universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Generation of inducible epigenome editing tools“

verfasst von / submitted by

Kalina Bozhkova BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2021 / Vienna, 2021

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 877
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Genetik und Entwicklungsbiologie
Betreut von / Supervisor:	Ao. Univ.-Prof. Dr. Christian Seiser

ABSTRACT

Studying the role of DNA methylation in early preimplantation development and reprogramming of cellular potency will unravel fundamental questions about early development and further help to advance novel strategies for regenerative medicine. Only recently targeted manipulation of DNA methylation has become available due to the discovery of the CRISPR/Cas9 system, and the utilization of a dCas9. In the course of this master thesis an inducible epigenome editing tool for the targeted manipulation of DNA methylation of selected genomic loci has been created, by utilizing a dCas9-system and a PB-transposon vector for generating stable transgenic cell lines. The potential of the tool to edit DNA methylation for a prolonged time has been tested. Additionally, multiplexed editing of several genomic loci in the same cell has been examined, by creating a single "CARGO"-vector harboring multiple small guide RNAs. As targets important key factors involved in the induction of "2C-like cells" (2CLCs) have been selected to study the effects of the newly created epigenome editing system on the 2CLC-population in mouse embryonic stem cells. The results generated during this master thesis revealed new insights into the design and efficiency of novel epigenome editing tools, which will greatly help to generate improved epigenome editing tools to study fundamental epigenetic questions of DNA methylation reprograming in development and disease.

Key words: DNA methylation, epigenome editing, 2C-like cells, cellular potency, CRISPR/dCas9

Die Erforschung der Rolle der DNA-Methylierung in der frühen Präimplantationsentwicklung und der Reprogrammierung von zellulärer Potenz wird dazu beitragen, grundlegende Fragen zur frühen Entwicklung zu beantworten und zu neuen Strategien für die regenerative Medizin beitragen. Erst seit kurzem ist eine gezielte Manipulation der DNA-Methylierung aufgrund der Entdeckung des CRISPR/Cas9-Systems und der Verwendung eines dCas9 möglich. Im Rahmen dieser Masterarbeit wurde mit Hilfe eines dCas9- und eines PB-Transposon-Vektors ein induzierbares Epigenom-Editing-Tool zur gezielten Manipulation von DNA Methylierung erstellt. Das Potenzial dieses neuartigen Tools zur Manipulierung der DNA-Methylierung wurde über einen längeren Zeitraum getestet. Darüber hinaus wurde die Möglichkeit zum „Multiplexed Editing" mehrerer genomischer Loci in derselben Zelle untersucht, indem ein einzelner Expressionsvektor erzeugt wurde, der mehrere gRNAs beherbergt. Als Targets für das Editing-Tool wurden wichtige Schlüsselfaktoren ausgewählt, die an der Induktion von „2C-like cells" (2CLCs) beteiligt sind. Schließlich wurden die Auswirkungen des neu geschaffenen Epigenom-Editing-Systems auf die 2CLC-Population analysiert. Die Ergebnisse dieser Masterarbeit ergaben neue Erkenntnisse über das Design und die Effizienz neuartiger Werkzeuge zur Epigenom-Editierung, die wesentlich dazu beitragen werden, effiziente neuartige Epigenome-Editing Tools herzustellen, welche dazu beitragen werden grundlegende epigenetische Fragen der DNA-Methylierungs-Reprogrammierung in Entwicklung und Krankheit zu verstehen.

Schlagwörter: DNA-Methylierung, Epigenom-Editierung, 2C-ähnliche Zellen, zelluläre Potenz, CRISPR/dCas9

ACKNOWLEDGMENTS

Above all I want to thank my supervisor Dr. Mark Wossidlo for the opportunity to conduct my master thesis in his laboratory and providing me with an interesting project to work on. I want to thank him for encouraging me to develop my skills and become a better student. He was always ready to answer any question and showed immense patience on my way to learning.

I would also like to express my gratitude to Prof. Seiser for agreeing to be my senior supervisor and Prof. Ivessa for serving as a link to University of Vienna. Without their help, this master's thesis would not be possible.

Furthermore, I would like to thank Dr. Julia Arand for her guidance and serving as an example of an excellent scientist. Likewise, I want to thank Kristeli Eleftheriou for helping me when it was needed both morally and in regard to any scientific question.

Finally, I am grateful for the welcoming and supporting environment provided by my colleagues from the Center of Anatomy and Cell Biology of the Medical University of Vienna.

TABLE OF CONTENTS

1 INTRODUCTION

1.1 EPIGENETICS AND CELLULAR POTENCY

The term epigenetics is generally accepted today as the study of heritable changes in gene function that do not change the DNA sequence (Dupont *et al*., 2009). All cells in an organism contain a similar genome, whereas different morphology and function are granted by their epigenome: the modifications 'sitting on top' (coming from the Greek syllable "epi", meaning on top) of the genome. Epigenetic modifications can be deposited on the DNA itself, such as DNA methylation, or posttranslationally modified histone tails, which together regulate chromatin accessibility and the function of transcription factors (Portela and Esteller, 2010). DNA methylation is among the most studied epigenetic modifications and is the focus of this master thesis.

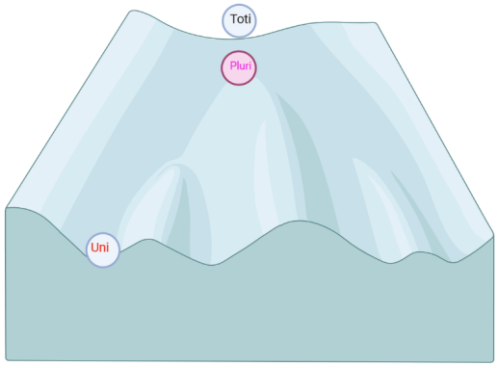


Figure 1: Waddington's epigenetic landscape. *Cell states are resembled as marbles rolling down a hill. On the top of the hill cells are totipotent, as they roll down to more differentiated states their ability to contribute to all tissues is lost. Created with BioRender.com*

In 1942 Conrad H. Waddington introduced the metaphor "epigenetic landscape" to describe developmental processes (Waddington *et al*., 1942). The epigenetic landscape consists of valleys and ridges, each valley represents a specific differentiation program. The cells, represented as marbles rolling down a hill, start with the full potential to roll in any direction / undergo any differentiation program. As they roll down, the variety of fates they can assume narrows down (Figure 1). On the top of the hill are totipotent cells, characterized with open chromatin and global hypomethylation. As they roll into different valleys, they become more specialized until they reach the final destination and become terminally differentiated, with methylated promotor areas, and more inaccessible chromatin.

1.2 DNA METHYLATION

DNA methylation is an epigenetic modification characterized by the covalent binding of a methyl group on the 5th carbon of a cytosine residue (5mC, Li & Zhang, 2014). In mammals DNA methylation is mainly observed at CpG dinucleotides (Ramsahoye *et al*., 2000). It is usually characterized as a repressive mark depending on the genomic context. Promoter areas of silenced genes are hypermethylated at CpG positions and demethylation of the area can result in derepression of gene expression (Eckersley-Maslin *et al.*, 2019). In recent years, a potential correlation between methylation levels and cellular potency has become more clear (Wossidlo *et al*., 2011; Eckersley-Maslin *et al*., 2018). Totipotent cells (e.g., the zygote) are hypomethylated, terminally differentiated cells are hypermethylated at a global scale.

1.2.1 DNMTs

Positioning and maintenance of DNA methylation is facilitated by the DNA methyltransferase (DNMTs) enzyme family (Smith *et al*., 2013). The maintenance of methylation is carried on by DNMT1, which recognizes hemi-methylated DNA upon replication; notably, a recent study also suggests a *de novo* methyltransferase activity of DNMT1 (Haggerty *et al*., 2021). DNMT3s are responsible for *de novo* DNA methylation and DNMT3a and DNMT3b are catalytically active and expressed in embryonic tissues. DNMT3L does not show catalytic activity and studies suggest that it interacts with DNMT3a and b stimulating their activity (Gowher *et al*., 2005, Siddique *et al*., 2013) and is important for the establishment of genomic imprints in germ cells (Hata *et al*., 2002).

1.2.2 TETs

There are two main pathways for a removal of DNA methylation – passive and active DNA demethylation. In passive DNA demethylation, 5mC is diluted out in each replication round due to an impairment/absence of the DNMT1 function. Active DNA demethylation is facilitated by the ten eleven translocation (TET) enzymes, which stepwise oxidize 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC), which can be recognized by Thymine DNA glycosylase (TGD) and removed by base excision repair. Each of these Tet enzyme mediated 5C-modifications can also be diluted out passively during replication (Wu and Zhang 2017, Figure 2).

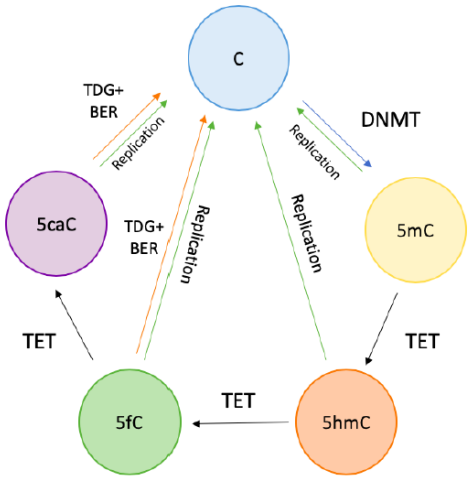


Figure 2: DNA methylation cycle in mammals. *Unmodified cytosine is methylated by DNMTs to 5mC (blue arrow). TET enzymes can oxidize 5mC to 5hmC and further 5fC, and 5caC (black arrows). 5caC is repaired by thymine DNA glycosylase (TDG) and together with base excision repair (BER) unmodified cytosine is restored (orange arrows). At any stage passive demethylation by non-maintenance during replication can lead to an unmodified cytosine again (green arrows). Image adapted from Wu and Zhang 2017.*

1.3 DNA METHYLATION IN EARLY DEVELOPMENT

In mammals life begins when a sperm fertilizes an oocyte creating a zygote with two pronuclei. The zygote divides into a two-cell embryo, four-cell embryo and further cleavage stages forming the morula, further to form the blastocyst with inner cell mass (ICM) and trophectoderm (TE), from which embryonic stem cells (ESCs) can be derived (Niakan *et al*. 2012). Maternal mRNAs and proteins derived from the oocyte help shortly after fertilization to transition from a state of no transcription to thousands of genes being transcribed at the same time in the two-cell embryo. This process, when the embryos own genome is activated for the first time, is termed embryonic genome activation (EGA) (Eckersley-Maslin *et al*., 2018). EGA occurs at different cleavage stages in different mammals. In mouse the first wave of EGA appears in the transition from S to G2 phase in the zygote (EGA) (Aoki *et al*., 1997) and the second wave of EGA begins as early as the two-cell embryo, where in human the second wave of EGA occurs at the eight-cell stage (Hamatani *et al*. 2004). These developmental transitions are accompanied with drastic changes of DNA methylation. The germ cells are hypermethylated, where the paternal genome in the sperm is more methylated than the maternal in the oocyte. A major wave of demethylation occurs from zygote to blastocyst stage. In the zygote the paternal pronucleus undergoes an active demethylation driven by TET enzymes, while the maternal pronucleus is protected by STELLA/Dppa3 (Wossidlo *et al*.,

2011) and undergoes a slower, passive loss of DNA methylation by exclusion of DNMT1 from the nucleus (Eckersley-Maslin *et al.*, 2018). Later in the blastocyst, DNA methylation is reestablished (Figure 3). Although, early development is accompanied by massive changes of the methylome, the exact role of DNA methylation reprogramming in establishing cellular potency in early development remains to be unraveled.

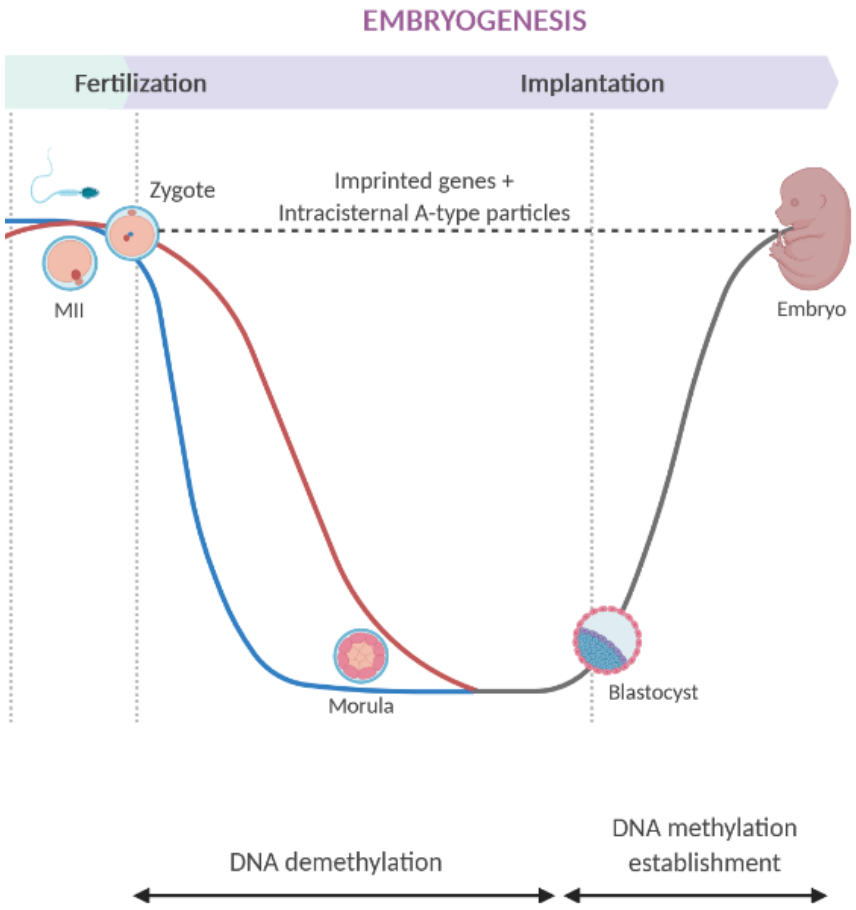


Figure 3: DNA methylation levels during mammalian development. *The red line represents the methylome of the oocyte and maternal genome, the blue line represents the methylome of the sperm and the paternal genome; they combine into the gray line representing the methylome of the embryo. The sperm and oocyte are hypermethylated, after the zygote is formed a wave of demethylation occurs, more rapid in the paternal pronucleus in comparison to the maternal pronucleus. Some specific genes, e.g. imprinted genes, and transposable elements (IAPs) remain methylated during the entire process (dashed line). After the formation of the blastocyst DNA methylation is reestablished. Created with BioRender.com.*

1.4 2C-LIKE STATE

Mouse embryonic stem cells (mESCs), which are derived from the ICM of the blastocyst, are self-renewing and pluripotent (Evans & Kaufman, 1981). Having the ability to give rise to every tissue

in the embryo have made them the first model system to study reprogramming (Nichols & Smith 2012). However, these cells lack the ability to contribute to extraembryonic lineages.

In 2012, Macfarlan *et al.* described a special rare population of cells within every mESC culture, which expresses genes characteristic for the 2-cell stage embryo, hence named 2C-like genes and 2C-like cells (2CLCs) (Figure 4). These cells are characterized by the upregulation of totipotency associated genes like Zscan4 and downregulation of pluripotency associated factors like Oct4. Macfarlan *et al.* created a stable cell line by genetically tagging a tdTomato-reporter linked to the MERVL retrotransposons, which are activated only at the 2-cell stage (Falco *et al.*, 2007). They showed that every cell within an ESC population cycles in and out of this state, suggestion an intrinsic memory of pluri- and totipotency (Macfarlan *et al.*, 2012).

In a 2019 paper, Eckersley-Maslin *et al.* searched for potential key regulators of the 2CLCs-population. By overexpressing candidate targets, they identified twelve factors able to increase the 2CLC-population (Eckersley-Maslyn *et al.* 2019, see Figure 21).

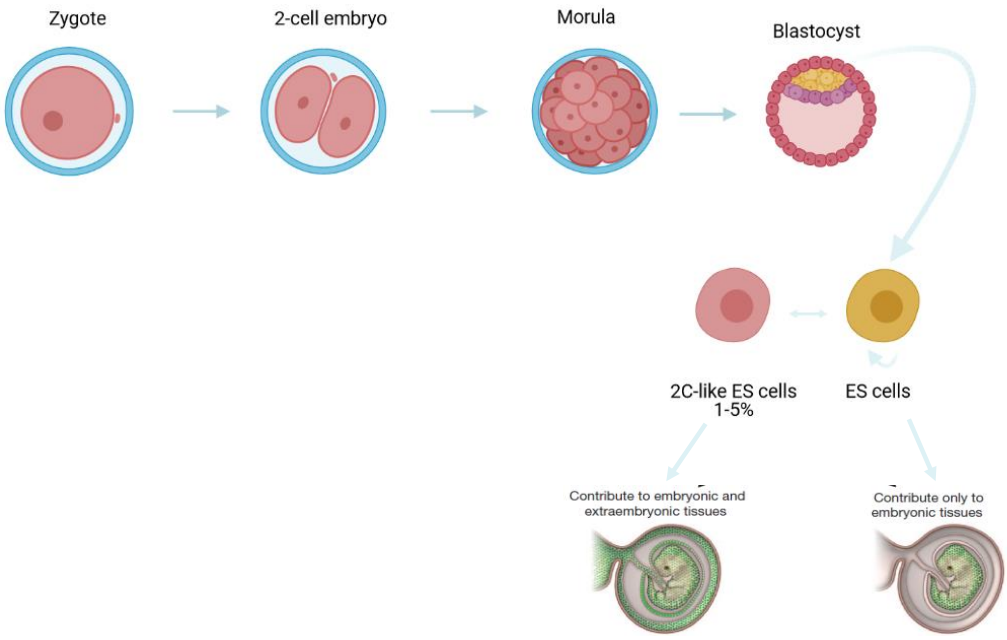


Figure 4: 2C-like cells - a transient state in mESCs that resembles totipotency. *During early embryonic development the zygote divides and forms the blastocyst, which is composed of the ICM and the TE. mESCs are derived from the ICM of the blastocyst, they are pluripotent and can contribute to all embryonic tissues. Within every mESCs culture a small population of cells fluctuate in and out of a totipotent state, resembling the 2-cell embryo, named thus 2C-like cells. 2C-like cells can contribute both to all embryonic tissues and extraembryonic annexes. Image adapted from Macfarlan et al. 2012, Created with BioRender.com.*

Before the discovery of 2CLCs the only model organism to totipotency was the embryo itself. However high throughput analyses on preimplantation embryos are not very feasible, due to the difficulty of obtaining large numbers of embryos. The study of 2CLCs aims to understand the early processes that form totipotent cells like EGA, epigenetic programing and differentiation.

1.5 EPIGENOME EDITING TOOLS

The first attempts to study global reprogramming of DNA methylation was achieved by either knockout or a knockdown of epigenetic modifiers or by the use of DNMT inhibitors such as 5-aza-2′-deoxycytidine (Christman *et al*., 2002). However, to understand dynamic and locus specific changes in DNA methylation levels, an approach for site-specific editing of specific CpGs was needed. A first method for specific locus targeting was using meganucleases to digest DNA at specific methylated/unmethylated genomic loci (Qomi *et al*., 2019). The methods further evolved to be able to recognize specific sequences in the genome and to design readers of genomic loci fused to epigenetic modifiers. First zinc finger nucleases (ZFN) were used, where zinc finger domains would recognize a specific DNA sequence and foc1 nuclease would facilitate the restriction. Each Zink finger can recognize 3 to 4 base pairs (bp) and multiple ZFNs could be paired to bind to up to 18 bp (Gupta *et al*., 2014) Among the limitations of this methods are the difficulty and lengthy procedure to design the tools and the low number of potential target regions and high off-target activities (Bozorg *et al*., 2019; Gupta *et al*., 2014). The next generation of editing tools were the transcription activator-like effector nucleases (TALENs) (Thakore *et al*., 2016; Bogdanove *et al*., 2011). They were easier to design and had no limitation in length. TALENs could be coupled to epigenome editing proteins to act as shuttles (Thakore *et al*., 2016), but the design still took several days and established tools were limited for each specific target sequence.

1.5.1 CRISPR

With the discovery of the bacterial adaptive immune system: clustered regularly interspaced short palindromic repeats or short “CRISPR”, began the era of next generation genome and epigenome editing tools. The most used type II CRISPR from *Streptococcus pyogenes* facilitates a Cas9 protein and a 20bp long guide RNA (gRNA) (Qi *et al*., 2013). Its advantage in comparison to ZFN and TALENs is that targeting of any part of the genome is possible with a 20bp protospacer in a sequence specific small gRNA (Gupta *et al*., 2014). The system has been further adapted by

mutations in the RuvC and HNH catalytic domains used to induce double strand breaks generating a catalytically inactive or 'dead' Cas9 (dCas9). The mutated enzyme recognizes the gRNA sequence without cleaving the DNA transforming the system into a shuttle for epigenome editing when fused to epigenetic modifiers (Qi *et al*., 2013; Wang *et al*., 2018).

Alongside its many advantages, the biggest issues of the CRISPR/dCas9 system are potential off-target effects. To address this problem, Tanenbaum *et al.* designed a system called SunTag which recruits multiple copies of the desired protein to the targeted area. They fused a green fluorescent protein (GFP) to single-chain variable fragment (scFv) antibodies designed to bind to a short peptide array. The peptide array recruits multiple copies of the antibodies, thus increasing the on-target and reducing the off-target effects (Tanenbaum *et al*., 2014). Combining the SunTag with CRISPR showed to decrease off-target effects (Pflueger *et al*., 2018). The system was further improved by Morita *et al*., 2016 by adapting the length and spaces between the GCN4 arrays, to recruit optimal number of copies of the catalytical domain (CD) of the TET enzyme Tet1 to its genomic target sequence (Figure 5).

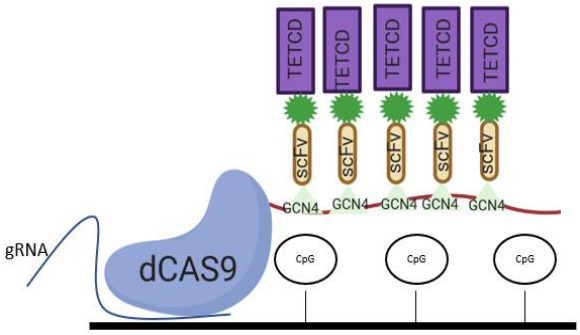


Figure 5: Schematic representation of the CRISPR-dCas9-SunTag-TET system. *The dCas9 is fused to a GCN4 array (green triangles), which recruits the scFV antibodies (yellow ovals), bound to a GFP tag (green stars and the TET1-CD (purple squares). The gRNA is represented as a blue line. The DNA is depicted as a black line with CpGs represented as lollipops. Created with BioRender.com.*

1.5.2 Inducible epigenome editing tools

Recently, an All-in-one-tool harboring the SunTag-dCas9 system, a DNA methylation modifier (DNMT or TET) and a gRNA altogether in a single vector has been established in the Wossidlo laboratory. The advantage of the "all-in-one" approach is the defined stochiometric delivery of all components needed for epigenome editing (EE) when transfected into cell culture. The All-in-one based expression vector, however, only allows transient expression of the system. An inducible EE-tool could enable studying the effects of EE over a prolonged time period.

There are several drug inducible systems that have been used in combination with a CRISPR/Cas9 system. On the posttranscriptional level, the chemically induced proximity (CIP) approach relies on two binding proteins to come close to each other in the presence of a drug activator. The components of the system are fused to the binding proteins, and thus editing is initiated once the drug has been administrated. Among the most frequently used CIP systems are those induced by rapamycin and the more recently plant derived hormones S-(+)-abscisic acid (ABA)-inducible ABI-PYL1 and gibberellin (GA)-inducible GAI-GID1 systems (Zhang *et al.*, 2019). Among the potential drawbacks of the CIP systems is the observed self-assembly in the absence of the induction-drug (Zhang *et al.,* 2019) and the possible side effects of rapamycin (Chan *et al.,* 1995). A concern associated with the ABA- and GA-inducible systems is the presence of the plant derived S-(+)-abscisic acid and gibberellin in the mouse diet, rendering them unsuitable for later use in mouse.

A popular drug-inducible system, induced at the transcription level is the Tet-On system. It was adapted from the *Escherichia coli* regulatory system initially used in bacteria to repress gene expression. In the modified Tet-On version binding of Tetracycline or its derivative Doxycycline (Dox) induces the expression of genes downstream of the promoter (Zhou *et al*., 2006). An improved version named Tet-On 3G system, was used in this master thesis. By utilizing the Cytomegalovirus (CMV) minimal promoter it possesses increased Dox sensitivity, minimized background expression and improved dynamic expression range (Zhang *et al*., 2019; Randolph *et al*., 2017). Furthermore, the Tet-On 3G system can be used in combination with a PiggyBac (PB) transposon system to create stable cell lines (Randolph *et al*., 2017). The PB system consists of two components: a transposon and a transposase. On the transposon large cargoes (up to 200kb) can be packed and the transposase facilitates its integration multiple times into the genome at ‘TTAA’ sites (Nakazawa *et al*., 2019). This master thesis aimed to generate an inducible EE-tool by cloning the components of the established All-in-one-tool, described above, under the Tet-On 3G promoter on a PB vector and transfecting generated constructs into mESCs.

1.5.3 Multiplex editing: the CARGO system

The Dox inducible system stably integrates the dCas9-TET tool into the genome, thus allowing prolonged and conditional EE. The system requires a specific gRNA to be introduced separately, which allows for flexibility in the editing of different genomic targets. Therefore, to further explore the potential of the inducible PB-dCas9-TET system, the simultaneous editing of multiple targets

(multiplexed EE) was attempted. In 2018, Gu *et al.* developed a method for the stochiometric delivery of multiple gRNAs into one vector termed "chimeric array of gRNA oligonucleotides" (CARGO). In the published CARGO system, up to eighteen gRNAs have been cloned in a single vector, each gRNA equipped with its own promotor and followed by a scaffold and termination sequence. The construction of the CARGO delivery system is based on the golden gate assembly method. It is a clever puzzle-like strategy, which utilizes the characteristic of the type IIS restriction enzymes to cut outside of their recognition motif, thus creating unique sticky ends and allowing for scarless cloning. In the CARGO system, multiple gRNAs can be packed into a single vector backbone, hypothetically enabling the dCas9 system to modify multiple genes in one delivery. In this master thesis the CARGO was used to test the potential of multiplexed EE.

1.6 REPROGRAMING OF SOMATIC CELLS

For the purpose to manipulate cellular potency *in vitro*, we questioned whether the PB-inducible EE-tool and the multitargeting CARGO system have the potential to induce reprogramming of somatic cells into totipotent 2CLCs. In 2018, Hernandez *et al*. showed that overexpression of chromatin-associated factors Dppa2 and Dppa4 is beneficial for reprograming mouse embryonic fibroblasts (MEFs) to induced pluripotent stem cells (iPSCs). Dppa2/4 are expressed in ESCs and the 2-cell stage mouse embryo and have not been detected in somatic cells. According to Hernandez *et al.,* knockdown of either Dppa2 or Dppa4 hinders the transition of MEFs into iPSCs, and, interestingly, overexpression of the two factors is able to boost the generation of iPSCs (Hernandez *et al*., 2018). The paper further states that combined overexpression of Dppa2/4 promotes faster transition by inducing a more permissive chromatin state.

In the Wossidlo laboratory, Marlene Musial, under the supervision of Dr. Julia Arand, created a heat map combining RNAseq data from Macfarlan *et al*., 2012 analyzing 2C-like stage genes with RNAseq data from Hernandez *et al*., 2018 of non-reprogrammed MEFs with and without overexpression of Dppa2/4. In Figure 6, Dppa2/4 overexpressing cells are depicted in green and the control group in black. Each column represents cells at a different day on a reprograming program and each row represents a different gene. The observed accumulation of overexpressed genes in the indicated red box suggests that overexpression of Dppa2/4 correlates with expression of 2C-like genes.

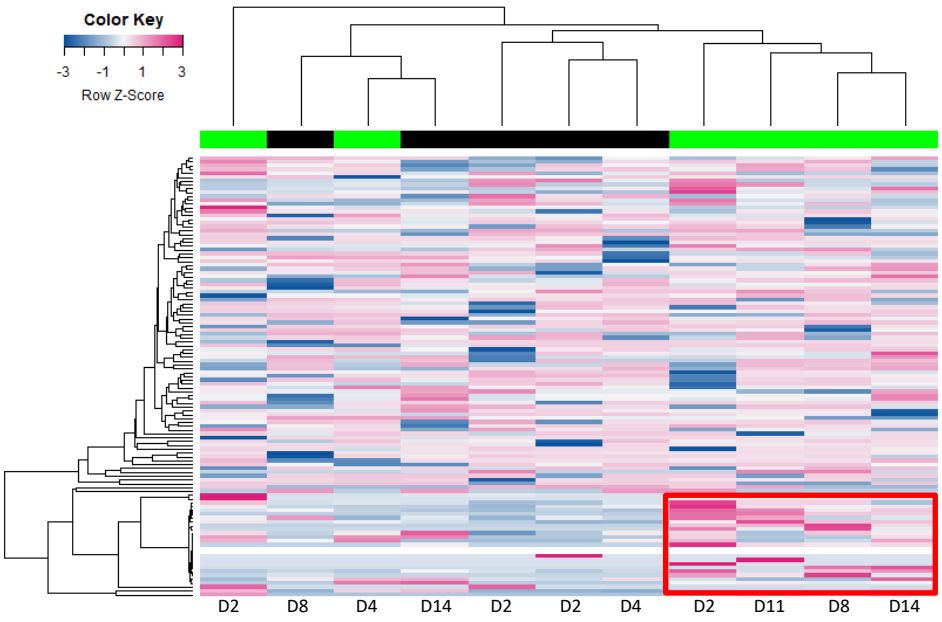


Figure 6: Heat map of RNAseq data for MEFs with and without overexpression of Dppa2/Dppa4. *The heat map was generated with R-studio combining RNAseq data from Macfarlan et al., 2012 and Hernandez et al., 2018. Each row represents a single gene expressed in mouse two-cell embryos. The columns show non-reprogramed cells on different days on a 14-day time course. In green are cells with overexpression of Dppa2/4, in black control cells without overexpression. The red box depicts a group of 2C-like genes, which are induced upon Dppa2/4 overexpression.*

2 AIM

After fertilization, as the newly derived one-cell embryo divides and further advances into development, the cells lose their totipotency and progresses to more specialized stages. This switch in cellular potency is marked by dramatic modifications of the epigenetic landscape. One prominent epigenetic mark that is drastically changing is in this process is DNA methylation. However, the role of DNA methylation in embryonic development and reprogramming of cellular potency is not fully understood.

