Investigation of DNA methylation turnover in pluripotency and early differentiation

Inaugural-Dissertation

to obtain the academic degree

Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry, Pharmacy

of Freie Universität Berlin

by

Leah Haut

Berlin, 2023

The work was carried out between 02/2020 and 08/2023 under the supervision of Prof. Dr. Alexander Meissner at the Max Planck Institute for Molecular Genetics - Department of Genome Regulation in Berlin

Reviewer 1:
Prof. Dr. Alexander Meissner
Max Planck Institute for Molecular Genetics - Department of Genome Regulation
Department of Biology, Chemistry, Pharmacy of Freie Universität Berlin

Reviewer 2:
Prof. Dr. Daniel Schubert
Department of Biology, Chemistry, Pharmacy of Freie Universität Berlin

Date of Defense: 21.12.2023

ACKNOWLEDGEMENTS

First of all, I would like to thank Prof. Dr. Alexander Meissner for giving me the opportunity to do my doctoral thesis in your group. I am very thankful for your trust, patience, guidance, support, and the freedom you gave me to grow both as a scientist and as a person.

Also, I am very thankful for advice, suggestions, and support from Prof. Dr. Daniel Schubert and Dr. Edda Schulz.

A special “thank you” goes out to Philine Guckelberger and Dr. Helene Kretzmer who both played a very important role during this journey. Not just because you are beautiful people, but also because our “turnover to HELLS” meetings were a joy and helped a lot for the development of my project. Every discussion and cup of coffee we shared was appreciated!

I would also like to thank Dr. Adriano Bolondi, who is an amazing mentor and friend. Your scientific advice, support in the lab, and guidance through the writing process were of utmost importance.

The next “thank you” is for Dr. Alexandro Landshammer for being the best early-morning-cell-culture buddy. You taught me how to keep my cells alive and besides that, your humor always made me laugh.

I also want to thank the whole Meissner lab, especially Abhi, Raha, Christina, Anja, Danny, and again Philine, Adri, and Sandro for all your support at the bench but also as friends.

I am also very grateful to Helene, Sara, Stefan, Alexandre, and Pay for your guidance through the computational jungle. Before this project, I never would have imagined to find joy in doing computational analysis, but thanks to you I did.

For the amazing support with various aspects of administration and experiments, another big thanks goes to Cordula, Maria, Sabine, Birgit, Jen, Claudi, Uta, and Erwin.

The Boehringer Ingelheim Fonds also deserves a big thanks, first of all for my fellowship, but also for creating a supportive community, with a special thanks to Anja Petersen.

A heartfelt “thank you” goes out to the beautiful people whom I am happy to call my friends - Sophie, Elodie, Hannah, Sarah, Sena, Lena, Laura, Lilly, Maria, Pia, and Wasim. I am grateful for every one of you, your open warm hearts, your friendly nature, and your beautiful souls.

The biggest gratitude goes out to my parents, Sabine and Steffen, and my sister, Ann-Marie. Your love, patience, and your sympathetic ears were there for me every single day, even across the distance. There are no words to express how thankful I am to have been born into our family!

Declaration of Independence:

Herewith I certify that I have prepared and written my thesis independently and that I have not used any sources and aids other than those indicated by me.

This dissertation has not yet been presented to any other examination authority in the same or a similar form and has not yet been published.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ADD	ATRX-DNMT3-DNMT3L domain
BS	Bisulfite
CGI	CpG island
CpG	Cytosine-phosphate-guanine
DKO	DNMT3A/B double knockout
DNMT1	DNA methyltransferase 1
DNMT3A/B	DNA methyltransferase 3 A/B
hESC	Human embryonic stem cells
hiPSC	Human induced pluripotent stem cells
hPSC	Human pluripotent stem cells
ICR	Imprinting control regions
LINE	Long interspersed nuclear element
mESC	Mouse embryonic stem cells
MTase	Methyltransferase domain
oxBS	oxidation + bisulfite
PGC	Primordial germ cell
PWWP	Pro-Trp-Trp-Pro
SINE	Short interspersed nuclear elements
SVA	SINE-VNRT-Alu
TET	Ten-eleven translocation dioxygenase
TF	Transcription factor
TKO	TET1/2/3 triple knockout
WGBS	Whole genome bisulfite sequencing
5caC	5-carboxylcytosine
5hmC	5-hydroxymethylcytosine
5fC	5-formylcytosine
5mC	5-methylcytosine

LIST OF FIGURES

ZUSAMMENFASSUNG

Während der Embryonalentwicklung durchlaufen Zellen eine Reihe von Zelltyp-Entscheidungen, die von morphologischen und funktionellen Veränderungen begleitet werden und schließlich in der Entstehung eines vollständigen Organismus gipfeln. Dieser komplizierte Prozess wird durch ein komplexes Zusammenspiel verschiedener genetischer und epigenetischer Mechanismen, einschließlich der DNA Methylierung, gesteuert. Nach großen Veränderungen, die das somatische DNA-Methylom im präimplantierten Embryo prägen, bleibt diese Modifikation global stabil, wobei lokale Veränderungen gewebsspezifisch auftreten, oft an mutmaßlichen genetischen Regulationselementen. In humanen pluripotenten Stammzellen (hPSCs) werden jedoch tausende hoch methylierte Regionen von DNA-Demethylasen (TETs) anvisiert, deren dortige lokale Demethylierungsaktivität durch De-novo-Methyltransferasen (DNMT3s) ausgeglichen wird, was zu einem empfindlichen Gleichgewicht führt, das als DNA-Methylierungs-Turnover bezeichnet wird. Der molekulare Mechanismus und seine funktionelle Rolle während der Pluripotenz und der fortschreitenden Entwicklung sind ungeklärt. In meiner Doktorarbeit habe ich experimentelle und analytische Ansätze kombiniert, um das Auftreten und die Regulierung des DNA-Methylierungs-Turnovers während der Pluripotenz und früher Zelldifferenzierung zu untersuchen. Ich konnte zeigen, dass dieser dynamische Mechanismus erheblich an Regionen auftritt, die während der in-vitro Differenzierung der drei Keimblätter demethyliert werden, dass er aber auch an genomischen Loci aktiv ist, die mit späteren, reiferen Zelltyp-Entscheidungen verknüpft sind. Zum ersten Mal habe ich die de-novo Etablierung des DNA-Methylierungs-Turnovers in transienten Vorläuferpopulationen beschrieben, was auf eine erweiterte regulative Rolle des DNA-Methylierungs-Turnovers über die Pluripotenz hinaus hinweisen könnte. Darüber hinaus liefere ich funktionelle Beweise dafür, dass Regionen, die während der Pluripotenz mit dem DNA-Methylierungs-Turnover assoziiert sind, in differenzierten Zellen eine Enhancer-Aktivität aufweisen, was auf eine funktionelle regulatorische Rolle des Turnovers hindeuten könnte. In Bezug auf Transposons bestätigen meine Analysen, dass der DNA-Methylierungs-Turnover Ziel-spezifisch ist. Insbesondere zeige ich, dass die evolutionär jungen ERV1 LTR7up1/up2 und die hominoid-spezifischen ERVK LTR5-Hs Unterfamilien der Long Terminal Repeat (LTR) Retrotransposons besonders stark vom DNA-Methylierungs-Turnover in hPSCs betroffen sind. Interessanterweise wurde zuvor gezeigt, dass genau diese Unterfamilien von Pluripotenzfaktoren, einschließlich NANOG, gebunden werden, was einen möglichen Mechanismus darstellt, der die Dynamik während der Plruipotenz steuert. Schließlich habe ich verschiedene genetisch veränderte hPSC-Linien hergestellt, um die funktionelle Rolle von TETs und DNMT3s an Zielregionen des Turnovers zu untersuchen. Somit bietet meine Arbeit ein wertvolles Werkzeug und einen bisher unerforschten Blickwinkel auf die Ziel-spezifische Regulierung des DNA-Methylierungs-Turnovers und unterstreicht dessen potenzielle Rolle für die menschliche Zelldifferenzierung während der Embryonalentwicklung.

ABSTRACT

Throughout embryonic development, cells undergo a series of lineage decisions, accompanied by morphological and functional changes, culminating in the formation of a complete organism. This intricate process is orchestrated by a complex interplay of diverse genetic and epigenetic mechanisms, including DNA methylation. After major changes shaping the somatic DNA methylome in the pre-implantation embryo, this modification remains globally stable, with local alterations occurring in a tissue-specific manner, often associated with putative genetic regulatory elements. However, in human pluripotent stem cells (hPSCs), thousands of highly methylated regions are targeted by DNA demethylases (TETs), whose local demethylation activity is counteracted by *de novo* methyltransferases (DNMT3s), resulting in a delicate balance referred to as DNA methylation turnover. What is the molecular mechanism and its functional role during pluripotency and developmental progression remains elusive. In my doctoral work, I combined experimental and analytical approaches to investigate the emergence and regulation of DNA methylation turnover during human pluripotency and early differentiation. I revealed that this dynamic mechanism substantially occurs at regions that undergo demethylation during *in vitro* three germ-layer differentiation, but that it is also active at genomic loci linked to mature lineage decisions. Importantly, I described the establishment of *de novo* DNA methylation turnover in transient progenitor populations for the first time, suggesting an extended regulative role of the DNA methylation turnover beyond pluripotency. Furthermore, I provide functional evidence that pluripotency-associated DNA methylation turnover regions have enhancer activity in differentiated cells, implying a potential functional regulatory role of the turnover. Regarding transposable elements, my analysis confirms that the DNA methylation turnover is highly target-specific. In particular, I reveal that the evolutionary young ERV1 LTR7up1/2 and the hominoid-specific ERVK LTR5-Hs subfamilies of the long terminal repeat (LTR) retrotransposons are prominently targeted by the DNA methylation turnover in hPSCs. Interestingly, specifically these subfamilies were previously shown to be bound by pluripotency factors, including NANOG, providing a possible underlying mechanism behind the turnover during pluripotency. Lastly, I generated various genetically modified hPSCs lines to experimentally dissect the functional role of TETs and DNMT3s at turnover targets. Thus, my work provides a valuable toolkit and an unexplored analytical angle into the target-specific regulation of DNA methylation turnover, emphasizing its potential role for human cell differentiation during embryonic development.

BACKGROUND

Mammalian genomes display high levels of cytosine methylation in the CpG context

In vertebrate genomes, DNA methylation primarily occurs at cytosines in a CpG sequence context. While CpGs are 4- to 5-times less frequent in vertebrates compared to invertebrates, a significant proportion of vertebrate CpGs are methylated, ranging from 55% to 90%, compared to 0-30% in invertebrates (Swartz et al., 1962; Josse et al., 1961; Bird, 1980). Notably, in the model organism *Drosophila melanogaster*, CpGs are only methylated during early embryogenesis, and in *Caenorhabditis elegans* and *Saccharomyces cerevisiae*, CpGs are entirely unmethylated (Urieli-Shoval et al., 19821; Lyko et al., 2000; Simpson et al., 1986; Capuano et al., 2014). Remarkably, CpG methylation levels in the human and mouse genomes vary from nearly totally methylated to unmethylated (bimodal distribution), depending on the developmental stage and cell type (Greenberg & Bourc'his, 2019). Furthermore, CH methylation is prevalent in vertebrates as well, yet at low levels. Notably, neurons possess the highest CH methylation among all cell types of the major human organ systems (Schultz et al., 2015). Interestingly, methylated cytosines can spontaneously be deaminated, leading to point mutations from cytosine to thymine, providing a potential explanation for the reduced abundance of CpGs in highly methylated mammalian genomes (Holliday & Grigg, 1993). In contrast to vertebrates, the invertebrate *Ciona intestinalis* and flowering plants generally exhibit reduced levels of CpG methylation (Feng et al., 2010). Instead, in flowering plants, cytosines in other sequence contexts are more frequently methylated, specifically CHG and CHH (Feng et al., 2010).

In mammals, elevated levels of CpG methylation are observed in various genomic contexts, including transposons, pericentromeric satellite repeats, imprinted genes, and inactivated X-chromosomes and gene bodies, with a slight enrichment at exons (Greenberg & Bourc'his, 2019; Smith & Meissner, 2013; Feng et al., 2010). However, the generally high levels of CpG methylation in mammalian genomes are interrupted by CpG-dense regions, known as CpG islands (CGIs), which are mostly unmethylated and typically coincide with gene promoters (Bird et al., 1985; Gardiner-Garden & Frommer, 1987). Generally, CpG-density negatively correlates with CpG methylation (Lienert et al., 2011; Illingworth et al., 2008; Weber et al., 2007).

Stability and plasticity of DNA methylation during mammalian embryonic development

During embryonic development, the DNA methylation landscape undergoes local, as well as global changes, while other patterns are maintained consistent over long developmental periods. DNA methyltransferase 1(DNMT1) primarily ensures the inheritance and stability of DNA methylation patterns, with some maintenance activity attributed to the canonical *de novo* methyltransferases DNMT3A and B (**Figure 1**) (Chen et al., 2003; Gruenbaum et al., 1982; Liao et al., 2015). In contrast, the plasticity of DNA methylation arises from enzymatic addition and removal, as well as passive loss of DNA methylation. DNMT3A and B are the canonical *de novo* methyltransferases, responsible for catalyzing the addition of a methyl group to a CpG that lacks methylation on both DNA strands (Okano et al., 1999). Notably, also DNMT1 has been shown to have *de novo* activity in specific genomic contexts (Haggerty et al., 2021; Li et al., 2018). DNA demethylation occurs through passive dilution of methylated cytosines upon replication and subsequent cell division, due to reduced DNMT1 activity, or through active demethylation mechanisms, which involves oxidation catalyzed by ten-eleven translocation (TET1,2,3) dioxygenases (**Figure 1**) (Tahiliani et al., 2009; Ito et al., 2010, 2011; He et al., 2011). TETs iteratively oxidize 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC). Eventually, thymine-DNA glycosylase (TDG) and base-excision repair (BER) re-establish the unmodified cytosine (Maiti & Drohat, 2011).

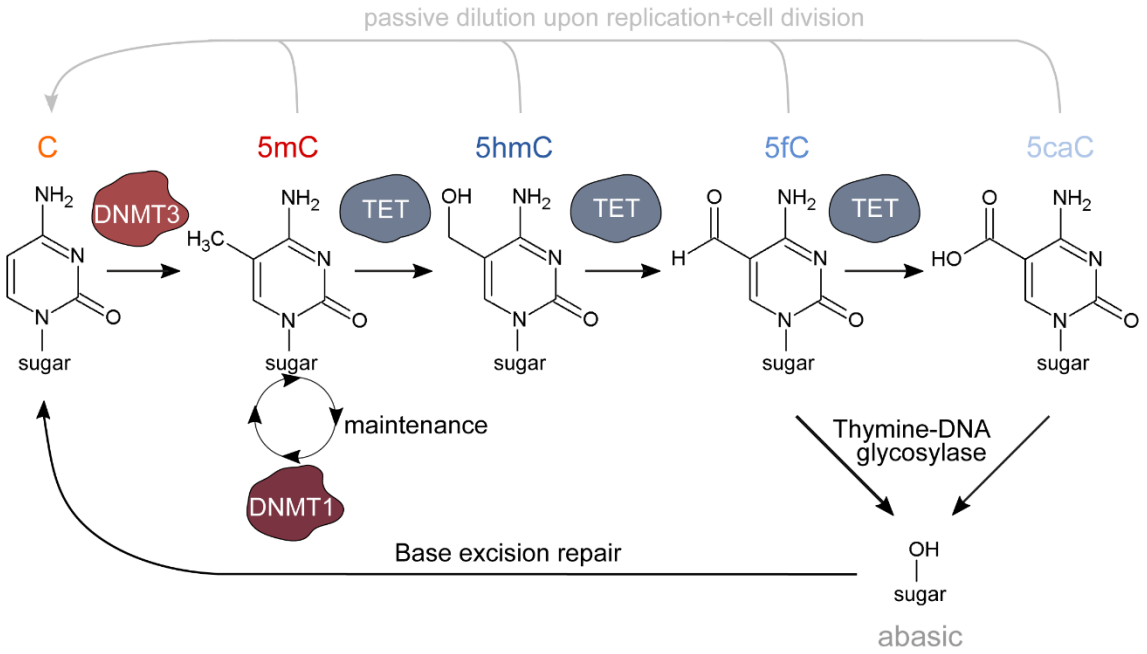

Figure 1: Functionality of DNA methyltransferases and demethylases. The schematic displays the functionality of the canonical maintenance methyltransferase DNMT1, the *de novo* methyltransferases DNMT3A/B and the ten-eleven translocation demethylases TET1-3. While all cytosine modifications can get diluted upon replication and subsequent cell division, a methyl group on a CpG (5mC) can be stepwise oxidized by TETs into hydroxymethylcytosine (5hmC), formylcytosine (5fC), and eventually carboxylcytosine (5caC). 5fC and 5caC can be deaminated by the thymine-DNA glycosylase (TDG) and subsequent base excision repair enables conversion to an unmodified cytosine.

There are two phases of drastic global epigenetic reprogramming during mammalian development, occurring in the germline and in the pre-implantation embryo (**Figure 2**) (Greenberg & Bourc'his, 2019). In the specification of primordial germ cells (PGCs), a two-step demethylation process reduces DNA methylation levels from 70-80% to below 10% (Wang et al., 2014). Initially, the majority of global methylation is passively depleted, likely due to downregulation of UHRF1, a crucial co-factor of DNMT1, followed by subsequent demethylation of imprinting control regions (ICRs) and germline-specific genes by TET1/TET2 activity (Kurimoto et al., 2008). Evolutionarily young retrotransposons, such as human-specific long interspersed nuclear elements (LINEs) and SINE-VNRT-Alu (SVA), tend to retain partial methylation, (Tang et al., 2015a). After PGC demethylation, a phase of re-methylation occurs to establish male or female-specific germline landscapes. Eventually, DNA methylation patterns in sperm resemble somatic cells, reaching ~80%, while oocytes are predominantly methylated in gene bodies with final genomic levels around ~50% (Kobayashi et al., 2012; Wang et al., 2014; Greenberg & Bourc'his, 2019). The oocyte methylome is influenced by the retention of DNMT1 and UHRF1 in the cytoplasm, and a link between *de*

novo DNA methylation and active gene transcription (Veselovska et al., 2015; Li et al., 2018). After fertilization, a second phase of epigenetic reprogramming initiates to remove germ cell-specific DNA methylation patterns in the early embryo. Although limitations exist in studying early human embryo demethylation, studies in mice have revealed active demethylation of both pronuclei by TET3, with a stronger effect on the paternal compared to the maternal pronucleus (Gu et al., 2011; Shen et al., 2014). Additionally, the retention of DNMT1 in the cytoplasm contributes to the erasure of the germ cell methylome in a replication-dependent manner (Carlson et al., 1992). Concomitant to global demethylation, evolutionarily young transposable elements (mainly Alus of the short interspersed nuclear element (SINE) family) are being *de novo* methylated in both parental human genomes (Zhu et al., 2018). Other evolutionarily young transposons, particularly SVAs in humans, as well as certain ICRs, retain gametic methylation in the pre-implantation blastocyst (Wang et al., 2014; Guo et al., 2014). The blastocyst enters re-methylation upon implantation when composed of the inner cell mass, trophectoderm, and primitive endoderm. As development progresses, the trophectoderm and primitive endoderm give rise to the extra-embryonic ectoderm and visceral endoderm respectively, both of which display global hypomethylation due to decreased DNMT3 expression, alongside hypermethylation of specific CGIs (Zhang et al., 2018; Smith et al., 2017). In contrast, the epiblast genome is increasingly methylated, reaching a global level of 70-80%. During the differentiation of somatic cell types, the majority of DNA methylation changes occur focally at regions frequently enriched with transcription factor (TF) binding sites, often associated with predicted enhancers and putative promoters (Schultz et al., 2015; Ziller et al., 2013).

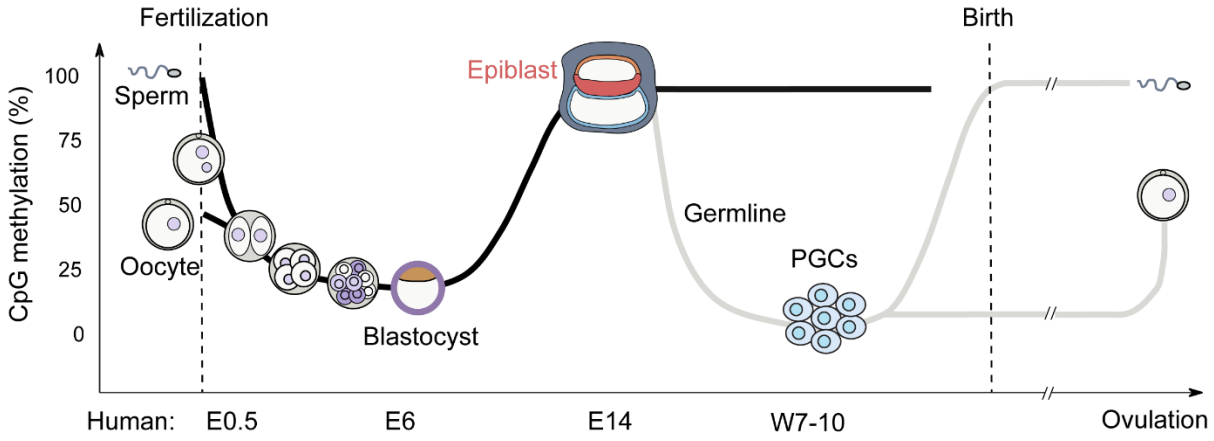

Figure 2: Global CpG methylation levels during human embryonic development. The schematic displays the global loss of CpG methylation after fertilization and the subsequent global re-methylation. While somatic cells maintain highly methylated throughout the rest of development, primordial germ cell (PGC) formation involves a second step of global demethylation. While male germ cells become hypermethylated again, female PGCs only regain DNA methylation (around 50%) upon ovulation. The figure is adapted from (Greenberg & Bourc'his, 2019).

DNA methyltransferases are necessary for mammalian embryonic development

In human pluripotent stem cells (hPSCs), all three DNMTs are active, with the expression of the predominant DNMT3B isoform, DNTM3B1, being notably higher than the levels of DNMT1 and the main DNMT3A variant, DNMT3A2 (Liao et al., 2015). As cells undergo differentiation, they maintain similar levels of DNMT1 and DNMT3A, while expression of DNMT3B is substantially reduced and switches to a catalytically inactive isoform, known as DNMT3B3 (**Figure 4**) (Liao et al., 2015). In mice, DNMT3A and DNMT3B are prominently expressed during pluripotency and are downregulated in embryoid bodies and adult tissues, although their expression is still sustained (Okano et al., 1998).

Depletion of all three DNMTs (DNMT1, DNMT3A/B) leads to cell death in hPSCs. However, a double knockout of the *de novo* methyltransferases only (DKO) does not disrupt viability, expression of pluripotency markers, cellular morphology, or their potential to differentiate into the three germ layers in directed or random embryoid body differentiation(Liao et al., 2015). Nevertheless, over successive cell passaging and despite the continued presence of DNMT1, DKO hPSCs gradually undergo genome-wide loss of CpG methylation, suggesting a role for DNMT3A/B in maintenance methylation (Liao et al., 2015). In contrast to DNMT3A/B, homozygous DNMT1 knockout hPSCs are lethal, underscoring that DNMT1 is indispensable for hPSCs (Liao et al., 2015).

Unlike hPSCs, mouse embryonic stem cells (mESCs), which represent a more naïve state of pluripotency, can survive in the absence of all three DNMTs and maintain self-renewability and chromosome stability (Tsumura et al., 2006). However, DNMT1-knockout mESCs fail to undergo lineage commitment, and inducing differentiation ultimately leads to cell death, suggesting that DNMT1 activity is essential for a more primed state of pluripotency found in hPSCs and somatic cell types, while it is dispensable for a more naïve-like state (Panning & Jaenisch, 1996; Jackson et al., 2004). Furthermore, DKO mESCs fail to form teratomas after 70 cell passages, while they are still capable of doing so within a shorter period after the knockout, implying that the DNA methylation levels, rather than the presence of the enzymes, are crucial for successful differentiation (Chen et al., 2003). Moreover, knockout studies of the three methyltransferases in mouse embryos have demonstrated their essential role for proper embryonic development, organogenesis, and body growth since their depletion leads to developmental delay, increased cell death, and ultimately results in prenatal death. In fact, DNMT1-, DNMT3B- and DNMT3A/B-homozygous knockout embryos are embryonically lethal, while lack of DNMT3A leads to postnatal death (Li et al., 1992; Okano et al., 1999).

DNA methylation defects are also frequently associated with various diseases. Mutations in genes encoding players of the DNA methylation machinery are often found in different types of cancer. For example, mutations in DNMT3A have been observed in acute myeloid leukemia (AML), and hypomorphic DNMT1 mouse mutants have been shown to develop T cell lymphomas (Yan et al., 2011; Walter et al., 2011; Gaudet et al., 2003). Furthermore, various human cancer types display altered DNMT3B isoform expression and global hypomethylation (Ehrlich, 2009; Gopalakrishnan, Van Emburgh, et al., 2009; Ostler et al., 2007; J. Wang et al., 2006; Xie et al., 1999). Additionally, patients with immunodeficiency, centromeric instability, facial anomalies syndrome (ICF) carry mutations in the DNMT3B gene (Okano et al., 1999; Xu et al., 1999; Hansen et al., 1999).

DNA demethylases are necessary for mammalian embryonic development

TET enzymes are differentially expressed throughout development and play essential roles for proper embryogenesis. In hPSCs, all three TET genes- TET1, TET2, and TET3- are expressed (Verma et al., 2018). In contrast, mESCs only express TET1 and TET2 at high and medium levels, respectively, while TET3 expression is notably low (Koh et al., 2011). In mice, upon differentiation, TET1 is downregulated, accompanied by an isoform switch from the full-length to a shortened variant (Zhang et al., 2016; Koh et al., 2011). In humans, an orthologous TET1 isoform has been identified, which is upregulated in various cancer types (Good et al., 2017). TET2 and TET3 fulfill cell type-specific functions (see below).

TETs are responsible for the generation of all 5hmC, which is completely depleted from the genome in their absence (Verma et al., 2018). Furthermore, TET deficiency leads to local hypermethylation, particularly at enhancers and bivalent promoters in hPSCs. Still, hPSCs depleted of all three TETs (TKO) maintain their typical morphology, self-renewal capacity, and expression of the pluripotency marker genes NANOG, OCT4, and SOX2 (Verma et al., 2018; Dawlaty et al., 2014). In contrast, in mESCs, TET1 is crucial for Nanog expression (Ito et al., 2010).

Upon random differentiation into embryoid bodies, TKO-hPSCs exhibit altered expression of lineage-specific genes. This includes the downregulation of genes associated with neuroectodermal differentiation, such as OTX2, PAX6, FOXG1, SOX10, and the endodermal marker FOXA2 (Verma et al., 2018). Additionally, changes in the transcription of the mesodermal markers GSC and Brachyury (T) are observed (Verma et al., 2018). Similarly, embryoid bodies derived from TKO-mESCs display decreased expression of both mesodermal and endodermal markers, suggesting a conserved role for TETs in mammals (Dawlaty et al., 2014). In line with this, human TKO cells cannot form teratomas, and mouse TKO teratomas lack endodermal and certain mesodermal tissues, as well as more advanced ectodermal structures (Dawlaty et al., 2014). By performing knockdown experiments targeting individual TET enzymes in mESCs, distinct and specific roles for each TET enzyme was unraveled (Koh et al., 2011).

Knockdown of TET1 leads to elevated expression of trophectodermal markers (Cdx2, Eomes, Hand1) and of the mesodermal marker Brachyury (T), along with downregulation of neuroectodermal markers (Pax6, Neurod1, Lefty1, Lefty2). Similarly, TET3 knockdown

causes reduced expression of Lefty2. In contrast, neuroectodermal markers are slightly upregulated in the absence of TET2 (Koh et al., 2011).

Furthermore, TET1 knockout mice are viable and fertile (Dawlaty et al., 2011). However, the majority of embryos are developmentally delayed, characterized by fewer somites, reduced body size, and lower weight compared to wildtype embryos of matching age, indicating only a partial compensation by TET2 and TET3 for the loss of TET1 (Dawlaty et al., 2011). Mice with an additional knockout of TET2 are found at a reduced frequency and do not follow the Mendelian ratio (Dawlaty et al., 2013). Most of the born mice die within the first two days after birth, displaying conditions such as exencephaly, head hemorrhage, and growth retardation.

TET2 is prominently expressed in both stem and mature hematopoietic cells, and various mutations in the TET2 gene have been linked to myeloid malignancies (Langemeijer et al., 2009; Jankowska et al., 2009). The knockout of TET2 in mouse embryos results in an augmented population of hematopoietic stem and progenitor cells within the bone marrow and spleen (Ko et al., 2011). This suggests a role for TET2 in limiting self-renewal capacity and proliferation of these cells (Ko et al., 2011). Additionally, TET2 plays a role in slowing down the differentiation of hematopoietic stem and progenitor cells, thereby preventing premature commitment towards monocytes/macrophages and myelo-proliferation (Ko et al., 2011; Castro-Diaz et al., 2014). Notably, hematopoietic stem cells also depend on DNA methylation, with DNMT1 hypomorphism leading to an imbalance in cell fate specification (Bröske et al., 2009; Smith & Meissner, 2013)

Oocytes and zygotes express TET3 but not TET1 and TET2 (Gu et al., 2011). Following fertilization, the maternal TET3 catalyzes the oxidation of 5mC into 5hmC at the paternal genome (Gu et al., 2011). The conditional depletion of the maternal TET3 allele in the zygote leads to a defect in the demethylation process of paternal OCT4 and NANOG alleles, coming along with diminished gene activation. In line with this, females with a conditional TET3 knockout in their oocytes are less fertile. Nevertheless, TET3 is dispensable for the development of both male and female germ cells. Notably, homozygous deletion of TET3 is neonatally lethal (Gu et al., 2011).

DNMT3s and TETs co-occur at thousands of genomic loci in pluripotent cells

While the DNA methylome undergoes changes during cell state transitions, such as differentiation or reprogramming, the DNA methylation landscape appeared to be highly stable within a given cell state. This picture was supported by the preservation of global patterns even upon depletion of DNMT3s and TET enzymes (Liao et al., 2015; Verma et al., 2018). However, a more intricate scenario emerged upon closer examination of hPSC depleted of both *de novo* methyltransferases DNMT3A/B (DKO), revealing focal DNA methylation dynamics at thousands of genomic loci, which persist independent of any cell state transition (**Figure 3**) (Charlton et al., 2020; Ginno et al., 2020).

In DKO hPSCs, over 11,000 regions, with an average size of 700bp, are rapidly hypomethylated by focal TET activity, while overall global levels exhibit only minor changes (Charlton et al., 2020). The active process of DNA demethylation at these sites becomes evident, as their methylation levels remain high when the DKO is executed in the background of a TET1/2/3 triple knockout (TKO) (resulting in a pentuple knockout (PKO)). Furthermore, the presence of 5hmC and TETs at these loci in hPSCs implies ongoing oxidation of 5mC, necessitating local *de novo* methylation by DNMT3s to uphold CpG methylation (Charlton et al., 2020). Either DNMT3A or B activity is sufficient to counteract this demethylation process, with one being largely compensatory in the absence of the other. This competitive interplay between DNMT3s and TETs has been termed DNA methylation turnover, and the regions influenced by this phenomenon are referred to as DKO-differentially methylated regions (DKO-DMRs). Of note, the occurrence of the DNA methylation turnover is largely conserved in mESCs, and the extent of DKO-associated hypomethylation is less prominent in adult mouse tissues and *in vitro* differentiated cells (Charlton et al., 2020).

Over half of the DKO-DMRs overlap with H3K4me1 Chromatin-immunoprecipitation peaks) (ChIP) (Charlton et al., 2020). However, a relatively low percentage of them additionally coincide with H3K27me3 or H3K27Ac marks, implying the absence of a classical poised chromatin state or a state of active enhancers, respectively. Additionally, a slight enrichment in chromatin accessibility was observed, along with increased CpG density compared to the background. Nonetheless, none of these features is unique to DKO-DMRs. Furthermore, neither a common sequence motif for a known TF, nor enrichment of any specific TF, was identified within DKO-DMRs (Charlton et al., 2020). Still, DKO-DMRs are bound by tissue-specific TFs upon differentiation, suggesting a possible enhancer function. In line with this, 85% of DKO-DMRs overlap with putative somatic enhancers and alterations of the

transcriptome upon loss of DNMT3s in hPSCs are frequently associated with DKO-DMRs, However, not all these enhancers coincide with a DKO-DMR. Furthermore, the 5'UTRs of evolutionarily young LINE1 elements, in particular L1HS and L1PA, are targets of the DNA methylation turnover (Charlton et al., 2020). This observation has led to the hypothesis that DNA methylation turnover could play a role in transposon regulation.

Nevertheless, the mechanism and function underlying the target-specificity of the DNA methylation turnover across diverse genomic endeavors are not yet understood.

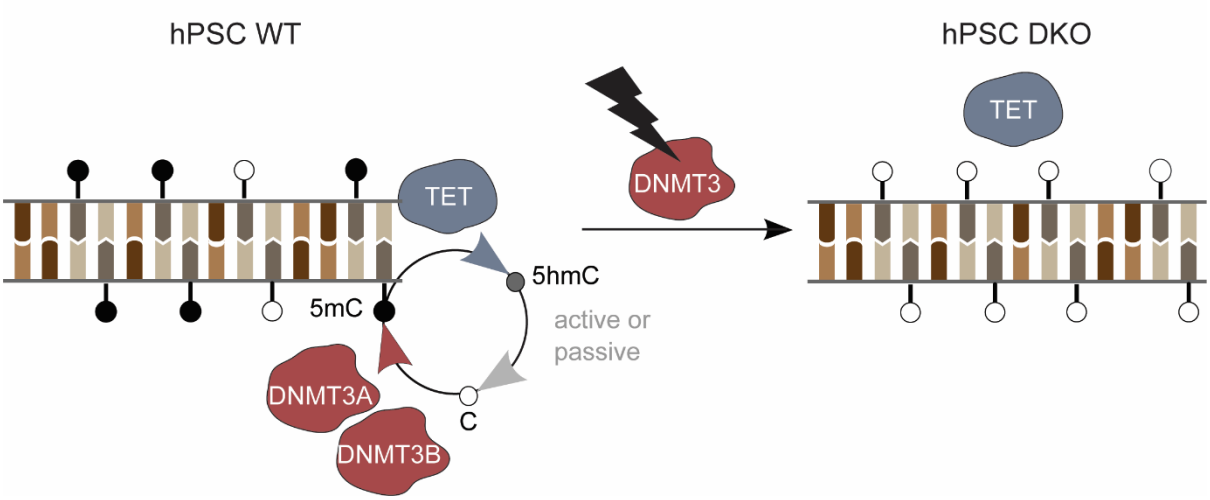

Figure 3: DNA methylation turnover in hPSCs. The schematic displays the co-occurrence of DNMT3A/DNMT3B and TETs at an overall highly methylated genomic locus in hPSCs. Their opposing activities enable high levels of 5mC and 5hmC. In the absence of DNMT3s (DKO hPSCs), TETs hypomethylate those loci, which is why they are termed DKO-DMRs (Charlton et al., 2020).

Protein domains and targeting of DNA methyltransferases

The C-terminus of mammalian DNMT1 and DNMT3A/B contains a methyltransferase (MTase) domain, similar in structure to the bacterial restriction methyltransferases (**Figure 4**) (Bestor et al., 1988; Bestor, 1990). Conversely, the co-factor of *de novo* methyltransferases, DNMT3L, lacks an MTase domain (Aapola et al., 2000).

Utilizing their cysteine-rich ADD domain (ATRX-DNMT3-DNMT3L), DNMT3A/B and DNMT3L preferentially bind to the unmodified histone H3 tail or H3K9me3-modified tail, while its binding is hindered in the presence of H3K4me3 (**Figure 4**) (Zhang et al., 2010b; Ooi et al., 2007; Baubec et al., 2015). In cases where the ADD domain is unable to bind to the H3 tail (e.g. by H3K4me3), it instead binds to the MTase domain within its own C-terminus, leading to auto-inhibition (Guo et al., 2015). The targeting of DNMT3s is often associated with specific histone modifications and the enzymes responsible that catalyze them; the relationship varies depending on the genomic context (Li et al., 2008; Zuo et al., 2012; Epsztejn-Litman et al., 2008). The intertwined connections between DNA methylation and other epigenetic mechanisms are the subject of a later paragraph.

Furthermore, the N-terminus of DNMT3s contains a PWWP domain (Pro-Trp-Trp-Pro), which is crucial for DNA binding (**Figure 4**) (Suetake et al., 2011; Jeong et al., 2009). However, there is limited understanding on how the DNA sequence determines target-specificity of DNMT3s. *De novo* methylation activity does not occur on nucleosome-associated DNA, has a positive correlation with the density of methylated CpGs, and conversely, it negatively correlates with unmethylated CpG-density (Baubec et al., 2015). Furthermore, several studies have investigated DNA sequence preferences of DNMT3. Among the identified motifs for DNMT3B is (NTCpGGN), a finding validated through a biochemical assay (Loaeza-Loaeza et al., 2020). Additionally, DNMT3 and DNMT1 activities are repressed by G-quadruplexes, a four-stranded secondary DNA structure that predominantly occurs in CGIs due to their high GC content (Mao et al., 2018; Loiko et al., 2022).

Aside from some degree of target-specificity, DNMT3A and DNMT3B are largely redundant in their function in both hPSCs and mESCs (Liao et al., 2015; Okano et al., 1999).

In contrast to DNMT3s, DNMT1's primary guiding mechanism involves its interaction with UHRF1 during S-phase of the cell cycle (Bostick et al., 2007; Sharif et al., 2007). The SET- and RING-associated (SRA) domain of UHRF1 specifically recognizes hemi-methylated CpGs at replication forks, while its Tudor domain (TDD) binds to unmethylated H3K4 and

H3K9me3, (Nady et al., 2011; Rothbart et al., 2012). Moreover, UHRF1's RING finger domain ubiquitylates various amino acids on the histone H3 tail, facilitating DNMT1 binding (Nishiyama et al., 2013; Qin et al., 2015). The replication foci targeting sequence (RFTS) of DNMT1, hidden within its MTase domain, is released upon recruitment by UHRF1, enabling DNMT1 to bind the replicating DNA (Takeshita et al., 2011; Song et al., 2011). Notably, the catalytic activity of DNMT1 is diminished within complexes including oxidized forms of 5mC (Seiler et al., 2018).

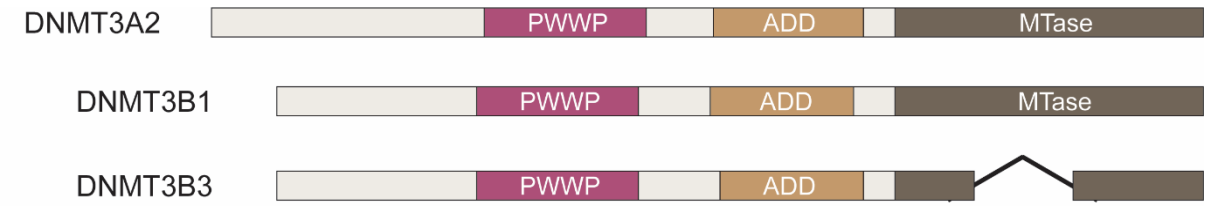

Figure 4: Protein domains of human DNMT3A and DNMT3B. The schematic displays the major protein domains of the predominantly expressed DNMT3A isoform, DNMT3A2, and DNMT3B which undergoes an isoform switch from DNMT3B1 to DNMT3B3 upon exit of pluripotency. PWWP (Pro-Trp-Trp-Pro). ADD (ATRX-DNTM3-DNMT3L). MTase (methyltransferase). Adapted from (Ren et al., 2018; Del Castillo Falconi et al., 2022).

Protein domains and targeting of Ten-eleven translocation dioxygenases

The catalytic domains of TET1-3 encompass a double-stranded β-helix (DSBH) and a cysteine-rich region that is characteristic of 2OG-Fe(II) oxygenases (**Figure 5**) (Tahiliani et al., 2009). Additionally, TET1 and TET3 contain a C-terminal CXXC-type zinc-finger domain (two cysteines separated by two amino acids). This domain enables other enzymes, such as H3K4 methyltransferase MLL, CXXC zinc finger-type protein CFP1, and DNMT1, to bind to unmethylated CpG (Frauer et al., 2011; Lee et al., 2001; Birke et al., 2002). Nevertheless, the functional role of the CXXC domain in TET1 and TET3 is controversially discussed (Frauer et al., 2011; Xu et al., 2012; Zhang et al., 2010a). In contrast to TET1 and TET3, the DNA sequence coding for the CXXC domain of TET2 has been separated from the rest of the gene due to a chromosomal inversion. However, the resulting independent intracellular DNAX-activating protein IDAX (also known as CXXC4) maintains an interaction with TET2 (Ko et al., 2013).

TET targeting has frequently been associated with various TFs. In murine cells, TET1 physically interacts with NANOG, and during reprogramming of mouse fibroblasts, genomic

targeting of TET1 is impaired upon Nanog depletion (Costa et al., 2013). This connection between pluripotency factors and TET targeting was also observed in hPSCs, where active demethylation preferentially occurs around loci bound by NANOG, OCT4, and SOX2, in comparison to targets of other TFs (Ginno et al., 2020). Notably, the catalytic domains of murine TET proteins exhibit a preference for sites bound by methylation-sensitive immediate-early transcription factors, such as c-MYC, CREB, JUN/FOS, which often carry basic helix-loop-helix (bHLH) or basic leucine zipper (bZIP) domains (Ravichandran et al., 2022). Conversely, OCT4 binding sites are less favored by TETs (Ravichandran et al., 2022). In murine retinal cells, a Tet3 isoform lacking the CXXC domain interacts with the transcriptional repressor REST and supports its catalytic activity, suggesting that REST may guide Tet3 to its genomic targets (Perera et al., 2015).

Moreover, TET activity positively correlates with chromatin accessibility, although the underlying causative relationship has not been examined yet (Ginno et al., 2020). Additionally, DNA binding of TET2 has been found to be enhanced by a site-specific monoubiquitylation (Nakagawa et al., 2015). In murine cells, TET1 occupancy positively correlates with CpG density, particularly showing a preference for binding CpG-rich promoter sequences (Wu et al., 2011a; Williams et al., 2011).

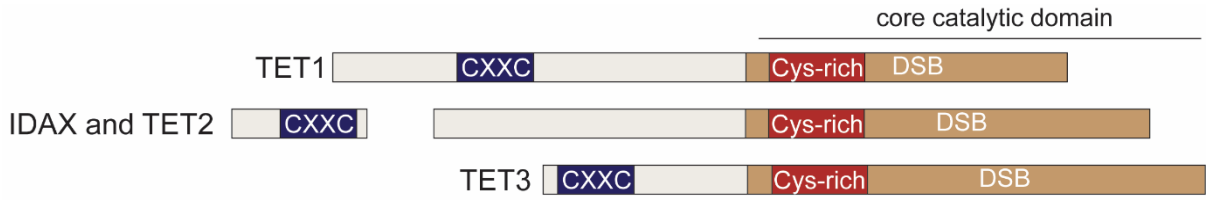

Figure 5: Protein domains of human TET1, 2 and 3. The schematic displays the major protein domains of the predominantly expressed TET1, TET2 and TET3 isoforms. TET2 lacks a CXXC domain which is instead encoded by IDAX. CXXC (two cysteines separated by two amino acids), DSBH (double-stranded β helix domain). Adapted from (Ravichandran et al., 2018).

DNA methylation patterns in different genomic contexts and their functional role

Back in the 1970s, studies conducted in mouse, rabbit and chicken revealed an inverse relationship between the expression of globin genes and DNA methylation across various cell types (Christman et al., 1977; McGhee & Ginder, 1979; Waalwijk & Flavell, 1978; Doerfler, 1983; Bird, 1984). In line with these findings, multiple studies involving distinct virus DNA and cellular host systems, such as adenovirus type 12 (Ad12)-transformed hamster cells, consistently show a negative correlation between DNA methylation and expression of the virus particles (Jähner et al., 1982; Sutter & Doerfler, 1980). However, DNA methylation does not universally correlate with gene repression, nor does its absence always imply activation. In fact, their relationship is more complex and intertwined with the function of other epigenetic mechanisms and the binding of transcription factors.

DNA methylation and gene silencing

DNA methylation represses gene transcription, primarily by targeting its promoter, but not through its coding region (Busslinger et al., 1983; Fradin et al., 1982; Kruczek & Doerfler, 1983). During development, promoter DNA methylation remains relatively stable and constitutes less than 5% of differentially methylated regions among various tissues (Hon et al., 2013; Ziller et al., 2013). Conversely, regions within 2kb of promoter CGIs, known as CGI shores, show more variability (Ziller et al., 2013; Irizarry et al., 2009).

Silenced promoters

DNA methylation is thought to play a role in long-term repression rather than being the primary mechanism for gene silencing. For example, G9a-mediated histone methylation of H3K9 is the first step in the inhibition of the pluripotency factor Oct4 during differentiation of mESCs (Epsztejn-Litman et al., 2008; Feldman et al., 2006). This is followed by *de novo* DNA methylation which prevents reprogramming of the differentiated cells (Epsztejn-Litman et al., 2008; Feldman et al., 2006). Notably, the histone methyltransferase G9a is involved in DNMT3 recruiting in this genomic context (Epsztejn-Litman et al., 2008)

On the other hand, CGI-containing promoters are often silenced by trimethylation of histone 3 lysine 27 (H3K27me3) catalyzed by the polycomb repressive complex 2 (PRC2) (Schwartz & Pirrotta, 2007; Orlando et al., 2012; Shao et al., 1999; Dellino et al., 2004). H3K27me3 hinders

transcription initiation machinery and ATP-dependent nucleosome remodeling by the SWI/SNF complex (Schwartz & Pirrotta, 2007). Initially, PRCs were discovered in *D. melanogaster*, where CpG methylation is largely absent during most stages of development, and therefore, it cannot contribute to silencing (Schwartz & Pirrotta, 2007). Moreover, in humans, EZH2, a member of the PRC2 complex, physically interacts with DNMT3s and DNMT1 and recruits them to gene promoters (Viré et al., 2006). In line with this, in mice, DNMT3A is recruited to shores of H3K27me3-promoters by the histone modification itself (Manzo et al., 2017). However, EZH2-mediated DNMT recruitment is not sufficient for *de novo* methylation at H3K27me3-positive promoters (Rush et al., 2009). In fact, both epigenetic modifications are frequently mutually exclusive, with *de novo* methylation restricting H3K27me3 and preventing its spread into those CGIs which are methylated (Lynch et al., 2012). In line with this, 5hmC displays particular accumulation at PRC2-occupied promoters, implying target-specificity of TETs to silenced promoters (Wu et al., 2011b). Furthermore, TET1 associates with components of the Sin3a corepressor complex, such as Sin3a itself and HDAC1, which mediates transcriptional silencing through histone deacetylation (Williams et al., 2011). Intriguingly, murine Dnmt3a has also been shown to interact with the histone deacetylase HDAC1 (Fuks et al., 2001; Taunton et al., 1996). This interaction can be disrupted by a post-translational modification of Dnmt3a, namely sumoylation, which can even impact Dnmt3a's repressive transcriptional effect (Ling et al., 2004). Thus, promoter silencing involves a complex and intertwined relationship between histone modifiers, DNA methyltransferases and demethylases.

Furthermore, DNA methylation at promoters, and also enhancers, can hinder the binding of certain TFs to their DNA binding motifs, as revealed by various *in vitro* and *in vivo* studies (Mann et al., 2013; Gaston & Fried, 1995; Watt & Molloy, 1988; Yin et al., 2017; Domcke et al., 2015). For instance, members of the bHLH, bZIP, and ETS families, along with NRF1, exhibit reduced binding to methylated DNA. Conversely, factors from the NFAT, homeodomain, POU, and NKX families, among others, tend to prefer methylated DNA, while other TFs, like YY1, show no preference for either methylated or unmethylated DNA (Mann et al., 2013; Gaston & Fried, 1995; Watt & Molloy, 1988; Yin et al., 2017; Domcke et al., 2015).

Active promoters

In certain genomic contexts, DNA demethylation is sufficient to activate gene expression, while others remain repressed. In a genome-wide study conducted in peri-implantation mouse embryos, repressed germline genes and one pluripotency gene were activated through promoter hypomethylation in the absence of DNMT3B, while the loss of promoter methylation of other pluripotency and hematopoietic genes did not lead to gene activation (Borgel et al., 2010). Moreover, actively transcribed genes are often marked by H3K4 trimethylation, a modification that is typically mutually exclusive with DNA methylation, just as the repressive H3K27me3 (Barski et al., 2007; Lynch et al., 2012). The presence of H3K4me3 on active promoters can explain their hypomethylation, given DNMT3's autoinhibitory mechanism as a consequence of their incapability to bind to H3K4me3 (Baubec et al., 2015). Furthermore, TET1 frequently occupies H3K4me3-promoters in mice to enhance transcriptional activity (Wu et al., 2011a). In line with this, 5hmC occurs in a bimodal pattern at active promoters, with depletion directly over transcription start sites and accumulation at its neighboring regions (Szulwach et al., 2011a). Interestingly, genes expressed at intermediate levels are more enriched for 5hmC than those with the highest expression levels, which is in line with the loss of 5hmC at promoters devoid of the repressive PRC2 (Szulwach et al., 2011a; Wu et al., 2011b). Similar to DNA demethylation, depletion of H3K4me3 was shown to only have minor effects on gene activity (Clouaire et al., 2012). Furthermore, histone H3 and H4 acetylation at promoters correlate with transcriptional activity (Bernstein et al., 2002). Apart from epigenetic modifications, promoter activity depends on TF availability and their DNA binding, the latter of which is facilitated by nucleosome depletion (Lee et al., 2004; Zaret & Carroll, 2011).

Bivalent promoters

The so-called bivalent domains carry the repressive histone mark H3K27me3, while they are simultaneously enriched for the active histone mark H3K4me3 and TET-mediated 5hmC, together with DNA hypomethylation (Barski et al., 2007; Bernstein et al., 2006; Pastor et al., 2011; Wu et al., 2011a). In contrast to promoters enriched for H3K4me3-only, genes associated with bivalent promoters exhibit low expression levels. Further, rather than promoting activity as seen for H3K4m3-only promoters, TET1 binding at bivalent promoters facilitates EZH2 recruitment, leading to gene repression (Wu et al., 2011a). Interestingly, as for H3K4me3- and H3K27me3-only promoters, bivalent CGIs are usually unmethylated and, conversely, methylated CGIs largely lack H3K27me3 and H3K4me3 (Lynch et al., 2012).

