (continued). All primers were obtained from SigmaAldrich. This table contains primers for Gibson assembly (GA), Sanger sequencing (SEQ) and site-directed mutagenesis (MUT). All sequences are given in 5' to 3' orientation.

Construct	Primer name	Type	Sequence
		GA	...CCATGCGCACTGGTTCTGCGCTGGG
	attL2_Dnmt1_R	GA	ccaactttgtacaagaaagctgggt...
			...CTAGTCCTTGGTAGCAGCCTCCTCT
	gB1checkR4	SEQ	GTCTGCCATTTCTGCTCTCC
	mDnmt1_E4_F	SEQ	TCAAAAAGCCAACGGTTGTCCC
	mDnmt1_E17_F	SEQ	CAGTGTGTACTGCAGTCGCGGG
	mDnmt1_E18_R	SEQ	TGATTGGCCCGAGGTTTTTGCCA
	mDnmt1_E22_F	SEQ	AGTGTGGGAAGTGCAAGGCGTG
	mDnmt1_E23_R	SEQ	TCGTCGTCTGCCTCCTTCACCG
	mDnmt1_E27_F	SEQ	AATTGAGGAGGTGGATGGCCGG
	mDnmt1_E28_R	SEQ	CTTTGGGCGTTTCACGGGGCTA
	NB_Dnmt1_F	SEQ	GGCACCACAGTGTTCACAGA
	mDnmt1_E32_R	SEQ	TCACCTCCCCAGCCATGACCAG
	mDnmt1_E35_F	SEQ	CCAGCCCATCCTCAGGGACCAT
	mDnmt1_E36_R	SEQ	GGCTATGACGCCATCTCCCAGC

The destination vector pPB mCherry-3XFLAG-attR v1 was derived by cloning the mCherry sequence amplified from PB_lox_mChe_lox_GFP (in-house), the 3XFLAG sequence amplified from pX458 (Addgene #48138), and the ccdB cassette from pLIX_403 (Addgene #41395) into the backbone of the PiggyBac vector PB_lox_mChe_lox_GFP (in-house) linearized with XbaI (NEB R0145S) and SacII (NEB R0157S) using Gibson assembly (SOPs B.5, B.7). For replacing the GS-linker between mCherry and the 3xFLAG with P2A, creating pPB mCherry-P2A-3XFLAG-attR v1, pPB mCherry-3XFLAG-attR v1 was digested with SpeI (NEB R3133S) and assembled with the P2A sequence amplified from pX458 (Addgene #48138). The second generation destination vector pPB mCherry-3XFLAG-attR v2 was cloned analogous to pPB mCherry-3XFLAG-attR v1, but using an extra-long forward primer to amplify mCherry, which includes additional 41 bp from the promoter downstream sequence of the original backbone PB_lox_mChe_lox_GFP. One Shot ccdB Survival 2 T1R Competent cells (Invitrogen A10460) were used for transformation with destination vectors (SOP B.10), to avoid toxicity from the ccdB gene. The correct assembly was confirmed using Sanger sequencing with the primers AB_EF1_colony_F to bGH_pA_ seq_R from Table 6.6.

Tab. 6.6.: Primers to construct the destination vectors. All primers were obtained from SigmaAldrich. This table contains primers for Gibson assembly (GA) and Sanger sequencing (SEQ). All sequences are given in 5' to 3' orientation.

Construct	Primer name	Type	Sequence
pPB mCherry-3XFLAG-attR v1	PB_Kz_mCherry_F	GA	ccGCTAGCGAATTCCTTAAGTctagAgccacc...
			...ATGGTGAGCAAGGGCGAG
	3xFlag_SGSG_mCherry_R	GA	cgtcgtggtccttatagtccatggtagctccggat...
		GA	...ccactagtCTTGTACAGCTCGTCCATGCCGCCG
	mCherry_SGSG_3xFlag_F	GA	cgtcgtggtccttatagtccatggtagctccggat...
			...ccactagtCTTGTACAGCTCGTCCATGCCGCCG
	3xFlag_attR_F	GA	tgatattgattacaaagacgatgacgataa...
			...gacaagtttgtacaaaaaagctgaa
	PB_attR2_R	GA	ccaccgcggtggagcgctgggccct...
			...GTTAaccactttgtacaagaaagct
pPB mCherry-3XFLAG-attR v2	PB_msc_v2	GA	AAGCTGTGACCGGCGCCTACTCTAGA...
			...GCTAGCGAATTCattaagggttccg...
			...gatcctcgaggccaccATGGTGAGCAAGGG
all	AB_EF1_colony_F	SEQ	GACCCTGCTTGCTCAACTCT
	ALM_RFP_IF	SEQ	GATGAACTTCGAGGACGGCGGC
	ALM_RFP_IR	SEQ	GCCGCCGTCCTCGAAGTTCATC
	ccdb_seq_R1	SEQ	agctcctgaaaatctcgacgga
	bGH_pA_seq_R	SEQ	ggacagcaaggggaggattg

Expression clones were generated by combining the respective entry and destination vectors in an LR clonase reaction using the Gateway LR Clonase II Enzyme mix (Invitrogen 11791020) according to the manufacturer's recommendation. The correct assembly of the expression clones was verified by Sanger sequencing using the primers d1CheckR to d2CheckF5 for the codon-optimized cDNA and gB1checkR4 to mDnmt1_E36_R for the endogenous cDNA, respectively, given in Table 6.5. A culture of 50 mL of LB medium supplemented with ampicillin was inoculated from the glycerol stock of the expression clones and incubated overnight at 37 °C overnight, shaking at 200 rpm. The plasmids were isolated according to SOP B.11.

The transposon vectors with introns were generated based on the pPB mCherry-P2A-Dnmt1 v2 expression clone. The Dnmt1 cDNA was removed from the plasmid by digest with SpeI (NEB R3133S) and HpaI (NEB R0105S) (SPO B.5). The cDNA was amplified in pieces (up- and downstream of the future intron location) using Gibson primers from pPB mCherry-P2A-Dnmt1 v2. The introns were amplified from V6.5 mESC genomic DNA using Gibson primers. The P2A-3xFLAG sequence was amplified from pPB mCherry-P2A-Dnmt1 v2. All Gibson primers are listed in Table 6.7. All fragments were then cloned back into the backbone containing the promoter, mCherry and polyA sequences using Gibson assembly (SOPs B.7 and B.10). The correct assembly of the expression clones was verified by Sanger sequencing using the primers gB1checkR4 to mDnmt1_E36_R given in Table 6.5. A culture of 50 mL of LB medium supplemented with ampicillin was inoculated from the glycerol stock of the expression clones and incubated overnight at 37 °C overnight, shaking at 200 rpm. The plasmids were isolated according to SOP B.11.

Tab. 6.7.: Primers to construct the destination vectors with introns. All primers were obtained from SigmaAldrich. This table contains primers for Gibson assembly (GA) and Sanger sequencing (SEQ). All sequences are given in 5' to 3' orientation.

Construct	Primer name	Type	Sequence
pPB "End" mCherry-P2A-3xFLAG-Dnmt1	Dnmt1_e2_e1_R	GA	GCCATCTCTTTCCAAGTCTTTGAGC...
			...CGCCTGCGGACATGGTCCGGGAGCG
	Dnmt1_e1_e2_F	GA	CGCTCCCGGACCATGTCCGCAGGCG...
			...GCTCAAAGACTTGGAAAGAGATGGC
	Dnmt1_e5_e4_R	GA	ACGGAACTAGGTGAAGTTTCAACTG...
			...AAAGGGTGTCACTGTCCGACTTGCT
	Dnmt1_e4_e5_F	GA	AGCAAGTCGGACAGTGACACCCTTT...
			...CAGTTGAAACTTCACCTAGTTCCGT
pPB "Igh" mCherry-P2A-3xFLAG-Dnmt1	Ighm_i1_Dnmt1_e18_R	GA	CCTGCTGGGAGGGTTTGGTTCTTAC...
			...CAGTGGAGAAGCCAATGAGCACCTT
	Dnmt1_e18_Ighm_i1_F	GA	AAGGTGCTCATTGGCTTCTCCACTG...
			...GTAAGAACCAAACCCTCCCAGCAGG
	Dnmt1_e19_Ighm_i1_R	GA	TCCATCAAAATGTATTCAGCAAATG...
			...CTGGAATGAAAGGTCAAGGTGTGAG
	Ighm_i1_Dnmt1_e19_F	GA	CTCACACCTTGACCTTTCATTCCAG...
			...CATTTGCTGAATACATTTTGATGGA
	Ighe_i3_Dnmt1_e28_R	GA	CCCACTCTCCACCTCCTGTACTCAC...
			...CTGTAGAACTTGTAGAGCCTCAGCT
	Dnmt1_e28_Ighe_i3_F	GA	AGCTGAGGCTCTACAAGTTCTACAG...

Tab. 6.7.: Primers to construct the destination vectors with introns (continued). All primers were obtained from SigmaAldrich. This table contains primers for Gibson assembly (GA) and Sanger sequencing (SEQ). All sequences are given in 5' to 3' orientation

Construct	Primer name	Type	Sequence
			...GTGAGTACAGGAGGTGGAGAGTGGG
	Dnmt1_e29_Ighe_i3_R	GA	GTAGGACCTGTGGGTATTCTCAGGC...
			...CTGTGGGCAGACATGAGTTAACCAG
	Ighe_i3_Dnmt1_e29_F	GA	CTGGTTAACTCATGTCTGCCCACAG...
			...GCCTGAGAATACCCACAGGTCCTAC
pPB "Hbb" mCherry-P2A-3xFLAG-Dnmt1	Hbbi2_Dnmt1_e3_R	GA	CCCAGGAGGTGCCCATCAGACTCAC...
			...CTCAGATAATTCCTCTTTATGTAAT
	Dnmt1_e3_Hbbi2_F	GA	ATTACATAAAGAGGAATTATCTGAG...
			...GTGAGTCTGATGGGCACCTCCTGGG
	Dnmt1_e4_Hbbi2_R	GA	ACTTGACTTTAGCCAGGTAGCCTTC...
			...CTGTGGGAAGATGGAAGAATCATCA
	Hbbi2_Dnmt1_e4_F	GA	TGATGATTCTTCCATCTTCCCACAG...
			...GAAGGCTACCTGGCTAAAGTCAAGT
pPB "Hbb" mCherry-P2A-3xFLAG-ΔcIDR2	dnmt1_e15_e6_R	GA	CAGGGTGCTGCTGGTACTTCAGGTT...
			...CTCAGCCGAGTTCCCCTCTTCCGAC
	dnmt1_e6_e15_F	GA	GTCGGAAGAGGGGAACTCGGCTGAG...
			...AACCTGAAGTACCAGCAGCACCCTG
pPB "Hbb" mCherry-P2A-3xFLAG-ΔROI	dnmt1_cIDR_1_R	GA	TCTCCTCCCTCTCATCCTCGTCTCT...
			...CTCAGCCGAGTTCCCCTCTTCCGAC

Tab. 6.7.: Primers to construct the destination vectors with introns (continued). All primers were obtained from SigmaAldrich. This table contains primers for Gibson assembly (GA) and Sanger sequencing (SEQ). All sequences are given in 5' to 3' orientation

Construct	Primer name	Type	Sequence
	dnmt1_cIDR_2_F	GA	GTCGGAAGAGGGGAACTCGGCTGAG...
			...AGAGACGAGGATGAGAGGGAGGAGA
pPB "Hbb" mCherry-P2A-3xFLAG-ΔPIP	IDR_pdb_R	GA	GTGAAGTGAGCCGTGATGGTGGTCT...
			...CCCTGGTGGTTCTCCTCGTAGCCAC
	IDR_pbd_F	GA	GTGGCTACGAGGAGAACCACCAGGG...
			...AGACCACCATCACGGCTCACTTCAC
all	pPB_Dnmt1_R	GA	ccgcggtggagcgctgggccctGTTAac...
			...CTAGTCCTTGGTAGCAGCCTCCTCT
	mCherry_P2A_F	GA	CGGCGGCATGGACGAGCTGTACAAG...
			...GGCTCCGGAGCCACGAACTTCTCTC
	Dnmt1_3XFLAG_R	GA	GGGCTGGAGCTGTTCGCGCTGGCAT...
			...cttatcgtcatcgtctttgtaatca
	3xFLAG_Dnmt1_F	GA	tgattacaaagacgatgacgataag...
			...ATGCCAGCGCGAACAGCTCCAGCCC

Transfection and screening

To derive over-expression lines, 1 million single cells from the KH2 TKO C5, or the Dnmt1 degron clone C39 were transfected with 8 μg of pFN21A Transposase-BFP (in-house) and the respective expression clone at a molar ratio of 1:2 using nucleofection (SOP A.7). After 48 h, single BFP+mCherry+ cells were sorted from the population using FACS (BD FACSFusion, operated by the FACS & Flow Cytometry Core of the Max Planck Institute for Molecular Genetics). Cells were placed back on 6 cm dishes covered with feeders at a density of about 8000 cell/plate. After 5 d, single colonies were picked according to SOP A.8. Genomic DNA was isolation and screened for the insertion of the over-expression construct on the basis of the mCherry fluorescence using flow cytometry (BD FACSCelesta Cell Analyzer). Clones with the strongest fluorescence signal were expanded and further characterized.

For the over-expression lines using the intron-containing expression vectors, the population was enriched for the top 10 % mCherry+ cells 9 d and 16 d post-transfection before the isolation of single clones 23 d post-transfection using FACS (BD FACSFusion, operated by the FACS & Flow Cytometry Core of the Max Planck Institute for Molecular Genetics).

6.6 Cell line characterization

Expression analysis

RNA was isolated and converted into cDNA using SOPs B.12 and B.13. All qPCR primers with the respective target genes are given in Table 6.8. Expression levels were measured using qPCR according to SOP B.14. *Gapdh* was used as a housekeeping gene control.

Tab. 6.8.: Primers for qPCR. All primers were obtained from SigmaAldrich. All sequences are given in 5' to 3' orientation.

Target locus	Primer name	Sequence
Dnmt1 exons 5-6	ALM_Dnmt1_ex5_F	GACCACCATCACGGCTCAC
	ALM_Dnmt1_ex6_R	CCCCTCTTCCGACTCTTCCT
Dnmt1 exons 35-36	Dnmt1_e35_qPCR_F	GCGAGGATCACACTACCAGC
	Dnmt1_e36_qPCR_R	CTGTATGTTGGGCAGGTCAC
mCherry	mCherry_qPCR_F	CCGACATCCCCGACTACTTG
	mCherry_qPCR_R	ATGAACTCGCCGTCCTGC
Gapdh	ALM_GAPDH_ex5_F2	ccttccgtgttcctacccc
	ALM_GAPDH_ex6_R2	atgcctgcttcaccaccttc
Tfrc	Tfrc_qPCR_F	CTCAACCAAATGGTTCGTACAG
	Tfrc_qPCR_R	ACATCTCATAGTCCAGGTTCAAT

Measurement of copy number

Transposon insertions of the Dnmt1 over-expression construct were counted by qPCR on genomic DNA (25 ng per well), using the mCherry qPCR primers and the *Tfrc* gene (two copies) as a reference. The resulting FI was used to count the mCherry integrations (copynumber = 2xFI). For technical replicates, the error was calculated using Gaussian error propagation.