The main goal of this master thesis was to create an inducible epigenome editing tool for targeted manipulation of DNA methylation. The tool needed to allow the manipulation of DNA methylation of key target sequences under controlled conditions over a prolonged time period. Thus, it would enable studying the impact of DNA methylation changes of specific loci on cellular potency.

First, stable cell lines should be generated, which are able to conditionally express the epigenome editing tool to manipulate key target sequences and test the stability of induced epimutations over time.

The next goal was to test the possibility of induced multiplexed epigenome editing of several genomic targets in the same cell. Here, the goal was to create a CARGO vector harboring several gRNAs and to test whether this system is able to manipulate the methylation levels of seven selected candidate 2C-genes simultaneously.

The final goal was to test whether editing of DNA methylation of key regulators of the 2C-like state can manipulate cellular potency of pluripotent ESCs and unipotent MEFs. As DNA methylation at the promoter area is associated with gene silencing, we questioned whether induced demethylation of the associated promotor area of 2C-like genes can induce the expression of the respective gene. Ultimately, we aimed to test demethylation of key regulatory genes could lead to changes in cellular potency.

3 MATERIALS AND METHODS

3.1 MATERIALS

3.1.1 Plasmids

3.1.1.1 All-in-one-TET-ng-tool

This plasmid has been established in the in the Wossidlo laboratory by Kristeli Eleftheriou and is derived from Morita *et al.,* 2018. The plasmid contains the catalytically inactive dCas9 fused with the SunTag system (linker length: 22 amino acids (aa)), the super-folded Green Fluorescent Protein (sfGFP), the CD of the human TET1 protein and the gRNA expression system. This plasmid was used as a template for the generation of the inducible EE-TET-tool.

3.1.1.2 All-in-one-DNMT-ng-tool

This plasmid has been generated in the Wossidlo laboratory by Kristeli Eleftheriou by replacing the TET CD of the All-in-one-TET-ng-tool with a DNMT3 CD. This plasmid was used for the generation of the inducible EE-DNMT-tool.

3.1.1.3 pB\tetO\Otx2\UbqC\Puro

This plasmid was a gift from Dr. Christa Bücker. It contained a PB expression cassette, an Otx2 gene body under a 3G-Tet-On promoter to allow for Dox inducible activation. An Ampicillin resistance cassette for bacterial selection and a Puromycin resistance cassette for selection in ESCs. After removal of the Otx2 gene, this plasmid was used for the generation of the inducible EE-tools.

3.1.1.4 03_09_scFv-GCN4-sfGFP-GB1-NLS_v2_3aR887E-3L

This plasmid was a gift from Prof. Dr. Albert Jeltsch. It contains a DNMT3A-DNMT3L fusion protein with a 3aR887E-3L mutation, for lower off-target effects, and a scFv-array with 10 scFv repeats. This plasmid was used to generate the inducible DNMT EE-tool.

3.1.1.5 02_77_dCas9-10XSunTag-BFP-short_10kb

This plasmid was a gift from Prof. Dr. Albert Jeltsch. It contains a dCas9 fused a GSN4-v4 array with 10 GSN4-v4 repeats with a 5 aa spacer. This plasmid was used to generate the inducible DNMT EE-tool.

3.1.1.6 pGEM-T vector

The pGEM-T vector, purchased from Promega, served as a subcloning vector for the generation of the constant region of the CARGO-system.

3.1.1.7 pX330-U6-Chimeric_BB-CB

This plasmid was a gift from Dr. Julia Arand. It harbors a gRNA expression system and a Cas9 cassette. The Cas9 cassette was replaced by a mCherry-reporter and the plasmid was used as a backbone for the CARGO system.

3.1.1.8 pCI-mCherry

This plasmid was a gift from the laboratory of Dr. Luc Snyers. It harbors a mCherry expression cassette. The plasmid was used to clone the mCherry tag into the CARGO system.

3.1.2 Oligonucleotide sequences

The oligonucleotide sequences used for this thesis were ordered from Microsynth (Table 1).

***Table 1: Oligonucleotide sequences used in this thesis**. Left column: the oligonucleotide names, right column: the sequence (5' > 3').*

KB-Seq-Pbin	GTCGAGTTTACTCCCTATCAG
KE-seq-gfp-r	GCCCATTAACATCACCATC
KB-Seq-dnmt	TAGTGGCGGTGGTGGTTCA

KB-Seq-F-cas9/ST	CCCCAAGTGAATATCGTGAA
KB-Seqp-F-cargo	CGATTTTTGTGATGCTCGTC
KB-Seqp-R-cargo	CAATGGAAAGTCCCTATTGGC
KB-Seqp-cargo-r2	CCGTAAGTTATGTAACGGGTAC
Kb-Pbinsert-F	TAAGCAACCGGTAGAATTCCCCCGCGAT
Kb-Tetng-R	ATCGATGAATTCGCGCGCCAAGATACAT
Pb-Mcs-F	CCGGTGGTACCGACTCGAATTCCGCCCGGGCTCGAGATCTACG CGT
Pb-Mcs-R	ACGCGTAGATCTCGAGCCCGGGCGGAATTCGAGTCGGTACCA CCGGTG
Kb-Gol-U6-F	TTTTTTCTAGAGGGCCTATTTCCCATGA
Kb-Gol-U6-R	TCGTCTCACGGTGTTTCGTCCTTT
Kb-Gol-Sc-F	TAAGCCGTCTCAGTTTTAGAGCTAGAAATAGC
Kb-gol-sc-t-R	AGGCCCTCTAGAAAAAAAGCACC
Kb-mch-r-gol	TCGATGAATTCCTTGTACAGCTCGTCCATGC
Kb-mch-f-gol	ATAACAACCGGTATGGTGAGCAAGGGCGA
Kb-Mut-Mch2-F	ATGCAGAAAAAGACCATGGGCTGGGAGGCCTCCTCCG
Kb-Mut-Mch2-R	ATGGTCTTTTTCTGCATTACGGGGCCGTCGGAGG
Kb-Cargo-Gfap_Cont-Oligo1	ACCGGGATGCCAGGAAAGTCTTCATATGAAGACAACATGTG
Kb-Cargo-Gfap_Cont-Oligo2	AAACCACATGTTGTCTTCATATGAAGACTTTCCTGGCATCC
Kb-Cargo-Gfap_Cont-Oligo3	ACCGTTTTTTTTCTAGAAGTCTTCATATGAAGACAAAGGATGT CAGCCC
Kb-Cargo-Gfap_Cont-Oligo4	AAACGGGCTGACATCCTTTGTCTTCATATGAAGACTTCTAGAA AAAAAA
Kb-Duxf3-Oligo1	CACCGATTGGCACCATGGCAGAAGC
Kb-Sp110-Bis-New-F	TCTTTCCCTACACGACGCTCTTCCGATCTTGTATAGTATTATTA AAGTGATTAGGAG

Kb-Sp110-Bis-R-New	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAAACTTTAAAAACAATATCCAATACC
Kb-Dppa2-Bis-F	TCTTTCCCTACACGACGCTCTTCCGATCTTAGGAAGATAGGAAGTAGTTATTTAT
Kb-Dppa2-Bis-R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAAAATATCAAACAAAAAAAATTCCAAA
Dppa4-Bis-F	TCTTTCCCTACACGACGCTCTTCCGATCTATATATAAAGCGTATTTTTGGTTAGGT
Dppa4-Bis-R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAAAAATAAATAAATCCTCCGAAAAAAAA
Kb-Trp63-Bis-F	TCTTTCCCTACACGACGCTCTTCCGATCTAAGTTTAGAGATGTTTATAAATTGTTAAT
Kb-Trp63-Bis-R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAATAAAATCAAATACCATCAAACATAA
Kb-Dppa2-2-Bis-F	TCTTTCCCTACACGACGCTCTTCCGATCTATCTGCTAGGGCATCTTCAGG
Kb-Dppa2-2-Bis-R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGATAAGGCTAACTTGGAACTCAC
Dppa4-F-Bs2	ATATATAAAGCGTATTTTTGGTTAGGT
Dppa4-R-Bs2	AAAAATAAATAAATCCTCCGAAAAAAAA
Dppa2-2-F-Bs2	TAGTGATTGAAGTTTGTGTTTTAGAT
Dppa2-2-R-Bs2	CAAATAAAACTAACTTAAAACTCACAAAA
Kb-Dppa4 Qpcr-F	TGAACCTGATTCACCGAGATGT
Kb-Qpcr-Dppa4 R	TGCTGCTCACTCGTTTCTTCTG
Kb-Trp63-Qpcr-F	AAAGAACGGCGATGCTTTCC
Kb-Trp63-Qpcr-R	GGCTCCACAAGCTCATTCCT
Ke-Gfap-Bi-F:	GTGAGGTATATGGTAAGGGTAGTTT
Ke-Gfap-Bi-R:	CAAACAACCAAACCTTATCTATAAA
Zscan4bi-Adap-F	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGAATAGGAGTTTTGATAGAATTAT

Zscan4bi-Adap-R	TCTTTCCCTACACGACGCTCTTCCGATCTCACATATACTTACACATTCAAA
Dux-Bi-Seq-F-Ill	TCTTTCCCTACACGACGCTCTTCCGATCTAGTAGGGTTTGTTTTTGTTAGGTTTG
Dux-Bi-Seq-R-Ill	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCCTAATACCCTAATATCAAATACCA
Gfap-Bi-Seq-F-Ill	TCTTTCCCTACACGACGCTCTTCCGATCTGTGAGGTATATGGTAAGGGTAGTTT
Gfap-Bi-Seq-R-Ill	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAAACAACCAAACCTTATCTATAAA
Kb-Duxf3-Oligo2	AAACGCTTCTGCCATGGTGCCAATC
Cargo-2c Oligo 1	ACCGATCCTAATAGAAAAGTCTTCATATGAAGACAACATGTG
Cargo-2c Oligo 2	AAACCACATGTTGTCTTCATATGAAGACTTTTCTATTAGGAT
Cargo-2c Oligo 3	ACCGAATAGTCACAGTCAAGTCTTCATATGAAGACAAAGAACCTCAGAT
Cargo-2c Oligo 4	AAACATCTGAGGTTCTTTGTCTTCATATGAAGACTTGACTGTGACTATT
Cargo-2c Oligo 5	ACCGCATCACAATTAAGTCTTCATATGAAGACAAAGTCCCTGATA
Cargo-2c Oligo 6	AAACTATCAGGGACTTTGTCTTCATATGAAGACTTAATTGTGATG
Cargo-2c Oligo 7	ACCGCGGCCTGGAGACTAAGTCTTCATATGAAGACAAAATTTTGTTGAGGG
Cargo-2c Oligo 8	AAACCCCTCAACAAAATTTTGTCTTCATATGAAGACTTAGTCTCCAGGCCG
Cargo-2c New Oligo 9	ACCGGTATTGGGATCAAAGTCTTCATATGAAGACAAGACTTTCAACG
Cargo-2c New Oligo 10	AAACCGTTGAAAGTCTTGTCTTCATATGAAGACTTTGATCCCAATAC

Cargo-2c New Oligo 11	ACCGGCTGGTAAGAATAAGTCTTCATATGAAGACAAATCAAAGCGAGC
Cargo-2c New Oligo 12	AAACGCTCGCTTTGATTTGTCTTCATATGAAGACTTATTCTTACCAGC
Cargo-2c New Oligo 13	ACCGATTGGCACCATGGAAGTCTTCATATGAAGACAAGAATTGACGGTT
Cargo-2c New Oligo 14	AAACAACCGTCAATTCTTGTCTTCATATGAAGACTTCCATGGTGCCAAT
Cargo-2c New Oligo 15	ACCGTTTTTTTTCTAGAAGTCTTCATATGAAGACAAATGGCAGAAGC
Cargo-2c New Oligo 16	AAACGCTTCTGCCATTTGTCTTCATATGAAGACTTCTAGAAAAAAAA
Kb-Scfv-Fsei-F	ATATAGGCCGGCCCATGGGCCCCGACATCGT
Kb-Gb1-Not1-R	ATTCGCGGCCGCTTACACCTTGCGCTTC
Kb-Suntag-Kfli-F	TATGGGACCCTAAGAAGTACGGCGGCTTCGACAGC
Kb-Suntag-Fsei-R	ATAGGCCGGCCAGGTCCAGGGTTCTCCTCCA
Kb-Pb-Dnmt-R-Ecori	AATGAATTCGCGCGCCAAGATACATTGATGAGTT
Kb-Mutate-Gcn-F	TGGCTCGTCTAAAGAAAGGTTC
Kb-Mutate-Gcn-R	CTTCGTTCTCAAGATGATAATTC
Kb-Pb-Dnmt-F-Ecori	TATATAAAGAATTCCCCCGCGATCGCGCCA
Kb-Pbdnmt-Fsei-F	AACCCTGGACCTGGCCGGCC

3.1.3 Buffers

Table 2:Buffers used in this thesis. *The left column contains the buffers names and the right buffer's composition.*

Buffer	Composition
P1 Mini Prep Buffer	50 mM Tris 10 mM EDTA RNase A (100 ng/ml)

P2 Mini Prep Buffer	0.2 M NaOH 0.2 M SDS
P3 Mini Prep Buffer	3 M Potassium Acetate
50X TAE Buffer	50 mM EDTA disodum salt 2 M Tris 1 M Acetic acid
DNA loading buffer	30% (v/v) glycerol 0.25% (w/v) bromophenol blue 0.25% (w/v) xylene cyanol FF
Fix-solution	3-parts Methanol 1-part Acetic acid

3.1.4 Chemicals

Table 3: Chemicals used in the thesis. *On the left: chemicals names, middle: abbreviation (if applicable), right column: distributor.*

Name	Abbreviation	Distributor
Acetic acid 100%		PanReac AppliChem
Agarose		LabQ
Ampicillin		PAA
Beta-mercaptoethanol		Sigma
Dimethyl sulfoxide	DMSO	Sigma
Doxycycline	Dox	PAA
Dulbecco's Phosphate-Buffered Saline	DPBS	Gibco
Ethidium Bromide	EtBr	Thermo Fischer Scientific
Ethylendiamintetracetate	EDTA	PanReac AppliChem
Ethanol 96%	EtOH	Merck
Fetal Bovine Serum	FBS	Sigma
Glycerol		Sigma
Kanamycin		PAA

L-Glutamine		Sigma
Lipofectamine 3000		Invitrogen
Methanol Isopropyl alcohol		VWR Loba Feinchemie
Magnesium chloride	MgCl2	Merck
Magnesium sulfate		Merck
Opti-MEM		Gibco
Penicillin/Streptomycin		Sigma
Polyacrylamide		Roth
Potassium chloride	KCl	Merck
Sodium chloride	NaCl	Merck
Sodium dodecyl sulfate	SDS	Sigma
Trypsin 0,25%		Gibco

3.2 BACTERIAL CULTURE

3.2.1 Bacterial strains and culturing

The *E. coli* strain DH5α chemically competent cells were used. The bacteria were cultured in LB-media supplemented with the corresponding antibiotic.

3.2.2 Transformation of bacterial cells with plasmid DNA

The *E. coli* strain DH5α was used for transformations. Cells were originally kept at -80°C. Before transformation bacteria were thawed on ice. 1µl DNA was added to the cells and mixed gently by flicking the tube. Then cells were placed on ice for 30 min. Next, the cells were heat-shocked at 42°C for 1 minute and placed back on ice. 200µl SOC media was added to the cells and they were incubated for 1h at 37°C.

Finally, transformed bacteria were plated on an agar plate containing the appropriate antibiotic and incubated at 37°C overnight. On the next day colonies were picked and placed in 2 ml LB medium containing the appropriate antibiotic and incubated again at 37°C overnight, shaking.

3.2.3 Plasmid DNA extraction - mini prep

The *E. coli* culture was transferred to 2 ml Eppendorf tubes and centrifuged for 10 min at 4000g. The supernatant was removed and 250 µl of buffer P1 (50 mM Tris, 10 mM EDTA, RNAse A (final concentration 100 ng/ ml) were added, the cells were vortexed until the pellet was resuspended. Next, 250 µl of P2 (0,2 M NaOH, 0,2 M SDS) were added, the tubes were inverted a few times and incubated for eight minutes at room temperature. Next, 350 µl P3 (3M potassium acetate) were added and the cells were incubated on ice for 10 min. In the meantime, new tubes with 700 µl isopropanol were prepared. The P3 Eppendorf tubes were centrifuged at 4°C at full speed for 5 min. 700 µl of the supernatant were added to the 700 µl isopropanol and incubated at -20°C for 30 min. The Eppendorf tubes were then centrifuged at full speed at 4°C for 15 min and the supernatant was removed. 500 µl EtOH were added and the Eppendorf tubes were centrifuged for 5 min full speed. The EtOH was completely removed and the pellet was dried for 5 min at 37°C. The pellet was eluted in 50 µl ddH2O and shaken at 300 rpm at 37°C for 30 min.

3.2.4 Plasmid DNA extraction-midi prep

The *E. coli* culture was transferred to 50 ml falcon and centrifuged for 10 min at 4000g. The supernatant was removed and the pellet was resuspended in 2ml of P1. Next, 2 ml of P2 (0,2 M NaOH, 0,2 M SDS) were added, the tubes were inverted a few times and incubated for 5 minutes at room temperature. Next, 2,5 ml P3 (3M potassium acetate) were added and the cells centrifuged at 4°C at full speed for 30 min. 7 ml of the supernatant were added to the 7ml ice-cold isopropanol and incubated at -20°C for 30 min. A 30 min centrifugation step followed and the supernatant was removed. 5 ml EtOH were added and the Eppendorf tubes were centrifuged for 15 min full speed. The EtOH was completely removed and the pellet was dried for 5 min at 37°C. The pellet was eluted in 200 µl ddH2O and shaken at 300rpm at 37°C for 30 min. Finally, the DNA was purified with magnetic beads (see section 3.2.6).

3.3 MOLECULAR CLONING

3.3.1 Cloning of the CARGO/Golden Gate assembly method

The CARGO system was assembled according to Gu *et al.*, 2018 and consists of three main components: constant region, variable region and a backbone. To create the constant region primers were designed to bind the All-in-one-tool to encompass the U6 promoter and a second pair for the

scaffold and termination sequence. The two PCRs were combined by a bridge-PCR with the forward primer for the scaffold and the reverse primer for the U6 promoter to ensure the desired order of sequences. The construct was then cloned into a pGEM-T vector (Promega) following the manufacturer's instructions.

The variable regions are 50 to 60 bp long oligonucleotides (ordered from Microsynth, see section 3.6.2), each comprising the second halve of a gRNA, two inverted to each other BsbI (type IIS restriction enzyme) recognition sequences and the first halve of the following gRNA.

For a backbone a pX330 vector was selected. The Cas9 insert removed by restriction digestion with the AgeI and EcoRI restriction enzymes and replaced by a mCherry selection marker via ligation. All BsbI recognition sites were mutation by a mutagenesis PCR (see section 3.2.3). The newly generated pX-mCherry plasmid was then digested with PsiI and XbaI restriction enzymes and the 4,5 kb band was extracted. Next, the constant region was freed from the pGEM-T vector using the BsmBI restriction enzyme and the 3,4 kb band was gel extracted. Then, minicircles were formed by ligating the constant region and the aligned and dephosphorylated oligos (following manufacturer's instructions). Each minicircle was treated by plasmid-safe Exonuclease V (NEB) following the manufacturer's instructions; the minicircles were then combined and cleaned with magnetic beads (see section 3.2.5). In the final reaction the minicircles and the backbone were combined together with a ligase and the BbsI restriction enzyme in a single reaction tube. The following program was selected on a thermocycler: (37°C for 5', 20°C for 5') x 50. The final CARGO vector was then transformed into DH5α competent bacteria.

3.3.2 Annealing and dephosphorylation of oligonucleotides

Oligonucleotides were resolved in recommended from the manufacturer volume of ddH_2O. 1µl of 100µM forward and reverse oligo were mixed with 1µl T4 ligase buffer, 0,5µl polynucleotide kinase and 6,5µl ddH_2O. The suspension was placed in a thermocycler under the following program:

1. 37 °C for 30 min
2. 98 °C for 5 min
3. ramp @-0,1 °C/s to 16°C
4. hold at 16 °C

3.3.3 Polymerase chain reaction (PCR)

PCRs were performed using the Q5 High-Fidelity Polymerase (NEB). The reaction mix and PCR program were chosen according to the manufacturer's instructions. For bisulfite converted samples the HotFirePool polymerase was used, following the manufacturer's protocol.

3.3.4 Mutagenesis PCR

Primer for site-directed mutagenesis were designed using NEBs BaseChanger with the forward primer comprising the nucleotide to be changed and the reverse primer starting at the beginning of the forward primer (Table 1). The annealing temperature suggested by the program was used.

3.3.5 Gel electrophoresis

To separate DNA fragments of various sizes gel electrophoresis was performed. The gel concentration varied depending on the fragment size from 1 to 2%. Agarose was dissolved in TAE buffer by boiling it in the microwave. Ethidium bromide was added to a final concentration of 0,5 µg/ml. To visualize the samples, loading dye was added in a 1 to 6 ratio. The separation was performed under 100 V for 30 minutes. Samples were then visualized under a UV-lamp.

3.3.6 PCR amplicon purification

To purify the samples MagBinding Beads (Zymo Research) were used. Magnetic beads were added to the PCR product in 1:1 ratio and incubated for 5 min at room temperature. The samples were then put on a magnetic stand and the supernatant was removed. Two washing steps with 200 µl 70% ethanol were performed. Next, the ethanol was removed, and the samples were dried on the magnetic stand at room temperature for 5 minutes. To elute the samples 20 µl ddH2O was used.

3.3.7 Gel extraction

To isolate genomic DNA from an agarose gel, the samples were visualized under UV light. The desired fragment was cut out with a razor blade and placed in an Eppendorf tube. The weight of the gel piece was measured, and gel extraction was performed with the Monarch® DNA Gel Extraction kit (NEB) following the manufacturer's protocol.

3.3.8 Addition of dA overhangs

For successful TA-cloning, dA overhangs were added at the 3' end of the amplicons using the OneTaq DNA Polymerase (NEB), following the protocol provided by the manufacturer. The samples were incubated for 20 minutes at 72°C

3.3.9 pGEM-T cloning

To increase the efficiency of the DNMT-tool cloning, the DNMT-fragment was first cloned in to a pGEM-t vector (Promega) following the manufacturer's protocol and incubated overnight at 16°C.

3.3.10 T4 ligation

To ligate two DNA fragments the T4 ligase (NEB) was used. The reaction was performed overnight at 16°C. The vector to insert ratio was 1:3. Manufacturer's protocol was followed.

3.3.11 Sequencing

For sequencing the Sanger-based service from Microsynth was used. Samples were diluted to the concentrations recommended by the manufacturer and sent for sequencing (Annex figure 1).

3.3.12 Restriction digestion

Restriction digests were performed with NEB enzymes following the manufacturer's protocol. 1 µg DNA was digested in a 20 µl reaction with the appropriate buffer. The reactions were incubated for 1 to 2 hours at 37°C. For 2-steps digestions the product of the first digestion was purified and incubated again with the second enzyme and the corresponding buffer. To stop the reaction the sample was incubated for the appropriate time and at the corresponding inactivation temperature as advised by the manufacturer.

3.4 CELL CULTURE

3.4.1 Culturing of mouse embryonic stem cells and fibroblasts

Mouse embryonic stem cells (mESCs) from the E14 cell line were cultured in a dedicated stem cell incubator at 37°C under 5% CO2 in Dulbecco's Modified Eagle's Medium (DMEM)-high glucose (D6546- 500ML, Sigma) supplemented with 16% Fetal Bovine Serum (FBS), 1% non-essential amino acids (M7145, Sigma), 1% penicillin-streptomycin (P4333, Sigma), 1% L-Glutamine

(G7513, Sigma), 1% sodium pyruvate (S8636, Sigma), 0.1mM β-mercaptoethanol (Sigma-Aldrich), and 1000 units/ml LIF. On a quarterly basis, mycoplasma tests were scheduled at the Vienna Biocenter as part of a regular service agreement.

Mouse embryonic fibroblasts (MEFs) were cultured at 37°C under 5% CO2 in DMEM-high glucose (D6546- 500ML, Sigma) supplemented with 10% FBS, 1% penicillin-streptomycin (P4333, Sigma), 1% L-Glutamine (G7513, Sigma).

3.4.2 Generation of PB-expressing cell lines

E14 mESCs were transfected with 0,75 μg PB-TET/ PB-DNMT and 0,25 μg transposase using Lipofectamine 3000 following the manufacturer's instructions (see section 3.4.4.1). 24h post transfection 1μg/ml puromycin was added to the media and the cells were selected on puromycin for eight days. Upon puromycin removal, 1μg/ml Dox was added to the media and GFP expressing cells were sorted 24h after using BioRad3 FACS sorter. The sorted GFP-positive cells were expanded on a 10 cm dish and single colonies were then picked on a 96-well plate on Dox. Homogenous colonies were expanded and kept for further use.

3.4.3 Karyotyping

For Karyotyping cells were grown in one well of 6 well-plate until they reached 70-80% confluency. One drop (+/- 100μl) of Colcemid (10μg/ml) was added to arrest the cells at the metaphase state and incubated for 3h at 37°C. Following centrifugation at 1000rpm x 8min, the supernatant was aspirated (1:1 Supernatant: Pellet ratio), and the pellet was resuspended in the remaining supernatant by flicking the tube or gently pipetting up and down. Next, 300 μl of 0.075M KCl were added, mixed, and another 500μl 0.075M KCl were mixed, until a volume of up to 4-6ml KCl was reached. Cells were incubated at RT x 15min. Next, 500 μl of Fix solution was added and the cells were spun down at 1000 rpm x 8min and the supernatant was aspirated (1:1 ratio). To the pellet a few 300 μl of fixative were added, resuspend, then filled up to 4ml with Fix and incubated at least 30min. at RT. Cells were spun down at 1000rpm x 8min.

To prepare the metaphase spreading a beaker with dH_2O, double frosted slides in methanol, Fix solution and Pasteur pipets were needed. The supernatant was aspirated until 1:1 ratio of supernatant to pellet was left in the tube, cells were then resuspended by flicking. A few drops of Fix were added, depending on the size of the cell pellet, until a light milky suspension was created. Cells were allowed to stand for 15min at RT. Glass slides were taken out from methanol and placed

in dH_2O for a few sec. To remove excess water from the glass slide, the edge of the slide were blotted with a paper towel.

Holding the glass slide in a 30-45° angel to the bench and from about 10 centimeters above the slide, 3 drops of cell suspension were placed next to each other, letting the drops rinse from the upper edge down toward the lower edge of the slide.

Next, the slide was flooded, drop wise, with Fix-solution letting it rinse off the slide from the upper edge toward the lower edge. The slide was then drained for a moment, then the lower edge blotted and the back of the slide dried with a paper towel. The slide was left to dry on a moist paper towel. Finally, the slide was baked at 90°C for 30 min.

3.4.4 Transfections

3.4.4.1 Lipofection of ESCs

24 hours prior to transfection 60 000 cells per well were seeded on a 24-well plate. Using Lipofectamine 3000 cells were transfected according to the manufacturer's protocols. Cells were sorted after 24 hours using fluorescence-activated cell sorting (FACS) (see section 3.4.6).

3.4.4.2 Electroporation of MEFs

MEFs were transfected using the Neon® Transfection System following the manufacturer's instructions. Briefly: MEFs media (10%FBS) with no antibiotics was warmed up at 37°C. 1 x $10^{\wedge 6}$ cells were trypsinized, washed twice with PBS and resuspended in 100 µl R buffer. 20 µg DNA was added to the suspension and the cells were electroporated in E2 at 1400V 10ms 3 pulses. The cells were then pipetted into the warm media and incubated at 37°C.

3.4.5 Flow cytometry

Cells were trypsinized and plated in 200 µl of media on a 96-well plate. To analyze the number of cells expressing a fluorescent protein the Beckman Coulter Life Sciences CytoFLEX benchtop flow cytometer was used. The CytExpert Software for the CytoFLEX Platform was used to analyze the data. The gates were set according to the non-transfected cells and the same gating was kept throughout the experiments.

3.4.6 Fluorescence-activated cell sorting (FACS) of mESCs

To select cells expressing a fluorescent protein FACS was performed with a S3e Cell Sorter (BioRad). Preceding sorting cells were trypsinized, centrifuged for 8 minutes at 1000g, passed

through 40 µM cell strainer and kept in 300 µl DPBS. While sorting (Annex figure 1) the cells were collected on 200 µl media and placed on ice until further analysis.