DNA methylation at gene bodies

During gene transcription, SETD2 catalyzes trimethylation of H3K36 (H3K36me3) within the coding region (Sun et al., 2005; Krogan et al., 2003). Further, H3K36me3 is recognized and bound by the PWWP domain of DNMT3s, leading to DNA methylation at actively transcribed gene bodies (Dhayalan et al., 2010). Within gene bodies, exons exhibit higher levels of DNA methylation in comparison to introns, a pattern that persists beyond methylation in the CpG context (Lister et al., 2009; Laurent et al., 2010). Several studies investigated the relationship between DNA methylation and gene expression, uncovering DNA methylation-dependent exon exclusion or inclusion in different genomic contexts based on the sensitivity of TFs to DNA methylation (Maunakea et al., 2013). Moreover, DNA methylation of intragenic CGIs regulates tissue-specific alternative promoter usage (Maunakea et al., 2010). Detailed localization and *de novo* activity analysis have highlighted that primarily DNMT3B, rather than DNMT3A, binds to gene bodies and orchestrates their DNA methylation, a process that relies on the presence of H3K36me3 (Baubec et al., 2015). In contrast, DNMT3A demonstrates a higher sensitivity to di-methylated H3K36 in intergenic regions of euchromatin (Weinberg et al., 2019). Additionally, genes with high expression levels and correspondingly elevated H3K36me3 exhibit lower levels of 5hmC within their gene bodies in hPSCs, as compared to genes with intermediate expression levels (Szulwach et al., 2011b). Conversely, in the mouse brain, the enrichment of 5hmC within gene bodies is positively correlated with gene expression (Szulwach et al., 2011b). Of note, the oxidation of methylcytosine in gene bodies is mainly catalyzed by TET2 in mESCs (Huang et al., 2014).

DNA methylation at enhancers

Enhancers are among the most variable regions in the genome when it comes to tissue-specific differential methylation (**Figure 6**) (Hon et al., 2013; Ziller et al., 2013). Moreover, enhancer hypomethylation positively correlates with the expression of their associated target genes (Stadler et al., 2011). Traditionally, the presence of tissue-specific H3K4me1 and H3K27Ac marks indicate active enhancers, while those marked with H3K4me1-only are considered poised enhancers (Rada-Iglesias et al., 2011; Heintzman et al., 2007; Creyghton et al., 2010; Heintzman et al., 2009). However, exceptions exist where enhancers deemed active lack H3K27Ac enrichment, and conversely, enhancers rich in H3K27Ac might not always be active (Heintzman et al., 2009; Zhu et al., 2013). A subgroup of enhancers carries H3K27me3 in addition to H3K4me1, which downregulates the expression of associated genes below the

levels of genes linked to enhancers with H3K4me1-only (Zentner et al., 2011). Notably, adult tissue-specific enhancers lack H3K27me3 (Zentner et al., 2011; Rada-Iglesias et al., 2011). Certain H3K27Ac-positive enhancers, whether located within a gene or outside, additionally possess H3K36me3. Furthermore, H3K27me-positive enhancers are also marked with H3K9me3 (Zentner et al., 2011). Unlike promoters, active enhancers generally show minimal or no presence of H3K4me3. Moreover, enhancers are enriched for 5hmC in ESC, which is diminished in the absence of TET2, leading to reduced enhancer activity (Stroud et al., 2011; Szulwach et al., 2011a; Pastor et al., 2011; Hon et al., 2014).

Of note, unmethylated enhancers in adult tissues are not necessarily active, but can also be a remnant of embryonic development where hypomethylation of certain enhancers was functionally relevant (Hon et al., 2013; Jadhav et al., 2019). Interestingly, these so-called 'decommissioned' enhancers can be reactivated in the absence of PRC2 (Jadhav et al., 2019), indicating that a final shutdown might only be accomplished through gaining methylation.

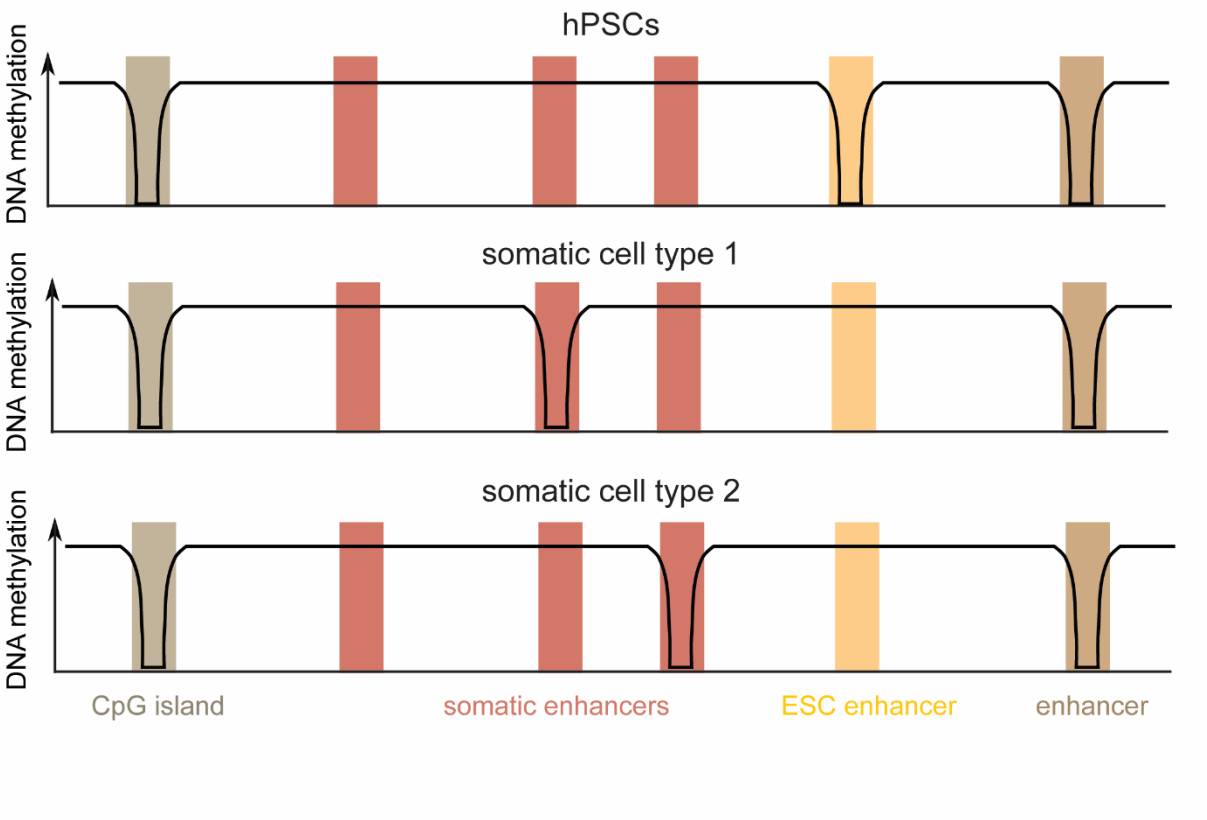

Figure 6: Cell type-specific DNA methylation at enhancers. The schematic displays cell type-specific DNA methylation patterns at enhancers. While some enhancers a constitutively demethylated, others only become demethylated upon differentiation into a specific cell type. Similarly, enhancers which are active and unmethylated in PSCs are hypermethylated upon differentiation.

DNA methylation at imprinting control regions

DNA methylation plays a pivotal role in genomic imprinting, where imprinted genes are mono-allelically expressed in a parental-origin-specific manner (Ferguson-Smith, 2011). These genes are organized in clusters, each of which is controlled *in cis*- by an imprinting control region (ICR). Maternally expressed genes are linked with maternally methylated ICRs located in promoters that repress noncoding RNAs and other multifunctional transcripts, eventually allowing expression of the protein-coding genes (Bartolomei et al., 1993; Stöger et al., 1993; Ferguson-Smith et al., 1993; Sleutels et al., 2002; Mancini-DiNardo et al., 2006). Conversely, hypomethylation at the corresponding paternal ICR and expression of the non-coding transcript lead to silencing of paternal protein-coding alleles. On the other hand, the paternal expression of imprinted genes is associated with paternally methylated ICRs located in intergenic regions (Bell & Felsenfeld, 2000; Hark et al., 2000). Remarkably, even during global hypomethylation in the pre-implantation embryo, ICRs remain methylated due to selective DNA binding of ZFP57-KAP1 (Li et al., 2008; Olek & Walter, 1997). Imprints are erased only in the germline to enable the re-establishment of male or female patterns during gametogenesis, occurring in the late fetus or postnatally, respectively, through the combined activity of DNMT3A and DNMT3L (Kaneda et al., 2004; Bourc'his et al., 2001; Yamaguchi et al., 2013). Other imprints are only established after fertilization in a tissue-specific manner (Feil et al., 1994). The critical role of DNA methylation in genomic imprinting was underscored by the loss of DNMT1, which leads to the de-repression of imprinted genes in mouse embryos, ultimately resulting in embryonic lethality at E11 (Li et al., 1994).

A unique example of genomic imprinting is X-chromosome inactivation (XCI). Occurring shortly after fertilization, in the four- to eight-cell stage, the paternal X chromosome undergoes imprinted XCI, primarily due to the expression of the noncoding RNA Xist, while the maternal Xist allele is silenced by H3K27me3 (Okamoto et al., 2004; Inoue et al., 2017). Xist coats its chromosome of origin, namely the paternal X chromosome, which comes along with the removal of activating histone marks, such as H3K4me3, H3K9ac, and H4ac. Concurrently, repressive modifications, such as H3K27me3, are added, eventually resulting in the silencing of X-linked genes (Okamoto et al., 2004, 2005). Upon developmental progression, the reversion of these epigenetic alterations reactivates the paternal X chromosome in the inner cell mass which gives rise to the embryo proper (Mak et al., 2004; Loda et al., 2022). Subsequent random XCI is also initiated by monoallelic Xist expression from either the maternal or paternal X chromosome. This process is also associated with a transition from

active to repressive histone marks (Chaumeil et al., 2006; Żylicz et al., 2019). However, random XCI is not reversed throughout development; on the contrary, Xist silencing on the active chromosome is maintained by DNA methylation. Loss of DNMT1 activity leads to the reactivation of Xist on the active X chromosome, subsequently resulting in the silencing of respective X-linked alleles in somatic cells (Panning & Jaenisch, 1996; Beard et al., 1995).

DNA methylation at transposable elements

Transposable elements constitute nearly half of the mammalian genome (International Human Genome Sequencing Consortium, 2001; Mouse Genome Sequencing Consortium, 2002). The most prevalent class of transposons are retrotransposable elements, which are categorized into long terminal repeats (LTRs), long interspersed elements (LINEs), and short interspersed elements (SINEs), based on their evolutionary origin (**Figure 7**) (Goodier & Kazazian, 2008). LTRs and LINEs encode their own machinery for retrotransposition, regulated by promoters situated in their LTR or the 5'UTR, respectively (Goodier & Kazazian, 2008). In contrast, SINEs are non-autonomous and rely on proteins encoded by LINE1s for their own mobilization (Goodier & Kazazian, 2008). In addition to transposons, the mammalian genome also contains simple repeats, also referred to as microsatellites, constituting ~3% of the human genome. Simple repeats typically consist of 1- to 6-mer nucleotide tandem repeats that, when mutated, can contain interruptions in their tandem structure (Tóth et al., 2000; International Human Genome Sequencing Consortium, 2001; Taylor et al., 1999; Ellegren, 2004).

Retrotransposable elements exhibit varying degrees of enrichment for the repressive marks H3K9me2/H3K9me3 and DNA methylation (Meissner et al., 2008; Martens et al., 2005; Bulut-Karslioglu et al., 2014). In mESCs, ERVK LTRs are notably enriched for SETDB1-mediated H3K9me3. The loss of SETDB1 or its recruiting factor KAP1 results in de-repression of ERVK LTRs (Matsui et al., 2010; Karimi et al., 2011; Rowe et al., 2010). Similarly, ERVL-LTRs are suppressed through G9a/GLP-mediated H3K9me2, while Suv39h-dependent trimethylation of H3K9 silences intact LINEs in pluripotent cells (Maksakova et al., 2013; Bulut-Karslioglu et al., 2014). Additionally, certain human-specific LINE-1 elements, which emerged after the human-chimpanzee divergence, are targeted by KAP1 and are de-repressed upon its depletion (Castro-Diaz et al., 2014). In contrast to LTRs and LINEs, SINEs exhibit relatively modest H3K9me3 signals (Bulut-Karslioglu et al., 2014).

Furthermore, the CpGs within LTRs and LINEs are often highly methylated, whereas methylation levels at SINEs tend to correlate with CpG-density (Meissner et al., 2008). However, the precise role of DNA methylation in transposon silencing is complex. During the reprogramming of primed mESCs into a more naïve state, the majority of LTRs, LINEs, and SINEs undergo demethylation, which initially comes along with de-repression in all classes. However, subsequently, retrotransposons are alternatively silenced through the targeted placement of H3K27me3 (Walter et al., 2016). Still, the loss of DNMTs in mESCs leads to the upregulation of LINE1 but only to minor de-repression of the evolutionarily youngest mouse LTRs, namely IAP ERVK (Matsui et al., 2010; Karimi et al., 2011; Hutnick et al., 2010). Of note, in mESCs depleted of DNMTs, IAP ERVKs and the 5'UTRs of certain LINE1 elements are protected from complete demethylation, likely due to the presence of H3K9m3 (Leung et al., 2014). In line with this, IAP activation is more pronounced in the absence of SETDB1, suggesting a stronger role for H3K9me3-mediated IAP silencing (Matsui et al., 2010). In hPSCs, particularly the evolutionarily youngest human-specific L1HS elements were shown to be repressed by DNA methylation (Castro-Diaz et al., 2014).

In somatic and germ cells, the role of DNA methylation for transposon repression becomes more apparent. Loss of DNMT1 in mouse embryos leads to widespread expression of IAP ERVKs throughout the embryo (Walsh et al., 1998). Of note, DNMT1 also operates as *de novo* methyltransferase in this particular genomic context, which makes DNMT3s dispensable for IAP methylation in mouse embryos (Haggerty et al., 2021). In line with this, during *in vitro* differentiation of DNMT1-deficient mESCs, IAP expression is activated, compared to its modest upregulation during pluripotency (Hutnick et al., 2010). Of note, cell commitment is accompanied by a reduction of H3K9me3, the primary IAP-silencing mechanism during pluripotency, suggesting a transition towards alternative silencing mechanisms, such as DNA methylation (Bulut-Karslioglu et al., 2014). In mouse male germ cells, DNMT3L-mediated methylation is crucial for the silencing of LINE and IAP transposons (Bourc'his & Bestor, 2004).

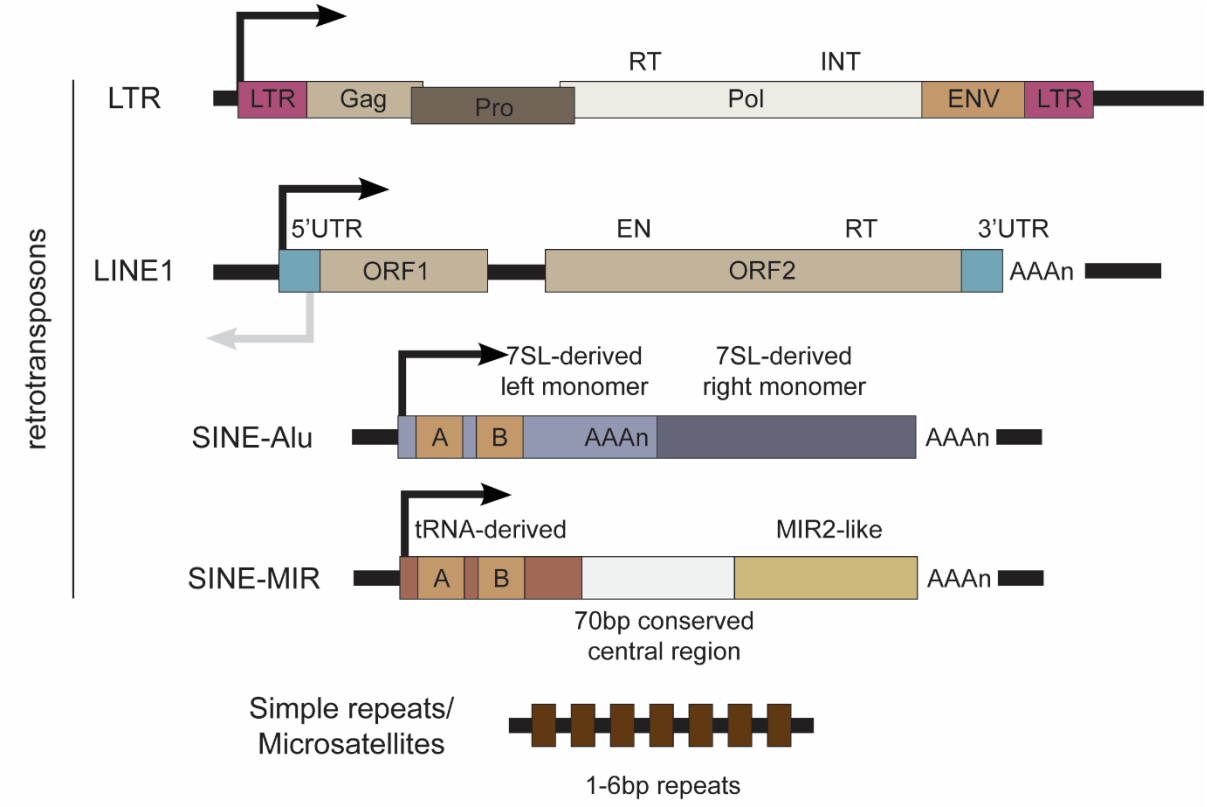

Figure 7: Structure of retrotransposon families and simple repeats. The schematic displays the structure of all class 2 transposons, which are retrotransposable elements, including LTR and non-LTR families (LINE1, SINEs-Alu, SINE-MIR). Further, the tandem repeat structure of simple repeats (microsatellites) is shown. LTR (long terminal repeat). Gag (group-specific antigen), Pro (protease), Pol (polymerase). RT (reverse transcriptase), INT (Integrase), Env (envelop). ORF1/2 (open reading frame 1/2). EN (endonuclease). A and B (tandem monomers). Adapted from (Mills et al., 2007; Goodier & Kazazian, 2008).

Centromeric and pericentromeric satellite repeats

Major and minor satellite repeats are tandem repeats located in the pericentromeric and the centromeric regions, respectively. In humans, major satellites can extend over several megabases, whereas minor satellites are typically found in the kilobase range (Thakur et al., 2021).

The heterochromatic environment of pericentromeric regions is marked by H3K9me3, H3K27me1, and H3K9me3-dependent H4K20me3, alongside with high levels of DNA methylation (Lehnertz et al., 2003; Liao et al., 2015; Peters et al., 2003; Schotta et al., 2004). Depletion of the histone lysine methyltransferase Suv39h results in accumulation of H3K27me1 and H3K9me1, along with a modest upregulation of major satellite expression (Peters et al., 2003; Schotta et al., 2004; Lehnertz et al., 2003). Interestingly, loss of Suv39h

also leads to DNA hypomethylation of major satellite repeats (Lehnertz et al., 2003). Conversely, H3K9me3 deposition at pericentromeric regions do not depend on DNMTs (Lehnertz et al., 2003). Furthermore, depletion of DNMT1 or DNMT3A/B does not result in reactivation of major satellites *in vitro*, suggesting that DNA methylation is not the primary mechanism for their repression (Lehnertz et al., 2003).

Like major satellites, minor satellite repeats exhibit enrichment of H3K27me1 and H3K9me3, although the latter is absent in human cells (Schotta et al., 2004; Sullivan & Karpen, 2004). The loss of SUV39H does not lead to a decrease in DNA methylation at centromeric repeats, which is in contrast to major satellites (Lehnertz et al., 2003). Of note, deficiency of DNMT3B is associated with chromosomal mispairing during mitosis and the immunodeficiency, centromeric instability and facial anomaly (ICF) syndrome, which is characterized by complete loss of methylation at satellite repeats (Gopalakrishnan et al., 2009; Xu et al., 1999).

Methods to study DNA methylation and hydroxymethylation

Bisulfite (BS) conversion followed by sequencing is considered the gold standard method for studying DNA methylation and was first established in 1992 using human DNA (Frommer et al., 1992). This technique involves treatment of DNA with sodium bisulfite, which deaminates all cytosines into uracil except for 5mC and 5hmC (**Figure 8**). Formyl and carboxyl groups on cytosines do not protect from conversion. Subsequent PCR amplification manifests the conversion by pairing adenines with the uracils. When the sequenced bisulfite-treated DNA is aligned to a non-treated reference genome, it reveals the cytosines with and without methylation/hydroxymethylation.

To distinguish between 5mc and 5hmC, an additional step of oxidation before bisulfite treatment (oxBS) converts 5hmC into 5fC, which is not protected from deamination into uracil by bisulfite treatment (Booth et al., 2013). Thus, oxBS reveals the true DNA methylation signal, and the difference between oxBS and BS further allows to identify 5hmC.

While BS/oxBS sequencing provides a detailed map of DNA methylation on a single-base level, methylated-DNA immunoprecipitation (MeDIP) followed by sequencing locates DNA methylation on a larger scale of around 100-150bp (Weber et al., 2005). Additionally, 5hmC-, 5fC- and 5caC-specific antibodies allow for their genomic location to be studied as well (Williams et al., 2011).

However, both BS-seq and MeDIP-seq are based on DNA isolation and provide only a snapshot of the DNA methylation landscape at a specific moment in the life of a cell population. The only current method for tracking DNA methylation dynamics in living cells involves a genetic system in which a fluorescent protein is expressed under the control of a DNA methylation-sensitive minimal promoter (Stelzer et al., 2015). Changes in the DNA methylation level are reported by altered expression of the fluorescent protein.

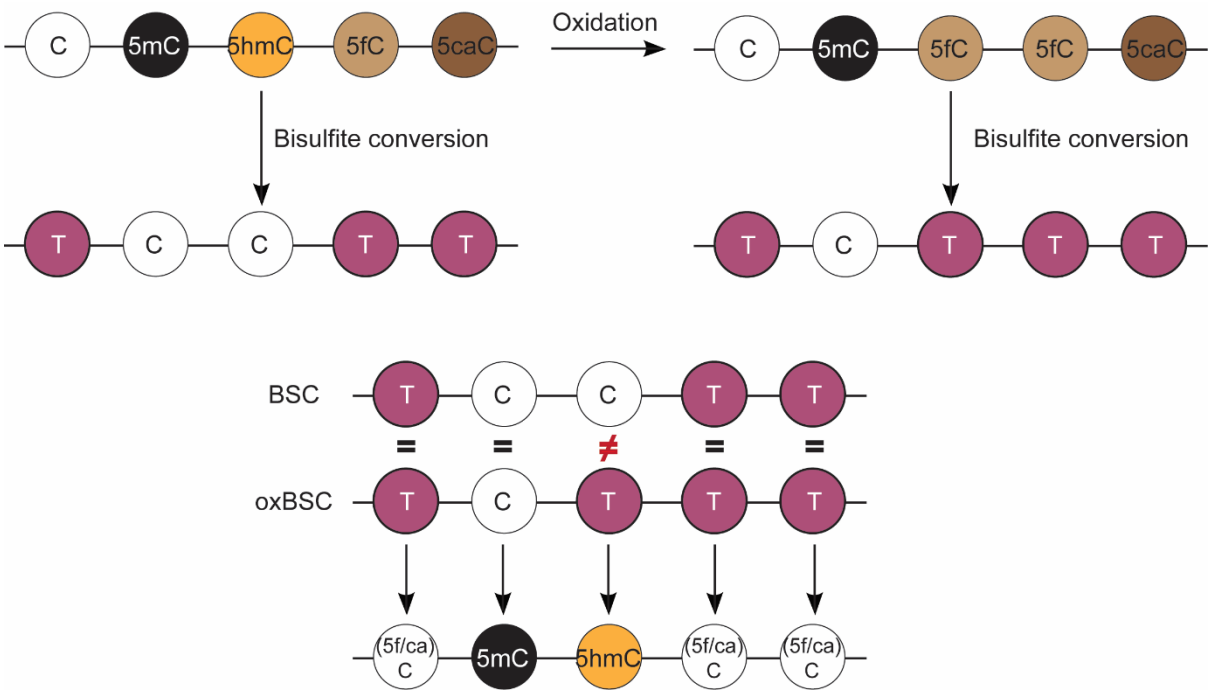

Figure 8: Bisulfite and oxidative bisulfite conversion. The schematic displays the molecular changes that happen to the different cytosine modifications during bisulfite conversion (BSC) and BSC with a preceding oxidation step (oxBSC). The differences in the DNA sequence after BSC and oxBSC allows to distinguish between 5mC and 5hmC.

Aims of the project

Given the intricate target-specificity of the DNA methylation turnover, the work of this thesis aims to explore the functional role of the DKO-DMRs, the association of the DNA methylation turnover with cellular differentiation and its selectivity for evolutionarily young retrotransposons. The main goals are summarized by the following core objectives:

Clarify enhancer functionality: Given the histone modification patterns at DKO-DMRs observed at DKO-DMRs, and their strong overlap with potential somatic enhancers, an investigation into the functional role of DKO-DMRs as enhancers in differentiated cells is undertaken.

Examine the relationship between the DNA methylation turnover in hPSCs and differentiation-related hypomethylation: Due to the pronounced overlap with potential somatic enhancers, the research is focused on dissecting the potential correlation between DNA methylation turnover and the demethylation events that occur during differentiation. This involves a qualitative and quantitative analysis, initially focusing on *in vitro* differentiated cells across the three germ layers and subsequently extending to more mature cell types and adult tissues.

Investigate the DNA methylation turnover in differentiated cells: Building upon the insights gained in the second aim, the objective of the third aim is to investigate the continuity of the DNA methylation turnover from hPSCs into differentiated cells and explore the possibility of the *de novo* establishment of the turnover in differentiated cells.

Explore retrotransposon target specificity: The thesis's second focus is to delve into the target-specificity of the DNA methylation turnover among various retrotransposon subfamilies. Identifying more differentially targeted subfamilies potentially opens avenues to learn about potential mechanisms guiding the DNA methylation turnover.

Establishing cellular systems for investigations of the mechanism and function of the DNA methylation turnover in the future: To investigate and understand the mechanism governing the DNA methylation turnover and its functional role in-depth, I aimed to create various cellular systems, encompassing knockout cell lines and lineages with tagged methylation and demethylation players.

By addressing these aims, this work endeavors to contribute to a better understanding of the DNA methylation turnover phenomenon and its significance for cellular differentiation, ultimately enriching our knowledge of epigenetic regulation and its implications for development.

MATERIAL AND METHODS

Experimental Methods and Approaches

Each experiment in this study used the human induced pluripotent stem cell (hiPSC) line ZIP34K14, which was generated from adult male fibroblasts by the Müller Lab at the Zentrum für Integrative Psychiatrie (unpublished) as previously described (Tandon et al., 2018).

Culturing of hiPSCs

The hiPSCs were cultured in mTeSR Plus medium (STEMCELL Technologies, 100-0276), supplemented with a final concentration of 100 U/ml Penicillin/Streptomycin (Thermo Fisher Scientific, 15140122). The cells were cultured on Matrigel-precoated plates (Corning, 354234) under standard conditions of 37°C and 5% CO2. The Matrigel coating was prepared by diluting it at a ratio of 1:100 in Knockout DMEM (Gibco, 10829-018), followed by an incubation period of at least 45 minutes at 37°C, and subsequent washing with DPBS.

For routine maintenance, cells were passaged every 2-3 days in clumps. This was achieved by incubating the cells with a final concentration of 5mM EDTA pH 8.0 (Thermo Fisher Scientific, 15575-038) in DPBS for 5 minutes at 37°C and 5% CO2. After aspirating the EDTA solution, cells were detached through gentle scraping with a spatula, and a portion was transferred to a new plate. During experiments and for generating new cell lines, cells were incubated in mTeSR Plus supplemented with 10uM Y-27632 (Tocris, 1254) prior to single-cell passaging using Accutase (Sigma-Aldrich, A6964) supplemented with 10uM Y-27632. Accutase treatment was performed for 10 minutes at 37°C and 5% CO2, followed by quenching the reaction with twice the volume of mTeSR Plus. The resulting cell suspension was then centrifuged at 300 x g for 5 minutes at room temperature (RT), and the pellet was re-suspended in mTeSR Plus for subsequent procedures. Cell counting was carried out by diluting a cell sample 1:1 in 0.4% Trypan Blue staining solution (Thermo Fisher Scientific, 15250061) and utilizing the Countess II automated cell counter (Thermo Fisher Scientific). The desired cell number was plated onto new wells coated with Matrigel in mTeSR Plus/10uM Y-27632.

For freezing cells, they were detached from the plate in clumps using EDTA, gently scraped, centrifuged, and resuspended in 10% DMSO in Knockout-serum replacement supplemented with 10uM Y-27632. The cell suspension was then stored in a freezing container at -80°C for

a minimum of 24 hours before being transferred to liquid nitrogen for long-term storage. The freezing medium was pre-sterilized and supplemented with Y-27632 before use.

During thawing, cells were placed in a 37°C water bath until most of the ice was melted. Subsequently, at least three times the volume of mTeSR media was added, and centrifugation at 300 x g for two minutes was performed. In order to maintain the clumps, the cell pellet was gently re-suspended in mTeSR media with 10uM Y-27632. The clumps were then plated on a Matrigel-coated dish.

Generation of genetically modified hiPSC cell lines

Molecular cloning

All constructs, including donor plasmids and Cas9/sgRNA-targeting plasmids, were synthesized using Gibson Assembly (NEB, E2621L) at 50°C for 1 hour in a heat block, unless stated otherwise. Following synthesis, the plasmids were introduced into DH5α bacteria. Individual colonies were selected, and plasmid purification was carried out using Mini-Prep (QIAGEN, 27106), followed by Sanger sequencing to validate the accuracy of the sequence.

Molecular cloning of the DNMT3B-FLAG construct

For the donor plasmid intended for homology-directed repair, a Gibson Assembly approach was employed, involving the assembly of five PCR products along with the linearized pUC19 plasmid backbone (XbaI digested; NEB, R0145S) (**Figure 9**). The FLAG-sequence was amplified using PCR primers containing homology overhangs to the GS-linker sequence or to the homology arm downstream of the DNTM3B stop codon, respectively. Amplification of the FLAG sequence was performed using PCR primers containing overhangs to the GS-linker sequence on one end and the homology arm downstream of the DNMT3B stop codon on the other end. The GS-linker sequence was amplified using PCR primers that incorporated overhangs corresponding to the genomic sequence upstream of the DNMT3B stop codon and the FLAG sequence, respectively (**Table 1**). Further, to prevent the re-cutting of the genomic target site post-successful homology-directed repair, an artificial oligo was amplified, containing silent mutations for the PAM and sgRNA sequence, using PCR primers with overhangs corresponding to the homology arm upstream of the DNMT3B stop codon or the GC-linker sequence, respectively. The two homology arms were amplified using genomic DNA from ZIP34K14 and PCR primers containing overhangs for the sequences upstream and downstream of the homology arms (**Table 1**). To facilitate the linearization of the repair

construct and its genomic integration post-transfection, the same PAM and sgRNA used for Cas9-targeting of the endogenous DNMT3B locus were included in the PCR primers for amplifying the homology arms. The PCR products and the linearized pUC19 were subjected to gel extraction and subsequently assembled using Gibson assembly.

The DNA oligo encoding the sgRNA for Cas9-mediated targeting of the C-Terminus of DNMT3B with overhangs homology overhangs to px458 was cloned into px458 by Gibson Assembly. To this end. px458 was linearized by BbsI-HF (NEB, R3539L) and gel extracted.

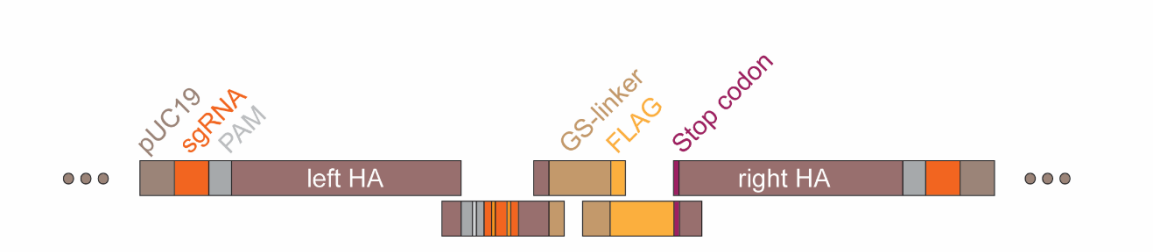

Figure 9: Gibson Assembly for generating the DNMT3B-FLAG donor construct. The schematic illustrates the approach utilized for constructing the DNMT3B-FLAG donor plasmid by Gibson Assembly. The five input fragments were obtained via PCR using distinct primers and templates (

Table 1). The repair construct is flanked by the identical sgRNA and PAM sequences employed for Cas9-mediated targeting of the endogenous DNMT3B locus. Further, the repair construct includes mutations in the PAM and sgRNA sequences (denoted by light grey and light orange breaks in the PAM and sgRNA, respectively) to prevent re-cutting after successful genomic integration.

Primer/oligo name	Sequence 5'-3'
Left_HA_3B_F	ACGCCAAGCTTGCATGCCTGCAGGTCGACTGCGGGCACCACG GCCCATGTTGGTTGGTGCAGATGGCTGACA
Left_HA_3B_R	TGTGTAGTGCACAGGAAAGCC
Right_HA_3B_F	TTCCAGCCAGGCCCCAA
Right_HA_3B_R	ATTCGAGCTCGGTACCCGGGGATCCTCTAGGCGGGCACCACG GCCCATGTTGGGAGCTCCCCAGAGGGTTCTA
FLAG (used as PCR template)	GACTATAAGGACCACGACGGAGACTACAAGGATCATGATATT GATTACAAAGACGATGACGATAAG
FLAG_GC_Linker_F	GGTGGCGGTGGCTCGGGCGGTGGTGGGTCGGACTATAAGGAC CACGAC
FLAG_R	CCCAGTGGGCTTGGGGCCTGGCTGGAACTACTTATCGTCATCG TCTTTG
GC_Linker_overhangs (used as PCR template)	AGGACTACTTTGCATGTGAAGGTGGCGGCGGATCTGGTGGCG GTGGCTCGGGCGGTGGTGGGTCGGACTATAAGGACCACGACG
GC_Linker_F	AGGACTACTTTGCATGTG
GC_Linker_R	CGTCGTGGTCCTTATAG
mutated_CRISPR_target_3B (used as PCR template)	CTAATATGGGACGAGGTGCCCGTCAGAAGCTGCTGGGAAGGT CCTGGAGCGTGCCTGTCATCCGACACCTCTTCGCCCCTCTGAA GGACTACTTTGCATGTGAA
mutated_CRISPR_target_F	CTTTGGCTTTCCTGTGCACTACACAGACGTGTCTAATATGGGA CGAGGTG
mutated_CRISPR_target_R	CGCCGCCACCTTCACATGCAAAGTAGTCC
GA_sgRNA_3B_FLAG	TTATATATCTTGTGGAAAGGACGAAACACCGCGGGCACCACG GCCCATGTGTTTTAGAGCTAGAAATAGCAAGTTAAAAT

sgRNA	Sequence	Cloned into	Published
DNMT3B-FLAG	GCGGGCACCACGGCCCA TGT	px458	-

Table 1: Cloning primers/oligos and sgRNA sequences for generating DNMT3B-FLAG hiPSCs. The table lists the sequences of all primers and oligos used to amplify the fragments for Gibson assembly of the DNMT3B-FLAG donor construct (top). Further, it displays the sgRNA sequence used for Cas9-mediated homology repair (bottom).

Molecular cloning of the TET1-FKBP-FLAG construct

The TET1-FKBP-FLAG repair plasmid was assembled by Gibson Assembly (NEB, E2621L) using the linearized pUC19 as backbone and three fragments, resulting in the incorporation of the following sequences: left homology arm (upstream of Cas9-target site) - mutated Cas9 - target site - GS-linker - FLAG-sequence - GS-linker - FKBP-domain - right homology arm (**Figure 10**). In particular, the individual fragments were obtained through PCR amplification from either oligonucleotides or genomic DNA as template, and pUC19 was linearized through XbaI digest (NEB, R0145S).

For the Cas9-mediated targeting of the N-Terminus of TET1, the DNA sequence encoding the sgRNA was cloned into px459 (linearized by BbsI-HF (NEB, R3539L)), as previously described for incorporating the sgRNA sequence into px458 (Addgene #48138) for generating the DNMT3B-FLAG hiPSC line.

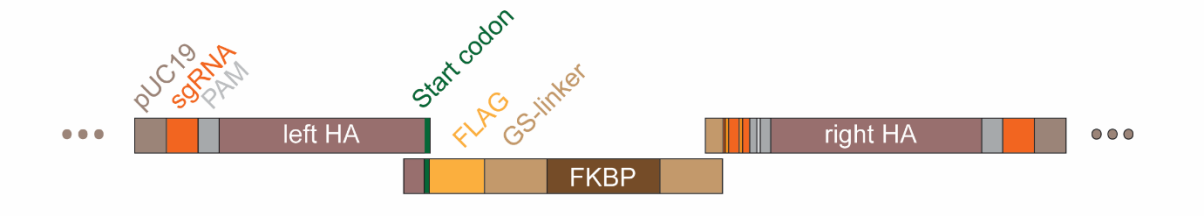

Figure 10: Gibson Assembly for generating the TET1-FKBP-FLAG donor construct. The schematic illustrates the approach utilized for constructing the TET1-FKBP-FLAG donor plasmid by Gibson Assembly. The three input fragments were obtained via PCR using distinct primers and templates (**Table 1**). The repair construct is flanked by the identical sgRNA and PAM sequences employed for Cas9-mediated targeting of the endogenous TET1 locus. Further, the repair construct includes mutations in the PAM and sgRNA sequences (denoted by light grey and light orange breaks in the PAM and sgRNA, respectively) to prevent re-cutting after successful genomic integration.

Primer/oligo name	Sequence 5'-3'
TET1_FKBP_HA_Left_F	CCAAGCTTGCATGCCTGCAGGTCGACTTGTCTCGATCCCGCCATGCAAGGCTGGTCTCAAACTCCTGACCTC
TET1_FKBP_HA_Left_F	AGAGTAGGAAAAAATGAAAGGGCGCAGGAAACAGAGTCATTGGTCCTTTGGA
TET1_FKBP_first_GS_Link_F	GCGCCCTTTCATTTTTTCCTACTCTGTAGCTATGGACTATAAGGACCACGACGGAGAC
TET1_FKBP_first_GS_Link_R	CGACCCACCACCGCCCGAGCCACCGCCACCAGATCCGCCGCCACCTTCCAGTTTTAGAAGCTCCACATCG
TET1_FKBP_HA_Right_F	CGGTGGCTCGGGCGGTGGTGGGTCGTCTAGATCTCGCCATGCAAGACCTTCCAGATTAGTCAGGAAGGAAG
TET1_FKBP_HA_Right_R	TTCGAGCTCGGTACCCGGGGATCCTTGTCTCGATCCCGCCATGCAAGGGGTCGCTTGTTTTGAGCCTG

sgRNA	Sequence	Cloned into	Published
TET1-FKBP-FLAG	TGTCTCGATCCCGCCATGCA	px459	-

Table 2: Cloning primers/oligos and sgRNA sequences for generating TET1-FKBP-FLAG hiPSCs. The table lists the sequences of all primers and oligos used to amplify the fragments for Gibson assembly of the TET1-FKBP-FLAG donor construct (top). Further, it displays the sgRNA sequence used for Cas9-mediated homology repair (bottom).

Molecular cloning of the TET1/2/3 knockout constructs

To establish the TET-TKO hiPSC line, the DNA sequences encoding the sgRNAs designed for the Cas9-mediated targeting of TET1, TET2 and TET3 were integrated into px459 or 2xpx459 using Gibson Assembly (**Table 3**) (Verma et al., 2018). While px459 (Addgene #62988) holds a single gRNA scaffold, 2xpx459 contains a second gRNA scaffold for incorporating an additional sgRNA within a single vector (unpublished by Dr. Alexandro Landshammer). The linearization of px459/2xpx459 was achieved by BbsI digestion, followed by gel extraction for further processing. In the case of integrating the second sgRNA sequence into 2xpx459, the vector containing the first sgRNA was subjected to SapI digestion (NEB, R0569S), followed by gel extraction and a subsequent round of Gibson Assembly.

sgRNA	**Sequence**	**Cloned into**	**Published**
TET1 ex5_1	TATTATACACACCTTGGGGC	2xpx459 (with TET1 ex5_2)	(Verma et al., 2018)
TET1 ex5_2	GGCCCATATTATACACACCT	2xpx459 (with TET1 ex5_1)	(Verma et al., 2018)
TET2 ex3	CTTATGGTCAAATAACGACT	px459	(Verma et al., 2018)
TET3 ex5_1	GTCATCTACACGGGGAAGGA	2xpx459 (with TET3 ex5_2)	(Verma et al., 2018)
TET3 ex5_2	GATCGAGAAGGTCATCTACA	2xpx459 (with TET3 ex5_1)	(Verma et al., 2018)

Table 3: sgRNA sequences for generating TET-TKO hiPSCs. The table lists the sequences of the sgRNA sequences used for establishing TET1/2/3-triple knockout in hiPSCs by Cas9-mediated targeting (Verma et al., 2018).

Molecular cloning of the DNMT3A/B knockout constructs

The DNA sequence corresponding to sgRNAs utilized to induce mutations in the DNMT3A and DNTM3B genes were integrated into the px458 or 2xpx458 vector, employing the same methodology as previously described for the TET knockout constructs. Analogous to 2xpx459, 2xpx458 contains two gRNA scaffolds allowing integration of two sgRNAs into a single vector (unpublished by Dr. Alexandro Landshammer).

sgRNA	Sequence	Cloned into	Published
DNMT3A ex7_1	GATGCTGTGGAAGAAAACCAG	2xpx458 (with DNMT3A ex19)	Simon Lauer (unpublished)
DNMT3A ex7_2	GTGACCCCGCATCCCCCACTG	2xpx458 (with DNMT3A ex19)	Simon Lauer (unpublished)
DNMT3A ex19	GCATGATGCGCGGCCCAAGG	2xpx458 (same vector as DNMT3A ex7_1)	Simon Lauer (unpublished)
DNTM3B ex2_1	GAGACTCGATCCTCGTCAACG	2xpx458 (with DNMT3B ex21)	Simon Lauer (unpublished)
DNMT3B ex2_2	GCGACTCGCCCCCAATCCTGG	2xpx458 (with DNMT3B ex21)	Simon Lauer (unpublished)
DNMT3B ex21	GAATGATAAACTCGAGCTGC	2xpx458 (with DNMT3B ex2_1 or ex2_2)	Simon Lauer (unpublished)

Table 4: sgRNA sequences for generating DNMT3-DKO hiPSCs. The table lists the sequences of the sgRNA sequences used for establishing DNMT3A/B-double knockout in hiPSCs by Cas9-mediated targeting.

Transfection

The respective hiPSC line underwent single-cell passaging (as described previously), and was subsequently plated into individual wells of a 6-well plate coated with Matrigel, in mTeSR Plus/10uM Y-27632, at a density of 300k-500k cells per well. Following attachment to the plate, cells were allowed to grow for 24h. On the subsequent day, the culture medium was changed to a mixture of fresh mTeSR Plus and conditioned mTeSR Plus in a 1:1 ratio, supplemented with 10um Y-27632. The conditioned medium was obtained by allowing the medium to incubate on ZIP34K14 cells for a minimum of 24h, followed by filtration through a 0.2um filter (Sarstedt, 831826001).

The transfection mixture was prepared as follows: Plasmid DNA was introduced into Opti-MEM (Thermo Fisher Scientific, 31985062), with a final volume of 250ul per 6well, prepared

as a master mix to be used across all wells. The amount of plasmid utilized varied among different experiments (**Table 5**). In parallel, 5ul Lipofectamine Stem Transfection reagent (Thermo Fisher Scientific, STEM0003) was added to 245ul of Opti-MEM, similarly set up as a master mix. The DNA/Opti-MEM mixture was added into the Lipo/Opti-MEM mixture, gently mixed by flipping, and incubated for 15min at RT. Each well was exposed to 500ul of the transfection mixture at 37°C, 5%CO2, overnight for a maximum of 24h. For each transfection experiment, between 2 to 6 6xwells were transfected, with an additional control well, lacking plasmid DNA.

Cell line	**Amount of plasmid DNA (per 6xwell)**
DNMT3AKO	1.25ug 2xpx458_sgRNA_DNMT3A_1 1.25ug 2xpx458_sgRNA_DNMT3A_2 Or 2.5ug 2xpx458_sgRNA_DNMT3A_1
DNMT3BKO	1.25ug 2xpx458_sgRNA_DNMT3B_1 1.25ug 2xpx458_sgRNA_DNMT3B_2 Or 2.5ug 2xpx458_sgRNA_DNMT3B_1
TET1+TET3-DKO	2.5ug 2xpx459_sgRNA_TET1_ex5 2.5ug 2xpx459_sgRNA_TET3_ex5
TET-TKO	5ug px459_sgRNA_TET2_ex3
DNMT3-FLAG	1.7ug px458_sgRNA_DNMT3_CTerm 0.8ug donor_DNMT3B_FLAG
TET-FLAG-FKBP	2ug px459_sgRNA_TET1_NTerm 3ug donor_TET1_FLAG_FKBP

Table 5: Amount of transfected plasmid DNA for each transfection experiment. Listed is the amount of plasmid DNA that ZIP34K14 cells have been transfected with in order to generate the different knockout and knockin cell lines.

FACS sorting

Following transfection with either px458 or 2xpx458, the cells were allowed to grow for an additional two days. During this period, the culture medium was exchanged daily, maintaining a mixture of 1:1 fresh and conditioned mTeSR Plus, supplemented with 10uM Y-27632. On the third day post-transfection, the cells were detached from the plate using Accutase treatment, followed by a single wash in DPBS. Subsequently, the cells were resuspended in FACS buffer containing 0.5mM EDTA and 2%FBS in DPBS, with an additional supplementation of 20um Y-27632. The subsequent FACS sorting procedure took place at 4°C, utilizing the Aria II or Aria Fusion (Becton Dickinson). The negative control well was utilized to set the sorting gate.

Following FACS sorting, the single-cell suspension was plated onto dishes that had been pre-coated with Matrigel. The culture medium consisted of a mixture of 1:1 fresh and conditioned mTeSR Plus/20uM Y-27632. Subsequently, the concentration of Y-27632 was gradually reduced over the following days to 10uM and 5uM, eventually transitioning the cells to fresh mTeSR Plus without Y-27632.

Puromycin selection

For cells transfected with px459/2xpx459, the process of selection for puromycin resistance was initiated on the day following transfection. Puromycin-Dihydrochloride was added to the culture medium at a concentration of 2ug/ml (Thermo Fisher Scientific, A1113802), in a mixture of 1:1 fresh and conditioned mTeSR Plus/10uM Y-27632. Based on the observed survival rate of the control cells that had not been transfected with plasmid DNA, the puromycin concentration was either maintained at 2ug/ml or lowered to 1ug/ml the following day. On the third day after selection began, the culture medium was changed to a mixture of 1:1 fresh and conditioned mTeSR Plus/10uM Y-27632, excluding the usage of Puromycin-Dihydrochlorid. Starting the next day, the cells were grown in fresh mTeSR Plus sans Y-27632.

Genotyping

Selected clones were carefully isolated and distributed for both genotyping and further cultivation. For genotyping purposes, genomic DNA was extracted from the clones and genotyping PCR was carried out using the Phire Animal Tissue Direct PCR Kit (Thermo Fisher Scientific, F140WH). The resulting PCR products were loaded onto a 2% agarose gel, and those of interest were extracted from the gel and sent for Sanger sequencing. The obtained

sequencing data were analyzed using the online tool Synthego Performance Analysis, ICE Analysis. Alternatively, PCR products with the desired size were cloned into the pJET1.2 vector (Thermo Fisher Scientific, K1231) and subsequently transformed into DH5α bacteria. From the transformed bacteria, ten colonies were selected for Mini-Prep, followed by Sanger sequencing of the inserted PCR product.

Western Blot

To prepare protein samples for Western blot analysis, cultured cells were incubated with a final concentration of 5mM EDTA pH 8.0 (Thermo Fisher Scientific, 15575-038) in DPBS for 5 minutes at 37°C and 5% CO2. Afterwards, the cells (approximately 3mio) were scraped off the culture plate, followed by a single wash with DPBS, centrifugation at 300 x g for 5min at 4°C and aspiration of the supernatant. Subsequently, the cells were lysed in RIPA buffer (Thermo Fisher Scientific, 89900), supplemented with 1x Halt Protease/Phosphatase inhibitor (Thermo Fisher Scientific, 78443). This lysis step was performed for 30min on ice, followed by centrifugation at 12,000 x g, 10min at 4°C. The protein concentration in the resulting supernatant was determined using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific, 23225), following the manufacturer's instructions.

For the subsequent Western Blot, 20ug of each protein sample was denatured by boiling in a final 1x Laemmli Buffer (BioRad, 1610747), containing 10% 2-Mercaptoethanol (Sigma-Aldrich, M6250) for 10min at 95°C and subsequently cooled on ice for 5min. The prepared samples were then loaded onto a NuPAGE 4-12% Bis-Tris Mini Protein Gel (Thermo Fisher Scientific, NP0322BOX). An initial voltage of 80V was applied for 10min, followed by a 1h run at 130V, and an additional 15min at 130V in 1x NuPAGE MOPS SDS Running Buffer (Thermo Fisher Scientific, NP0001) supplemented with 1:400 NuPAGE Antioxidant (Thermo Fisher Scientific, NP0005). Subsequently, separated proteins were transferred from the gel to a PVDF membrane using the iBlot 2 Starter Kit (Thermo Fisher Scientific, IB21002S) with the P0 program, following the manufacturer's instructions.

Following the transfer, the membrane was blocked with blocking buffer consisting of 5% Blotting-Grade Blocker (BioRad, 1706404) in 1x TBST (Thermo Fisher Scientific, 28360) for 1h at RT. Subsequently, the membrane was incubated with the primary antibody diluted in the blocking buffer overnight at 4°C (**Table 6**). On the following day, the primary antibody was washed off with three 10-minute washes at RT using TBST. The membrane was then incubated

with a secondary antibody coupled to horse radish peroxidase, diluted in blocking buffer for 1h at RT (**Table 6**). After three washing steps with TBST à 10min at RT, the membrane was exposed to SuperSignal West Dura Extended Duration Substrate (Thermo Fisher Scientific, 34075). The chemiluminescence signal, catalyzed by the horse radish peroxidase, was imaged using the BioRad ChemiDoc XRS+ system.

Antibody name	**Company**	**Catalog number**	**Dilution**
Primary antibodies			
Rabbit anti-DNMT3B	Cell Signaling Technologies	#67259	1:1000
Rabbit anti-DNMT3A	Abcam	ab188470	1:2000
Mouse anti-beta Actin	Abcam	ab8226	1:1000
Secondary antibodies			
Donkey anti-rabbit IgG (H+L) Peroxidase AffiniPure	Jackson Immunoresearch	711-035-152	1:10000
Donkey anti-mouse IgG (H+L) Peroxidase AffiniPure	Jackson Immunoresearch	715-035-152	1:10000

Table 6: Antibodies used for Western Blots. The table lists all primary and secondary antibodies used for western blot detection of DNMT3A and DNMT3B and their dilutions in blocking buffer. Anti-beta Actin was used as loading control.