Assessment of splicing efficiency

Removal of the *Hbb* intron from the pPB "Hbb" mCherry-P2A-3xFLAG-Dnmt1, -ΔcIDR2, -ΔROI and -ΔPIP constructs' mRNA by splicing was assessed using PCR

on cDNA isolated from expressing clones using SOPs B.12, B.13 and B.4. The chosen primers (Dnmt1_E3_F (F): GCCAGTTGTGTGACTTGGAA; gb1cR4 (R): GTCTGCCATTTCTGCTCTCC) flank the intron and produce a 188 bp amplicon when the intron is removed completely and a 842 bp amplicon when the intron is retained.

Degron-mediated degradation of DNMT1

To degrade endogenous DNMT1, the cells were treated with 250 mM (Figures 4.3, 4.4, and 4.6) or 500 mM (Figure 4.18) dTAG-13 (Torcis/Bio Techne 6605) dissolved in DMSO (Sigma D2650-100ML) to 1 mM.

Western blot

All Western blots were carried out on nuclear protein extracted using SOP B.15. The procedures for the SDS PAGE, transfer and blotting are described in SOPs B.16 and B.17. Protein extraction, SDS PAGE and blotting were performed by Sabine Otto (Meissner Lab, Max Planck Institute for Molecular Genetics). The following antibodies were used: anti-DNMT1 (abcam ab87654, 1:1000), anti-FLAG (SIGMA F1804-50UG, 1:1000), anti-Lamin B (Santa Cruz, sc-374015, 1:300), anti-rabbit IgG horseradish peroxidase conjugate (Jackson 211-032-171, 1:5000; or alternatively TrueBlot Anti-Rabbit IgG HRP, 18-8816-31 1:2000), anti-mouse IgG horseradish peroxidase conjugate (Jackson 115-035-174, 1:5000). The blots were imaged using the Bio-Rad ChemiDoc MP Imaging System (1708280; with Version 5.2.1 build 1.1. of the Image Lab software).

Reduced-representation bisulfite sequencing

Genomic DNA was quantified with a Qubit 3.0 Fluorometer and diluted to a concentration of 12 ng μL^{-1}. RRBS libraries were prepared using 100 ng input per sample with the NuGen Ovation RRBS Methyl-Seq System (Tecan 0353) according to the supplied protocol, using 12 PCR cycles. Bisulfite conversion was carried out using the Qiagen EpiTect Fast Bisulfite Conversion kit (Qiagen 59824). The libraries were purified in two rounds using Agencourt RNAclean XP beads (Beckman Coul-

ter A63987) and their quality was assessed using the Agilent 4150 Tape-Station HS D1000 ScreenTape. The RRBS libraries for the dTAG-13 time courses were prepared by Dr. Raha Weigert (Meissner Lab, Max Planck Institute for Molecular Genetics). The RRBS libraries were sequenced on the NovaSeq6000 (Illumina) with 100 bp single end reads by the Sequencing Core Facility of the Max Planck Institute for Molecular Genetics.

The resulting data were processed and analyzed by Sara Hetzel (Meissner Lab, Max Planck Institute for Molecular Genetics). Aadapter and quality trimming of raw reads was performed using cutadapt (version 2.4; parameters: quality-cutoff 20, overlap 5, minimum-length 25; Illumina TruSeq adapter) [247], followed by NuGEN diversity adapter trimming (https://github.com/nugentechnologies/NuMetRRBS). We used BSMAP (version 2.90; parameters: v 0.1, s 12, q 20, w 100, S 1, u, R, D C, CGG) to align the trimmed reads to the mouse reference genome mm10 [248]. Aligned reads were deduplicated based on unique molecular identifiers using NuDup (https://github.com/nugentechnologies/nudup; parameters: start 6, length 6). Mcall from the MOABS package (version 1.3.2; default parameters) was used to call methylation rates [249].

For analysis, the genome was annotated with the mm10 reference downloaded from GENCODE (VM19) and then segmented into 1 kb tiles using bedtools makewindows (parameters: -w 1000 -s 1000). All analyses were restricted to autosomes and CpGs covered by at least 10 and at maximum 150 reads. Promoter regions were defined as 1500 bp upstream and 500 bp downstream of the TSS. The location of CGIS and repeats were obtained from UCSC. CGI shores were defined as 2 kb flanking either side of a CGI, and CGI shelves were defined as 2 kb flanking each of these shores. Open water was defined as any genomic regions not classified as CGI, shelf or shore. Coordinates of imprinted DMRs and DKO^0 DMRs were obtained from Gigante et al. (2019) and Haggerty et al. (2021), respectively [33, 250]. All analyses were carried out on the subset of CGps covered in all samples. Mean methylated per feature was calculated if the feature contained at least three covered CpGs. Recovering elements were defined as features with an average methylation difference > 0.1 between recovery and treatment and as co-occurring with DKO^0 DMRs if they overlapped by at least 1 bp. The heatmap was generated using the package pheatmap.

Whole-genome bisulfite sequencing

Genomic DNA was quantified using the spectrophotometry (Thermo Scientific, NanoDrop 2000/2000c, ND-2000) and 800 n]g were diluted in 50 μL low-EDTA buffer (Swift Biosciences DLILMMS-12). The DNA was sheared using a Covaris sonicator (10 % duty cycle, 200 cycles per burst, 5 % intensity, 2x45 s) in Covaris micro TUBE AFA Fiber Pre-Slit Snap-Cap tubes (SKU 520045), and subsequently concentrated using the DNA clean and Concentrator Kit (Zymo D4004) according to the manufacturer's instructions. The DNA was eluted in 20 μL low-EDTA buffer. The size distribution was assessed using the Agilent Tape Station. Bisulfite conversion of 100 ng was carried out using the EZ DNA Methylation-Gold Kit (Zymo D5005) according to the manufacturer's instructions and eluted in 15 μL low-EDTA buffer. WGBS libraries were prepared using the Accel-NGS Methyl-seq DNA library kit (Swift Biosciences DLILMMS-12) as recommended by the manufacturer and cleaned using Agencourt AMPure XP beads (Beckman Coulter A63881). The DNA concentration was quantified using the Qubit 3.0 Fluorometer. The correct size distribution of fragments was analyzed with the Agilent 4150 Tape-Station HS D1000 ScreenTape. The WGBS libraries for V6.5 wt, KH2 TKO and KH2 TKO + V5-Dnmt1^{co} C37 were prepared by Dr. Alexandra Mattei (Meissner Lab, Max Planck Institute for Molecular Genetics). The WGBS library for J1 DKO_0 was prepared by Chuck Haggerty (Meissner Lab, Max Planck Institute for Molecular Genetics). The WGBS libraries were sequenced on the NovaSeq6000 (Illumina) with 150 bp paired end reads by the Sequencing Core Facility of the Max Planck Institute for Molecular Genetics. The resulting data were processed by Dr. Helene Kretzmer (Meissner Lab, Max Planck Institute for Molecular Genetics) as described in Haggerty et al. (2021) [33].

7

References

[1] T. B. Johnson and R. D. Coghill. “The discovery of 5-methyl-cytosinc in tuberculinic acid, the nucleic acid of the tubercle bacillus”. In: *Journal of the American Chemical Society* 47.11 (1925), pp. 2838–2844. DOI: 10.1021/ja01688a030 (cit. on p. 1).

[2] Hotchkiss and R. D. Hotchkiss. “The quantitative separation of purines, pyrimidines and nucleosides by paper chromatography”. In: *Journal of Biological Chemistry* 175.1 (1948), pp. 315–332. DOI: 10.1016/S0021-9258(18)57261-6 (cit. on p. 1).

[3] D. B. Dunn and J. D. Smith. “The occurrence of 6-methylaminopurine in deoxyribonucleic acids”. In: *Biochemical Journal* 68.4 (1958), pp. 627–636. DOI: 10.1042/bj0680627 (cit. on p. 1).

[4] M. Ehrlich et al. “DNA methylation in thermophilic bacteria: N4-methylcytosine, 5-methylcytosine, and N6-methyladenine”. In: *Nucleic Acids Research* 13.4 (1985), pp. 1399–1412. DOI: 10.1093/nar/13.4.1399 (cit. on p. 1).

[5] D. Billen and R. Hewitt. “Influence of starvation for methionine and other amino acids on subsequent bacterial deoxyribonucleic acid replication”. In: *Journal of Bacteriology* 92.3 (1966), pp. 609–617. DOI: 10.1128/jb.92.3.609-617.1966 (cit. on p. 1).

[6] D. Billen. “Methylation of the Bacterial Chromosome: an Event at the “Replication Point”?” In: *Journal of Molecular Biology* 31.3 (1968), pp. 477–486. DOI: 10.1016/0022-2836(68)90422-1 (cit. on p. 1).

[7] C. Lark. “Studies on the in vivo methylation of DNA in Escherichia coli 15T”. In: *Journal of Molecular Biology* 31.3 (1968), pp. 389–399. DOI: 10.1016/0022-2836(68)90416-6 (cit. on p. 1).

[8] E. Fleissner and E. Borek. “A new enzyme of RNA synthesis: RNA methylase”. In: *PNAS* 48.7 (1962), pp. 1199–1203. DOI: 10.1073/pnas.48.7.1199 (cit. on p. 1).

[9] M. Gold et al. “The enzymic methylation of RNA and DNA. II. On the species-specificity of the methylation enzymes”. In: *PNAS* 50.1 (1963), pp. 164–169. DOI: 10.1073/pnas.50.1.164 (cit. on p. 2).

[10] W. Arber. “Host-Controlled Modification of Bacteriophage”. In: *Annual Review of Microbiology* 19 (1965), pp. 365–378. DOI: 10.1146/annurev.mi.19.100165.002053 (cit. on p. 2).

[11] R. J. Roberts et al. “REBASE–a database for DNA restriction and modification: enzymes, genes and genomes”. In: *Nucleic Acids Research* 43.Database issue (2015), pp. D298–9. DOI: 10.1093/nar/gku1046 (cit. on p. 2).

[12] T. S. Balganesh et al. “Construction and use of chimeric SPR/phi 3T DNA methyltransferases in the definition of sequence recognizing enzyme regions”. In: *EMBO Journal* 6.11 (1987), pp. 3543–3549. DOI: 10.1002/j.1460-2075.1987.tb02681.x (cit. on p. 2).

[13] T. A. Trautner et al. “Chimeric multispecific DNA methyltransferases with novel combinations of target recognition”. In: *Nucleic Acids Research* 16.14A (1988), pp. 6649–6658. DOI: 10.1093/nar/16.14.6649 (cit. on p. 2).

[14] K. Wilke et al. “Sequential order of target-recognizing domains in multispecific DNA-methyltransferases”. In: *EMBO Journal* 7.8 (1988), pp. 2601–2609. DOI: 10.1002/j.1460-2075.1988.tb03110.x (cit. on p. 2).

[15] J. Pósfai et al. “Predictive motifs derived from cytosine methyltransferases”. In: *Nucleic Acids Research* 17.7 (1989), pp. 2421–2435. DOI: 10.1093/nar/17.7.2421 (cit. on p. 2).

[16] L. Chen et al. “Direct identification of the active-site nucleophile in a DNA (cytosine-5)-methyltransferase”. In: *Biochemistry* 30.46 (1991), pp. 11018–11025. DOI: 10.1021/bi00110a002 (cit. on p. 2).

[17] M. J. Blow et al. “The Epigenomic Landscape of Prokaryotes”. In: *PLOS Genetics* 12.2 (2016), e1005854. DOI: 10.1371/journal.pgen.1005854 (cit. on p. 2).

[18] S. Tweedie et al. “Methylation of genomes and genes at the invertebrate-vertebrate boundary”. In: *Molecular and Cellular Biology* 17.3 (1997), pp. 1469–1475. DOI: 10.1128/MCB.17.3.1469 (cit. on p. 3).

[19] X. Zhang et al. “Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis”. In: *Cell* 126.6 (2006), pp. 1189–1201. DOI: 10.1016/j.cell.2006.08.003 (cit. on p. 3).

[20] D. Schübeler. “Function and information content of DNA methylation”. In: *Nature* 517.7534 (2015), pp. 321–326. DOI: 10.1038/nature14192 (cit. on p. 3).

[21] H. Zhang et al. “Dynamics and function of DNA methylation in plants”. In: *Nature Reviews Molecular Cell Biology* 19.8 (2018), pp. 489–506. DOI: 10.1038/s41580-018-0016-z (cit. on p. 3).

[22] A. de Mendoza et al. “Evolution of DNA Methylome Diversity in Eukaryotes”. In: *Journal of Molecular Biology* (2019). DOI: 10.1016/j.jmb.2019.11.003 (cit. on pp. 3–5).

[23] D. Buitrago et al. “Impact of DNA methylation on 3D genome structure”. In: *Nature Communications* 12.1 (2021), p. 3243. DOI: 10.1038/s41467-021-23142-8 (cit. on p. 3).

[24] M. M. Suzuki and A. Bird. “DNA methylation landscapes: provocative insights from epigenomics”. In: *Nature Reviews Genetics* 9.6 (2008), pp. 465–476. DOI: 10.1038/nrg2341 (cit. on p. 3).

[25] S. Sarda et al. “The evolution of invertebrate gene body methylation”. In: *Molecular Biology and Evolution* 29.8 (2012), pp. 1907–1916. DOI: 10.1093/molbev/mss062 (cit. on p. 3).

[26] V. J. Simpson et al. “Caenorhabditis elegans DNA does not contain 5-methylcytosine at any time during development or aging”. In: *Nucleic Acids Research* 14.16 (1986), pp. 6711–6719. DOI: 10.1093/nar/14.16.6711 (cit. on p. 3).

[27] T. E. Keller et al. “Evolutionary Transition of Promoter and Gene Body DNA Methylation across Invertebrate-Vertebrate Boundary”. In: *Molecular Biology and Evolution* 33.4 (2016), pp. 1019–1028. DOI: 10.1093/molbev/msv345 (cit. on p. 3).

[28] K. E. Webster et al. “Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis”. In: *PNAS* 102.11 (2005), pp. 4068–4073. DOI: 10.1073/pnas.0500702102 (cit. on p. 3).

[29] R. A. Rollins et al. “Large-scale structure of genomic methylation patterns”. In: *Genome Research* 16.2 (2006), pp. 157–163. DOI: 10.1101/gr.4362006 (cit. on pp. 3, 5).

[30] J. Klughammer et al. “Comparative analysis of genome-scale, base-resolution DNA methylation profiles across 580 animal species”. In: *bioRxiv* (2022). Preprint. DOI: 10.1101/2022.06.18.496602 (cit. on pp. 3, 105).

[31] M. V. C. Greenberg and D. Bourc'his. "The diverse roles of DNA methylation in mammalian development and disease". In: *Nature Reviews Molecular Cell Biology* 20.10 (2019), pp. 590–607. DOI: 10.1038/s41580-019-0159-6 (cit. on pp. 5, 106).

[32] Q. Wang et al. "Imprecise DNMT1 activity coupled with neighbor-guided correction enables robust yet flexible epigenetic inheritance". In: *Nature Genetics* 52.8 (2020), pp. 828–839. DOI: 10.1038/s41588-020-0661-y (cit. on pp. 5, 115).

[33] C. Haggerty et al. "Dnmt1 has de novo activity targeted to transposable elements". In: *Nature Structural & Molecular Biology* 28.7 (2021), pp. 594–603. DOI: 10.1038/s41594-021-00603-8 (cit. on pp. 5, 73, 78, 115–117, 153, 154).