3.5 DNA METHYLATION ANALYSIS

3.5.1 Designing primers for bisulfite conversion

Primers for bisulfite conversion were designed using the online platforms Benchling and Oligo7. The primers were 25-28 nucleotides long with Tm between 48 and 56°C. The amplicon size was between 150 and 400 bp (Table 1).

3.5.2 Primers for next generation sequencing

Illumina adaptors were added to the bisulfite primers to allow for NGS of generated amplicons.

3.5.3 DNA extraction and bisulfite conversion

To collect cells for DNA extraction the media was removed and the cells were washed with PBS. Trypsin was added and the cells were incubated for 5' at 37°C. The Trypsin was deactivated by serum-containing media and the cells were spun down for 10' at 4000g. The supernatant was removed and the cells were frozen at -20°C for at least 10'. To lyse the collected cells 10µl 2x Digestion buffer (Zymo® Research), 1 µl Proteinase K and H2O up to 20 µl. The cells were incubated at 50°C for 20 min. In the meantime, the conversion mix was prepared: fresh 3 M NaOH; Bisulfite/NaOH: 58mg Na-metabisulfite, 76,35 µl ddH2O; 18,3 µl 3 M NaOH; Scavenger: 1,4 mg Trolox, 39,4 µl Dioxan. For the Sulfonation and deamination in 0,2 ml tubes 94 µl Bisulfite/NaOH, 37 µl Scavenger and 20 µl DNA were added. The reaction was incubated in a thermocycler under the following program:

99°C	15 min
50°C	30 min
99°C	5 min
50°C	90 min
99°C	5 min
50°C	90 min

For the washing, desulfonation and elution 600 µl Binding Buffer was added to 10 µl MagBinding beads (Zymo® Research) in a 1,5 ml Eppendorf tube. The samples were added to the

mixture, vortexed, and incubated for 5 min at RT. Next, they were moved to a magnetic stand and incubated another 5 min, the supernatant was removed. The samples were then washed with 400 µl Washing Buffer. Next, 200 µl Desulfonation Buffer was added and the samples were incubated for 15 min at RT. Two washing steps were performed next and the samples were eluted in 20 µl ddH2O.

3.5.4 Bisulfite PCR amplification

For amplification of the desired locus Hot FIREPol DNA Polymerase (SolisBiodyne) was used. As template 3 µl of the bisulfite converted DNA was used. Table shows the reaction mix:

Hot FIREPol	1.25U/50 µl
Buffer B1	2.5 µl
MgCl2	2 µM
Forward Primer	0.5 µl
Reverse Primer	0.5 µl
dNTPs	0.5µl
Template	3µl
ddH_2O	15.875µl

The following PCR program was used with 40 cycles:

95°C	15 min
95°C	1 min
56°C	1.5 min
72°C	1 min
72°C	5 min
12°C	hold

The Tm was selected depending on the primers according to the manufacturer's recommendations.

3.5.5 DNA Methylation analysis by Combined Bisulfite Restriction Analysis (COBRA)

Combined bisulfite restriction analysis (COBRA) is a method combining bisulfite conversion of DNA, polymerase chain reaction and restriction digestion. This method is used for quantitative measuring of DNA methylation levels of specific CpGs of a DNA sequence of interest (Eads and Laird, 2002). DNA methylation protects methylated cytosines from bisulfite conversion while unmethylated CpGs are converted to uracils. In the next step the uracils are replaced by thymines by PCR amplification. Next, restriction enzymes with CG in the recognition site as motif are used to analyze the methylation status of a locus of interest. Briefly, genomic DNA is bisulfite-converted (as described above) and the locus of interest is amplified with PCR. Upon purification, the amplicon is digested using restriction enzymes whose recognition site is affected by the methylation levels of the DNA. Through gel electrophoresis, the fragments of the digestion can be analyzed, through densitometric analysis and the methylation level can be determined (Figure 7).

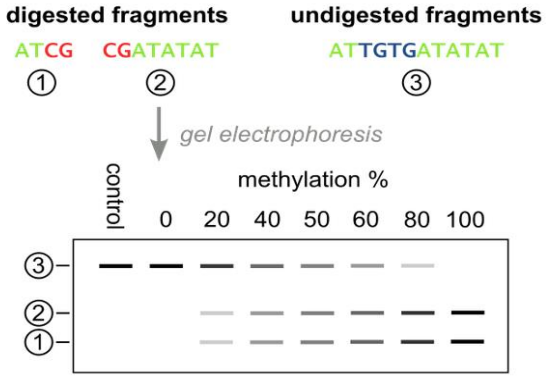


***Figure 7: Schematic quantification of COBRA**. DNA is digested by restriction enzymes and analyzed by gel electrophoresis. Densitometric analysis is used to determine the methylation status. Methylated fragments are digested while unmethylated remain undigested. DNA methylation levels are calculated as the percentage of intensity of the cleaved fragments to all bands (Image adapted from Wikipedia).*

3.5.6 Densitometric analysis using ImageJ

The PCRs of the bisulfite treated DNA were purified with magnetic beads and digested with an appropriate restriction enzyme. The digested PCRs were separated on an agarose gel and imaged for signal quantification. The signal intensity of digested bands was measured by the ImageJ gel analysis tool. Using the formula: (Σ digested bands/ Σ all bands)*100 methylation was calculated.

3.5.7 Methylation analysis by next generation sequencing (NGS)

DNA was bisulfite converted as described earlier and targeted regions were amplified with FusionPrimers including part of TruSeq Adaptors using HotFire Pol (SolisBioDyne). The amplicons were purified and their concentration measured using nanodrop. Amplicons were diluted to 5 nM, if applicable, equimolarly mixed and libraries were prepared by amplifying with TruSeq adaptors including barcodes on the i7 and i5 adaptor in five cycles using HotFire Pol. The amplification products with different barcode combinations were mixed and purified using "clean-up" beads and subsequently submitted for sequencing to the Vienna Bio Center sequencing facility to be sequenced in a 2x 125 bp run. Amplicons were separated by barcode using custom-made scripts, trimmed for adaptor sequences using cutadapt and sorted by primer sequences for the different amplicons. On average 15.708 reads were obtained for each sample. The data was analyzed using the BiQ Analyzer software.

3.6 *IN SILICO* DESIGN OF GRNAS

Methylome data of 2-cell-like cells and mESCs from (Eckersley-Maslin *et al*. 2016) and MEFs (Gao *et al*. 2018) were analyzed using SeqMonk (Babraham Institute). The Running Window Generator was used to generate tiling probes across the genome with size of 500 bp. Read count quantitation was then performed on both strands. To create a quantitation pipeline for bisulfite methylation over features, a minimum count of 1 was used to include in position and the minimum observations to include over feature. The promoter areas (+/- 500bp) of selected genes were analyzed (Annex figure 2).

Once a suitable area at the promotor was selected, DNA sequence was obtained from the UCSC Genome Browser (http://genome.ucsc.edu.). For gRNA design the online platforms CHOPCHOP (Montague *et al*., 2014) and Benchling were used, gRNAs with >65 on-target and >70 off-target score were selected.

3.7 STATISTICAL ANALYSIS

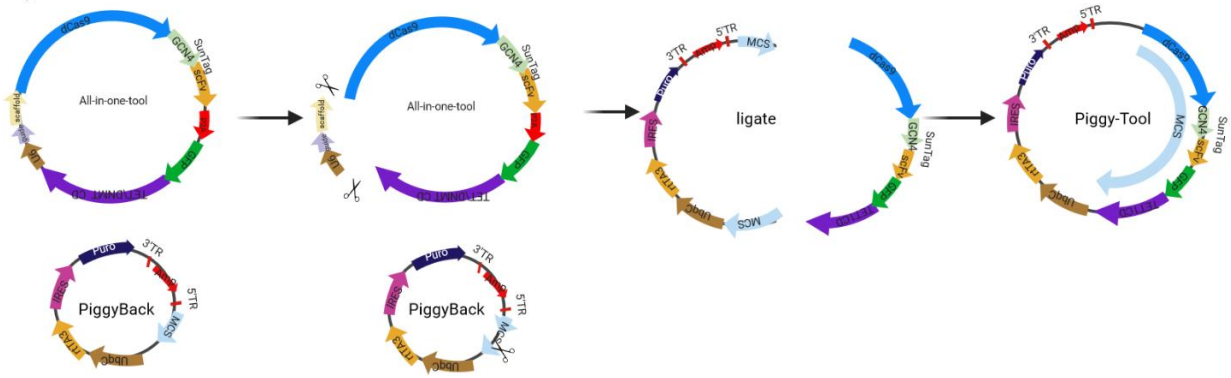
One-way ANOVA and T- test were performed to evaluate the results using GraphPad Prism (version 9.1.1 for Windows, GraphPad Software, San Diego, California USA, www.graphpad.com").

4 RESULTS

4.1 GENERATING AN INDUCIBLE EPIGENOME EDITING SYSTEM

The first aim of this master thesis was to create an inducible EE-tool. For this purpose, the All-in-one-tool previously established in the Wossidlo laboratory (without the gRNA expression system) was cloned under a Tet-On 3G promoter on a PB transposon vector (Figure 8A). The newly synthesized construct was then transfected into mESCs and the cells were selected on puromycin for eight days. After the selection Dox was added to the media to activate the EE system and GFP positive cells were FACS sorted. Finally, the sorted cells were single-plated on a 96-well plate and selected colonies were further expanded. Once the stable cell line was created its ability to manipulate the epigenome was tested by subsequent transfection with gRNA and methylation analysis of the gRNA and system expressing (double positive) cells (Figure 8B).

A)



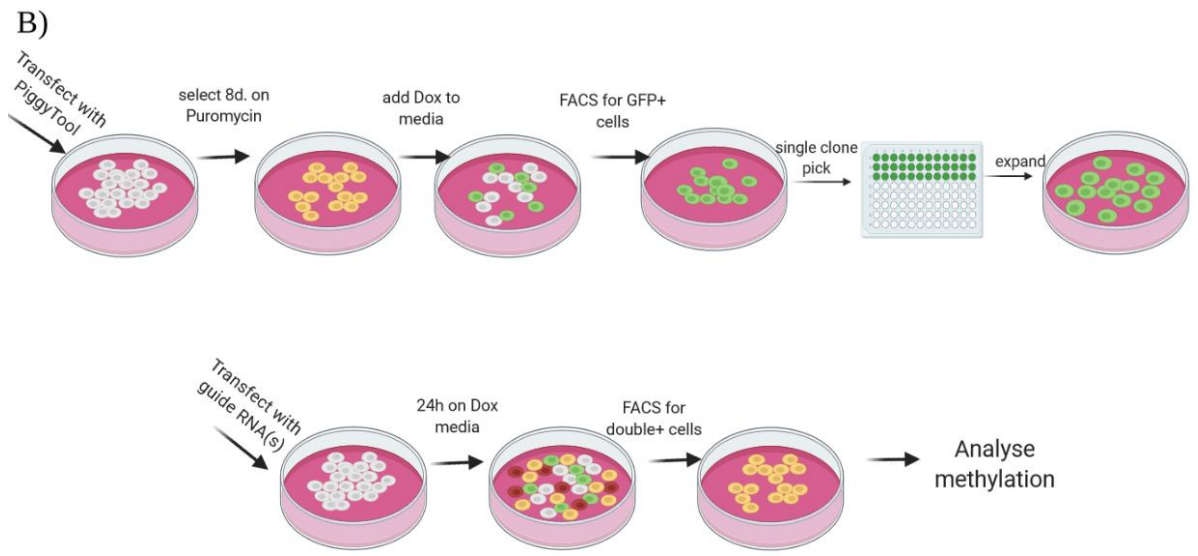


Figure 8: Experimental work flow. *A) Cloning of the Piggy-Tool. An All-in-one-tool is digested to remove the gRNA system. The multiple cloning site (MSC) of the PB vector is opened and the two plasmids are ligated. B) Creating Piggy-Tool expressing stable cell line. mESCs are transfected with the Piggy-Tool and selected on media containing puromycin for 8 days. On day 8 Dox is added to the media. On day 9 GFP positive cells are FACS sorted. Single clones are picked on a 96 well-plate. Homogeneous-looking clones are expanded. Next, the stable cell line is transfected with gRNAs and Dox is added to the media. Double positive cells are FACS sorted and cultured for one day, then their methylome is analyzed. Figure created with BioRender.com*

4.1.1 Cloning Doxycycline inducible PB-TET vector

A PiggyBac (PB) transposase and a PB vector (PB\tetO\Otx2\UbqC\Puro) was received from Dr. Christa Bücker. The Otx2 gene was removed from the multiple cloning site (MCS) using the AgeI and EcoRI restriction enzymes. Next, a 34bp long pair of oligonucleotides (Table 1) was ordered to fill in the MCS identical to the PB_TRE_MCS_Ubqc_rtta_Ires_Puro vector. The MCS was then digested with the EcoRI restriction enzyme and the vector was dephosphorylated. From the TET-All-in-one-ng-tool, already generated in the Wossidlo laboratory, the dCas9, SunTag, TET CD and GFP-reporter were amplified by PCR. Finally, the PB was ligated to the All-in-one-tool (Figure 8A) and transformed into competent bacterial cells. Upon plasmid DNA extraction restriction analysis was performed using the KpnI enzyme to ensure correct orientation of the insert (Figure 9). The DNA was then sequenced to check for potential single nucleotide polymorphisms and no mutation was detected (Annex figure 2A).

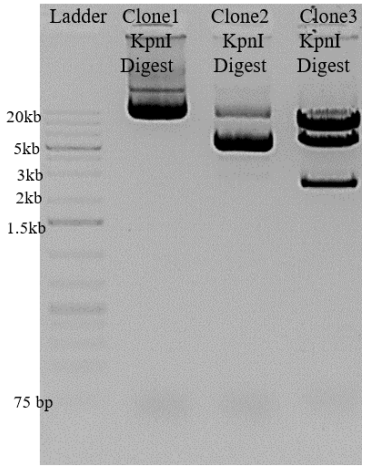


***Figure 9. Test digest of selected clones for the PB-TET vector**, expected sizes: 12 kb, 6 kb and 100 bp. Clone 2 shows correct orientation.*

4.1.2 Cloning Doxycycline inducible PB-DNMT vector

A Doxycycline inducible DNMT system was created in collaboration with Prof. Dr. Albert Jeltsch. To ensure higher efficiency Jeltsch *et al.* fused the DNMT3A and DNMT3L CDs (Siddique, A *et al*., 2013). In addition, the Arginine residue 887 has been mutated to Glutamate (R887E) to enhance DNA binding, allowing for lower off-target effects (Hofacker d, *et al*, 2019). Two plasmids were obtained from Prof. Dr. Albert Jeltsch: 03_09_scFv-GCN4-sfGFP-GB1-NLS_v2_3aR887E-3L harboring the mutated DNMT3A::DNMT3L fusion and a scFV array; 02_77_dCas9-10XSunTag-BFP-short_10kb with a 10 repeats long GCN4 array.

First, the All-in-one-DNMT-tool was improved by replacing the 5x SunTag with the 10x SunTag from the Prof. Dr. Albert Jeltsch-plasmid. Next, the DNMT3 was replaced with the mutated DNMT3A::DNMT3L fusion (Figure 11). From the ready All-in-one-tool, the steps for the generation of the PB-TET- tool was repeated as described in section 4.1.1 and Figure 8A. Test restriction was performed using BaeI, the selected clone was sequenced (Annex figure 2B).

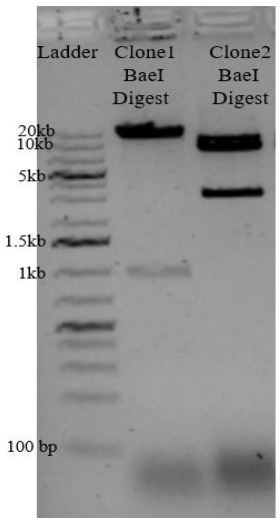


***Figure 10: Test digest of selected clones for the PB-DNMT vector**, expected sizes: 16,4 kb, 964 bp and 33 bp. Clone 1 shows correct orientation*

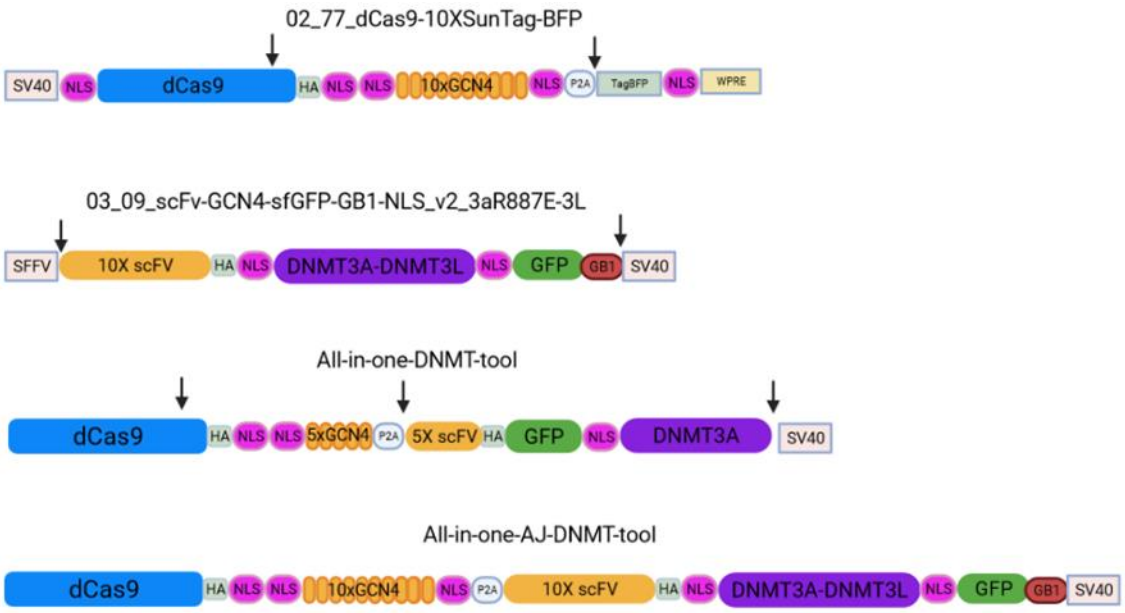


***Figure 11: Creating an enhanced DNMT-tool with plasmids from Prof. Dr. Albert Jeltsch** (AJ). From the dCas9-10xSunTag plasmid the 10xGCN4 array was cut out with restriction enzymes. From the scFV-GCN4-sfGFP-GB1-NLS_v2_3aR887E-3L the scFV array and the DNMT3A::DNMT3L was cut out with restriction enzymes. From the All-in-one-DNMT-tool the part from the dCas9 until the DNMT was removed with restriction enzymes. Arrows indicated the places, where the restriction enzymes cut. To generate the All-in-one-AJ-DNMT-tool, the digested All-in-one-DNMT-tool was ligated with the 10xGCN4 array and the DNMT3A::DNMT3L fragments from the plasmids obtained from Prof. Dr. Albert Jeltsch. Figure created with BioRender.com*

4.1.3 Puromycin Death curve

To determine the optimal concentration of puromycin for selection 40 000 wild type E14 cells were seeded in three 24 well-plates. Plate one received 2µg/ml puromycin, plate two – 1µg/ml and plate three – 0,5µg/ml puromycin. After 24h all cells from plate one were dead, plates two and three had surviving cells, media was changed and the same amount of puromycin added. 72h after the addition of puromycin all cell from plate two were dead, thus, 1µg/ml was determined as the optimal puromycin concentration (Figure 12).

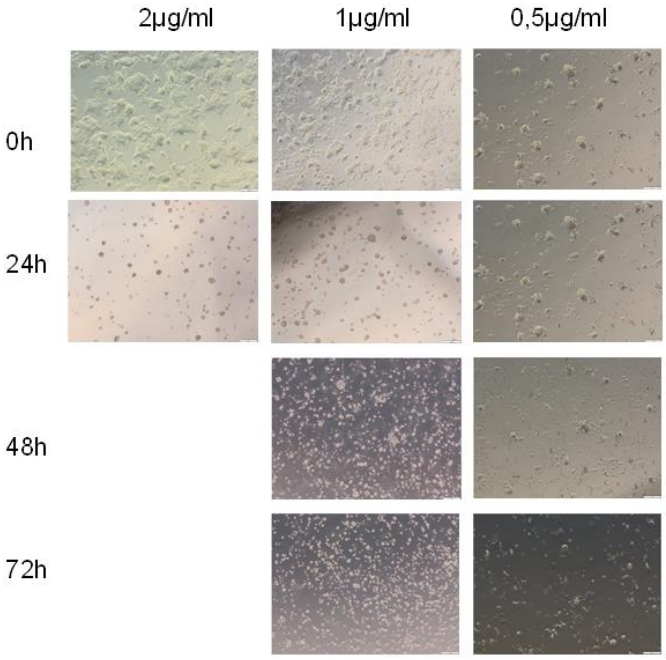


Figure 12: Puromycin death curve. *To 70% confluent 24 wells puromycin was added in three different concentrations: 2µg/ml, 1µg/ml and 0,5 µg/ml. Images of the cells were taken every 24 h for three days. Magnification 40x, scale bar = 50µm.*

4.1.4 Creating a stable cell line

To create a Dox inducible PB-TET stable cell line the PB-TET and transposase vectors were co-transfected into mouse embryonic stem cells (mESCs). The cells were then selected on puromycin for eight days. Next, upon addition of Dox GFP-expressing cells were sorted (Figure 8B, Annex figure 1) The newly created cell line was named “PiggyTET”.

4.2 CHARACTERIZATION OF THE GENERATED INDUCIBLE EE CELL LINES

4.2.1 GFP expression upon Dox addition

The first important property of the newly created cell line to test was at which timepoint upon addition of Dox to the medium the system is activated. 1µg/ml of Dox was added to the medium (following the literature Bencsik *et al*, 2016) and cells were imaged for GFP expression under a fluorescence microscope every six hours. In some cells the tool was activated as early as six hours upon Dox addition. After 24 hour all cells expressed the tool (Figure 13).

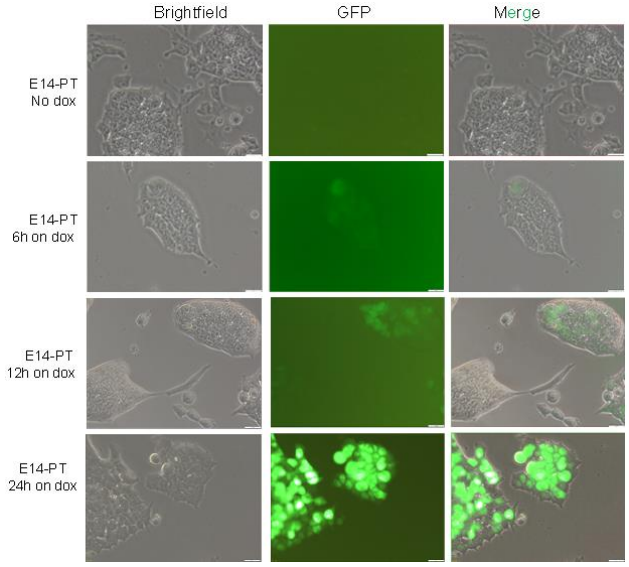


Figure 13: Time track of GFP expression upon addition of Dox to the stem cell media*. Dox with concentration of 1 µg/ml was added to the stem cell media and cells were observed under a microscope 0, 6, 12 and 24 h upon addition. Images were taken with 40x objective and 500ms exposure time. Brightfield and GFP images were merged using imageJ. Magnification 40x, scale bar 20µm.*

4.2.2 gRNA expression time track

The Dox-inducible system provides the opportunity to induce overexpression of the EE-tool over a longer period of time compared to transient transfection of the All-in-one plasmid. Thus, the limiting factor for a prolonged EE would be the expression of the respective gRNA. To test how long the gRNA is expressed in the cells, PiggyTETs were transfected with a mCherry co-expressing gRNA-expression vector and testing derived cells for mCherry expression was used as a proxy for gRNA expression. The number of mCherry-positive cells was followed via flow cytometry for five days (Figure 14). Starting with ~50% mCherry positive cells on day 1 the expression of mCherry

gradually declined in the next days, disappearing completely by day five (Figure 15) suggesting a cell division-dependent dilution of the gRNA expression vector over time.

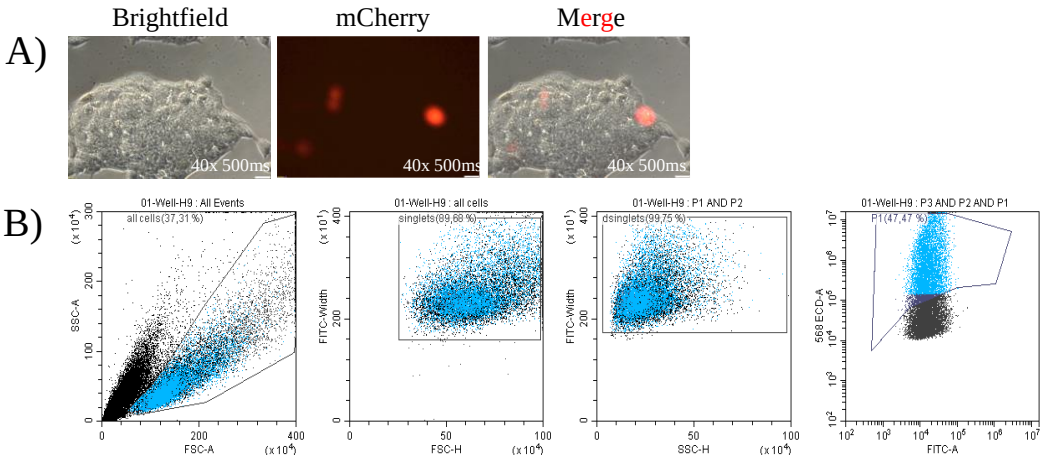


Figure 14. gRNA expression test: *A) mCherry expressing cells on day 2 after transfection. Images were taken with fluorescent microscope with 40x objective and 500ms exposure time, scale 20µm. B) Flow cytometry charts. In the first panel forward (FSC) and side (SSC) scatter were used to distinguish living cells. In the next two panels single and double single cells were isolated. In the last panel mCherry positive cells are selected.*

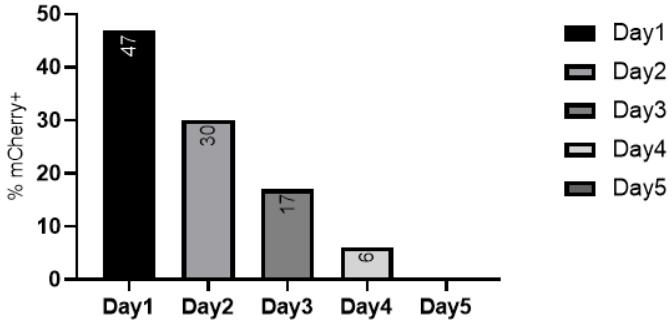


Figure 15. gRNA expression over 5 days. *The y-axis shows the percentage of mCherry positive cells, the x-axis represents the days. The figure was created with GraphPad Prism.*

4.2.3 Karyotyping of derived cell lines

The PB transposon system integrates the transgene at random positions into the genome. To test whether the integration of the TET-tool has interfered with the chromosomal integrity of the cells, chromosomal spreads of PiggyTET cells (Figure 16) to those of the wild type E14 cell line (E14-WT) were compared. Mouse ESCs have 40 chromosomes, since the E14 cell line used during this master thesis was already many passages old, 53.3% of the 30 analyzed cells had 40 chromosomes. From the PT-E14 38% of the 21 counted cells displayed normal chromosomal count (Table 4).

Since there was no significant difference between the two cell lines, the PT-E14 were used in further experiments.

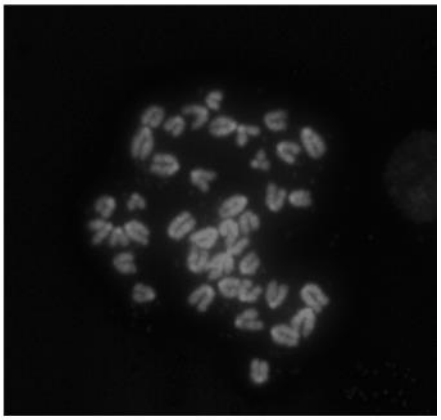


Figure 16: Chromosomal spread of PiggyTET cells. *The expected number of 40 chromosomes is observed.*

Table 4: Chromosome count of E14-WT and PT-E14 cells. *The cell line, n= number of chromosomes counted, and total number of counted cells is given. Cells counted with respective number of chromosomes and their percentage is estimated below.*

E14-WT	n=38	n=39	n=40	n=41	n=42	n=43	n=44	n=45	n=46	Total count
Cells count	2	4	16	1	1	2	3	1	2	30
Percentage	6,6	13,3	53,3	3,3	3,3	6,6	10	3,3	6,6	
PiggyTET-E14	n=37	n=39	n=40	n=41	n=42	n=43	n=44			Total count
Cell count	1	1	8	4	4	2	1			21
Percentage	4,7	4,7	38	19	19	9,5	4,7			

4.3 MULTIPLEXED GRNA DELIVERY SYSTEM: CHIMERIC ARRAY OF GRNA OLIGONUCLEOTIDES (CARGO)

As described in section 4.1, the PB-inducible tool requires separate delivery of gRNA. To further explore the reprograming potential of this system, multiplexed editing of several genomic loci in the same single cell was attempted. Gu *et al.* have developed a clever cloning system for assembling multiple gRNAs in one vector, enabling their stochiometric delivery to all cells transfected (Gu *et al.*, 2018) (Figure 17).

As backbone the Px330 vector was adapted by replacing the Cas9 region with a mCherry reporter (Figure 18A) and mutating all preexisting BbsI recognition sites (Figure 18B).