Directed differentiation into CD184+ endoderm and FACS sorting

Directed endoderm differentiation was performed using the STEMdiff Trilineage Differentiation Kit Medium (Stemcell Technologies, 05233). Initially, hiPSCs cultured in mTeSR Plus were treated with Accutase for 10min at 37°C and 5%CO2 to obtain a single-cell suspension. Subsequently, 100k or 1mio cells per well were seeded into 24-well or 6-well plates, respectively, in mTeSR Plus/10uM Y-27632. Control wells, maintained in mTeSR Plus/10uM Y-27632 throughout the differentiation process, were seeded with 50k or 125k cells, respectively.

The following day, differentiation was initiated by switching to STEMdiff Trilienage Endoderm Medium, and the media was refreshed daily for the next three days, according to the manufacturer's protocol. On the fifth day of differentiation, the cells were harvested by Accutase treatment for 10min at 37°C and 5% CO2.

The Accutase reaction was quenched by dilution in FACS-buffer (0.5mM EDTA, 2%FBS in DPBS) and the single-cell suspension was centrifuged at 300 x g for 5min at 4°C; after which the supernatant was discarded. A small fraction of the differentiated, along with the undifferentiated control cells, were stained with an igG control antibody diluted in FACS-buffer (1:200; igG2A,k-ISO-PE antibody (Biolegend, 400218)). The majority of cells was stained with an antibody against CXCR4-PE (1:200, Biolegend, 306506), an endoderm surface marker. Each antibody staining was combined with a DAPI staining (final concentration 5ug/ml) to distinguish between dead and live cells. The cells were incubated with the staining solutions for 15min at 4°C. Afterwards, the cells were washed once in FACS-buffer and resuspended in FACS-buffer in preparation for the final FACS sorting.

FACS sorting was performed as previously described. Briefly, the sorting was conducted at 4°C using the Aria Fusion (Becton Dickinson). The undifferentiated hiPSCs stained for CXCR4 and the igG controls were employed for setting the negative gate for CXCR4-positive sorting. Following FACS sorting, the cells were used in the oxBS/BS Amplicon-Sequencing protocol.

oxBS/BS Amplicon-Sequencing

Bisulfite-Primers were designed using MethPrimer2.0 and reviewed for potential unspecific PCR products using BiSearch (**Table 7**). The genomic DNA from both hiPSCs and CXCR4-positive endoderm cells was isolated using the GeneJET Genomic DNA Purification Kit (Thermo Fisher Scientific, K0722) and then eluted in water.

The bisulfite (BS) and oxidative bisulfite (oxBS) conversion were performed using the TrueMethyl oxBS Module kit (Nugen/Tecan, 0414-32) according to the manufacturer's instructions. After the final elution step of the cleaned-up oxBS/BS-converted DNA, each region of interest was amplified in 4-5xPCR reactions of 27cycles utilizing the bisulfite-primers and the EpiTaq (Takara, R110A). The PCR products from each region of interest were subsequently pooled and subjected to purification using the MinElute PCR Purification Kit (Qiagen, 28004). Eventually, next-generation sequencing libraries were prepared using the NEBNext Ultra II DNA Library Kit (NEB, E7645L), and sequencing was performed using MiSeq.

Genomic location	Closest gene	Size [bp]	Primer F 5'-3'	Primer R 5'-3'
chr5:112539019-112539305	MCC	173	TTTAGTAGTTTTTTA ATGAATTTAGTGATT	AAAACAAAATAATCT AATCTAAACC
chr11:68147852-68148077	LRP5	281	GAGGTAGGTTGGTTT TTTATAGTTG	CTCCAAAAAACCTCA TACAAACAA
chr7:101741050-101741616	CUX1	197	TTGTGGATATATTAA GGTTTTAGATATATT	AAAAATTCCTTTATTC TTCCCTAAC
chr9:137326607-137327145	RXRA	210	GGGTATTTTTTGGTT TTTTGATTTT	ACTACCTACCACCCC CTCCTAT

Table 7: Selected genomic loci for oxBS/BS Amplicon-Seq. The table lists the genomic location (hg19), and the primer pairs used to amplify the genomic regions of interest after bisulfite conversion.

Luciferase enhancer assay

Molecular cloning of Luciferase-enhancer constructs

In the pGL4.27[luc2P/minP/Hygro] (Promega, E845A), regions of interest were inserted upstream of a Firefly luciferase gene driven by a minimal promoter. In particular, pGL4.27 was linearized by EcoRV-HF (*NEB*, R3195S) digestion and the regions of interest were amplified by PCR from genomic DNA of ZIP34K14, using primers with homology overhangs for pGL4.27 (**Table 8**). The final Luciferase-enhancer constructs were synthesized for 1h at 50°C using Gibson Assembly (*NEB*, E2621L). The control pGL4.27-constructs containing eSOX17 or NANOG-3UTR were provided by Dr. Alexandro Landshammer. Following synthesis, all constructs were transformed into DH5α bacteria and individual colonies were selected for plasmid purification, followed by Sanger-sequencing to confirm the sequence.

Performing the luciferase assay

The enhancer assay was conducted differently based on whether it was performed in hiPSCs or endoderm cells:

For performing the enhancer assay in pluripotent cells, hiPSCs were single-cell passaged using Accutase treatment for 10min at 37 °C and 5%CO2. For each region of interest cloned into pGL4.27, 3x24wells were seeded with 200k cells each, to obtain three biological replicates. These cells were cultured overnight in mTeSR Plus/10uM Y-27632. The next day, they were co-transfected with the pGL4.27-construct and pRL-TK (Promega, E2241) which contains a constitutively expressed Renilla luciferase gene for normalization, along with a transfection

control plasmid constitutively expressing mCherry. On the next day, the Dual-Glo Enhancer Assay (Promega, E2940) was performed according to the manufacturer's instructions. In particular, each biological replicate was split into three technical replicates. Luminescence was measured using the pre-installed DualGlo protocol of the GloMax-Multi Detection System (Promega).

Prior to conducting the enhancer assay, a test sample of differentiated endoderm cells was stained for CXCR4-PE (1:200, Biolegend, 306506) and an IgG2a, κ-control (1:200, Biolegend 400218) in order to examine differentiation efficiency. The unprocessed data was then analyzed using FlowJo (Beckton Dickinson, v10.7.2). As the proportion of CXCR4-positive cells was remarkably high, differentiated cells were not sorted before carrying out the enhancer assay.

For performing the enhancer assay in endoderm differentiated cells, hiPSCs cultured in mTeSR Plus were single-cell passaged using Accutase treatment for 10min at 38°C and 5%CO2. Subsequently, 3x24wells were seeded à 100k cells for each pGL4.27-construct. These cells were maintained in mTeSR Plus/10um Y-27632 until the following day. Then, directed endoderm differentiation was performed using the STEMdiff Endoderm Differentiation Medium (Stemcell technologies, 05233) was performed, in accordance with both previous descriptions and the manufacturer's instructions up until day 4. On day 4 of differentiation, cells were co-transfected with the pGL4.27-construct, pRL-TK, and the plasmid of constitutively expressing mCherry, as outlined earlier. On day 5 of differentiation, the Dual-Glo Enhancer assay (Promega, E2940) was performed, according to the manufacturer's protocol. Luminescence was measured using the pre-installed DualGlo protocol of the GloMax-Multi Detection System (Promega).

Name	Genomic location (hg19)	Primer F	Primer R
Hypo_endo1	chr4:6138918-6139386	ACCTGAGCTCGCTAGCCTCGAGGATGGAATCGAGCCACATGCAG	TTGGCCGCCGAGGCCAGATCTTGATAGCCTCCGCATTTCAGGATA
Hypo_endo3	chr13:111193420-111194023	ACCTGAGCTCGCTAGCCTCGAGGATCCACGCTTCAGCCTCTCATC	TTGGCCGCCGAGGCCAGATCTTGATGAAGCTGAGCCCCAGGCAG
eSOX17	chr8:55,136,923-55,138,19	-	-
NANOG-3UTR	chr12:7,951,265-7,951,899	-	-

Table 8: Selected genomic loci for the Luciferase enhancer assay. The table lists the genomic location (hg19), and the primer pairs used to amplify the genomic regions of interest for subsequent cloning into the Firefly-Luciferase plasmid.

Data processing

Initially, the ratio between the Luciferase and the Renilla activity was calculated for each technical replicate (**Figure 11**). Subsequently, the technical replicates were averaged within each biological replicate. Next, the average across the three biological replicates was calculated to obtain one ratio (Firefly/Renilla activity) for each region of interest. Finally, the relative Luciferase activity ratio (LAR) was determined by comparing each vector containing a specific region of interest to the empty vector. A ratio exceeding 1 indicates an enhanced Luciferase expression resulting from the presence of the respective region of interest.

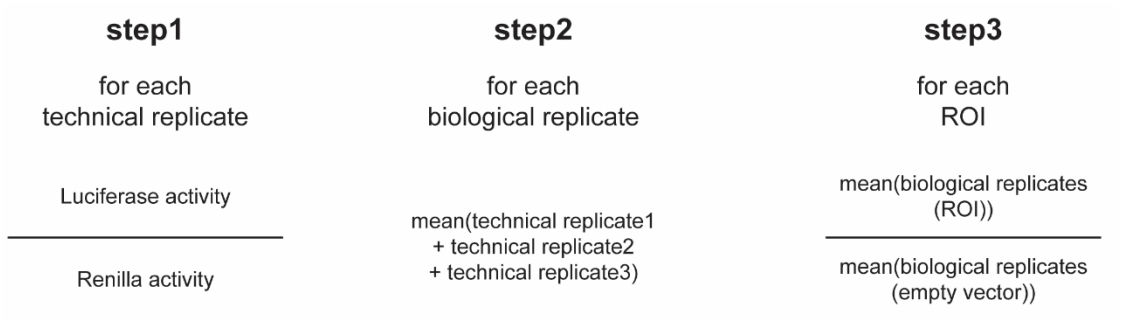

Figure 11: Calculation to obtain the enhancer-based relative luminescence activity ratio (LAR). For each technical replicate, Luciferase and Renilla firefly activity have been experimentally measured. For each technical replicate, the Luciferase activity was divided by the Renilla activity and subsequently, the mean of the technical replicates gives the biological replicate. This procedure was repeated for each region of interest (ROI) and for the empty vector. Eventually, the ratio between the mean of the biological replicate of the ROI and the empty vector informs about the enhancer potential of the ROI.

Computational Methods and Approaches

oxWGBS/WGBS data processing

Published WGBS and oxWGBS data sets were downloaded from the Gene expression Omnibus and underwent a defined processing pipeline (**Table 9**). Initially, raw reads were subjected to quality-filtering and trimming of Illumina adapters using cutadapt (version 2.4; parameters: --quality-cutoff 20 --overlap 5 --minimum-length 25 --adapter illumine) (Martin, 2011). The filtered reads were then aligned to the human genome reference hg19 using BSMAP (version 4.1.4.1.; parameters: -V 2 -v 0.1 -s 16 -q 20 -w 100 -S 1 -u -R) (Xi & Li, 2009). Subsequently, duplicate reads were removed using MarkDuplicates from GATK (version 4.1.4.1.; parameters: coordinate --REMOVE_DUPLICATES true --VALIDATION_STRINGENCY LENIENT) (McKenna et al., 2010). Next, the mcall command from MOABS was employed to compute methylation levels for individual CpGs (version 1.3.2.; default parameters) (Sun et al., 2014). Eventually, CpGs located on chromosome X, Y, or M were excluded, as were those with a coverage below 10 and greater than 150.

Cell type	Cell line	Data type	GEO accession number	Paper
hESC WT	HUES64	WGBS	GSM4458668	(Charlton et al., 2020)
hESC WT	HUES8	WGBS	GSM3618718	(Charlton et al., 2020)
hESC DKO	HUES64	WGBS	GSM4458669	(Charlton et al., 2020)
hESC DKO	HUES64	WGBS	GSM4458670	(Charlton et al., 2020)
hESC DKO	HUES8	WGBS	GSM4458672	(Charlton et al., 2020)
DE	HUES8	WGBS	GSM4221047	(Alvarez-Dominguez et al., 2020)
PP1	HUES8	WGBS	GSM4221048	(Alvarez-Dominguez et al., 2020)
PP2	HUES8	WGBS	GSM4221049	(Alvarez-Dominguez et al., 2020)
EN	HUES8	WGBS	GSM4221050	(Alvarez-Dominguez et al., 2020)
Scbeta	HUES8	WGBS	GSM4221051	(Alvarez-Dominguez et al., 2020)
hESC	H9	WGBS	GSM1521762	(Ziller et al., 2015)
NE	H9	WGBS	GSM1521763	(Ziller et al., 2015)
ERG	H9	WGBS	GSM1521764	(Ziller et al., 2015)
MRG	H9	WGBS	GSM1521765	(Ziller et al., 2015)
ECTO	HUES64	WGBS	GSM1112820	(Gifford et al., 2013)
ENDO	HUES64	WGBS	GSM1112848	(Gifford et al., 2013)
MESO	HUES64	WGBS	GSM1112839	(Gifford et al., 2013)
MN day16	HUES64	WGBS	GSM4368689	(Charlton et al., 2020)
MN day16	HUES64	oxWGBS	GSM4368690	(Charlton et al., 2020)
MN day60	HUES64	WGBS	GSM4368691	(Charlton et al., 2020)
MN day60	HUES64	oxWGBS	GSM4368692	(Charlton et al., 2020)
Liver	Adult tissue	WGBS	GSM1058027	(Roadmap Epigenomics Consortium et al., 2015)
Lung	Adult tissue	WGBS	GSM983647	(Roadmap Epigenomics Consortium et al., 2015)
Spleen	Adult tissue	WGBS	GSM983652	(Roadmap Epigenomics Consortium et al., 2015)
Small intestine	Adult tissue	WGBS	GSM983646	(Roadmap Epigenomics Consortium et al., 2015)

Table 9: Utilized published WGBS data sets. The table lists the cell types, originating cell lines and the GEO accession number of publicly available WGBS data sets which have been used in this thesis. hESC (human embryonic stem cells).

oxBS/BS Amplicon-Seq data processing

oxBS/BS Amplicon-Seq data were processed using the same workflow that has been employed for oxWGBS/WGBS processing, except for the last step, which removes all CpGs with coverage greater than 150 (see previous paragraph).

Calculating 5hmC levels

To determine 5hmC levels in both whole-genome or Amplicon-Seq data, the levels derived from oxBS (5mC) were subtracted from those of BS (5mC+5hmC). This calculated difference corresponds to 5hmC levels.

Calling of DKO-DMRs

First, WGBS data for HUES64 WT, HUES8 WT, HUES64 DKO-cloneA, HUES64 DKO-cloneB and HUES8 DKO were filtered to include only CpGs which were covered in all the data sets. Next, differentially methylated regions (DMRs) were called using METILENE (v0.2-8; parameters: -d 0.2 -c 2) and only regions with FDR<0.05 were retained (Jühling et al., 2016). Lastly, among the hypomethylated DMRs, those with a WT DNA methylation >=0.7 and Δmeth of <-0.6 were selected.

Calling of tissueDMRs and splitting them according to their DNA methylation turnover behavior

The primary procedure to identify DMRs between hPSCs and in vitro differentiated cell types, as well as adult tissues, followed the same methodology as the identification of DKO-DMRs. Initially, WGBS data from the specific differentiated cell type and from the respective progenitor hPSC line was merged into a single file. The average methylation levels of HUES64 WT and HUES8 WT were used as the hPSC reference in the adult tissue comparison. For calling tissueDMRs between day 16 and day 60 of MN differentiation, the WGBS of those two cell states were integrated instead of hPSC data. Only CpGs that were covered in both datasets were retained. Subsequently, DMRs were called using METILENE (v0.2-8; parameters: -d 0.2

-c 2)and those meeting the criteria of FDR <0.05, a WT methylation level >= 0.3, and a Δmeth of <-0.3 were selected for further analysis (Jühling et al., 2016). Further, for in vitro stepwise differentiated cell types, I omitted DMRs called in a given cell type that had already emerged in progenitor cells. To achieve this, I called a set of DMRs for each cell type of the pancreatic and radial glial protocol, applying less stringent criteria (FDR<1.0; WT DNAmeth >=0.3; Δmeth <-0.3). Using bedtools intersect (v2.30.0; parameters: -v), I excluded those DMRs from the set of tissueDMRs of a given cell type, which were identified as "loose DMR" in any progenitor cell type (Quinlan & Hall, 2010). Moreover, bedtools intersect was generally used to determine overlaps between tissueDMRs of different cell types (v2.30.0; parameters: -u), such as for the three germ layers, and to identify DMRs that are specific to a cell type cell t (v2.30.0; parameters: -v).

To identify those tissueDMRs which are targeted by the DNA methylation turnover in hPSCs, I determined the average DNA methylation level in WT and DKO hPSCs at each DMR, using the UCSC tool bigWigAverageOverBed (v2; default parameters). By calculating the difference between DKO-WT hPSCs, I categorized tissueDMRs into "turnover" (Δmeth<-0.3), "no turnover" (Δmeth>-0.2) and "in between" (Δmeth>=-0.3 and <=-0.2); except for MN-D60-DMRs, which were split into "turnover" (Δmeth<-0.3), "no turnover" (Δmeth>=-0.3).

Calculating Δmethylation at DKO-DMRs between two cell types

To determine the rate of DNA methylation loss at DKO-DMRs during differentiation, I utilized the UCSC tool bigWigAverageOverBed (v2; default parameters) to determine the average DNA methylation level in hPSCs and a differentiated cell type or adult tissue. As described above, I then determined the delta between the differentiated cell type-hPSCs.

Determining the proportion of repeats with DNA methylation turnover within a family or subfamily of retrotransposons

All analyses regarding retrotransposons and simple repeats are based on the UCSC RepeatMasker annotations for hg19. Repeats were filtered for containing at least two CpGs and those which have, were split into "methylated" (DNAmeth WT hPSCs >0.3) and "unmethylated" (DNAmeth WT hPSCs <=0.3) using the UCSC tool bigWigAverageOverBed (v2; default parameters). Next, I excluded all subfamilies with less than 30 methylated repeat elements. Further, the average DNA methylation level in DKO hPSCs at repeats was

determined and for further analysis, the proportion of elements with turnover (DKO-WT hPSC Δmeth<-0.3) was considered.

In the special case of LINE1, I additionally extracted full-length elements with a size of at least 4.5kb. The filtering was slightly less stringent, as full-length elements often range between 5kb-7kb (Swergold, 1990). The first sixth of each full-length element was defined as 5'end.

The CpG-density at repeat elements was determined using the countPattern function from the Biostrings package in R (v2.66.0) (Pagès et al., 2022). The GC content (excluding CpGs) was calculated by first deleting CpGs from the fasta sequences using the gsub function in R, followed by counting Gs and Cs.

ATAC-seq data processing

Raw reads of one ATAC-seq replicate for hiPSCs (Philine Guckelberger, unpublished data) were subjected to quality and adaptor trimming using cutadapt (version 2.4; parameters: --quality-cutoff 20 --overlap 5 --minimum-length 25) (Martin, 2011). Alignment of the filtered reads to the human genome reference hg19 was done using the 'mem' command of the BWA software package (version 0.7.17; default parameters) (H. Li & Durbin, 2009). Duplicates were filtered out using GATK (version 4.1.4.1.; parameters: --REMOVE_DUPLICATES true --VALIDATION_STRINGENCY LENIENT) (McKenna et al., 2010). Narrow peaks were called using MACS2 (version 2.1.2; default parameters) (Zhang et al., 2008). To identify DKO-DMRs overlapping at least one ATAC-peak, I utilized bedtools intersect (version 2.30.0; parameters: -u). For generating profile plots of the ATAC-seq signal, the signal was normalized to genomic features using the function 'normalizeToMatrix' from the R package EnrichedHeatmap (parameters: extend = c(1500, 1500), mean_mode = "wo", w = 10, target_ratio = 0.25) (Gu et al., 2018).

ChIP-seq data processing

ChIP-seq data sets were downloaded from the Gene expression omnibus and processed according to the procedure described for ATAC-seq data, with the exception of calling broad instead of narrow peaks using the additional parameter '--broad' in MACS2 (version 2.1.2) (**Table 10**) (Zhang et al., 2008). To determine intersections between DKO-DMRs and GSM1521721 or GSM772800, I utilized bedtools intersect (version 2.30.0; parameters: -u).

Target	Cell type	Cell line	GEO accession number	Publication
H3K27Ac	hESC	H1	GSM466732	(Heintzman et al., 2009)
H3K27Ac	hESC	H9	GSM1521721	(Roadmap Epigenomics Consortium et al., 2015)
H3K27Ac	ENDO	HUES64	GSM1112831	(Gifford et al., 2013)
H3K4me1	hESC	HUES64	GSM772800	(Roadmap Epigenomics Consortium et al., 2015)

Table 10: Utilized published ChIP-seq data sets. The table lists the cell types, originating cell lines and the GEO accession number of publicly available ChIP-seq data sets which have been used in this thesis.

Data visualization

Violin plots, boxplots, scatter plots, bar graphs, and pie charts were created using the R package ggplot2 (Wickham, 2016).

RESULTS

DKO-DMRs are functional enhancers in hepatic endoderm cells

Regions that become actively demethylated by TETs in the absence of *de novo* DNA methyltransferases in hPSCs (DKO-DMRs), co-recruit DNMT3s and TETs in wildtype hPSCs, resulting in the DNA methylation turnover. The majority of DKO-DMRs overlap with putative somatic enhancers, while the overlap with pluripotent-specific enhancers is minor (Charlton et al., 2020). Here, the definition of putative somatic enhancers is based on the presence of the active enhancer mark H3K27Ac in a somatic cell context, often cell type-specific (Hnisz et al., 2013). In contrast, only a minority of DKO-DMRs are enriched for H3K27Ac in hPSCs, indicating an inactive status of DKO-DMRs in hPSCs but a potential activation in certain somatic tissues. Further, the enrichment of H3K4me1 points towards a poised chromatin state at DKO-DMRs in hPSCs (Charlton et al., 2020). In line with this, high levels of DNA methylation, which is one characteristic of DKO-DMRs, are generally associated with inactive enhancers (Stadler et al., 2011). Moreover, in the context of DNA methylation turnover, the loss of DNMT3s in hPSCs was shown to cause changes in the transcriptome, with two-thirds of differentially expressed genes being associated with DKO-DMRs (Charlton et al., 2020). However, H3K27Ac is only a predictor of enhancer activity and it has been shown that some H3K27Ac-positive regions do not enhance gene expression in an experimental setup (Zhu et al., 2013). To investigate whether DKO-DMRs have the potential to drive gene expression, I conducted a Luciferase enhancer assay for three selected DKO-DMRs in human pluripotent stem cells (hPSCs) and *in vitro* differentiated cells.

The DKO-DMRs were selected based on their DNA methylation levels, H3K27Ac signal, and chromatin accessibility in hPSCs and in vitro differentiated hepatic endoderm cells (ENDO) (**Figure 12**). Based on publicly available WGBS data sets for hPSCs and ENDO, I selected DKO-DMRs which are hypomethylated upon differentiation, indicative of potential enhancer activation (Charlton et al., 2020; Gifford et al., 2013). A second indicator for somatic enhancer activity is the opening of chromatin upon differentiation, suggesting facilitated transcription factor binding. Based on ATAC-seq data, I refined the selection of candidate DKO-DMRs to those situated within accessible chromatin linked to ENDO differentiation (data unpublished, generated by Philine Guckelberger) (**Figure 12**). Furthermore, two of the three selected DKO-DMRs display an increased enrichment of H3K27Ac upon differentiation based on publicly

available H3K27Ac ChIP-seq data for hPSCs and ENDO (DKO-DMR1+3) (Heintzman et al., 2009). The DKO-DMRs1-3 that were ultimately chosen are situated in the first intron of JAKMIP1 (DMR1), 80kb/155kbp upstream of DTWD1/ATP8B4 (DMR2), and the RAB20 intron (DMR3), respectively. (**Figure 12**). None of these genes are known to play a specific role for endoderm differentiation; however, enhancers do not necessarily regulate the most proximal gene promoter (Tang et al., 2015b; Rao et al., 2014). Instead, JAKMIP1 is involved in neuronal translation regulation and RAB20 plays a role in controlling neurite outgrowth, while DTWD1 is a tumor suppressor (Oguchi et al., 2018; Ma et al., 2015; Berg et al., 2015). However, long-range promoter-enhancer interactions can occur and thus, the potential enhancer function of the selected DKO-DMRs might affect other genes than their nearest neighbor. Along with DMR1-3, a confirmed enhancer in hPSC-derived ENDO cells located 240kb upstream of the SOX17 gene served as a positive control for enhancer activity (referred to as eSOX17) (Landshammer et al., 2023) (**Figure 12**). As a potential positive control in hPSCs, I included a region located in NANOG's 3'UTR (referred to as NANOG-3UTR) that is substantially enriched for H3K27Ac and exhibits an ATAC-peak in hPSCs. The regions of interest were tested for enhancer activity in a Dual-Luciferase assay to assess their enhancer-based relative luminescence activity ratio (LAR) in hPSCs and hPSC-derived ENDO cells (**Figure 11**).

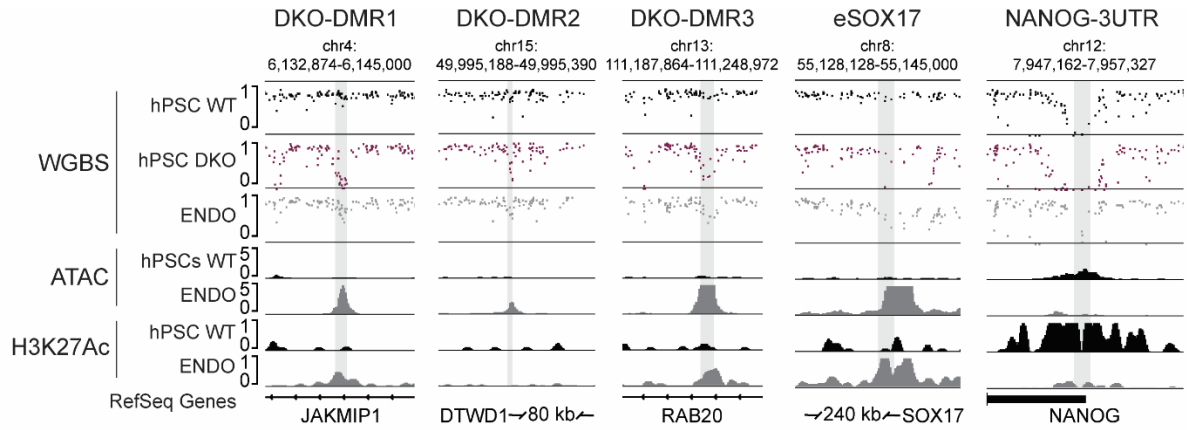

Figure 12: Selection of DKO-DMRs for enhancer assay based on epigenetic features. The browser tracks display the location of DKO-DMRs within the genome (light-grey bars) and the DNA methylation levels, ATAC-signal and H3K27Ac-signal at the DKO-DMRs and their flanking regions in hPSCs (black) and hPSC-derived ENDO cells (grey). Further, the second WGBS track (purple) displays hypomethylation at DKO-DMRs in DKO hPSCs as confirmation for the local DNA methylation turnover.

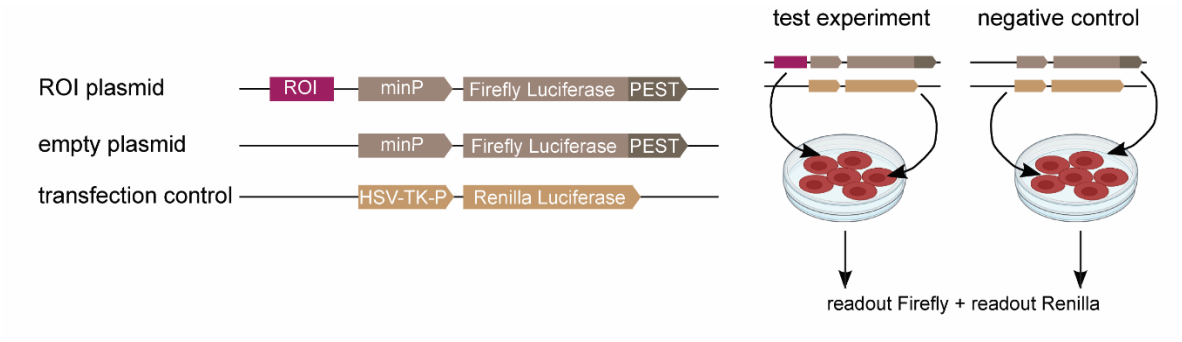

Figure 13: Schematic of the luciferase enhancer assay. The schematic displays the two vector systems of the Luciferase assay. One contains the Firefly luciferase gene, fused to a PEST domain and driven by a minimal promoter (minP). The PEST-sequence fused to the Firefly luciferase destabilizes the enzyme for assessing the current enhancer activity (Rechsteiner & Rogers, 1996). The regions of interest (ROI) are cloned upstream of the minP to assess their enhancing activity onto the minP. In parallel, an empty vector without a region of interest was used to measure the leakiness of the minP. For normalizing differences in transfection efficiency of different ROI-Firefly-vectors, a second vector with a Renilla luciferase gene under the control of the constitutively active HSV-TK-promoter was co-transfected with each test vector. See also **Figure 11** for a detailed explanation on how to calculate the relative luminescence activity ratio (LAR).

NANOG-3UTR shows an increased LAR in hPSCs but only by 3.34-fold. Compared to ENDO-enhancer eSOX17 which doubles reporter gene activity in hPSCs, NANOG-3UTR enhancer activity is only slightly higher, indicative for NANOG-3UTR being a weak enhancer in hPSCs (**Figure 14**). To assess enhancer activity in hPSC-derived ENDO cells, I efficiently differentiated hPSCs into ENDO (97.7% CXCR4+ cells), and performed luciferase enhancer assay (**Figure 14** A). As expected, eSOX17 enhanced reporter activity immensely (LAR 109) in ENDO cells (**Figure 14** B). In contrast, NANOG-3UTR displayed minor enhancer activity in ENDO, similar to the LAR observed in hPSC (**Figure 14** B), suggesting a generally weak enhancer potential of the NANOG-3UTR.

Similar to NANOG-3UTR, DKO-DMR2 increases reporter activity almost equally (3.02 LAR), indicating a weak enhancer activity in hPSCs as well (**Figure 14** B). In contrast, reporter activity is almost unchanged between the presence and absence of DKO-DMR1 and DKO-DMR3 (1.28 and 1.27, respectively), speaking against them being enhancers in hPSCs (**Figure 14** B). Interestingly, in ENDO cells, all three DKO-DMRs increase reporter gene activity more than NANOG-3UTR does. Most strikingly the presence of DKO-DMR1 upregulates reporter gene activity by 88-fold which lays below the enhancer potential of eSOX17 but substantially exceeds NANOG-3UTR, indicating a strong enhancing capability of DKO-DMR1 in ENDO

(**Figure 14** B). DKO-DMR2 and DKO-DMR3 enhance reporter gene activity as well but to a lesser extent (12 LAR and 19 LAR, respectively) (**Figure 14** B).

In conclusion, the in vitro enhancer experiment demonstrated that all three DKO-DMRs have substantial enhancer activity in ENDO. In contrast, only DKO-DMR1 functions as a weak enhancer in hPSCs. This observation shows that DKO-DMRs may operate as somatic rather than pluripotency enhancers, which is consistent with enrichment for the poised enhancer mark H3K4m1 at DKO-DMRs in hPSCs.

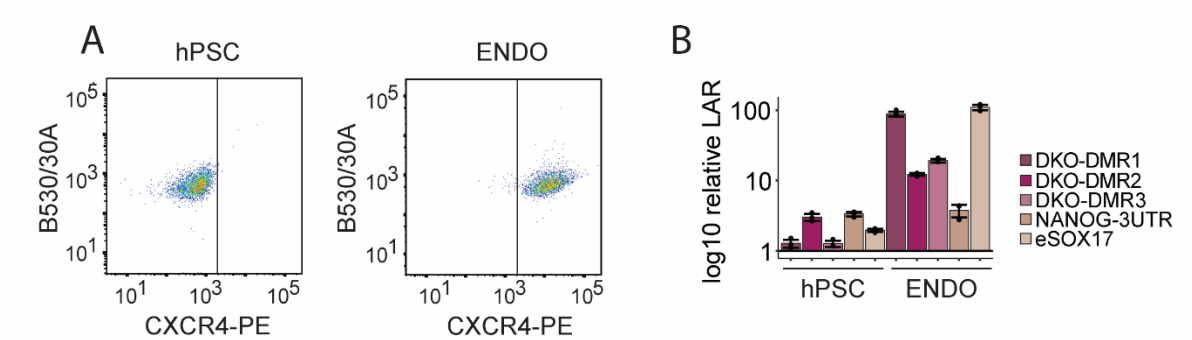

Figure 14: DKO-DMRs display enhancer activity in ENDO cells. A The FACS plot displays the efficiency of in vitro differentiation of ENDO cells derived from hPSCs. **B** The bar graph displays the relative luminescence activity ratio (LAR) (log10) of the Firefly luciferase activity over the transfection control in comparison to the empty plasmid over the transfection control, in hPSCs and ENDO cells. Error bars display the standard deviation.

Investigating the relationship between the DNA methylation turnover in hPSCs and differentiation-related hypomethylation

Upon three germ layer differentiation local hyper- and hypomethylation events are happening, of which the majority is lineage-specific, showing that pluripotent DNA methylation patterns are plastic during development and can already undergo changes upon the first lineage decisions (Gifford et al., 2013). Interestingly, DNA demethylation mainly occurs at intergenic regions and regions enriched for H3K4me1 or H3K27Ac which coincides with DKO-DMRs features (Charlton et al., 2020). This opens the hypothesis that DKO-DMRs might undergo demethylation upon trilineage differentiation as well. Moreover, since the DNA methylation turnover is defined by the competition between TETs and DNMT3s at highly methylated regions this DNA methylation dynamic could serve as the driving motor to maintain local DNA methylation plasticity in hPSCs. Beyond, local alterations of the DNA methylation landscape are not restricted to early differentiation but also occur later during development (Lowdon et al., 2014; Alvarez-Dominguez et al., 2020) which expands the question of association between the DNA methylation turnover in hPSCs and differentiation-related hypomethylation to later stages of development.

To investigate the relationship between the DNA methylation turnover in hPSCs and hypomethylation upon differentiation, I pursued a two-sided approach. First, I focused on DKO-DMRs aiming to reveal potential DNA methylation changes upon differentiation into cell types that arise early after the exit of pluripotency and in mature tissues, using published WGBS data of *in vitro* differentiated cells of the three germ layers and *in vivo* adult tissues. On the other hand, I investigated to which extent regions which become hypomethylated upon differentiation (referred to as tissueDMRs) happen to be subject to the DNA methylation turnover in hPSCs. Here, I focused on the three germ layers and on stepwise differentiated more mature cell types. For consistency and comparability, I re-called DKO-DMRs (in the following DKO-DMRs refers to the new set) and tissueDMRs of the different cell types and tissues using the published pipeline METILENE (Jühling et al., 2016). The original calling strategies distinguish by the use of diverse statistical tests for examining significant DNA methylation changes, as well as varied criteria for merging DMRs and the minimum number of CpGs contained in a DMR (Gifford et al., 2013; Ziller et al., 2015; Alvarez-Dominguez et al., 2020; Charlton et al., 2020).

Identification of a refined set of DKO-DMRs and differentiation-induced hypomethylated regions

Calling DKO-DMRs with METILENE reveals a larger set of DNA methylation turnover regions in hPSCs

As a quality check for the newly called DKO-DMRs using *METILENE*, I examined chromatin features that are characteristic of DKO-DMRs, like the strong overlap with putative somatic enhancers, an increased chromatin accessibility and enrichment of poised enhancer marks (Charlton et al., 2020).

Both sets of DKO-DMRs contain ~ 11k regions (**Figure 15** A). Interestingly, even though the original and the new sets were both filtered for a minimum loss of 0.6 DNA methylation, only half of the regions of each set were also found in the other. Thus, with the new approach, I was able to identify ~5k turnover regions in hPSCs that were not included in the original set, suggesting that the DNA methylation turnover targets far more genomic loci than initially proposed. This reveals that based on the calling strategy for DMRs, other sets of regions are defined to become significantly hypomethylated in early passage DKO hPSCs, suggesting that the DNA methylation turnover targets far more genomic loci than initially proposed (**Figure 15** A) (Charlton et al., 2020). Yet, both sets display a highly similar frequency to overlap with peaks of H3K4me1, H3K27Ac, or both, suggesting a similar chromatin context (**Figure 15** B). In line with the strong enrichment for H3K4me1 in both DKO-DMR sets, more than 85% of the DMRs are located in a putative somatic enhancer, indicative of the DKO-DMRs of both sets being in a poised enhancer state (**Figure 15** C). Additionally, around half of the regions of the original DKO-DMR set and 40% of the new set overlap with an ATAC-peak (**Figure 15** D). However, the ATAC-signal at both DKO-DMR sets is relatively weak compared to the signal of ATAC-peaks at unmethylated loci (**Figure 15** E).

First of all, the newly called DKO-DMRs display similar chromatin features and a strong overlap with putative somatic enhancers and thus, represent the original DKO-DMRs well. However, calling DKO-DMRs with the METILENE pipeline revealed a more frequent occurrence of immediate strong loss of DNA methylation upon DNMT3 depletion, indicative of a broader TET activity in hPSCs than initially assumed.

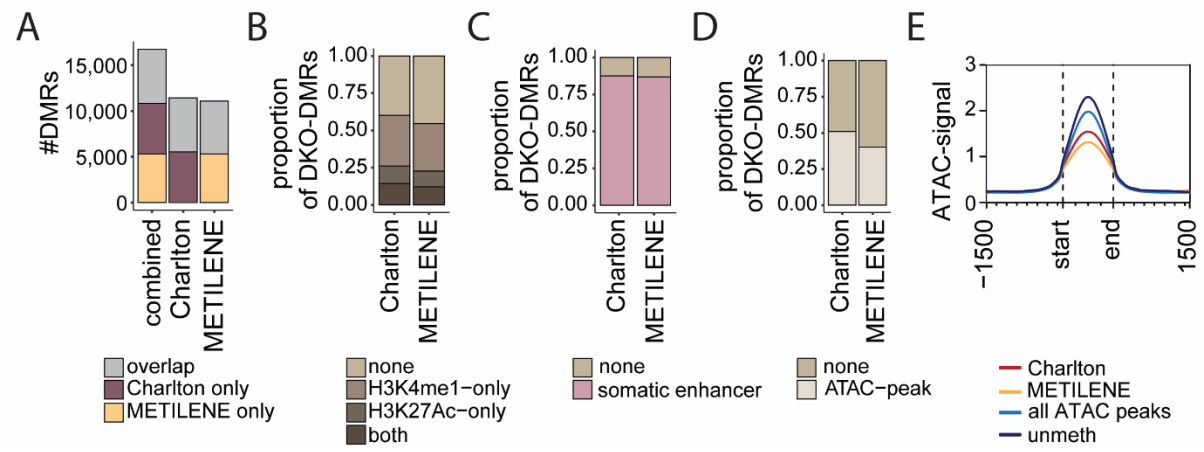

Figure 15: The new set of DKO-DMRs lies in putative somatic enhancers. A The bar chart displays the overlap between the original set of DKO-DMRs (Charlton et al., 2020) and the ones called with METILENE and vice versa. **B+C+D** The bar graphs compare the overlap of the two DMR sets with ChIP-seq peaks of H3K4me1, H3K27Ac and both (B), putative somatic enhancers (C) (Hnisz et al., 2013) and ATAC-seq peaks (D). **E** The profile plot displays the ATAC-signal of all ATAC-peaks and their neighboring regions 1.5kb up- and downstream. Furthermore, the plot displays the ATAC-signal at ATAC-peaks specifically overlapping with DKO-DMRs (Charlton et al., 2020), the METILENE-DKO-DMRs, and at unmethylated loci.

Re-called three germ layer tissueDMRs reproduce hypomethylation bias towards ECTO cells

As a basis to dissect the relationship between the DNA methylation turnover in hPSCs and early differentiation-related hypomethylation, I called DMRs which become hypomethylated (<-0.3) upon differentiation of hPSCs into the three germ layers, CD56+ ectoderm (ECTO), CD184+ endoderm (ENDO) (the same as hepatic ENDO) and CD56+ mesoderm (MESO) (**Figure 16**).

ECTO displays more than twice as many hypomethylated DMRs as ENDO and MESO (1,564, 586, and 718, respectively), with only a minority found in all three lineages (262) (**Figure 16** B) (Gifford et al., 2013). Similarly, a small number of tissueDMRs is only found in two of the three germ layers (ECTO/ENDO: 113; ECTO/MESO: 152; MESO/ENDO: 59), showing that the number of shared tissueDMRs is similar between the three germ layers. The strong difference in the total number of ectoDMRs compared to endoDMRs and mesoDMRs arises from loci which are uniquely hypomethylated in one lineage only (**Figure 16** B). In ECTO, the unique tissueDMRs represent the biggest group with 66.1% (1037), while ENDO and MESO have 23.7% (152) and 33.8% (245) uniquely hypomethylated loci, respectively (**Figure 16** B) (Gifford et al., 2013).

In line with the originally called trilineage tissueDMRs, the new sets reproduce that ECTO cells most frequently undergo local demethylation as compared to ENDO and MESO where the majority of hypomethylated DMRs are shared with at least one other of the three germ layers (Gifford et al., 2013).

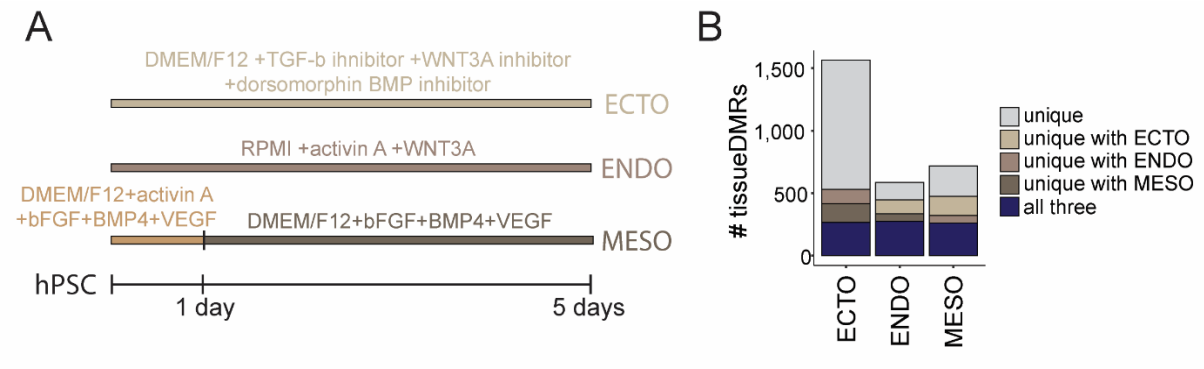

Figure 16: Frequent hypomethylation upon ECTO differentiation. A The timeline displays the differentiation protocols for obtaining hPSC-derived CD56+ ectoderm (ECTO), CD184+ endoderm (ENDO), and CD56+ mesoderm (MESO) cells. **B** The bar graph shows the number of hypomethylated DMRs in each of the three germ layers and indicates which ones are shared between lineages and which are unique to only one lineage.

Accumulation of hypomethylated DMRs in adult tissues

As a counterpart to investigate DNA methylation changes at DKO-DMRs upon early differentiation represented by the three germ layers, I further examined the methylome of adult human tissues. To this end, I called hypomethylated DMRs, utilizing published WGBS data sets of liver, lung, spleen, and the small intestine (Roadmap Epigenomics Consortium et al., 2015).

In each of the investigated tissues, 24.5k-28k regions were defined to be hypomethylated compared to hPSCs, which indicates an accumulation of hypomethylation events over development when comparing to the three germ layers (**Figure 16** B, **Figure 17** A). Further, at least 20% of adult tissueDMRs are unique to their respective tissue wherein tissue specificity is again shown.

In summary, adult tissueDMRs were successfully defined and can be used for analysis regarding DNA methylation changes at DKO-DMRs in mature cells.

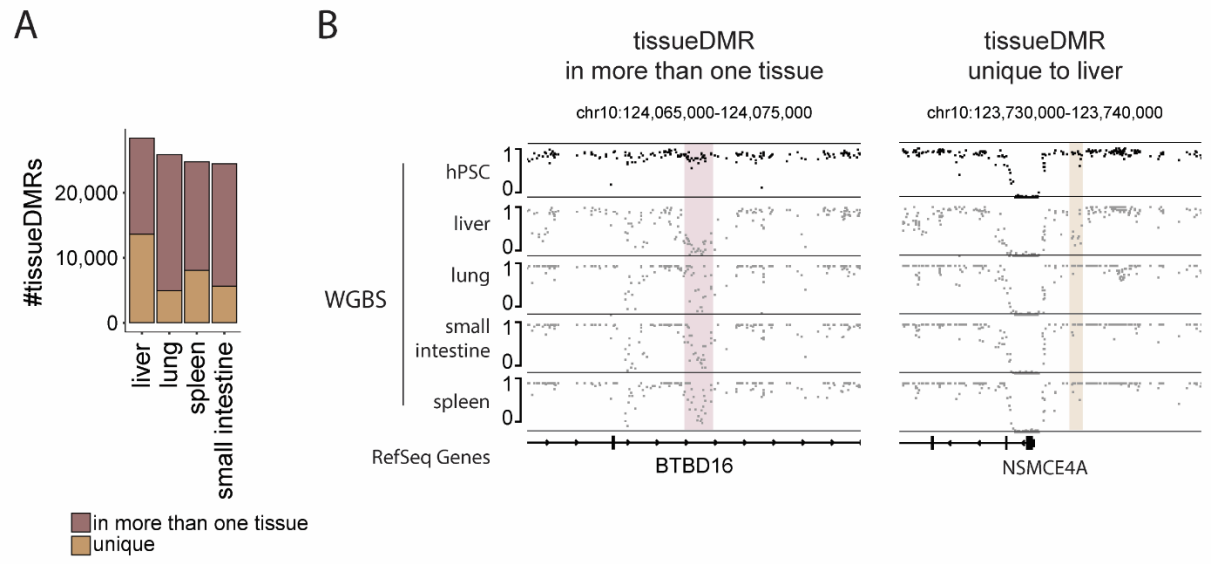

Figure 17: Calling hypomethylated DMRs in adult human tissues. A The bar graph displays hypomethylated DMRs called between hPSCs and adult tissues, with the colors indicating whether a DMR is found in one or multiple tissues. **B** The genome browser images display examples of tissueDMRs, which are shared between the adult tissues (left) and which are unique to only one of them (right).

Calling tissueDMRs for stepwise in vitro differentiation of pancreatic and radial glial cells

To investigate whether the relationship between differentiation-related hypomethylation and DNA methylation turnover in hPSCs is dependent on the developmental progress of the differentiated cell type, I called hypomethylated DMRs for stepwise differentiation of hPSCs into mid radial glial cells (MRG) and stem cell-derived beta cells (SCbeta), representing neuroectodermal and endodermal lineage decisions, respectively (Ziller et al., 2015; Alvarez-Dominguez et al., 2020). For adequate correlation between DNA methylation turnover and distinct states of cell maturation, I only included hypomethylated DMRs that can be properly attributed to one step of differentiation.

In the earliest measured step of radial glial differentiation, between hPSCs and neuroepithelial cells (NE; 12 days), ~300 regions are hypomethylated. A similar number of hypomethylated DMRs occur between hPSCs and early radial glial cells (ERGs), which are differentiated for two more days (14 days total). However, only eight of the ERG-DMRs can be assigned as *de novo* hypomethylated in ERG, while the majority already undergoes slight demethylation in NE which was however not sufficient to be called as NE-DMR (**Figure 18** B). In contrast, ~2k genomic loci significantly lose DNA methylation between hPSCs and MRG cells (35 days), of which ~700 are *de novo* hypomethylated from ERG to MRG (**Figure 18** B, C). During *in vitro*

pancreatic differentiation, almost 200 regions become hypomethylated during the first three days of differentiation into pancreatic endoderm (DE; 3 days), while the majority of DMRs between hPSCs and the intermediate cell types pancreatic progenitor 1 (PP1; 8 days), pancreatic progenitor 2 (PP2; 14 days) and endocrine cells (EN; 18 days) either already occur in their respective progenitors or become demethylated over multiple differentiation steps (**Figure 18** B, C). Eventually, in the last step of stem cell-derived beta cell (SCbeta; 25+ days) derivation, an additional 340 DMRs were demethylated. A minority of all DMRs between hPSCs and SCbeta or MRG, respectively, overlap with each other (4.5% and 1.7%, respectively), indicating cell type-specific DNA methylation changes (**Figure 18** B). Importantly, the cellular identity of pancreatic endoderm (DE) differs from the previously investigated hepatic endoderm (ENDO) (Alvarez-Dominguez et al., 2020; Gifford et al., 2013; Toivonen et al., 2013).

In summary, for two independent successive differentiation systems, I defined lists of regions which become uniquely hypomethylated during one of the differentiation steps but in none of the prior ones. This will allow to reveal whether there is a bias for early or late differentiation-related hypomethylation to be associated with the DNA methylation turnover in hPSCs.

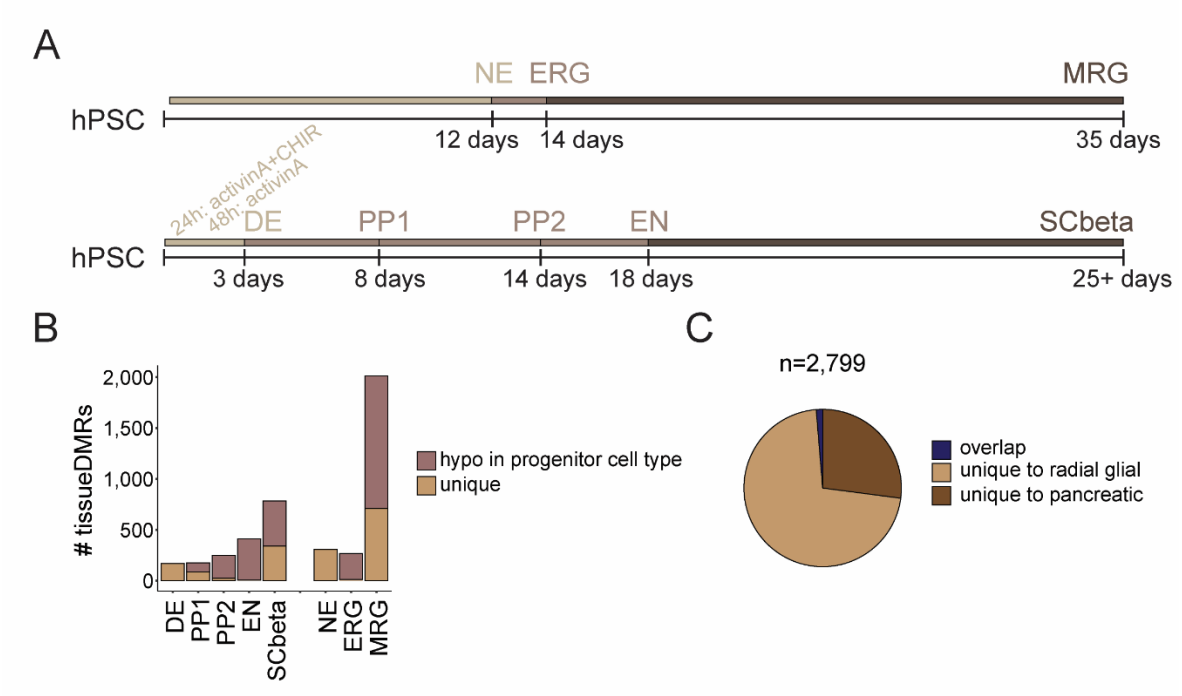

Figure 18: Cell type-specific hypomethylation during pancreatic and radial glial differentiation. **A** The timeline displays the differentiation protocols of hPSCs into mid radial glial (MRG) and stem-cell derived beta cells (SCbeta) with their respective intermediate cell types. Precursor cell types are neuroepithelial (NE) or definitive endoderm (DE), pancreatic progenitors 1/2 (PP1/PP2), and endocrine cells (EN), respectively. **B** The bar graph displays the number of all tissueDMRs called between hPSCs and SCbeta or MRG, respectively, and their overlap. **C** The bar graph displays the number of tissueDMRs which are uniquely assigned to the cell types of pancreatic and radial glial differentiation.