[34] X. Guo et al. "Structural insight into autoinhibition and histone H3-induced activation of DNMT3A". In: *Nature* 517.7536 (2015), pp. 640–644. DOI: 10.1038/nature13899 (cit. on p. 5).

[35] M. Weber et al. "Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome". In: *Nature Genetics* 39.4 (2007), pp. 457–466. DOI: 10.1038/ng1990 (cit. on p. 5).

[36] T. Baubec et al. "Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation". In: *Nature* 520.7546 (2015), pp. 243–247. DOI: 10.1038/nature14176 (cit. on p. 5).

[37] D. N. Weinberg et al. "The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape". In: *Nature* 573.7773 (2019), pp. 281–286. DOI: 10.1038/s41586-019-1534-3 (cit. on p. 5).

[38] Q. Xu et al. "SETD2 regulates the maternal epigenome, genomic imprinting and embryonic development". In: *Nature Genetics* 51.5 (2019), pp. 844–856. DOI: 10.1038/s41588-019-0398-7 (cit. on pp. 5, 111).

[39] K. Shirane et al. "NSD1-deposited H3K36me2 directs de novo methylation in the mouse male germline and counteracts Polycomb-associated silencing". In: *Nature Genetics* 52.10 (2020), pp. 1088–1098. DOI: 10.1038/s41588-020-0689-z (cit. on pp. 5, 111).

[40] C. Mertineit et al. "Sex-specific exons control DNA methyltransferase in mammalian germ cells". In: *Development* 125.5 (1998), pp. 889–897. DOI: 10.1242/dev.125.5.889 (cit. on pp. 6, 109).

[41] T. Chen et al. "A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active de novo methylation". In: *The Journal of Biological Chemistry* 277.41 (2002), pp. 38746–38754. DOI: 10.1074/jbc.M205312200 (cit. on p. 6).

[42] Y. Tadokoro et al. "De novo DNA methyltransferase is essential for self-renewal, but not for differentiation, in hematopoietic stem cells". In: *The Journal of Experimental Medicine* 204.4 (2007), pp. 715–722. DOI: 10.1084/jem.20060750 (cit. on p. 6).

[43] D. J. Weisenberger et al. "Role of the DNA methyltransferase variant DNMT3b3 in DNA methylation". In: *Molecular Cancer Research* 2.1 (2004), pp. 62–72 (cit. on p. 6).

[44] R. Hirasawa et al. "Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development". In: *Genes & Development* 22.12 (2008), pp. 1607–1616. DOI: 10.1101/gad.1667008 (cit. on pp. 5, 107–109).

[45] J. Peters. "The role of genomic imprinting in biology and disease: an expanding view". In: *Nature Reviews Genetics* 15.8 (2014), pp. 517–530. DOI: 10.1038/nrg3766 (cit. on p. 7).

[46] S. Quenneville et al. “In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions”. In: *Molecular Cell* 44.3 (2011), pp. 361–372. DOI: 10.1016/j.molcel.2011.08.032 (cit. on pp. 7, 107).

[47] N. Takahashi et al. “ZFP57 and the Targeted Maintenance of Postfertilization Genomic Imprints”. In: *Cold Spring Harbor Symposia on Quantitative Biology* 80 (2015), pp. 177–187. DOI: 10.1101/sqb.2015.80.027466 (cit. on p. 7).

[48] D. M. Messerschmidt et al. “DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos”. In: *Genes & Development* 28.8 (2014), pp. 812–828. DOI: 10.1101/gad.234294.113 (cit. on pp. 7, 107).

[49] D. N. Weinberg et al. “Two competing mechanisms of DNMT3A recruitment regulate the dynamics of de novo DNA methylation at PRC1-targeted CpG islands”. In: *Nature Genetics* 53.6 (2021), pp. 794–800. DOI: 10.1038/s41588-021-00856-5 (cit. on pp. 7, 23, 32).

[50] T. Gu et al. “The disordered N-terminal domain of DNMT3A recognizes H2AK119ub and is required for postnatal development”. In: *Nature Genetics* (2022). DOI: 10.1038/s41588-022-01063-6 (cit. on pp. 7, 9, 98).

[51] M. R. Rountree et al. “DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci”. In: *Nature Genetics* 25.3 (2000), pp. 269–277. DOI: 10.1038/77023 (cit. on p. 7).

[52] F. Ding and J. R. Chaillet. “In vivo stabilization of the Dnmt1 (cytosine-5)- methyltransferase protein”. In: *PNAS* 99.23 (2002), pp. 14861–14866. DOI: 10.1073/pnas.232565599 (cit. on pp. 7, 109).

[53] J. Essers et al. “Nuclear dynamics of PCNA in DNA replication and repair”. In: *Molecular and Cellular Biology* 25.21 (2005), pp. 9350–9359. DOI: 10.1128/MCB.25.21.9350-9359.2005 (cit. on p. 8).

[54] L. S. Chuang et al. “Human DNA-(Cytosine-5) Methyltransferase-PCNA Complex as a Target for p21WAF1”. In: *Science* 277.5334 (1997), pp. 1996–2000. DOI: 10.1126/science.277.5334.1996 (cit. on p. 8).

[55] T. Iida et al. “PCNA clamp facilitates action of DNA cytosine methyltransferase 1 on hemimethylated DNA”. In: *Genes to Cells* 7.10 (2002), pp. 997–1007. DOI: 10.1046/j.1365-2443.2002.00584.x (cit. on p. 8).

[56] F. Spada et al. “DNMT1 but not its interaction with the replication machinery is required for maintenance of DNA methylation in human cells”. In: *Journal of Cell Biology* 176.5 (2007), pp. 565–571. DOI: 10.1083/jcb.200610062 (cit. on pp. 8, 95, 103).

[57] L. Schermelleh et al. “Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation”. In: *Nucleic Acids Research* 35.13 (2007), pp. 4301–4312. DOI: 10.1093/nar/gkm432 (cit. on pp. 8, 59, 68, 84, 95, 103, 111).

[58] A. Nishiyama et al. “Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication”. In: *Nature* 502.7470 (2013), pp. 249–253. DOI: 10.1038/nature12488 (cit. on pp. 8, 103).

[59] W. Qin et al. “DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination”. In: *Cell Research* 25.8 (2015), pp. 911–929. DOI: 10.1038/cr.2015.72 (cit. on pp. 8, 92).

[60] S. Ishiyama et al. “Structure of the Dnmt1 Reader Module Complexed with a Unique Two-Mono-Ubiquitin Mark on Histone H3 Reveals the Basis for DNA Methylation Maintenance”. In: *Molecular Cell* 68.2 (2017), 350–360.e7. DOI: 10.1016/j.molcel.2017.09.037 (cit. on pp. 8, 92, 102).

[61] T. Li et al. “Structural and mechanistic insights into UHRF1-mediated DNMT1 activation in the maintenance DNA methylation”. In: *Nucleic Acids Research* 46.6 (2018), pp. 3218–3231. DOI: 10.1093/nar/gky104 (cit. on p. 8).

[62] A. Nishiyama et al. “Two distinct modes of DNMT1 recruitment ensure stable maintenance DNA methylation”. In: *Nature Communications* 11.1 (2020), p. 1222. DOI: 10.1038/s41467-020-15006-4 (cit. on pp. 8, 41, 54, 57, 102, 103).

[63] X. Liu et al. “UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9”. In: *Nature Communications* 4 (2013), p. 1563. DOI: 10.1038/ncomms2562 (cit. on p. 8).

[64] W. Ren et al. “Direct readout of heterochromatic H3K9me3 regulates DNMT1-mediated maintenance DNA methylation”. In: *PNAS* 117.31 (2020), pp. 18439–18447. DOI: 10.1073/pnas.2009316117 (cit. on pp. 8, 9).

[65] F. Syeda et al. “The replication focus targeting sequence (RFTS) domain is a DNA-competitive inhibitor of Dnmt1”. In: *The Journal of Biological Chemistry* 286.17 (2011), pp. 15344–15351. DOI: 10.1074/jbc.M110.209882 (cit. on p. 8).

[66] M. Fatemi et al. “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”. In: *Journal of Molecular Biology* 309.5 (2001), pp. 1189–1199. DOI: 10.1006/jmbi.2001.4709 (cit. on pp. 8, 102).

[67] M. Pradhan et al. “CXXC domain of human DNMT1 is essential for enzymatic activity”. In: *Biochemistry* 47.38 (2008), pp. 10000–10009. DOI: 10.1021/bi8011725 (cit. on p. 8).

[68] J. Song et al. “Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation”. In: *Science* 331.6020 (2011), pp. 1036–1040. DOI: 10.1126/science.1195380 (cit. on p. 8).

[69] N. Yang and R.-M. Xu. “Structure and function of the BAH domain in chromatin biology”. In: *Critical Reviews in Biochemistry and Molecular Biology* 48.3 (2013), pp. 211–221. DOI: 10.3109/10409238.2012.742035 (cit. on p. 9).

[70] E. Borowczyk et al. “Identification of a region of the DNMT1 methyltransferase that regulates the maintenance of genomic imprints”. In: *PNAS* 106.49 (2009), pp. 20806–20811. DOI: 10.1073/pnas.0905668106 (cit. on pp. 9, 21, 22, 68, 92, 103).

[71] L. D’Aiuto et al. “Dissection of structure and function of the N-terminal domain of mouse DNMT1 using regional frame-shift mutagenesis”. In: *PLOS One* 5.3 (2010), e9831. DOI: 10.1371/journal.pone.0009831 (cit. on pp. 9, 21, 68).

[72] B. Shaffer et al. “The DNMT1 intrinsically disordered domain regulates genomic methylation during development”. In: *Genetics* 199.2 (2015), pp. 533–541. DOI: 10.1534/genetics.114.173609 (cit. on pp. 9, 21).

[73] Y. Shin and C. P. Brangwynne. “Liquid phase condensation in cell physiology and disease”. In: *Science* 357.6357 (2017). DOI: 10.1126/science.aaf4382 (cit. on pp. 10, 14).

[74] C. P. Brangwynne et al. “Germline P granules are liquid droplets that localize by controlled dissolution/condensation”. In: *Science* 324.5935 (2009), pp. 1729–1732. DOI: 10.1126/science.1172046 (cit. on p. 10).

[75] C. P. Brangwynne et al. “Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes”. In: *PNAS* 108.11 (2011), pp. 4334–4339. DOI: 10.1073/pnas.1017150108 (cit. on p. 10).

[76] P. Li et al. “Phase transitions in the assembly of multivalent signalling proteins”. In: *Nature* 483.7389 (2012), pp. 336–340. DOI: 10.1038/nature10879 (cit. on pp. 10, 12).

[77] A. Molliex et al. “Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization”. In: *Cell* 163.1 (2015), pp. 123–133. DOI: 10.1016/j.cell.2015.09.015 (cit. on p. 10).

[78] A. Patel et al. “A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation”. In: *Cell* 162.5 (2015), pp. 1066–1077. DOI: 10.1016/j.cell.2015.07.047 (cit. on p. 10).

[79] S. F. Banani et al. “Biomolecular condensates: organizers of cellular biochemistry”. In: *Nature Reviews Molecular Cell Biology* 18.5 (2017), pp. 285–298. DOI: 10.1038/nrm.2017.7 (cit. on pp. 10–12, 14, 15, 102).

[80] S. Alberti et al. “Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates”. In: *Cell* 176.3 (2019), pp. 419–434. DOI: 10.1016/j.cell.2018.12.035 (cit. on pp. 11, 12, 14, 37, 42, 96, 97, 102).

[81] Z. Gao et al. “Liquid-Liquid Phase Separation: Unraveling the Enigma of Biomolecular Condensates in Microbial Cells”. In: *Frontiers in Microbiology* 12 (2021), p. 751880. DOI: 10.3389/fmicb.2021.751880 (cit. on p. 11).

[82] M. M. Babu. “The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease”. In: *Biochemical Society Transactions* 44.5 (2016), pp. 1185–1200. DOI: 10.1042/BST20160172 (cit. on pp. 12, 21).

[83] A. C. Muntau et al. “Innovative strategies to treat protein misfolding in inborn errors of metabolism: pharmacological chaperones and proteostasis regulators”. In: *Journal of Inherited Metabolic Disease* 37.4 (2014), pp. 505–523. DOI: 10.1007/s10545-014-9701-z (cit. on p. 13).

[84] O. M. Morris et al. “Intrinsically disordered proteins: modes of binding with emphasis on disordered domains”. In: *Open Biology* 11.10 (2021), p. 210222. DOI: 10.1098/rsob.210222 (cit. on pp. 12, 13, 21, 37, 94, 96).

[85] E. W. Martin and T. Mittag. “Relationship of Sequence and Phase Separation in Protein Low-Complexity Regions”. In: *Biochemistry* 57.17 (2018), pp. 2478–2487. DOI: 10.1021/acs.biochem.8b00008 (cit. on pp. 13, 37, 41).

[86] K. V. Morris et al. “Small interfering RNA-induced transcriptional gene silencing in human cells”. In: *Science* 305.5688 (2004), pp. 1289–1292. DOI: 10.1126/science.1101372 (cit. on p. 13).

[87] E. W. Martin and A. S. Holehouse. “Intrinsically disordered protein regions and phase separation: sequence determinants of assembly or lack thereof”. In: *Emerging Topics in Life Sciences* 4.3 (2020), pp. 307–329. DOI: 10.1042/ETLS20190164 (cit. on pp. 14, 41).

[88] D. E. Hickford et al. “DDX4 (VASA) is conserved in germ cell development in marsupials and monotremes”. In: *Biology of Reproduction* 85.4 (2011), pp. 733–743. DOI: 10.1095/biolreprod.111.091629 (cit. on p. 14).

[89] S. Elbaum-Garfinkle et al. “The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics”. In: *PNAS* 112.23 (2015), pp. 7189–7194. DOI: 10.1073/pnas.1504822112 (cit. on p. 14).

[90] T. J. Nott et al. “Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles”. In: *Molecular Cell* 57.5 (2015), pp. 936–947. DOI: 10.1016/j.molcel.2015.01.013 (cit. on p. 14).

[91] M.-T. Wei et al. “Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles”. In: *Nature Chemistry* 9.11 (2017), pp. 1118–1125. DOI: 10.1038/nchem.2803 (cit. on p. 14).

[92] S. Das et al. “Comparative roles of charge, π, and hydrophobic interactions in sequence-dependent phase separation of intrinsically disordered proteins”. In: *PNAS* 117.46 (2020), pp. 28795–28805. DOI: 10.1073/pnas.2008122117 (cit. on p. 14).

[93] K. J. Meaburn and T. Misteli. “Cell biology: chromosome territories”. In: *Nature* 445.7126 (2007), pp. 379–781. DOI: 10.1038/445379a (cit. on p. 14).

[94] L. Guelen et al. “Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions”. In: *Nature* 453.7197 (2008), pp. 948–951. DOI: 10.1038/nature06947 (cit. on p. 14).

[95] D. Peric-Hupkes et al. “Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation”. In: *Molecular Cell* 38.4 (2010), pp. 603–613. DOI: 10.1016/j.molcel.2010.03.016 (cit. on p. 14).