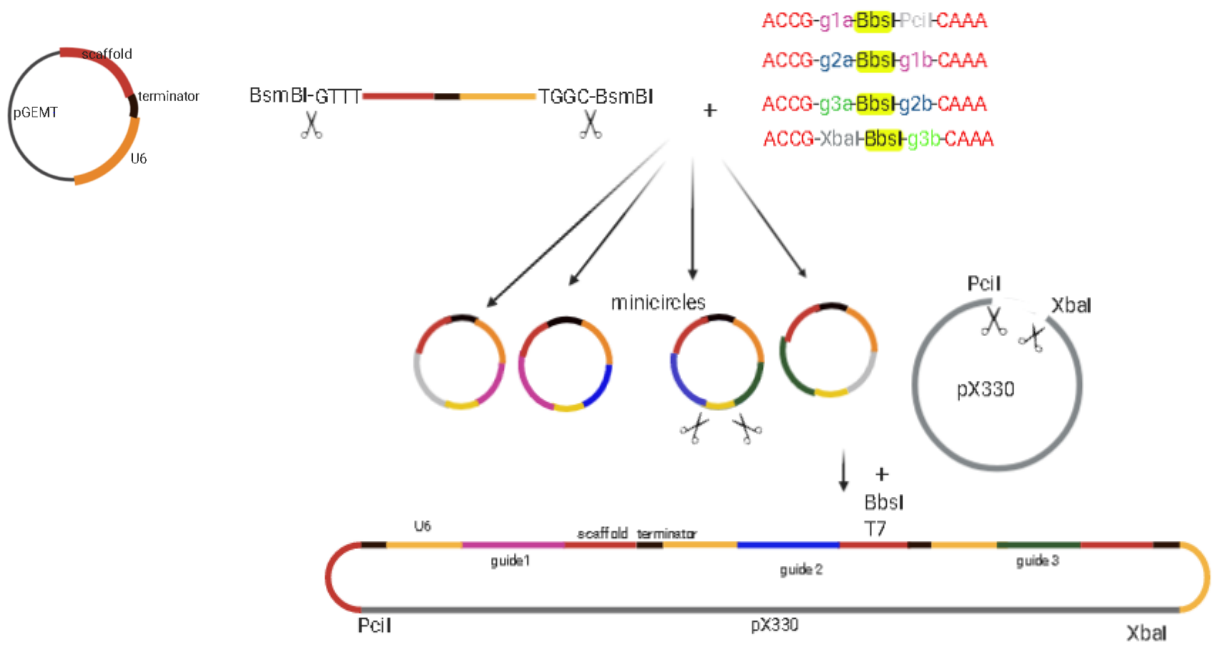


Figure 17: CARGO cloning strategy. *The constant region comprising the gRNA scaffold (brown line), termination sequence (black line) and U6 promoter (yellow line) is cut out of a pGEM-T vector with the BsmBI Type IIS restriction enzyme. The constant region is then ligated to the variable region: 50bp long oligonucleotides comprising compatible with the constant region sticky ends, the first halve of a gRNA and the second halve of the following gRNA, separated with two mirrored BbsI recognition sites. The first and last variable regions contain complementary sequences to the final vector backbone. The ligation of the variable to the constant regions results in 400bp long minicircles. The minicircles and the digested final vector backbone pX330 (gray circle) are combined together with the BbsI restriction enzyme and a ligase in a single tube reaction. After series of restrictions and ligation a final product is created consisting of the backbone, and each gRNA equipped with its own U6 promoter, a scaffold and termination sequence. Figure created with BioRender.com*

The variable region (gRNA seed sequences) was ordered as 50-60 bp long oligonucleotides, which were aligned and de-phosphorylated. To create the constant region two primer sets were designed on the All-in-one-tool. The first set encompassing the U6 promoter and the second pair – the scaffold and termination sequence. The two PCRs were then combined by a bridge-PCR with the forward primer for the constant region and the reverse primer for the U6 promoter. The construct was then cloned into a pGEM-T vector with recognition sites for the type IIS restriction enzyme BsmBI to create unique sticky ends compatible with the variable region. The backbone was then digested with PsiI and Xbai restriction enzymes and the 4,5 kb band was extracted (Figure 19A). Next, the constant region was freed from the pGEM-T vector using the BsmBI restriction enzyme (Figure 19A). The minicircles were then formed by ligating the constant region and the aligned and dephosphorylated oligos (Figure 19B). In the final reaction the minicircles and the backbone were combined together with a ligase and the BbsI restriction enzyme. The construct was transformed

into competent bacterial cells. Upon plasmid DNA extraction restriction analysis was performed using the NdeI enzyme to ensure correct incorporation of all components (Figure 19C). The DNA was then sequenced to exclude potential mutations (Annex figure 2C).

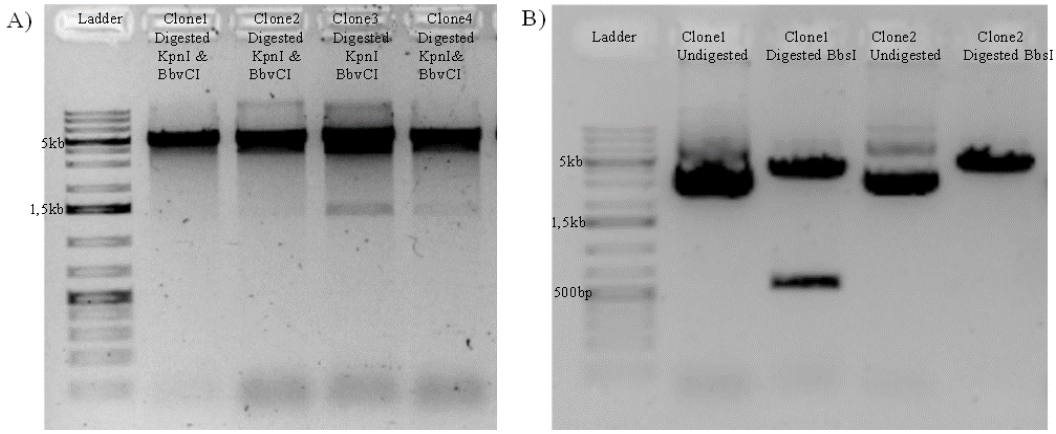


Figure 18: CARGO backbone vector. *A) Px330 with mCherry insert. Test digest with KpnI and BbvCI. Plasmid size: 5kb, expected sizes of the digest: 3,6kb and 1,3 kb. B) BbsI recognition site mutated. Plasmid size: 5kb. In the case of an unsuccessful mutation a 500 bp piece is to be observed (Clone1), in Clone2 the enzyme is not cutting, indicating that the recognition site is mutated.*

First, this protocol was tested by creating a single-gene targeting Gfap-CARGO (Figure 19). Once a successful protocol was established a 2C-CARGO (section 4.4) with seven gRNA as described in section 4.4 was created. The DNA was then sequenced to exclude potential mutations (Annex figure 2D).

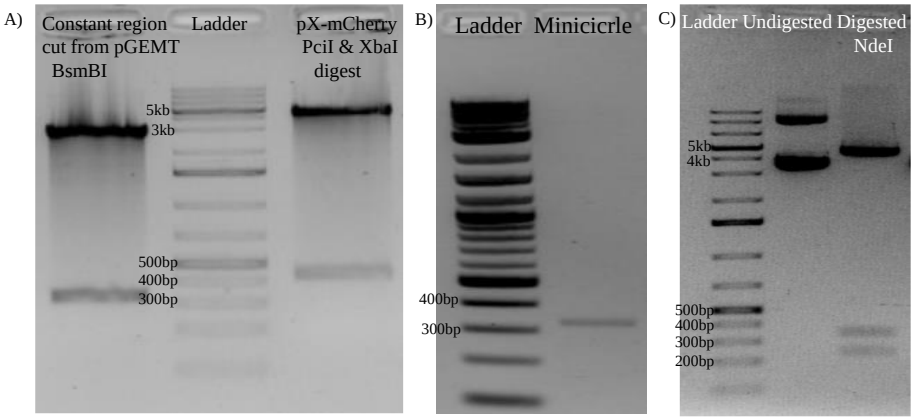


Figure 19. Creating the GFAP-CARGO. *A) Left lane: digesting the constant region out of the pGEM-T backbone with BsmBI, size of the constant region: 365 bp, size of the pGEM-T: 3 kb; middle lane: ladder; right lane: digesting the final backbone with PciI and XbaI, expected sizes: 4,2 kb and 400bp; B) Minicircle, size: 365bp; C) Test digest of the assembled CARGO with NdeI. Expected sizes: 4,5 kb, 350 bp and 250 bp.*

4.4 gRNA SELECTION FOR INDUCTION OF THE 2C POPULATION

After the CARGO-cloning protocol was established, the next step was to create a multitarget CARGO for manipulation of the 2CLC-population named accordingly "2C-CARGO". The target selection for the 2C-CARGO was based on Eckersley-Maslyn *et al*., 2019. Methylome data in the promoter area close to the transcriptional starting site (Eckersley-Maslin *et al*., 2016; Gao R *et al*., 2018) of genes, whose overexpression showed most upregulation of the 2C-like population (Figure 21), was analyzed using SeqMonk (Babraham Institute). Since different genome assembly versions were used by Eckersley-Maslin *et al*. and Gao R *et al*., the MEF data is shown separately (Figure 22). The following six candidate promoter areas of genes with differently methylated in 2CLCs compared to ESCs and MEFs were selected as target for the respective gRNAs (Figure 22).

Developmental pluripotency associated 2 (Dppa2) and Dppa4 are both expressed at the 2C-like stage and important for EGA (Eckersley-Maslin *et al*., 2019; De Iaco *et al*., 2019). Knockdown of either of them reduces 2CLCs population, where knockout completely abolishes it (Eckersley-Maslyn *et al*., 2019). The promoter area of both genes is lowly methylated in ESCs (10% for Dppa2 and 0% for Dppa4); however, since they are methylated in MEFs (promoter areas of both genes ~80% methylated) (Figure 22E-G), these genes are of interest to target with a TET-tool to test reactivation of Dppa2/4 in somatic cells. In this context, studies have shown that demethylation of the promoter area can lead to the expression of these genes (Eckersley-Maslyn *et al*., 2019; Hernandez *et al*., 2018) – hence, they were included in the 2C-CARGO as candidate factors for studying MEFs. As the study of Eckersley-Maslyn *et al*., 2019 was conducted with an older version of the mouse assembly genome (mm9), the positions of the targets were compared to mm10 and Dppa2 appeared around 6kb upstream of its previous position. Thus, it was decided to create two gRNAs for Dppa2, one based on mm10 and one on the older assembly, to test for more efficient editing (Figure 20, Figure 22E-G).

The next target Zscan4c activates early embryonic genes and is essential for 2-cell stage progression in the embryo, is expressed only at the 2C-stage and in ESCs (Falco *et al*., 2007). The Zscan4 gene family contains six isoforms named a to f, where Zscan4c is most commonly expressed form in ESCs. The overexpression of Zscan4c increases 2CLC-population (Hirata *et al*., 2012; Eckersley-Maslin *et al*., 2019). According to the methylome data, its promotor area is hypermethylated in ESCs (86%) and less methylated in 2CLCs (~13%) (Figure 22C). Unfortunately, no publicly available methylation data was available for Zscan4c in MEFs.

The transcription factor Double Homeobox (Dux) is conserved among placental animals and drives EGA (De Iaco *et al*., 2017). It binds to Dppa2 promoter and totipotency genes like Zscan4c and activates 2CLC population in ESCs cultures (Yang *et al*., 2020). Dux knock outs do not express MERVL and Zscan4c (De Iaco *et al*. 2017). Dux binding regulates Zscan4c (Tagliaferri *et al*., 2020; Hendrickson *et al*., 2017), therefore, even as it was not part of the list of 2C regulating genes provided by Eckersley-Maslyn *et al*., 2019, it was included on the 2C-CARGO. The promoter area of Dux is 80% methylated in mESCs and 50% methylated in 2CLCs and MEFs (Figure 20, Figure 22A).

The next 2C-CARGO target was the Sp110 nuclear body protein: regulator of gene expression and ribosome biogenesis (Leu *et al.*, 2018). Sp110 is upregulated in 2CLCs. Its promotor area is 75% methylated in ESCs, and 50% in 2CLCs, there was no publicly available methylation data for Sp110 in MEFs (Figure 20, Figure 22B).

The last candidate, Trp63, also known as p63, encodes tumor protein p63, a member of the p53 family of transcription factors involved in cellular responses to stress and development. Knockout of p63 in mice results in developmental defects such as lack of limbs and other tissues (Assefnia *et al*., 2014). Trp63 was expected to be 90% methylated in ESCs, 80% in MEFs and ~50% methylated in 2CLCs (Figure 20, Figure 22D).

As an unrelated control gRNA the STAT3 binding region of Gfap was selected. GFAP is important for astrocytes development, not the 2C-state. Thus, it was interesting to analyze whether an unrelated gRNA can decrease potential off-target effects without interfering with the 2C-population (Morita *et al*., 2016). The expected DNA methylation levels of targeted CpGs at the Gfap locus are 90% in MEFs.

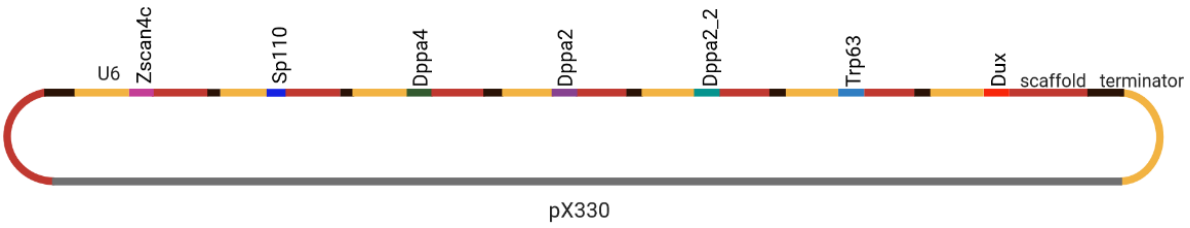


Figure 20: 2C-CARGO. *On a pX330 vector backbone the seven gRNAs (Zscan4c, Sp110, Dppa4, Dppa2, Dppa2_2, Trp63 and Dux) are cloned. Each gRNA starts with an U6 promoter (yellow line) and is followed by a scaffold (brown line) and a termination sequence (black line). Figure created with BioRender.com*

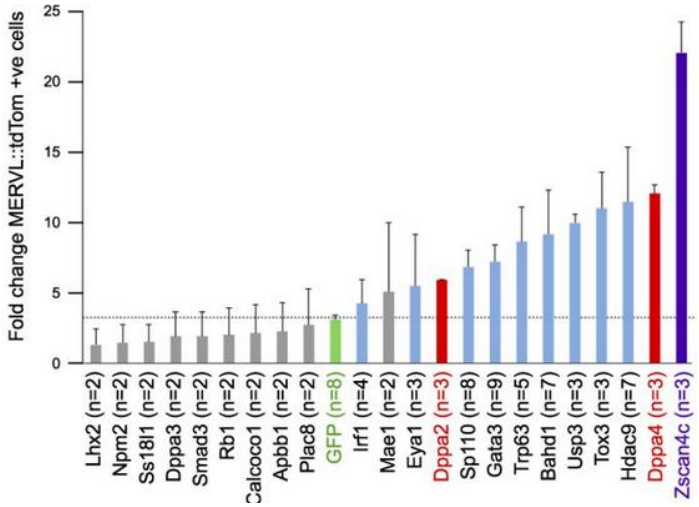
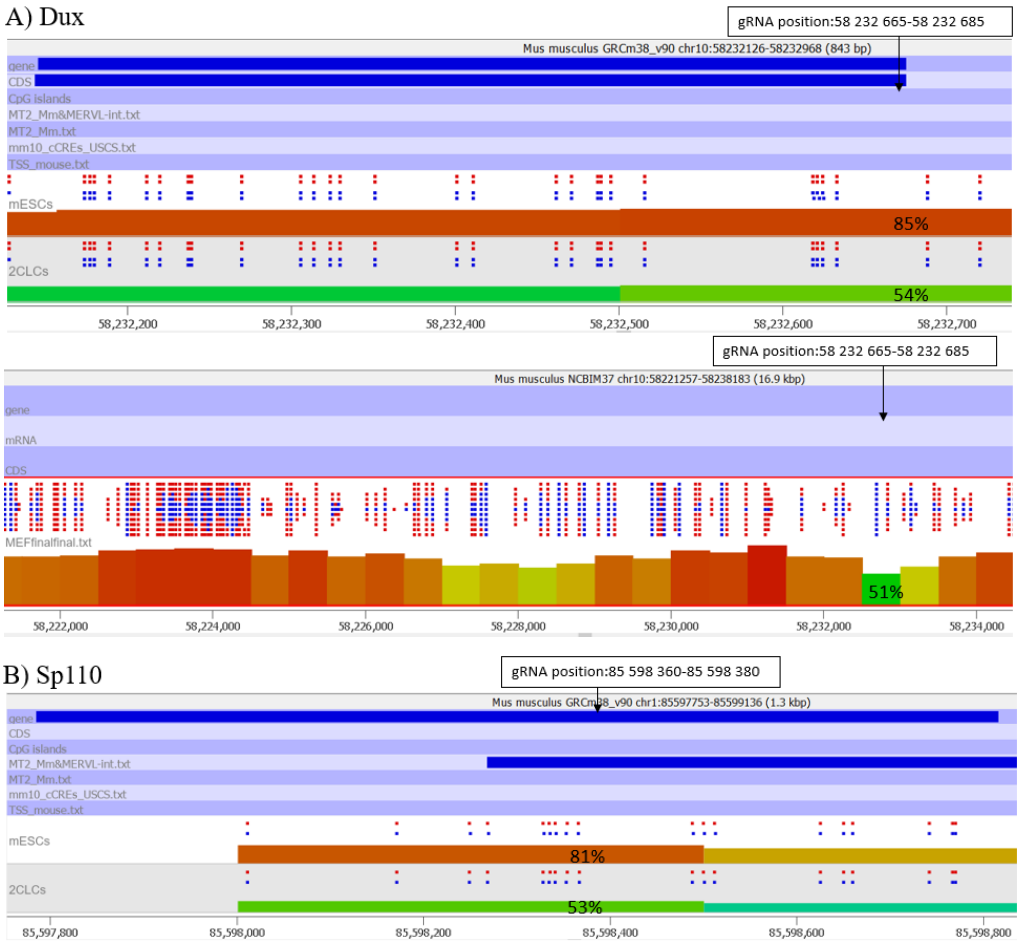
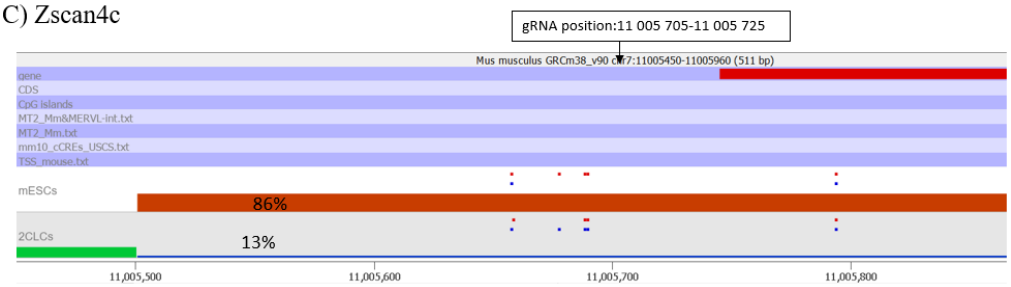


Figure 21: Fold change of MERVL::tdTomato positive cells. *Shown are transfected and untransfected cells. In green is the control transfected only with GFP. Data shown are average and Standard deviation (STD). Replicates number is shown under each gene. Image is taken from Eckersley-Maslin et al., 2019*



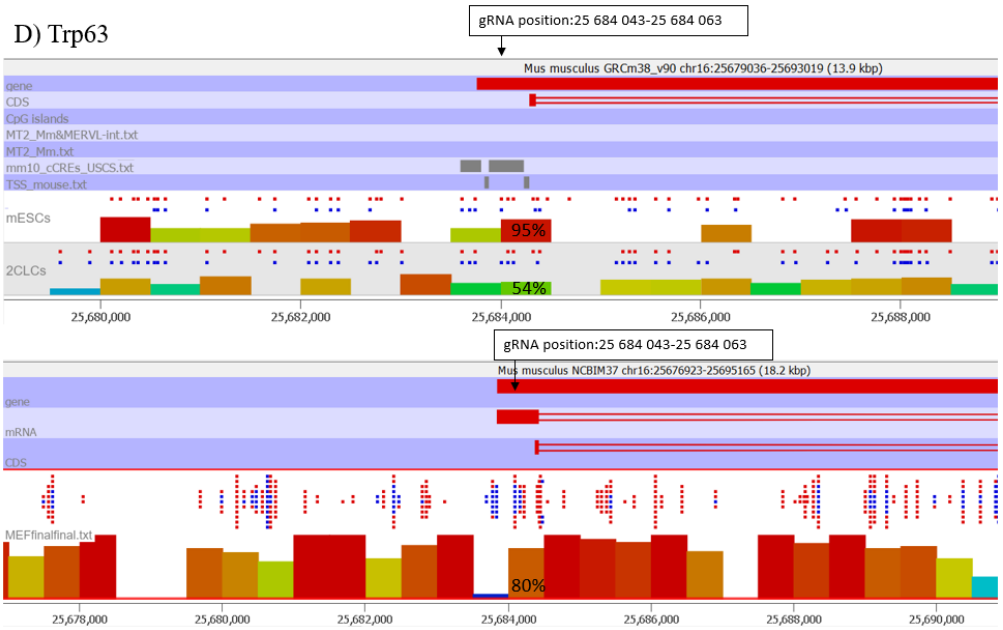
MEFs: methylome data missing

C) Zscan4c

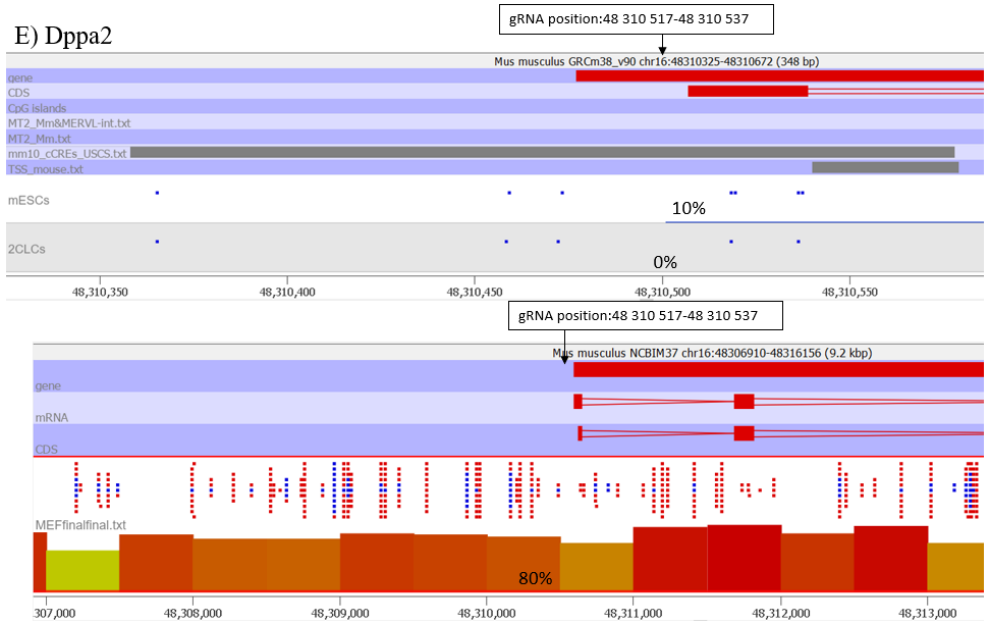


MEFs: methylome data missing

D) Trp63



E) Dppa2



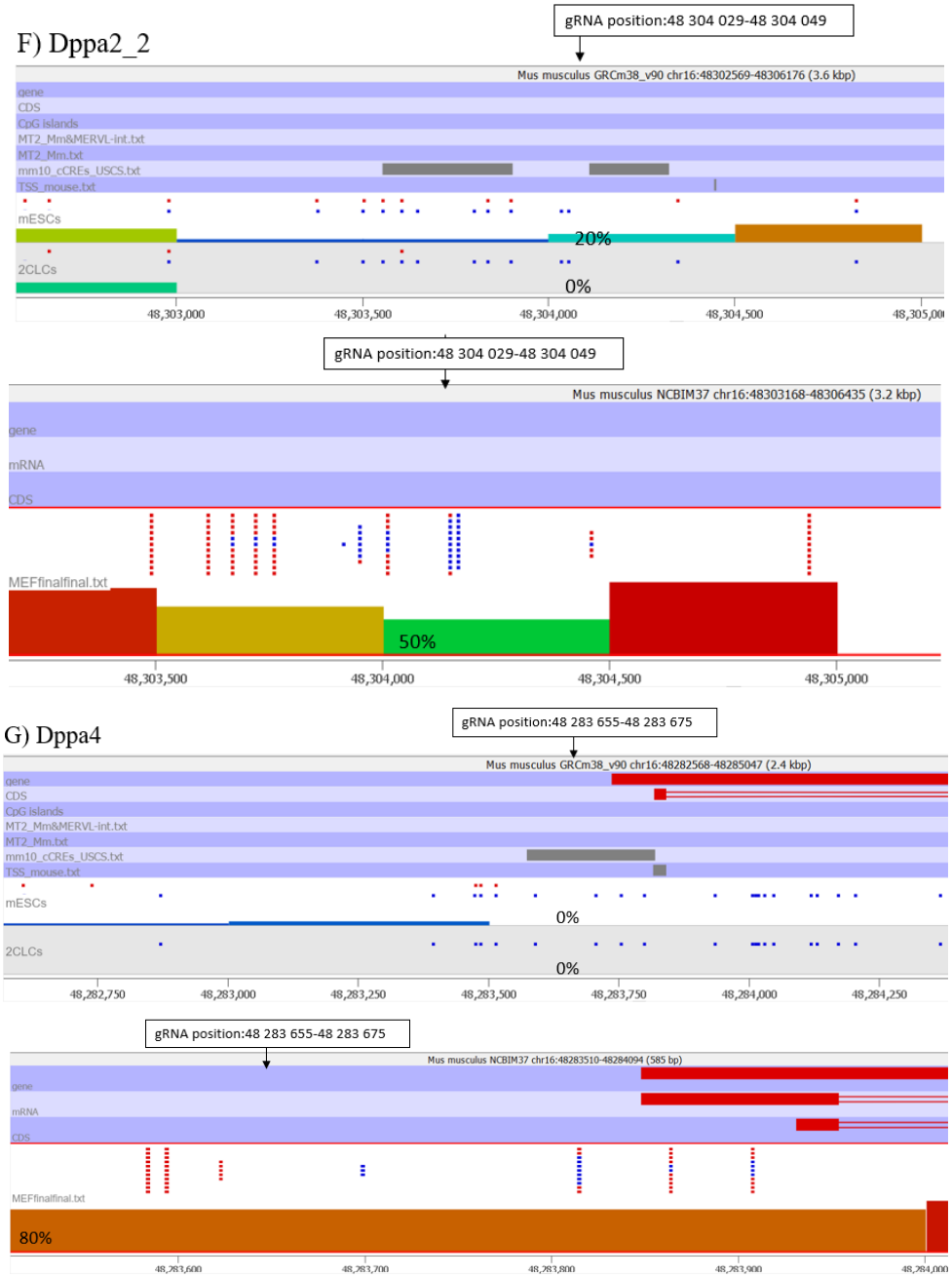


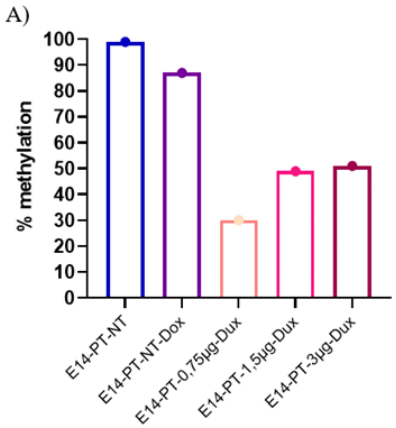
Figure 22: SeqMonk screenshots of the promotor area of the 2C-CARGO targets. *Above the chromosome (chr) and gene coordinates are shown. gRNA position is indicated with a black arrow, with coordinates written next to it. The corresponding gene is shown as a line (blue for reverse transcription of the gene, red for forward). TSS are shown as gray squares. CpGs are depicted as dots, red for the forward and blue for the reverse strand. Shown are tiling probes with size of 500 bp; above, in red, are the methylation levels for the mESCs; below, in green, the 2CLC methylation levels. Position on the chr is indicated at the bottom. Underneath is a screenshot showing the methylation levels in MEFs. The methylation percentage of the targeted tile is shown. Methylome date obtained from Eckerslev-Maslin et al., 2016 and Gao et al., 2018.*

4.5 MANIPULATION OF THE METHYLATION STATE OF TARGETED GENES

4.5.1 Determining the optimal amount of gRNA expression vectors

First, to achieve efficient editing of DNA methylation by the generated inducible EE-tool, the optimal amount of gRNA expression vector for the transfection was determined. PiggyTET-E14 cells were seeded on 24-well plates and transfected with 0,75µg, 1,5µg and 3µg of Dux gRNA expressing plasmid. As negative and off-target controls untransfected PiggyTET-E14 and PiggyTET-E14 with Dox (no gRNA) added to the media were used. Reporter-expressing cells were sorted (Annex figure 1) and changes in DNA methylation were analyzed on day three after transfection with combined bisulfite restriction analysis (COBRA) and next-generation-sequencing (NGS). According to the COBRA analysis results, the most efficient editing was achieved with the lowest gRNA dosage (Figure 23A).