The DNA methylation turnover in hPSCs and its relationship to three germ layer differentiation

The majority of DKO-DMRs remain highly methylated in the three germ layers

DKO-DMRs are characterized by their permanent TET occurrence which allows them to switch from high to low levels of DNA methylation in hPSCs within a few days after DNMT3 depletion (Charlton et al., 2020). However, the switch in methylation could potentially happen even faster. In mESCs, the velocity of DNA demethylation by the catalytic domain of TET3 revealed an increase in 5hmC already after 6h (Ravichandran et al., 2022). In contrast, passive demethylation is replication-dependent and since hPSCs divide every 30-37h, the first possible time point to lose DNA methylation is after this time only (Yong et al., 2008). Since the first lineage decision of hPSCs after exit from pluripotency comes along with local DNA methylation changes, it might be that the DNA methylation turnover in hPSCs prepares DKO-DMRs for immediate hypomethylation when differentiation into one of the three germ layers

is initiated. To investigate this hypothesis, I examined to which extent DKO-DMRs become demethylated upon trilineage differentiation.

Surprisingly, only 7% of DKO-DMRs (806 out of 11,089) are overlapping at least one of the trilineage tissueDMRs, indicating that the majority of DKO-DMRs stably maintain high DNA methylation levels upon trilineage differentiation (**Figure 19** A). Indeed, only 379 of those DKO-DMRs which do not overlap with any trilineage tissueDMR lose more than 0.3 DNA methylation in at least one of the three differentiated cell types, confirming that the majority of DKO-DMRs are not demethylated upon trilineage differentiation (**Figure 19** B).

Further, I found that DKO-DMRs which overlap a tissueDMR show a bias towards being hypomethylated in ECTO. More than half overlap with tissueDMRs unique for ECTO, and an additional 5%, 8%, and 11% with ECTO+ENDO, ECTO+MESO, and ECTO+ENDO+MESO tissueDMRs, respectively (**Figure 19**A). Least, DKO-DMRs overlap with tissueDMRs unique to ENDO or ENDO+MESO (3% each). However, these biases mostly disappear when considering the group size of each tissueDMR category, except for a remaining slightly smaller overlap of DKO-DMRs with tissueDMRs unique to ENDO (**Figure 19**A). This could imply that the preference for overlapping ectoDMRs is purely coincidental and has no biological meaning.

In summary, the great majority of DKO-DMRs do not lose DNA methylation upon differentiation into the three germ layers, indicating that trilineage differentiation-related hypomethylation is not necessarily a consequence of the DNA methylation turnover. However, this does not exclude that the DNA methylation turnover maintains plasticity of a genomic region and might be necessary to enable a quick switch in DNA methylation upon differentiation.

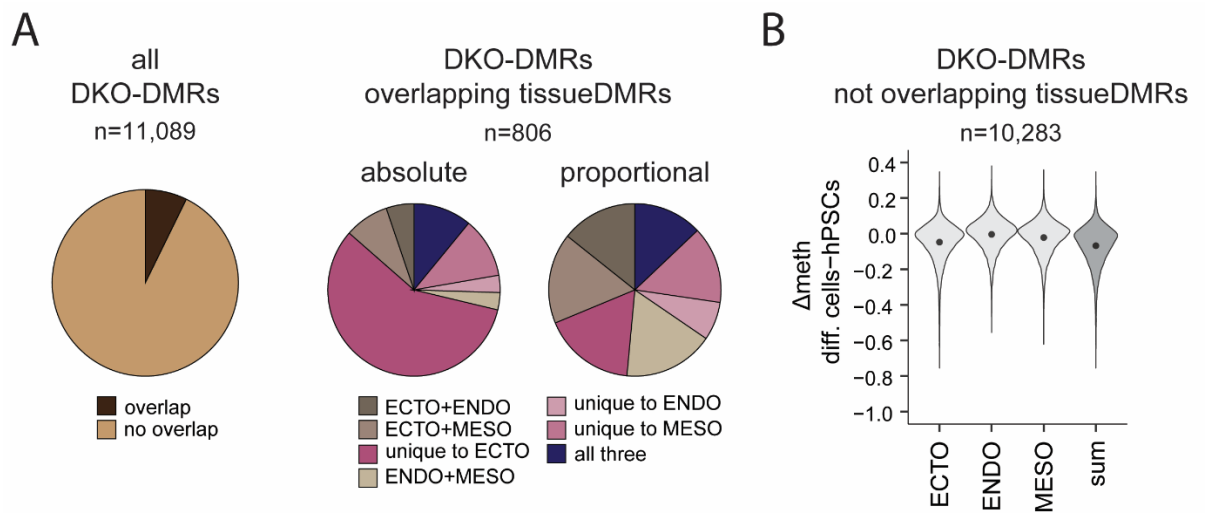

Figure 19: DKO-DMRs mostly remains highly methylated in the three germ layers. A On the left: The pie chart displays the number of DKO-DMRs overlapping with at least one tissueDMR. On the top right, top: The pie chart displays with which tissueDMRs the DKO-DMRs overlap. On the right, bottom: Pie chart displaying the overlap of DKO-DMRs with the different groups of tissueDMRs in proportion to the absolute number of tissueDMRs in each group. **B** The violin plot shows the DNA methylation difference between hPSCs and the respective differentiated cell type for those DKO-DMRs which do not overlap a tissueDMR. A negative value means loss of DNA methylation upon differentiation. Sum displays the strongest level of hypomethylation observed between hPSCs and the differentiated cell types for each DKO-DMR.

The majority of trilineage tissueDMRs are subject to the DNA methylation turnover in hPSCs

The majority of regions which do lose DNA methylation upon trilineage differentiation are enriched for the enhancer marks H3K4me1 and H3K27Ac in their respective differentiated states, indicating that hypomethylation frequently happens at putative enhancers (Gifford et al., 2013). Considering the DNA methylation turnover as some kind of poised DNA methylation state opens the question of whether differentiation-related demethylation is associated with local DNA methylation turnover occurring in the progenitor cell type. To this end, I investigated the trilineage tissueDMRs for their DNA methylation turnover potential in hPSCs.

Corresponding to those DKO-DMRs previously reported to become hypomethylated in at least one of the three germ layers, more than one-third of tissueDMRs are overlapping at least one DKO-DMR, (**Figure 20** A, B). In accordance with the fact that DKO-DMRs are most frequently hypomethylated in ECTO, ectoDMRs have the greatest overlap (43%) with DKO-DMRs, followed by mesoDMRs (39%), and finally endoDMRs (32%). Interestingly, the hypomethylation rate from WT to DKO hPSCs at those tissueDMRs which do not overlap a DKO-DMR is bimodally distributed with a bottleneck at ~-0.2 and an accumulation of

tissueDMRs with a loss bigger than 0.2 (**Figure 20** D). Specifically, non-DKO-DMR-overlapping ectoDMRs and mesoDMRs display similarly high rates of demethylation (-0.545 and -0.535, respectively), while endoDMRs have marginally better DNA methylation maintenance (-0.480) (**Figure 20** D). In line with this, endoDMRs have the highest average DNA methylation level in WT hPSCs (0.079) and also retain the highest level in DKO hPSCs (0.216) (**Figure 20** C). Further, the great majority of ectoDMRs, endoDMRs, and mesoDMRs lose more than <-0.3 DNA methylation in DKO hPSCs (91.4%, 82.1%, and 90.9%, respectively), indicating that trilineage-related hypomethylation is strongly associated with the DNA methylation turnover in hPSCs (**Figure 20** E).

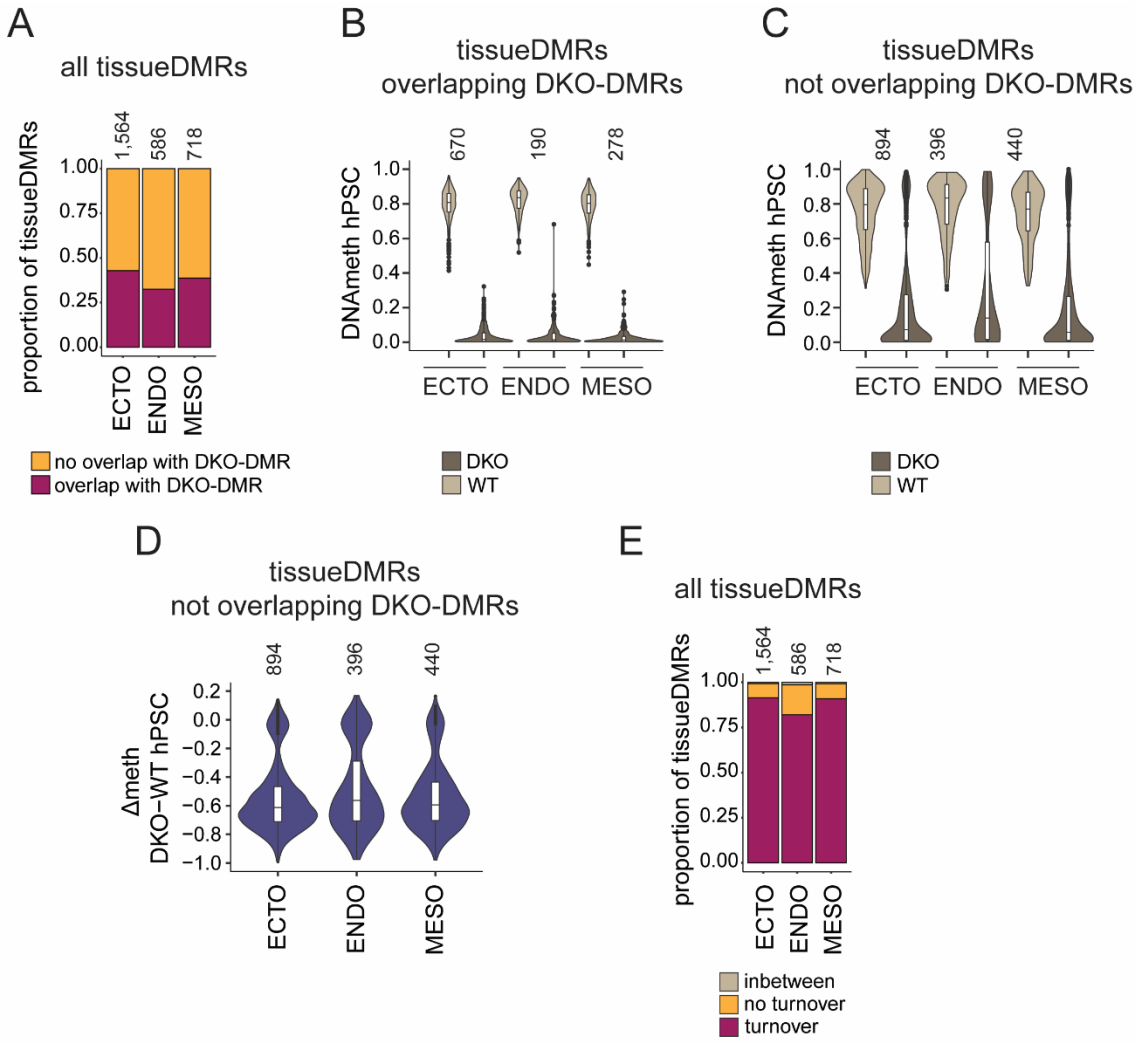

Figure 20: DNA methylation turnover at tissueDMRs of the three germ layers. A The bar graph displays the proportion of ECTO, ENDO, and MESO tissueDMRs overlapping at least one DKO-DMR. **B+C** The violin plots show absolute DNA methylation levels in WT and DKO hPSCs at tissueDMRs overlapping DKO-DMRs (B) and not-overlapping DKO DMRs (C). **D** The violin plot shows the absolute change in DNA methylation between WT and DKO hPSCs at tissueDMRs which do not overlap a DKO-DMR (negative value means hypomethylation from WT to DKO). **E** The bar graph displays the proportion of tissueDMRs which lose >0.3 DNA methylation (red), <=0.2, and those in between upon DNMT3 depletion.

Even though tissueDMRs of all three germ layers strongly associate with the DNA methylation turnover in hPSCs, there is a slight bias against endoDMRs. Indeed, only 19% of tissueDMRs unique to ENDO are overlapping with DKO-DMRs, while this is more than twice as frequent for tissueDMRs specific to ECTO or MESO (**Figure 21** A). In line with this, endoDMRs become on average less demethylation (-0.407), as compared to ectoDMRs (-0.642) and mesoDMRs (-0.623) (**Figure 21** B). Further, only two-thirds of endoDMRs lose <-0.3 DNA

methylation in DKO, while this is the case for almost all tissueDMRs unique to ECTO or MESO (**Figure 21** B, C). TissueDMRs shared between two or all three lineages have a similar association with the DNA methylation turnover in hPSCs as tissueDMRs unique to ECTO or MESO (**Figure 21** A, B).

In conclusion, the DNA methylation turnover in hPSCs does not determine which regions become hypomethylated upon three germ layer differentiation, indicated by the modest number of DKO-DMRs demethylated in ECTO, MESO, or ENDO. However, trilineage-related hypomethylation is strongly linked to the DNA methylation turnover in hPSCs. Thus, the DNA methylation turnover may still play a role for differentiation, with a character being more about preserving the ability to transition to low levels of DNA methylation rather than dictating. Furthermore, endoDMRs are slightly more resistant to demethylation in DKO hPSCs, implying that DNA methylation turnover in hPSCs may be more crucial for ECTO and MESO differentiation than for ENDO.

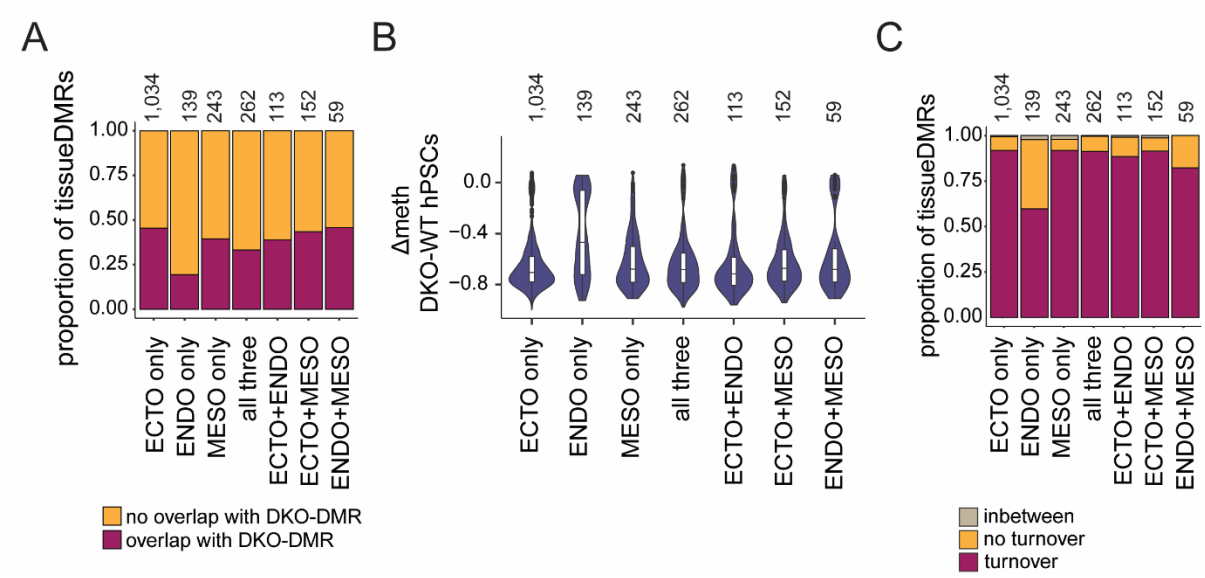

Figure 21: DNA methylation turnover at shared and unique tissueDMRs of the three germ layers. ECTO, ENDO, and MESO tissueDMRs are grouped by their shared or unique occupancy. **A** Bar graph displays the proportion of the categorized tissueDMRs overlapping at least one DKO-DMR (ECTO only: 45%; ENDO only: 19%; MESO only: 39.5%). **B** Violin plot shows the absolute change in DNA methylation between WT and DKO hPSCs at grouped tissueDMRs (negative value means hypomethylation from WT to DKO). **C** Bar graph displays the proportion of grouped tissueDMRs which lose >0.3 DNA methylation (red), <=0.2 and those in between upon DNMT3 depletion (ECTO only: 92%; MESO only: 92%; ENDO only: 60%).

The DNA methylation turnover in hPSCs and its association to hypomethylation in more advanced cell types

Additional DKO-DMRs are lowly methylated in adult tissues

The majority of DKO-DMRs remains highly methylated in the three germ layers. However, local demethylation events are not restricted to the earliest steps of differentiation but are also observed in more mature cell types (Lowdon et al., 2014; Alvarez-Dominguez et al., 2020). Further, changes in DNA methylation in somatic cells happen in a lineage-specific manner (Lowdon et al., 2014). Thus, the DNA methylation turnover at DKO-DMRs might be associated with differentiation-related hypomethylation in general, rather than being confined to cell states that are developmentally close to hPSCs. To this end, I investigated potential hypomethylation of DKO-DMRs in adult tissues (liver, lung, spleen, and small intestine), as representatives for mature cell types (Roadmap Epigenomics Consortium et al., 2015).

Remarkably, almost one-fourth of the DKO-DMRs overlap at least one lung tissueDMRs and similar observations can be made for the small intestine (23%), spleen (20%), and liver (19%) tissueDMRs (**Figure 22** A). In total, almost 40% of DKO-DMRs overlap with at least one of the adult tissueDMRs, with the majority being hypomethylated in more than one tissue (63%) (**Figure 22** B). However, only a minority of DKO-DMRs is demethylated in all four investigated tissues, indicating lineage-and tissue-specificity. Notably, the actual loss of DNA methylation at DKO-DMRs does not necessarily happen in the final steps of adult tissue maturation, but in any developmental stage between hPSCs and the mature tissue. Still, compared to the small number of DKO-DMRs becoming hypomethylated upon trilineage differentiation, this indicates that DKO-DMRs tend to lose DNA methylation in more mature cell types rather than immediately upon exit of pluripotency (**Figure 22** C). Together with the DKO-DMRs which overlap trilineage tissueDMRs, almost half of the DKO-DMRs become hypomethylated in at least one of the investigated differentiated cell types (**Figure 22** D). Noteworthy, more than half of those DKO-DMRs demethylated in one of the three germ layers maintain low levels of DNA methylation in adult tissues (**Figure 22** D)

In summary, investigating adult tissues to reveal differentiation-related DNA methylation changes at DKO-DMRs substantially increased the number of DKO-DMRs that are demethylated in a cell-type specific manner during development. This suggests that including more differentiated cell types into this analysis may reveal additional DKO-DMRs that are hypomethylated in some context of cellular differentiation. However, it is uncertain whether this will eventually encompass the majority of DKO-DMRs.

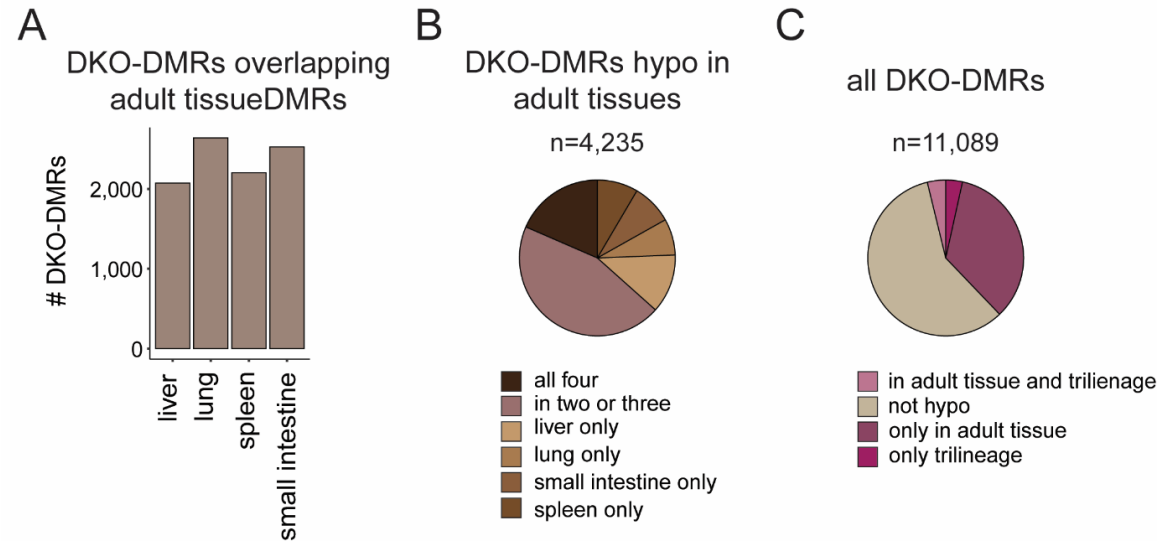

Figure 22: A substantial number of DKO-DMRs is hypomethylated between hPSCs and adult tissues. A The bar graph shows the overlap between DKO-DMRs and tissueDMRs of each adult tissue type. **B** The pie chart displays whether a DKO-DMR overlaps tissueDMRs of more than one tissue or whether it is specific to only one tissue. **C** The pie chart displays all DKO-DMRs and whether they overlap adult tissueDMRs, trilineage tissueDMRs, or both.

Early differentiation-related hypomethylation is more frequently associated with the DNA methylation turnover in hPSCs than late

The DNA methylome alters locally in more mature cell types, just as it does in the three germ layers, emphasizing the need to maintain DNA methylation flexibility later in development as well (Lowdon et al., 2014; Alvarez-Dominguez et al., 2020). The prior observation that plasticity of the DNA methylation landscape is significantly related to DNA methylation turnover in hPSCs during three germ layer differentiation provides a similar possibility for later stages of development... Furthermore, the fact that some adult tissueDMRs are subject to DNA methylation turnover in hPSCs suggests that late hypomethylation events in general may be related to the DNA methylation turnover in hPSCs (**Figure 22** D). To compare the relationship between the DNA methylation turnover in hPSCs and early versus late hypomethylation events, I evaluated hypomethylated DMRs of stepwise-*in vitro* differentiated progenitor cell types and their respective derivatives (**Figure 18**) (Ziller et al., 2015; Alvarez-Dominguez et al., 2020). Due to the small number of tissueDMRs gained in the intermediate cell types of pancreatic and radial glial differentiation, I focused the analysis on the first and last cell types of each differentiation protocol.

Upon radial glial differentiation, a similar number of NE- and MRG-DMRs overlap with at least one DKO-DMR (134 and 118, respectively) (**Figure 23** A). However, the total number of tissueDMRs found in NE is lower than in SCbeta which is why proportionally more NE tissueDMRs are associated with the DNA methylation turnover in hPSCs. This is also reflected on the level of average demethylation rate of NE- (-0.655) and MRG-DMRs (-0.363) in DKO hPSCs and strikingly, almost all NE-DMRs are hypomethylated by <-0.3 in DKO hPSCs, compared to around half of MRG tissueDMRs, supporting that proportionally early tissueDMRs are more associated with the DNA methylation turnover in hPSCs, but not on an absolute level (**Figure 23** C, D). Of note, in WT hPSCs, NE-DMRs have a slightly lower average DNA methylation level (0.774) as compared to MRG (0.822), suggesting that early tissueDMRs are generally more susceptible to DNA demethylation in hPSCs (**Figure 23** B). In contrast to radial glial differentiation, only a small number of DE- and SCbeta-DMRs overlap at least one DKO-DMR (10 (5.8%) and 23 (6.7%), respectively), indicative of a smaller association between the DNA methylation turnover in hPSCs and pancreatic-related hypomethylation (**Figure 23** A). Still, just as for NE and MRG, DE-DMRs are on average stronger demethylated in DKO hPSCs than SCbeta-DMRs (-0.300 and -0.166, respectively) (**Figure 23** A). In line with this, the proportion of SCbeta-DMRs which are hypomethylated <-0.3 in DKO hPSCs is lower) as compared to DE-DMRs (21% and 49%, respectively) (**Figure 23** D). However, this bias for a link between the DNA methylation turnover in hPSCs and early pancreatic DMRs is again only proportionally, whereas the total number of DE-DMRs and SCbeta-DMRs with DNA methylation turnover is similar (83 and 73, respectively). Of note, early hypomethylated regions during pancreatic differentiation are on average slightly lower methylated in WT hPSCs as compared to later (DE-DMRs: 0.777; SCbeta-DMRs: 0.884), as it was observed for radial glial differentiation (**Figure 23** B).

In summary, demethylation at later stages is less commonly related to DNA methylation turnover in hPSCs along both examined differentiation trajectories; nonetheless, the absolute occurrences are similar between early and late. Still, this may indicate a more important role for the DNA methylation turnover in hPSCs during early differentiation. Furthermore, radial glial differentiation is more frequently connected with the DNA methylation turnover in hPSCs than pancreatic differentiation, suggesting that the DNA methylation turnover is less crucial for certain lineages.

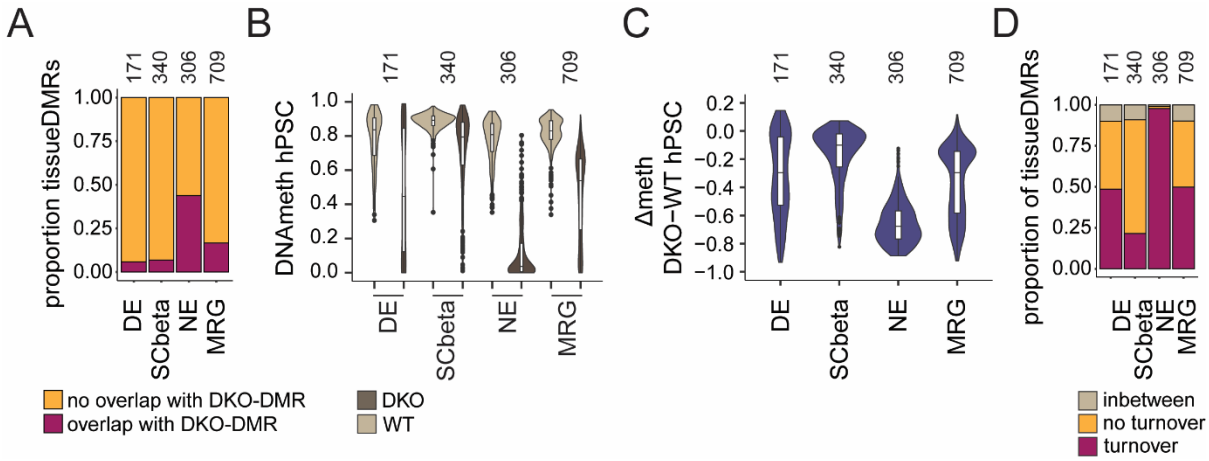

Figure 23: DNA methylation turnover at early and late tissueDMRs. A The bar graph displays the overlap between tissueDMRs of pancreatic and radial glial differentiation with DKO-DMRs. **B+C** Violin plot shows absolute DNA methylation levels in WT and DKO hPSCs at cell type-specific tissueDMRs (B) and the DNA methylation difference between DKO and WT (C) (negative value means hypomethylation from WT to DKO). **D** Bar graph displays the proportion of tissueDMRs which lose >0.3 DNA methylation (red), <=0.2, and those in between upon DNMT3 depletion.

Exploring the connection between the rate of differentiation-related hypomethylation and the DNA methylation turnover in hPSCs

By pre-targeting TETs in hPSCs, tissueDMRs with DNA methylation turnover set themselves apart from other highly methylated genomic sites that undergo hypomethylation upon differentiation. Considering this, along with the assumption that TETs are necessary for demethylating tissueDMRs with DNA methylation turnover, these tissueDMRs may experience stronger hypomethylation upon initiation of differentiation, while other genomic regions might require *de novo* TET recruitment prior or even depend on passive demethylation. In order to analyze the relationship between the DNA methylation turnover in hPSCs and the rate of demethylation upon differentiation, I examined tissueDMRs with and without DNA methylation turnover and further put their differentiation-related hypomethylation rates into correlation with the rate of demethylation in DKO hPSCs. To this end, tissueDMRs of each investigated cell type were split by their rate of demethylation between WT and DKO hPSCs (referred to as tissue DMRs “with turnover” and “without turnover”) (see Material and Methods for cutoffs).

Differentiation-related hypomethylation of tissueDMRs is stronger when associated with DNA methylation turnover during pluripotency

Interestingly, the average rate of hypomethylation upon differentiation at tissueDMRs with DNA methylation turnover tends to be higher as compared to tissueDMRs without DNA methylation turnover, except for DE- and SCbeta-DMRs (**Figure 24** A, B). Especially in the neuroectodermal lineages tissueDMRs with DNA methylation turnover lose more DNA methylation upon differentiation as compared to respective tissueDMRs without DNA methylation turnover, with an average increased loss by 0.069, 0.130, and 0.133 in ETCO, NE and MRG, respectively, suggesting an impact of the DNA methylation turnover in hPSCs on the rate of differentiation-related hypomethylation. This effect is weaker in ENDO (delta: 0.040) and MESO (delta: 0.016) (**Figure 24** B). Vice versa, DE- and SCbeta-DMRs without DNA methylation turnover tend to lose more DNA methylation upon differentiation as compared to those associated with DNA methylation turnover in DE and SCbeta (DE: without turnover -0.511, with turnover -0.455; SCbeta: without turnover -0.572, with turnover -0.540) (**Figure 24** B). This is in line with the rarer targeting of pancreatic tissueDMRs by the DNA methylation turnover in hPSCs, further supporting a lineage-specific bias of the DNA methylation turnover in hPSCs.

Three germ layer-DMRs, as well as NE-DMRs without DNA methylation turnover, tend to have higher methylation levels in WT hPSCs as compared to their counterparts with DNA methylation turnover (**Figure 24** A, C). Conversely, MRG-DMRs show the opposite pattern (**Figure 24** C). This suggests that TET activity not only leads to DNA methylation turnover in hPSCs but might also reduce DNA methylation levels, consistent with higher local DNA methylation levels in TET-TKO hPSCs (Charlton et al., 2020). This trend towards higher DNA methylation levels for tissueDMRs with DNA methylation turnover continues in their respective differentiated cell type, with MRG-DMRs following a similar pattern (**Figure 24** C). Particularly noteworthy are the substantial differences in DNA methylation levels between tissueDMRs with and without DNA methylation turnover for NE-DMRs (with turnover: 0.233; without turnover: 0.494), followed by ECTO- and ENDO-DMRs. However, while the average DNA methylation levels in hPSCs at DE -and SCbeta-DMRs follow the same trend as the other tissueDMRs, with elevated levels in tissueDMRs without DNA methylation turnover, the levels in their respective differentiated cell types are very similar, forming the basis inversed trend of increased hypomethylation at pancreatic DMRs without DNA methylation turnover upon differentiation (**Figure 24** C).

In summary, tissueDMRs which are associated with the DNA methylation turnover in hPSCs lose slightly more DNA methylation as compared to tissueDMRs without DNA methylation turnover; most strongly observed for MRG, followed by NE and ECTO. This supports the hypothesis that premature presence of TETs may increase demethylation efficiency upon induction of differentiation. In contrast to cell types of ectodermal origin, DE and SCbeta tissueDMRs display the opposite tendency, supporting previous observations that the targeting and function of the DNA methylation turnover might be lineage-specific.

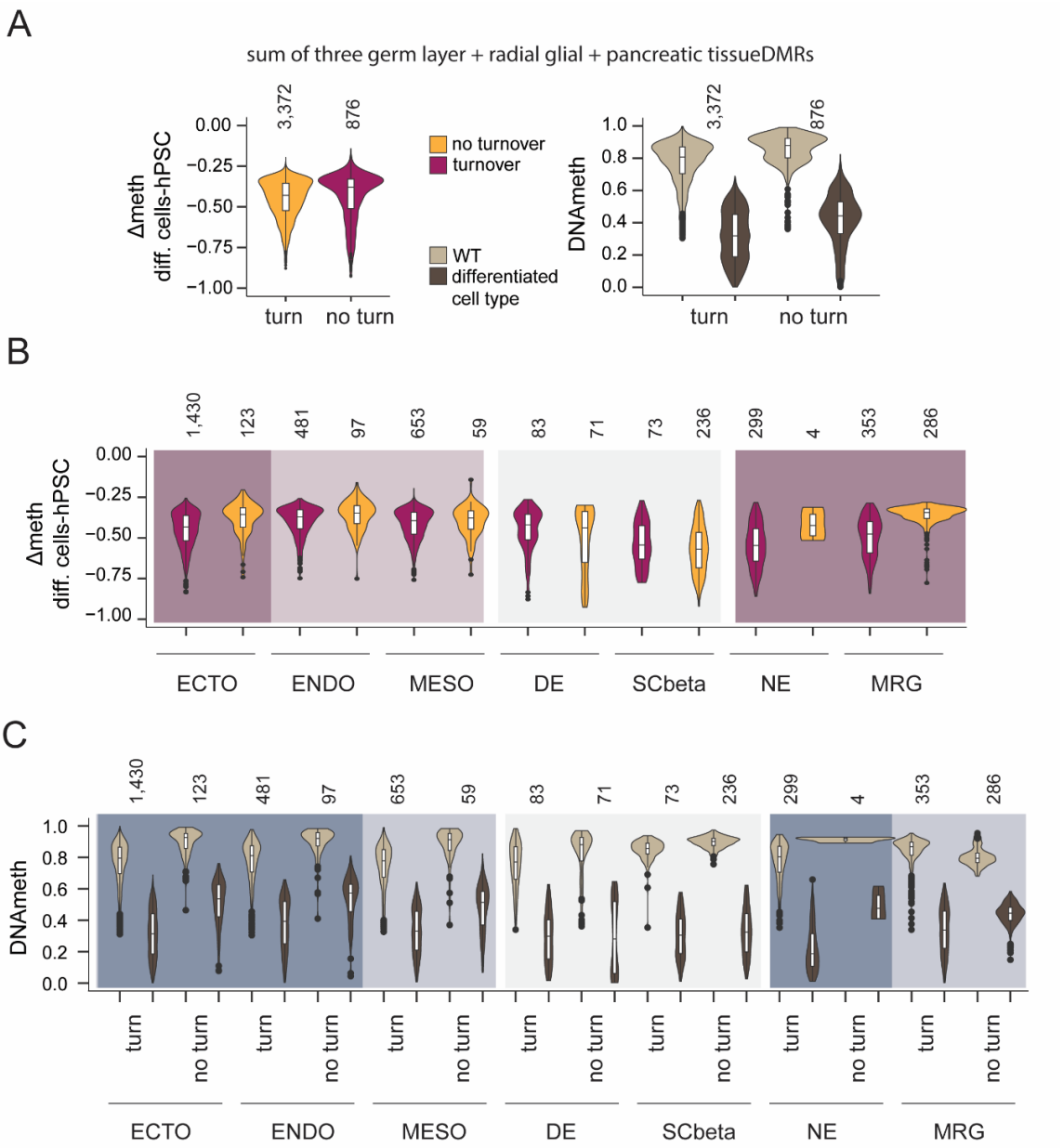

Figure 24: Differentiation-related hypomethylation at tissueDMRs with and without DNA methylation turnover. TissueDMRs of the three germ layers, as well as DE, SCbeta, NE, and MRG, are separated by having DNA methylation turnover in hPSC or not. **A** All tissueDMRs are summarized and the violin plots display their DNA methylation levels in WT hPSCs and the respective differentiated cell type (right), as well as the rate of hypomethylation upon loss of DNMT3s in hPSCs (left); (negative value means hypomethylation from WT to DKO). Important to note, that the method for determining tissueDMRs only considers CpGs that are covered by both data sets, as opposed to computing the DNA methylation difference between two cell types that bases on all CpGs covered in the WGBS of hPSCs and MRG. As a result, average DNA methylation loss during differentiation at a tissueDMR can potentially be lower than the cutoff of -0.3 that was utilized for DMR calling. The violin plots in **B** and **C** distinguish between tissueDMRs of the different cell types. Dark red/ dark blue highlights the three groups of tissueDMRs with the biggest difference in the hypomethylation rate or DNA methylation level in differentiated cells between those with and without DNA methylation turnover in hPSCs, respectively. TissueDMRs underlaid with grey do not follow the general trend of increased hypomethylation upon differentiation at tissueDMRs with DNA methylation turnover in hPSCs.

The upper limit of hypomethylation during differentiation is comparable to the level of TET activity during pluripotency

The observation of a trend for categorized tissueDMRs to become more demethylated during differentiation when subjected to the DNA methylation turnover in hPSCs prompted the investigation of a potential linear correlation between the hypomethylation rate associated with differentiation and the one associated with DNMT3-depletion in hPSCs. The following analysis includes only tissueDMRs with DNA methylation turnover in hPSCs.

However, the rate of demethylation upon differentiation does not linearly increase with the rate of hypomethylation in DKO hPSCs in any of the cell types, revealing that there is no linear relationship between the level of TET activity at tissueDMRs in hPSCs and its prospective activity during differentiation (**Figure 25** A). Still, it is interesting that the majority of ECTO- (92%), ENDO- (84%), and MESO-DMRs (91%) do not lose more DNA methylation upon differentiation than they do in DKO hPSCs, while the opposite can occur (**Figure 25** B). This observation that the rate of differentiation-related hypomethylation is at most equal to or below the demethylation rate upon DNMT3-depletion in hPSCs, suggests that TET activity in hPSCs limits demethylation activity upon differentiation. In contrast, in later cell types like MRG and SCbeta, tissueDMRs frequently lose more DNA methylation upon differentiation as compared to DKO hPSCs (MRG: 65%; SCbeta: 77%), while tissueDMRs of their progenitor cell types, NE and DE, rather follow the trend observed for the three germ layers (20% and39%, respectively) (**Figure 25** A).

In summary, though hypomethylation during differentiation is not linearly related to demethylation in DKO hPSCs, the analysis revealed that the rate of differentiation-related hypomethylation rarely exceeds the demethylation rate upon DNMT3-knockout in hPSCs. Assuming that TET activity is responsible for demethylation upon differentiation, TET activity in hPSCs might create a range of possible hypomethylation rates occurring during early differentiation.

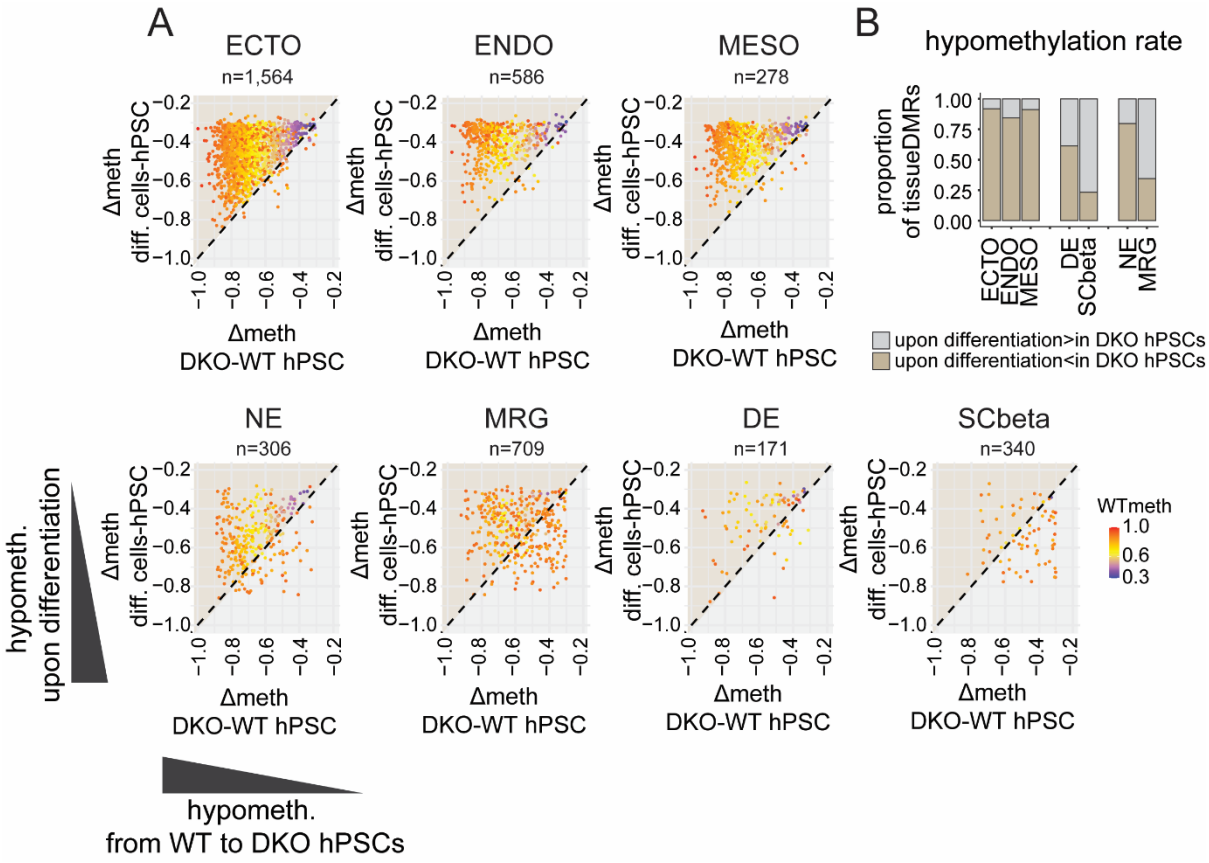

Figure 25: Relationship between the DNA methylation turnover and hypomethylation upon differentiation. A The scatter plots display ECTO, ENDO, MESO, NE, MRG, DE, and SCbeta tissueDMRs and their level of hypomethylation from WT hPSCs to DKO hPSCs (x-axis) and from WT hPSCs to the respective differentiated cell types (y-axis). The rainbow color scale indicates the DNA methylation level in WT hPSCs. Further, the diagonal is drawn into each plot. **B** The bar graph displays the proportion of tissueDMRs with stronger hypomethylation in DKO hPSCs compared to the differentiated cells (red) and vice versa (yellow).

The DNA methylation turnover dynamic in differentiated cells

The DNA methylation turnover continues at highly methylated DKO-DMRs in ENDO cells

The two studies which investigated the DNA methylation turnover primarily focus on the pluripotent state (Ginno et al., 2020; Charlton et al., 2020). Nevertheless, in adult mouse tissues, still a small number of genomic loci were shown to be demethylated in the absence of DNMT3s and similarly, a small fraction of the DKO-DMRs called in hPSCs continue to turn over in differentiated human motor neurons (Charlton et al., 2020). This indicates that even in mature tissues the DNA methylation turnover still takes place, even though to a smaller extent. However, so far, no insights have been gained into DNA methylation turnover dynamics in the three germ layers, as the lineage decision closest to pluripotency where the DNA methylation

turnover was primarily described. In this work, the majority of DKO-DMRs was shown to maintain high DNA methylation levels in the three germ layers (**Figure 19**). Still, many of them lose methylation later in development, revealed by low DNA methylation levels at DKO-DMRs in adult tissues (**Figure 22**). This indicates that the DNA methylation switch at DKO-DMRs happens during various developmental time points and thus raises the question of whether the DNA methylation turnover continues until the switch in DNA methylation levels happens. Considering the DNA methylation turnover as a way to prepare regions for possible demethylation during later stages of differentiation, DKO-DMRs that become demethylated in one lineage but remain highly methylated in another may maintain their turnover until the cell has committed to one of them. Only after this decision has taken place, the DNA methylation turnover might stop in the lineage where the region does not become demethylated in any later stage of development. As a first step towards investigating this hypothesis, I examined TET activity at selected DKO-DMRs in *in vitro* differentiated ENDO cells, which represent an early stage of embryonic development after exiting pluripotency.

To examine DNMT and TET activity during pluripotency and upon differentiation, I conducted oxBS/BS-amplicon-seq for a selection of DKO-DMRs in hPSC and hPSC-derived ENDO cells. Based on published WGBS data of WT hPSCs (HUES8 and HUES64), I chose two DKO-DMRs which remain highly methylated upon ENDO differentiation but become hypomethylated in adult liver tissue, showing that the DNA methylation switch is developmentally still ahead (**Figure 26**) (Charlton et al., 2020). Both DKO-DMRs are located in an intron of RXRA or LRP5, respectively. Additionally, I included two DKO-DMRs that already lose DNA methylation in ENDO cells to examine different levels of 5mC and 5hmC. These two lie in introns of CUX1 or MCC, respectively (**Figure 26**).

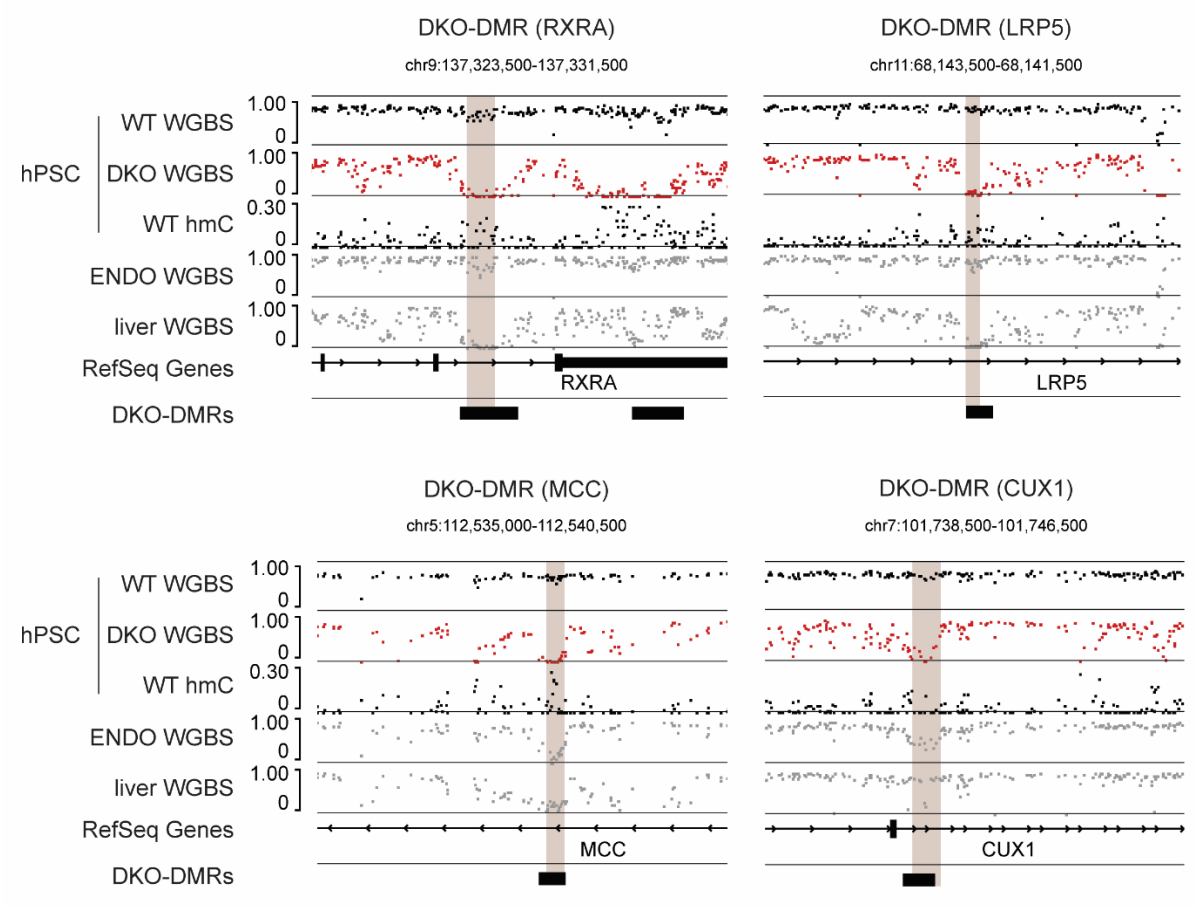

Figure 26: Selection of DKO-DMRs for investigating the DNA methylation turnover in ENDO. The browser tracks display WGBS data for WT hPSCs, DKO hPSCs, ENDO, and LIVER, as well as the 5hmC signal in WT hPSCs at the selected DKO-DMRs. Those CpGs covered by the Amplicon-seq are highlighted with a brown background. DKO-DMR (RXRA) and DKO-DMR (LRP5) remain highly methylated in ENDO while demethylation at the control DKO-DMRs (MCC and CUX1) is already initiated in ENDO.

In hPSCs, all four DKO-DMRs are highly methylated (between 0.592 and 0.894) and exhibit varying levels of hydroxymethylation (between 0.021 and 0.1). The highest and lowest 5hmC levels, respectively, are found at DKO-DMR (RXRA) and DKO-DMR (LRP5) (**Figure 27** B). Furthermore, within DKO-DMRs, there are major differences in the 5mC and 5hmC levels between individual CpGs. Here, DKO-DMR (CUX1) exhibits the widest range of methylation (0.562-0.816), whereas DKO-DMR (RXRA) displays the greatest variations in hydroxymethylation (0.032-0.183). These findings are consistent with WGBS and oxWGBS data of hPSCs (Charlton et al., 2020).