[96] M. Wijgerde et al. “Transcription complex stability and chromatin dynamics in vivo”. In: *Nature* 377.6546 (1995), pp. 209–213. DOI: 10.1038/377209a0 (cit. on p. 15).

[97] N. Dillon et al. “The Effect of Distance on Long-Range Chromatin Interactions”. In: *Molecular Cell* 1.1 (1997), pp. 131–139. DOI: 10.1016/S1097-2765(00)80014-3 (cit. on p. 15).

[98] Z. Peng et al. “Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life”. In: *Cellular and Molecular Life Scienes* 72.1 (2015), pp. 137–151. DOI: 10.1007/s00018-014-1661-9 (cit. on pp. 15, 93).

[99] K. Zhou et al. “Nucleosome structure and dynamics are coming of age”. In: *Nature Structural & Molecular Biology* 26.1 (2019), pp. 3–13. DOI: 10.1038/s41594-018-0166-x (cit. on p. 15).

[100] D. P. Satijn et al. “RING1 is associated with the polycomb group protein complex and acts as a transcriptional repressor”. In: *Molecular and Cellular Biology* 17.7 (1997), pp. 4105–4113. DOI: 10.1128/MCB.17.7.4105 (cit. on p. 15).

[101] A. J. Saurin et al. “The human polycomb group complex associates with pericentromeric heterochromatin to form a novel nuclear domain”. In: *The Journal of Cell Biology* 142.4 (1998), pp. 887–898. DOI: 10.1083/jcb.142.4.887 (cit. on p. 15).

[102] A. Piunti and A. Shilatifard. “The roles of Polycomb repressive complexes in mammalian development and cancer”. In: *Nature Reviews Molecular Cell Biology* 22.5 (2021), pp. 326–345. DOI: 10.1038/s41580-021-00341-1 (cit. on p. 15).

[103] A. R. Strom et al. “Phase separation drives heterochromatin domain formation”. In: *Nature* 547.7662 (2017), pp. 241–245. DOI: 10.1038/nature22989 (cit. on pp. 15–17, 41, 97, 102).

[104] A. G. Larson et al. “Liquid droplet formation by HP1a suggests a role for phase separation in heterochromatin”. In: *Nature* 547.7662 (2017), pp. 236–240. DOI: 10.1038/nature22822 (cit. on pp. 15, 16, 41, 97, 102, 119).

[105] S. Sanulli et al. “HP1 reshapes nucleosome core to promote phase separation of heterochromatin”. In: *Nature* 575.7782 (2019), pp. 390–394. DOI: 10.1038/s41586-019-1669-2 (cit. on pp. 15, 16).

[106] J. D. Lewis et al. “Purification, sequence, and cellular localization of a novel chromosomal protein that binds to Methylated DNA”. In: *Cell* 69.6 (1992), pp. 905–914. DOI: 10.1016/0092-8674(92)90610-O (cit. on p. 16).

[107] C. H. Li et al. “MeCP2 links heterochromatin condensates and neurodevelopmental disease”. In: *Nature* 586.7829 (2020), pp. 440–444. DOI: 10.1038/s41586-020-2574-4 (cit. on pp. 16, 41, 97, 100).

[108] W. Zhang et al. “The ageing epigenome and its rejuvenation”. In: *Nature Reviews Molecular Cell Biology* 21.3 (2020), pp. 137–150. DOI: 10.1038/s41580-019-0204-5 (cit. on p. 16).

[109] A. Becker et al. “Direct homo- and hetero-interactions of MeCP2 and MBD2”. In: *PLOS One* 8.1 (2013), e53730. DOI: 10.1371/journal.pone.0053730 (cit. on p. 16).

[110] L. Wang et al. “Rett syndrome-causing mutations compromise MeCP2-mediated liquid-liquid phase separation of chromatin”. In: *Cell Research* 30.5 (2020), pp. 393–407. DOI: 10.1038/s41422-020-0288-7 (cit. on pp. 16, 100).

[111] R. P. Ghosh et al. “Unique physical properties and interactions of the domains of methylated DNA binding protein 2”. In: *Biochemistry* 49.20 (2010), pp. 4395–4410. DOI: 10.1021/bi9019753 (cit. on p. 16).

[112] H. Zhang et al. “MeCP2-induced heterochromatin organization is driven by oligomerization-based liquid-liquid phase separation and restricted by DNA methylation”. In: *Nucleus* 13.1 (2022), pp. 1–34. DOI: 10.1080/19491034.2021.2024691 (cit. on pp. 16, 17, 97, 100, 101).

[113] H. Fan et al. “BAHCC1 binds H3K27me3 via a conserved BAH module to mediate gene silencing and oncogenesis”. In: *Nature Genetics* 52.12 (2020), pp. 1384–1396. DOI: 10.1038/s41588-020-00729-3 (cit. on pp. 16, 100).

[114] M. M. Keenen et al. “HP1 proteins compact DNA into mechanically and positionally stable phase separated domains”. In: *eLife* 10 (2021). DOI: 10.7554/eLife.64563 (cit. on p. 16).

[115] A. D. Stephens et al. “Chromatin’s physical properties shape the nucleus and its functions”. In: *Current Opinion in Cell Biology* 58 (2019), pp. 76–84. DOI: 10.1016/j.ceb.2019.02.006 (cit. on p. 16).

[116] M. W. Parker et al. “A new class of disordered elements controls DNA replication through initiator self-assembly”. In: *eLife* 8 (2019). DOI: 10.7554/eLife.48562 (cit. on pp. 16–18, 37, 38, 41, 97, 100, 102, 119).

[117] V. Tsakraklides and S. P. Bell. “Dynamics of pre-replicative complex assembly”. In: *The Journal of Biological Chemistry* 285.13 (2010), pp. 9437–9443. DOI: 10.1074/jbc.M109.072504 (cit. on p. 17).

[118] M. W. Parker et al. “Molecular determinants of phase separation for Drosophila DNA replication licensing factors”. In: *eLife* 10 (2021). DOI: 10.7554/eLife.70535 (cit. on pp. 17, 18, 100, 102).

[119] M. Findeisen et al. “Cyclin A-dependent kinase activity affects chromatin binding of ORC, Cdc6, and MCM in egg extracts of Xenopus laevis”. In: *European Journal of Biochemistry* 264.2 (1999), pp. 415–426. DOI: 10.1046/j.1432-1327.1999.00613.x (cit. on p. 17).

[120] K. Y. Lee et al. “Phosphorylation of ORC2 protein dissociates origin recognition complex from chromatin and replication origins”. In: *The Journal of Biological Chemistry* 287.15 (2012), pp. 11891–11898. DOI: 10.1074/jbc.M111.338467 (cit. on p. 17).

[121] C.-J. Li et al. “Role for Cdk1 (Cdc2)/cyclin A in preventing the mammalian origin recognition complex’s largest subunit (Orc1) from binding to chromatin during mitosis”. In: *Molecular and Cellular Biology* 24.13 (2004), pp. 5875–5886. DOI: 10.1128/MCB.24.13.5875-5886.2004 (cit. on p. 17).

[122] A. Rowles et al. “Changes in association of the Xenopus origin recognition complex with chromatin on licensing of replication origins”. In: *Journal of Cell Science* 112 (Pt 12) (1999), pp. 2011–2018. DOI: 10.1242/jcs.112.12.2011 (cit. on p. 17).

[123] A. J. Kuo et al. “The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome”. In: *Nature* 484.7392 (2012), pp. 115–119. DOI: 10.1038/nature10956 (cit. on p. 17).

[124] H. Leonhardt et al. “Dynamics of DNA replication factories in living cells”. In: *The Journal of Cell Biology* 149.2 (2000), pp. 271–280. DOI: 10.1083/jcb.149.2.271 (cit. on p. 18).

[125] A. Sporbert et al. “DNA Polymerase Clamp Shows Little Turnover at Established Replication Sites but Sequential De Novo Assembly at Adjacent Origin Clusters”. In: *Molecular Cell* 10.6 (2002), pp. 1355–1365. DOI: 10.1016/S1097-2765(02)00729-3 (cit. on p. 18).

[126] A. Maya-Mendoza et al. “S phase progression in human cells is dictated by the genetic continuity of DNA foci”. In: *PLOS Genetics* 6.4 (2010), e1000900. DOI: 10.1371/journal.pgen.1000900 (cit. on p. 18).

[127] G. M. Cooper and C. D. Brown. “Qualifying the relationship between sequence conservation and molecular function”. In: *Genome Research* 18.2 (2008), pp. 201–205. DOI: 10.1101/gr.7205808 (cit. on p. 21).

[128] Y. Liu et al. “A comprehensive review and comparison of existing computational methods for intrinsically disordered protein and region prediction”. In: *Briefings in bioinformatics* 20.1 (2019), pp. 330–346. DOI: 10.1093/bib/bbx126 (cit. on p. 22).

[129] J. J. Ward et al. “Prediction and functional analysis of native disorder in proteins from the three kingdoms of life”. In: *Journal of Molecular Biology* 337.3 (2004), pp. 635–645. DOI: 10.1016/j.jmb.2004.02.002 (cit. on pp. 22, 92).

[130] H. M. Berman et al. “The Protein Data Bank”. In: *Nucleic Acids Research* 28.1 (2000), pp. 235–242. DOI: 10.1093/nar/28.1.235 (cit. on pp. 22, 92).

[131] J. T. Nielsen and F. A. A. Mulder. “Quality and bias of protein disorder predictors”. In: *Scientific Reports* 9.1 (2019), p. 5137. DOI: 10.1038/s41598-019-41644-w (cit. on pp. 22, 92).

[132] R. J. Emenecker et al. “Metapredict: a fast, accurate, and easy-to-use predictor of consensus disorder and structure”. In: *Biophysical Journal* 120.20 (2021), pp. 4312–4319. DOI: 10.1016/j.bpj.2021.08.039 (cit. on pp. 22, 123).

[133] J. Jumper et al. “Highly accurate protein structure prediction with AlphaFold”. In: *Nature* 596.7873 (2021), pp. 583–589. DOI: 10.1038/s41586-021-03819-2 (cit. on pp. 22, 23).

[134] M. Varadi et al. “AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models”. In: *Nucleic Acids Research* 50.D1 (2022), pp. D439–D444. DOI: 10.1093/nar/gkab1061 (cit. on pp. 23, 92).

[135] N. Malhis et al. “Improved measures for evolutionary conservation that exploit taxonomy distances”. In: *Nature Communications* 10.1 (2019), p. 1556. DOI: 10.1038/s41467-019-09583-2 (cit. on p. 27).

[136] NCBI. *National Center for Biotechnology Information (NCBI)*. Bethesda (MD), 1988 (cit. on pp. 27, 124).

[137] R. M. Roberts et al. “The evolution of the placenta”. In: *Reproduction* 152.5 (2016), R179–89. DOI: 10.1530/REP-16-0325 (cit. on pp. 29, 106).

[138] The UniProt Consortium. “UniProt: the universal protein knowledgebase in 2021”. In: *Nucleic Acids Research* 49.D1 (2021), pp. D480–D489. DOI: 10.1093/nar/gkaa1100 (cit. on pp. 31–33, 92, 123, 124).

[139] R. K. Gaur. “Amino acid frequency distribution among eukaryotic proteins”. In: *IIOAB Journal* 2 (2014), pp. 6–11 (cit. on p. 37).

[140] H. Hou et al. “Single-cell pH imaging and detection for pH profiling and label-free rapid identification of cancer-cells”. In: *Scientific Reports* 7.1 (2017), p. 1759. DOI: 10.1038/s41598-017-01956-1 (cit. on p. 37).

[141] A. J. Plys et al. “Phase separation of Polycomb-repressive complex 1 is governed by a charged disordered region of CBX2”. In: *Genes & Development* 33.13-14 (2019), pp. 799–813. DOI: 10.1101/gad.326488.119 (cit. on p. 38).

[142] R. F. Ryan et al. “KAP-1 corepressor protein interacts and colocalizes with heterochromatic and euchromatic HP1 proteins: a potential role for Krüppel-associated box-zinc finger proteins in heterochromatin-mediated gene silencing”. In: *Molecular and Cellular Biology* 19.6 (1999), pp. 4366–4378. DOI: 10.1128/MCB.19.6.4366 (cit. on p. 38).

[143] L. Wang et al. “Histone Modifications Regulate Chromatin Compartmentalization by Contributing to a Phase Separation Mechanism”. In: *Molecular Cell* 76.4 (2019), 646–659.e6. DOI: 10.1016/j.molcel.2019.08.019 (cit. on p. 38).

[144] M. Kato et al. “Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels”. In: *Cell* 149.4 (2012), pp. 753–767. DOI: 10.1016/j.cell.2012.04.017 (cit. on p. 41).

[145] Y. Shin et al. “Spatiotemporal Control of Intracellular Phase Transitions Using Light-Activated optoDroplets”. In: *Cell* 168.1-2 (2017), 159–171.e14. DOI: 10.1016/j.cell.2016.11.054 (cit. on pp. 41, 46, 48).

[146] O. Mortusewicz et al. “Recruitment of DNA methyltransferase I to DNA repair sites”. In: *PNAS* 102.25 (2005), pp. 8905–8909. DOI: 10.1073/pnas.0501034102 (cit. on pp. 41, 102).

[147] K. Ha et al. “Rapid and transient recruitment of DNMT1 to DNA double-strand breaks is mediated by its interaction with multiple components of the DNA damage response machinery”. In: *Human Molecular Genetics* 20.1 (2011), pp. 126–140. DOI: 10.1093/hmg/ddq451 (cit. on p. 41).

[148] V. Spegg and M. Altmeyer. “Biomolecular condensates at sites of DNA damage: More than just a phase”. In: *DNA Repair* 106 (2021), p. 103179. DOI: 10.1016/j.dnarep.2021.103179 (cit. on pp. 41, 102).

[149] H. Guo et al. “Regulation of flowering time by Arabidopsis photoreceptors”. In: *Science* 279.5355 (1998), pp. 1360–1363. DOI: 10.1126/science.279.5355.1360 (cit. on p. 46).

[150] P. Más et al. “Functional interaction of phytochrome B and cryptochrome 2”. In: *Nature* 408.6809 (2000), pp. 207–211. DOI: 10.1038/35041583 (cit. on p. 46).

[151] L. J. Bugaj et al. “Optogenetic protein clustering and signaling activation in mammalian cells”. In: *Nature Methods* 10.3 (2013), pp. 249–252. DOI: 10.1038/nmeth.2360 (cit. on p. 46).

[152] W. Mo et al. “Arabidopsis cryptochrome 2 forms photobodies with TCP22 under blue light and regulates the circadian clock”. In: *Nature Communications* 13.1 (2022), p. 2631. DOI: 10.1038/s41467-022-30231-9 (cit. on p. 46).

[153] Q. Wang and C. Lin. “A structural view of plant CRY2 photoactivation and inactivation”. In: *Nature Structural & Molecular Biology* 27.5 (2020), pp. 401–403. DOI: 10.1038/s41594-020-0432-6 (cit. on p. 47).

[154] K. L. Tucker et al. “Germ-line passage is required for establishment of methylation and expression patterns of imprinted but not of nonimprinted genes”. In: *Genes & Development* 10.8 (1996), pp. 1008–1020. DOI: 10.1101/gad.10.8.1008 (cit. on pp. 68, 113).