The NGS data provides a more detailed picture (Figure 23B). Interestingly, every CpG-position behaves differently but with a similar trend, with some more prone to editing that other CpGs. Unfortunately, the CpG, for which the COBRA analysis for Dux was performed, was not detected by the NGS-analysis, making it difficult to draw conclusions regarding its editability. What also becomes prominent from the data is that activating the tool alone without a gRNA produces a small but significant amount (~10%) of off-target effects (E14-PT-Dox). However, providing a gRNA to the EE system results in much stronger editing (~30%). Comparing the results from the NGS and COBRA data, there is a small discrepancy when it comes to the editing efficiency of 0,75µg versus 1,5µg of gRNA. It is important to keep in mind, however, that COBRA analyses a single CpG-position only and, as it is apparent from the NGS data, there is a difference in the methylation status and accessibility between neighboring CpGs.



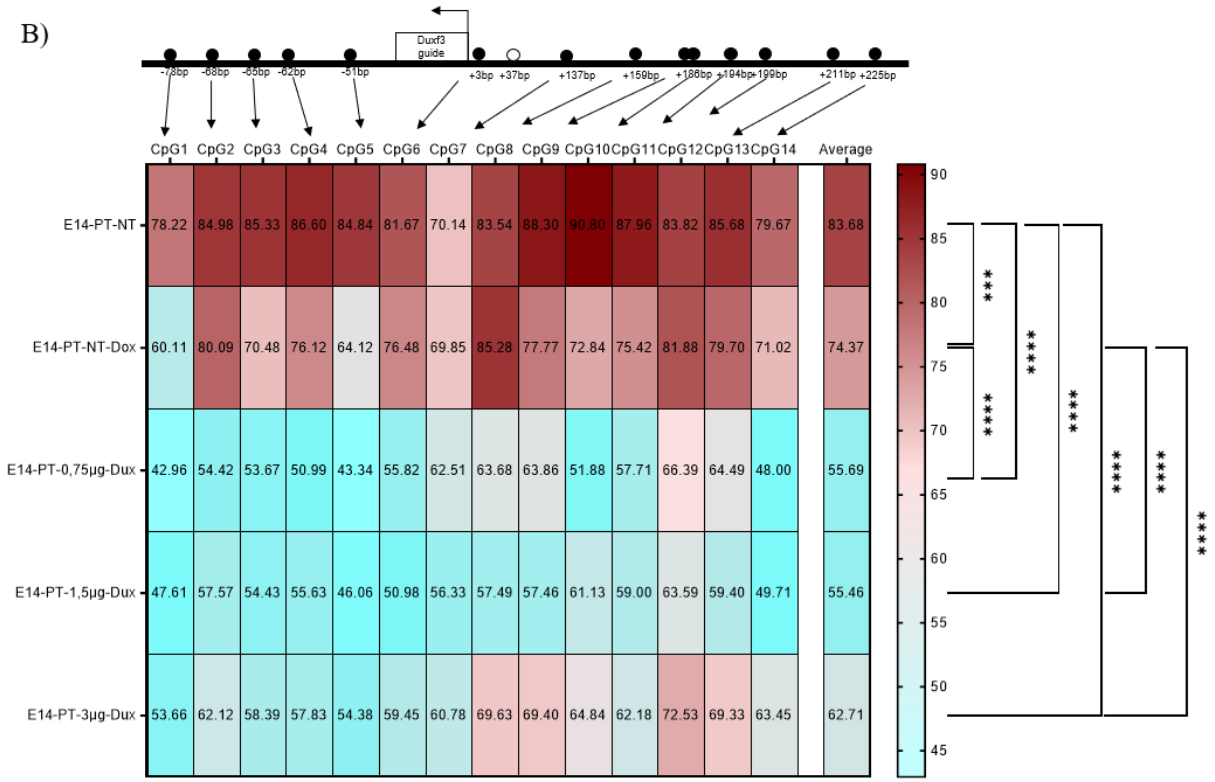


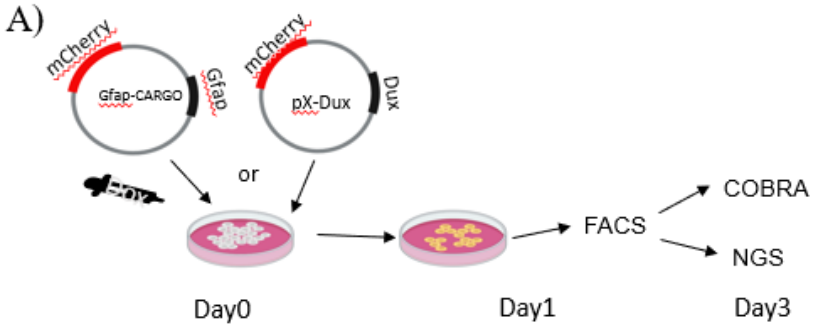
Figure 23: The effects of different amounts of gRNA expression vector on editing efficiency. *A) COBRA results in blue non-transfected PT cells (E14-PT-NT); purple: E14-PT-NT with Dox in the media; orange: PT cells transfected with 0,75µg gRNA; pink: PT cells transfected with 1,5µg gRNA; dark-red: PT cells transfected with 3µg. Data shown is one technical replicate for each condition. B) Heat map of NGS data showing CpG methylation at targeted loci. At the top a map of the targeted region, where each CpG is represented as a black dot, with its distance from the gRNA underneath. The CpG which the enzyme is cutting used for the COBRA was not detected by the NGS analysis and is thus shown as a hollow circle. P value > 0.05 was determined as not significant (ns), P value ≤ 0.05 was marked with *, P value ≤ 0.01 was marked with **, P value ≤ 0.001 was marked with **, P value ≤ 0,0001 was marked with ****. Figures were created using GraphPad Prism.*

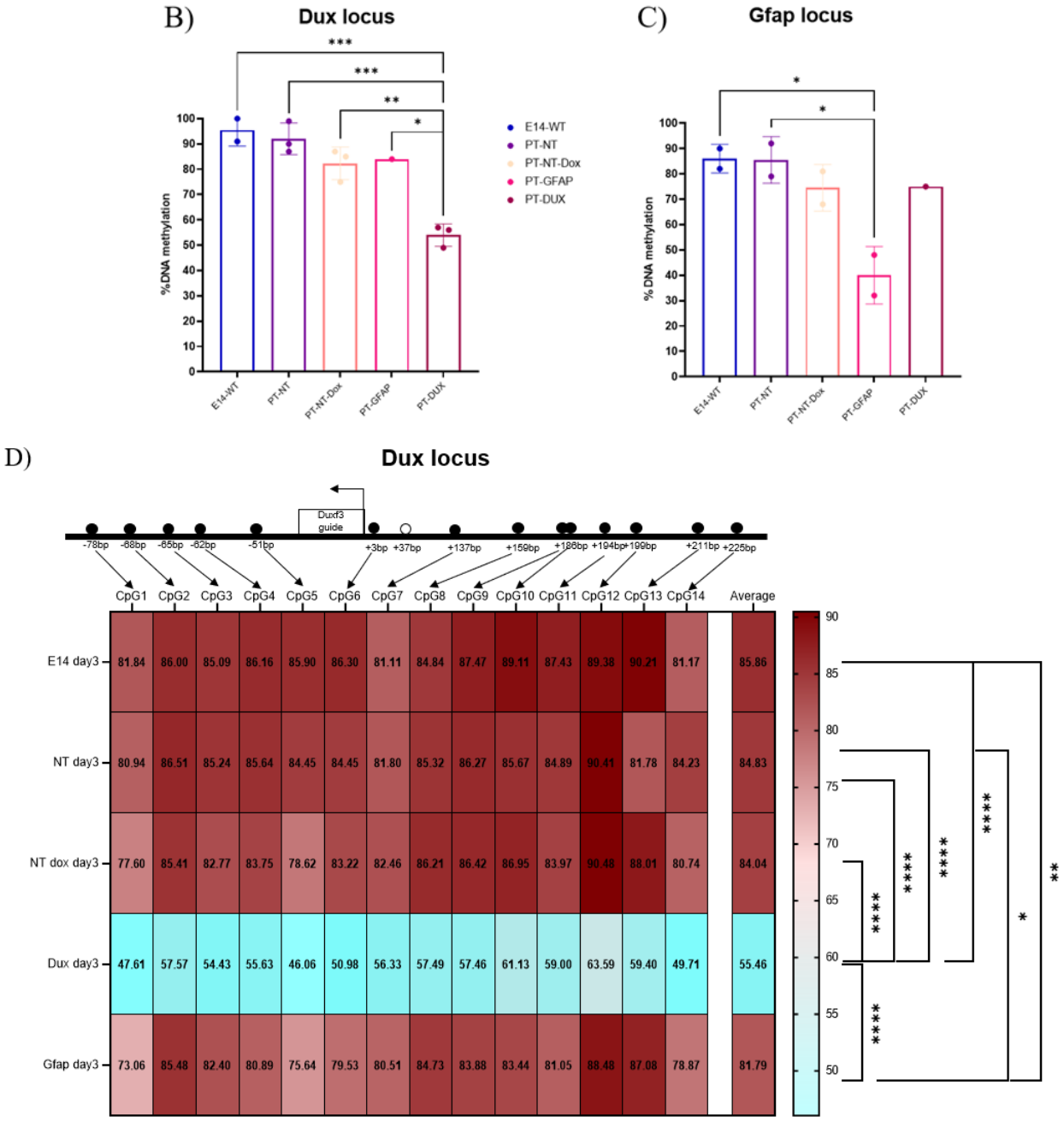
4.5.2 Testing the editing-efficiency of the PB system

In the past an All-in-one-plasmid has been used in the Wossidlo laboratory, because it provides a defined stochastic delivery of all required components of the editing system. To test if the separate delivery of a gRNA can induce efficient editing also for longer time, E14-PT cells were transfected with a Dux gRNA on a pX-mCherry expression vector (section 3.2.1) and as an off-target control the CARGO-cloned Gfap gRNA was used. On the day of the transfection (day 0) Dox was added to the medium. On day one post transfection, cells double positive for GFP and mCherry were sorted (Annex figure 1) and further cultured. Subsequently, DNA extraction and bisulfite conversion were performed on day three. The conversion was analyzed both by COBRA and NGS (Figure 24). WT-E14 and non-transfected (NT) PiggyTETs were used to control for off-target

effects. Further, NT-E14-PiggyTETs with Dox added to the media served as to control for off-target effects induced by the activation of the system alone. To test if the gRNA is able to meliorate off-targets, non-targeted loci were analyzed (Dux locus on Gfap transfected cells and Gfap locus on Dux transfected cells).

According to both COBRA and NGS data on day three post transfection there was no significant difference in the methylation status of E14-WT cell and PiggyTET cells (Figure 24B-E) indicating that the tool is inactive in the absence of Dox. Activation with Dox without gRNA, shows significant off-target effects at the Gfap locus according to the NGS analysis (~15% lower methylation, Figure 24E); no such effects were detected at the Dux locus, or by COBRA (Figure 24C-D). When the Dux locus was examined, where cells where targeted with Gfap gRNA, a 10% decrease of methylation was detected by the NGS analysis (Figure 24E). There were no off-target effects detected at the Gfap locus in cells targeted with Dux gRNA, however. Overall, there is highly significant difference between the methylation status of targeted and untargeted loci, regardless of the transfected gRNA (Figure 24D-E). Dux and Gfap demonstrated highly significant editing efficiencies compared to non-transfected PiggyTET and E14-WT cells according to both the COBRA and NGS analysis (Figure 24B-E), and highly significant editing efficiency compared to all control conditions according to the NGS data (Figure 24D-E). These results demonstrate that efficient editing can be achieved using the PiggyTET cell line. Furthermore, the CARGO tool is validated as a reliable gRNA delivery system.




B)
Dux locus
%DNA methylation
E14-WT
PT-NT
PT-NT-Dox
PT-GFAP
PT-DUX
C)
Gfap locus
%DNA methylation
E14-WT
PT-NT
PT-NT-Dox
PT-GFAP
PT-DUX
D)
Dux locus
Duxf3 guide
-78bp -88bp -85bp -82bp -51bp +3bp +37bp +137bp +159bp +186bp +194bp +199bp +211bp +225bp
CpG1 CpG2 CpG3 CpG4 CpG5 CpG6 CpG7 CpG8 CpG9 CpG10 CpG11 CpG12 CpG13 CpG14 Average
E14 day3 81.84 86.00 85.09 86.16 85.90 86.30 81.11 84.84 87.47 89.11 87.43 89.38 90.21 81.17 85.86
NT day3 80.94 86.51 85.24 85.64 84.45 84.45 81.80 85.32 86.27 85.67 84.89 90.41 81.78 84.23 84.83
NT dox day3 77.60 85.41 82.77 83.75 78.62 83.22 82.46 86.21 86.42 86.95 83.97 90.48 88.01 80.74 84.04
Dux day3 47.61 57.57 54.43 55.63 46.06 50.98 56.33 57.49 57.46 61.13 59.00 63.59 59.40 49.71 55.46
Gfap day3 73.06 85.48 82.40 80.89 75.64 79.53 80.51 84.73 83.88 83.44 81.05 88.48 87.08 78.87 81.79

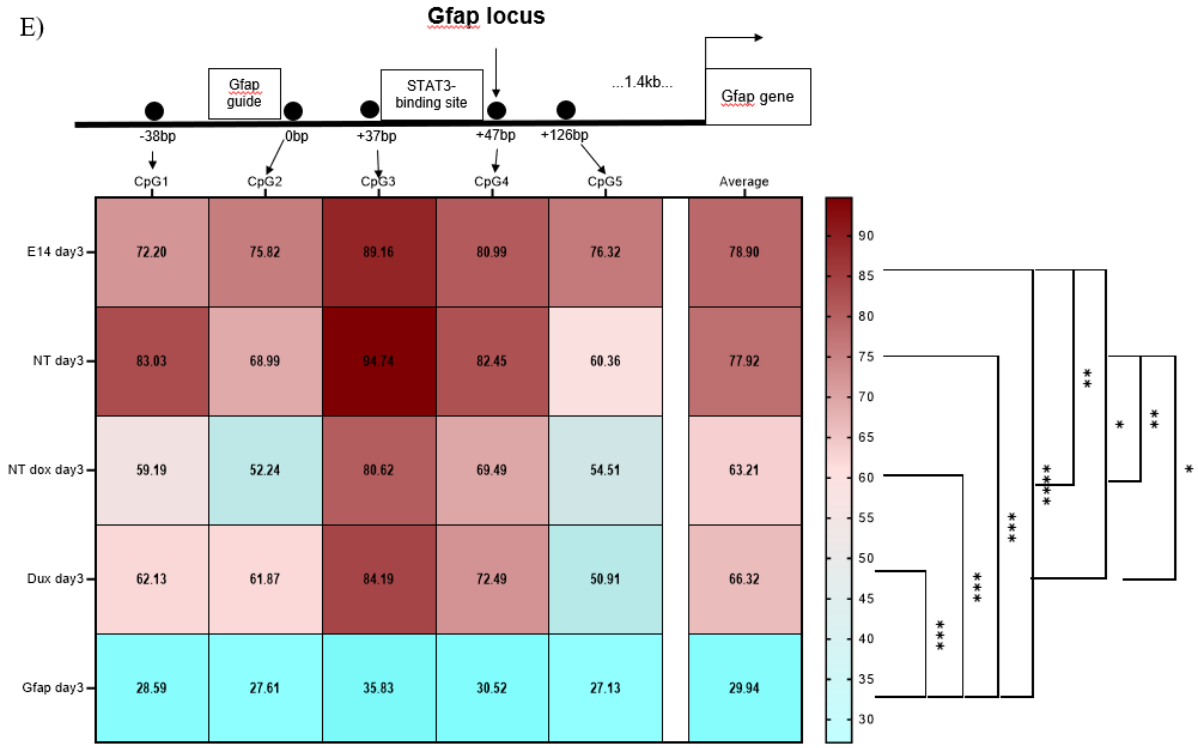


Figure 24 Testing the PB-system for successful epigenome editing. *A) Experimental set-up. Cells were transfected on day zero, cells double positive for GFP (TET-tool) and mCherry (gRNA) were FACS sorted on day one; methylation analysis was performed on day three. B-C) COBRA results of the Dux and Gfap locus respectively In blue WT-E14 cells; purple: PT non-transfected cells (PT-NT); orange: PT-NT with Dox in the media; pink: PT cells transfected with a Gfap-gRNA; dark-red: PT cells transfected with a Dux gRNA. Data shown are mean ±STD. Technical replicates are shown as dots. D-E) Heat map of NGS data showing CpG methylation at targeted loci for Dux and Gfap respectively. On top we see map of the targeted region, where each CpG is represented as a black dot, with its distance from the gRNA underneath. The CpG which the enzyme used for the COBRA of the Dux locus is cutting is not detected by the NGS analysis and is thus shown as a hollow circle. The CpG which the enzyme used for the COBRA of the Gfap locus is marked by an arrow. P value > 0.05 was determined as not significant (ns), P value ≤ 0.05 was marked with *, P value ≤ 0.01 was marked with **, P value ≤ 0.001 was marked with **, P value ≤ 0,0001 was marked with ****. Figure created with GraphPad Prism.*

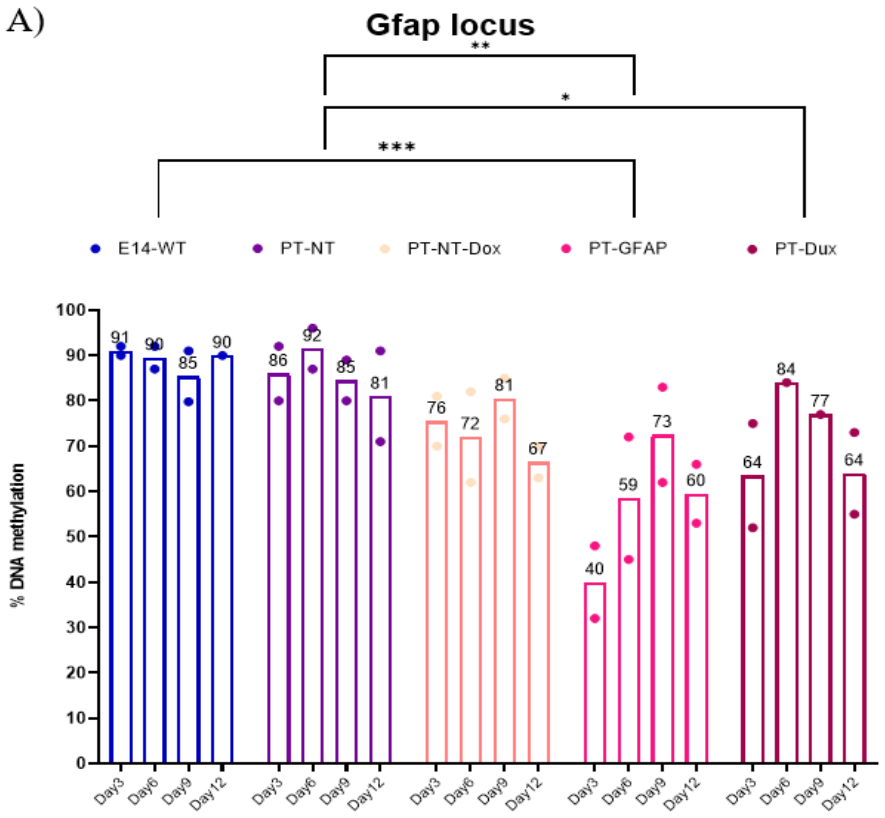
4.5.3 Testing long-term editing using the PB system

Next, it was tested whether stable EE under constant Dox-mediated induction can be achieved. PiggyTET-E14 cells were transfected with Dux-gRNA cloned in a pX-mCherry expression vector or Gfap-CARGO. One day post transfection cells were double sorted for GFP and mCherry expression (Annex figure 1). On days three, six, nine and twelve after transfection the methylation levels of the Dux and Gfap loci were analyzed by COBRA and NGS.

When comparing the mean methylation values between the different conditions, a significant overall difference in the methylation levels between the cells transfected with the

appropriate gRNA and the rest of the conditions was observed (Figure 25A, Figure 26A). On days three and six significant decrease of DNA methylation on the targeted locus of transfected cells in comparison with non-transfected cells was detected (Figure 25B-C, Figure 26B-C). This data was also confirmed by the NGS-analysis (Figure 27). At days nine and twelve no significant change in DNA methylation was measured (Figure 25D-E, Figure 26D-E). It is interesting to note, however, that despite the lack of substantial editing detected on later days, when comparing the methylation levels of the edited loci at days three to twelve, no significant difference was detected between them (Figure 25A, Figure 26A). Small, but significant off-target effects under Dox treatment are, again, to be noted (15%) at the Gfap locus (Figure 27B). From this data it was concluded that the induction of the EE-tool by Dox is solely responsible for the off-target effects and these are not reduced when the system is provided with an (unrelated) gRNA. In accordance with previous data (Figure 24) the methylation levels of wild-type and PT cells without Dox were similar, implying that the PB-system remains inactive in the absence of Dox.

In conclusion, the PiggyTET cell line allows for efficient EE for up to six days with transient transfection of gRNA expression vectors.



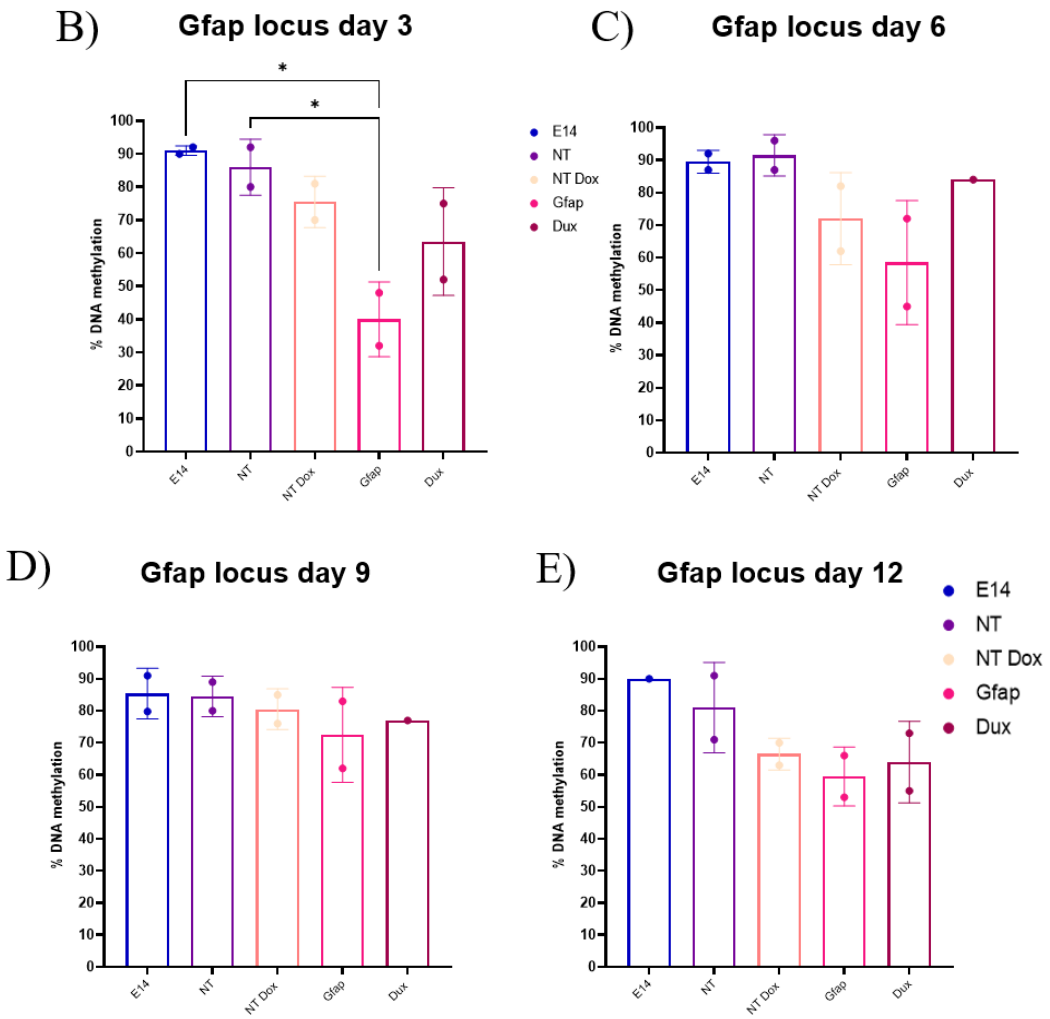
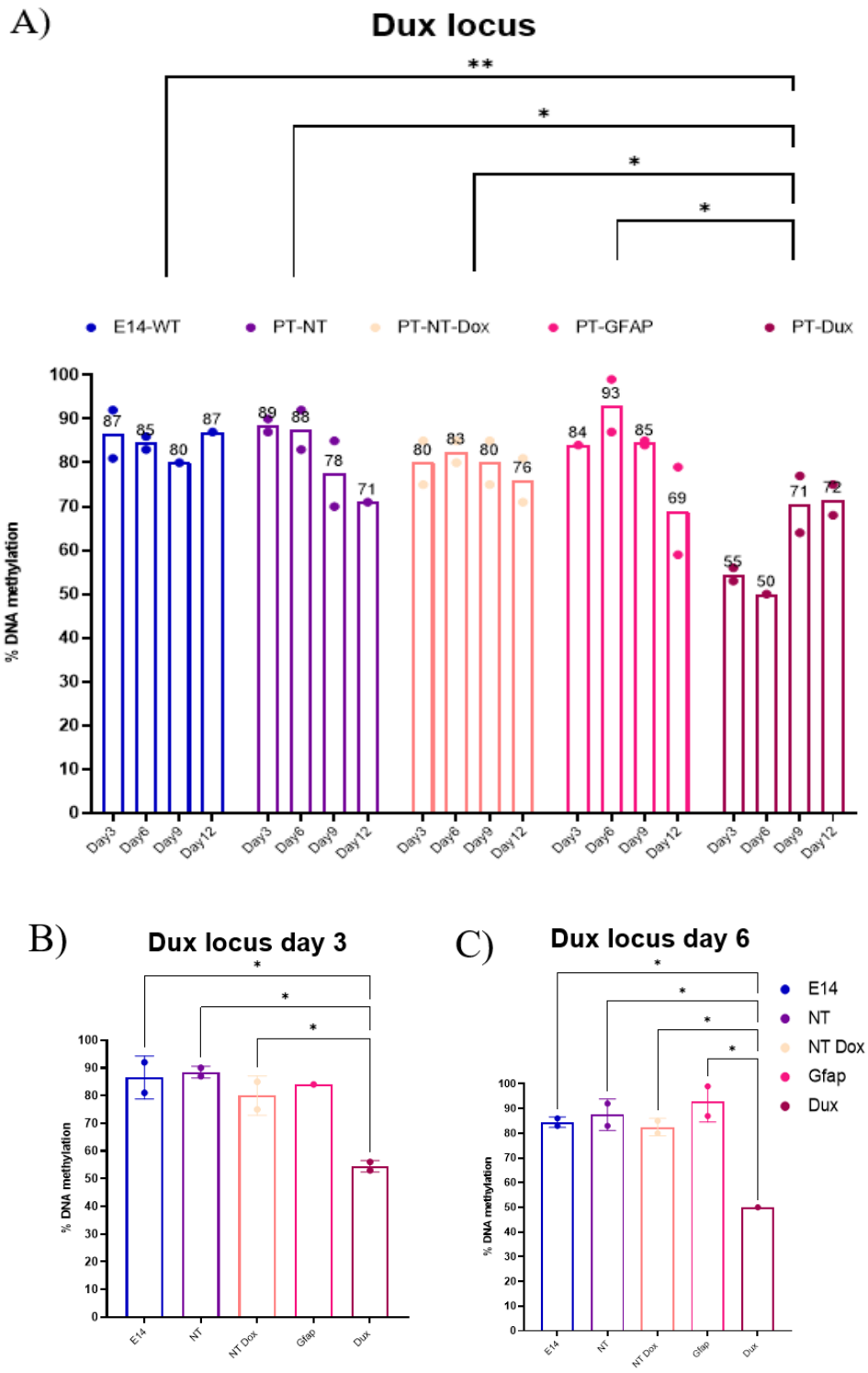


Figure 25: DNA methylation editing of the Gfap locus over the course of 12 days. *A) Methylation levels over the course of 12 days under different conditions. In blue WT-E14 cells; purple: PT non-transfected cells (PT-NT); orange: PT-NT with Dox in the media; pink: PT cells transfected with a Gfap-gRNA; dark-red: PT cells transfected with gRNA targeting the Dux promoter area. B -E) Methylation levels on the Gfap locus on days 3 to 12 respectively on WT-E14, PT-NT, PT-NT with Dox in the media; PT cells transfected with a Gfap-gRNA and PT cells transfected with a Dux gRNA. P value > 0.05 was determined as not significant (ns), P value ≤ 0.05 was marked with *, P value ≤ 0.01 was marked with **, P value ≤ 0.001 was marked with **, P value ≤ 0,0001 was marked with ****. Figure was created using GraphPad Prism.*


A)
Dux locus
**
*
*
*
E14-WT
PT-NT
PT-NT-Dox
PT-GFAP
PT-Dux
% DNA methylation
100
90
80
70
60
50
40
30
20
10
0
87
85
80
87
89
88
78
71
80
83
80
76
84
93
85
69
55
50
71
72
Day3
Day6
Day9
Day12
B)
Dux locus day 3
% DNA methylation
E14
NT
NT Dox
Gfap
Dux
C)
Dux locus day 6
% DNA methylation
E14
NT
NT Dox
Gfap
Dux