In hPSC-derived CXCR4+-ENDO cells, DKO-DMR (MCC) is fully demethylated (0.006) and simultaneously loses nearly all 5hmC (0.004) (**Figure 27** A, B). Similarly, DKO-DMR (CUX1)

only retains low 5mC levels at two of four covered CpGs (0.056) (**Figure 27** B). However, the hydroxymethylation level at DKO-DMR (CUX1) is not only maintained between hPSCs and ENDO but even slightly increases from 0.063 to 0.077. In particular, the distribution of high and low 5hmC levels over the CpGs changes, indicating that the DNA methylation turnover is still ongoing even though methylation levels decreased (**Figure 27** B). In contrast to DKO-DMR (MCC) and (CUX1), DKO-DMR (RXRA) only slightly lose 5mC during the transition from hPSCs (0.592) to ENDO (0.481) (**Figure 27** B). Strikingly, DKO-DMR (RXRA) retained its average 5hmC level between hPSCs (0.1) and ENDO (0.122), indicating that the DNA methylation turnover is locally maintained upon differentiation. Of note, on the single CpG-level, the DKO-DMR (RXRA) experiences 5mC and 5hmC alterations. Particularly, the observed 5mC loss in ENDO is mainly due to strong hypomethylation of two out of seven CpGs, which have the highest levels of 5hmC in hPSCs, suggesting that accumulation of 5hmC may play a role for subsequent demethylation (**Figure 27** B). Additionally, those CpGs within DKO-DMR (RXRA) that exhibit the most constant 5mC values upon differentiation, gain 5hmC alongside. Compared to all other tested DKO-DMRs, DKO-DMR (LRP5) maintains 5mC levels between hPSCs (0.894) and ENDO (0.878) best (**Figure 27** B). Furthermore, even though the average 5hmC level at DKO-DMR (LRP5) in hPSCs is relatively low (0.0212), it is maintained on a very similar level in ENDO (0.0212), with a similar range over single CpGs (hPSCs: 0.004-0.054; ENDO: -0.011-0.068), indicating that the local DNA methylation turnover is continued at DKO-DMR (RXRA) in ENDO.

In conclusion, both of the examined DKO-DMRs that exhibit sustained high methylation in ENDO, while undergoing demethylation in the adult liver, maintain hydroxymethylation in ENDO cells, indicating continuous TET activity. Furthermore, the persistence of high 5mC levels despite the TET activity implies the presence of DNMT3s at these DKO-DMRs in ENDO cells. Thus, this work supports the idea that the DNA methylation turnover plays a role in maintaining DNA methylation flexibility also after the first lineage decision and during other developmental stages following the pluripotent state

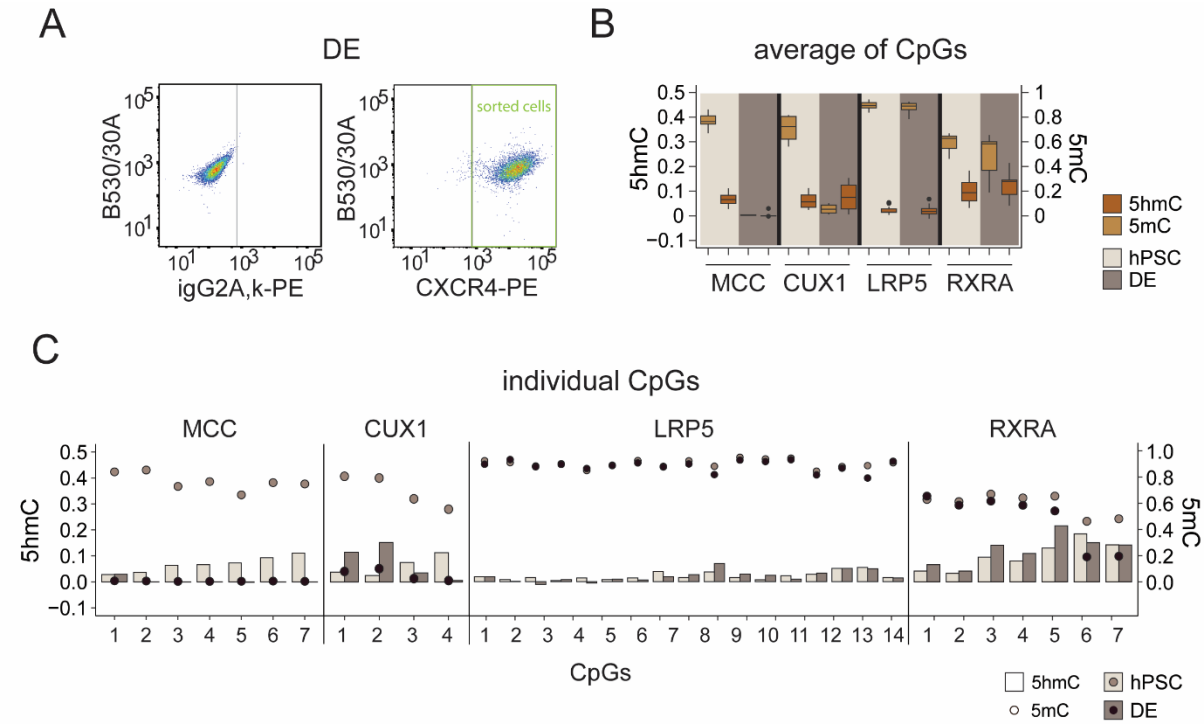

Figure 27: DNA methylation and hydroxymethylation at DKO-DMRs upon differentiation. A The FACS-plot displays the sorting strategy for CXCR4-positive ENDO cells, gated in comparison to an igG2A,k-PE control. **B** The boxplots display the distribution of DNA methylation and hydroxymethylation levels at individual CpGs within one DKO-DMR (MCC, CUX1, LRP5, RXRA) in hPSCs and hPSC-derived ENDO cells. **C** The bar graphs show DNA methylation and hydroxymethylation on the single CpG-level for each DKO-DMR in hPSCs and hPSC-derived ENDO cells. DNA methylation levels are obtained from oxBS-Amplicon sequencing, while hydroxymethylation was calculated as the difference between BS- and oxBS-Amplicon sequencing.

Increased hydroxymethylation in differentiated cells prior to differentiation-related demethylation

As of yet, we have learned that the link between differentiation-related hypomethylation and the DNA methylation turnover in hPSCs is decreased as development progresses (**Figure 23**). However, in theory, the factors required for co-targeting DNMT3s and TETs are not restricted to pluripotency and are present in differentiated cells as well (**Figure 27**). According to the idea that DNA methylation plasticity is generally associated with the DNA methylation turnover, regions that become demethylated later in development might be co-targeted by DNMT3s/TETs in more progressed progenitor cells rather than already in hPSCs. To this end, I investigated whether regions that are hypomethylated on day 60 of motor neuron (MN) differentiation and are not subject to the DNA methylation turnover in hPSCs establish the dynamic *de novo* in day 16 progenitor cells, prior to the actual demethylation (Charlton et al., 2020).

In particular, I determined genomic regions that are significantly demethylated from day 16 to day 60 of MN differentiation (MN-D60-DMRs) and distinguished between those with and without DNA methylation turnover in hPSCs, based on published WGBS data (Charlton et al., 2020). Further, I utilized oxWGBS data for hPSCs and day16 MN to analyze changes in 5hmC levels as an indicator for TET activity (Charlton et al., 2020). Here, Moreover, I called a set of background regions that are comparable in size and CpG-density to the MN-D60-DMRs but are neither demethylated during MN differentiation nor subject to the DNA methylation turnover in hPSCs (**Figure 28** B-E).

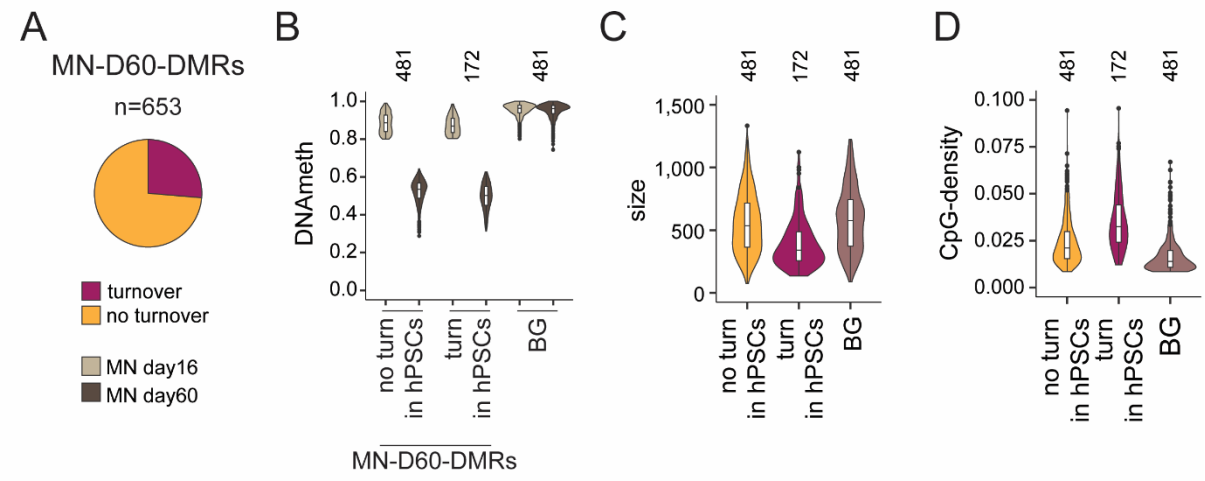

Figure 28: Increased hydroxymethylation in MN-D16 at MN-D60 hypomethylated DMRs. A Pie chart displays the proportion of MN-D60 DMRs with and without DNA methylation turnover in hPSCs. **B** The violin plot shows similarly high levels of DNA methylation in MN (day 16) at the different regions and strong demethylation at MN-D60 DMRs in MN (day 60), but not at the background regions (BG). **C+D** Violin plots display the similarities in size (C) and CpG-density (D) of MN-D60 DMRs with and without DNA methylation turnover in hPSCs and the background regions.

Between MN day 16 and day 60 653 genomic loci become significantly demethylated, the majority (74%) of which are not targeted by the DNA methylation turnover in hPSCs (**Figure 28** A, B). As expected, MN-D60 DMRs without DNA methylation turnover in hPSCs have on average lower 5hmC levels in hPSCs (0.034) than those with DNA methylation turnover (0.059), confirming that the DNA methylation turnover causes an enrichment of 5hmC (**Figure 29**). Hydroxymethylation at the background regions is similarly low as at MN-D60 DMRs without DNA methylation turnover in hPSCs (0.036).

Strikingly, on day16 MN, 5hmC substantially accumulates at MN-D60 DMRs without DNA methylation turnover in hPSCs (0.132), while the background regions retain similarly low levels as compared to hPSCs (0.037) (**Figure 29**). This does suggest that the DNA methylation turnover was established *de novo* in day16 MN Additionally, MN-D60 DMRs with DNA methylation turnover in hPSCs exhibit a similar increase in average 5hmC level in day16 MN (0.167) as seen for MN-D60-DMRs without DNA methylation turnover in hPSCs (**Figure 29**). This indicates that 5hmC may generally accumulate prior to demethylation.

In summary, regions that undergo demethylation in progressed cell types newly establish the DNA methylation turnover in progenitor states. Additionally, TET activity also rises at loci with DNA methylation turnover in hPSCs when approaching the developmental state in which it becomes demethylated. The broader question of whether differentiation-related hypomethylation generally involves accumulation of 5hmC before hypomethylation requires further investigations in the future.

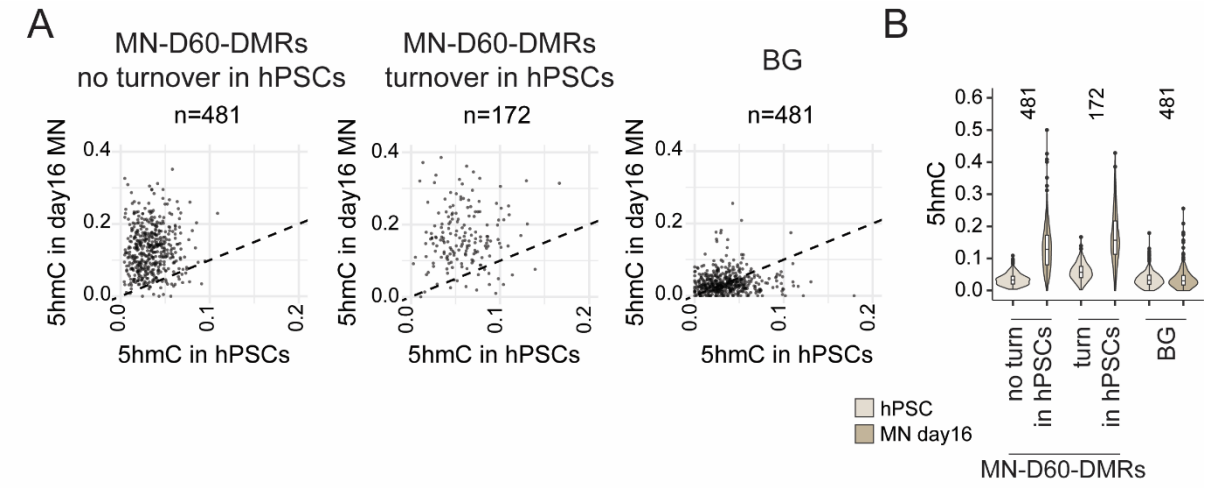

Figure 29: Increased hydroxymethylation in MN-D16 at MN-D60 hypomethylated DMRs. A Scatter plots display the 5hmC levels in hPSCs (x-axis) and MN (day16) (y-axis) at individual MN-D60 DMRs with and without DNA methylation turnover in hPSCs, as well as at background regions (BG). The dotted line marks the diagonal on which the levels are unchanged from hPSCs to MN (day 16). **B** The violin plot shows the increase of 5hmC levels from hPSCs to MN (day 16) at MN-D60 DMRs with and without DNA methylation turnover (TO) but not at the background regions.

tissueDMRs without DNA methylation turnover are enriched for simple repeats and SINE-Alus

TissueDMRs of the same cell type share a key characteristic: they experience hypomethylation related to the same differentiation step, possibly due to similar DNA-sequence features that respond to common differentiation signals. However, despite this, these regions exhibit substantial differences in terms of the DNA methylation turnover in hPSCs, raising the question of the genetic background responsible for these variations. The molecular and genetic mechanisms that control the DNA methylation turnover remain largely unknown. In addition to H3K4me1 enrichment, DKO-DMRs are located in slightly open chromatin in hPSCs and also exhibit a higher CpG-density compared to the background (**Figure 15**) (Charlton et al., 2020). However, none of these features are unique to DKO-DMRs. When delving deeper into the sequence context, it was discovered that the 5'UTRs of evolutionarily young LINE1 retrotransposons tend to be targeted by the DNA methylation turnover more frequently than older ones (Charlton et al., 2020). This is currently the most notable and well-described connection between a genomic element and the DNA methylation turnover in hPSCs. Additionally, other DKO-DMRs were found to overlap with retrotransposons from the SINE and LTR families (Charlton et al., 2020). In order to determine whether repeat elements could be the foundation for variations in DNA methylation turnover-targeting between the ECTO-, ENDO-, and MESO-DMRs, I analyzed their intersections with different repeat classes.

Across all germ layers, tissueDMRs with DNA methylation turnover less frequently overlap any type of repetitive element (37-51%), as compared to those without (76-78%), suggesting that the majority of repeats are not associated with the DNA methylation turnover in hPSCs (**Figure 30** A). Interestingly, the proportion of simple repeats (33%) and SINE-Alus (50%) is higher in repeats intersected by ECTO-DMRs without DNA methylation turnover than in repeats overlapping ECTO-DMRs with DNA methylation turnover (9% and 22%, respectively) (**Figure 30** B). Notably, simple repeats account for the smallest proportion of nucleotides in the genome among all repeat classes (1.7%), making the substantial overlap considerably less likely to occur merely by chance (**Figure 30** B). Similarly, SINE-Alus only contribute to 20% of the genomic nucleotides associated with repeat elements, despite their high copy number, due to their small individual size., This implies, that SINE-Alu abundance in ECTO-DMRs lacking the DNA methylation turnover is not a random occurrence. Of note, the absolute number of SINE-Alus located within ECTO-DMRs without DNA methylation turnover is smaller (n=70) as compared to with (n=162), which might be due to the smaller count of ECTO-

DMRs without DNA methylation turnover than with. In contrast, ECTO-DMRs without DNA methylation turnover have a lower percentage of intersected repeats that are LTR retrotransposons (6%) compared to ECTO-DMRs with DNA methylation turnover (20%) (**Figure 30** B). Given the proportion of nucleotides covered by LTR retrotransposons genome-wide (17%), this is an unusually low prevalence of LTRs in ECTO-DMRs without DNA methylation turnover compared to other repeat classes. On the other hand, SINE-MIRs are proportionally strongly enriched in ECTO-DMRs with DNA methylation turnover (17%) (**Figure 30** B). Similar trends can be observed for ENDO- and MESO-DMRs, except for the fraction of SINE-Alus being comparable between MESO-DMRs with and without DNA methylation turnover (**Figure 30** B).

In summary, especially simple repeats and SINE-Alus are enriched in the trilineage tissueDMRs without DNA methylation turnover, whereas those with DNA methylation turnover more frequently contain LTRs and SINE-MIRs. This suggests that LTRs and SINE-MIRs may tend to be targets of the DNA methylation turnover, in addition to the 5'UTRs of young LINE1 elements, while SINE-Alus and simple repeats are more resistant.

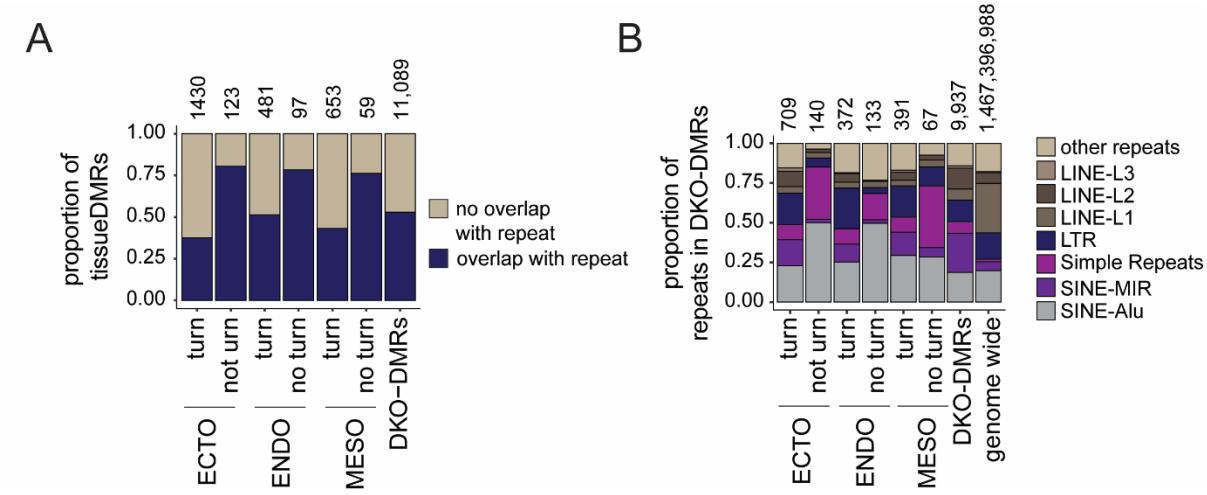

Figure 30: tissueDMRs with and without DNA methylation turnover are enriched for different repeat classes. A+B TissueDMRs of the three germ layers are separated by having DNA methylation turnover in hPSCs or not. The bar plots display the proportion of tissueDMRs that overlap at least one repeat element (retrotransposons, DNA-transposons, simple repeats) (A) and to which repeat class the intersected elements belong (B).

Efforts towards dissecting the relationship between retrotransposons and the DNA methylation turnover

The previous observation that the DNA methylation turnover in hPSCs has a preference for 5'UTRs of evolutionarily young L1 elements, together with the finding that tissueDMRs with DNA methylation turnover are enriched for LTR and SINE-MIR retrotransposon, but rather devoid of simple repeats and SINE-Alus, suggests a recruiting mechanism involving repeat-associated features (Charlton et al., 2020). To identify the predominant subfamilies of repetitive elements targeted by the DNA methylation turnover, I conducted a comprehensive analysis of the DNA methylation turnover across repetitive elements genome-wide in hPSCs.

Minor variability in DNA methylation turnover at repeat families negatively correlates with DNA methylation levels in hPSCs

Within the various repeat families, there are elements lacking CpGs and those with low levels of DNA methylation (**Figure 31** A). To determine families with significant DNA methylation turnover, I excluded those elements with less than 2 CpGs or DNA methylation levels <=0.3 (see Material and Methods for details). Of note, the majority of the internal parts and LTRs of the ERVK family (82.8%, and 94%, respectively), as well as SINE-Alus (85.5%), contain CpGs, whereas simple repeats are notably depleted of CpGs (9%) (**Figure 31** A). The remaining repeat families fall within this range. Moreover, those elements containing CpGs are almost universally highly methylated in hPSCs, independent of the repeat family (**Figure 31** A).

Across each family, the vast majority of individual elements is independent of the DNA methylation turnover in hPSCs (**Figure 31** B). After SINE-MIR family members (11%), the LTRs of ERVK and ERV1 are most likely to experience the DNA methylation turnover in hPSCs. This underscores a potential biological meaning for the previously observed enrichment of SINE-MIRs and LTR elements in tissueDMRs with DNA methylation turnover (**Figure 30** B). Furthermore, simple repeats display a similar trend towards a stronger association with the DNA methylation turnover in hPSCs, as do L2 and CR1 elements (**Figure 31** B). The least connection is seen for L1 elements and internal parts of ERVL and ERVL-MaLR, together with SINE-Alus (all three 2%) (**Figure 31** B). SINE-Alus' independence from the DNA methylation turnover is consistent with the observation of tissueDMRs lacking the DNA methylation turnover, which revealed an increased enrichment of SINE-Alus as compared to tissueDMRs targeted by the turnover (**Figure 30** B). Interestingly, for each LTR

family, the DNA methylation turnover tends to target the LTRs themselves more frequently than the corresponding internal parts (**Figure 31** B). This is in line with 5'ends of full-length LINE1s being preferably subject of the turnover compared to the rest of the element, supporting the idea that promoter sequences of transposable elements may be its main target (Charlton et al., 2020).

Furthermore, the average DNA methylation level at elements of a repeat family and their proportion of turnover are negatively correlated, with a notable exception for simple repeats (Spearman's ρ =-0.736, p=0.006, simple repeats excluded), which is in line with previous observations for tissueDMRs (**Figure 24** and **Figure 31** B). This suggests that there is a potential association between the DNA methylation turnover or related features and the reduction of DNA methylation levels at repeat elements in hPSCs.

In summary, while some repeat families slightly more frequently display DNA methylation turnover than others, these differences are generally minor. This indicates that either repeats are not the primary targets of the DNA methylation turnover or, as found for the LINE1 family, only specific subfamilies are. Nonetheless, these subtle differences are negatively correlated with the DNA methylation levels at repeat families.

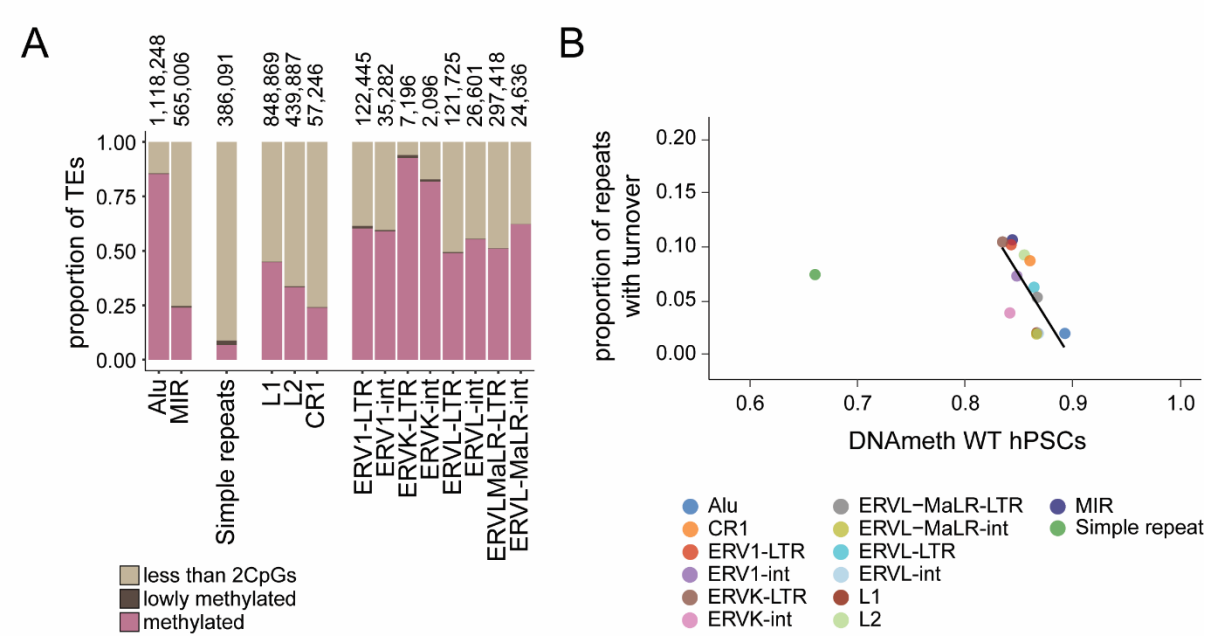

Figure 31: SINE-MIRs and LTRs of ERVK and ERV1 are most strongly associated with the DNA methylation turnover in hPSCs. A The bar graph displays the proportion of elements with less than 2 CpGs and of those being lowly methylated (<=0.3) or methylated (>0.3) in hPSCs for each repeat family, sorted by the repeat classes (SINEs, LINEs, LTRs). **B** The scatter plot displays the proportion of methylated elements that are subject to the DNA methylation turnover in hPSCs (Δmeth DKO-WT hPSCs<-0.3) (y-axis) compared to the average DNA methylation level of all elements with at least 2 CpGs (x-axis) for each repeat family. The regression line displays the linear model for all data points, except for simple repeats.

DNA methylation turnover targeting substantially varies among subfamilies of ERVK LTRs, ERV1 LTRs, simple repeats and 5'ends of LINE1s

Upon subdividing the repeat elements into their distinct subfamilies, substantial variations in DNA methylation turnover targeting become apparent (**Figure 32** A). Particularly, LTR subfamilies of ERVK and ERV1 diverge the most, with the hominoid-specific ERVK LTR5-Hs and ERV1 LTR7 being most frequently associated with the DNA methylation turnover, out of all investigated repeat subfamilies (59.7% and 54%, respectively) (**Figure 32** A-C). On the contrary, subfamilies, such as ERVK MER11A, exhibit minimal dependence on the DNA methylation turnover (0.2%) and ERV1 LTR12B elements are completely unaffected by the DNA methylation turnover (**Figure 32** A, B, C). Other subfamilies of ERVK and ERV1 range between these extremes (**Figure 32** B, C). Of note, even closely related subfamilies, like ERVK LTR5-Hs and their evolutionary progenitors, LTR5A (0.4%) and LTR5B (10.5%), can display pronounced discrepancies. In contrast, other clades like the HERV9 clade, including ERV1 LTR12B (0%), LTR12 (0.2%), LTR12_ (0.4%), and LTR12D (0.5%), demonstrate consistency

in showing substantial resistance against the DNA methylation turnover in hPSCs (**Figure 32** B) (Subramanian et al., 2011; Kojima, 2018). In contrast to the subfamilies of ERV1 and ERVK LTRs, those belonging to ERVL and ERVL-MaLR are similarly unaffected by the DNA methylation turnover (**Figure 32** D, E). In conclusion, the substantial effect of the DNA methylation turnover on specific subfamilies of ERV1 LTRs and ERVK LTRs suggests a subfamily-specific DNMT3/TET recruiting mechanism.

Furthermore, the DNA methylation turnover targeting of the internal parts of ERVK and ERV1 LTR transposons is more similar between subfamilies as compared to the LTRs themselves, underscoring a generally reduced association of internal fragments with the DNA methylation turnover in hPSCs (**Figure 32** B, C). In many cases, the targeting tendencies of internal fragments and their respective LTRs align. For instance, HERVK22-int, displaying the highest proportion of elements with DNA methylation turnover among all internal subfamilies of ERVK, mirrors the strong association of LTR22B and LTR22C with the DNA methylation turnover in hPSCs (**Figure 32** C). (Kojima, 2018). Similarly, LTR7 and their related internal subfamily HERVH-int are both more frequently targeted by the DNA methylation turnover compared to other internal and LTR subfamilies (**Figure 32** B). Further, the internal fragment of ERV1 LTR12 (HERV9-int) ranks among the least associated internal parts, similar to the pattern observed for LTR12. However, LTR14C and HERVK14C-int display opposite trends, with none of the HERVK14C-int copies being subject to the DNA methylation turnover while LTR14C was ranked within the highest associated LTRs of ERVK (**Figure 32** C). Thus, these observations indicate that internal fragments of LTR retrotransposons generally miss targeting mechanisms for the DNA methylation turnover.

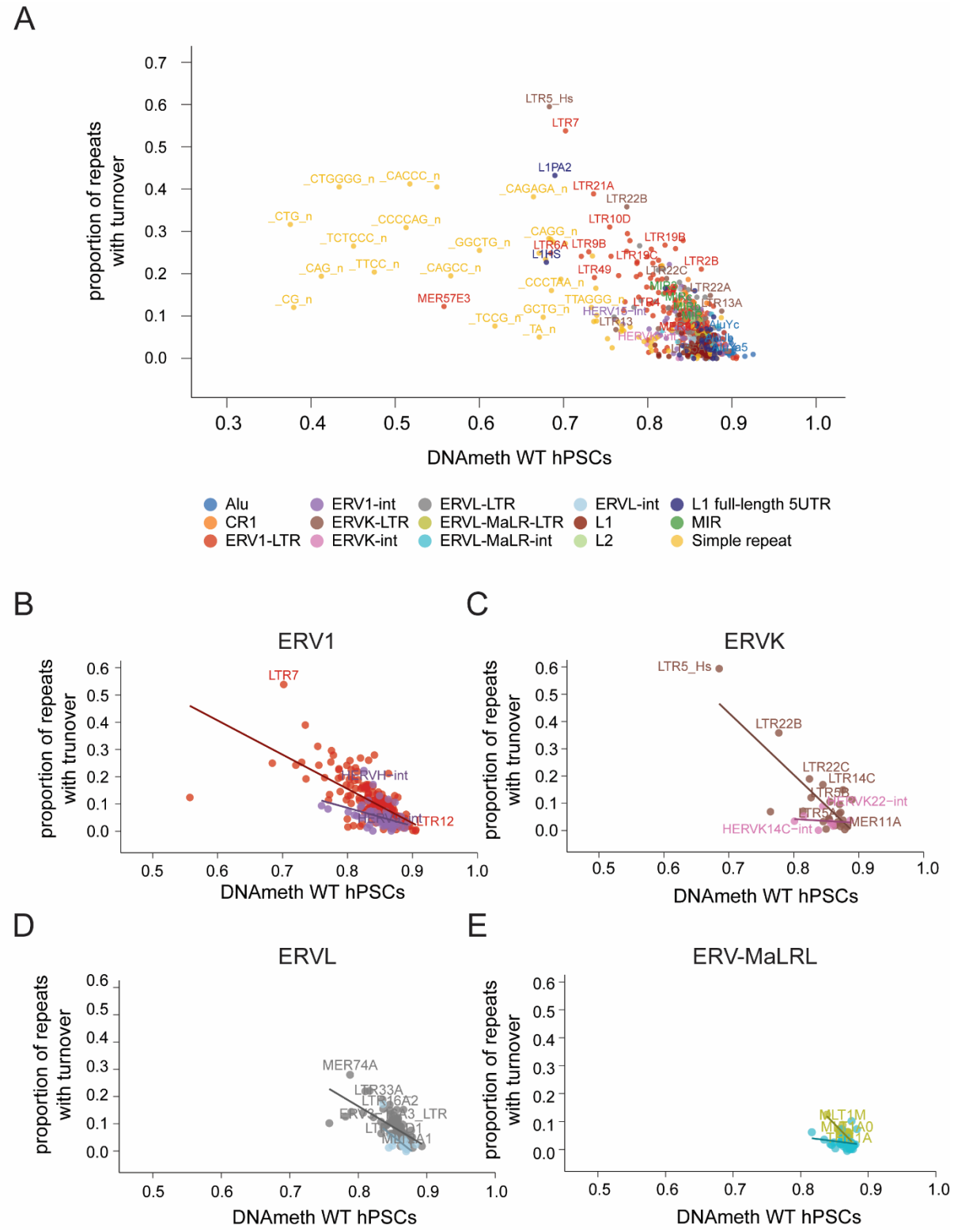
A
proportion of repeats with turnover
DNAmeth WT hPSCs
LTR5_Hs
LTR7
L1PA2
LTR21A
LTR22B
LTR10D
LTR19B
LTR2B
LTR22A
LTR13A
LTR12
MER57E3
Alu
CR1
ERV1-LTR
ERV1-int
ERVK-LTR
ERVK-int
ERVL-LTR
ERVL-MaLR-LTR
ERVL-MaLR-int
ERVL-int
L1
L2
L1 full-length 5UTR
MIR
Simple repeat
B
ERV1
C
ERVK
HERVK14C-int
MER11A
D
ERVL
MER74A
E
ERV-MaLRL

Figure 32: ERVK LTR5-Hs and ERV1 LTR7 are substantially associated with the DNA methylation turnover in hPSCs. A-E The scatter plots display the proportion of methylated elements which are subject to the DNA methylation turnover in hPSCs (Δmeth DKO-WT hPSCs<-0.3) (y-axis) compared to the average DNA methylation level of all elements with at least 2 CpGs (x-axis) for all repeat subfamilies, colored by their belonging to their respective repeat class (Alus, MIRs, L1, L2, CR1, ERV1, ERVK, ERVL, ERVL-MaLR and simple repeats) (A) or for all subfamilies belonging to ERV1 (B), ERVK (C), ERVL (D) or ERVL-MaLR (E). Additionally, the plots in (B-E) include regression lines for LTR and internal fragments. ERV1-LTRs: Spearman's ρ =-0.730, p<0.0001. ERV1-int: Spearman's ρ =-0.479, p=0.0004. ERVK-LTRs: Spearman's ρ =-0.482, p=0.018. ERVK-int: Spearman's ρ =-0.428, p=0.354. ERVL-LTRs: Spearman's ρ =-0.786, p<0.0001. ERVL-int: Spearman's ρ =-0.537, p<0.028. ERVL-MaLR-LTRs: Spearman's ρ =-0.647, p<0.0001. ERVL-MaLR-int: Spearman's ρ =-0.298, p=0.070.

Similar to prior observations, the 5'ends of full-length LINE1 elements from distinct subfamilies exhibit varied targeting of the DNA methylation turnover in hPSCs (**Figure 33** A) (Charlton et al., 2020). Moreover, among L1PA subfamilies, there is a striking discrepancy, with certain subfamilies showing pronounced association with the DNA methylation turnover, while others a minimally targeted. In particular, nearly half of the L1PA2 5'ends are subject to the DNA methylation turnover in hPSCs (43.4%), followed by L1HS (L1PA1) 5'ends (22.8%) and L1PA3 (16.7%), whereas only a minority of e.g. L1PA16 (0.8%) and L1PA13 (0.8%) are. Similar patterns are seen for the L1PB subfamilies, with some being slightly related to the DNA methylation turnover (L1PB1) and others not at all (L1PB4) (**Figure 33** A). Of note, considering all the genomic fragments belonging to LINE1, subfamilies typically have negligible associations with the DNA methylation turnover, with only minor variations among them (**Figure 33** A). This is consistent with all LINE1 element parts other than 5'UTRs being rather independent of the DNA methylation turnover (Charlton et al., 2020). Furthermore, no subfamily of LINE2 and LINE-CR1 exhibits a noticeably stronger relationship with the DNA methylation turnover in hPSCs as compared to others, indicating that none of them is specifically targeted by the DNA methylation turnover (between 7.1%-12.4% and 6.4%-18.8%, respectively) (**Figure 33** A).

Furthermore, among the 38 investigated SINE-Alu subfamilies, none displayed an outstandingly high proportion of copies being subject to the DNA methylation turnover (between 0.5%-7.1%) (**Figure 33** B). This suggests that Alu elements are generally resistant to demethylation in DKO hPSCs. The relatively short full length of SINE-Alus (average ~261bp) makes it unlikely that studying isolated parts of Alu elements would yield divergent findings, contrasting with LINE1 observations (Charlton et al., 2020). Similarly, turnover targeting to

the four SINE-MIR subfamilies displays minor variations (between 8.2% for SINE-MIRs and 15.3% for SINE-MIR3). In contrast, subfamilies of simple repeats greatly vary in their relationship with the DNA methylation turnover in hPSCs (**Figure 33** B). Interestingly, as the GC content within the repetitive unit of a subfamily rises, the proportion of simple repeats with DNA methylation turnover follows (Spearman's ρ =-0.67, p<0.0001) (**Figure 33** C).

In summary, among all investigated subfamilies, elements of ERVK LTR5-Hs and ERV1 LTR7 are most frequently associated with the DNA methylation turnover in hPSCs. This suggests that there may be specific mechanisms governing this preferential targeting.

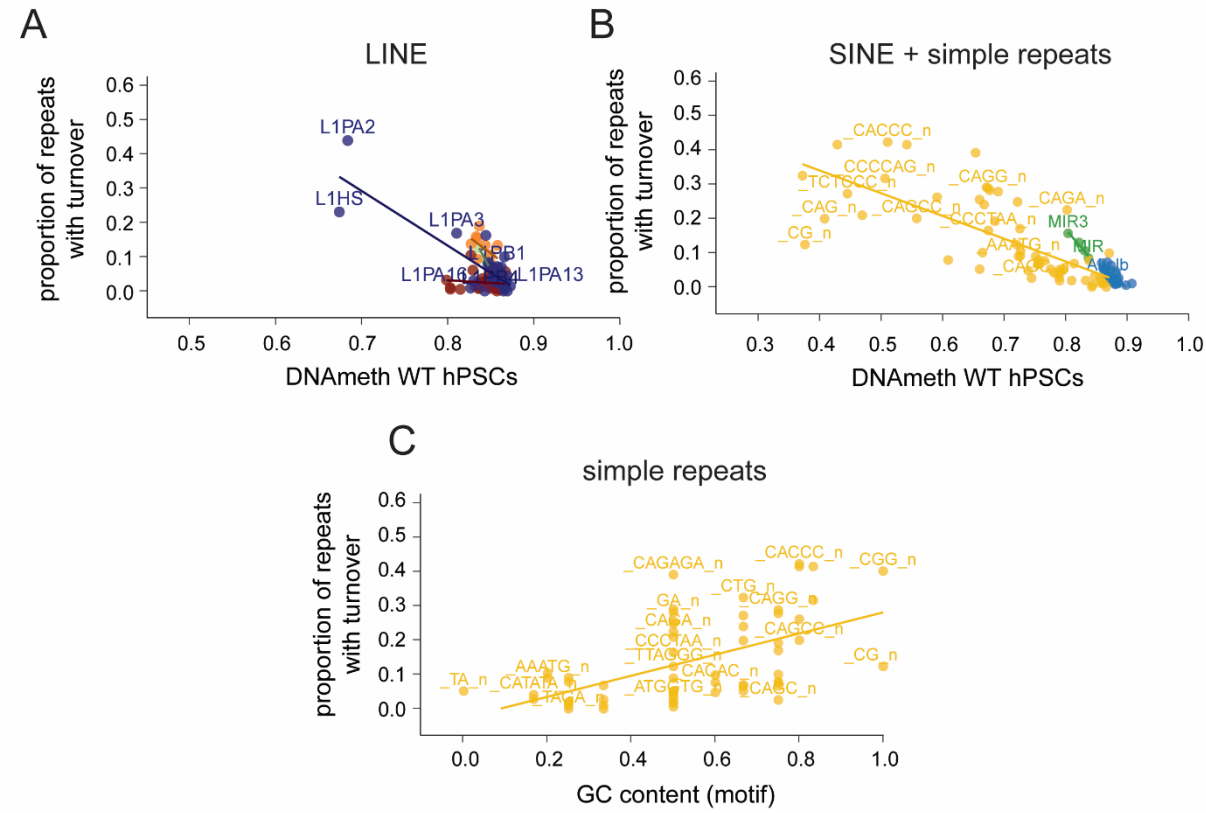

Figure 33: Subfamilies of LINE1-5'ends and simple repeats are differentially targeted by the DNA methylation turnover. A+B The scatter plots display the proportion of methylated elements which are subject to the DNA methylation turnover in hPSCs (Δmeth DKO-WT hPSCs<-0.3) (y-axis) compared to the average DNA methylation level of all elements with at least 2 CpGs (x-axis) for all repeat subfamilies belonging to LINE1/LINE2/CR1 (A) or SINE-Alu/-MIR/simple repeats (B), with additional regression lines. LINE1-5'ends: Spearman's ρ =-0.441, p=0.013. LINE1: Spearman's ρ =-0.253, p=0.006. LINE2: Spearman's ρ =-1, p=0.083. LINE1-CR1: Spearman's ρ =-0.552, p=0.104. SINE-Alu: Spearman's ρ =-0.686, p<0.0001. SINE-MIR: Spearman's ρ =-1, p=0.083. Simple repeats: Spearman's ρ =-843, p<0.0001. **C** The scatter plot displays the proportion of methylated elements to the GC content within the repetitive sequence of simple repeats for all subfamilies of simple repeats. (regression line: Spearman's ρ =-0.67, p<0.0001).

The DNA methylation levels at repeat subfamilies in hPSCs negatively correlate with the DNA methylation turnover

Reflecting the pattern observed at the level of repeat classes, there is a general inclination for subfamilies with higher rates of DNA methylation turnover to exhibit reduced DNA methylation levels in hPSCs (**Figure 32** and **Figure 33**). Particularly, subfamilies of ERV1-, ERVL- and ERVL-MaLR-LTRs, as well as those of internal parts of ERV1, SINE-Alus and simple repeats display a significant negative correlation between DNA methylation levels and the proportion of elements with DNA methylation turnover (ERV1-LTR: Spearman's ρ =-0.730, $p<0.0001$; ERVL-LTRs: Spearman's ρ =-0.786, $p<0.0001$; ERVL-MaLR-LTRs: Spearman's ρ =-0.647, $p<0.0001$; ERV1-int: Spearman's ρ =-0.479, $p=0.0004$; SINE-Alu: Spearman's ρ =-0.686, $p<0.0001$; simple repeats: Spearman's ρ =-843, $p<0.0001$). Furthermore, subfamilies of other repeat classes display similar trends, however, statistical significance is comparatively weak (Spearman's ρ >-0.5- or $p>0.01$) (**Figure 32** and **Figure 33**).

In summary, these observations suggest that the DNA methylation turnover may have an impact on DNA methylation levels of repeat elements.

DNA methylation turnover at simple repeat and LINE1-5'end subfamilies positively correlates with CpG-density and GC-content

In other genomic contexts, CpG-density and DNA methylation levels have been demonstrated to be negatively linked, which means that regions enriched for CpGs are more likely to be lowly methylated (Lienert et al., 2011; Weber et al., 2007; Illingworth et al., 2008). Interestingly, DKO-DMRs have higher CpG-densities, which suggests it may contribute to targeting of the DNA methylation turnover (Charlton et al., 2020). However, CpG density alone is neither predictive for DNA methylation levels nor for the DNA methylation turnover at DKO-DMRs (Lienert et al., 2011; Charlton et al., 2020). To rule out that repeat elements are an exception and that the DNA methylation turnover does not locally correlate with CpG-density, I looked into their relationship at different repeat families and subfamilies. Additionally, I examined the relationship between the GC content of repeats, outside of the CpG context, and the DNA methylation turnover, prompted by the observation that simple repeat subfamilies with high GC enrichment in their repetitive unit are more frequently associated with the DNA methylation turnover in hPSCs (**Figure 33** C).

The differential targeting of various repeat families by the DNA methylation turnover in hPSCs does not correlate with either CpG-density or GC content (Spearman's ρ =0.090, p=0.761) (**Figure 34**). Notably, even though ERVK-LTRs and ERV1-LTRs display relatively strong associations with the DNA methylation turnover, their CpG-densities fall within intermediate ranges, indicating that high CpG densities are not a prerequisite for the DNA methylation turnover (**Figure 34**). In line with this, SINE-Alus exhibit the highest average CpG-density while being least subject to the DNA methylation turnover. Furthermore, the comparable CpG-density shared between ERV1-LTR, LINE2, and LINE-CR1, which are yet distinctly targeted, underscore that DNA methylation turnover and CpG-density at repeats do not follow a linear correlation (**Figure 34**). Thus, this suggests that the CpG-density alone may not determine for targeting by the DNA methylation turnover.

Similarly, the GC content at repeat families also does not correlate with their susceptibility to the DNA methylation turnover (Spearman's ρ =0.301, p=0.295) (**Figure 34**). In particular, ERV1-LTRs and ERVK-LTRs display intermediate GC contents, which are similar to those of ERVL-MaLR-ints and L2 elements - families with a strong and a weak link to the DNA methylation turnover, respectively. Thus, as for the CpG-density, this implies that the GC content alone is not the key factor guiding the DNA methylation turnover to its targets.

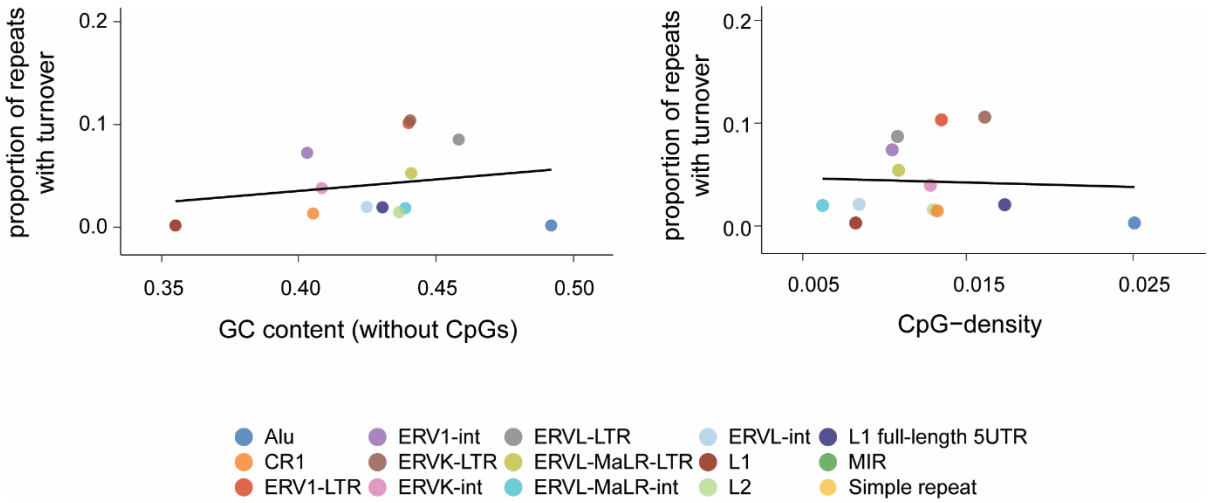

Figure 34: DNA methylation turnover and CpG-density/GC content are not correlated on the level of repeat families. The scatter plots display the proportion of methylated elements with DNA methylation turnover (y-axis) compared to the average GC content (left) or CpG-density (right) at repeat families, with additional regression lines (GC content: Spearman's ρ =0.090, p=0.761; CpG-density: Spearman's ρ =0.090, p=0.761).

However, among subfamilies of simple repeats, the proportion of repeats experiencing the DNA methylation turnover strongly correlates with the GC content of the entire repeat sequence, excluding CpGs, (Spearman's ρ =0.679, $p<0.0001$) and moderately increases with the CpG-density (Spearman's ρ =0.369, p=0.0029) (**Figure 35**). Similar trends are observed at subfamilies of LINE1-5'ends, wherein a more pronounced positive correlation is seen with the CpG-density compared to the GC content (CpG-density; Spearman's ρ =0.623, p=0.00018; GC content; Spearman's ρ =0.547, p=0.00143) (**Figure 35**). In contrast, neither the GC content nor the CpG-density of ERV1-LTR and ERVK-LTR subfamilies correlates with their levels of association with the DNA methylation turnover in hPSCs (**Figure 35**).

In summary, these findings rule out that the CpG-density or GC content are general key factors governing the target-specificity of the DNA methylation turnover to certain repeat elements in hPSCs. Nevertheless, these DNA sequence features might still play a role for specific repeat families, such as simple repeats and the 5'ends of LINE1s.

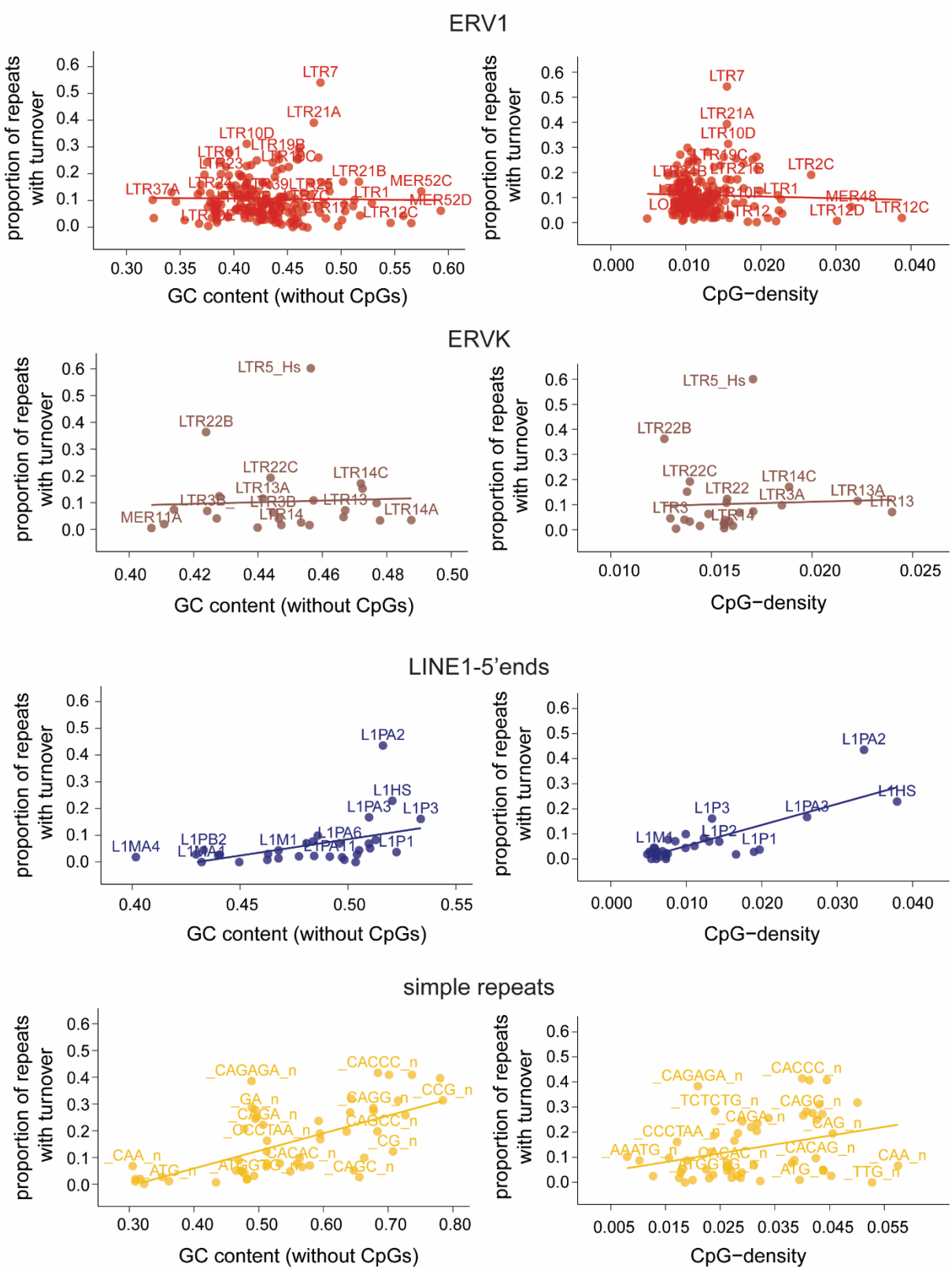
ERV1
proportion of repeats with turnover
GC content (without CpGs)
CpG−density
LTR7
LTR21A
LTR10D
LTR21B
MER52C
MER52D
LTR1
LTR12C
LTR2C
MER48
LTR12D
ERVK
LTR5_Hs
LTR22B
LTR22C
LTR14C
LTR13A
LTR14A
MER11A
LTR22
LTR3A
LTR13
LINE1-5'ends
L1PA2
L1HS
L1PA3
L1P3
L1MA4
L1PB2
L1M1
L1PA6
L1P1
L1P2
simple repeats
_CAGAGA_n
_CACCC_n
_CCG_n
_CAGG_n
_CAGCC_n
_CCCTAA_n
_CAA_n
_CACAC_n
_CAGC_n
_TCTCTG_n
_CACAG_n
_AAATG_n
_TTG_n

Figure 35: DNA methylation turnover and CpG-density/GC content positively correlate at simple repeats and L1-5'ends. The scatter plots display the proportion of methylated elements with DNA methylation turnover (y-axis) compared to the average GC content (left) or CpG-density (right) at subfamilies of ERV1-LTR, ERVK-LTR, LINE1-5'ends, and simple repeats, with additional regression lines. ERV1-LTR: GC content: Spearman's ρ =-0.081, p=0.267; CpG-density: Spearman's ρ =-0.082, p=0.263. ERVK-LTR: GC content: Spearman's ρ =-0.1365, p=0.523; CpG-density: Spearman's ρ =-0.245, p=0.247. L1-5'ends: GC content: Spearman's ρ =-0.574, p=0.0014; CpG-density: Spearman's ρ =-0.623, p=0.00018. Simple repeats: GC content: Spearman's ρ =0.679, p<0.0001; CpG-density: Spearman's ρ =0.369, p=0.0029.