[155] R. G. Garvilles et al. “Dual Functions of the RFTS Domain of Dnmt1 in Replication-Coupled DNA Methylation and in Protection of the Genome from Aberrant Methylation”. In: *PLOS One* 10.9 (2015), e0137509. DOI: 10.1371/journal.pone.0137509 (cit. on p. 68).

[156] B. Nabet et al. “The dTAG system for immediate and target-specific protein degradation”. In: *Nature Chemical Biology* 14.5 (2018), pp. 431–441. DOI: 10.1038/s41589-018-0021-8 (cit. on p. 69).

[157] Z. D. Smith et al. “Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer”. In: *Nature* 549.7673 (2017), pp. 543–547. DOI: 10.1038/nature23891 (cit. on pp. 70, 111).

[158] T. Dahlet et al. “Genome-wide analysis in the mouse embryo reveals the importance of DNA methylation for transcription integrity”. In: *Nature Communications* 11.1 (2020), p. 3153. DOI: 10.1038/s41467-020-16919-w (cit. on pp. 70, 111).

[159] Q. Chen et al. “Structural basis of seamless excision and specific targeting by piggyBac transposase”. In: *Nature Communications* 11.1 (2020), p. 3446. DOI: 10.1038/s41467-020-17128-1 (cit. on p. 74).

[160] M. Seczynska et al. “Genome surveillance by HUSH-mediated silencing of intronless mobile elements”. In: *Nature* 601.7893 (2022), pp. 440–445. DOI: 10.1038/s41586-021-04228-1 (cit. on pp. 82, 113).

[161] I. A. Tchasovnikarova et al. “Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells”. In: *Science* 348.6242 (2015), pp. 1481–1485. DOI: 10.1126/science.aaa7227 (cit. on p. 82).

[162] A. Lacy-Hulbert et al. “Interruption of coding sequences by heterologous introns can enhance the functional expression of recombinant genes”. In: *Gene Therapy* 8.8 (2001), pp. 649–653. DOI: 10.1038/sj.gt.3301440 (cit. on pp. 82, 83).

[163] C. L. Worth et al. “Structural and functional constraints in the evolution of protein families”. In: *Nature Reviews Molecular Cell Biology* 10.10 (2009), pp. 709–720. DOI: 10.1038/nrm2762 (cit. on p. 93).

[164] T. Zarin et al. “Proteome-wide signatures of function in highly diverged intrinsically disordered regions”. In: *eLife* 8 (2019). DOI: 10.7554/eLife.46883 (cit. on pp. 93, 104).

[165] A. Schlessinger et al. “Protein disorder–a breakthrough invention of evolution?” In: *Current Opinion in Structural Biology* 21.3 (2011), pp. 412–418. DOI: 10.1016/j.sbi.2011.03.014 (cit. on p. 93).

[166] G. L. Dignon et al. “Sequence determinants of protein phase behavior from a coarse-grained model”. In: *PLOS Computational Biology* 14.1 (2018), e1005941. DOI: 10.1371/journal.pcbi.1005941 (cit. on p. 93).

[167] F. G. Quiroz and A. Chilkoti. “Sequence heuristics to encode phase behaviour in intrinsically disordered protein polymers”. In: *Nature Materials* 14.11 (2015), pp. 1164–1171. DOI: 10.1038/nmat4418 (cit. on p. 93).

[168] N. S. González-Foutel et al. “Conformational buffering underlies functional selection in intrinsically disordered protein regions”. In: *Nature Structural & Molecular Biology* 29.8 (2022), pp. 781–790. DOI: 10.1038/s41594-022-00811-w (cit. on p. 94).

[169] A. Prestel et al. “The PCNA interaction motifs revisited: thinking outside the PIP-box”. In: *Cellular and Molecular Life Scienes* 76.24 (2019), pp. 4923–4943. DOI: 10.1007/s00018-019-03150-0 (cit. on pp. 94, 95).

[170] T. Jimenji et al. “Structure of PCNA in complex with DNMT1 PIP box reveals the basis for the molecular mechanism of the interaction”. In: *Biochemical and Biophysical Research Communications* 516.2 (2019), pp. 578–583. DOI: 10.1016/j.bbrc.2019.06.060 (cit. on p. 94).

[171] G.-L. Moldovan et al. “PCNA, the maestro of the replication fork”. In: *Cell* 129.4 (2007), pp. 665–679. DOI: 10.1016/j.cell.2007.05.003 (cit. on p. 94).

[172] W. Strzalka and A. Ziemienowicz. “Proliferating cell nuclear antigen (PCNA): a key factor in DNA replication and cell cycle regulation”. In: *Annals of Botany* 107.7 (2011), pp. 1127–1140. DOI: 10.1093/aob/mcq243 (cit. on p. 94).

[173] A. J. Horsfall et al. “Unlocking the PIP-box: A peptide library reveals interactions that drive high-affinity binding to human PCNA”. In: *The Journal of Biological Chemistry* 296 (2021), p. 100773. DOI: 10.1016/j.jbc.2021.100773 (cit. on p. 94).

[174] S. Chong and M. Mir. “Towards Decoding the Sequence-Based Grammar Governing the Functions of Intrinsically Disordered Protein Regions”. In: *Journal of Molecular Biology* 433.12 (2021), p. 166724. DOI: 10.1016/j.jmb.2020.11.023 (cit. on p. 96).

[175] J. Wang et al. “Structural insights into DNMT5-mediated ATP-dependent high-fidelity epigenome maintenance”. In: *Molecular Cell* 82.6 (2022), 1186–1198.e6. DOI: 10.1016/j.molcel.2022.01.028 (cit. on p. 97).

[176] K. Schneider et al. “Dissection of cell cycle-dependent dynamics of Dnmt1 by FRAP and diffusion-coupled modeling”. In: *Nucleic Acids Research* 41.9 (2013), pp. 4860–4876. DOI: 10.1093/nar/gkt191 (cit. on p. 99).

[177] J. Zhang et al. “S phase-dependent interaction with DNMT1 dictates the role of UHRF1 but not UHRF2 in DNA methylation maintenance”. In: *Cell Research* 21.12 (2011), pp. 1723–1739. DOI: 10.1038/cr.2011.176 (cit. on p. 99).

[178] F. Schönenberger et al. “Discrimination of cell cycle phases in PCNA-immunolabeled cells”. In: *BMC Bioinformatics* 16 (2015), p. 180. DOI: 10.1186/s12859-015-0618-9 (cit. on p. 99).

[179] A. Hermann et al. “The Dnmt1 DNA-(cytosine-C5)-methyltransferase Methylates DNA Processively with High Preference for Hemimethylated Target Sites”. In: *Journal of Biological Chemistry* 279.46 (2004), pp. 48350–48359. DOI: 10.1074/jbc.M403427200 (cit. on pp. 102, 121).

[180] R. Goyal et al. “Accuracy of DNA methylation pattern preservation by the Dnmt1 methyltransferase”. In: *Nucleic Acids Research* 34.4 (2006), pp. 1182–1188. DOI: 10.1093/nar/gkl002 (cit. on pp. 102, 121).

[181] S. Adam et al. “DNA sequence-dependent activity and base flipping mechanisms of DNMT1 regulate genome-wide DNA methylation”. In: *Nature Communications* 11.1 (2020), p. 3723. DOI: 10.1038/s41467-020-17531-8 (cit. on pp. 102, 114, 121).

[182] G. Vilkaitis et al. “Processive methylation of hemimethylated CpG sites by mouse Dnmt1 DNA methyltransferase”. In: *Journal of Biological Chemistry* 280.1 (2005), pp. 64–72. DOI: 10.1074/jbc.M411126200 (cit. on pp. 102, 121).

[183] B. Alberts. *Molecular Biology of the Cell*. 4th ed. New York: Garland Science and Garland, 2002 (cit. on p. 103).

[184] S. Pradhan et al. “Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation”. In: *Journal of Biological Chemistry* 274.46 (1999), pp. 33002–33010. DOI: 10.1074/jbc.274.46.33002 (cit. on p. 103).

[185] J. Charlton et al. “Global delay in nascent strand DNA methylation”. In: *Nature Structural & Molecular Biology* 25.4 (2018), pp. 327–332. DOI: 10.1038/s41594-018-0046-4 (cit. on pp. 103, 120).

[186] C. Xu and V. G. Corces. “Nascent DNA methylome mapping reveals inheritance of hemimethylation at CTCF/cohesin sites”. In: *Science* 359.6380 (2018), pp. 1166–1170. DOI: 10.1126/science.aan5480 (cit. on p. 103).

[187] O. R. P. Bininda-Emonds et al. “The delayed rise of present-day mammals”. In: *Nature* 446.7135 (2007), pp. 507–512. DOI: 10.1038/nature05634 (cit. on p. 103).

[188] E. D. Ingles and J. E. Deakin. “Global DNA Methylation patterns on marsupial and devil facial tumour chromosomes”. In: *Molecular Cytogenetics* 8 (2015), p. 74. DOI: 10.1186/s13039-015-0176-x (cit. on p. 105).

[189] M. A. Gama-Sosa et al. “The 5-methylcytosine content of DNA from human tumors”. In: *Nucleic Acids Research* 11.19 (1983), pp. 6883–6894. DOI: 10.1093/nar/11.19.6883 (cit. on p. 106).

[190] C. Popp et al. “Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency”. In: *Nature* 463.7284 (2010), pp. 1101–1105. DOI: 10.1038/nature08829 (cit. on p. 106).

[191] D. I. Schroeder et al. “The human placenta methylome”. In: *PNAS* 110.15 (2013), pp. 6037–6042. DOI: 10.1073/pnas.1215145110 (cit. on p. 106).

[192] D. I. Schroeder and J. M. LaSalle. “How has the study of the human placenta aided our understanding of partially methylated genes?” In: *Epigenomics* 5.6 (2013), pp. 645–654. DOI: 10.2217/epi.13.62 (cit. on p. 106).

[193] W. Zhou et al. “DNA methylation loss in late-replicating domains is linked to mitotic cell division”. In: *Nature Genetics* 50.4 (2018), pp. 591–602. DOI: 10.1038/s41588-018-0073-4 (cit. on p. 106).

[194] D. I. Schroeder et al. “Early Developmental and Evolutionary Origins of Gene Body DNA Methylation Patterns in Mammalian Placentas”. In: *PLOS Genetics* 11.8 (2015), e1005442. DOI: 10.1371/journal.pgen.1005442 (cit. on p. 106).

[195] M. A. M. Cleaton et al. “Phenotypic outcomes of imprinted gene models in mice: elucidation of pre- and postnatal functions of imprinted genes”. In: *Annual Review of Genomics and Human Genetics* 15 (2014), pp. 93–126. DOI: 10.1146/annurev-genom-091212-153441 (cit. on p. 106).

[196] C. A. Edwards et al. “The origins of genomic imprinting in mammals”. In: *Reproduction, Fertility, and Development* 31.7 (2019), pp. 1203–1218. DOI: 10.1071/RD18176 (cit. on pp. 106–108).

[197] M. B. Renfree et al. “The origin and evolution of genomic imprinting and viviparity in mammals”. In: *Philosophical Transactions of the Royal Society B* 368.1609 (2013), p. 20120151. DOI: 10.1098/rstb.2012.0151 (cit. on p. 106).

[198] A. C. Ferguson-Smith and D. Bourc’his. “The discovery and importance of genomic imprinting”. In: *eLife* 7 (2018). DOI: 10.7554/eLife.42368 (cit. on p. 107).

[199] R. Strogantsev et al. “Allele-specific binding of ZFP57 in the epigenetic regulation of imprinted and non-imprinted monoallelic expression”. In: *Genome Biology* 16 (2015), p. 112. DOI: 10.1186/s13059-015-0672-7 (cit. on p. 107).

[200] F. Gaudet et al. “Dnmt1 Expression in Pre- and Postimplantation Embryogenesis and the Maintenance of IAP Silencing”. In: *Molecular and Cellular Biology* 24.4 (2004), pp. 1640–1648. DOI: 10.1128/MCB.24.4.1640-1648.2004 (cit. on p. 108).

[201] Y. Wei et al. “Unfaithful maintenance of methylation imprints due to loss of maternal nuclear Dnmt1 during somatic cell nuclear transfer”. In: *PLOS One* 6.5 (2011), e20154. DOI: 10.1371/journal.pone.0020154 (cit. on p. 108).

[202] Y. Kurihara et al. “Maintenance of genomic methylation patterns during preimplantation development requires the somatic form of DNA methyltransferase 1”. In: *Developmental Biology* 313.1 (2008), pp. 335–346. DOI: 10.1016/j.ydbio.2007.10.033 (cit. on pp. 108, 109).

[203] D. Bourc’his et al. “Delayed and incomplete reprogramming of chromosome methylation patterns in bovine cloned embryos”. In: *Current Biology* 11.19 (2001), pp. 1542–1546. DOI: 10.1016/S0960-9822(01)00480-8 (cit. on p. 108).

[204] W. Dean et al. “Conservation of methylation reprogramming in mammalian development: Aberrant reprogramming in cloned embryos”. In: *PNAS* 98.24 (2001), pp. 13734–13738. DOI: 10.1073/pnas.241522698 (cit. on p. 108).

[205] X. Yang et al. “Nuclear reprogramming of cloned embryos and its implications for therapeutic cloning”. In: *Nature Genetics* 39.3 (2007), pp. 295–302. DOI: 10.1038/ng1973 (cit. on p. 108).

[206] X. Wang et al. “Epigenetic Reprogramming During Somatic Cell Nuclear Transfer: Recent Progress and Future Directions”. In: *Frontiers in Genetics* 11 (2020), p. 205. DOI: 10.3389/fgene.2020.00205 (cit. on p. 108).

[207] C. Y. Howell et al. “Genomic Imprinting Disrupted by a Maternal Effect Mutation in the Dnmt1 Gene”. In: *Cell* 104.6 (2001), pp. 829–838. DOI: 10.1016/S0092-8674(01)00280-X (cit. on p. 109).

[208] S. Ratnam et al. “Dynamics of Dnmt1 Methyltransferase Expression and Intracellular Localization during Oogenesis and Preimplantation Development”. In: *Developmental Biology* 245.2 (2002), pp. 304–314. DOI: 10.1006/dbio.2002.0628 (cit. on p. 109).

[209] M. C. Cirio et al. “Preimplantation expression of the somatic form of Dnmt1 suggests a role in the inheritance of genomic imprints”. In: *BMC Developmental Biology* 8 (2008), p. 9. DOI: 10.1186/1471-213X-8-9 (cit. on p. 109).

[210] B. Min et al. “Dnmt1 binds and represses genomic retroelements via DNA methylation in mouse early embryos”. In: *Nucleic Acids Research* 48.15 (2020), pp. 8431–8444. DOI: 10.1093/nar/gkaa584 (cit. on p. 109).

[211] S. Röth et al. “Advances in targeted degradation of endogenous proteins”. In: *Cellular and Molecular Life Scienes* 76.14 (2019), pp. 2761–2777. DOI: 10.1007/s00018-019-03112-6 (cit. on p. 110).

[212] E. F. Douglass et al. “A comprehensive mathematical model for three-body binding equilibria”. In: *Journal of the American Chemical Society* 135.16 (2013), pp. 6092–6099. DOI: 10.1021/ja311795d (cit. on p. 110).