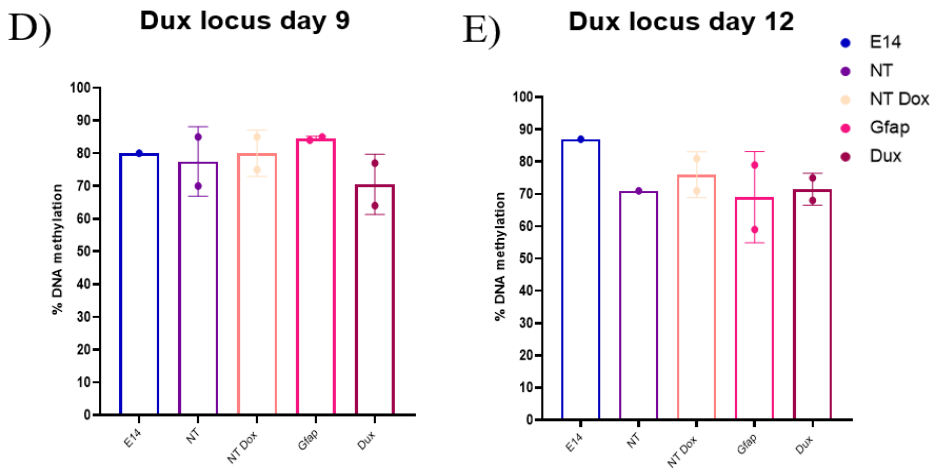
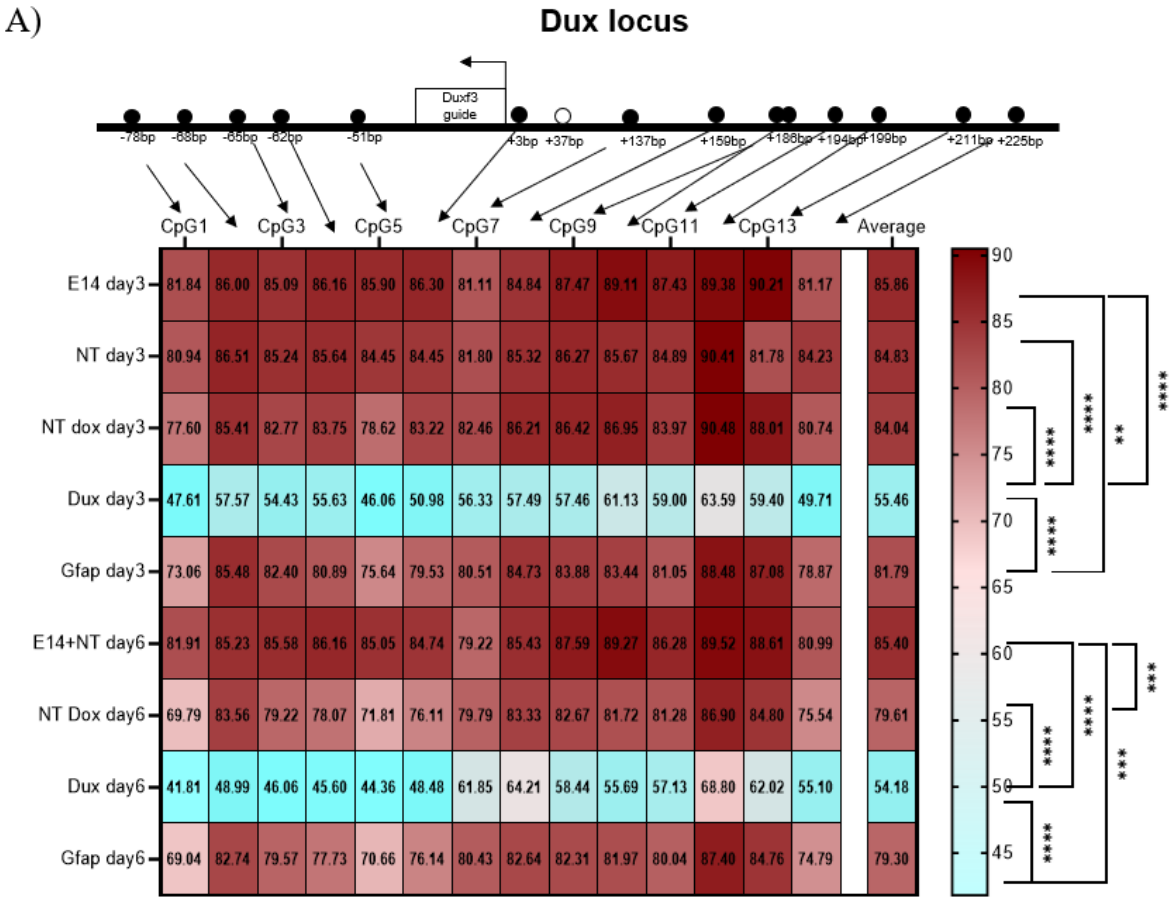


Figure 26: DNA methylation editing of the Dux locus over the course of 12 days. *A) Methylation levels over the course of 12 days under different conditions. In blue WT-E14 cells; purple: PT non-transfected cells (PT-NT); orange: PT-NT with Dox in the media; pink: PT cells transfected with a Gfap-guide; dark-red: PT cells transfected with a Dux guide. B) -E) Methylation levels on the Dux locus on days 3 to 12 respectively on WT-E14, PT-NT, PT-NT with Dox in the media; PT cells transfected with a Gfap-guide and PT cells transfected with a Dux guide. P value > 0.05 was determined as not significant (ns), P value ≤ 0.05 was marked with *, P value ≤ 0.01 was marked with **, P value ≤ 0.001 was marked with **, P value ≤ 0,0001 was marked with ****. Figure was created using GraphPad-Prism.*



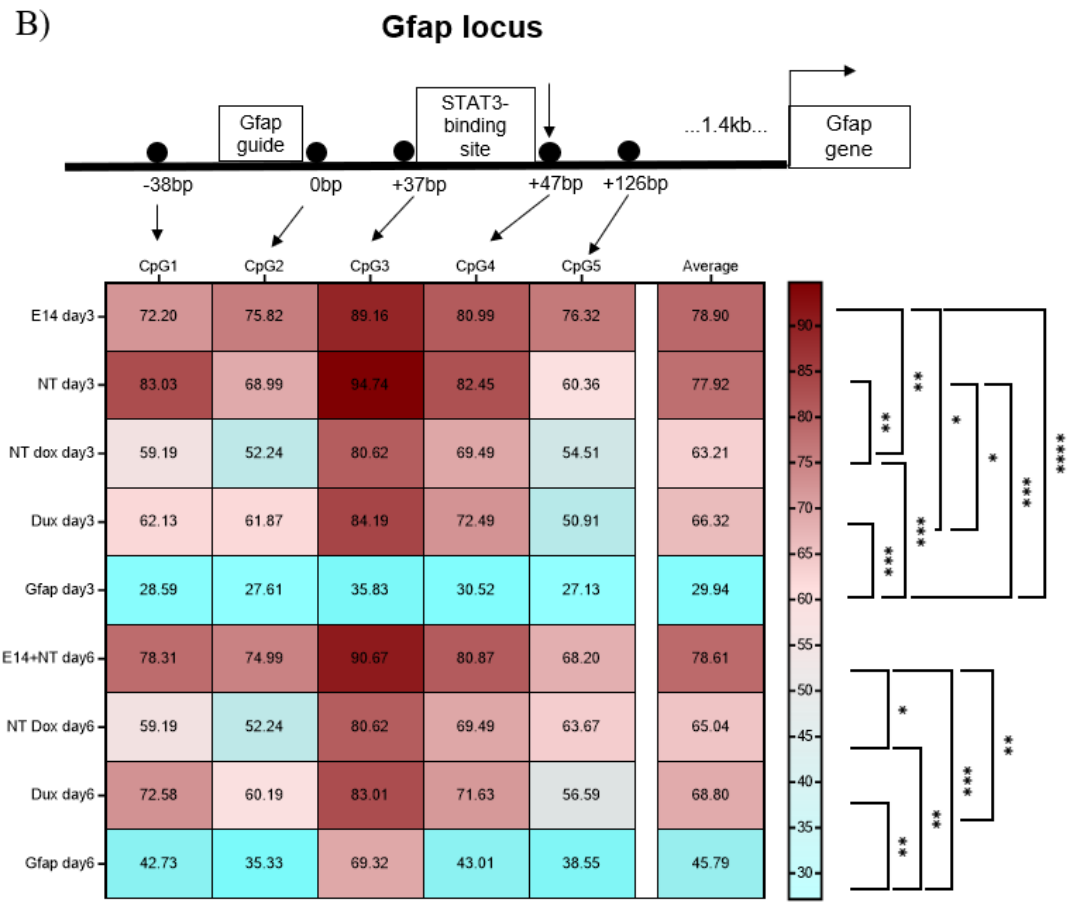


Figure 27: Heat map of NGS data showing CpG methylation at targeted loci for A) Dux and B) Gfap respectively. *On top a map of the targeted region, where each CpG is represented as a black dot, with its distance from the guide underneath. The CpG which the enzyme used for the COBRA of the Dux locus is cutting is not detected by the NGS analysis and is thus shown as a hollow circle. The CpG which the enzyme used for the COBRA of the Gfap locus is marked by an arrow. P value > 0.05 was determined as not significant (ns), P value ≤ 0.05 was marked with *, P value ≤ 0.01 was marked with **, P value ≤ 0.001 was marked with **, P value ≤ 0,0001 was marked with ****. Figure created using GraphPad Prism.*

4.5.4 Testing the 2C-CARGO on PiggyTET-E14 cells

The next goal was to test whether multiple genomic loci can be edited at the same time applying generated EE-tools. PiggyTETs were transfected with the 2C-CARGO construct expressing seven gRNAs simultaneously, described earlier. Double positive cells for GFP (TET-Tool) and mCherry (2C-CARGO) were sorted on day one and analyzed by COBRA on day three post transfections. The average transfection efficiency of the 2C-CARGO was 20% and 50 000 cells were sorted on average. As controls E14-WT cells, non-transfected PTs (NT), NT on Dox (ND) and PTs transfected with the Gfap-CARGO were used (Figure 28).

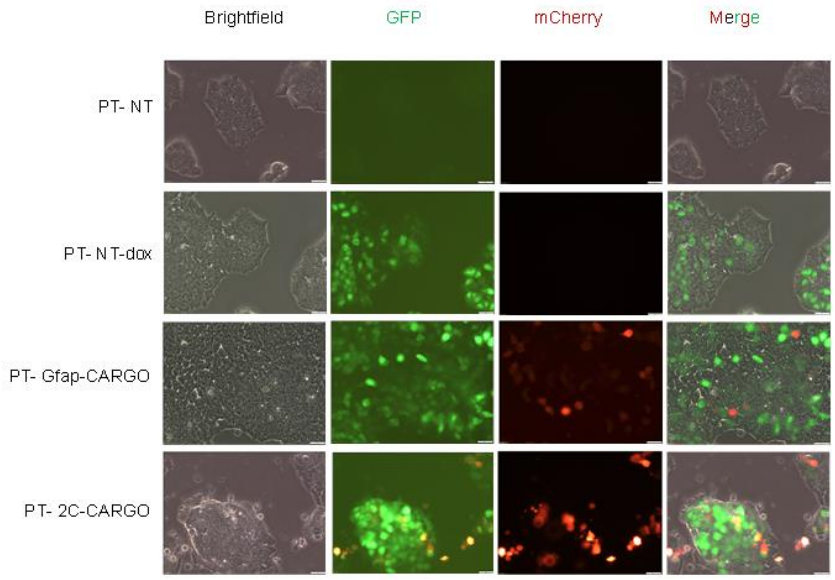


Figure 28: PT-E14 cells 24h post transfection and Dox activation: *top to bottom: non-transfected cells (PT-NT), PT-NT with the with Dox in the media (PT-NT-dox), PT transfected with the Gfap-CARGO, PT transfected with the 2C-CARGO. Left to right: brightfield, GFP, mCherry and a merged image. All pictures were taken with a 40x objective and 500ms exposure time. Scale bar = 50µm.*

According to the COBRA analysis, only the Dux gRNA achieved significant editing in the 2C-CARGO transfected cell with a decrease of 30%, in accordance with the previous experiment with single Dux gRNA (Figure 29, Figure 30A, Figure 27A). For the other targets of the 2C-CARGO no significant changes in DNA methylation between 2C-CARGO group and control groups were detected. Moreover, for some targets the overall observed methylation levels were lower than the predicted levels reported in public available methylome data of mESCs. Here, the reported methylation level for Sp110 was around 80% and similar methylation levels we observed in all experimental groups, regardless of the treatment (Figure 29, Figure 30B). For Trp63, 95% methylation was reported, however, the COBRA data shows very low methylation levels of that locus in the control groups fluctuating between the different experiments, with highest value of 52% on cells transfected with the Gfap-CARGO (Figure 29, Figure 30C). Zscan4c also behaves differently in the different replicates, with only one replica of the wild type cells coming close to the reported methylation levels of 86%. Overall, the cells fluctuate around 50% methylation under the different conditions at the Zscan4c locus (Figure 29, Figure 30D). The reported methylation levels for Dppa2/4 are very low (10-0% respectively). Following the lower trend of all previously analyzed loci, all of the samples show 0% methylation for Dppa2/4 (Figure 29, Figure 30E-G).

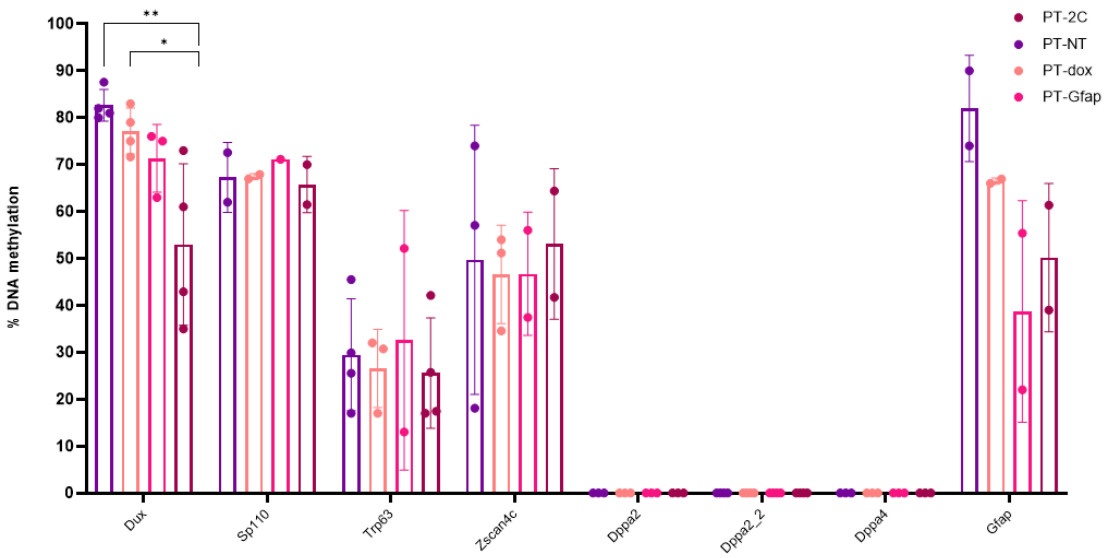
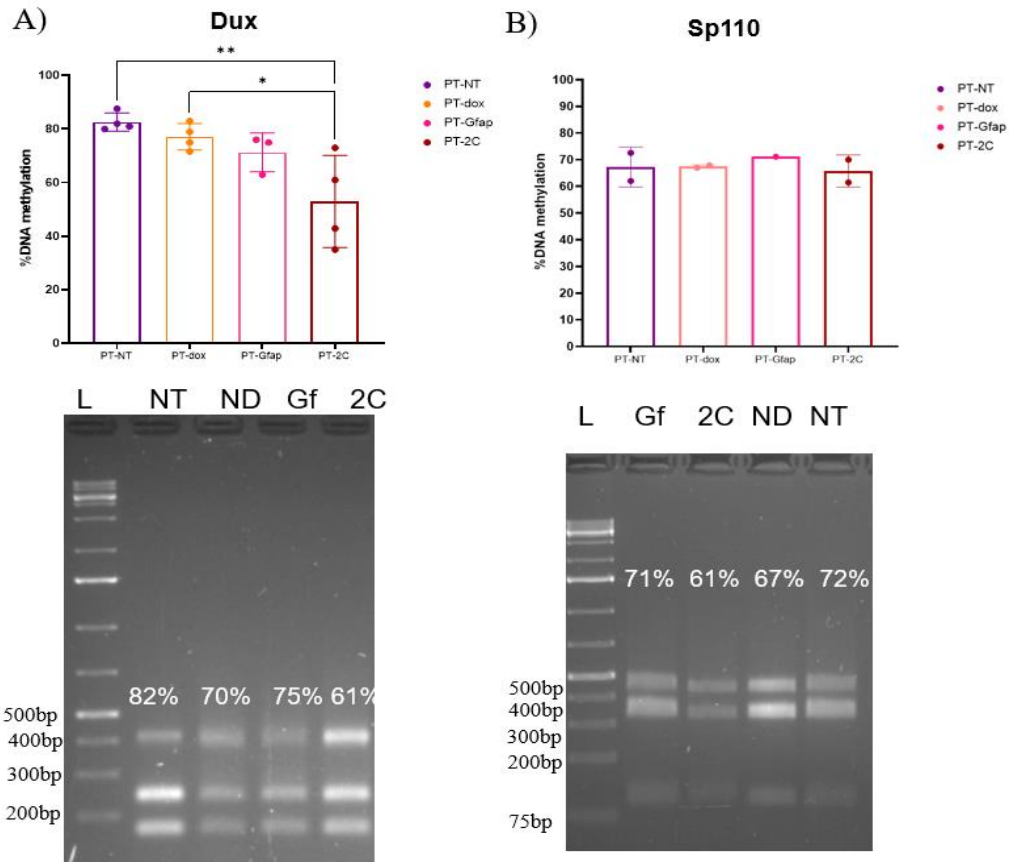
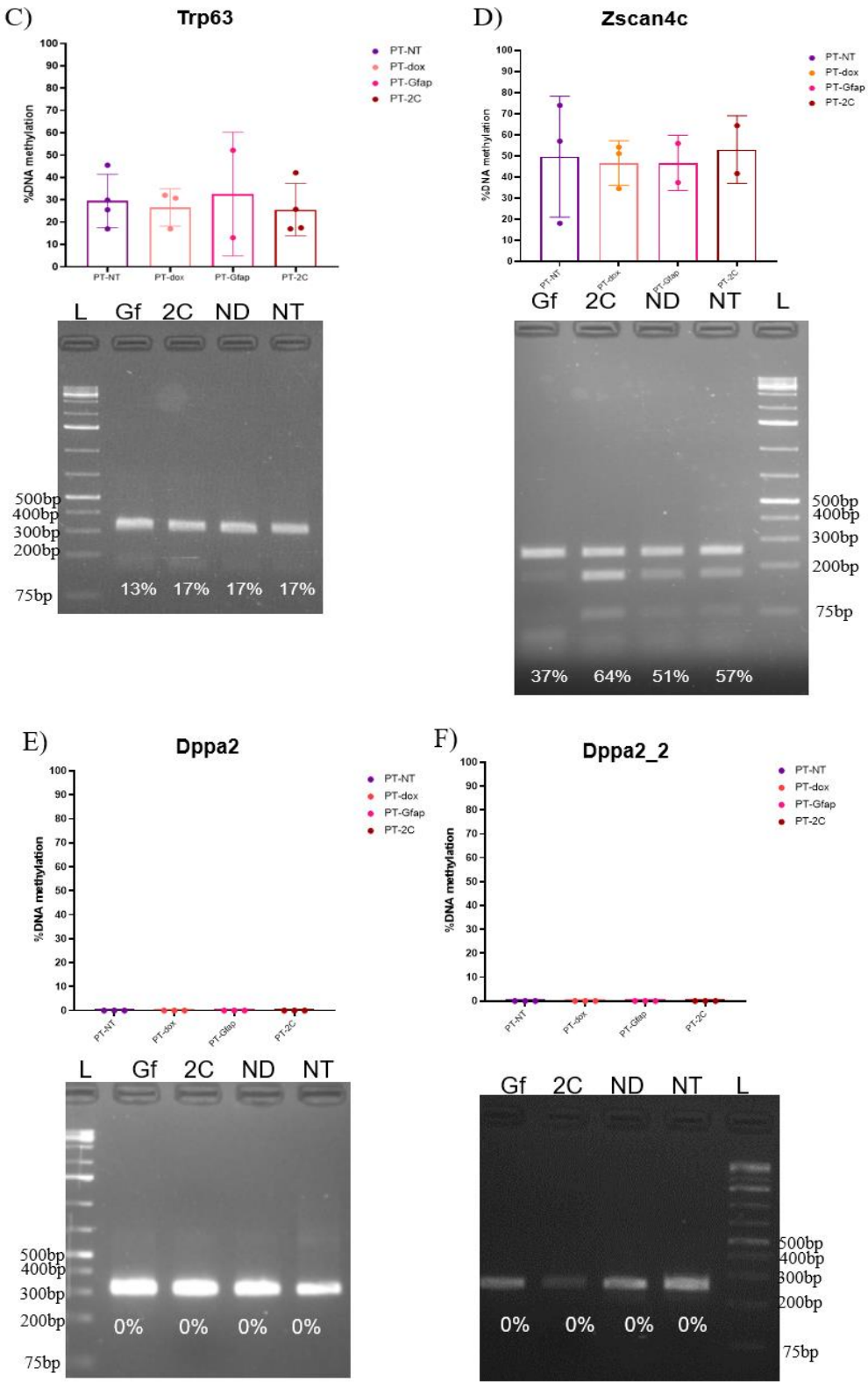


Figure 29: DNA methylation editing of the 2C-CARGO targets in mESCs. *In purple: PT non-transfected cells (PT-NT); orange: PT-NT with Dox; pink: PT cells transfected with a Gfap-CARGO; dark-red: PT cells transfected with a 2C-CARGO and analyzed for DNA methylation changes on day three by COBRA. Data shown are mean and STD, each technical replicate is represented with a dot. P value > 0.05 was determined as not significant (ns), P value ≤ 0.05 was marked with *, P value ≤ 0.01 was marked with **. Figure created using GraphPad Prism*




C)
Trp63
%DNA methylation
PT-NT
PT-dox
PT-Gfap
PT-2C
L Gf 2C ND NT
500bp
400bp
300bp
200bp
75bp
13% 17% 17% 17%
D)
Zscan4c
%DNA methylation
PT-NT
PT-dox
PT-Gfap
PT-2C
Gf 2C ND NT L
500bp
400bp
300bp
200bp
75bp
37% 64% 51% 57%
E)
Dppa2
%DNA methylation
PT-NT
PT-dox
PT-Gfap
PT-2C
L Gf 2C ND NT
500bp
400bp
300bp
200bp
75bp
0% 0% 0% 0%
F)
Dppa2_2
%DNA methylation
PT-NT
PT-dox
PT-Gfap
PT-2C
Gf 2C ND NT L
500bp
400bp
300bp
200bp
75bp
0% 0% 0% 0%

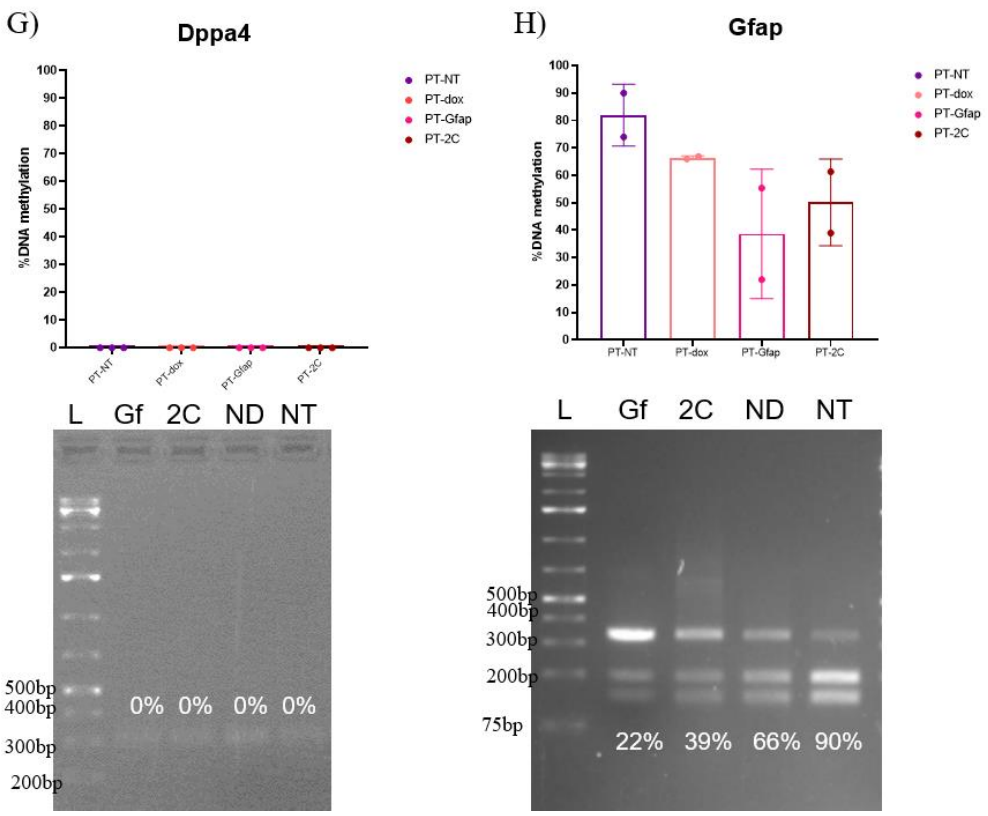


Figure 30: DNA methylation editing of the 2C-CARGO targets. *A-H) 2C-CARGO and Gfap-CARGO targets. On top a column-graph showing the methylation levels under different conditions. In purple: PT non-transfected cells (PT-NT); orange: PT-NT with Dox; pink: PT cells transfected with Gfap-CARGO; dark-red: PT cells transfected with 2C-CARGO. Data shown are mean and STD, each technical replicate is represented with a dot. P value > 0.05 was determined as not significant (ns), P value ≤ 0.05 was marked with *, P value ≤ 0.01 was marked with **, P value ≤ 0.001 was marked with **, P value ≤ 0,0001 was marked with ****. Graphs were created with GraphPad Prism. Underneath each graph an agarose gel of the performed COBRA analysis is shown. A) Expected bands for Dux: uncut: 466bp, cut:222bp and 144bp. B) Expected bands for Sp110: uncut:475bp, cut:404bp, 171bp; C) Expected bands for Trp63: Uncut: 347, cut:28, 229; D) Expected bands forZscan4: uncut: 246, cut 176 and 70; E) Expected bands for Dppa2: uncut 355, cut: 299 and 156;F) Expected bands for Dppa2_2: uncut301 cut 252 and 49; G) Expected bands for Dppa4: uncut: 315, cut: 229 and 86; H) Expected bands for Gfap: uncut: 372, cut:160 and 112.*

The off-target control – the Gfap-CARGO, demonstrated results close to those obtained with the experiments described in 4.5.2 and 4.5.3 (Figure 30H) suggesting that the poor editing efficiency of most of the 2C-CARGO targets was due to either inefficient expression of the gRNAs or to inefficient gRNA design.

Overall, the results show that different loci display different editing susceptibility with inefficient editing on most gRNAs. Furthermore, it is important to note that there is a high heterogeneity between the results from different experiments, raging to up to 60% as observed on

the Zscan4c locus (Figure 29) suggesting that more experiments could shed more light on the efficiency of the system.

4.5.5 Testing the 2C-CARGO in MEFs

This master thesis was aiming to test the ability of the multi-guided CARGO system to influence cell fate. MEFs were co-transfected with the 2C-CARGO and an All-in-one-ng-TET-tool plasmid (Figure 31). Cells were sorted for the combined expression of GFP and mCherry on day one and analyzed for changes in DNA methylation by COBRA on day three. The average transfection efficiency of the TET-tool together with the 2C-CARGO was 10 % and 19.000 MEFs were sorted on average. No changes in morphology were detected three days post transfection, which was expected considering that reprograming of somatic cells by Yamanaka factors takes up to 14 days (Yamanaka *et al*., 2006; Hernandez *et al*., 2018). Unfortunately, COBRA analysis was not successful for all tested loci in MEFs, due to difficulties in collecting enough DNA for the analysis. Data was obtained for the Dux and Trp63 loci only (Figure 32, Figure 33) and no significant decrease in methylation levels on the targeted loci in comparison with the untransfected cells were detected (Figure 32). According to the publicly available methylation data for MEFs, the analyzed Dux locus is expected to be 50% methylated and Trp63 80% methylation have been reported. However, fluctuations of up to 60% were observed between the different replicates on both loci leading to lower average methylation levels (30% on Dux and 50% on the Trp63 locus). (Figure 32, Figure 33). Due to time constraints, the experiments could not be repeated in order to obtain more replicas.