Evolutionary young repeat subfamilies tend to be more associated with the DNA methylation turnover in hPSCs

The DNA methylation turnover in hPSCs has been demonstrated to more frequently target evolutionary young LINE1 elements than older relatives. This is supported by the finding of my analysis that the majority of the hominoid-specific ERVK LTR5-Hs exhibit substantial susceptibility to the DNA methylation turnover in hPSCs, while the older predecessor subfamily ERVK LTR5A is almost entirely independent of the turnover (**Figure 35**). To investigate whether *de novo* repeat insertions tend to be accompanied by the establishment of the DNA methylation turnover which may diminish over evolutionary time, I examined the relationship between the evolutionary age of retrotransposons and the occurrence of the DNA methylation turnover in hPSCs. In order to accomplish that, I analyzed the clades of L1PA and L1PB (5'ends), as well as ERVK LTR5 and ERV1 LTR7, the latter of which was recently reclassified based on their predicted evolutionary ages (Carter et al., 2022; Subramanian et al., 2011; Khan et al., 2006). Of note, the new categorization of LTR7 elements differs slightly from the classical RepBase classification used in my previous analysis (**Figure 34** and **Figure 35**).

Indeed, the DNA methylation turnover in hPSCs at the 5'ends of LINE1 L1PA subfamilies displays a significant negative correlation with their evolutionary age (Spearman's ρ =-0.718, p=0.0036) (**Figure 36**) (Khan et al., 2006). Specifically, the youngest subfamilies of the L1PA clade, including L1HS, L1PA2, and L1PA3 (2.3-15.8mya), show a higher propensity for the DNA methylation turnover, while older subfamilies are less frequently targeted (**Figure 36**). Similarly, the association with DNA methylation turnover at the older L1PB clade (34.3-121.6mya) declines from evolutionarily younger (L1PB1: 7.0%) to older subfamilies (L1PB4: 0%) (**Figure 36**) (Khan et al., 2006). However, for L1PB, this correlation lacks statistical significance, likely due to the small number of subfamilies (Spearman's ρ =-1, p=0.0833).

As indicated by my previous analysis, the DNA methylation turnover in hPSCs targets LTR5-Hs elements (6-8mya) substantially more frequently than their progenitors LTR5A (15-21mya) or LTR5B (19-25mya) (**Figure 36**) (Subramanian et al., 2011). However, like for L1PB, this association is not statistically significant (Spearman's ρ =-0.5, p=1). Furthermore, despite the older age predicted for LTR5B as compared to LTR5A, LTR5B's slightly more pronounced association with the DNA methylation turnover may not contradict the negative relationship between the DNA methylation turnover and evolutionary age, considering that the predicted time windows of both subfamilies overlap.

Consistent with the finding for LTR5 and L1PA-5'ends, the DNA methylation turnover at LTR7 subfamilies also occurs less frequently with increasing evolutionary age (**Figure 36**). In particular, the proportion of elements exhibiting the DNA methylation turnover increases from the oldest subfamily, LTR7c (11.2%) (>25mya), to the most recent one, LTR7up2 (67.6%) (10-14mya) (**Figure 36**) (Carter et al., 2022). Of note, LTR7o and LTR7bc arose at a similar time as LTR7c (all three: >25mya), with LTR7bc likely being slightly younger than LTR7o, which itself was predicted to be younger than LTR7c (**Figure 36**) (Carter et al., 2022). Similarly, LTR7up1's predicted emergence, at 10-14mya, parallels that of LTR7up2 but is possibly slightly delayed. These nuanced age predictions align with the diminishing DNA methylation turnover targeting as evolutionary age increases (**Figure 36**). Similar as for other clades with a small number of subfamilies, this relationship lacks statistical significance (Spearman's ρ =-0.866, p=0.058).

When examining subfamilies across all investigated clades collectively, the DNA methylation turnover follows an exponential decline with increasing evolutionary age and is largely absent at those which originated around 20mya and earlier (**Figure 36**). Of note, only LTR7bc and LTR7o, which were predicted to be older than 25mya, seem to defy this tendency.

In summary, for all investigated clades of retrotransposable elements, the DNA methylation turnover in hPSCs at subfamilies decreases with an increased amount of time that the subfamily is predicted to be present in the genome.

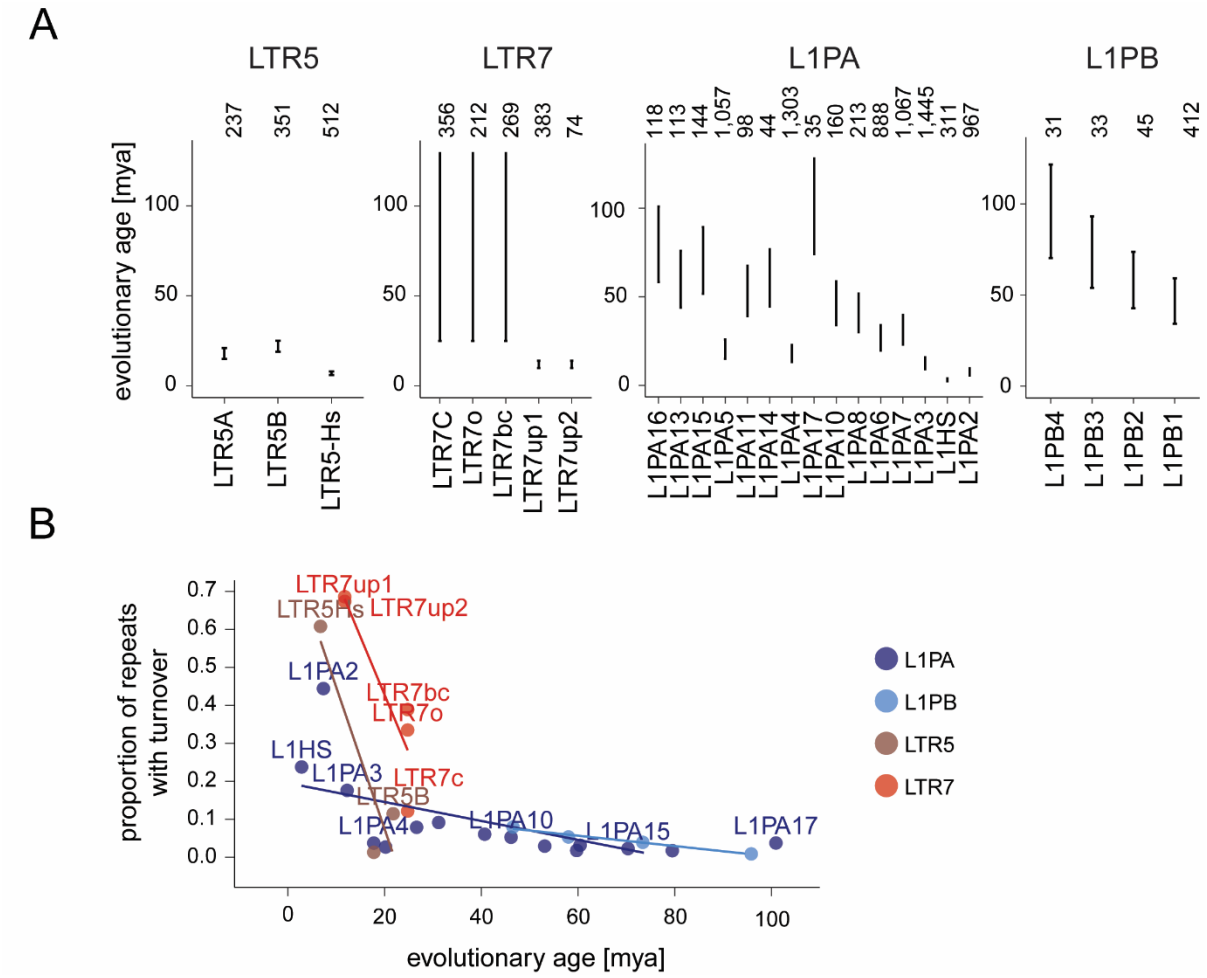

Figure 36: Increased DNA methylation turnover at evolutionarily younger ERVK, ERV1 and LINE1 subfamilies. A The plots display the time windows in which retrotransposon subfamilies are predicted to have arisen (mya = million years ago). The number of methylated elements of each subfamily is indicated above the plots. **B** The scatter plot displays the proportion of methylated elements with DNA methylation turnover (y-axis) compared to their evolutionary age (mean-age of the predicted time windows or the youngest predicted age for LTR7c, LTR7o, and LTR7bc which arose >25mya) at subfamilies belonging to ERV1-LTR7, ERVK-LTR5, and LINE1-5'ends of L1PA and L1PB, with additional regression lines. L1PA-5'ends: Spearman's ρ =-0.718, p=0.0036. L1PB-5'ends: Spearman's ρ =-1, p=0.0833. ERVK LTR5: Spearman's ρ =-0.5, p=1. ERV1 LTR7: Spearman's ρ =-0.866, p=0.058.

Further experimental efforts toward studying the role and mechanism of the DNA methylation turnover

Shedding light on the complex relationship between DNA methylation turnover and differentiation-related hypomethylation, as well as its differential occurrence at subfamilies of retrotransposons, prompted a series of fundamental questions. Looking into these questions promises a deeper understanding of the mechanisms and functional implications underlying the DNA methylation turnover.

One crucial question revolves around the function of TET enzymes in differentiation-related hypomethylation, both at regions with and without DNA methylation turnover. The extent to which this depends on TET activity remains elusive. One way to shed light on this involves the examination of hypomethylation during differentiation under conditions where TETs are depleted. Furthermore, the importance of DNMT3s counteracting TETs and maintaining high levels of DNA methylation at DKO-DMRs could be investigated in hPSCs depleted of DNMT3s. Moreover, unraveling the mechanistic components behind the selective co-targeting of TETs and DNMT3s could be achieved by investigating the DNA sequence contribution. Alternatively, the identification of interaction partners of DNMT3s and TETs could offer valuable insights into this intriguing selective mechanism.

As a next step towards dissecting the function and mechanism of the DNA methylation turnover in the context of these particular questions, I generated TET1- and DNMT3B-tagged lines in hiPSCs. Furthermore, I re-created a TET-triple knockout and DNMT3A/B-double knockout in an isogenic background to ensure comparability across experiments.

Generating DNMT3A/B-DKO hiPSCs to study the role of DNA methylation at DKO-DMRs

The absence of DNMT3s in hPSCs significantly impacts the cellular transcriptome, with a majority of differentially expressed genes being linked to DKO-DMRs (Charlton et al., 2020). This suggests that DNA methylation at DKO-DMRs plays a role in gene expression regulation. Nevertheless, beyond its potential impact on the cellular transcriptome, the precise role of high DNA methylation levels at DKO-DMRs remains enigmatic. Previous studies have shown that DNA methylation counteracts chromatin opening at certain enhancers in mESCs (Kreibich et al., 2022). Additionally, DNMT3A's interaction with histone deacetylases suggests that demethylation of DKO-DMRs could impact various chromatin features in these regions

(Taunton et al., 1996; Fuks et al., 2001). Moreover, distinct TFs exhibit a preference for methylated DNA binding sites, while DNA methylation prevents other TFs from DNA binding, implying that loss of DNA methylation at DKO-DMRs could have complex consequences on TF binding patterns (Yin et al., 2017; Kaluscha et al., 2022). With the ultimate goal to study consequences of DKO-DMR demethylation, I engineered single-knockouts and a double-knockout for DNMT3A and DNMT3B in hiPSCs using Cas9-mediated gene targeting, the latter of which causes the initially observed local demethylation of DKO-DMRs (Charlton et al., 2020).

Generating DNMT3AKO hiPSCs

For mutating DNMT3A in hiPSCs, exon7, and exon19 were co-targeted since the three major protein-coding transcript variants of DNMT3A contain both of these exons; only the short isoform of DNMT3A does not (**Figure 37** A) (Simon Lauer, unpublished). After transfection and selection, genotyping for exon7 revealed a high knockout rate, with four out of seven clones carrying homozygous out-of-frame mutations (C2, C6, C7), implying pre-mature stop-codons (**Figure 37** B) (**Table 11**). Indeed, a western blot for DNMT3A confirmed depletion on the protein level in C2 and C7, but not C6 (**Figure 37** C) Precisely, except for one weak band, all wildtype bands (>64kDa) disappeared in C2 and C7. Thus, the genotyping and western blot results indicate successful knockout of DNMT3A in C2 and C7.

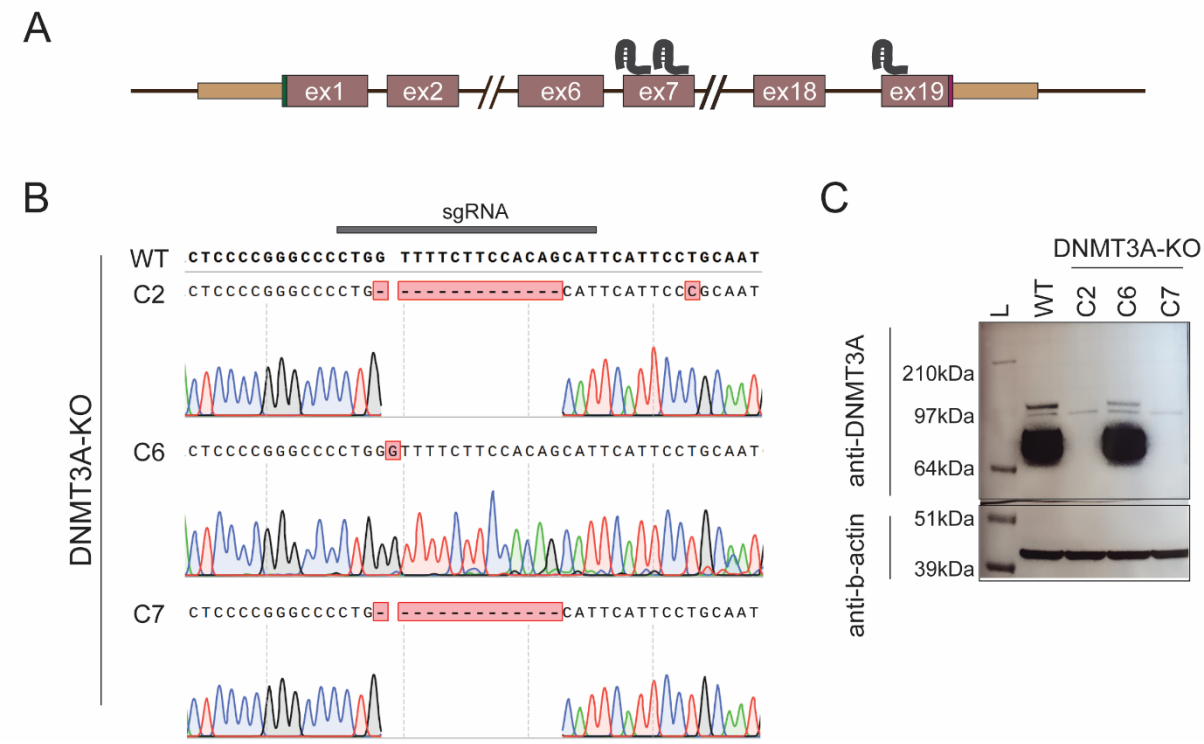

Figure 37: Genotyping of DNMT3A knockout in hiPSCs. **A** The schematic displays the strategy for mutating exon7 and exon19 of DNMT3A using Cas9-mediated targeting (design by Simon Lauer, unpublished). **B** The sequencing tracks display a part of exon7 surrounding the targeting region of one of the sgRNAs in WT and DNMT4A-KO clones C2, C6, and C7. The red labeling highlights differences between WT and the respective knockout clone. **C** The western blot displays the loss of DNMT3A bands in DNMT3A-KO C2 and C7, but not in C7. L: ladder.

Generating DNMT3BKO hiPSCs

Disruption of DNMT3B was obtained by Cas9-targeting to exon2 and exon21 using the same general strategy as for DNMT3A (design and confirmation by Simon Lauer, unpublished) (**Figure 38** A). Exon 2 is contained in all protein-coding transcript variants of DNMT3B and thus, allows to deplete all DNMT3B isoforms simultaneously. Similar to the DNMT3A knockout, the rate of homozygous frameshift mutations in exon2 was high (four out of nine clones) (**Figure 38** B) (**Table 11**). Subsequently, three potentially homozygous clones (C1, C5, C6) were examined for DNMT3B depletion on the protein level in a western blot (**Figure 38** C). In line with the genotyping results, each wildtype band of DNMT3B vanished in each of the three knockout candidates, confirming the successful knockout.

Generating DNMT3A/B-DKO hiPSCs

To create the DNMT3 double-knockout in hiPSCs, DNMT3B was targeted by Cas9 in the DNMT3A-KO(C2) using the same approach as for DNMT3B single-knockouts (**Figure 37** A and **Figure 38** A). The efficiency to obtain homozygous out-of-frame mutations was high (eight out of sixteen) and a subsequent western blot, confirmed successful depletion of the DNMT3B protein in all three examined homozygous-knockout clones (C3, C7, C11) (**Figure 38** C) (**Table 11**).

To summarize, I generated single-, and double-knockouts of DNMT3A/B in hiPSCs. These cell lines may be used in comparative experiments with wildtype hiPSCs to reveal causal links between DNA methylation levels at DKO-DMRs and other local chromatin features. Further, for each genotype, multiple clones are available to rule out clone-specific effects.

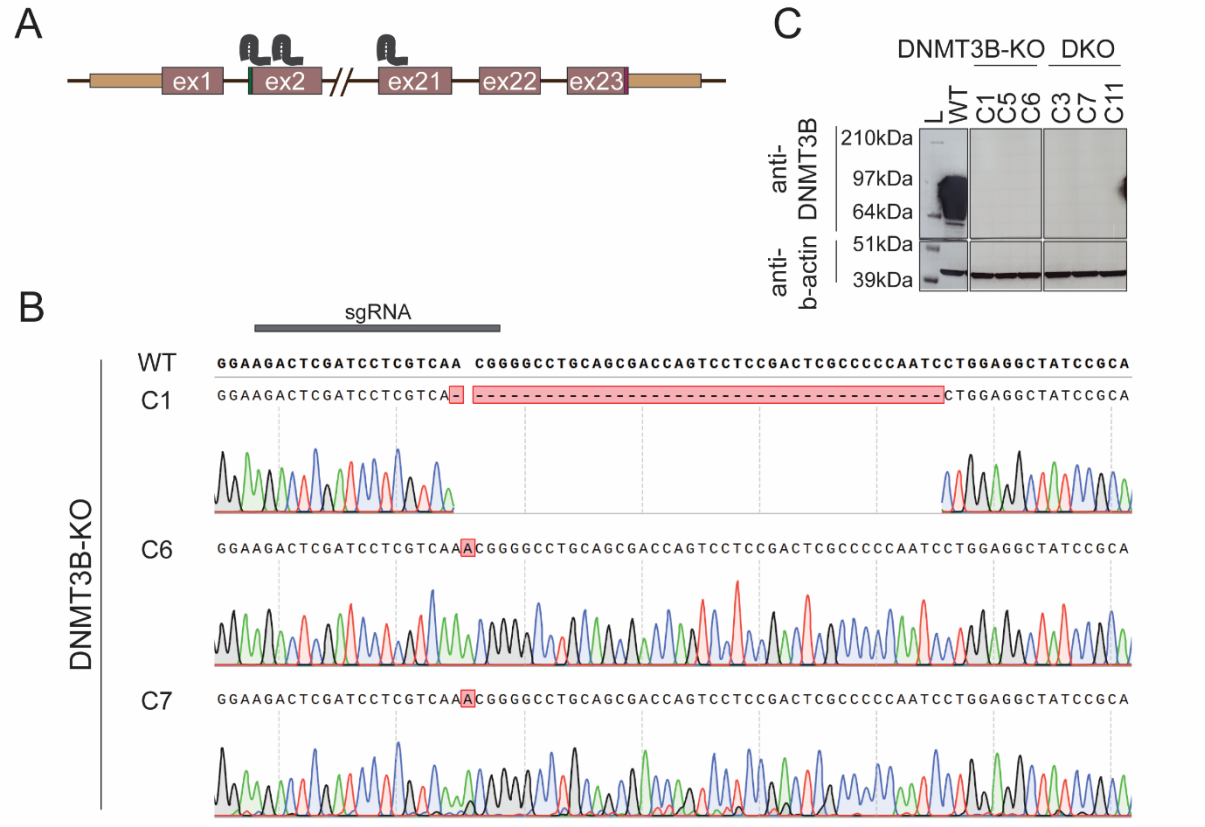

Figure 38: Genotyping of DNMT3B knockout in WT and DNMT3A-KO hiPSCs. **A** The schematic displays the strategy for mutating exon2 and exon21 of DNMT3B using Cas9-mediated targeting (the start codon is shown in exon2 because exon1 is not contained in the majority of DNMT3B isoforms including the majorly expressed ones). **B** The sequencing tracks display a part of exon2 surrounding the targeting region of one of the sgRNAs for WT and DNMT3B-KO (single knockout) clones C1, C6, and C7. The red labeling highlights differences between WT and the respective knockout clone. The DNMT3B exon2 of DNMT3A/B-DKO hiPSCs have the same genotype as DNMT3B-KO C6 and C7 (not shown). **C** The western blot displays the loss of DNMT3B bands in all DNMT3B single-knockout clones C1, C5, and C6 and the DNMT3A/B-DKO clones C3, C7 and C11 and C7. L: ladder.

DNMT3A targeting in WT (7 genotyped)

WT	hom	het	unclear
1	0 at least one iF 3 both with FS	0 iF 1 FS	2

DNMT3B targeting in WT (9 genotyped)

WT	hom	het	unclear
0	1 at least one iF 4 both with FS	0 iF 0 FS	4

DNMT3B targeting in DNMT3A-KO (16 genotyped)

WT	hom	het	unclear
4	0 at least one iF 8 both with FS	0 iF 2 FS	2

Table 11: Genotyping quantification of DNMT3A- and DNMT3B-single and -double knockouts. The table displays the number of genotyped clones which did not display mutations in the targeted exons (WT), those that either carry heterozygous (het) or homozygous (hom) mutations, and those with unclear genotypes. Further, within the homozygous and heterozygous clones, I distinguished frameshift (FS) mutations from those in frame (iF).

Efforts towards studying the role of TET activity for early and late differentiation

Genomic regions that undergo hypomethylation during differentiation are frequently subject to DNA methylation turnover in hPSCs, particularly when demethylation occurs in early cell types rather than more mature stages (**Figure 23**). Nevertheless, the necessity of TET activity for demethylation during differentiation is limited to a handful of studies, including investigations of hematopoietic cells, B-cells, and Purkinje cells (Orlanski et al., 2016; Suzuki et al., 2017; Stoyanova et al., 2021). To examine the process of active demethylation upon exit of pluripotency, I generated a triple-knockout of all three TETs in hiPSCs. This approach has the potential to distinguish between TET-dependent and TET-independent passive demethylation. However, investigating demethylation in later stages of differentiation is limited within the TET-TKO hiPSC model, as early differentiation may already be compromised. To circumvent this limitation, I additionally initiated the generation of an inducible TET-triple-knockout system.

Generating TET-TKO hiPSCs

Using the Cas9 system, the TET-triple knockout was generated in hiPSCs, employing established sgRNAs (**Figure 39** A) (Verma et al., 2018). Initially, the focus was on simultaneously targeting TET1 and TET3, with subsequent efforts to disrupt TET2 (**Figure 39** B) Remarkably, the rate of homozygous frameshift mutations in the TET1 or TET3 gene was high (TET1: 10 of 75; TET3: 13 of 26), of which seven occurred pairwise, resulting in seven clones with homozygous frameshift mutations in both genes (**Figure 39** C) (**Table 12**). To establish the TET-TKO hiPSC line, TET2 was targeted in two TET1/3-DKO clones (Cp1#12 and Cp2#16). Similar to TET1 and TET3, the efficiency to obtain clones with homozygous frameshift mutations was high (4 of 13 genotyped clones), resulting in the creation of four TET-TKO hiPSC lines. Since TET2 and TET3 expression levels are relatively low in hPSCs, only TET1 depletion on the protein level can be validated in these cells in the future.

In summary, the successful establishment of TET triple-knockout in hiPSCs was confirmed on the genetic level across four clones. This enables the investigation of the role of TET activity in shaping the DNA methylome during early lineage decisions of hPSCs.

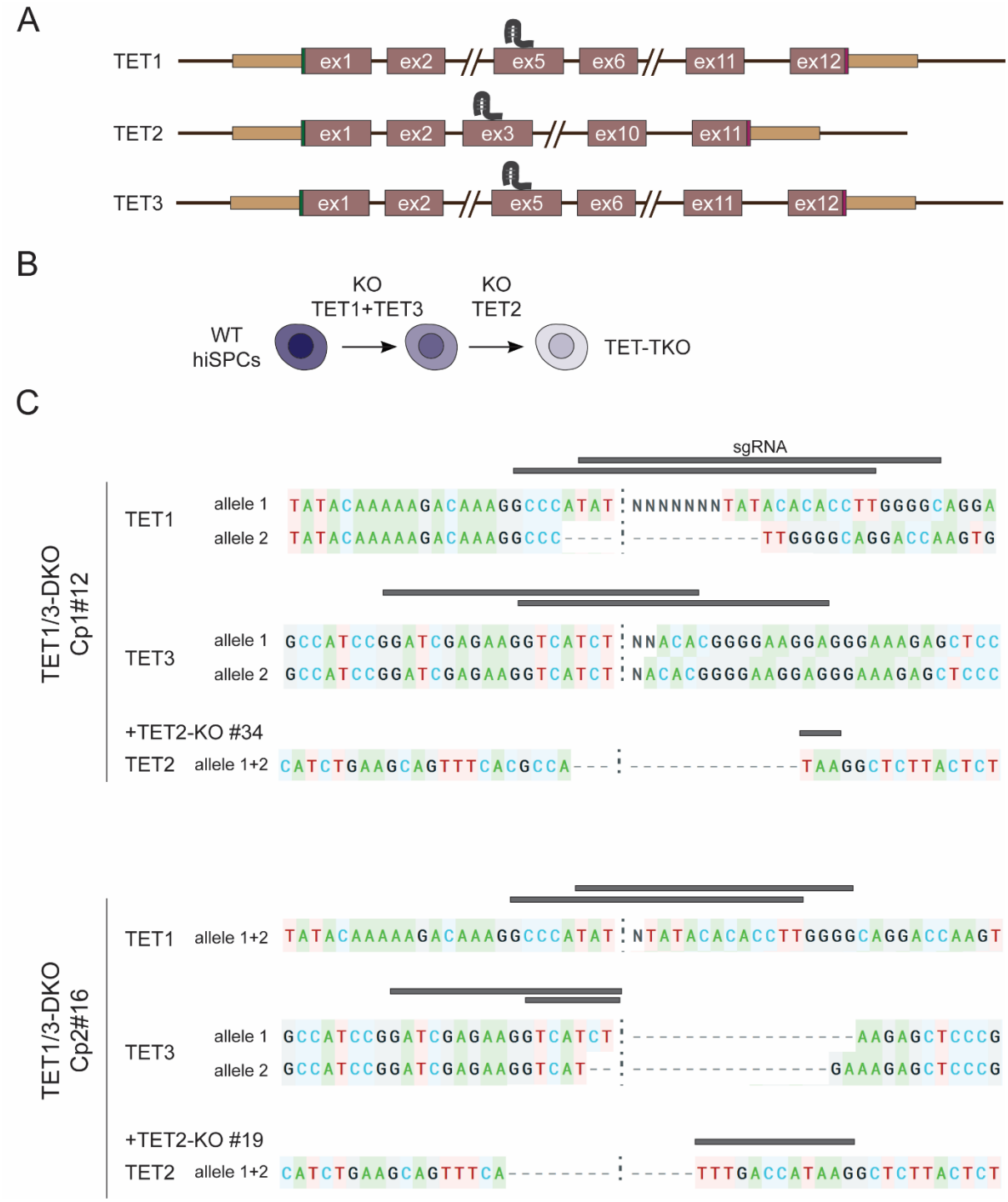

Figure 39: Genotyping of TET triple-knockout hiPSC lines. A+B The schematics display the strategy for mutating TET1 exon5, TET2 exon3, and TET3 exon5 using Cas9-mediated targeting (A) and the stepwise procedure of co-targeting TET1+TET3 first, followed by TET2 (B). **C** displays the Sanger-sequencing results for genotyping of TET1/3-DKO clones Cp1#12 and Cp2#16 evaluated by Synthego ICE. Horizontally dashed lines display deletions and (N)n mark insertions. The wildtype sequence is not shown. The black bars label the alignment of respective sgRNAs. In case a part of the sequence has been deleted, the sgRNA is shortened. Further, for each TET1/3-DKO clone, one example of additional TET2 knockout is shown.

TET1+TET3 targeting in WT

TET1 (75 genotyped)				TET3 (26 genotyped)			
WT	hom	het	unclear	WT	hom	het	unclear
31	4 at least one iF 10 both with FS	4 iF 9 FS	16	5	1 at least one iF 13 both with FS	0 iF 4 FS	3

TET2 targeting in TET1/3-DKO (13 genotyped)

WT	hom	het	unclear
	1 at least one iF 4 both with FS	0 iF 0 FS	8

Table 12: Genotyping quantification of TET1-3 triple knockout in hiPSCs. The table displays the number of genotyped clones which did not display mutations in the targeted exons (WT), those that either carry heterozygous (het) or homozygous (hom) mutations, and those with unclear genotypes. Further, within the homozygous and heterozygous clones, I distinguished frameshift (FS) mutations from those in frame (iF).

Efforts towards generating an inducible TET-TKO hiPSC line

The approach to create an inducible TET-TKO hiPSC line combines a system for the acute depletion of TET1 and the full knockout of TET2 and TET3. Specifically, the objective was to introduce an FKBP into the endogenous locus of TET1, in order to enable its degradation upon the addition of dTag13 to the culture medium. Subsequently, the knockout of TET2 and TET3 has to be implemented in the future to complete the system.

To begin with the proposed strategy, Cas9-mediated homology-directed repair was employed to knockin the FKBP-FLAG-cassette downstream of the TET1 start codon (**Figure 40** A). Genotyping via PCR identified potentially homozygous clones (6 of 46 clones), characterized by a shifted band size in the absence of the wildtype band (**Figure 40** B) (**Table 13**). Additionally, many other clones displayed insertion bands; however, heterozygosity is insufficient for complete degradation of TET1. Subsequently, three of the potentially homozygous clones (C12, C18, C39) underwent genotyping by Sanger sequencing, confirming the correct sequence of the integrated fragment and its integration site (not shown). In the future, the functional degradation of TET1 through addition of dTag13 requires verification, followed by the knockout of TET2 and TET3. Once these steps are completed, this system will allow the controlled regulation of TET activity to study their role for demethylation during cellular differentiation.

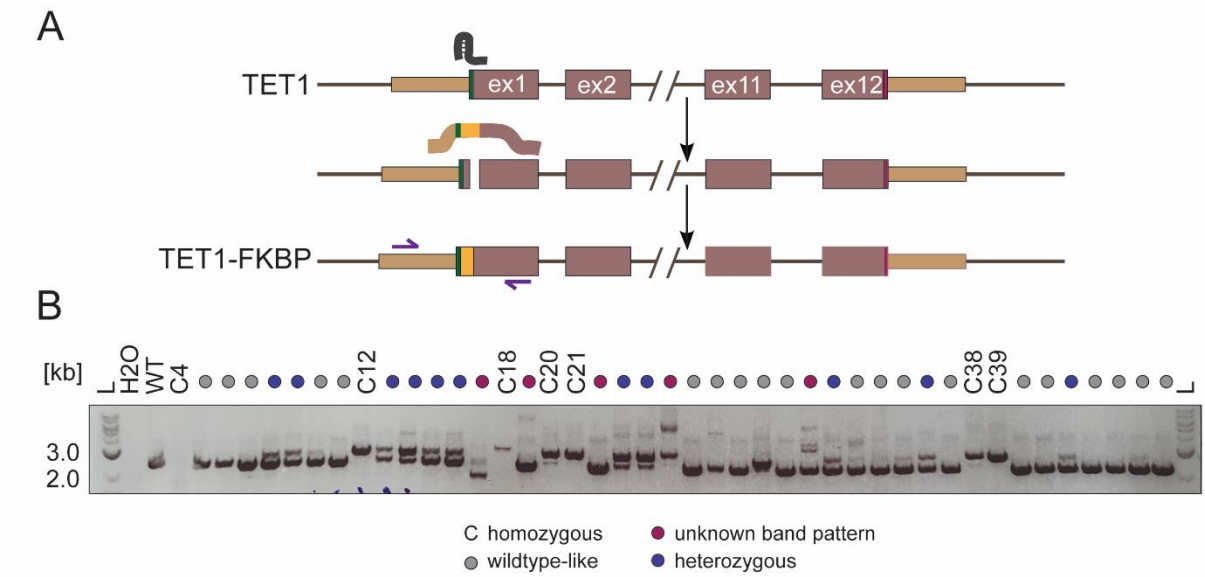

Figure 40: Genotyping of TET1-FKBP-FLAG candidate clones. A The schematic shows homology-directed repair induced by Cas9-targeting of exon1, downstream of the TET1 start codon (green) (stop codon in red). A repair construct with homology arms to the upstream and downstream sequence of the start codon enables the insertion of the FKBP-FLAG-domain (yellow) by homology-directed repair. Purple arrows indicate the annealing sites of the genotyping primers used in (B). **B** The agarose gel contains the amplified exon1 of TET1 of each clone. The wildtype band has a size of 2.2kb, while a band size of 2.7kb indicates successful insertion of the FKBP-FLAG-cassette. Wildtype-like clones are indicated with grey dots, potential heterozygous insertions with blue dots, and band patterns of unknown origin with pink dots. Potential homozygous candidates are labeled with their clone number (except for C4). L: ladder.

TET1-FKBP-FLAG (46 genotyped)
23 WT
11 potential het
6 potential hom
6 unknown band patterns

Table 13: Genotyping quantification of TET1-FKBP-FLAG candidate hiPSCs. The table displays the number of genotyped clones targeted for insertion of an FKBP-FLAG cassette behind the start codon of the endogenous TET1 gene. The quantification is based on the agarose gel in **Figure 40** and groups the clones with a wildtype (WT) band pattern, a wildtype band plus integration band (heterozygous, het) and integration band only (homozygous, hom)

Generating a DNMT3B-FLAG hiPSC line to study DNMT3B interaction partners

Our current understanding of DNA methylation turnover is constrained by the lack of knowledge concerning not only its functional relevance but also the mechanism that governs the specificity of DNMT3/TET co-recruitment. Generally, insights into the DNA sequence contribution or chromatin features of the target loci essential for DNMT3 and TET targeting are limited (Zhang et al., 2010b; Ooi et al., 2007; Baubec et al., 2015; Loaeza-Loaeza et al., 2020). Furthermore, the possibility exists that sequence-specific factors, such as transcription factors or long non-coding RNAs, play a role in guiding DNMT3s/TETs to their destination. Supported by the fact that DKO-DMRs are enriched for multiple TF binding motifs, I pursued finding protein interaction partners of DNMT3s and TETs (Charlton et al., 2020). In an effort to identify the protein-interactome of DNMT3B in hiPSCs, I introduced an endogenous FLAG-tag to DNMT3B, enabling a clean protein-pulldown procedure. In the future, co-immunoprecipitation followed by mass spectrometry analysis will be conducted to determine the protein interaction partners of DNMT3B. As for TET1, the previously generated TET1-FKBP hiSPC line is suitable for this experimental strategy, as it already contains a FLAG-tag.

Once again, the knockin strategy for introducing a FLAG-sequence upstream of the DNMT3B stop codon utilized Cas9-mediated homology-directed repair (**Figure 41** A). Genotyping identified a potential homozygous clone (C19) and five heterozygous clones (C22, C24 and C5, C13, C15 are not shown) out of 24 clones in total (**Figure 41** B) (**Table 14**). Subsequent Sanger sequencing confirmed the correct sequence of the inserted fragment and insertion site in the homozygous clone (not shown). Furthermore, the FLAG epitope was detected on a western blot at a size similar to DNMT3B (>64kDa), validating the presence of the FLAG at the endogenous DNMT3B (**Figure 41** C).

This achievement sets the foundation for future identification of the protein-interactome of DNMT3B in hiPSCs, potentially providing valuable insights into mechanisms guiding DNMT3B to specific genomic loci.

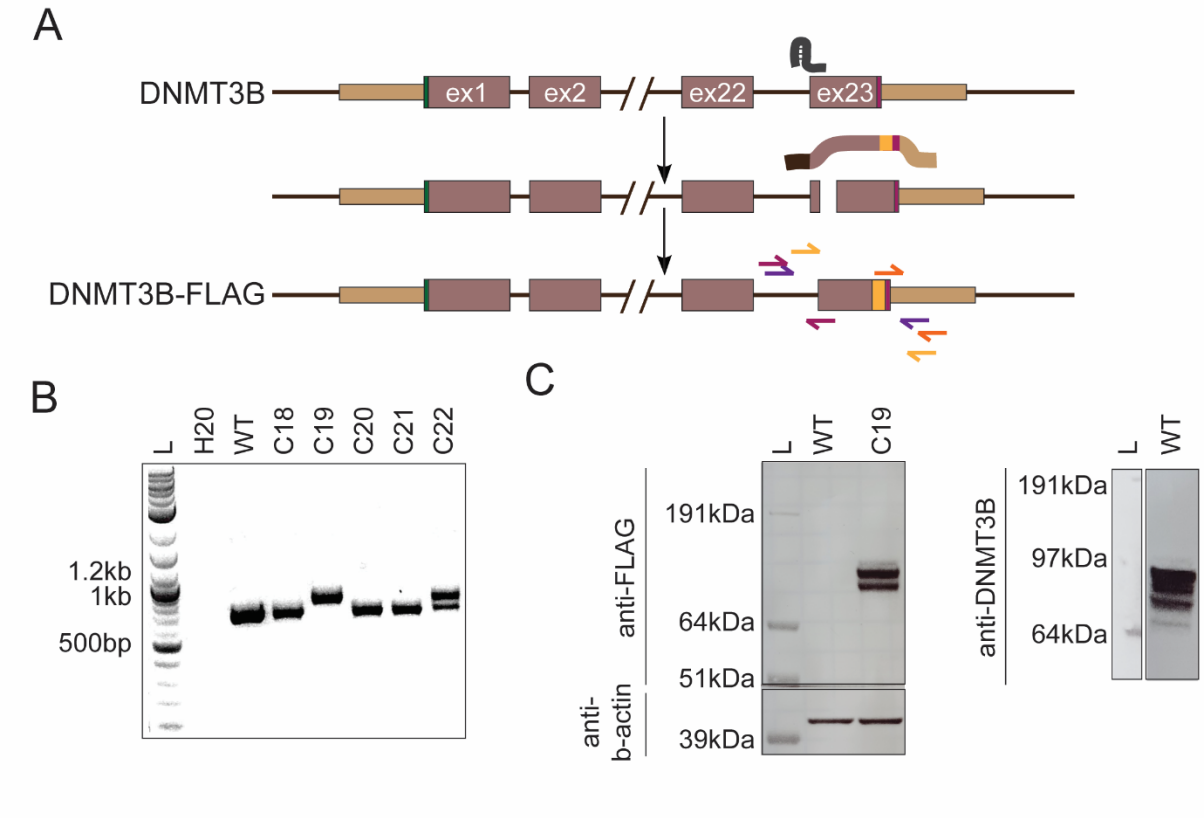

Figure 41: FLAG-tagging of the endogenous DNMT3B. A The schematic shows the strategy for inserting a FLAG-sequence before the stop codon (red) of the endogenous DNMT3B. Cas9-targeting of exon1 mediates homology-directed repair with a cassette consisting of the FLAG-sequence, nested between homology arms to the upstream and downstream sequence of the cutting site. Colored arrows indicate the annealing sites of the genotyping primers. (start codon in green) **B** The genotyping gel contains the amplified exon23 (yellow primer pair in (A)) of an unmodified wildtype control (WT) and multiple candidate clones. The wildtype band has a size of 695bp and successful insertion of the FLAG-sequence shifts it to 804bp. **C** The western blot with an antibody against FLAG reveals the FLAG-epitope in DNMT3B-FLAG(C19) but not wildtype hiPSCs (left). The second western blot against DNMT3B shows the protein size of the endogenous DNMT3B protein (right). L: ladder.

DNMT3B-FLAG in hiPSCs (24 genotyped)
18 WT
5 potential het
1 potential hom

Table 14: Genotyping quantification of FLAG-tagging DNMT3B in hiPSCs. The table displays the number of genotyped clones targeted for insertion of a FLAG-sequence before the stop codon of the endogenous DNMT3B gene. The quantification is based on the agarose gel in **Figure 41** and groups the clones with a wildtype (WT) band pattern, a wildtype band plus integration band (heterozygous, het) and integration band only (homozygous, hom)

DISCUSSION AND OUTLOOK

The study of DNA methylation and its role in various processes during mammalian embryogenesis has been a subject of extensive research. However, despite this considerable attention, generalizing its function and the underlying mechanisms guiding methyltransferases and demethylases to their genomic targets remain enigmatic. The dynamic nature of the DNA methylation turnover is particularly intriguing, given the simultaneous and balanced activity of DNMT3s and TETs at co-targeted genomic locations. My project was motivated by the significant overlap between DNA methylation turnover and putative somatic enhancers, prompting an investigation into the role of this dynamic during cellular differentiation. Moreover, the observed turnover at evolutionary young LINE1 L1PA repeat elements led to my in-depth exploration of the bias of DNA methylation turnover towards certain groups of repeats. Overall, I found a remarkable association between local hypomethylation events during early differentiation and the DNA methylation turnover in hPSCs. Notably, regions that undergo hypomethylation in more mature cell types are less frequent targets of the DNA methylation turnover in hPSCs. Instead, these regions tend to initiate turnover activity after exiting from pluripotency, in what I defined as "*de novo* DNA methylation turnover", an important molecular dynamic that I present in this study for the first time. Moreover, the study highlights specific differentiation systems, like stepwise *in vitro* derivation of mid radial glial and pancreatic cells, that can be further utilized to dissect the function of the DNA methylation turnover. Finally, I established novel cell systems to investigate the DNA methylation turnover at the mechanistic level in the future. While studying this phenomenon at various genomic locations, I confirmed the association between evolutionary young retrotransposons and the DNA methylation turnover and extended this observation to encompass elements belonging to LINE L1PB, ERV1 LTR7, and ERVK LTR5. Previously, evolutionary young LINE1 L1PA subfamilies were shown to miss a specific DNA sequence motif, which occurs in older elements, hypothesized to influence the DNA methylation turnover in this genomic context(Charlton et al., 2020; Hermant & Torres-Padilla, 2021). Along this line, my in-depth analysis, which includes all retrotransposon families and their subfamilies, offers a new lens to identify potentially relevant sequences within the repeats which could be responsible for the differential DNA methylation turnover observed in young and old repetitive elements.

DNA methylation turnover: a potential mechanism to maintain DNA methylation plasticity during development

The remarkable association between DNA methylation turnover and differentiation-related hypomethylation, along with its biases towards early differentiation or against certain lineages sparks speculation about its potential functions during differentiation. Furthermore, the overall stable DNA methylation levels of DKO-DMRs in the three germ layers contribute to this intriguing exploration of the DNA methylation turnover's role in cellular differentiation.

The DNA methylation turnover precedes early differentiation-related hypomethylation

The first lineage decision of hPSCs is their commitment to one of the three germ layers. Strikingly, the vast majority of hypomethylation events that take place during these transitions occur at DNA methylation turnover targets in hPSCs, suggesting that the DNA methylation turnover could play a role for differentiation-related hypomethylation, at least in some regions. In contrast, I observed a decreased association between the DNA methylation turnover in hPSCs and regions that become hypomethylated in more mature cells. These contradicting findings could be explained by the diverse temporal dynamics with which DNA methylation turnover is initiated. For instance, a minority of regions that become hypomethylated during motor neuron differentiation between day 16 and day 60 are DNA methylation turnover targets in hPSCs. However, the other regions tend to establish DNA methylation turnover *de novo* after exiting pluripotency, in intermediate progenitor states. On the other hand, regions which are subject to the DNA methylation turnover in hPSCs might maintain the dynamic in subsequent progenitor states, before undergoing tissue-specific hypomethylation. Importantly, I could confirm continued turnover for a subset of DKO-DMRs upon the first lineage decision after exiting pluripotency. However, whether DNA methylation turnover rates are constant until demethylation takes place is yet unknown and requires further and more time-resolved investigation in the future. It is reasonable to think that targeted alterations of the methylation landscape during differentiation could depend on other molecular features altered in a tissue-specific manner, including the binding of tissue-specific transcription factors, remodeling of chromatin accessibility and epigenetic state, as well as the action of long non-coding RNAs (Tsankov et al., 2015; Chen & Dent, 2014; Stergachis et al., 2013; Fatica & Bozzoni, 2014).

My investigation of DNA methylation turnover during pancreatic differentiation revealed the peculiarity of this lineage as compared to the other two germ layers and even hepatic endoderm.

In fact, very few hypomethylated regions along the pancreatic trajectory show DNA methylation turnover in hPSCs. Given that pancreatic endoderm emerges quickly, merely three days after initiating differentiation, there is little opportunity for DNA methylation turnover to potentially establish before tissue-specific hypomethylation. This implies that these regions establish tissue-specific hypomethylation independent of the DNA methylation turnover and that other mechanisms could be in place, including the ones mentioned above. Further, these findings suggest that target-specificity of the DNA methylation turnover might be permitted by factors involved in neuroectodermal development, but unrelated to pancreatic development.

Overall, my analysis revealed how DNA methylation turnover and differentiation-related hypomethylation are associated, with variable behaviors of different regions within tissues and temporal windows.

The DNA methylation turnover at the majority of DKO-DMRs starts long before hypomethylation occurs

Another interesting finding are the largely stable DNA methylation levels at DKO-DMRs upon three germ layer differentiation. Intriguingly, my analysis revealed a temporal discrepancy between the initiation of the DNA methylation turnover at DKO-DMRs and their tissue-specific demethylation later throughout development. One explanation might be that DKO-DMRs also become hypomethylated in alternative earlier-entered lineages which were not subject to my analysis. The idea that TETs target turnover regions prior to functional demethylation because of potential difficulties in the *de novo* recruitment of TETs in a heavily methylated environment is somewhat unlikely. Recent work has demonstrated that re-targeting DKO-DMRs by TETs in cells deficient for the endogenous demethylases re-establish an almost wildtype-like DNA methylation landscape (Charlton et al., 2020). However, not all DKO-DMRs are being successfully re-targeted by TET1 and TET2, demonstrating that the *de novo* establishment of the DNA methylation turnover at DKO-DMRs in hPSCs is constrained, at least for TET1 and TET2. Anyhow, it appears that the DNA methylation turnover rather serves as a mechanism to prime tissue-specific hypomethylation during differentiation, which itself is likely to be governed by other variables. Further hints on the likelihood that the DNA methylation turnover contributes to DNA methylation plasticity would be provided by determining whether the DNA methylation turnover stops at locations unrelated to one lineage upon entering that lineage.

In conclusion, my investigations revealed that the DNA methylation turnover does not necessarily entail immediate tissue-specific hypomethylation. Rather, my findings suggest a role for priming those genomic locations which become hypomethylated in any lineage and cell type during development.

Does the DNA methylation turnover play a role for the differentiation-related demethylation rate?

Further, I revealed that tissue-specific demethylation is slightly more pronounced at regions which are DNA methylation turnover targets in hPSCs, particularly severe for neuroectodermal lineages. Moreover, TET activity during pluripotency at turnover targets seems to set the boundary for their local DNA methylation loss upon differentiation. This suggests that the DNA methylation turnover may impact the rate of differentiation-related demethylation due to the local persistence of TET enzymes. In contrast, regions without DNA methylation turnover might undergo passive demethylation, which could result in lower rates and speed of demethylation given its dependency on replication and cell division speed. However, this is highly speculative since comparative investigations of passive and active demethylation are challenging, given that they largely differ in their genomic targets (Ginno et al., 2020). Further, dependency on active or passive demethylation as the driving force of hypomethylation can also be tissue-specific (Suzuki et al., 2017; Stoyanova et al., 2021; Vincenzetti et al., 2019). However, we cannot exclude a direct contribution of the TET enzymes to turnover-independent demethylation during three germ layer differentiation. To shed light on this, time-resolved differentiation experiments of TET-TKO hPSCs into the three germ layers could reveal whether TETs are directly required for demethylation of early somatic enhancers, including how their processivity can influence demethylation rates and speed.

In summary, my analysis reveals a slight association between DNA methylation turnover and rates of differentiation-related hypomethylation, which implies a potential role of TET pre-targeting for differentiation.