[213] A. Kanhere and M. Bansal. “Structural properties of promoters: similarities and differences between prokaryotes and eukaryotes”. In: *Nucleic Acids Research* 33.10 (2005), pp. 3165–3175. DOI: 10.1093/nar/gki627 (cit. on p. 112).

[214] M. J. Luo and R. Reed. “Splicing is required for rapid and efficient mRNA export in metazoans”. In: *PNAS* 96.26 (1999), pp. 14937–14942. DOI: 10.1073/pnas.96.26.14937 (cit. on p. 113).

[215] N. Kataoka et al. “Pre-mRNA Splicing Imprints mRNA in the Nucleus with a Novel RNA-Binding Protein that Persists in the Cytoplasm”. In: *Molecular Cell* 6.3 (2000), pp. 673–682. DOI: 10.1016/S1097-2765(00)00065-4 (cit. on p. 113).

[216] Z. Zhou et al. “The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans”. In: *Nature* 407.6802 (2000), pp. 401–405. DOI: 10.1038/35030160 (cit. on p. 113).

[217] Y. Huang and J. A. Steitz. “Splicing Factors SRp20 and 9G8 Promote the Nucleocytoplasmic Export of mRNA”. In: *Molecular Cell* 7.4 (2001), pp. 899–905. DOI: 10.1016/S1097-2765(01)00233-7 (cit. on p. 113).

[218] J. P. Rodrigues et al. “REF proteins mediate the export of spliced and unspliced mRNAs from the nucleus”. In: *PNAS* 98.3 (2001), pp. 1030–1035. DOI: 10.1073/pnas.98.3.1030 (cit. on p. 113).

[219] M. L. Luo et al. “Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly”. In: *Nature* 413.6856 (2001), pp. 644–647. DOI: 10.1038/35098106 (cit. on p. 113).

[220] P. Valencia et al. “Splicing promotes rapid and efficient mRNA export in mammalian cells”. In: *PNAS* 105.9 (2008), pp. 3386–3391. DOI: 10.1073/pnas.0800250105 (cit. on p. 113).

[221] P. Björk and L. Wieslander. “Integration of mRNP formation and export”. In: *Cellular and Molecular Life Scienes* 74.16 (2017), pp. 2875–2897. DOI: 10.1007/s00018-017-2503-3 (cit. on p. 113).

[222] S. Kramer. “Nuclear mRNA maturation and mRNA export control: from trypanosomes to opisthokonts”. In: *Parasitology* 148.10 (2021), pp. 1196–1218. DOI: 10.1017/S0031182021000068 (cit. on p. 113).

[223] K. L. Tucker et al. “Complementation of methylation deficiency in embryonic stem cells by a DNAmethyltransferaseminigene”. In: *PNAS* 93.23 (1996), pp. 12920–12925. DOI: 10.1073/pnas.93.23.12920 (cit. on p. 113).

[224] F. Gaudet et al. “A short DNA methyltransferase isoform restores methylation in vivo”. In: *Journal of Biological Chemistry* 273.49 (1998), pp. 32725–32729. DOI: 10.1074/jbc.273.49.32725 (cit. on p. 113).

[225] D. Biniszkiewicz et al. “Dnmt1 overexpression causes genomic hypermethylation, loss of imprinting, and embryonic lethality”. In: *Molecular and Cellular Biology* 22.7 (2002), pp. 2124–2135. DOI: 10.1128/MCB.22.7.2124-2135.2002 (cit. on p. 113).

[226] M. Damelin and T. H. Bestor. “Biological functions of DNA methyltransferase 1 require its methyltransferase activity”. In: *Molecular and Cellular Biology* 27.11 (2007), pp. 3891–3899. DOI: 10.1128/MCB.00036-07 (cit. on p. 113).

[227] O. Yarychkivska et al. “BAH domains and a histone-like motif in DNA methyltransferase 1 (DNMT1) regulate de novo and maintenance methylation in vivo”. In: *The Journal of Biological Chemistry* 293.50 (2018), pp. 19466–19475. DOI: 10.1074/jbc.RA118.004612 (cit. on p. 113).

[228] D. Ortega-Alarcon et al. “Stabilization Effect of Intrinsically Disordered Regions on Multidomain Proteins: The Case of the Methyl-CpG Protein 2, MeCP2”. In: *Biomolecules* 11.8 (2021). DOI: 10.3390/biom11081216 (cit. on p. 114).

[229] Ö. Deniz et al. “Regulation of transposable elements by DNA modifications”. In: *Nature Reviews Genetics* 20.7 (2019), pp. 417–431. DOI: 10.1038/s41576-019-0106-6 (cit. on p. 115).

[230] C. P. Walsh and T. H. Bestor. “Cytosine methylation and mammalian development”. In: *Genes & Development* 13.1 (1999), pp. 26–34. DOI: 10.1101/gad.13.1.26 (cit. on p. 117).

[231] V. Asimi et al. “Hijacking of transcriptional condensates by endogenous retroviruses”. In: *Nature Genetics* 54.8 (2022), pp. 1238–1247. DOI: 10.1038/s41588-022-01132-w (cit. on p. 117).

[232] S. Sridharan et al. “Systematic discovery of biomolecular condensate-specific protein phosphorylation”. In: *Nature Chemical Biology* 18.10 (2022), pp. 1104–1114. DOI: 10.1038/s41589-022-01062-y (cit. on p. 119).

[233] Z. Du et al. “DNMT1 stability is regulated by proteins coordinating deubiquitination and acetylation-driven ubiquitination”. In: *Science Signaling* 3.146 (2010), ra80. DOI: 10.1126/scisignal.2001462 (cit. on p. 120).

[234] P.-O. Estève et al. “A methylation and phosphorylation switch between an adjacent lysine and serine determines human DNMT1 stability”. In: *Nature Structural & Molecular Biology* 18.1 (2011), pp. 42–48. DOI: 10.1038/nsmb.1939 (cit. on p. 120).

[235] P. V. Hornbeck et al. “PhosphoSitePlus, 2014: mutations, PTMs and recalibrations”. In: *Nucleic Acids Research* 43.Database issue (2015), pp. D512–20. DOI: 10.1093/nar/gku1267 (cit. on p. 120).

[236] H. Mohammed et al. “Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes”. In: *Nature protocols* 11.2 (2016), pp. 316–326. DOI: 10.1038/nprot.2016.020 (cit. on p. 120).

[237] L. Schermelleh et al. “Trapped in action: direct visualization of DNA methyltransferase activity in living cells”. In: *Nature Methods* 2.10 (2005), pp. 751–756. DOI: 10.1038/nmeth794 (cit. on p. 120).

[238] B. C. Campbell et al. “mGreenLantern: a bright monomeric fluorescent protein with rapid expression and cell filling properties for neuronal imaging”. In: *PNAS* 117.48 (2020), pp. 30710–30721. DOI: 10.1073/pnas.2000942117 (cit. on p. 120).

[239] M. Hirano et al. “A highly photostable and bright green fluorescent protein”. In: *Nature Biotechnology* 40.7 (2022), pp. 1132–1142. DOI: 10.1038/s41587-022-01278-2 (cit. on p. 120).

[240] C. A. Müller et al. “Capturing the dynamics of genome replication on individual ultra-long nanopore sequence reads”. In: *Nature Methods* 16.5 (2019), pp. 429–436. DOI: 10.1038/s41592-019-0394-y (cit. on p. 121).

[241] NCBI. *NCBI Datasets*. Bethesda, MD, 2004 (cit. on p. 124).

[242] P. R. Mühlhausen. “Sequence conservation of DNA - Methyltransferase 1 acros genetic landscapes of placental and non-placental animals”. Berlin, 2021 (cit. on p. 124).

[243] F. Sievers et al. “Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega”. In: *Molecular Systems Biology* 7 (2011), p. 539. DOI: 10.1038/msb.2011.75 (cit. on p. 124).

[244] J. Huerta-Cepas et al. “ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data”. In: *Molecular Biology and Evolution* 33.6 (2016), pp. 1635–1638. DOI: 10.1093/molbev/msw046. (cit. on p. 124).

[245] J. A. Capra and M. Singh. “Predicting functionally important residues from sequence conservation”. In: *Bioinformatics* 23.15 (2007), pp. 1875–1882. DOI: 10.1093/bioinformatics/btm270 (cit. on p. 124).

[246] S. Basu et al. “Unblending of Transcriptional Condensates in Human Repeat Expansion Disease”. In: *Cell* 181.5 (2020), 1062–1079.e30. DOI: 10.1016/j.cell.2020.04.018 (cit. on p. 124).

[247] A. Kechin et al. “cutPrimers: A New Tool for Accurate Cutting of Primers from Reads of Targeted Next Generation Sequencing”. In: *Journal of Computational Biology* 24.11 (2017), pp. 1138–1143. DOI: 10.1089/cmb.2017.0096 (cit. on p. 153).

[248] Y. Xi and W. Li. “BSMAP: whole genome bisulfite sequence MAPping program”. In: *BMC Bioinformatics* 10 (2009), p. 232. DOI: 10.1186/1471-2105-10-232 (cit. on p. 153).

[249] D. Sun et al. “MOABS: model based analysis of bisulfite sequencing data”. In: *Genome Biology* 15.2 (2014), R38. DOI: 10.1186/gb-2014-15-2-r38 (cit. on p. 153).

[250] S. Gigante et al. “Using long-read sequencing to detect imprinted DNA methylation”. In: *Nucleic Acids Research* 47.8 (2019), e46. DOI: 10.1093/nar/gkz107 (cit. on p. 153).

A

Tissue culture SOPs

A.1 Thawing cells

1. Prepare 5 mL of PBS (Gibco 10010023) in a 15 mL centrifuge tube.
2. Remove a frozen stock vial of cells from liquid nitrogen and keep on dry ice.
3. Thaw cells in a 37 °C water bath. Once thawed, remove vial without delay from water bath and resuspend the cells in the prepared PBS.
4. To remove the cryoprotectant from the medium, centrifuge cells at 1000 rpm for 4 min.
5. Resuspend the pellet in the appropriate medium.

A.2 Seeding Feeder Cells

1. For mESC culture, coat a 6-well or 12-well plate with 0.2 % gelatin (Sigma G1890-100G; autoclaved and sterile filtered) and incubate at 37 °C for at least 10 min.
2. Resuspend 2.5 million inactive MEFs (CD1) in 12 mL mESC medium.
3. Aspirate the gelatin solution from the plate. Add 2 mL cell suspension to each well.
4. Shake plate vigorously in order to evenly distribute the cells.
5. Place the cells in the incubator.

6. Wait 1 d for the cells to settle and spread.

A.3 Culture and Passage of mESCs

1. Thaw a frozen stock vial of mESC cells.
2. Plate cells onto feeder cells in a 6-well plate at a density of about 50000 cells/well in mESC medium. Shake the plate vigorously to evenly distribute the cells.
3. Keep the cells at 37 °C and 5 % CO_2.
4. The medium is changed every day. Cells are split when they have reached about 80 % confluency (every 2 d to 3 d).
5. To split the cells, aspirate the medium and wash the cells once with PBS (Gibco 10010023).
6. Aspirate PBS and cover cells with TrypLE Express enzyme (Gibco 12604021) (about 1/5 of the filling volume of the well). Incubate 37 °C for 5 min.
7. Add 1 mL of mESC medium to the cells and generate a single-cell suspension by pipetting up and down about 20 to 30 times.
8. Transfer the cell suspension into a 15 mL conical bottom tube containing 5 mL PBS. Pellet the cells by centrifugation at 1000 rpm for 4 min.
9. Aspirate the supernatant and resuspend the cells in mESC medium.
10. Replace the MEF medium on a 6-well plate of feeder cells with 2 mL mESC medium. Add 50000 mESCs in suspension to one well.
11. Shake the plate vigorously to evenly distribute the cells.

Tab. A.1.: Composition ofmESC medium. This medium is used to cultivate and passage mESC. When preparing the medium, the components must be passed through a 0.22 mm pore-sized PES membrane. Add 2-Mercaptoethanol when almost all other components have passed the filter.

Component	Volume
Non-essential amino acids (Thermo 11140-035)	5.0 mL
GlutaMAX (Thermo 35050-038)	5.0 mL
FBS (Pan Biotech, P30-2602)	75 mL
KnockOut DMEM (Gibco 10829018)	404.5 mL
Penicillin/Streptomycin (Gibco 15140148)	5.0 mL
2-Mercaptoethanol (Thermo 21985023)	500 µL
LIF (produced in-house)	50.0 µL

A.4 Freezing mESCs

1. Transfer the cells into a suspension following steps 5-9 in SOP A.3.
2. Aliquot a cell suspension volume equivalent to about 120000 cells into a cryogenic vial. Top up the volume with mESC medium to 250 µL. Add 250 µL 2X freezing medium.
3. Freeze cells without delay at –80 °C at –1 °C min^{-1}, *e.g.*, by using a polystyrene box.
4. Store temporarily at –80 °C or in liquid nitrogen for long-term.

Tab. A.2.: Composition of2X freezing medium. When preparing the medium, all components but the DMSO must be passed through a 0.22 mm pore-sized PES membrane.

Component	Volume
FBS (Pan Biotech, P30-2602)	20 mL
KnockOut DMEM (Gibco 10829018)	60 mL
DMSO (Sigma D2650-100ML)	20 mL

A.5 Feeder Cell Depletion

1. Transfer the cells into a suspension following steps 4-6 in Section A.3.
2. Dilute the cell suspension in 10 mL mESC medium in a 10 cm dish.
3. Place the dish in the incubator for 40 min.

4. Transfer the medium into a 15 mL conical bottom tube.

A.6 Culture of HEK293T cells

Plate HEK293T cells in KnockOut DMEM (Gibco 10829018) supplemented with 10 % FBS (Pan Biotech, P30-2602) and Penicillin/Streptomycin (Gibco 15140148). Cells are grown to confluency and split in a ratio of 1:4 to 1:5 using TrypLE Express Enzyme (Gibco 12604021).

A.7 Nucleofection

1. Bring the P3 Primary Cell Nucleofector Solution and Supplement 1 from the P3 Primary Cell 4D-Nucleofector X Kit (Lonza V4XP-3024) to RT.
2. Prepare of plasmid DNA for the transfection in a microcentrifuge tube.
3. Pellet 1 million cells in a 15 mL conical bottom tube by centrifugation at 1000 rpm for 4 min.
4. Combine 82 μL Nucleofector Solution and 18 μL Supplement 1 in a microcentrifuge tube.
5. Aspirate the supernatant and resuspend the cells in 1 mL PBS (Gibco 100100-23). Pellet again by centrifugation at 1000 rpm for 4 min.
6. Add the nucleofection solution to the plasmid DNA and mix by pipetting
7. Aspirate the supernatant from the cells and resuspend them in 100 μL DNA-nucleofection mix.
8. Transfer the cell suspension to a Nucleocuvette vessel.
9. Apply electrical current with in the 4D-NucleofectorTM X Unit (Lonza AAF-1003X).
10. Without delay, transfer the cell suspension using the provided pipettes into 4 mL mESC medium. Distribute the cell suspension over two wells of a 6-well plate with feeders.