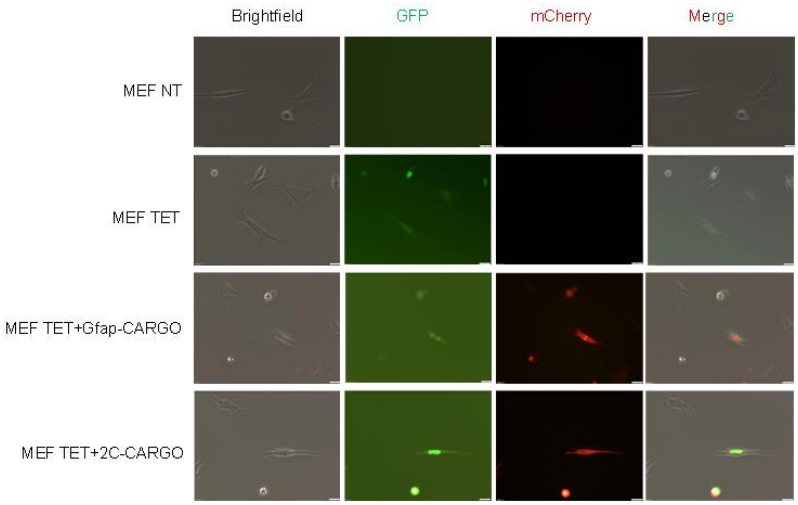


Figure 31: MEFs 24h post transfection: *top to bottom non-transfected cells (MEF NT), MEFs transfected with the All-in-one-ng-TET-tool (MEF TET), MEFs co-transfected with the All-in-one-ng-TET-tool and Gfap-CARGO, MEFs co-transfected with the All-in-one-ng-TET-tool and 2C-CARGO. Left to right: brightfield, GFP, mCherry and a merged image. All images were taken with a 40x objective and 500ms exposure time. Scale bar = 20µm*

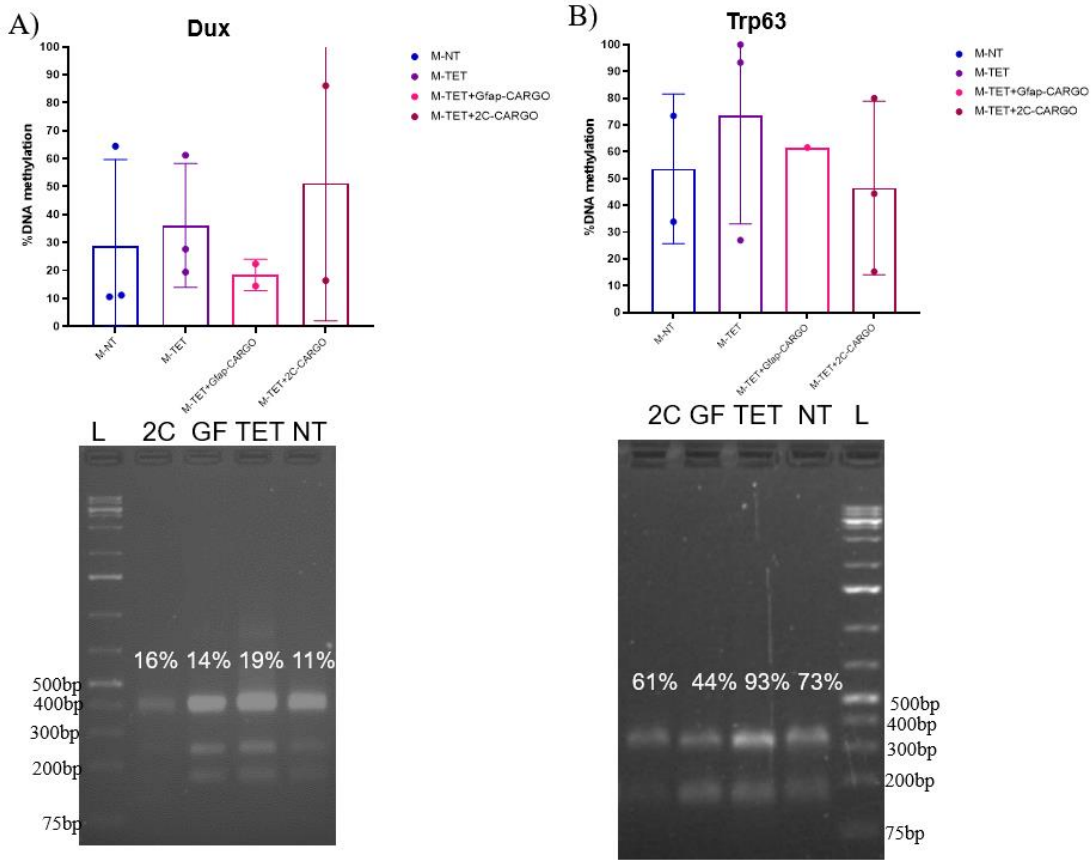


Figure 32: DNA methylation of the 2C-CARGO targets Dux and Trp63 in MEFs three days after induction of EE. *A-B) Dux and Trp63. On top a column-graph showing the methylation levels under different conditions. In blue WT-MEFs (M-NT); purple: MEFs transfected with the All-in-one-ng-TET-tool (M-TET); pink: MEFs co-transfected with the All-in-one-ng-TET-tool and Gfap-CARGO; dark red: MEFs co-transfected with the All-in-one-ng-TET-tool and 2C-CARGO. Data shown are mean and STD, each technical replicate is represented with a dot. Graphs were created using GraphPad Prism. Underneath each graph an agarose gel of the performed COBRA analysis is shown. A) Expected bands for Dux: uncut: 466bp, cut:222 and 144. B) Expected bands for Trp63: Uncut: 347, cut:28, 229.*

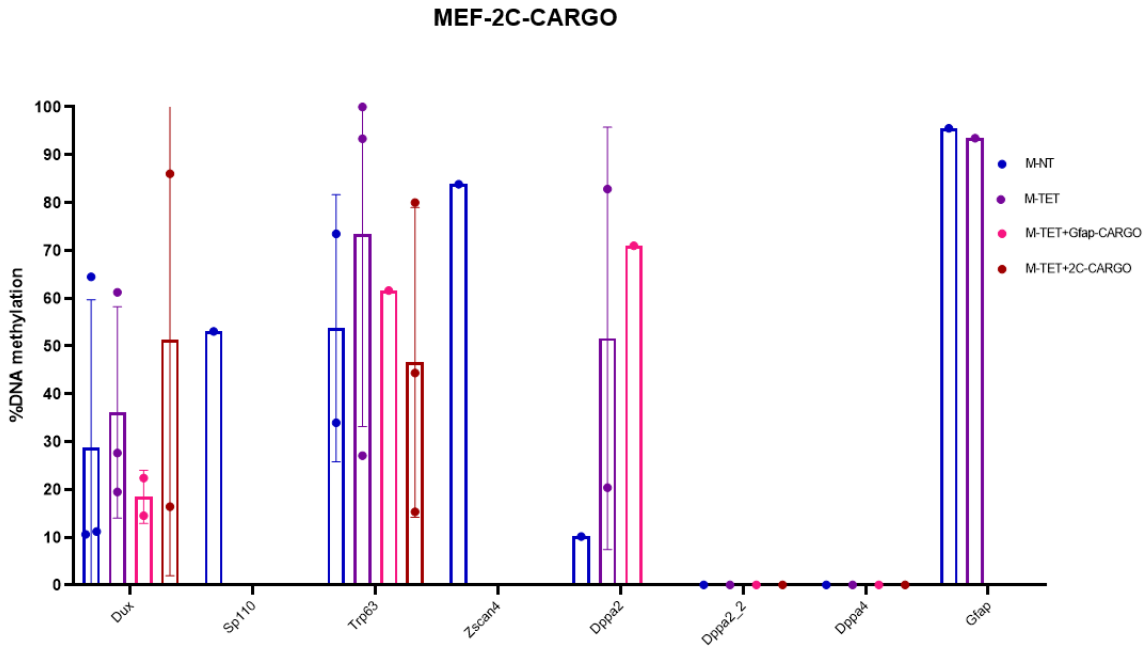


Figure 33: DNA methylation editing of the 2C-CARGO targets on MEFs. after three days of EE *In blue WT-MEF cells (M-NT), purple: MEFS transfected with the All-in-one-ng-TET-tool (M-TET); pink: MEFs co- transfected with the All-in-one-ng-TET-tool and Gfap-CARGO; dark-red: MEFs co-transfected with the All-in-one-ng-TET-tool 2C-CARGO. Data shown are mean and STD, each technical replicate is represented with a dot. Figure created using GraphPad Prism.*

4.6 MANIPULATION OF THE 2C POPULATION

The final goal of this thesis was to test whether manipulation of the cellular epigenome could lead to changes in the 2CLC-population. First, the 2CLC-population within a non-transfected (NT) ESCs culture was followed for seven days to analyze the reported fluctuations of 2CLCs in mESCs. The observed data was consistent with the literature, reporting 2CLCs fluctuating between 1-5% (Figure 34) (Macfarlan *et al*., 2012).

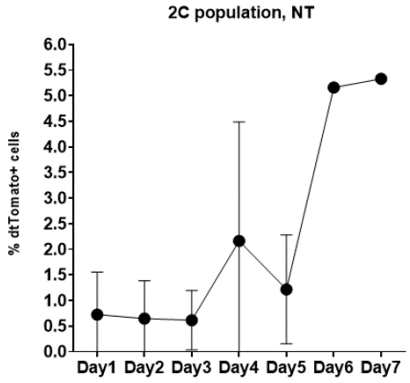


Figure 34: Fluctuation of the 2CLCs population within a mESCs culture*. The fluctuation of the 2CLCs population was followed via flow cytometry over the course of seven days. Cells were split every day and the percentage of the tdTomato positive (activation of MERVL-reporter = 2CLC) cells was calculated. The x-axis shows days of observation. On the y-axis the percentage of 2CLCs detected. Data shown are mean ±STD. Each time point had four technical replicates. Figure created using GraphPad Prism*

Next, it was examined whether the 2C-CARGO system can increase of the percentage of 2CLCs. Therefore, PiggyTET cells were transfected with the 2C-CARGO and Dox was added to the media (Figure 35). Over the course of five days mCherry expressing cells were counted using flow cytometry to determine the amount of 2CLCs (Figure 36, Figure 37). To control whether the change in the percentage of 2CLCs within the population was only due to transfection stress, Gfap-CARGO transfected cells were analyzed. Non-transfected PiggyTET cells under Dox treatment served to monitor whether the activation of the TET system alone might change the percentage of 2CLCs. Additionally, the percentage of 2CLCs in PiggyTETs to WT cells was compared. To locate the correct population of 2CLCs, the samples were sorted first for single cells, then double singlets. Out of those GFP-positive cells were selected (cells expressing the TET-tool). From the GFP-positive cells, mCherry-positive cells were sorted (cells that express the tool and have received a gRNA). Finally, the tdTomato positive population (activation of MERVL-reporter) was determined (Figure 37).

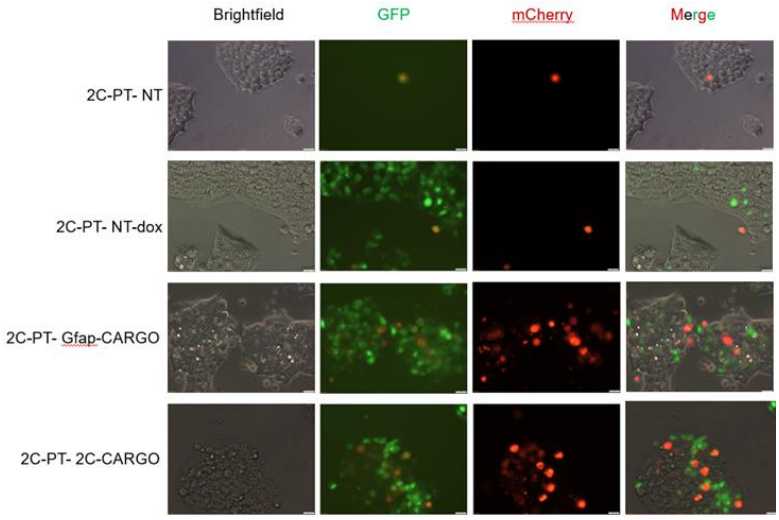


Figure 35: 2C-PiggyTET cells 24h post transfection and Dox activation: *top to bottom untransfected cells (2C-PT-NT), untransfected with Dox in the media (2C-PT-NT-dox), transfected with the Gfap-CARGO and the 2C-CARGO. Left to right: brightfield, GFP, mCherry, and a merged image. All images were taken with a 40x objective and 500ms exposure time. Scale bar = 20µm.*

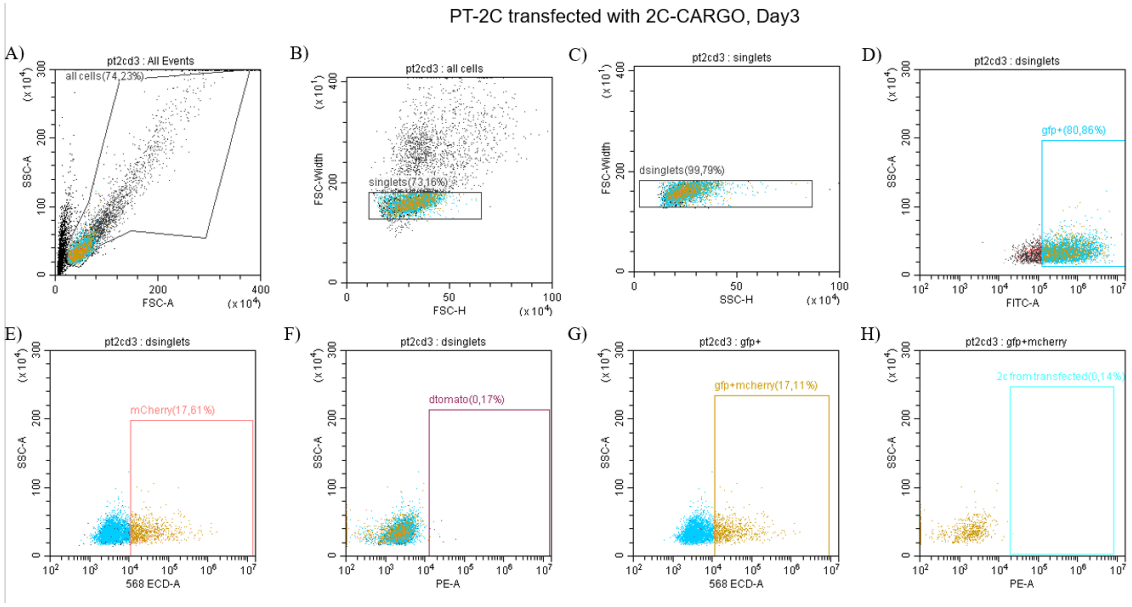


Figure 36: Flow cytometry of 2C-PiggyTET cells on day three upon transfection with 2C-CARGO. *A) Using forward (FSC) and side (SSC) scatter the living cell population was selected. B) Out of the living cells FSC-width and height were used to select single cells. C) For more stringency, FSC-width and SSC-height was used to select double single cells. D-F) Out of the double single cells FITC-, ECD-, and PE- lasers were used to select for GFP, mCherry and tdTomato positive cells, respectively. G) GFP and mCherry positive cells. H) Triple positive cells.*

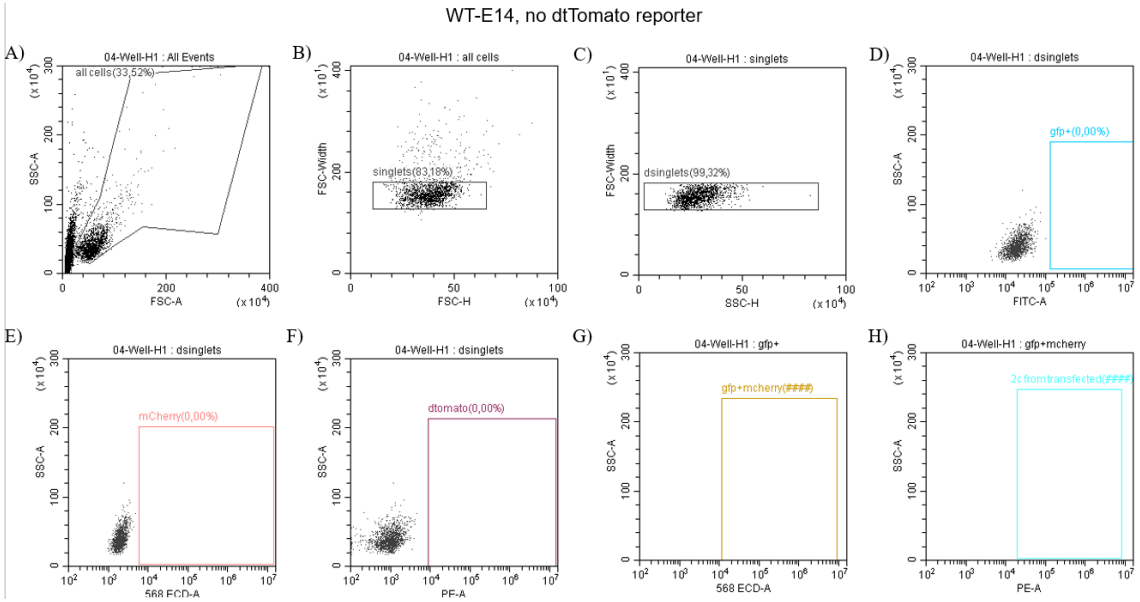


Figure 37: Flow cytometry of WT-E14 cells expressing no reporter to determine the correct setting of the gates. *A) Using forward (FSC) and side(SSC) scatter the living cells population was selected. B) Out of the living cells FSC-width and height were used to select single cells. C)For more stringency FSC-width and SSC-height was used to select double single cells. D-F)Out of the double single cells FITC-, ECD-, and PE-lasers were used to select for GFP, mCherry and tdTomato positive cells, respectively. G) GFP and mCherry positive cells. H) Triple positive cells.*

The characteristic fluctuation of the 2CLCs population was evident under all conditions (Figure 38). No significant difference in the percentage of 2CLCs was observed between the WT, PiggyTETs or PiggyTETs on Dox, indicating that the PiggyTET system is not affecting 2CLCs population. Some increase in tdTomato positive cells was detected on days one and two in the transfected cells, however since the same changes occur both in the Gfap- and 2C-CARGO transfected cells, they are likely a transfection-induced stress response (Figure 38).

One final experiment was conducted to follow more precisely the 2CLCs population of cells both expressing the TET system and having received the 2C-CARGO. Prior to flow cytometry counting, cells were double sorted for GFP and mCherry expression 24h post transfection. Consistent with the previous findings, no significant diffeence was observed between the different conditions, indicating that there was either no overexpression of the desired genes, or their overexpression didn't lead to an incerase in the 2CLC- population (Figure 39).

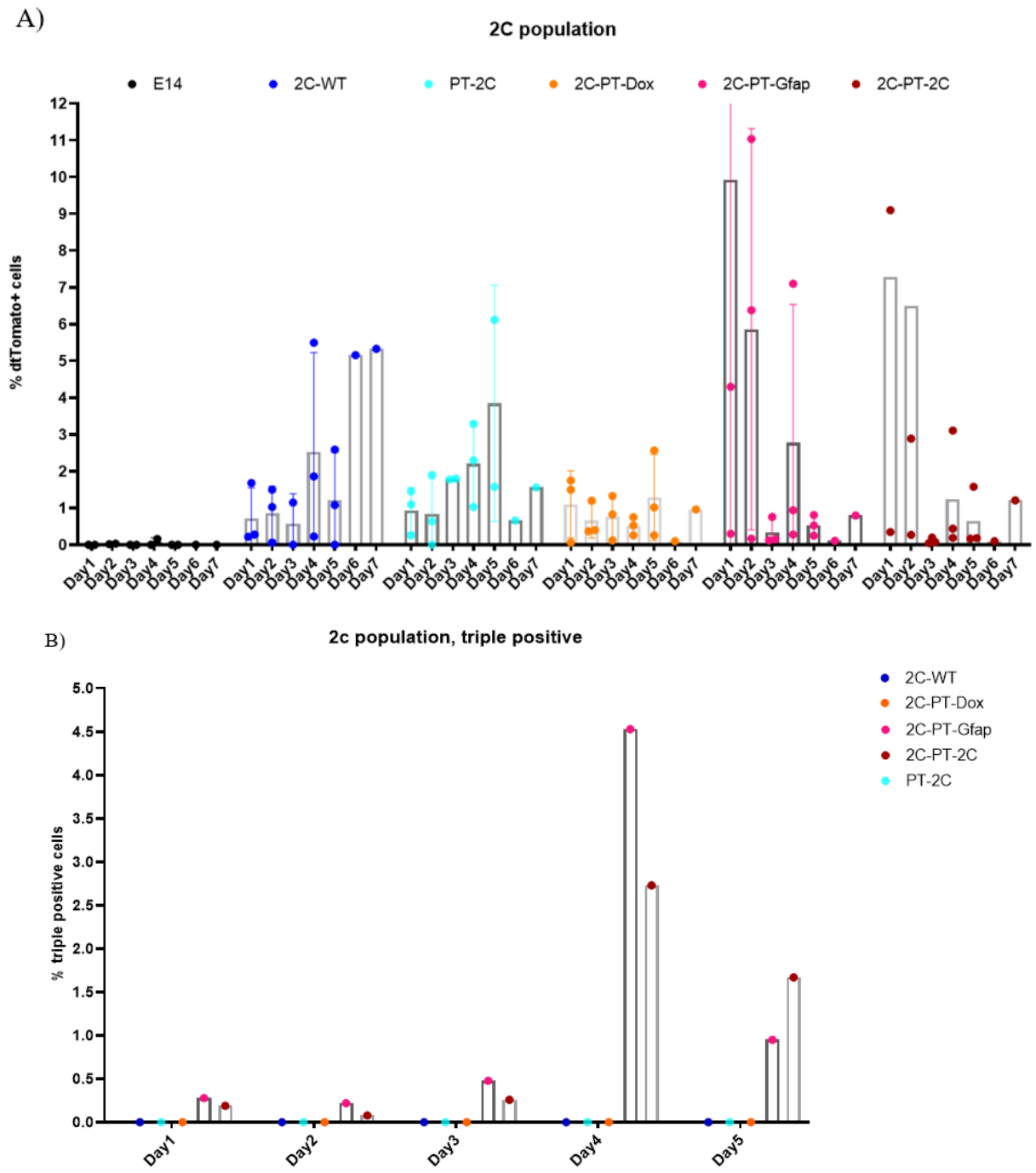


Figure 38: Following the 2CLC-population in transfected and untransfected cells for five days. *2C-WT (dark blue): wild type mESCs with a tdTomato reporter; PT-2C (blue): mESCs with a tdTomato reporter and a PiggyTET-tool; 2C-PT-Dox (orange): PT-2C with Dox in the media; 2C-PT-Gfap (pink): PT-2C with Dox in the media, transfected with Gfap-CARGO gRNA; 2C-PT-2C (dark red): PT-2C with Dox in the media, transfected with 2C-CARGO gRNA. A) Percentage of tdTomato positive cells (2CLCs); B) Percentage of triple positive cells. The x-axis shows days of observation. On the y-axis the percentage of 2CLCs is detected. Only transfected cells (2C-PTpGfap and 2C-PT-2C) were expected to be triple positive. Data shown are mean ±STD. Technical replicates are represented as points. Figure created using GraphPad Prism*

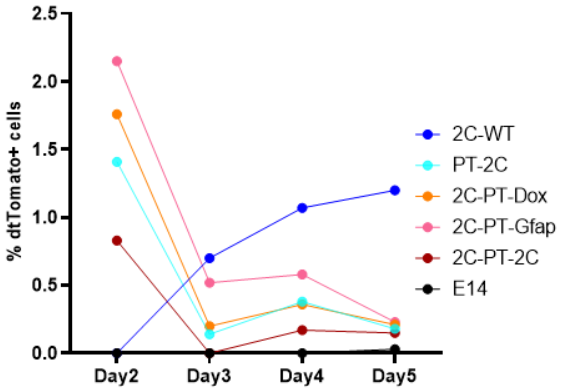


Figure 39: Fluctuation of the 2CLC-population of cells sorted for GFP and mCherry. *The percentage of tdTomato positive cells (activation of MERVL-reporter = 2CLCs) was followed over the course of five days. E14 (black): wild type mESCs; 2C-WT (dark blue): wild type mESCs with a tdTomato reporter; PT-2C (blue): mESCs with a tdTomato reporter and a PiggyTET-tool; 2C-PT-Dox (orange): PT-2C with Dox in the media; 2C-PT-Gfap (pink): PT-2C with Dox in the media, transfected with Gfap-CARGO gRNA; 2C-PT-2C (dark red): PT-2C with Dox in the media, transfected with 2C-CARGO gRNA. The x-axis shows days of observation. On the y-axis the percentage of 2CLCs is detected. Data shown are mean ±STD. Each time point had one technical replicate. Figure created using GraphPad Prism.*

5 DISCUSSION

DNA methylation is an important epigenetic modification that dynamically changes during developmental progression. Nevertheless, the role of DNA methylation reprogramming in early development and cellular potency is not fully understood, since until recently easy-to-use tools for manipulating DNA methylation of specific genomic loci were lacking. With the discovery of CRISPR/Cas9 and further utilizing a catalytically inactive dCas9 the targeted manipulation of DNA methylation has become possible. Previously in the Wossidlo laboratory, an All-in-one-tool combining a dCas9-SunTag system fused to DNA methylation modifiers and a gRNA in a single expression vector was established. However, such a system is only suitable for transient expression of the EE-system. Therefore, during this master thesis an inducible EE-tool was generated to allow monitoring the effects of DNA methylation reprograming over an extended period. The selected system integrates all components of the All-in-one-tool, except of the gRNA, in a PB transposon vector. This Dox inducible tool was tested for its potential to manipulate DNA methylation of selected loci with a single gRNA. Since the new tool requires separate gRNA delivery, it provided the opportunity to test multiplexed targeting of several candidate loci. Changes in cellular potency are orchestrated by expression changes in a multitude of genes, thus an important aspect of this thesis was to select a combination of genes with the potential to interfere with cellular potency. Following the CARGO protocol by Gu *et al.*, 2018 the selected candidate targets were combined

into a single gRNA-expression vector (Gu *et al*., 2018). Finally, the effect of this system on cellular potency was examined, although, due to time-constrains, further experiments are needed to draw meaningful conclusions. The data generated during this thesis can be used to further improve platforms allowing to study the role of DNA methylation in establishing cellular potency.

As a feasible model organism to study cellular plasticity the 2C-like state was selected, since it permits examining the transition from pluripotency to totipotency and enables high-throughput analysis not possible on actual two-cell embryos, as obtaining large quantities of embryos is difficult. 2CLCs are characterized by the expression of 2-cell stage genes like MERVL, and the repression of pluripotency genes like Oct4. In this master thesis, 2CLCs were used to test important features of totipotency, such as whether by changing the DNA methylation status of promoter areas of key genes, the differentiation state of cells could be reprogramed leading to a higher potent state. The quest for reprograming via manipulation of DNA methylation was taken one step further, by attempting to reprogram unipotent MEFs to a 2C-like state. It has been reported that Dppa2/4 are supporting efficient reprograming of somatic cells (Hernandez *et al*., 2019). A knockout of DNMT3b in embryos has been shown to provoke the expression of Dppa2/4, supporting the hypothesis that demethylation of the promoter areas of these genes could induce their gene expression (Eckersley-Maslin *et al*., 2019). Moreover, an analysis by Dr. Julia Arand and Marlenie Musial, showed that Dppa2/4 overexpression correlates with expression of 2C-genes (section 1.6, Figure 6). Therefore, ESCs and MEFs were selected as model cell lines to study the role of DNA methylation in reprogramming cellular potency in the course of this master thesis.

To study the effects of DNA methylation reprograming by EE over a longer time period compared to transient manipulations, a ESC cell line with an inducible EE-tool was created. A PB transposon vector was selected as experimental approach, since it can harbor large DNA fragments and insert them into the cell's genome. As the PB transposase integrates the desired transposon randomly (and potentially multiple times) inside the cell's genome, intact genome integrity was verified by karyotyping. Furthermore, generating cell lines with multiple integrations of the EE-system can reduce the risk of silencing the system by the host cells (Randolph *et al*., 2017) and may thus be seen as advantageous. To achieve a tight control of expressing the EE-system, the modified Tet-On 3G system was employed, due to it's tunability (gene expression depends on the dosage of Dox), and the fast on/off response upon Dox addition/withdraw. Expression of the integrated EE-tools was detected as early as 6h upon Dox administration. Unfortunately, due to

time constraints, the optimal concentration of Dox to achieve best editing results was not tested. The ability to precisely control the expression of the system would be particularly useful in the future to create transgenic mice and study the effects of targeted DNA methylation *in vivo*. Since the system is placed under a minimal CMV promoter potential leakiness was further examined. Analyses showed no difference between the DNA methylation levels of PiggyTETs and WT cells. Hence, it was concluded that the PB-Tet-On 3G system is a tight platform for inducible EE-tools.

A major concern associated with CRISPR editing are the so-called "off-target effects", often resulting in the editing of untargeted sites by unbound modifiers (Pflueger *et al.*, 2019). To address this issue, the EE-tools used in this thesis are equipped with a SunTag system. It utilizes a protein array to recruit multiple copies of the methylation modifier to the targeted locus, thus minimizing off-targets and increasing efficiency. As discussed above, similar DNA methylation levels were detected in PiggyTET and WT cell lines for the Dux locus in the absence of Dox. It is important to note, though, that administration of Dox alone, without gRNA, led to 10% demethylation of off-target loci. Indicating that an active EE-system without gRNA behaves like an overexpression of the epigenetic modifier (Pflueger *et al.*, 2019). It was further assessed whether supplying the EE-system with an unrelated gRNA (Gfap) could reduce off-target effects, which proved to be unsuccessful, however. Therefore, a better approach to reduce potential off-targets could be to manipulate the CD of the utilized epi-editing enzyme to e.g., reduce the binding efficiency to its substrate (Hofacker *et al.*, 2019).