DKO-DMRs can act as somatic enhancers

The majority of DKO-DMRs in hPSCs intersect with putative somatic enhancers, are enriched for histone marks canonically associated with poised enhancers (e.g. H3K4me1), exhibit slight chromatin opening, and in hPSCs, they lack a chromatin landscape usually found at active

elements (H3K27Ac) (Charlton et al., 2020). In this study, I tested the activity of DKO-DMRs experiencing hypomethylation and increased chromatin opening during hepatic differentiation. This experiment showed how these regions exhibit enhancer activity during *in vitro* differentiation, providing functional evidence for the association between DKO-DMRs and somatic enhancers. It is unclear, though, how enhancer activation is impacted by DNA methylation turnover or whether the switch in DNA methylation levels (high in hPSCs and low somatic cells) serve a direct role in enhancer activation. To this end, I engineered multiple cell lines that will allow to investigate these questions in greater detail and in the context of human early differentiation, providing a solid toolkit towards the functional characterization of DNA methylation turnover at somatic enhancers in the future.

Importantly, my results provide functional evidence that DKO-DMRs can act as enhancers in somatic cells, opening various avenues to dissect the role of DNA methylation turnover.

Potential molecular functions of the DNA methylation turnover at somatic enhancers

Transcription factor binding

DNA methylation is known to impact the binding of certain TFs to their sequence motifs, either positively or negatively (Yin et al., 2017). Moreover, multiple studies in mice have also shown how the TETs-induced oxidized derivatives of 5mC, including 5hmC, 5caC, and 5fC, influence TF binding (Iurlaro et al., 2013; Yang et al., 2019; Golla et al., 2014; Ray et al., 2020; Sayeed et al., 2015; Verma et al., 2018; Wu et al., 2011b). In this specific context, DNA methylation turnover may therefore facilitate the binding of certain TFs with special affinity for 5hmC, 5caC, or 5fC, as well as prevent the binding of others. Moreover, in hPSCs and differentiated cells, DNA methylation turnover may orchestrate the controlled, defined, and time-sensitive sequence of TF binding events at somatic enhancers, therefore playing a crucial role in spatiotemporal gene activation during development (Parry et al., 2021).

An example of this comes from the TF TCF4 (TCF7L2), an effector of the WNT signaling pathway and an important player in various differentiation processes, which was shown to preferentially bind to unmethylated DNA and to have an affinity to carboxylated CpGs *in vitro*(Glass et al., 2005; Trompouki et al., 2011; Yang et al., 2019; Yin et al., 2017). Interestingly, TCF4 is slightly enriched at some DKO-DMRs in hPSCs (Charlton et al., 2020). A possible explanation could be TCF4's aversion against methylated cytosines, which may prevent a stronger binding, and its preference for carboxylated cytosine on the other hand, still

allowing weak binding in hPSCs. This idea would reinforce the necessity of a precise balance between DNMT3 and TET activity at DKO-DMRs in hPSCs. Moreover, hypomethylation of those regions during differentiation may allow rapid and strong binding of TCF4, due to the loss of methylated CpGs and a TET-mediated enrichment of carboxylcytosine, which may further support and drive lineage specification. DNA methylation turnover may therefore prevent excessive binding of TCF4 to DKO-DMRs in hPSCs (high methylation levels), but still preserve the necessary epigenetic priming for rapid accumulation upon differentiation. To learn more about how the DNA methylation turnover affects TCF4 binding, one might examine alterations in TCF4-binding patterns in DKO and TKO hPSCs where either CpGs are unmethylated or the DNA methylation turnover is stopped, respectively.

In addition, asynchronous DNA methylation turnover among individual cells may produce cell-to-cell variability of the 5mC pattern and its derivatives at DKO-DMRs, variably exposing TF binding sites in differential modified CpG contexts (Parry et al., 2021). Therefore, DNA methylation turnover could represent a molecular mechanism to create heterogeneity within a cell population, allowing for differential responsiveness to differentiation stimuli and therefore potentially instructing multi-lineage differentiation during development (Parry et al., 2021). In line with this hypothesis, TF binding to certain motifs within a subset of active enhancers in mESCs is specific to unmethylated DNA, which is diminished in the absence of TETs, as the number of methylated molecules increases (Kreibich et al., 2023). This suggests DNA methylation-heterogeneity on the single-molecule level, which causes differential TF binding patterns. Furthermore, DNMTs and TETs at DKO-DMRs were shown to increase the number of hemi-methylated DNA double-strands, thus promoting heterogeneity (Charlton et al., 2020).

Hence, in various systems and genomic contexts, cytosine modifications have an impact on TF binding. These causalities could similarly extend to the DKO-DMRs. Considering their enhancer functionality, the DNA methylation turnover may have a pivotal role in establishing complex binding patterns, regulating and priming enhancer activity.

Chromatin accessibility

The majority of DKO-DMRs, even though heavily methylated, lay in open chromatin; yet, my analysis revealed that unmethylated genomic loci display a greater level of chromatin opening (Charlton et al., 2020). Typically, DNA methylation and chromatin accessibility are negatively correlated (Clark et al., 2018; Thurman et al., 2012). Moreover, the depletion of all three

DNMTs in mESCs leads to chromatin opening in specific genomic loci, including a subset of active enhancers, suggesting a locus-specific causal relationship between the two epigenetic features (Domcke et al., 2015; Kreibich et al., 2023). In line with this, TET activity has been associated with chromatin accessibility genome-wide (Ginno et al., 2020). This suggests that maintaining high DNA methylation levels at DKO-DMRs, through DNMT3s counteracting TET activity, might be crucial in preserving their moderately opened chromatin from further de-condensation. As proposed for TF binding, tissue-specific demethylation of DKO-DMRs may increase their chromatin accessibility, further facilitating TF binding. Experiments like ATAC-seq in TKO and DKO hPSCs are necessary to eventually identify a potential causal relationship between DNMT3s, TETs and chromatin accessibility at turnover regions.

In summary, my results highlight a state of intermediate chromatin accessibility at DKO-DMRs, which potentially depends on a balanced DNA methylation turnover.

DNA methylation turnover and its relationship with transposable elements

My analysis underscores a general tendency of the DNA methylation turnover to target evolutionary young transposable elements, beyond the previously described bias for LINE1-5'ends (Charlton et al., 2020). Specifically, young subfamilies of ERV1 LTR7, ERVK LTR5, and LINE1 L1PA are major targets. Intriguingly, older subfamilies of these clades accumulated changes in their DNA sequences during evolution, including TF binding motifs, which could potentially be involved in their reduced DNA methylation turnover targeting (Carter et al., 2022; Khan et al., 2006; Grow et al., 2015). Furthermore, evolutionary young LTR7, LTR5, and L1PA elements are transcribed in hPSCs, while their older relatives are usually inactive (Smith et al., 2014; Grow et al., 2015; Carter et al., 2022). Thus, my finding opens new avenues to dissect the mechanism permitting target-specificity and to speculate about the turnover's potential function at transposons, which will be discussed in the following.

Learning from transposons about a potential mechanism behind the target-specificity of the DNA methylation turnover in hPSCs

In the previous section, I discussed the potential role of the DNA methylation turnover for the binding of TFs that are sensitive to cytosine modifications. Vice versa, TFs may contribute to guiding DNMT3s and TETs to specific genomic loci, thus shaping the DNA methylation turnover landscape (Feldmann et al., 2013). However, among 63 TFs previously studied, none

of them displayed consistent enrichment across all DKO-DMRs, suggesting multiple TFs to be capable of orchestrating the turnover (Charlton et al., 2020). In addition to trying to identify TF candidates from scratch, the differential recruitment of the turnover to retrotransposon subfamilies could provide insights into targeting mechanisms (Charlton et al., 2020). Despite the shared sequence similarities inherent to subfamilies within the same clade of repetitive elements, discernable and focal DNA sequence differences persist (Hermant & Torres-Padilla, 2021). While the most conserved sequence segments are unlikely to play a role in the differential targeting of the DNA methylation turnover among these subfamilies, subfamily-specific motifs could potentially be involved. The following section undertakes a literature-based exploration of potential TF candidates that exhibit association with distinct subfamilies featuring or lacking DNA methylation turnover.

The LTR7up elements, which are frequently subject to the DNA methylation turnover, contain a SOX2/3 binding motif that is missing in non-DNA methylation turnover-associated LTR7 subfamilies (Carter et al., 2022). Furthermore, the actual binding of the pluripotency factor SOX2 is enriched at LTR7up elements compared to every other LTR7 subfamily. Similarly, young LINE1 L1PA subfamilies harbor two SRY-binding sites—potential targets for SOX2 (Khan et al., 2006). Given that SOX2 acts as a pioneer factor, capable of opening condensed chromatin and potentially supporting passive DNA demethylation, SOX2 could be hypothesized to play a role in establishing the DNA methylation turnover at LTR7up and young L1PA elements (Dodonova et al., 2020; Vanzan et al., 2021; Soufi et al., 2012).

In addition, the majority of LTR7up elements exhibit enriched binding of the pluripotency factor KLF4 (Carter et al., 2022). The role of KLF4 as a pioneer factor during the reprogramming of somatic cells back to pluripotency, coupled with its reported physical interaction with TET2, suggests a potential function for KLF4 in recruiting the DNA methylation turnover machinery to LTR7up elements (Sardina et al., 2018; Stadhouders et al., 2018; Soufi et al., 2012).

Similarly, NANOG exhibits notable enrichment at LTR7up elements (Carter et al., 2022). In line with this, another pluripotency-associated TF, OCT4, displays an affinity for LTR7up elements, particularly for those with robust transcriptional activity (Carter et al., 2022). This trend aligns with observations made for LTR5-Hs, where an OCT4-binding site distinguishes them from LTR5A and LTR5B (Grow et al., 2015). Nevertheless, direct OCT4 enrichment at LTR5-Hs is not evident in primed hPSCs, implying a limited role of OCT4 in establishing the

DNA methylation turnover on these elements. Still, OCT4, just like NANOG, fulfills pioneer functions in developmental processes of *Xenopus tropicalis* and zebrafish and during early reprogramming, which implies a potential role for the chromatin opening at DNA methylation turnover targets (Pálfy et al., 2020; Veil et al., 2019; Soufi et al., 2012). Despite these insights, the specific impacts of NANOG and OCT4 on DNA methylation are largely unexplored.

The strong enrichment of pluripotency factors at retrotransposons with DNA methylation turnover is in line with previous observations in mESCs, where TET activity is substantially elevated in the vicinity of genomic sites bound by OCT4, SOX2, and NANOG, as compared to loci that are enriched for other TFs (Ginno et al., 2020). Intriguingly, murine NANOG was shown to physically interact with TET1/ TET2 (Costa et al., 2013). Moreover, TET1 binding to genomic targets is reduced upon NANOG depletion, suggesting a role for TET1 recruitment (Costa et al., 2013). Additionally, given that many pluripotency factors are capable of binding to compacted chromatin and facilitating nucleosome remodeling, this could potentially enhance chromatin accessibility also for DNMT3s/TETs. Thus, this would suggest that chromatin opening may be an important feature of turnover targets, rather than their association with any specific TF (Ginno et al., 2020). Furthermore, chromatin accessibility as a key prerequisite for the turnover would also be a possible explanation for the observation of the DNA methylation turnover in differentiated cell types, where pluripotency factor expression is typically diminished, or absent. Here, other pioneer factors expressed in somatic cell types could de-compact the chromatin, permitting access for DNMT3/TET. However indications of regulators other than pluripotency factors which are differentially enriched between retrotransposons with and without DNA methylation turnover remain scarce (Hermant & Torres-Padilla, 2021). In the future, assessing how pluripotency factors influence DNA methylation turnover and conducting a more in-depth exploration of distinct motifs distinguishing retrotransposon subfamilies with DNA methylation turnover from those lacking it could potentially shed light on the enigma surrounding the underlying mechanism.

In conclusion, my discovery that the DNA methylation turnover substantially targets recently evolved elements of ERVK LTR5 and ERV1 LTR7 and previous studies pointing out their specific enrichment for binding motifs of pluripotency factors provides a basis to explore pluripotency factors as potential drivers behind the target-specificity of the DNA methylation turnover in hPSCs in the future.

DNA methylation turnover: a byproduct of the transcriptional potential of evolutionary young transposable elements?

Intriguingly, identifying LTR5-Hs and LTR7up1/2 as prominent targets of the DNA methylation turnover, in addition to the previously found L1HS and L1PA2, aligns with their transcriptional activity in hPSCs (Smith et al., 2014; Grow et al., 2015; Carter et al., 2022; Charlton et al., 2020). This increases the likelihood of a link between transcriptional activity and the occurrence of the DNA methylation turnover. In the subsequent discussion, I will delve into the potential nature of their relationship and explore implications for the turnover's potential function arising from this connection.

I substantially observed the DNA methylation turnover at ERVK LTR5-Hs elements, which are transcribed in hPSCs (Grow et al., 2015). In contrast, my analysis revealed that their older relatives, LTR5A and LTR5B, are predominantly devoid of the DNA methylation turnover. Simultaneously, LTR5A and LTR5B were shown to be largely inactive in hPSCs (Grow et al., 2015). Given that chemically induced demethylation enhances transcription of ERVKs, TET-mediated demethylation of LTR5-Hs in the turnover context may serve their activation and entail DNMT3 targeting only to counteract this process, aiming to maintain suppression (Grow et al., 2015). Therefore, DNA methylation turnover could be a mere consequence of two opposing mechanisms used by the transposable element and the human host cell. The inactivity of LTR5A and LTR5B elements in hPSCs might suggest a lack of effective TET recruitment, which could be caused by sequence-specific differences between evolutionary young and old repeat families, as discussed previously.

This hypothesis could also apply to the LINE1 subfamilies. The DNA methylation turnover at L1HS and L1PA comes along with their transcription in hPSCs, whereas my analysis revealed that the repressed L1PA5, L1PA6, and L1PA7 subfamilies are largely independent of the DNA methylation turnover (Smith et al., 2014). Furthermore, similar to LTR5, the absence of DNA methylation at L1HS and L1PA2 results in their upregulation, suggesting that in the wildtype situation, DNMT3s might merely be active at L1HS and L1PA2 to counteract TET-mediated derepression (Castro-Diaz et al., 2014). In line with this, I observed substantial DNA methylation turnover at young LTR7 subfamilies (up1 and up2), which are transcriptionally active in hPSCs (Carter et al., 2022). In contrast, their older cognate LTR7c is transcriptionally silenced and my analysis revealed the absence of the turnover from most LTR7c elements. Even though rigorous studies examining the potential impact of DNA methylation on the repression of LTR7up1/up2 are currently lacking, the resemblance with LTR5 and LINE1

subfamilies could potentially suggest that the DNA methylation turnover at LTR7up1/2 may also result from the competing interests of the transposon and its host cell.

Given the evolutionary convergence of retrotransposons towards regulatory elements, a link between transcriptional activation and the DNA methylation turnover may have genome-wide applicability (Thompson et al., 2016; Sundaram & Wysocka, 2020). To be precise, regulatory elements which arose from transposons may still carry features to recruit TETs in order to mediate demethylation, aiming for transcriptional activation. Beyond, any genomic locus, which bears the potential to become transcribed in the absence of DNA methylation and contains features to target TETs, may potentially establish the DNA methylation turnover as a byproduct of DNMT3s trying to maintain DNA methylation levels high.

In conclusion, my observation that the DNA methylation turnover is specifically found at transcriptionally active retrotransposon subfamilies, while their silenced relatives are rather independent of it, suggests that the DNA methylation turnover could be a mere byproduct of active demethylation, serving transcriptional activation, and re-methylation aiming to maintain suppression.

CONCLUSION

The DNA methylation turnover, described as the first DNA methylation dynamic uncoupled from cellular transitioning into a different state (like during differentiation or reprogramming), raised many questions, particularly concerning the function and mechanism of its target-specificity to prospective somatic enhancers and certain groups of retrotransposable elements (Charlton et al., 2020).

My research work has revealed the closely intertwined association between DNA methylation turnover and local alterations in the DNA methylation landscape during cellular differentiation. Specifically, early hypomethylation events are strongly related to the dynamic state of the DNA methylation turnover during pluripotency. However, demethylation related to more advanced cell types often initiates the DNA methylation turnover *de novo* in differentiated progenitor cells. This implies that differentiation-related hypomethylation generally correlates with the DNA methylation turnover, which is established not only in hPSCs but also in differentiated cell types.

Furthermore, I successfully addressed the pivotal question about the attribution of putative enhancer function of the DKO-DMRs. The observation that all tested DKO-DMRs exhibit enhancer activity in differentiated cells highlights a potential functional role of the DNA methylation turnover in regulating enhancer functionality at these sites.

Interestingly, my study also highlighted DNA methylation turnover target-specificity in the context of retrotransposons, specifically for evolutionarily young elements. This paved the way for a comparative literature research, that led to the identification of pluripotency factors as potential candidates guiding TETs and/or DNMT3s to their shared genomic loci. Considering that pluripotency factors often serve as pioneer factors, it implies that TET targeting may also hinge on mere chromatin accessibility.

This work opened multiple avenues for deeper exploration of DNA methylation turnover both at the functional and mechanistic levels. To discern its functional significance, leveraging the DNMT3-depleted hiPSCs and the TET1/2/3-depleted hiPSCs could potentially unveil the influence of DNA methylation levels and the turnover on TF binding and chromatin accessibility. Moreover, knockdown experiments targeting specific pluripotency factors might reveal their impact on the DNA methylation turnover in hPSCs. Additionally, an extended comparative analysis of motifs among retrotransposon subfamilies with and without DNA methylation turnover could uncover further candidate factors that may also be expressed in differentiated cells.

In conclusion, the *de novo* establishment of the DNA methylation turnover in differentiated cells, alongside its ongoing dynamics, extends the phenomenon to a more expanded developmental context. This broader scope emphasizes the need to comprehend its functional implications. Overall, this study adds to the growing knowledge surrounding DNA methylation dynamics and their impact on cellular differentiation and developmental progression.

REFERENCES

Aapola, U., Shibuya, K., Scott, H. S., Ollila, J., Vihinen, M., Heino, M., Shintani, A., Kawasaki, K., Minoshima, S., Krohn, K., Antonarakis, S. E., Shimizu, N., Kudoh, J., & Peterson, P. (2000). Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. *Genomics*, *65*(3), 293–298. https://doi.org/10.1006/geno.2000.6168

Alvarez-Dominguez, J. R., Donaghey, J., Rasouli, N., Kenty, J. H. R., Helman, A., Charlton, J., Straubhaar, J. R., Meissner, A., & Melton, D. A. (2020). Circadian Entrainment Triggers Maturation of Human In Vitro Islets. *Cell Stem Cell*, *26*(1), 108-122.e10. https://doi.org/10.1016/j.stem.2019.11.011

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., & Zhao, K. (2007). High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell*, *129*(4), 823–837. https://doi.org/10.1016/j.cell.2007.05.009

Bartolomei, M. S., Webber, A. L., Brunkow, M. E., & Tilghman, S. M. (1993). Epigenetic mechanisms underlying the imprinting of the mouse H19 gene. *Genes and Development*, *7*(9), 1663–1673. https://doi.org/10.1101/gad.7.9.1663

Baubec, T., Colombo, D. F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A. R., Akalin, A., & Schübeler, D. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature*, *520*(7546), 243–247. https://doi.org/10.1038/nature14176

Beard, C., Li, E., & Jaenisch, R. (1995). Loss of methylation activates Xist in somatic but not in embryonic cells. *Genes and Development*, *9*(19), 2325–2334. https://doi.org/10.1101/gad.9.19.2325

Bell, A. C., & Felsenfeld, G. (2000). *Bell&Felsenfeld_2000*. *405*(May), 2–5.

Berg, J. M., Lee, C., Chen, L., Galvan, L., Cepeda, C., Chen, J. Y., Peñagarikano, O., Stein, J. L., Li, A., Oguro-Ando, A., Miller, J. A., Vashisht, A. A., Starks, M. E., Kite, E. P., Tam, E., Gdalyahu, A., Al-Sharif, N. B., Burkett, Z. D., White, S. A., … Geschwind, D. H. (2015). JAKMIP1, a Novel Regulator of Neuronal Translation, Modulates Synaptic Function and Autistic-like Behaviors in Mouse. *Neuron*, *88*(6), 1173–1191. https://doi.org/10.1016/j.neuron.2015.10.031

Bernstein, B. E., Humphrey, E. L., Erlich, R. L., Schneider, R., Bouman, P., Liu, J. S., Kouzarides, T., & Schreiber, S. L. (2002). Methylation of histone H3 Lys 4 in coding regions of active genes. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(13), 8695–8700. https://doi.org/10.1073/pnas.082249499

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S. L., & Lander, E. S. (2006). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell*, *125*(2), 315–326. https://doi.org/10.1016/j.cell.2006.02.041

Bestor, T. H. (1990). DNA methylation: evolution of a bacterial immune function into a regulator of gene expression and genome structure in higher eukaryotes. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *326*(1235), 179–187. https://doi.org/10.1098/rstb.1990.0002

Bestor, T., Laudano, A., Mattaliano, R., & Ingram, V. (1988). Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *Journal of Molecular Biology*, *203*(4), 971–983. https://doi.org/10.1016/0022-2836(88)90122-2

Bird, A. P. (1980). DNA methylation and frequency of CpG in animal DNA. *Nucleic Acids Research*, *8*(7), 1499–1504.

Bird, A. P. (1984). Gene expression: DNA methylation - How important in gene control? *Nature*, *307*(5951), 503–504. https://doi.org/10.1038/307503a0

Bird, A., Taggart, M., Frommer, M., Miller, O. J., & Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell*, *40*(1), 91–99. https://doi.org/10.1016/0092-8674(85)90312-5

Birke, M., Schreiner, S., García-Cuéllar, M., Mahr, K., Titgemeyer, F., & Slany, R. (2002). The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. *Nucleic Acids Research*, *30*(4), 958–965. https://doi.org/10.1093/nar/30.4.958

Booth, M. J., Ost, T. W. B., Beraldi, D., Bell, N. M., Branco, M. R., Reik, W., & Balasubramanian, S. (2013). Oxidative bisulfite sequencing of 5-methylcytosine and 5- hydroxymethylcytosine. *Nature Protocols*, *8*(10), 1841–1851. https://doi.org/10.1038/nprot.2013.115

Borgel, J., Guibert, S., Li, Y., Chiba, H., Schübeler, D., Sasaki, H., Forné, T., & Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. *Nature Genetics*, *42*(12), 1093–1100. https://doi.org/10.1038/ng.708

Bostick, M., Kim, J. K., Estève, P.-O., Clark, A., Pradhan, S., & Jacobsen, S. E. (2007). *UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells*. *317*, 1760–1764.

Bourc'his, D., & Bestor, T. H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature*, *431*(7004), 96–99. https://doi.org/10.1038/nature02886

Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B., & Bestor, T. H. (2001). Dnmt3L and the establishment of maternal genomic imprints. *Science*, *294*(5551), 2536–2539. https://doi.org/10.1126/science.1065848

Bröske, A. M., Vockentanz, L., Kharazi, S., Huska, M. R., Mancini, E., Scheller, M., Kuhl, C., Enns, A., Prinz, M., Jaenisch, R., Nerlov, C., Leutz, A., Andrade-Navarro, M. A., Jacobsen, S. E. W., & Rosenbauer, F. (2009). DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. *Nature Genetics*, *41*(11), 1207–1215. https://doi.org/10.1038/ng.463

Bulut-Karslioglu, A., DeLaRosa-Velázquez, I. A., Ramirez, F., Barenboim, M., Onishi-Seebacher, M., Arand, J., Galán, C., Winter, G. E., Engist, B., Gerle, B., O'Sullivan, R. J., Martens, J. H. A., Walter, J., Manke, T., Lachner, M., & Jenuwein, T. (2014). Suv39h-Dependent H3K9me3 Marks Intact Retrotransposons and Silences LINE Elements in Mouse Embryonic Stem Cells. *Molecular Cell*, *55*(2), 277–290. https://doi.org/10.1016/j.molcel.2014.05.029

Busslinger, M., Hurst, J., & Flavell, R. A. (1983). DNA methylation and the regulation of gene expression. *Cell*, *34*, 197–206. https://doi.org/10.1016/0306-9877(89)90109-6

Capuano, F., Mülleder, M., Kok, R., Blom, H. J., & Ralser, M. (2014). Cytosine DNA methylation is found in drosophila melanogaster but absent in saccharomyces cerevisiae, schizosaccharomyces pombe, and other yeast species. *Analytical Chemistry*, *86*(8), 3697–3702. https://doi.org/10.1021/ac500447w

Carlson, L. L., Page, A. W., & Bestor, T. H. (1992). Properties and localization of DNA methyltransferase in preimplantation mouse embryos: Implications for genomic imprinting. *Genes and Development*, *6*(12 B), 2536–2541. https://doi.org/10.1101/gad.6.12b.2536

Carter, T. A., Singh, M., Dumbović, G., Chobirko, J. D., Rinn, J. L., & Feschotte, C. (2022). Mosaic cis-regulatory evolution drives transcriptional partitioning of HERVH endogenous retrovirus in the human embryo. *ELife*, *11*, 1–29. https://doi.org/10.7554/ELIFE.76257

Castro-Diaz, N., Ecco, G., Coluccio, A., Kapopoulou, A., Yazdanpanah, B., Friedli, M., Duc, J., Jang, S. M., Turelli, P., & Trono, D. (2014). Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes and Development*, *28*(13), 1397–1409. https://doi.org/10.1101/gad.241661.114

Charlton, J., Jung, E., Mattei, A., Bailly, N., Lia, J., Martin, E., Giesselmann, P., Stamenova, E., Kiskinis, E., Gnirke, A., Smith, Z., & Meissner, A. (2020). TETs compete with DNMT3 activity in pluripotent cells at thousands of methylated somatic enhancers. *Nature Genetics*.

Chaumeil, J., Le Baccon, P., Wutz, A., & Heard, E. (2006). A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes and Development*, *20*(16), 2223–2237. https://doi.org/10.1101/gad.380906

Chen, T., & Dent, S. Y. R. (2014). Chromatin modifiers and remodellers: Regulators of cellular differentiation. *Nature Reviews Genetics*, *15*(2), 93–106. https://doi.org/10.1038/nrg3607

Chen, T., Ueda, Y., Dodge, J. E., Wang, Z., & Li, E. (2003). Establishment and Maintenance of Genomic Methylation Patterns in Mouse Embryonic Stem Cells by Dnmt3a and Dnmt3b. *Molecular and Cellular Biology*, *23*(16), 5594–5605. https://doi.org/10.1128/mcb.23.16.5594-5605.2003

Christman, J. K., Price, P., Pedrinan, L., & Acs, G. (1977). Correlation between Hypomethylation of DNA and Expression of Globin Genes in Friend Erythroleukemia Cells. *European Journal of Biochemistry*, *81*(1), 53–61. https://doi.org/10.1111/j.1432-1033.1977.tb11926.x

Clark, S. J., Argelaguet, R., Kapourani, C. A., Stubbs, T. M., Lee, H. J., Alda-Catalinas, C., Krueger, F., Sanguinetti, G., Kelsey, G., Marioni, J. C., Stegle, O., & Reik, W. (2018). ScNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells e. *Nature Communications*, *9*(1), 1–9. https://doi.org/10.1038/s41467-018-03149-4

Clouaire, T., Webb, S., Skene, P., Illingworth, R., Kerr, A., Andrews, R., Lee, J. H., Skalnik, D., & Bird, A. (2012). Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. *Genes and Development*, *26*(15), 1714–1728. https://doi.org/10.1101/gad.194209.112

Costa, Y., Ding, J., Theunissen, T. W., Faiola, F., Hore, T. A., Shliaha, P. V., Fidalgo, M., Saunders, A., Lawrence, M., Dietmann, S., Das, S., Levasseur, D. N., Li, Z., Xu, M., Reik, W., Silva, J. C. R., & Wang, J. (2013). NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature*, *495*(7441), 370–374. https://doi.org/10.1038/nature11925

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A., Boyer, L. A., Young, R. A., & Jaenisch, R. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(50), 21931–21936. https://doi.org/10.1073/pnas.1016071107

Dawlaty, M. M., Breiling, A., Le, T., Barrasa, M. I., Raddatz, G., Gao, Q., Powell, B. E., Cheng, A. W., Faull, K. F., Lyko, F., & Jaenisch, R. (2014). Loss of tet enzymes compromises proper differentiation of embryonic stem cells. *Developmental Cell*, *29*(1), 102–111. https://doi.org/10.1016/j.devcel.2014.03.003

Dawlaty, M. M., Breiling, A., Le, T., Raddatz, G., Barrasa, M. I., Cheng, A. W., Gao, Q., Powell, B. E., Li, Z., Xu, M., Faull, K. F., Lyko, F., & Jaenisch, R. (2013). Combined Deficiency of Tet1 and Tet2 Causes Epigenetic Abnormalities but Is Compatible with Postnatal Development. *Developmental Cell*, *24*(3), 310–323. https://doi.org/10.1016/j.devcel.2012.12.015

Dawlaty, M. M., Ganz, K., Powell, B. E., Hu, Y. C., Markoulaki, S., Cheng, A. W., Gao, Q., Kim, J., Choi, S. W., Page, D. C., & Jaenisch, R. (2011). Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell*, *9*(2), 166–175. https://doi.org/10.1016/j.stem.2011.07.010

Del Castillo Falconi, V. M., Torres-Arciga, K., Matus-Ortega, G., Díaz-Chávez, J., & Herrera, L. A. (2022). DNA Methyltransferases: From Evolution to Clinical Applications. *International Journal of Molecular Sciences*, *23*(16), 1–18. https://doi.org/10.3390/ijms23168994

Dellino, G. I., Schwartz, Y. B., Farkas, G., McCabe, D., Elgin, S. C. R., & Pirrotta, V. (2004). Polycomb silencing blocks transcription initiation. *Molecular Cell*, *13*(6), 887–893. https://doi.org/10.1016/S1097-2765(04)00128-5

Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R. Z., Ragozin, S., & Jeltsch, A. (2010). The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *Journal of Biological Chemistry*, *285*(34), 26114–26120. https://doi.org/10.1074/jbc.M109.089433

Dodonova, S. O., Zhu, F., Dienemann, C., Taipale, J., & Cramer, P. (2020). Nucleosome-bound SOX2 and SOX11 structures elucidate pioneer factor function. *Nature*, *580*(7805), 669–672. https://doi.org/10.1038/s41586-020-2195-y

Doerfler, W. (1983). DNA methylation and gene expression. *Annual Review Biochemistry*, *52*, 93–124. https://doi.org/10.1128/mmbr.55.3.451-458.1991

Domcke, S., Bardet, A. F., Adrian Ginno, P., Hartl, D., Burger, L., & Schübeler, D. (2015). Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature*, *528*(7583), 575–579. https://doi.org/10.1038/nature16462

Ehrlich, M. (2009). DNA hypomethylation in cancer cells. *Epigenomics*, *1*(2), 239–259. https://doi.org/10.2217/EPI.09.33

Ellegren, H. (2004). Microsatellites: Simple sequences with complex evolution. *Nature Reviews Genetics*, *5*(6), 435–445. https://doi.org/10.1038/nrg1348

Epsztejn-Litman, S., Feldman, N., Abu-Remaileh, M., Shufaro, Y., Gerson, A., Ueda, J., Deplus, R., Fuks, F., Shinkai, Y., Cedar, H., & Bergman, Y. (2008). De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. *Nature Structural and Molecular Biology*, *15*(11), 1176–1183. https://doi.org/10.1038/nsmb.1476

Fatica, A., & Bozzoni, I. (2014). Long non-coding RNAs: New players in cell differentiation and development. *Nature Reviews Genetics*, *15*(1), 7–21. https://doi.org/10.1038/nrg3606

Feil, R., Walter, J., Allen, N. D., & Reik, W. (1994). Developmental control of allelic methylation in the imprinted mouse Igf2 and H19 genes. *Development*, *120*(10), 2933–2943. https://doi.org/10.1242/dev.120.10.2933

Feldman, N., Gerson, A., Fang, J., Li, E., Zhang, Y., Shinkai, Y., Cedar, H., & Bergman, Y. (2006). G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nature Cell Biology*, *8*(2), 188–194. https://doi.org/10.1038/ncb1353

Feldmann, A., Ivanek, R., Murr, R., Gaidatzis, D., Burger, L., & Schübeler, D. (2013). Transcription Factor Occupancy Can Mediate Active Turnover of DNA Methylation at Regulatory Regions. *PLoS Genetics*, *9*(12). https://doi.org/10.1371/journal.pgen.1003994

Feng, S., Cokus, S. J., Zhang, X., Chen, P., Bostick, M., & Goll, M. G. (2010). *Conservation and divergence of methylation patterning in plants and animals*. https://doi.org/10.1073/pnas.1002720107

Ferguson-Smith, A. C. (2011). Genomic imprinting: The emergence of an epigenetic paradigm. *Nature Reviews Genetics*, *12*(8), 565–575. https://doi.org/10.1038/nrg3032

Ferguson-Smith, A. C., Sasaki, H., Cattanach, B. M., & Surani, M. A. (1993). *Parental-origin-specific epigenetic modification of the mouse H19 gene*. *362*(April), 751–755.

Fradin, A., Manley, J. L., & Prives, C. L. (1982). Methylation of simian virus 40 Hpa II site affects late, but not early, viral gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, *79*(17 I), 5142–5146. https://doi.org/10.1073/pnas.79.17.5142

Frauer, C., Rottach, A., Meilinger, D., Bultmann, S., Fellinger, K., Hasenöder, S., Wang, M., Qin, W., Söding, J., Spada, F., & Leonhardt, H. (2011). Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1. *PLoS ONE*, *6*(2). https://doi.org/10.1371/journal.pone.0016627

Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., Molloy, P. L., & Paul, C. L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(5), 1827–1831. https://doi.org/10.1073/pnas.89.5.1827

Fuks, F., Burgers, W. A., Godin, N., Kasai, M., & Kouzarides, T. (2001). Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. *EMBO Journal*, *20*(10), 2536–2544. https://doi.org/10.1093/emboj/20.10.2536

Gardiner-Garden, M., & Frommer, M. (1987). CpG Islands in Vertebrate Genomes. *Journal of Molecular Biology*, *196*, 261–282. https://doi.org/10.1016/S0040-4020(02)00396-4

Gaston, K., & Fried, M. (1995). CpG methylation and the binding of YY1 and ETS proteins to the Surf-1/Surf-2 bidirectional promoter. *Gene*, *157*(1–2), 257–259. https://doi.org/10.1016/0378-1119(95)00120-U

Gaudet, F., Hodgson, J. G., Eden, A., Jackson-Grusby, L., Dausman, J., Gray, J. W., Leonhardt, H., & Jaenisch, R. (2003). Induction of tumors in mice by genomic hypomethylation. *Science*, *300*(5618), 489–492. https://doi.org/10.1126/science.1083558

Gifford, C. A., Ziller, M. J., Gu, H., Trapnell, C., Donaghey, J., Tsankov, A., Shalek, A. K., Kelley, D. R., Shishkin, A. A., Issner, R., Zhang, X., Coyne, M., Fostel, J. L., Holmes, L., Meldrim, J., Guttman, M., Epstein, C., Park, H., Kohlbacher, O., … Meissner, A. (2013). Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell*, *153*(5), 1149–1163. https://doi.org/10.1016/j.cell.2013.04.037

Ginno, P. A., Gaidatzis, D., Feldmann, A., Hoerner, L., Imanci, D., Burger, L., Zilbermann, F., Peters, A. H. F. M., Edenhofer, F., Smallwood, S. A., Krebs, A. R., & Schübeler, D. (2020). A genome-scale map of DNA methylation turnover identifies site-specific dependencies of DNMT and TET activity. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-16354-x

Glass, D. A., Bialek, P., Ahn, J. D., Starbuck, M., Patel, M. S., Clevers, H., Taketo, M. M., Long, F., McMahon, A. P., Lang, R. A., & Karsenty, G. (2005). Canonical Wnt signaling in differentiated osteoblasts controls osteoclast differentiation. *Developmental Cell*, *8*(5), 751–764. https://doi.org/10.1016/j.devcel.2005.02.017

Golla, J. P., Zhao, J., Mann, I. K., Sayeed, S. K., Mandal, A., Rose, R. B., & Vinson, C. (2014). Carboxylation of cytosine (5caC) in the CG dinucleotide in the E-box motif (CGCAG|GTG) increases binding of the Tcf3|Ascl1 helix-loop-helix heterodimer 10-fold. *Biochemical and Biophysical Research Communications*, *449*(2), 248–255. https://doi.org/10.1016/j.bbrc.2014.05.018

Good, C. R., Madzo, J., Patel, B., Maegawa, S., Engel, N., Jelinek, J., & Issa, J. P. J. (2017). A novel isoform of TET1 that lacks a CXXC domain is overexpressed in cancer. *Nucleic Acids Research*, *45*(14), 8269–8281. https://doi.org/10.1093/nar/gkx435

Goodier, J. L., & Kazazian, H. H. (2008). Retrotransposons Revisited: The Restraint and Rehabilitation of Parasites. *Cell*, *135*(1), 23–35. https://doi.org/10.1016/j.cell.2008.09.022

Gopalakrishnan, S., Sullivan, B. A., Trazzi, S., Della Valle, G., & Robertson, K. D. (2009). DNMT3B interacts with constitutive centromere protein CENP-C to modulate DNA methylation and the histone code at centromeric regions. *Human Molecular Genetics*, *18*(17), 3178–3193. https://doi.org/10.1093/hmg/ddp256

Gopalakrishnan, S., Van Emburgh, B. O., Shan, J., Su, Z., Fields, C. R., Vieweg, J., Hamazaki, T., Schwartz, P. H., Terada, N., & Robertson, K. D. (2009). A novel DNMT3B splice variant expressed in tumor and pluripotent cells modulates genomic DNA methylation patterns and displays altered DNA binding. *Molecular Cancer Research*, *7*(10), 1622–1634. https://doi.org/10.1158/1541-7786.MCR-09-0018

Greenberg, M. V. C., & Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. *Nature Reviews Molecular Cell Biology*, *20*(10), 590–607. https://doi.org/10.1038/s41580-019-0159-6

Grow, E. J., Flynn, R. A., Chavez, S. L., Bayless, N. L., Wossidlo, M., Wesche, D. J., Martin, L., Ware, C. B., Blish, C. A., Chang, H. Y., Pera, R. A. R., & Wysocka, J. (2015). Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature*, *522*(7555), 221–246. https://doi.org/10.1038/nature14308

Gruenbaum, Y., Cedar, H., & Razin, A. (1982). Substrate and sequence specificity of a eukaryotic DNA methylase. *Nature*, *295*(5850), 620–622. https://doi.org/10.1038/295620a0

Gu, T. P., Guo, F., Yang, H., Wu, H. P., Xu, G. F., Liu, W., Xie, Z. G., Shi, L., He, X., Jin, S. G., Iqbal, K., Shi, Y. G., Deng, Z., Szabó, P. E., Pfeifer, G. P., Li, J., & Xu, G. L. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature*, *477*(7366), 606–612. https://doi.org/10.1038/nature10443

Gu, Z., Eils, R., Schlesner, M., & Ishaque, N. (2018). EnrichedHeatmap: An R/Bioconductor package for comprehensive visualization of genomic signal associations. *BMC Genomics*, *19*(1), 1–7. https://doi.org/10.1186/s12864-018-4625-x

Guo, H., Zhu, P., Yan, L., Li, R., Hu, B., Lian, Y., Yan, J., Ren, X., Lin, S., Li, J., Jin, X., Shi, X., Liu, P., Wang, X., Wang, W., Wei, Y., Li, X., Guo, F., Wu, X., … Qiao, J. (2014). The DNA methylation landscape of human early embryos. *Nature*, *511*(7511), 606–610. https://doi.org/10.1038/nature13544

Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., Tian, C., Wang, J., Cong, Y., & Xu, Y. (2015). Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature*, *517*(7536), 640–644. https://doi.org/10.1038/nature13899

Haggerty, C., Kretzmer, H., Riemenschneider, C., Kumar, A. S., Mattei, A. L., Bailly, N., Gottfreund, J., Giesselmann, P., Weigert, R., Brändl, B., Giehr, P., Buschow, R., Galonska, C., von Meyenn, F., Pappalardi, M. B., McCabe, M. T., Wittler, L., Giesecke-Thiel, C., Mielke, T., … Meissner, A. (2021). Dnmt1 has de novo activity targeted to transposable elements. In *Nature Structural and Molecular Biology* (Vol. 28, Issue 7). Springer US. https://doi.org/10.1038/s41594-021-00603-8

Hansen, R. S., Wijmenga, C., Luo, P., Stanek, A. M., Canfield, T. K., Weemaes, C. M. R., & Gartler, S. M. (1999). The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(25), 14412–14417. https://doi.org/10.1073/pnas.96.25.14412

Hark, A. T., Schoenherr, C. J., Katz, D. J., Ingram, R. S., Levorse, J. M., & Tilghman, S. M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature*, *405*(6785), 486–489. https://doi.org/10.1038/35013106

He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C.-X., Zhang, K., He, C., & Xu, G.-L. (2011). Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science*, *333*(6047), 1303 LP – 1307. https://doi.org/10.1126/science.1210944

Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., Ye, Z., Lee, L. K., Stuart, R. K., Ching, C. W., Ching, K. A., Antosiewicz-Bourget, J. E., Liu, H., Zhang, X., Green, R. D., Lobanenkov, V. V., Stewart, R., Thomson, J. A., Crawford, G. E., … Ren, B. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, *459*(7243), 108–112. https://doi.org/10.1038/nature07829

Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E., & Ren, B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics*, *39*(3), 311–318. https://doi.org/10.1038/ng1966

Hermant, C., & Torres-Padilla, M. E. (2021). TFs for TEs: The transcription factor repertoire of mammalian transposable elements. *Genes and Development*, *35*(1), 22–39. https://doi.org/10.1101/GAD.344473.120

Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-André, V., Sigova, A. A., Hoke, H. A., & Young, R. A. (2013). Super-enhancers in the control of cell identity and disease. *Cell*, *155*(4), 934. https://doi.org/10.1016/j.cell.2013.09.053

Holliday, R., & Grigg, G. W. (1993). DNA methylation and mutation. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, *285*(1), 61–67. https://doi.org/10.1016/0027-5107(93)90052-H

Hon, G. C., Rajagopal, N., Shen, Y., McCleary, D. F., Yue, F., Dang, M. D., & Ren, B. (2013). Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nature Genetics*, *45*(10), 1198–1206. https://doi.org/10.1038/ng.2746

Hon, G. C., Song, C. X., Du, T., Jin, F., Selvaraj, S., Lee, A. Y., Yen, C. A., Ye, Z., Mao, S. Q., Wang, B. A., Kuan, S., Edsall, L. E., Zhao, B. S., Xu, G. L., He, C., & Ren, B. (2014). 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. *Molecular Cell*, *56*(2), 286–297. https://doi.org/10.1016/j.molcel.2014.08.026

Huang, Y., Chavez, L., Chang, X., Wang, X., Pastor, W. A., Kang, J., Zepeda-Martínez, J. A., Pape, U. J., Jacobsen, S. E., Peters, B., & Rao, A. (2014). Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(4), 1361–1366. https://doi.org/10.1073/pnas.1322921111

Hutnick, L. K., Huang, X., Loo, T. C., Ma, Z., & Fan, G. (2010). Repression of retrotransposal elements in mouse embryonic stem cells is primarily mediated by a DNA methylation-independent mechanism. *Journal of Biological Chemistry*, *285*(27), 21082–21091. https://doi.org/10.1074/jbc.M110.125674

Illingworth, R., Kerr, A., DeSousa, D., Jørgensen, H., Ellis, P., Stalker, J., Jackson, D., Clee, C., Plumb, R., Rogers, J., Humphray, S., Cox, T., Langford, C., & Bird, A. (2008). A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biology*, *6*(1), 0037–0051. https://doi.org/10.1371/journal.pbio.0060022

Inoue, A., Jiang, L., Lu, F., & Zhang, Y. (2017). Genomic imprinting of Xist by maternal H3K27me3. *Genes and Development*, *31*(19), 1927–1932. https://doi.org/10.1101/gad.304113.117

International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*, 860–921. https://doi.org/10.1038/35087627

Irizarry, R. A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., Ji, H., Potash, J. B., Sabunciyan, S., & Feinberg, A. P. (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature Genetics*, *41*(2), 178–186. https://doi.org/10.1038/ng.298

Ito, S., Dalessio, A. C., Taranova, O. V., Hong, K., Sowers, L. C., & Zhang, Y. (2010). Role of tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, *466*(7310), 1129–1133. https://doi.org/10.1038/nature09303

Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C., & Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, *333*(6047), 1300–1303. https://doi.org/10.1126/science.1210597

Iurlaro, M., Ficz, G., Oxley, D., Raiber, E. A., Bachman, M., Booth, M. J., Andrews, S., Balasubramanian, S., & Reik, W. (2013). A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biology*, *14*(10), 1–11. https://doi.org/10.1186/gb-2013-14-10-r119

Jackson, M., Krassowska, A., Gilbert, N., Chevassut, T., Forrester, L., Ansell, J., & Ramsahoye, B. (2004). Severe Global DNA Hypomethylation Blocks Differentiation and Induces Histone Hyperacetylation in Embryonic Stem Cells. *Molecular and Cellular Biology*, *24*(20), 8862–8871. https://doi.org/10.1128/mcb.24.20.8862-8871.2004

Jadhav, U., Cavazza, A., Banerjee, K. K., Xie, H., O'Neill, N. K., Saenz-Vash, V., Herbert, Z., Madha, S., Orkin, S. H., Zhai, H., & Shivdasani, R. A. (2019). Extensive Recovery of Embryonic Enhancer and Gene Memory Stored in Hypomethylated Enhancer DNA. *Molecular Cell*, *74*(3), 542-554.e5. https://doi.org/10.1016/j.molcel.2019.02.024

Jähner, D., Stuhlmann, H., Stewart, C. L., Barbers, K., Lohler, J., Simon, I., & Jaenisch, R. (1982). *De novo methylation and expression of retroviral genomes during mouse embryogenesis*. *298*(August), 8–13.

Jankowska, A. M., Szpurka, H., Tiu, R. V., Makishima, H., Afable, M., Huh, J., O'Keefe, C. L., Ganetzky, R., McDevitt, M. A., & Maciejewski, J. P. (2009). Loss of heterozygosity 4q24 and TET2 mutations associated with myelodysplastic/myeloproliferative neoplasms. *Blood*, *113*(25), 6403–6410. https://doi.org/10.1182/blood-2009-02-205690

Jeong, S., Liang, G., Sharma, S., Lin, J. C., Choi, S. H., Han, H., Yoo, C. B., Egger, G., Yang, A. S., & Jones, P. A. (2009). Selective Anchoring of DNA Methyltransferases 3A and 3B to Nucleosomes Containing Methylated DNA. *Molecular and Cellular Biology*, *29*(19), 5366–5376. https://doi.org/10.1128/mcb.00484-09

Josse, J., Kaiser, A. D., & Kornberg, A. (1961). of Deoxyribonucleic Acid. *The Journal of Biological Chemistry*, *236*(3), 864–875.