11. Shake the plate vigorously to distribute the cells evenly.

A.8 Colony Screening

1. Coat two 96-well plates with 0.2 % gelatin (Sigma G1890-100G; autoclaved and sterile filtered) for at least 10 min.
2. Plate 2.5 million feeder cells on one of the plates according to SOP A.2.
3. Replace the gelatin solution on the other plate with 130 mL/well of mESC medium. This place will be used for DNA extraction.
4. The next day, replace the medium on the feeder plate with 170 mL/well mESC medium. This plate will be used to maintain the clones in culture during the screen.
5. Add 30 μL of TrypLE Express Enzyme (Gibco 12604021) in a 96-well plate with a conical bottom.
6. Pick a colony by scratching over it with a pipette tip and lift it off the plate in a volume of 5 μL. Put the colony in a well and pipette it up and down a few times.
7. Put the 96-well plate into the incubator for 5 min to 10 min.
8. Add 70 μL mESC medium to each well and pipette up and down around 20 times for dissociation of the colony.
9. Check under the light microscope whether cells are singled out.
10. Add 30 μL of cell suspension to the 96-well feeder plate and 70 μL of cell suspension to the gelatin-coated 96-well plate without feeders.
11. Maintain cells in culture for 3 d to 4 d.
12. Extract the DNA from the gelatin-only plate using SOP B.2

B

Molecular biology SOPs

B.1 Genomic DNA Extraction

The genomic DNA extraction is carried out with the PureLink Genomic DNA Mini Kit (Invitrogen K182002).

1. Pre-heat a heat block to 55 °C.
2. Resuspend pellet in 200 µL PBS.
3. Add 20 µL Proteinkase K and 20 µL RNase. Mix well by vortexing briefly and incubate at RT for 2 min.
4. Add 200 µL Genomic Lysis/Binding Buffer and mix well by vortexing until a homogenous solution is obtained (few seconds).
5. Incubate the lysate at 55 °C for 10 min.
6. Add 200 µL of 100 % ethanol. Mix well by vortexing for 5 s.
7. Transfer the lysate (640 µL) to the provided spin column. Centrifuge column at 10 000 g for 1 min.
8. Place the column in a new collection tube. Add 500 µL Wash Buffer 1.
9. Centrifuge at 10 000 g for 1 min.
10. Place the column in a new collection tube. Add 500 µL Wash Buffer 2.
11. Centrifuge at maximum speed for 3 min.
12. Place the column in a new microcentrifuge tube. Add 60 µL elution buffer to the column membrane. Incubate for 1 min at RT.

13. Centrifuge at the lowest speed for 1 min and then at maximum speed 1 min.

B.2 Genomic DNA Extraction (96-well format)

1. Grow the cells on the screening plate (96-well plate, without MEFs) until about 50 % to 80 % confluency (3 d to 4 d).
2. Carefully (*sic!*) wash the wells with PBS (Gibco 10010023) three times. Discard the buffer by carefully (*sic!*) inverting the plates over paper towels.
3. Add 50 µL DNA lysis buffer to each well.
4. Cover plate with parafilm, place in a zip-lock bag with some wet paper towels and incubate at 37 °C overnight.
5. Add 150 µL 5 M NaCl (Invitrogen AM9759) with 10 mL 100 % ice cold ethanol.
6. Add 100 µL NaCl/ethanol solution to each well.
7. Keep at room temperature for 2 h to 4 h. Avoid vibrations. The precipitating DNA is visible as colourless threads under the light microscope.
8. Wash the wells three times with 70 % ethanol. Discard the alcohol by carefully inverting the plate over paper towels.
9. Air dry the DNA for 15 min to 20 min. Keep the plate tilted.
10. Pre-warm an aliquot of Buffer EB to 65 °C.
11. Resuspend the DNA in 30 µL Buffer EB (Qiagen 19086), either by pipetting up and down or incubation overnight at 37 °C.

Tab. B.1.: Composition ofDNA lysis buffer. Add RNase and Proteinase K just prior to use.

Component	Volume
1 M Tris, pH 7.5 (Invitrogen 15567027)	25 µL
0.5 M EDTA, pH 8.0 (Invitrogen 15575020)	50 µL
5 M NaCl (Invitrogen AM9759)	5 µL
10 % SDS (Invitrogen 15553027)	125 µL
nuclease free H_2O (Invitrogen 10977049)	2070 µL
Proteinase K (Ambion AM2546; add fresh)	125 µL
RNase (Roche 11119915001; add fresh)	100 µL

B.3 Genotyping PCR (96-well format)

1. Amplify the required region from the genomic DNA extracted with SOP B.2 by PCR in 35 cycles.
2. Separate the PCR products on a 1 % TAE agarose gel containing SYBR Safe dye (Thermo S33102) at 120 V for 30 min. The addition of loading dye is not necessary, since the MangoMix already contains dye.

Tab. B.2.: Reagents for a genomic DNA PCR screen. This master mix is for screening one 96-well plate.

Reagent	Volume
10 μM forward primer	50 μL
10 μM reverse primer	50 μL
MangoMix (Bioline BIO-25034)	500 μL
nuclease free H_2O (Invitrogen 10977049)	200 μL
genomic DNA	2 μL

B.4 Standard High-Fidelity PCR

1. Combine the reagents in a 200 μL PCR tube.
2. Mix the reagents by gently flicking the tube and spin them down in a microfuge.
3. Run the standard PCR program in the thermocycler (Eppendorf Mastercycler Pro).
4. Add 10 μL loading dye (NEB B7024S) to the sample.
5. Separate the PCR products on a 1 % TAE agarose gel containing SYBR Safe dye (Thermo S33102) at 130 V for 45 min.

Tab. B.3.: Reagents for a standard high-fidelity PCR. When setting up several reactions, it is recommended to prepare a master mix containing all reagents but the template DNA and water.

Reagent	Volume
10 μM forward primer	2.5 μL
10 μM reverse primer	2.5 μL
template DNA	50 ng
Q5 Master Mix (NEB M0492S)	25 μL
nuclease free H_2O (Invitrogen 10977049)	to 50 μL

Tab. B.4.: Thermocycler standard PCR program. Determine the correct annealing temperature using the Thermo Fisher Tm Calculator at https://bit.ly/3M0QIjL.

Step	Temperature	Duration		
Initial Denaturation	98 °C	60 s		
Denaturation	98 °C	20 s	}	28 cycles
Annealing	variable	20 s		
Elongation	72 °C	15 s kb^{-1}		
Final Elongation	72 °C	120 s		
Storage	12 °C	∞		

B.5 Restriction Digest and Dephosphorylation

1. Combine the reagents in a 200 μL PCR tube.
2. Run the following program in the thermocycler. After the digestion step, pause the program and add 0.65 μL of phosphatase and 2.85 μL of phosphatase buffer (NEB M0289L). Continue the program.
3. Pool samples if applicable. Add 5 μL loading dye (NEB B7024S) to the sample.
4. Separate the fragments on a 0.7 % to 1.0 % TAE agarose gel containing SYBR Safe dye (Thermo S33102), using the narrow wells, at 130 V for 45 min.
5. Extract the DNA according to SOP B.6.

Tab. B.5.: Reagents for a restriction digest. When setting up several reactions, it is recommended to prepare a master mix containing all reagents but the template DNA and water. The choice of buffer and incubation temperature depends on the restriction enzyme used.

Reagent	Volume
10X buffer (*e.g.*, rCutSmart, NEB B600)	2.5 µL
Plasmid/fragment	3.0 µg
Restriction enzyme	1.0 µL
nuclease free H_2O (Invitrogen 10977049)	to 25 µL

Tab. B.6.: Thermocycler restriction digest and dephosphorylation program. The temperature used for digestion depends on the restriction enzyme used.

Step	Temperature	Duration
Digestion	37 °C	2 h
Addition of phosphatase	–	–
Dephosphorylation	37 °C	30 min
Heat inactivation	65 °C	20 min
Storage	12 °C	∞

B.6 Gel Extraction

Use the Qiagen gel extraction kit (Qiagen 28706) according to the manufacturer's recommendations with the following modifications. After the addition of QG buffer, place the samples on a multiholder vortexer and shake for 5 min at RT. To elute, use 20 µL of EB buffer pre-warmed to 65 °C. After incubation, centrifuge the column placed in the elution tube at 1 rfc to 100 rfc for 1 min. Then continue to the final elution step.

B.7 Gibson Assembly

1. Design primers overlapping for 25 bp with each the 3' and 5' end of the fragments to be joined.
2. Amplify the fragments with the primers for extending the 3' and 5' ends according to the Standard PCR protocol using Phusion polymerase (see SOP B.4) with an annealing temperature of 68 °C.

3. Purify the PCR products on a TAE agarose gel and extract the bands corresponding to the right size using SOP B.6.

4. Set up the assembly reaction on ice.
   The following formula can be used to convert the plasmid measurement from μg to pmol:

$$\text{DNA}(\mu\text{g}) \times \frac{\text{pmol}}{660\,\text{pmol}} \times \frac{10^6\,\text{pmol}}{1\,\mu\text{g}} \times \frac{1}{\text{DNA(bp)}} = \text{DNA(pmol)}$$

   A calculator is online at https://nebiocalculator.neb.com/#!/dsdnaamt.

5. Incubate the reaction at 50 °C for 30 min.

6. Use 5 μL for the transformation of 50 μL competent cells (*E. coli* 10β or DH5α) according to SOP B.10.

Tab. B.7.: Reagents for a Gibson assembly. For the assembly of up to three parts, provide each insert at a molar ratio of 1:5 with respect to the backbone.

Reagent	Volume
Linearized plasmid	50 ng
Insert(s)	1:5 ratio
NEBuilder HiFi DNA Assembly Master Mix (NEB E2621L)	10 μL
nuclease free H_2O	to 20 μL

B.8 Gibson Assembly for sgRNAs

1. Order the gRNA as a 60 bp primer containing 20 bp of choice flanked by 20 bp long overhands which match with the plasmid after cutting it with BbsI:
   TGTGGAAAGGACGAAAACACC **20 bp target** GTTTTAGAGCTAGAAATAGC
   If the guide sequence does not start with a G, add a G to its 5' end.

2. Cut the pSpCas9 plasmid (*e.g.*, pX458, Addgene #48138) with BbsI-HF (NEB R3539S) according to the Restriction Digest and Dephosphorylation protocol (SOP B.5), purify the product on a 0.70 % agarose gel and carry out a gel extraction (SOP B.6).

3. Dilute the 100 μM guide primer 1:70 (to a final concentration of 1.42 μM).

4. Set up the reaction on ice.

The following formula can be used to convert the plasmid measurement from μg to pmol:

$$\text{DNA}(\mu g) \times \frac{\text{pmol}}{660\,\text{pmol}} \times \frac{10^6\,\text{pmol}}{1\,\mu g} \times \frac{1}{\text{DNA(bp)}} = \text{DNA(pmol)}$$

5. Incubate the reaction at 50 °C for 30 min.
6. Continue with SOP B.10. Plate all of the culture on one plate containing ampicillin.

Tab. B.8.: Reagents for a Gibson assembly of an sgRNA. If the plasmid is not concentrated enough, increase the total volume of the reaction to 10 μL.

Reagent	Volume
1.42 μM sgRNA primer	1 μL
BbsI-digested plasmid	0.05 pmol
2X Gibson Master Mix	2.5 μL
nuclease free H_2O (Invitrogen 10977049)	to 5 μL

B.9 pJET Blunt-End Cloning

This pJET blunt-end cloning protocol uses the CloneJET PCR Cloning Kit (Thermo Scientific K1231).

1. Set up the ligation reaction on ice.
2. Vortex briefly and centrifuge for 3 s to 5 s.
3. Incubate the ligation mixture at room temperature for 5 min. For PCR products >3 kb, ligation can be prolonged to 30 min.
4. Use the ligation mixture directly for transformation or store at –20 °C. Thaw the mixture carefully on ice before using it.
5. Transform 5 μL *E. coli* 10β cells with 2 μL of the ligation product. Follow the Heat Shock Transformation protocol (see Section B.10). After the heat shock, add 100 μL SOC medium to each sample. Plate 50 μL of the sample of one half of an LB agar plate.

Tab. B.9.: Reagents for pJET blunt-end cloning.

Reagent	Volume
2X Reaction Buffer	10 µL
Non-purified PCR product	1 µL
or	
purified PCR product/other blunt-end DNA fragment	0.15 pmol ends
pJET1.2/blunt Cloning Vector (50 ng µL^{-1})	1 µL (0.05 pmol ends)
nuclease free H_2O	to 19 µL
T4 DNA Ligase	1 µL

B.10 Heat Shock Transformation

1. Retrieve competent *E. coli* cells from –80 °C storage. Thaw the cells on ice. Use DH5α or 10β cells for regular vectors, and stbl strains for viral vectors, such as lentiviral vectors. Use BL21(DE3) (NEB C2527H) for protein expression.
2. In a microfuge tube, combine 50 µL cell suspension with 5 µL DNA (*e.g.*, from ligation reaction).
3. Incubate on ice for 15 min.
4. Heat shock at 42 °C for 30 s.
5. Incubate on ice for 2.5 min.
6. Add 250 µL SOC medium.
7. Incubate the reaction at 250 °C for 1 h in a thermomixer (Eppendor 53820000-15) 1000 rpm.
8. In the meantime, warm up an LB plate containing the appropriate antibiotic to RT.
9. Plate the cell suspension onto the LB plate using sterile a hockey stick or glass beads to distribute the cell suspension evenly on the plate. For efficient cloning protocols and re-transformations, plate 50 µL to 100 µL only.
10. Incubate the plate top down at 37 °C overnight.

B.11 Midi Prep

Use the Qiagen Plasmid Plus Midi Kit (Qiagen 12945) according to the manufacturer's instructions with the following modifications to extract DNA from a 50 mL culture. Use 4 mL of P1, P2 and S3 buffer. After the addition of S3 buffer, centrifuge the solution at RT and 4000 rpm for 10 min. Decant the supernatant into the filter syringe supplied with the kit and squeeze the liquid into a new centrifuge tube. Add 2 mL of BB buffer and mix by inverting the tube. Proceed according to the supplied protocol until the elution step. Elute with 50 mL to 100 mL of EB pre-warmed to 65 °C. After 1 min of incubation, first centrifuge the tube at 1 rfc to 100 rfc for 1 min. Then continue to the final elution step.

B.12 RNA Extraction

This protocol uses the RNeasy Mini Kit (Qiagen 74106).

1. Suspend the cell pellet consisting of $< 5 \times 10^6$ cells in 350 μL RT buffer. Vortex thouroughly.
2. Add 350 μL fresh 70 % ethanol. Homogenize by pipetting.
3. Immediately 700 μL of the lysate, including any precipitate, to an RNeasy Mini spin column.
4. Close the lid of the column and centrifuge for 20 s using the centrifuge short program at maximum speed.
5. Place the column in a new collection tube. Add 350 μL RW1 buffer to the column, close lid, and centrifuge using the short program at maximum speed.
6. Place the column in a new collection tube.
7. Add 70 μL RDD buffer to a new microcentrifuge tube. Add 10 μL DNase I. Mix gently by inverting the tube. Spin down in a minifuge.
8. Add DNaseI mix to the membrane of the column. Incubate at RT for 15 min.
9. Add 350 μL RW1 buffer to the column. Close lid and centrifuge for 20 s using the centrifuge short program at maximum speed.