In this context, an improved version of the All-in-one-DNMT-tool was created in collaboration with Prof. Dr. Albert Jeltsch. It harbors a mutation at *A887E* in the DNMT3A-CD rendering it less efficient for DNA binding (Hofacker *et al.*, 2019). Assuming that unguided DNMT will be less efficient in binding and modifying DNA, while at the targeted locus the SunTag system will attract sufficient amount of DNMTs to enable editing. Additionally, a DNMT3A::DNMT3L-CDs fusion was utilized, to increase on-target efficiency (Siddique *et al*., 2013). Unfortunately, due to time limitations, the DNMT-system was not tested in cell lines during this master thesis. If such approach proved to be successful a similar mutation could be introduced into the TET-CD.

As mentioned before, an All-in-one-TET-tool has been previously established in the Wossidlo laboratory. However, results show that this EE-system was only suitable for transient EE and gradually lost effect after up to four days, dependent on the genomic context, when employed in mESCs (data not shown). The inducible PB-TET-tool was created to test prolonged EE;

nonetheless, one limiting factor of this system is the requirement for a gRNA to be delivered via transient transfection and it remained to be tested if such transient gRNA delivery can achieve long-term silencing. To determine the stability of EE, the methylation status of Dux and Gfap in PiggyTET cells transfected with a gRNA was monitored for twelve days. The system retained significant reduction in DNA methylation on the targeted loci in comparison to the untransfected cells for six days, and underwent gradual loss of editing after nine days (Figure 25). Therefore, providing a transiently transfected gRNA does not lead to stable EE of DNA methylation the selected targets. The achieved increase is showing an improved stability of the introduced epimutation compared to transient transfections with the All-in-one-tool and suggests a more stable inherit of the introduced epimutation than the expression of the gRNA itself, which is limited to three days (Figure 15). The return to the initial DNA methylation state could be caused by the surrounding chromatin landscape promoting *de novo* methylation. A subsequent analysis with CHIP-seq for active and repressive histone modifications could be performed in the future to test this hypothesis. Another explanation for the regain of methylation could be that the altered methylation status is harmful for the cells leading to cell death of modified cells and leaving the unmodified cells to proliferate further. This hypothesis could by tested by analyzing the number of dead cells between transfected and untransfected cells. To extend the editing time-window and potentially allow for more stable EE in the future, an inducible cell line, harboring the tool and the gRNA in its genome should be generated to remove the constrain of transient gRNA transfections. However, this would require the time-consuming generation of a specialized cell line for each gRNA of interest. A different method to extend gRNA expression could be via viral transduction of gRNA (Yamamoto *et al*., 2019). Additionally, recent studies have shown that a combination for CRISPR with chromatin regulators like KRAB, can lead to long term silencing (Nuñez *et al*., 2021). Further testing is needed to determine the best approach.

The PB system provides an EE-platform compatible with different gRNAs, thus first, the optimal amount of gRNA in the transfection mix was determined by COBRA and NGS analysis. Interestingly, better results were achieved using lower concentrations of gRNA expression vector. Possibly, with greater amounts of gRNA the targeted locus is blocked by the binding of too many free-floating gRNAs, preventing the editing machinery from accessing the locus. It is furthermore possible that excessive amount of gRNA can induce a metabolic burden on the cells and indirectly influence the activity of the EE-system.

Comparing the methylation levels between transfected and control cells, a significant loss of methylation on the Dux and Gfap loci was detected in the edited cells. The satisfactory results of the EE capability of the PB cell-line when using a single gRNA inspired the utilization of the system's flexibility to simultaneously target multiple genes for "multiplexed EE". As the transition into the 2C-like state conducts the expression of an array of genes (Macfarlan *et al*., 2012), targeting a number of them instead of one at a time could produce better results in changing cellular potency. Following the protocol by Gu *et al.*, 2018, a CARGO vector harboring seven gRNAs for genes, whose overexpression leads to increase of the 2C-population (Eckersley-Maslin *et al*., 2019), was created (2C-CARGO). The selected targets where Zscan4c, Trp63 and Sp110, whose promoter areas are heavily methylated in ESCs and MEFs, and hypomethylated in 2CLCs. As Dux regulates the expression of Zscan4c (De Iaco *et al*., 2017; Yang *et al*., 2020) it was also included in the 2C-CARGO to potentially enhance the induction of a 2C-like state. The next targets, Dppa2 and Dppa4 are expressed in ESCs and their promoters are hypomethylated in this cell type. Nevertheless, they were included in the 2C-CARGO as strong candidates for EE in MEFs, as discussed above. Overall, the design of the 2C-CARGO was optimized for use in ESCs as well as in MEFs in combination with the TET-PB, or in ESCs targeting the Dppa2/4 genes in combination with the DNMT-PB.

At the time this master thesis was conducted, there was no established protocol for gRNA design in the Wossidlo laboratory. Such more sophisticated pipeline was developed at the end of this study by Kristeli Eleftheriou in the laboratory. It combines methylome data from WGBS with ATAC-seq data to look for nucleosome positioning around the targeted area. According to the studies by Gilbert *et al*., 2013 and Qi *et al*., 2013 on CRISPRa, best editing is achieved at nucleosome-deprived regions immediately downstream of the TSS, likely due in part to the contribution of dCas9 directly interfering with early transcriptional initiation or elongation. The gRNA used in the 2C-CARGO in this study were examined according to the new pipeline. All gRNAs, with the exceptions of Gfap and Dppa2_2, were located in nucleosome-poor positions and are thus expected to produce efficient editing.

Prior to gRNA selection, the CARGO-cloning strategy was tested by creating a single guided Gfap-CARGO. NGS and COBRA analysis of ESCs transfected with the Gfap-CARGO and a Dux gRNA, cloned into a pX-mCherry vector, showed similar editing efficiency with both gRNAs. After the promising results of the single guided CARGO, the 2C-CARGO was tested both on ESCs and MEFs. Surprisingly, significant editing was only detected at the Dux locus with

substantial fluctuations between different replicates. As discussed above, the gRNAs were located in areas with low-nucleosome density. However, since only Dux, which has proved its efficiency in past experiments, showed efficient editing in the 2C-CARGO system, obtained data suggest that a better gRNA *in silico* design is required. It is worth mentioning that gRNA expression was only evaluated as a result from mCherry expression, which is independently expressed by the 2C-CARGO vector. Thus, it is unclear whether all gRNAs in the CARGO-vector were also transcribed, though Gu *et al.* have shown that the CARGO-system is able to provide up to twelve gRNAs for multiplexed labelling of twelve genomic loci (Gu *et a.l*, 2018). Furthermore, recent studies are making use of two neighboring gRNAs for each targeted locus to enhance on-target specificity (Nuñez *et al*., 2021). In the future the separate testing of each gRNA beforehand will help to draw better conclusions regarding the efficiency of multiplexed editing.

Furthermore, the analysis of the 2C-CARGO editing was performed by COBRA. This method examines only a single or few CpG-positions. The NGS data from previous experiments revealed mosaic editing of the different CpGs around the targeted area. Generally, CpGs with initially lower methylation appear to be more susceptible for editing. Likewise, CpGs that are more easily targeted are also more prone for off-target effects. Therefore, by presenting a broader picture, NGS would be a more optimal analysis for evaluating editing efficiencies, since it is possible that with COBRA only poorly-targetable CpGs are analyzed. Nonetheless, neither analysis reveals the chromatin state around the targeted area. Possibly next to non-responsive CpGs there are repressive chromatin marks, that prevent access of the demethylation machinery, or promoting repair of the changes shortly after they have been deposited.

Regarding the multiplexed EE-experiments in MEFs, COBRA analysis was not possible for all targeted candidate loci in MEFs due to difficulties in efficient double transfection of MEFs with the All-in-one-tool and 2C-CARGO, and collecting enough DNA for later analysis. The only loci, on which enough data was obtained were Dux and Trp63. Both showed up to 50% discrepancy in the methylation levels between different replicates (Figure 33) and no tendency for EE upon transfection. Possibly, the discrepancies in the results between the different replicates are due to differences in the handling of the cells. Further improvement of the MEF transfection method and DNA extraction protocol is needed in order to draw meaningful conclusions about this cell type.

The final goal of this master thesis was to test if targeting DNA methylation of selected 2C-genes can manipulate cellular potency by inducing the 2CLC-population in mESCs. ESCs,

harboring a tdTomato-reporter linked to the MERVL retrotransposons, transfected with the CARGO-system were monitored for seven days by flow cytometry. No difference in 2CLC-population between cells transfected with the 2C- and the Gfap-CARGO was detected. Given the COBRA analysis results of 2C-CARGO transfected PiggyTETs, it could be assumed that no DNA methylation changes were induced in order to alter the gene expression of the 2C-genes. Furthermore, demethylation of the promoter area has been associated with increase in gene expression (Eckersley-Maslin *et al*., 2019), however this has not been tested for all of the 2C-CARGO targets. Regrettably, due to time-constraints it was not possible to perform qPCRs on 2C-genes to test this assumption. As discussed earlier the overexpression of the candidate genes targeted by 2C-CARGO has not been tested yet. In the future, in order to determine an optimal cocktail of genes to induce 2CLCs, it would be necessary to achieve efficient editing on all targets first and subsequently to look for gene expression changes in response to editing.

In summary, the work of this master thesis serves as a platform for future improvements of EE of DNA methylation. The newly created cell lines allow to study the effects of long-term EE. A CARGO-vector for the multiplexed targeting of several genes at the same time was created and proved the possibility to assemble a multi-gRNA expression system. Ultimately, this work can serve as a basis for the unravelling of fundamental questions of the function of DNA methylation in early development. Namely, how DNA methylation affects cellular potency *in vitro*. Furthermore, created cell lines can be used for the generation of a transgenic mice harboring inducible EE-tools. Using the ability to conditionally activate the EE-system would allow to follow the role of DNA methylation of candidate loci in early development *in vivo*. Furthermore, the EE-system will allow to study stability and inheritance of epimutations and their role in disease and abnormal development. In the future, epigenetically reprograming cellular potency could help to derive personalized higher potent cell types for regenerative medicine and reproductive biology. As the current methods to create iPSC partly retain their somatic methylome creating problems, such as low capabilities to differentiate into other cell types or the development of cancer (Khoo *et al*., 2020).

6 REFERENCES

Assefnia, Shahin, *et al*. "Trp63 is regulated by STAT5 in mammary tissue and subject to differentiation in cancer." *Endocrine-related cancer* 21.3 (2014): 443.

Aoki, Fugaku, Diane M. Worrad, and Richard M. Schultz. "Regulation of transcriptional activity during the first and second cell cycles in the preimplantation mouse embryo." *Developmental biology* 181.2 (1997): 296-307.

Chan CC, Martin DF, Xu D, Roberge FG. Side effects of rapamycin in the rat. J Ocul Pharmacol Ther. 1995 Summer;11(2):177-81. doi: 10.1089/jop.1995.11.177. PMID: 8564638.

J. Am. Chem. Soc. 2017, 139, 33, 11337–11340. Publication Date:August 8, 2017. https://doi.org/10.1021/jacs.7b06555 Christman, Judith K. "5-Azacytidine and 5-aza-2′-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy." *Oncogene* 21.35 (2002): 5483-5495.

De Iaco, Alberto, *et al*. "DPPA2 and DPPA4 are necessary to establish a 2C-like state in mouse embryonic stem cells." *EMBO reports* 20.5 (2019): e47382.

Dupont C, Armant DR, Brenner CA. Epigenetics: definition, mechanisms and clinical perspective. Semin Reprod Med. 2009 Sep;27(5):351-7. doi: 10.1055/s-0029-1237423. Epub 2009 Aug 26. PMID: 19711245; PMCID: PMC2791696.

Eads, Cindy A., and Peter W. Laird. "Combined bisulfite restriction analysis (COBRA)." *DNA methylation protocols*. Springer, Totowa, NJ, 2002. 71-86.

Eckersley-Maslin MA, Svensson V, Krueger C, Stubbs TM, Giehr P, Krueger F, Miragaia RJ, Kyriakopoulos C, Berrens RV, Milagre I, Walter J, Teichmann SA, Reik W. MERVL/Zscan4 Network Activation Results in Transient Genome-wide DNA Demethylation of mESCs. Cell Rep. 2016 Sep 27;17(1):179-192. doi: 10.1016/j.celrep.2016.08.087. PMID: 27681430; PMCID: PMC5055476.

Eckersley-Maslin, M.A., Alda-Catalinas, C. & Reik, W. Dynamics of the epigenetic landscape during the maternal-to-zygotic transition. *Nat Rev Mol Cell Biol* **19,** 436–450 (2018). https://doi.org/10.1038/s41580-018-0008-z

Eckersley-Maslin M, Alda-Catalinas C, Blotenburg M, Kreibich E, Krueger C, Reik W. Dppa2 and Dppa4 directly regulate the Dux-driven zygotic transcriptional program. Genes Dev. 2019 Feb 1;33(3-4):194-208. doi: 10.1101/gad.321174.118. Epub 2019 Jan 28. PMID: 30692203; PMCID: PMC6362816.

Evans, Martin J., and Matthew H. Kaufman. "Establishment in culture of pluripotential cells from mouse embryos." *nature* 292.5819 (1981): 154-156.

Falco, Geppino, *et al*. "Zscan4: a novel gene expressed exclusively in late 2-cell embryos and embryonic stem cells." *Developmental biology* 307.2 (2007): 539-550.

Gao, Rui, *et al*. "Inhibition of aberrant DNA re-methylation improves post-implantation development of somatic cell nuclear transfer embryos." *Cell Stem Cell* 23.3 (2018): 426-435.

Gilbert, Luke A., *et al*. "CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes." *Cell* 154.2 (2013): 442-451.

Gowher H, Liebert K, Hermann A, Xu G, Jeltsch A. Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. J Biol Chem. 2005 Apr 8;280(14):13341-8. doi: 10.1074/jbc.M413412200. Epub 2005 Jan 24. PMID: 15671018.

Gu, Bo, *et al*. "Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements." *Science* 359.6379 (2018): 1050-1055.

Gupta, Rajat M., and Kiran Musunuru. "Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9." *The Journal of clinical investigation* 124.10 (2014): 4154-4161.

Haggerty, C., Kretzmer, H., Riemenschneider, C. et al. Dnmt1 has de novo activity targeted to transposable elements. Nat Struct Mol Biol 28, 594–603 (2021). https://doi.org/10.1038/s41594-021-00603-8

Hamatani, Toshio, *et al*. "Dynamics of global gene expression changes during mouse preimplantation development." *Developmental cell* 6.1 (2004): 117-131.

Hata K, Okano M, Lei H, Li E. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Development. 2002 Apr;129(8):1983-93. PMID: 11934864.

Hendrickson, Peter G., *et al*. "Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons." *Nature genetics* 49.6 (2017): 925-934.

Hernandez, Charles *et al*. “Dppa2/4 Facilitate Epigenetic Remodeling during Reprogramming to Pluripotency.” *Cell stem cell* vol. 23,3 (2018): 396-411.e8. doi:10.1016/j.stem.2018.08.001

Hirata, Tetsuya, *et al*. "Zscan4 transiently reactivates early embryonic genes during the generation of induced pluripotent stem cells." *Scientific reports* 2.1 (2012): 1-11.

Hofacker, Daniel, *et al*. "Engineering of effector domains for targeted DNA methylation with reduced off-target effects." *International journal of molecular sciences* 21.2 (2020): 502.
Khoo, Tze Sean, *et al*. "Retention of somatic memory associated with cell identity, age and metabolism in induced pluripotent stem (iPS) cells reprogramming." *Stem cell reviews and reports* 16.2 (2020): 251-261.

Leu, Jia-Shiun, *et al*. "Functional domains of SP110 that modulate its transcriptional regulatory function and cellular translocation." *Journal of biomedical science* 25.1 (2018): 34.

Li E, Zhang Y. DNA methylation in mammals. Cold Spring Harb Perspect Biol. 2014 May 1;6(5):a019133. doi: 10.1101/cshperspect.a019133. PMID: 24789823; PMCID: PMC3996472.

Macfarlan TS, Gifford WD, Agarwal S, Driscoll S, Lettieri K, Wang J, Andrews SE, Franco L, Rosenfeld MG, Ren B, Pfaff SL. Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. Genes Dev. 2011 Mar 15;25(6):594-607. doi: 10.1101/gad.2008511. Epub 2011 Feb 28. PMID: 21357675; PMCID: PMC3059833.

Macfarlan, T., Gifford, W., Driscoll, S. *et al*. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487,** 57–63 (2012). https://doi.org/10.1038/nature11244

Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M. & Valen, E. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. Nucleic Acids Res. 42, W401–7 (2014).

Morita, Sumiyo, *et al*. "Targeted DNA demethylation in vivo using dCas9–peptide repeat and scFv–TET1 catalytic domain fusions." *Nature biotechnology* 34.10 (2016): 1060-1065.

Nakazawa, Yozo, *et al*. "Optimization of the PiggyBac transposon system for the sustained genetic modification of human T-lymphocytes." *Journal of immunotherapy (Hagerstown, Md.: 1997)* 32.8 (2009): 826.

Niakan, Kathy K., *et al*. "Human pre-implantation embryo development." *Development* 139.5 (2012): 829-841.

Nichols, Jennifer, and Austin Smith. "Pluripotency in the embryo and in culture." *Cold Spring Harbor perspectives in biology* 4.8 (2012): a008128.

Nuñez, James K., *et al*. "Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing." *Cell* 184.9 (2021): 2503-2519.

Pflueger, Christian, *et al*. "A modular dCas9-SunTag DNMT3A epigenome editing system overcomes pervasive off-target activity of direct fusion dCas9-DNMT3A constructs." *Genome research* 28.8 (2018): 1193-1206.

Portela, A., Esteller, M. Epigenetic modifications and human disease. Nat Biotechnol 28, 1057–1068 (2010). https://doi.org/10.1038/nbt.1685

Qi, Lei S., *et al*. "Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression." *Cell* 152.5 (2013): 1173-1183.

Qomi, Saeed Bozorg, Amir Asghari, and Majid Mojarrad. "An overview of the CRISPR-Based genomic-and epigenome-editing system: function, applications, and challenges." *Advanced biomedical research* 8 (2019).

Randolph, L.N., Bao, X., Zhou, C. *et al.* An all-in-one, Tet-On 3G inducible PiggyBac system for human pluripotent stem cells and derivatives. *Sci Rep* **7,** 1549 (2017). https://doi.org/10.1038/s41598-017-01684-6.

Siddique AN, Nunna S, Rajavelu A, Zhang Y, Jurkowska RZ, Reinhardt R, Rots MG, Ragozin S, Jurkowski TP, Jeltsch A. Targeted methylation and gene silencing of VEGF-A in human cells by using a designed Dnmt3a-Dnmt3L single-chain fusion protein with increased DNA methylation

activity. J Mol Biol. 2013 Feb 8;425(3):479-91. doi: 10.1016/j.jmb.2012.11.038. Epub 2012 Dec 4. PMID: 23220192.

Smith ZD, Meissner A. DNA methylation: roles in mammalian development. Nat Rev Genet. 2013 Mar;14(3):204-20. doi: 10.1038/nrg3354. Epub 2013 Feb 12. PMID: 23400093.

Tagliaferri, Daniela, *et al*. "Retinoic Acid Induces Embryonic Stem Cells (ESCs) Transition to 2 Cell-Like State Through a Coordinated Expression of Dux and Duxbl1." *Frontiers in cell and developmental biology* 7 (2020): 385.

Takahashi, Kazutoshi, and Shinya Yamanaka. "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors." cell 126.4 (2006): 663-676.

Tanenbaum, Marvin E., *et al*. "A protein-tagging system for signal amplification in gene expression and fluorescence imaging." *Cell* 159.3 (2014): 635-646.

Waddington, C. Canalization Of Development And The Inheritance Of Acquired Characters. *Nature* **150,** 563–565 (1942). https://doi.org/10.1038/150563a0

Wang, Haifeng, *et al*. "CRISPR-mediated programmable 3D genome positioning and nuclear organization." *Cell* 175.5 (2018): 1405-1417.

Wossidlo, M., Nakamura, T., Lepikhov, K. *et al.* 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat Commun* **2,** 241 (2011). https://doi.org/10.1038/ncomms1240

Wu, X., Zhang, Y. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat Rev Genet* **18,** 517–534 (2017). https://doi.org/10.1038/nrg.2017.33

Yamamoto, Yusuke, Keizo Tomonaga, and Tomoyuki Honda. "Development of an RNA virus-based episomal vector capable of switching transgene expression." *Frontiers in microbiology* 10 (2019): 2485.

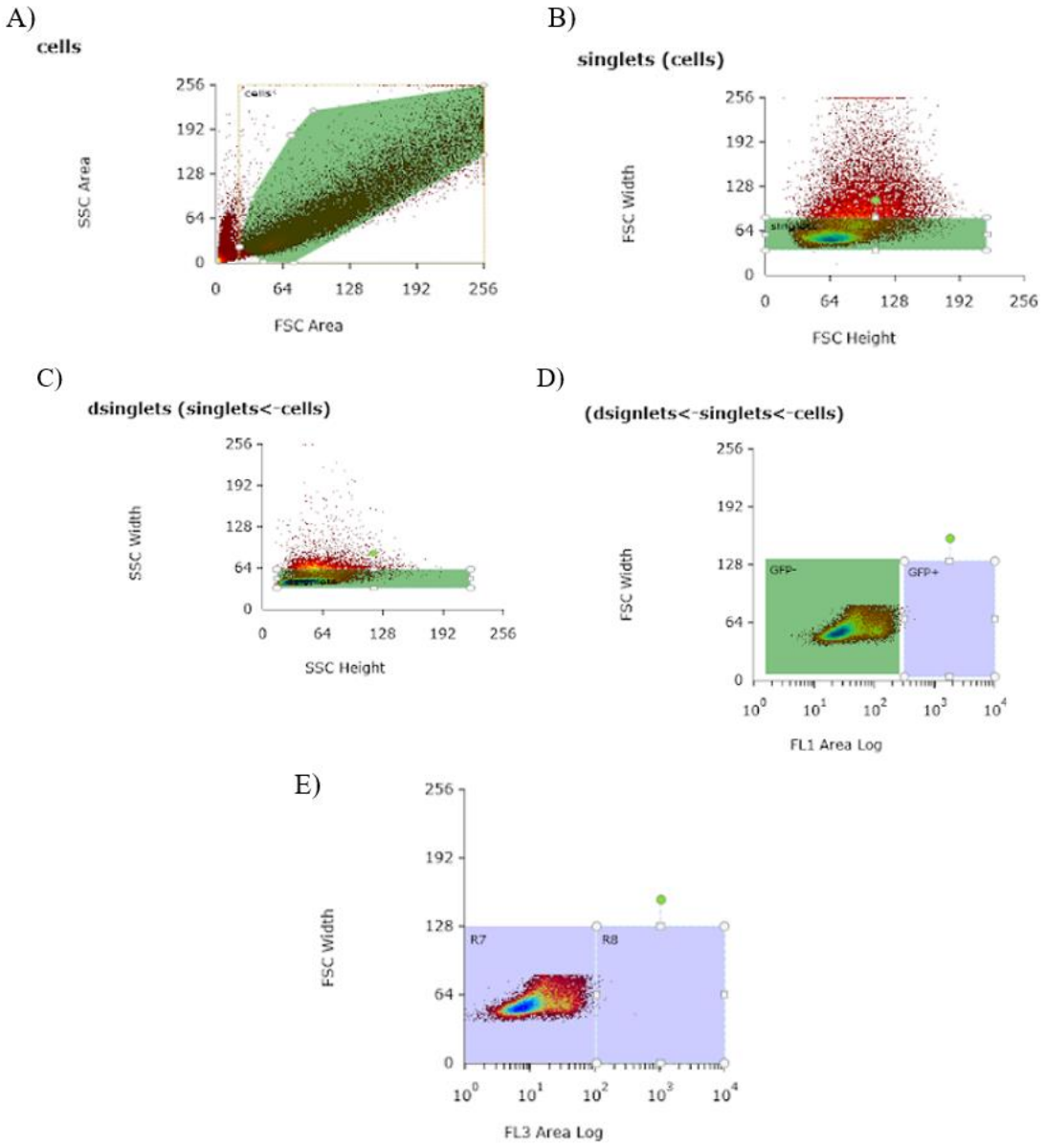
Yang, Guang, *et al*. "Dux-mediated corrections of aberrant H3K9ac during 2-cell genome activation optimize efficiency of somatic cell nuclear transfer." *Cell stem cell* 28.1 (2021): 150-163.

Zhang, Jingfang, *et al*. "Drug inducible CRISPR/Cas systems." *Computational and structural biotechnology journal* 17 (2019): 1171-1177.

Zhou, X., Vink, M., Klaver, B. *et al.* Optimization of the Tet-On system for regulated gene expression through viral evolution. *Gene Ther* **13,** 1382–1390 (2006). https://doi.org/10.1038/sj.gt.3302780

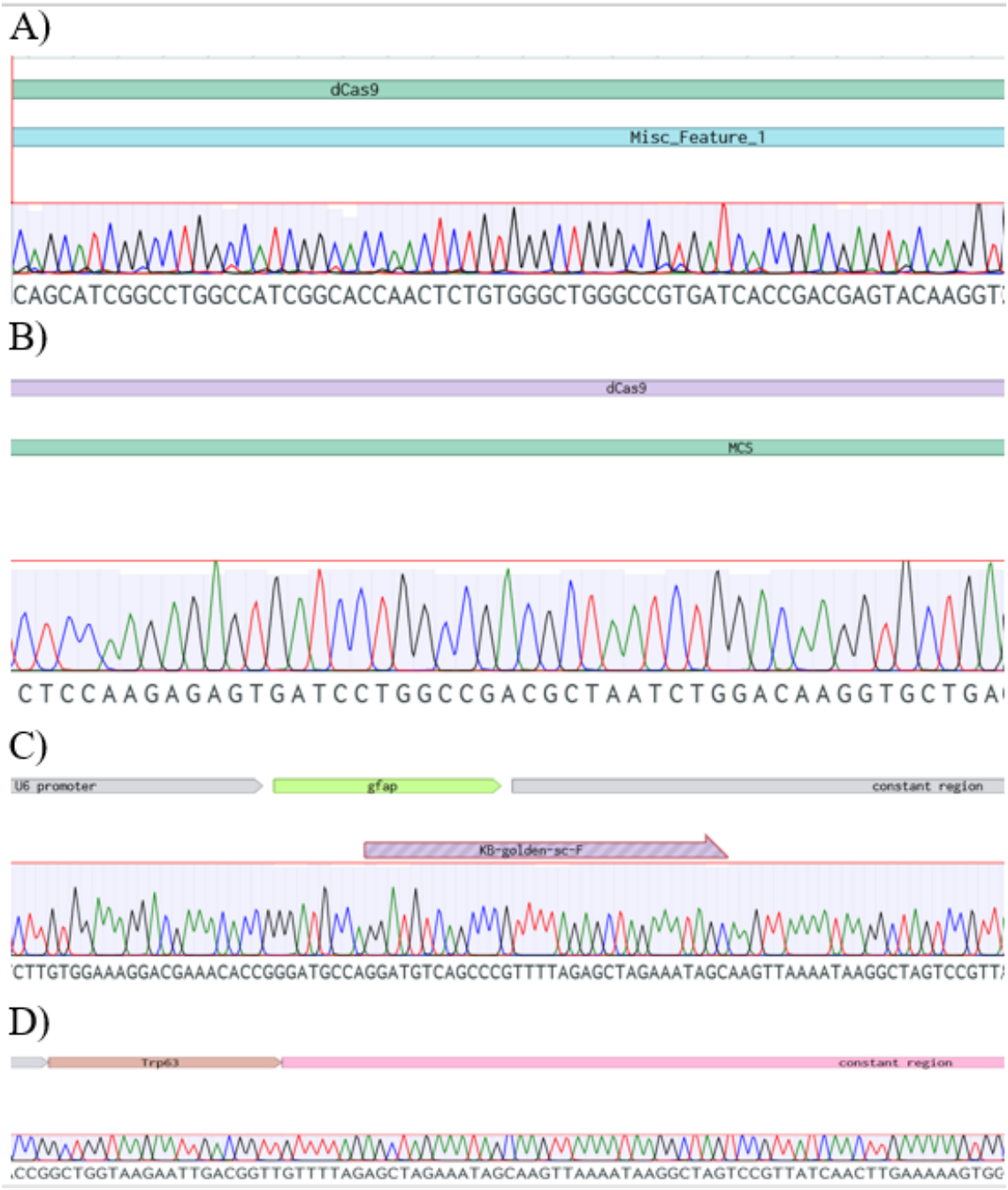
7 ANNEX

7.1 FLUORESCENCE-ACTIVATE CELL SORTING (FACS) STRATEGY



Annex figure 1: Representative image of gate setting strategy for FACS sorting. *PiggyTET cells were transfected with gRNA and Dox was added to the media. 24h post transfection cells were trypsinized and double sorted for GFP and mCherry expression. A) Distinguishing living cells from dead cells using forward and side scatter (FSC and SSC). B) Determining single cells using FSC width and height. C) Double sorting single cells using SSC width and height. D) Selecting GFP positive cells using FSC width and fluorescent protein 1 (GFP). E) Selecting mCherry positive cells using FSC width and fluorescent protein 3 (mCherry).*

7.2 SANGER SEQUENCING



Annex figure 2: Representative images of sanger sequencing results from sequencing *of A) PB-TET showing part of the dCas9; B) PB-DNMT showing part of the dCas9; C) Gfap-CARGO showing the gRNA encompassed by the constant region; D) 2C-CARGO showing part of the plasmid comprising the Trp63 gRNA encompassed by the constant region. After test digestion the selected DNA was sent for sequencing. In the chromatogram Adenine (A) is green, Thymine (T) – red, Cytosine (C) – blue and Guanine – black. Results were analyzed using Benchling.*