10. Place the column in a new collection tube. Add 500 µL RPE buffer to the column. Close lid and centrifuge for 20 s using the centrifuge short program at maximum speed.
11. Place the column in a new collection tube. Add 500 µL RPE buffer to the column. Close lid and centrifuge for 2 min at maximum speed.
12. Place the column in a new collection tube. Centrifuge at maximum speed for 1 min to dry the membrane.
13. Place the column in a clean, RNase-free microcentrifuge tube. Add 30 µL DNase-, RNase-free water to the membrane. Centrifuge at minimum speed for 1 min, then at maximum speed for 1 min.
14. Assess quantity and quality of the RNA using the Nanodrop spectrophotometer.
15. Keep eluate on ice. Store at –80 °C.

B.13 cDNA Synthesis

This protocol uses the RevertAid First Strand cDNA Synthesis Kit (Thermo K1622).

1. On ice, aliquot two times 2 µg of RNA in two PCR tubes, one as a sample and one for the RT minus control. When converting several samples, only one RT minus control is required.
2. Add 1 µL random hexamer primer. Then raise the volume of the reaction to 12 µL with nuclease-free water.
3. Incubate the reaction at 65 °C for 5 min. Cool the reaction on ice.
4. Add the following reagents in the given order to each reaction. For the RT minus control, replace the enzyme by 1 µL of nuclease-free H_2O (Invitrogen 10977049).
5. Mix the reagents by gently flicking the tube and spin them down in a microfuge.
6. Run the following program on the thermocycler.

7. Store at 4 °C for short-term storage and at –20 °C for long-term storage.

Tab. B.10.: Reagents for cDNA synthesis.

Reagent	Volume
RNA	2 µg
Random hexamer primer	1 µL
nuclease-free H_2O (Invitrogen 10977049)	up to 12 µL
5X reaction buffer	4.0 µL
RiboLock RNase Inhibitor	1.0 µL
10 mM dNTP Mix	2.0 µL
RevertAid RT	1.0 µL

Tab. B.11.: Thermocycler cDNA synthesis program.

Step	Temperature	Duration
Primer annealing	65 °C	5 min
Cooling	4 °C	∞
DNA polymerization	25 °C	5 min
	42 °C	60 min
Termination	70 °C	5 min
Storage	4 °C	∞

B.14 qPCR

Each reaction, including the housekeeping gene control, has to be carried out at least in three technical replicates. Quadruplicates are preferred so that at least one outlier can be removed. The cDNA concentration is considered the same as the input RNA concentration for the cDNA synthesis. Use the following qPCR primers for the housekeeping gene control: ALM_GAPDH_ex5_F2 (ccttccgtgttcctacccc) and ALM_GAPDH_ex6_R2 (atgcctgcttcaccaccttc).

1. Dilute the cDNA from SOP B.13 to 6.25 ng µL^{-1}.
2. Add 4 µL of the diluted cDNA to the respective wells of an optical 96-well plate.
3. Prepare the target master mix (one for each target gene, and one for the housekeeping gene control) on ice. Cover the tubes when possible because the PowerUp SYBR Green MasterMix is light sensitive.

4. Add 6 μL of the targert master mix to the respective wells.
5. Make sure there are no air bubbles in each well.
6. Cover the plate with qPCR film and seal the plate thouroughly.
7. Spin down the plate at 2000 rpm for 2 min.
8. Keep the plate on ice and covered, *e.g.*, with an ice bucket lid or aluminum foil.
9. Open the StepOne program. Click on the "Run experiment" panel. Load a pre-existing template and double check the settings (10 μL reaction volumne, SYBR Green, add melting curves at the end of 40 cycles). De-select empty wells if applicable. Save the file as ".eds"
10. Open the machine's drawer and insert the plate. Close drawer.
11. Start the qPCR run through the software (not on the machine's screen).
12. After the run, export the data as `.xls` file.
13. Calculate the mean and standard deviation for all sample triplicates/quadruplicates. The standard deviation should be less than 0.2.
14. Calculate the difference between the mean cycle number $\bar{C}_T$ of the target of interest and the reference target (housekeeping gene control) as well as the fold increase (FI).

$$\Delta C_T = \bar{C}_{T,\text{sample}} - \bar{C}_{T,\text{reference}}$$

$$FI = 2^{-\Delta C_T}$$

15. If a control sample, *e.g.*, wild-type or mock treatment, was included, normalize the FI to the control FI.

$$FI_{\text{norm}} = \frac{FI_{\text{sample}}}{FI_{\text{control}}}$$

Tab. B.12.: Reagents the qPCR target master mix.

Reagent	Amount per well
PowerUp SYBR Green MasterMix (Thermo A25777)	5 μg
10 μM qPCR F primer	0.5 μL
10 μM qPCR R primer	0.5 μL

B.15 Nuclear Protein Extraction

1. Pre-chill buffer A (Table B.13), buffer C (Table B.14) and RIPA buffer (Tables B.16) on ice.
2. To lyse the cells, suspend the cell pellet consisting of 2×10^6 cells in 500 µL ice cold buffer A and transfer the suspension into a 15 mL falcon tube. Add 3 mL buffer A und mix well by pipetting.
3. Spin at 1500 rpm at 4 °C for 5 min.
4. Aspirate the supernatant. The pellet contains the nuclei. Resuspend the pellet in 250 microL ice cold buffer C and transfer the suspension into a clean 1.5 µL microcentrifuge tube.
5. Centrifuge at 3000 g at 4 °C for 5 min.
6. Aspirate the supernatant. Resuspend the pellet in 55 µL RIPA buffer.
7. Incubate at 4 °C for 20 min on a vortexer equipped with a vertical multi-size tube holder.
8. Centrifuge at maximum speed at 4 °C. Transfer the supernatant into a clean 1.5 µL microcentrifuge tube.
9. Aliquot 2 µL for protein quantification.
10. Proceed with SDS PAGE or store the supernatant at –20 °C or –80 °C.

Tab. B.13.: Composition of buffer A. Reagent amount for 100 mL buffer. Add DTT and protease inhibitor only prior to use. Each sample requires 3.5 mL buffer A.

Reagent	Amount
0.5 M HEPES	50 mL
5 mM Mg_2Cl_2	0.5 mL
25 mM KCl	2.5 mL
0.05 mM EDTA	10 µL
10 % glycerol	10 mL
0.1 % IGEPAL	1 mL
nuclease-free H_2O	80 mL
100 mM DTT	1 mL
Protease inhibitor (cOmplete)	3 tablets

Tab. B.14.: Composition of buffer C. Reagent amount for 20 mL buffer. Add DTT and protease inhibitor only prior to use. To make 25X stock solution of protease inhibitor, dissolve one tablet in 2 mL H_2O. Each sample requires 250 µL buffer A.

Reagent	Amount
0.5 M HEPES	200 µL
3 mM Mg_2Cl_2	40 µL
100 mM KCl	2 mL
0.01 mM EDTA	4 µL
10 % glycerol	2 mL
nuclease-free H_2O	15.8 mL
100 mM DTT	30 µL
25X Protease inhibitor (cOmplete)	120 µL

Tab. B.15.: Composition of 2X RIPA stock solution. Reagent amount for 20 mL buffer.

Reagent	Amount
100 mM Tris-Cl pH 7.5	5 mL
300 mM NaCl	3 mL
2 % NP40 (Tergitol)	571 µL
1 % sodium deoxycholate	0.2 g
nuclease-free H_2O	11 mL
100 mM DTT	30 µL
25X Protease inhibitor (cOmplete)	120 µL

Tab. B.16.: Composition of RIPA buffer. Reagent amount for 2 mL buffer. Add DTT, protease inhibitor, DNase and RNase only prior to use.

Reagent	Amount
2X RIPA buffer	1 mL
nuclease-free H_2O	750 µL
100 mM DTT	20 µL
25X Protease inhibitor (cOmplete)	80 µL
RNase	100 µL
DNase	50 µL

B.16 SDS PAGE

1. Pre-heat the heat block with PCR tube adapter to 72 °C. If protein solution is frozen, thaw on ice.
2. Dilute 3 µg of protein in 37.5 µL RIPA buffer in a 0.2 mL PCR strip tube with flexible lid.

3. Add 12.5 µL 4X LDS sample buffer and 5.5 µL sample reducing agent.
4. Incubate at 72 °C for 10 min.
5. Remove the comb and the strip from the pre-cast gel.
6. Mount the gel in the gel tank.
7. Add 400 µL antioxidant to MOPS SDS running buffer (Table B.17) and fill the front chamber.
8. Fill the back chamber with MOPS SDS running buffer.
9. Rinse the gel's wells with the running buffer using a syringe.
10. Load 5 µL protein size marker into one well.
11. Load 50 µL of sample into each well.
12. Close the gel tank. Run the gel at 80 V for 10 min, then at 130 V for 1 h (up to 1.5 h).
13. Open the gel using the gel knife. Remove the pockets and bottom part of the gel. Carefully lift the gel with the gel knife at one corner. Using fingers, transfer the gel to a dish with H_2O.

Tab. B.17.: Composition of MOPS SDS running buffer. Reagent amount for 1 L buffer.

Reagent	Amount
NuPAGE MOPS SDS Running Buffer (20X)	50 mL
distilled H_2O	950 mL

B.17 Western Blot

1. Take the top two layers (mat and paper) off the stack.
2. Wet a filter paper with H_2O.
3. Place the gel on top of the membrane.
4. Using tweezers, place the wet filter paper on top of the gel.
5. Remove air bubbles by rolling over the stack with a roller, from the center outwars in all four directions.
6. Place paper and mat on top of the filter paper. Repeat air bubble removal.

7. Place the stack in the center of the iBlot (shackle to the top).
8. Place an adsorbant pad on top of the stack (shackle to the bottom left).
9. Close the iBlot apparatus.
10. Run the P0 program (20 V, 7 min) on the iBlot.
11. Remove the membrane from the stack.
12. Incubate the membrane in 0.5 % TBS-T (Table B.18) for at least 1 h.
13. Dilute the primary antibody in 5 % milk TBS-T. Left-over antibody solution can be stored at –20 °C.
14. Pour off the buffer and cut the membrane, so that loading control protein and protein of interest are separated.
15. Incubate each membrane piece with the respective antibody for at least 1 h or over night at 4 °C.
16. Pour off the antibody solution and wash the membrane with TBS-T for 10 min.
17. Repeat the washing step twice.
18. Dilute the secondary antibody in 5 % milk TBS-T.
19. Incubate the membrane with secondary antibody for at least 1 h or over night at 4 °C.

Tab. B.18.: Composition of TBS-T buffer. Reagent amount for 1 L buffer.

Reagent	Amount
20X TBS buffer	50 mL
TWEEN 20	10 mL
nuclease-free H_2O	940 μL

C

Protein isolation SOPs

C.1 Protein Overexpression

1. Inoculate 5 mL of MDG medium supplemented with ampicillin with a single colony BL21(DE3) carrying the fusion construct of interest.
2. Incubate at 37 °C and 200 rpm overnight.
3. Use the MDG pre-culture to inoculate 100 mL of ZYM-5052 medium supplemented with ampicillin.
4. Incubate at 37 °C and 200 rpm for 2 h, then 18 °C and 200 rpm for 48 h to 72 h, until the culture is pink.
5. Harvest the cells by centrifugation at 15 500 g, 30 min, 4 °C in 50 mL centrifuge tubes.
6. Discard the supernatant and freeze the tubes containing the pellets at –80 °C overnight.

Tab. C.1.: Composition of MDG medium. Dissolve all reagents in nuclease free H_2O (Invitrogen 10977049) and filter through a 0.22 mm pore-sized PES membrane.

Reagent	Final concentration
Na_2HPO_4	25 mM
KH_2PO_4	25 mM
NH_4Cl	50 mM
Na_2SO_4	5 mM
$MgSO_4$	2 mM
Trace Metal Mix	0.2x
Glucose	0.5 % [w/w]
Aspartic acid	0.25 % [w/w]

Tab. C.2.: Composition of ZYM-5052 medium. Dissolve all reagents in nuclease free H_2O (Invitrogen 10977049) and filter through a 0.22 mm pore-sized PES membrane.

Reagent	Final concentration
Na_2HPO_4	25 mM
KH_2PO_4	25 mM
NH_4Cl	50 mM
Na_2SO_4	5 mM
$MgSO_4$	2 mM
Trace Metal Mix	0.2x
Glucose	0.5 % [w/w]
Glycerol	0.05 % [w/w]
Glucose	0.5 % [w/w]
Lactose	0.2 % [w/w]
Y/Z broth	1x

C.2 Cell Lysis

1. Thaw the pellets from SOP C.1 for 20 min on ice.
2. Resuspend the pellets in 20 mL ice-cold lysis buffer supplemented with 1x protease inhibitor (pool all the pellets).
3. Lyse the cells by sonication (QSonica, Q700-110; equipped with the 1/16" microtip probe,4417) in 10 cycles (15 s on, 45 s off, amplitude 40 %).
4. Clear the lysate by centrifugation at 15 500 g, 30 min, 4 °C.
5. Transfer the supernatant to a new 50 mL centrifuge tube.

Tab. C.3.: Composition of lysis and wash buffer. Lysis and wash buffer is used for cell lysis and as wash buffer for the His-tag affinity chromatography. Dissolve all reagents in nuclease free H_2O (Invitrogen 10977049). For cell lysis, add 1 tablet of cOmplete protease inhibitor (Sigma 11873580001) to 50 mL of buffer. Degas overnight for use in chromatography.

Reagent	Final concentration
Tris-HCl pH 7.5	50 mM
NaCl	500 mM
Imidazole	10 mM
Triton-X-100 (optional for lysis)	0.1 %
Urea (optional for lysis)	1 M to 8 M

C.3 Protein Purification

1. Pre-equilibrate a cOmplete His-Tag purification column (Merck 6781543001) placed in the chromatography system with degassed wash buffer (Table C.3).
2. Load the supernatant onto the column and was with 15 column volumes of Buffer A.
3. Elute the bound fusion protein with 10 column volumes of degassed elution buffer.
4. Pool the fractions containing fluorescent signal in a 3000 MWCO Amicon Ultra centrifugal filters (Merck UFC803024).
5. Centrifuge at 7.500 g and 4 °C for 20 min.
6. Resuspend the protein bound to the filter in 4 mL storage buffer.
7. Centrifuge at 7.500 g and 4 °C for 20 min.
8. Resuspend the protein bound to the filter in 500 μL storage buffer.
9. Aliquot the purified protein into protein lo-bind microcentrifuge tubes at store at –80 °C.

Tab. C.4.: Composition of elution buffer. Dissolve all reagents in nuclease free H_2O (Invitrogen 10977049). Degas overnight for use in chromatography.

Reagent	Final concentration
Tris-HCl pH 7.5	50 mM
NaCl	500 mM
Imidazole	250 mM

Tab. C.5.: Composition of storage buffer. Dissolve all reagents in nuclease free H_2O (Invitrogen 10977049) and filter through a 0.22 mm pore-sized PES membrane.

Reagent	Final concentration
Tris-HCl pH 7.5	50 mM
NaCl	250 mM
DTT	1 mM
Glycerol	10 %

I declare to the Freie Universität Berlin that I have completed the submitted dissertation independently and without the use of sources and aids other than those indicated. The present thesis is free of plagiarism. I have marked as such all statements that are taken literally or in content from other writings. This dissertation has not been submitted in the same or similar form in any previous doctoral procedure.

I agree to have my thesis examined by a plagiarism examination software.

Berlin, 17.10.2022

Nina Bailly