Jühling, F., Kretzmer, H., Bernhart, S. H., Otto, C., Stadler, P. F., & Hoffmann, S. (2016). Metilene: Fast and sensitive calling of differentially methylated regions from bisulfite sequencing data. *Genome Research*, *26*(2), 256–262. https://doi.org/10.1101/gr.196394.115

Kaluscha, S., Domcke, S., Wirbelauer, C., Stadler, M. B., Durdu, S., Burger, L., & Schübeler, D. (2022). Evidence that direct inhibition of transcription factor binding is the prevailing mode of gene and repeat repression by DNA methylation. In *Nature Genetics* (Vol. 54, Issue 12). Springer US. https://doi.org/10.1038/s41588-022-01241-6

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, H., Li, E., & Sasaki, H. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature*, *429*(6994), 900–903. https://doi.org/10.1038/nature02633

Karimi, M. M., Goyal, P., Maksakova, I. A., Bilenky, M., Leung, D., Tang, J. X., Shinkai, Y., Mager, D. L., Jones, S., Hirst, M., & Lorincz, M. C. (2011). DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mescs. *Cell Stem Cell*, *8*(6), 676–687. https://doi.org/10.1016/j.stem.2011.04.004

Khan, H., Smit, A., Boissinot, S., & Schönfeld, F. (2006). Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Research*, *16*, 78–87. https://doi.org/10.1101/gr.4001406.1

Ko, M., An, J., Bandukwala, H. S., Chavez, L., Äijö, T., Pastor, W. A., Segal, M. F., Li, H., Koh, K. P., Lähdesmäki, H., Hogan, P. G., Aravind, L., & Rao, A. (2013). Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature*, *497*(7447), 122–126. https://doi.org/10.1038/nature12052

Ko, M., Bandukwala, H. S., An, J., Lamperti, E. D., Thompson, E. C., Hastie, R., Tsangaratou, A., Rajewsky, K., Koralov, S. B., & Rao, A. (2011). Ten-eleven-translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(35), 14566–14571. https://doi.org/10.1073/pnas.1112317108

Kobayashi, H., Sakurai, T., Imai, M., Takahashi, N., Fukuda, A., Yayoi, O., Sato, S., Nakabayashi, K., Hata, K., Sotomaru, Y., Suzuki, Y., & Kono, T. (2012). Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish Oocyte-specific heritable marks. *PLoS Genetics*, *8*(1). https://doi.org/10.1371/journal.pgen.1002440

Koh, K. P., Yabuuchi, A., Rao, S., Huang, Y., Cunniff, K., Nardone, J., Laiho, A., Tahiliani, M., Sommer, C. A., Mostoslavsky, G., Lahesmaa, R., Orkin, S. H., Rodig, S. J., Daley, G. Q., & Rao, A. (2011). Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell*, *8*(2), 200–213. https://doi.org/10.1016/j.stem.2011.01.008

Kojima, K. K. (2018). Human transposable elements in Repbase: Genomic footprints from fish to humans. *Mobile DNA*, *9*(1), 1–14. https://doi.org/10.1186/s13100-017-0107-y

Kreibich, E., Kleinendorst, R., Barzaghi, G., Kaspar, S., & Krebs, A. R. (2022). Single molecule multi-omics reveals context-dependent regulation of enhancers by DNA methylation. *BioRxiv*, 2022.05.19.492653. https://www.biorxiv.org/content/10.1101/2022.05.19.492653v1%0Ahttps://www.biorxiv.org/content/10.1101/2022.05.19.492653v1.abstract

Kreibich, E., Kleinendorst, R., Barzaghi, G., Kaspar, S., Krebs, A. R., Kreibich, E., Kleinendorst, R., Barzaghi, G., Kaspar, S., & Krebs, A. R. (2023). Resource Single-molecule footprinting identifies context- dependent regulation of enhancers by DNA methylation Single-molecule footprinting identifies context-dependent regulation of enhancers by DNA methylation. *Molecular Cell*, 1–16. https://doi.org/10.1016/j.molcel.2023.01.017

Krogan, N. J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D. P., Beattie, B. K., Emili, A., Boone, C., Buratowski, S., Greenblatt, J., & Shilatifard, A. (2003). *Methylation of Histone H3 by Set2 in Saccharomyces cerevisiae Is Linked to Transcriptional Elongation by RNA Polymerase II Downloaded from http://mcb.asm.org/ on September 18 , 2012 by IMPERIAL COLLEGE LONDON Methylation of Histone H3 by Set2 in Saccharom*. *23*(12), 4207–4218. https://doi.org/10.1128/MCB.23.12.4207

Kruczek, I., & Doerfler, W. (1983). Expression of the chloramphenicol acetyltransferase gene in mammalian cells under the control of adenovirus type 12 promoters: Effect of promoter methylation on gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, *80*(24 I), 7586–7590. https://doi.org/10.1073/pnas.80.24.7586

Kurimoto, K., Yamaji, M., Seki, Y., & Saitou, M. (2008). Specification of the germ cell lineage in mice: A process orchestrated by the PR-domain proteins, Blimp1 and Prdm14. *Cell Cycle*, *7*(22), 3514–3518. https://doi.org/10.4161/cc.7.22.6979

Landshammer, A., Bolondi, A., Kretzmer, H., Much, C., Buschow, R., Rose, A., Wu, H. J., Mackowiak, S. D., Braendl, B., Giesselmann, P., Tornisiello, R., Parsi, K. M., Huey, J., Mielke, T., Meierhofer, D., Maehr, R., Hnisz, D., Michor, F., Rinn, J. L., & Meissner, A. (2023). T-REX17 is a transiently expressed non-coding RNA essential for human endoderm formation. In *eLife* (Vol. 12). https://doi.org/10.7554/eLife.83077

Langemeijer, S. M. C., Kuiper, R. P., Berends, M., Knops, R., Aslanyan, M. G., Massop, M., Stevens-Linders, E., Van Hoogen, P., Van Kessel, A. G., Raymakers, R. A. P., Kamping, E. J., Verhoef, G. E., Verburgh, E., Hagemeijer, A., Vandenberghe, P., De Witte, T., Van Der Reijden, B. A., & Jansen, J. H. (2009). Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nature Genetics*, *41*(7), 838–842. https://doi.org/10.1038/ng.391

Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C. T., Low, H. M., Sung, K. W. K., Rigoutsos, I., Loring, J., & Wei, C. L. (2010). Dynamic changes in the human methylome during differentiation. *Genome Research*, *20*(3), 320–331. https://doi.org/10.1101/gr.101907.109

Lee, C. K., Shibata, Y., Rao, B., Strahl, B. D., & Lieb, J. D. (2004). Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nature Genetics*, *36*(8), 900–905. https://doi.org/10.1038/ng1400

Lee, J. H., Voo, K. S., & Skalnik, D. G. (2001). Identification and Characterization of the DNA Binding Domain of CpG-binding Protein. *Journal of Biological Chemistry*, *276*(48), 44669–44676. https://doi.org/10.1074/jbc.M107179200

Lehnertz, B., Ueda, Yoshihide, Derijck, A. A. H. A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, Taiping, Li, E., Jenuwein, T., & Peters, A. H. F. M. (2003). Suv39h-Mediated Histone H3 Lysine 9 Methylation Directs DNA Methylation to Major Satellite Repeats at Pericentric Heterochromatin. *Current Biology*, *13*, 1192–1200. https://doi.org/10.1016/S

Leung, D., Du, T., Wagner, U., Xie, W., Lee, A. Y., Goyal, P., Li, Y., Szulwach, K. E., Jin, P., Lorincz, M. C., & Ren, B. (2014). Regulation of DNA methylation turnover at LTR retrotransposons and imprinted loci by the histone methyltransferase Setdb1. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(18), 6690–6695. https://doi.org/10.1073/pnas.1322273111

Li, E., Beard, C., & Jaenisch, R. (1994). Role for DNA methylation in genomic imprinting. *Trends in Genetics*, *10*(3), 78. https://doi.org/10.1016/0168-9525(94)90228-3

Li, En, Bestor, T. H., & Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, *69*(6), 915–926. https://doi.org/10.1016/0092-8674(92)90611-F

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, X., Ito, M., Zhou, F., Youngson, N., Zuo, X., Leder, P., & Ferguson-Smith, A. C. (2008). A Maternal-Zygotic Effect Gene, Zfp57, Maintains Both Maternal and Paternal Imprints. *Developmental Cell*, *15*(4), 547–557. https://doi.org/10.1016/j.devcel.2008.08.014

Li, Y., Zhang, Z., Chen, J., Liu, W., Lai, W., Liu, B., Li, X., Liu, L., Xu, S., Dong, Q., Wang, M., Duan, X., Tan, J., Zheng, Y., Zhang, P., Fan, G., Wong, J., Xu, G. L., Wang, Z., … Zhu, B. (2018). Stella safeguards the oocyte methylome by preventing de novo methylation mediated by DNMT1. *Nature*, *564*(7734), 136–140. https://doi.org/10.1038/s41586-018-0751-5

Liao, J., Karnik, R., Gu, H., Ziller, M. J., Clement, K., Tsankov, A. M., Akopian, V., Gifford, C. A., Donaghey, J., Galonska, C., Pop, R., Reyon, D., Tsai, S. Q., Mallard, W., Joung, J. K., Rinn, J. L., Gnirke, A., & Meissner, A. (2015). Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. *Nature Genetics*, *47*(5), 469–478. https://doi.org/10.1038/ng.3258

Lienert, F., Wirbelauer, C., Som, I., Dean, A., Mohn, F., & Schübeler, D. (2011). Identification of genetic elements that autonomously determine DNA methylation states. *Nature Genetics*, *43*(11), 1091–1097. https://doi.org/10.1038/ng.946

Ling, Y., Sankpal, U. T., Robertson, A. K., McNally, J. G., Karpova, T., & Robertson, K. D. (2004). Modification of de novo DNA methyltransferase 3a (Dnmt3a) by SUMO-1 modulates its interaction with histone deacetylases (HDACs) and its capacity to repress transcription. *Nucleic Acids Research*, *32*(2), 598–610. https://doi.org/10.1093/nar/gkh195

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q. M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B., & Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, *462*(7271), 315–322. https://doi.org/10.1038/nature08514

Loaeza-Loaeza, J., Beltran, A. S., & Hernández-Sotelo, D. (2020). Dnmts and impact of cpg content, transcription factors, consensus motifs, lncrnas, and histone marks on dna methylation. *Genes*, *11*(11), 1–19. https://doi.org/10.3390/genes11111336

Loda, A., Collombet, S., & Heard, E. (2022). Gene regulation in time and space during X-chromosome inactivation. *Nature Reviews Molecular Cell Biology*, *23*(4), 231–249. https://doi.org/10.1038/s41580-021-00438-7

Loiko, A. G., Sergeev, A. V., Genatullina, A. I., Monakhova, M. V., Kubareva, E. A., Dolinnaya, N. G., & Gromova, E. S. (2022). Impact of G-Quadruplex Structures on Methylation of Model Substrates by DNA Methyltransferase Dnmt3a. *International Journal of Molecular Sciences*, *23*(18). https://doi.org/10.3390/ijms231810226

Lowdon, R. F., Zhang, B., Bilenky, M., Mauro, T., Li, D., Gascard, P., Sigaroudinia, M., Farnham, P. J., Bastian, B. C., Tlsty, T. D., Marra, M. A., Hirst, M., Costello, J. F., Wang, T., & Cheng, J. B. (2014). Regulatory network decoded from epigenomes of surface ectoderm-derived cell types. *Nature Communications*, *5*. https://doi.org/10.1038/ncomms6442

Lyko, F., Ramsahoye, B. H., & Jaenisch, R. (2000). DNA methylation in Drosophila melanogaster. *Nature*, *408*, 538–540. https://doi.org/10.1111/j.1600-0498.1982.tb00663.x

Lynch, M. D., Smith, A. J. H., De Gobbi, M., Flenley, M., Hughes, J. R., Vernimmen, D., Ayyub, H., Sharpe, J. A., Sloane-Stanley, J. A., Sutherland, L., Meek, S., Burdon, T., Gibbons, R. J., Garrick, D., & Higgs, D. R. (2012). An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. *EMBO Journal*, *31*(2), 317–329. https://doi.org/10.1038/emboj.2011.399

Ma, Y., Yue, Y., Pan, M., Sun, J., Chu, J., Lin, X., Xu, W., Feng, L., Chen, Y., Chen, D., Shin, V. Y., Wang, X., & Jin, H. (2015). Histone deacetylase 3 inhibits new tumor suppressor gene DTWD1 in gastric cancer. *American Journal of Cancer Research*, *5*(2), 663–673. http://www.ncbi.nlm.nih.gov/pubmed/25973305%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4396045

Maiti, A., & Drohat, A. C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: Potential implications for active demethylation of CpG sites. *Journal of Biological Chemistry*, *286*(41), 35334–35338. https://doi.org/10.1074/jbc.C111.284620

Mak, W., Nesterova, T. B., De Napoles, M., Appanah, R., Yamanaka, S., Otte, A. P., & Brockdorff, N. (2004). Reactivation of the Paternal X Chromosome in Early Mouse Embryos. *Science*, *303*(5658), 666–669. https://doi.org/10.1126/science.1092674

Maksakova, I. A., Thompson, P. J., Goyal, P., Jones, S. J. M., Singh, P. B., Karimi, M. M., & Lorincz, M. C. (2013). Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells. *Epigenetics and Chromatin*, *6*(1), 1–16. https://doi.org/10.1186/1756-8935-6-15

Mancini-DiNardo, D., Steele, S. J. S., Levorse, J. M., Ingram, R. S., & Tilghman, S. M. (2006). Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. *Genes and Development*, *20*(10), 1268–1282. https://doi.org/10.1101/gad.1416906

Mann, I. K., Chatterjee, R., Zhao, J., He, X., Weirauch, M. T., Hughes, T. R., & Vinson, C. (2013). CG methylated microarrays identify a novel methylated sequence bound by the CEBPB|ATF4 heterodimer that is active in vivo. *Genome Research*, *23*(6), 988–997. https://doi.org/10.1101/gr.146654.112

Manzo, M., Wirz, J., Ambrosi, C., Villaseñor, R., Roschitzki, B., & Baubec, T. (2017). Isoform-specific localization of DNMT3A regulates DNA methylation fidelity at bivalent CpG islands. *The EMBO Journal*, *36*(23), 3421–3434. https://doi.org/10.15252/embj.201797038

Mao, S. Q., Ghanbarian, A. T., Spiegel, J., Martínez Cuesta, S., Beraldi, D., Di Antonio, M., Marsico, G., Hänsel-Hertsch, R., Tannahill, D., & Balasubramanian, S. (2018). DNA G-quadruplex structures mold the DNA methylome. *Nature Structural and Molecular Biology*, *25*(10), 951–957. https://doi.org/10.1038/s41594-018-0131-8

Martens, J. H. A., O'Sullivan, R. J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., & Jenuwein, T. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO Journal*, *24*(4), 800–812. https://doi.org/10.1038/sj.emboj.7600545

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12.

Matsui, T., Leung, D., Miyashita, H., Maksakova, I. A., Miyachi, H., Kimura, H., Tachibana, M., Lorincz, M. C., & Shinkai, Y. (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature*, *464*(7290), 927–931. https://doi.org/10.1038/nature08858

Maunakea, A. K., Chepelev, I., Cui, K., & Zhao, K. (2013). Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Research*, *23*(11), 1256–1269. https://doi.org/10.1038/cr.2013.110

Maunakea, A. K., Nagarajan, R. P., Bilenky, M., Ballinger, T. J., Dsouza, C., Fouse, S. D., Johnson, B. E., Hong, C., Nielsen, C., Zhao, Y., Turecki, G., Delaney, A., Varhol, R., Thiessen, N., Shchors, K., Heine, V. M., Rowitch, D. H., Xing, X., Fiore, C., … Costello, J. F. (2010). Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature*, *466*(7303), 253–257. https://doi.org/10.1038/nature09165

McGhee, J. D., & Ginder, G. D. (1979). Specific DNA methylation sites in the vicinity of the chicken β-globin genes [26]. *Nature*, *280*(5721), 419–420. https://doi.org/10.1038/280419a0

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Cold Spring Harbor Laboratory Press*, *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110.20

Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B. E., Nusbaum, C., Jaffe, D. B., Gnirke, A., Jaenisch, R., & Lander, E. S. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, *454*(7205), 766–770. https://doi.org/10.1038/nature07107

Mills, R. E., Bennett, E. A., Iskow, R. C., & Devine, S. E. (2007). Which transposable elements are active in the human genome? *Trends in Genetics*, *23*(4), 183–191. https://doi.org/10.1016/j.tig.2007.02.006

Mouse Genome Sequencing Consortium. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature*, *420*(6915), 520–562. http://www.nature.com/nature/journal/v420/n6915/full/nature01262.html

Nady, N., Lemak, A., Walker, J. R., Avvakumov, G. V., Kareta, M. S., Achour, M., Xue, S., Duan, S., Allali-Hassani, A., Zuo, X., Wang, Y. X., Bronner, C., Chédin, F., Arrowsmith, C. H., & Dhe-Paganon, S. (2011). Recognition of multivalent histone states associated with heterochromatin by UHRF1 protein. *Journal of Biological Chemistry*, *286*(27), 24300–24311. https://doi.org/10.1074/jbc.M111.234104

Nakagawa, T., Lv, L., Nakagawa, M., Yu, Y., Yu, C., D'Alessio, A. C., Nakayama, K., Fan, H. Y., Chen, X., & Xiong, Y. (2015). CRL4VprBP E3 ligase promotes monoubiquitylation and chromatin binding of TET dioxygenases. *Molecular Cell*, *57*(2), 247–260. https://doi.org/10.1016/j.molcel.2014.12.002

Nishiyama, A., Yamaguchi, L., Sharif, J., Johmura, Y., Kawamura, T., Nakanishi, K., Shimamura, S., Arita, K., Kodama, T., Ishikawa, F., Koseki, H., & Nakanishi, M. (2013). Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. *Nature*, *502*(7470), 249–253. https://doi.org/10.1038/nature12488

Oguchi, M. E., Etoh, K., & Fukuda, M. (2018). Rab20, a novel Rab small GTPase that negatively regulates neurite outgrowth of PC12 cells. *Neuroscience Letters*, *662*(October 2017), 324–330. https://doi.org/10.1016/j.neulet.2017.10.056

Okamoto, I., Arnaud, D., Le Baccon, P., Otte, A. P., Disteche, C. M., Avner, P., & Heard, E. (2005). Evidence for de novo imprinted X-chromosome inactivation independent of meiotic inactivation in mice. *Nature*, *438*(7066), 369–373. https://doi.org/10.1038/nature04155

Okamoto, I., Otte, A. P., Allis, C. D., Reinberg, D., & Heard, E. (2004). Epigenetic Dynamics of Imprinted. *Science*, *303*(January), 644–649.

Okano, M., Bell, D. W., Haber, D. A., & Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, *99*(3), 247–257. https://doi.org/10.1016/S0092-8674(00)81656-6

Okano, M., Xie, S., & Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases Non-invasive sexing of preimplantation stage mammalian embryos. *Nature America Inc.*, *19*(july), 219–220.

Olek, A., & Walter, J. (1997). *The pre-implantation ontogeny of the H19 methylation imprint*. *17*(november), 275–276.

Ooi, S. K. T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S. P., Allis, C. D., Cheng, X., & Bestor, T. H. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature*, *448*(7154), 714–717. https://doi.org/10.1038/nature05987

Orlando, D. A., Guenther, M. G., Frampton, G. M., & Young, R. A. (2012). CpG island structure and trithorax/polycomb chromatin domains in human cells. *Genomics*, *100*(5), 320–326. https://doi.org/10.1016/j.ygeno.2012.07.006

Orlanski, S., Labi, V., Reizel, Y., Spiro, A., Lichtenstein, M., Levin-Klein, R., Koralov, S. B., Skversky, Y., Rajewsky, K., Cedar, H., & Bergman, Y. (2016). Tissue-specific DNA demethylation is required for proper B-cell differentiation and function. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(18), 5018–5023. https://doi.org/10.1073/pnas.1604365113

Ostler, K. R., Davis, E. M., Payne, S. L., Gosalia, B. B., Expósito-Céspedes, J., Beau, M. M. L., & Godley, L. A. (2007). Cancer cells express aberrant DNMT3B transcripts encoding truncated proteins. *Oncogene*, *26*(38), 5553–5563. https://doi.org/10.1038/sj.onc.1210351

Pagès, H., Aboyoun, P., Gentleman, R., & Debroy, S. (2022). Biostrings: Efficient manipulation of biological strings. *R Package Version 2.66.0*. https://bioconductor.org/packages/Biostrings

Pálfy, M., Schulze, G., Valen, E., & Vastenhouw, N. L. (2020). Chromatin accessibility established by Pou5f3, Sox19b and Nanog primes genes for activity during zebrafish genome activation. *PLoS Genetics*, *16*(1), 1–25. https://doi.org/10.1371/journal.pgen.1008546

Panning, B., & Jaenisch, R. (1996). DNA hypomethylation can activate Xist expression and silence X-linked genes. *Genes and Development*, *10*(16), 1991–2002. https://doi.org/10.1101/gad.10.16.1991

Parry, A., Rulands, S., & Reik, W. (2021). Active turnover of DNA methylation during cell fate decisions. *Nature Reviews Genetics*, *22*(1), 59–66. https://doi.org/10.1038/s41576-020-00287-8

Pastor, W. A., Pape, U. J., Huang, Y., Henderson, H. R., Lister, R., Ko, M., McLoughlin, E. M., Brudno, Y., Mahapatra, S., Kapranov, P., Tahiliani, M., Daley, G. Q., Liu, X. S., Ecker, J. R., Milos, P. M., Agarwal, S., & Rao, A. (2011). Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature*, *473*(7347), 394–397. https://doi.org/10.1038/nature10102

Perera, A., Eisen, D., Wagner, M., Laube, S. K., Künzel, A. F., Koch, S., Steinbacher, J., Schulze, E., Splith, V., Mittermeier, N., Müller, M., Biel, M., Carell, T., & Michalakis, S. (2015). TET3 is recruited by REST for context-specific hydroxymethylation and induction of gene expression. *Cell Reports*, *11*(2), 283–294. https://doi.org/10.1016/j.celrep.2015.03.020

Peters, A. H. F. M., Kubicek, S., Mechtler, K., O'Sullivan, R. J., Derijck, A. A. H. A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y., Martens, J. H. A., & Jenuwein, T. (2003). Partitioning and Plasticity of Repressive Histone Methylation States in Mammalian Chromatin. *Molecular Cell*, *12*(6), 1577–1589. https://doi.org/10.1016/S1097-2765(03)00477-5

Qin, W., Wolf, P., Liu, N., Link, S., Smets, M., Mastra, F. La, Forné, I., Pichler, G., Hörl, D., Fellinger, K., Spada, F., Bonapace, I. M., Imhof, A., Harz, H., & Leonhardt, H. (2015). DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination. *Cell Research*, *25*(8), 911–929. https://doi.org/10.1038/cr.2015.72

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., & Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, *470*(7333), 279–285. https://doi.org/10.1038/nature09692

Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S., & Aiden, E. L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, *159*(7), 1665–1680. https://doi.org/10.1016/j.cell.2014.11.021

Ravichandran, M., Jurkowska, R. Z., & Jurkowski, T. P. (2018). Target specificity of mammalian DNA methylation and demethylation machinery. *Organic and Biomolecular Chemistry*, *16*(9), 1419–1435. https://doi.org/10.1039/c7ob02574b

Ravichandran, Mirunalini, Rafalski, D., Davies, C. I., Ortega-Recalde, O., Nan, X., Glanfield, C. R., Kotter, A., Misztal, K., Wang, A. H., Wojciechowski, M., Rażew, M., Mayyas, I. M., Kardailsky, O., Schwartz, U., Zembrzycki, K., Morison, I. M., Helm, M., Weichenhan, D., Jurkowska, R. Z., … Jurkowski, T. P. (2022). Pronounced sequence specificity of the TET enzyme catalytic domain guides its cellular function. *Science Advances*, *8*(36), 1–16. https://doi.org/10.1126/sciadv.abm2427

Ray, S., Tillo, D., Ufot, A., Assad, N., Durell, S., & Vinson, C. (2020). BZIP dimers CREB1, AtF2, Zta, Atf3|cjun, and CFOs|cjun prefer to bind to some double-stranded DNA sequences containing 5-formylcytosine and 5-carboxylcytosine. *Biochemistry*, *59*(38), 3529–3540. https://doi.org/10.1021/acs.biochem.0c00475

Rechsteiner, M., & Rogers, S. W. (1996). PEST sequences and regulation by proteolysis. *Trends in Biochemical Sciences*, *21*(7), 267–271. https://doi.org/10.1016/S0968-0004(96)10031-1

Ren, W., Gao, L., & Song, J. (2018). Structural basis of DNMT1 and DNMT3A-mediated DNA methylation. *Genes*, *9*(12). https://doi.org/10.3390/genes9120620

Roadmap Epigenomics Consortium, ., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M. J., Amin, V., Whitaker, J. W., Schultz, M. D., Ward, L. D., Sarkar, A., Quon, G., Sandstrom, R. S., Eaton, M. L., … Moore, R. (2015). Integrative analysis of 111 reference human epigenomes. *Nature*, *518*, 317–330. https://doi.org/https://doi.org/10.1038/nature14248

Rothbart, S. B., Krajewski, K., Nady, N., Tempel, W., Xue, S., Badeaux, A. I., Barsyte-Lovejoy, D., Martinez, J. Y., Bedford, M. T., Fuchs, S. M., Arrowsmith, C. H., & Strahl, B. D. (2012). Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. *Nature Structural and Molecular Biology*, *19*(11), 1155–1160. https://doi.org/10.1038/nsmb.2391

Rowe, H. M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P. V., Layard-Liesching, H., Verp, S., Marquis, J., Spitz, F., Constam, D. B., & Trono, D. (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature*, *463*(7278), 237–240. https://doi.org/10.1038/nature08674

Rush, M., Appanah, R., Lee, S., Lam, L. L., Goyal, P., & Lorincz, M. C. (2009). Targeting of EZH2 to a defined genomic site is sufficient for recruitment of Dnmt3a but not de novo DNA methylation. *Epigenetics*, *4*(6), 404–414. https://doi.org/10.4161/epi.4.6.9392

Sardina, J. L., Collombet, S., Tian, T. V., Gómez, A., Di Stefano, B., Berenguer, C., Brumbaugh, J., Stadhouders, R., Segura-Morales, C., Gut, M., Gut, I. G., Heath, S., Aranda, S., Di Croce, L., Hochedlinger, K., Thieffry, D., & Graf, T. (2018). Transcription Factors Drive Tet2-Mediated Enhancer Demethylation to Reprogram Cell Fate. *Cell Stem Cell*, *23*(5), 727-741.e9. https://doi.org/10.1016/j.stem.2018.08.016

Sayeed, S. K., Zhao, J., Sathyanarayana, B. K., Golla, J. P., & Vinson, C. (2015). C/EBPβ (CEBPB) protein binding to the C/EBP|CRE DNA 8-mer TTGC|GTCA is inhibited by 5hmC and enhanced by 5mC, 5fC, and 5caC in the CG dinucleotide. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, *1849*(6), 583–589. https://doi.org/10.1016/j.bbagrm.2015.03.002

Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D., & Jenuwein, T. (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes and Development*, *18*(11), 1251–1262. https://doi.org/10.1101/gad.300704

Schultz, M. D., He, Y., Whitaker, J. W., Hariharan, M., Mukamel, E. A., Leung, D., Rajagopal, N., Nery, J. R., Urich, M. A., Chen, H., Lin, S., Lin, Y., Jung, I., Schmitt, A. D., Selvaraj, S., Ren, B., Sejnowski, T. J., Wang, W., & Ecker, J. R. (2015). Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature*, *523*(7559), 212–216. https://doi.org/10.1038/nature14465

Schwartz, Y. B., & Pirrotta, V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. *Nature Reviews Genetics*, *8*(1), 9–22. https://doi.org/10.1038/nrg1981

Seiler, C. L., Fernandez, J., Koerperich, Z., Andersen, M. P., Kotandeniya, D., Nguyen, M. E., Sham, Y. Y., & Tretyakova, N. Y. (2018). Maintenance DNA Methyltransferase Activity in the Presence of Oxidized Forms of 5-Methylcytosine: Structural Basis for Ten Eleven Translocation-Mediated DNA Demethylation. *Biochemistry*, *57*(42), 6061–6069. https://doi.org/10.1021/acs.biochem.8b00683

Shao, Z., Raible, F., Mollaaghababa, R., Guyon, J. R., Wu, C. T., Bender, W., & Kingston, R. E. (1999). Stabilization of chromatin structure by PRC1, a polycomb complex. *Cell*, *98*(1), 37–46. https://doi.org/10.1016/S0092-8674(00)80604-2

Sharif, J., Muto, M., Takebayashi, S. I., Suetake, I., Iwamatsu, A., Endo, T. A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., Tajima, S., Mitsuya, K., Okano, M., & Koseki, H. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature*, *450*(7171), 908–912. https://doi.org/10.1038/nature06397

Shen, L., Inoue, A., He, J., Liu, Y., Lu, F., & Zhang, Y. (2014). Tet3 and DNA replication mediate demethylation of both the maternal and paternal genomes in mouse zygotes. *Cell Stem Cell*, *15*(4), 459–471. https://doi.org/10.1016/j.stem.2014.09.002

Simpson, V. J., Johnson, T. E., & Hammen, R. F. (1986). Caenorhabditis elegans DNA does not contain 5-methylcytosine at any time during development or aging. *Nucleic Acids Research*, *14*(16), 6711–6719.

Sleutels, F., Zwart, R., & Barlow, D. P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature*, *415*(6873), 810–813. https://doi.org/10.1038/415810a

Smith, Z. D., Chan, M. M., Humm, K. C., Karnik, R., Mekhoubad, S., Regev, A., Eggan, K., & Meissner, A. (2014). DNA methylation dynamics of the human preimplantation embryo. *Nature*, *511*(7511), 611–615. https://doi.org/10.1038/nature13581

Smith, Z. D., & Meissner, A. (2013). DNA methylation: Roles in mammalian development. *Nature Reviews Genetics*, *14*(3), 204–220. https://doi.org/10.1038/nrg3354

Smith, Z. D., Shi, J., Gu, H., Donaghey, J., Clement, K., Cacchiarelli, D., Gnirke, A., Michor, F., & Meissner, A. (2017). Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer. *Nature*, *549*(7673), 543–547. https://doi.org/10.1038/nature23891

Song, J., Rechkoblit, O., Bestor, T. H., & Patel, D. J. (2011). Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science*, *331*(6020), 1036–1040. https://doi.org/10.1126/science.1195380

Soufi, A., Donahue, G., & Zaret, K. S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell*, *151*(5), 994–1004. https://doi.org/10.1016/j.cell.2012.09.045

Stadhouders, R., Vidal, E., Serra, F., Di Stefano, B., Le Dily, F., Quilez, J., Gomez, A., Collombet, S., Berenguer, C., Cuartero, Y., Hecht, J., Filion, G. J., Beato, M., Marti-Renom, M. A., & Graf, T. (2018). Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. *Nature Genetics*, *50*(2), 238–249. https://doi.org/10.1038/s41588-017-0030-7

Stadler, M. B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., Wirbelauer, C., Oakeley, E. J., Gaidatzis, D., Tiwari, V. K., & Schübeler, D. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*, *480*(7378), 490–495. https://doi.org/10.1038/nature10716

Stelzer, Y., Shivalila, C. S., Soldner, F., Markoulaki, S., & Jaenisch, R. (2015). Tracing Dynamic Changes of DNA Methylation at Single-Cell Resolution. *Cell*, *163*(1), 218–229. https://doi.org/10.1016/j.cell.2015.08.046

Stergachis, A. B., Neph, S., Reynolds, A., Humbert, R., Miller, B., Paige, S. L., Vernot, B., Cheng, J. B., Thurman, R. E., Sandstrom, R., Haugen, E., Heimfeld, S., Murry, C. E., Akey, J. M., & Stamatoyannopoulos, J. A. (2013). XDevelopmental fate and cellular maturity encoded in human regulatory DNA landscapes. *Cell*, *154*(4), 888–903. https://doi.org/10.1016/j.cell.2013.07.020

Stöger, R., Kubička, P., Liu, C. G., Kafri, T., Razin, A., Cedar, H., & Barlow, D. P. (1993). Maternal-specific methylation of the imprinted mouse Igf2r locus identifies the expressed locus as carrying the imprinting signal. *Cell*, *73*(1), 61–71. https://doi.org/10.1016/0092-8674(93)90160-R

Stoyanova, E., Riad, M., Rao, A., & Heintz, N. (2021). 5-Hydroxymethylcytosine-mediated active demethylation is required for mammalian neuronal differentiation and function. *ELife*, *10*, 1–23. https://doi.org/10.7554/elife.66973

Stroud, H., Feng, S., Morey Kinney, S., Pradhan, S., & Jacobsen, S. E. (2011). 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biology*, *12*(6). https://doi.org/10.1186/gb-2011-12-6-r54

Subramanian, R. P., Wildschutte, J. H., Russo, C., & Coffin, J. M. (2011). Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. *Retrovirology*, *8*, 90. https://doi.org/10.1186/1742-4690-8-90

Suetake, I., Mishima, Y., Kimura, H., Lee, Y. H., Goto, Y., Takeshima, H., Ikegami, T., & Tajima, S. (2011). Characterization of DNA-binding activity in the N-terminal domain of the DNA methyltransferase Dnmt3a. *Biochemical Journal*, *437*(1), 141–148. https://doi.org/10.1042/BJ20110241

Sullivan, B. A., & Karpen, G. H. (2004). Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. *Nature Structural and Molecular Biology*, *11*(11), 1076–1083. https://doi.org/10.1038/nsmb845

Sun, D., Xi, Y., Rodriguez, B., Park, H. J., Tong, P., Meong, M., Goodell, M. A., & Li, W. (2014). MOABS: model based analysis of bisulfite sequencing data. *Genome Biology*, *15*(R38). https://doi.org/10.1080/03610919308813135

Sun, X. J., Wei, J., Wu, X. Y., Hu, M., Wang, L., Wang, H. H., Zhang, Q. H., Chen, S. J., Huang, Q. H., & Chen, Z. (2005). Identification and characterization of a novel human histone H3 lysine 36-specific methyltransferase. *Journal of Biological Chemistry*, *280*(42), 35261–35271. https://doi.org/10.1074/jbc.M504012200

Sundaram, V., & Wysocka, J. (2020). Transposable elements as a potent source of diverse cis-regulatory sequences in mammalian genomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1795). https://doi.org/10.1098/rstb.2019.0347

Sutter, D., & Doerfler, W. (1980). Methylation of integrated adenovirus type 12 DNA sequences in transformed cells is inversely correlated with viral gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, *77*(1), 253–256. https://doi.org/10.1073/pnas.77.1.253

Suzuki, T., Shimizu, Y., Furuhata, E., Maeda, S., Kishima, M., Nishimura, H., Enomoto, S., Hayashizaki, Y., & Suzuki, H. (2017). RUNX1 regulates site specificity of DNA demethylation by recruitment of DNA demethylation machineries in hematopoietic cells. *Blood Advances*, *1*(20), 1699–1711. https://doi.org/10.1182/bloodadvances.2017005710

Swartz, M. N., Trautner, T. A., & Kornberg, A. (1962). of Deoxyribonucleic Acid * XI. FURTHER STUDIES ON NEAREST NEIGHBOR BASE SEQUENCES IN DEOXYRIBONUCLEIC ACIDS. *The Journal of Biological Chemistry*, *237*(6), 1961–1967. https://doi.org/10.1016/S0021-9258(19)73967-2

Swergold, G. D. (1990). Identification, Characterization, and Cell Specificity of a Human LINE-1 Promoter. *Molecular and Cellular Biology*, *10*(12), 6718–6729. https://doi.org/10.1128/mcb.10.12.6718-6729.1990

Szulwach, K. E., Li, X., Li, Y., Song, C. X., Han, J. W., Kim, S. S., Namburi, S., Hermetz, K., Kim, J. J., Rudd, M. K., Yoon, Y. S., Ren, B., He, C., & Jin, P. (2011). Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. *PLoS Genetics*, *7*(6). https://doi.org/10.1371/journal.pgen.1002154

Szulwach, K. E., Li, X., Li, Y., Song, C. X., Wu, H., Dai, Q., Irier, H., Upadhyay, A. K., Gearing, M., Levey, A. I., Vasanthakumar, A., Godley, L. A., Chang, Q., Cheng, X., He, C., & Jin, P. (2011). 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nature Neuroscience*, *14*(12), 1607–1616. https://doi.org/10.1038/nn.2959

Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., & Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, *324*(5929), 930–935. https://doi.org/10.1126/science.1170116

Takeshita, K., Suetake, I., Yamashita, E., Suga, M., Narita, H., Nakagawa, A., & Tajima, S. (2011). Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). *Proceedings of the National Academy of Sciences of the United States of America*, *108*(22), 9055–9059. https://doi.org/10.1073/pnas.1019629108

Tandon, R., Brändl, B., Baryshnikova, N., Landshammer, A., Steenpaß, L., Keminer, O., Pless, O., & Müller, F. J. (2018). Generation of two human isogenic iPSC lines from fetal dermal fibroblasts. *Stem Cell Research*, *33*(August), 120–124. https://doi.org/10.1016/j.scr.2018.10.004

Tang, W. W. C., Dietmann, S., Irie, N., Leitch, H. G., Floros, V. I., Bradshaw, C. R., Hackett, J. A., Chinnery, P. F., & Surani, M. A. (2015). A unique gene regulatory network resets the human germline epigenome for development. *Cell*, *161*(6), 1453–1467. https://doi.org/10.1016/j.cell.2015.04.053

Tang, Z., Luo, O. J., Li, X., Zheng, M., Zhu, J. J., Szalaj, P., Trzaskoma, P., Magalska, A., Wlodarczyk, J., Ruszczycki, B., Michalski, P., Piecuch, E., Wang, P., Wang, D., Tian, S. Z., Penrad-Mobayed, M., Sachs, L. M., Ruan, X., Wei, C. L., … Ruan, Y. (2015). CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell*, *163*(7), 1611–1627. https://doi.org/10.1016/j.cell.2015.11.024

Taunton, J., Hassig, C. A., & Schreiber, S. L. (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science*, *272*, 408–411.

Taylor, J. S., Durkin, J. M. H., & Breden, F. (1999). The death of a microsatellite: A phylogenetic perspective on microsatellite interruptions [2]. *Molecular Biology and Evolution*, *16*(4), 567–572. https://doi.org/10.1093/oxfordjournals.molbev.a026138

Thakur, J., Packiaraj, J., & Henikoff, S. (2021). Sequence, Chromatin and Evolution of Satellite DNA. *International Journal of Molecular Sciences*, *22*(4309).

Thompson, P. J., Macfarlan, T. S., & Lorincz, M. C. (2016). Long Terminal Repeats: From Parasitic Elements to Building Blocks of the Transcriptional Regulatory Repertoire. *Molecular Cell*, *62*(5), 766–776. https://doi.org/10.1016/j.molcel.2016.03.029

Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E., Sheffield, N. C., Stergachis, A. B., Wang, H., Vernot, B., Garg, K., John, S., Sandstrom, R., Bates, D., Boatman, L., Canfield, T. K., Diegel, M., Dunn, D., Ebersol, A. K., … Stamatoyannopoulos, J. A. (2012). The accessible chromatin landscape of the human genome. *Nature*, *489*(7414), 75–82. https://doi.org/10.1038/nature11232

Toivonen, S., Lundin, K., Balboa, D., Ustinov, J., Tamminen, K., Palgi, J., Trokovic, R., Tuuri, T., & Otonkoski, T. (2013). Activin A and Wnt-dependent specification of human definitive endoderm cells. *Experimental Cell Research*, *319*(17), 2535–2544. https://doi.org/10.1016/j.yexcr.2013.07.007

Tóth, G., Gáspári, Z., & Jurka, J. (2000). Microsatellites in different Potyvirus genomes: Survey and analysis. *Genome Research*, *10*(7), 967–981. https://doi.org/10.1016/j.gene.2011.08.016

Trompouki, E., Bowman, T. V., Lawton, L. N., Fan, Z. P., Wu, D. C., Dibiase, A., Martin, C. S., Cech, J. N., Sessa, A. K., Leblanc, J. L., Li, P., Durand, E. M., Mosimann, C., Heffner, G. C., Daley, G. Q., Paulson, R. F., Young, R. A., & Zon, L. I. (2011). Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. *Cell*, *147*(3), 577–589. https://doi.org/10.1016/j.cell.2011.09.044

Tsankov, A. M., Gu, H., Akopian, V., Ziller, M. J., Donaghey, J., Amit, I., Gnirke, A., & Meissner, A. (2015). Transcription factor binding dynamics during human ES cell differentiation. *Nature*, *518*(7539), 344–349. https://doi.org/10.1038/nature14233

Tsumura, A., Hayakawa, T., Kumaki, Y., Takebayashi, S. I., Sakaue, M., Matsuoka, C., Shimotohno, K., Ishikawa, F., Li, E., Ueda, H. R., Nakayama, J. I., & Okano, M. (2006). Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes to Cells*, *11*(7), 805–814. https://doi.org/10.1111/j.1365-2443.2006.00984.x

Urieli-Shoval, S., Gruenbaum, Y., Sedat, J., & Razin, A. (19821). The absence of detectable methylated bases in Drosophila melanogaster. *FEBS Letters*, *146*(1), 148–149. https://doi.org/10.7868/s0869565218110257

Vanzan, L., Soldati, H., Ythier, V., Anand, S., Braun, S. M. G., Francis, N., & Murr, R. (2021). High throughput screening identifies SOX2 as a super pioneer factor that inhibits DNA methylation maintenance at its binding sites. *Nature Communications*, *12*(1), 1–18. https://doi.org/10.1038/s41467-021-23630-x

Veil, M., Yampolsky, L. Y., Grüning, B., & Onichtchouk, D. (2019). Pou5f3, SoxB1, and Nanog remodel chromatin on high nucleosome affinity regions at zygotic genome activation. *Genome Research*, *29*(3), 383–395. https://doi.org/10.1101/gr.240572.118

Verma, N., Pan, H., Doré, L. C., Shukla, A., Li, Q. V., Pelham-Webb, B., Teijeiro, V., González, F., Krivtsov, A., Chang, C. J., Papapetrou, E. P., He, C., Elemento, O., & Huangfu, D. (2018). TET proteins safeguard bivalent promoters from de novo methylation in human embryonic stem cells. *Nature Genetics*, *50*(1), 83–95. https://doi.org/10.1038/s41588-017-0002-y

Veselovska, L., Smallwood, S. A., Saadeh, H., Stewart, K. R., Krueger, F., Maupetit Méhouas, S., Arnaud, P., Tomizawa, S. ichi, Andrews, S., & Kelsey, G. (2015). Deep sequencing and de novo assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape. *Genome Biology*, *16*(1), 1–17. https://doi.org/10.1186/s13059-015-0769-z

Vincenzetti, L., Leoni, C., Chirichella, M., Kwee, I., & Monticelli, S. (2019). The contribution of active and passive mechanisms of 5mC and 5hmC removal in human T lymphocytes is differentiation- and activation-dependent. *European Journal of Immunology*, *49*(4), 611–625. https://doi.org/10.1002/eji.201847967

Viré, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J. M., Bollen, M., Esteller, M., Di Croce, L., De Launoit, Y., & Fuks, F. (2006). The Polycomb group protein EZH2 directly controls DNA methylation. *Nature*, *439*(7078), 871–874. https://doi.org/10.1038/nature04431

Waalwijk, C., & Flavell, R. A. (1978). DNA methylation at a CCGG sequence in the large intron of the rabbit beta-globin gene: tissue-specific variations. *Nucleic Acids Research*, *5*(12), 4631–4641.

Walsh, C. P., Chaillet, J. R., & Bestor, T. H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nature Genetics*, *20*(october), 116–117.

Walter, M. J., Ding, L., Shen, D., Shao, J., Grillot, M., McLellan, M., Fulton, R., Schmidt, H., Kalicki-Veizer, J., O'Laughlin, M., Kandoth, C., Baty, J., Westervelt, P., Dipersio, J. F., Mardis, E. R., Wilson, R. K., Ley, T. J., & Graubert, T. A. (2011). Recurrent DNMT3A mutations in patients with myelodysplastic syndromes. *Leukemia*, *25*(7), 1153–1158. https://doi.org/10.1038/leu.2011.44

Walter, M., Teissandier, A., Pérez-Palacios, R., & Bourc'his, D. (2016). An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *ELife*, *5*, 1–30. https://doi.org/10.7554/elife.11418

Wang, J., Walsh, G., Liu, D. D., Lee, J. J., & Mao, L. (2006). Expression of ΔDNMT3B variants and its association with promoter methylation of p16 and RASSF1A in primary non-small cell lung cancer. *Cancer Research*, *66*(17), 8361–8366. https://doi.org/10.1158/0008-5472.CAN-06-2031

Wang, L., Zhang, J., Duan, J., Gao, X., Zhu, W., Lu, X., Yang, L., Zhang, J., Li, G., Ci, W., Li, W., Zhou, Q., Aluru, N., Tang, F., He, C., Huang, X., & Liu, J. (2014). Programming and inheritance of parental DNA methylomes in mammals. *Cell*, *157*(4), 979–991. https://doi.org/10.1016/j.cell.2014.04.017

Watt, F., & Molloy, P. L. (1988). Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes & Development*, *2*(9), 1136–1143. https://doi.org/10.1101/gad.2.9.1136

Weber, M., Davies, J. J., Wittig, D., Oakeley, E. J., Haase, M., Lam, W. L., & Schübeler, D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nature Genetics*, *37*(8), 853–862. https://doi.org/10.1038/ng1598

Weber, M., Hellmann, I., Stadler, M. B., Ramos, L., Pääbo, S., Rebhan, M., & Schübeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature Genetics*, *39*(4), 457–466. https://doi.org/10.1038/ng1990

Weinberg, D. N., Papillon-Cavanagh, S., Chen, H., Yue, Y., Chen, X., Rajagopalan, K. N., Horth, C., McGuire, J. T., Xu, X., Nikbakht, H., Lemiesz, A. E., Marchione, D. M., Marunde, M. R., Meiners, M. J., Cheek, M. A., Keogh, M. C., Bareke, E., Djedid, A., Harutyunyan, A. S., … Lu, C. (2019). The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature*, *573*(7773), 281–286. https://doi.org/10.1038/s41586-019-1534-3

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. In *Media* (Vol. 35, Issue July). http://link.springer.com/10.1007/978-0-387-98141-3

Williams, K., Christensen, J., Pedersen, M. T., Johansen, J. V., Cloos, P. A. C., Rappsilber, J., & Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature*, *473*(7347), 343–349. https://doi.org/10.1038/nature10066

Wu, H., D'Alessio, A. C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Eve Sun, Y., & Zhang, Y. (2011a). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature*, *473*(7347), 389–394. https://doi.org/10.1038/nature09934

Wu, H., D'Alessio, A. C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Eve Sun, Y., & Zhang, Y. (2011b). Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes & Development*, *25*, 679–684. https://doi.org/10.1038/nature09934

Xi, Y., & Li, W. (2009). BSMAP: Whole genome bisulfite sequence MAPping program. *BMC Bioinformatics*, *10*, 1–9. https://doi.org/10.1186/1471-2105-10-232

Xie, S., Wang, Z., Okano, M., Nogami, M., Li, Y., He, W. W., Okumura, K., & Li, E. (1999). Cloning, expression and chromosome locations of the human DNMT3 gene family. *Gene*, *236*(1), 87–95. https://doi.org/10.1016/S0378-1119(99)00252-8

Xu, G., Bestor, T. H., Bourc'his, D., Hsieh, C., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J. J., Viegas-pe, E., Malades, Ã. N., & Se, D. (1999). *Chromosome instability and immunodefficiency syndrome caused by mutat*. *402*(November), 187–191.

Xu, Y., Xu, C., Kato, A., Tempel, W., Abreu, J. G., Bian, C., Hu, Y., Hu, D., Zhao, B., Cerovina, T., Diao, J., Wu, F., He, H. H., Cui, Q., Clark, E., Ma, C., Barbara, A., Veenstra, G. J. C., Xu, G., … Shi, Y. G. (2012). Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for xenopus eye and neural development. *Cell*, *151*(6), 1200–1213. https://doi.org/10.1016/j.cell.2012.11.014

Yamaguchi, S., Shen, L., Liu, Y., Sendler, D., & Zhang, Y. (2013). Role of Tet1 in erasure of genomic imprinting. *Nature*, *504*(7480), 460–464. https://doi.org/10.1038/nature12805

Yan, X. J., Xu, J., Gu, Z. H., Pan, C. M., Lu, G., Shen, Y., Shi, J. Y., Zhu, Y. M., Tang, L., Zhang, X. W., Liang, W. X., Mi, J. Q., Song, H. D., Li, K. Q., Chen, Z., & Chen, S. J. (2011). Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nature Genetics*, *43*(4), 309–317. https://doi.org/10.1038/ng.788

Yang, J., Horton, J. R., Li, J., Huang, Y., Zhang, X., Blumenthal, R. M., & Cheng, X. (2019). Structural basis for preferential binding of human TCF4 to DNA containing 5-carboxylcytosine. *Nucleic Acids Research*, *47*(16), 8375–8387. https://doi.org/10.1093/nar/gkz381

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P. K., Kivioja, T., Dave, K., Zhong, F., Nitta, K. R., Taipale, M., Popov, A., Ginno, P. A., Domcke, S., Yan, J., Schübeler, D., Vinson, C., & Taipale, J. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*, *356*(6337). https://doi.org/10.1126/science.aaj2239

Yong, B. P., Yoon, Y. K., Sun, K. O., Sun, G. C., Ku, S. Y., Seok, H. K., Young, M. C., & Shin, Y. M. (2008). Alterations of proliferative and differentiation potentials of human embryonic stem cells during long-term culture. *Experimental and Molecular Medicine*, *40*(1), 98–108. https://doi.org/10.3858/emm.2008.40.1.98

Zaret, K. S., & Carroll, J. S. (2011). Pioneer transcription factors: Establishing competence for gene expression. *Genes and Development*, *25*(21), 2227–2241. https://doi.org/10.1101/gad.176826.111

Zentner, G. E., Tesar, P. J., & Scacheri, P. C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Research*, *21*(8), 1273–1283. https://doi.org/10.1101/gr.122382.111

Zhang, H., Zhang, X., Clark, E., Mulcahey, M., Huang, S., & Shi, Y. G. (2010). TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine. *Cell Research*, *20*(12), 1390–1393. https://doi.org/10.1038/cr.2010.156

Zhang, W., Xia, W., Wang, Q., Towers, A. J., Chen, J., Gao, R., Zhang, Y., Yen, C. an, Lee, A. Y., Li, Y., Zhou, C., Liu, K., Zhang, J., Gu, T. P., Chen, X., Chang, Z., Leung, D., Gao, S., Jiang, Y. hui, & Xie, W. (2016). Isoform Switch of TET1 Regulates DNA Demethylation and Mouse Development. *Molecular Cell*, *64*(6), 1062–1073. https://doi.org/10.1016/j.molcel.2016.10.030

Zhang, Yingying, Jurkowska, R., Soeroes, S., Rajavelu, A., Dhayalan, A., Bock, I., Rathert, P., Brandt, O., Reinhardt, R., Fischle, W., & Jeltsch, A. (2010). Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. *Nucleic Acids Research*, *38*(13), 4246–4253. https://doi.org/10.1093/nar/gkq147

Zhang, Yong, Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nussbaum, C., Myers, R. M., Brown, M., Li, W., & Shirley, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9). https://doi.org/10.1186/gb-2008-9-9-r137

Zhang, Yu, Xiang, Y., Yin, Q., Du, Z., Peng, X., Wang, Q., Fidalgo, M., Xia, W., Li, Y., Zhao, Z. A., Zhang, W., Ma, J., Xu, F., Wang, J., Li, L., & Xie, W. (2018). Dynamic epigenomic landscapes during early lineage specification in mouse embryos. *Nature Genetics*, *50*(1), 96–105. https://doi.org/10.1038/s41588-017-0003-x

Zhu, P., Guo, H., Ren, Y., Hou, Y., Dong, J., Li, R., Lian, Y., Fan, X., Hu, B., Gao, Y., Wang, X., Wei, Y., Liu, P., Yan, J., Ren, X., Yuan, P., Yuan, Y., Yan, Z., Wen, L., … Tang, F. (2018). Single-cell DNA methylome sequencing of human preimplantation embryos. *Nature Genetics*, *50*(1), 12–19. https://doi.org/10.1038/s41588-017-0007-6

Zhu, Y., Sun, L., Chen, Z., Whitaker, J. W., Wang, T., & Wang, W. (2013). Predicting enhancer transcription and activity from chromatin modifications. *Nucleic Acids Research*, *41*(22), 10032–10043. https://doi.org/10.1093/nar/gkt826

Ziller, M. J., Edri, R., Yaffe, Y., Donaghey, J., Pop, R., Mallard, W., Issner, R., Gifford, C. A., Goren, A., Xing, J., Gu, H., Cacchiarelli, D., Tsankov, A. M., Epstein, C., Rinn, J. L., Mikkelsen, T. S., Kohlbacher, O., Gnirke, A., Bernstein, B. E., … Meissner, A. (2015). Dissecting neural differentiation regulatory networks through epigenetic footprinting. *Nature*, *518*(7539), 355–359. https://doi.org/10.1038/nature13990

Ziller, M. J., Gu, H., Müller, F., Donaghey, J., Tsai, L. T. Y., Kohlbacher, O., De Jager, P. L., Rosen, E. D., Bennett, D. A., Bernstein, B. E., Gnirke, A., & Meissner, A. (2013). Charting a dynamic DNA methylation landscape of the human genome. *Nature*, *500*(7463), 477–481. https://doi.org/10.1038/nature12433

Zuo, X., Sheng, J., Lau, H. T., McDonald, C. M., Andrade, M., Cullen, D. E., Bell, F. T., Iacovino, M., Kyba, M., Xu, G., & Li, X. (2012). Zinc finger protein ZFP57 requires its co-factor to recruit DNA methyltransferases and maintains DNA methylation imprint in embryonic stem cells via its transcriptional repression domain. *Journal of Biological Chemistry*, *287*(3), 2107–2118. https://doi.org/10.1074/jbc.M111.322644

Żylicz, J. J., Bousard, A., Žumer, K., Dossin, F., Mohammad, E., da Rocha, S. T., Schwalb, B., Syx, L., Dingli, F., Loew, D., Cramer, P., & Heard, E. (2019). The Implication of Early Chromatin Changes in X Chromosome Inactivation. *Cell*, *176*(1–2), 182-197.e23. https://doi.org/10.1016/j.cell.2018.11.041

LIST OF PUBLICATIONS

Contributions

Veenvliet, J. V.*, Bolondi, A.*, Kretzmer, H.**, **Haut, L.****, Scholze-Wittler, M., Schifferl, D., Koch, F., Guignard, L., Sampath Kumar, A., Pustet, M., Heimann, S., Buschow, R., Wittler, L., Timmermann, B., Meissner, A.*** & Herrmann, B. G.*** (2020). Mouse embryonic stem cells self-organize into trunk-like structures with neural tube and somites. *Science*, *370*(6522), eaba4937.

*These authors contributed equally to this study

**These authors contributed equally to this study

***These authors contributed equally to this study

Bolondi, A., **Haut, L.**, Gassaloglu, S. I., Burton, P., Kretzmer, H., Buschow, R., Meissner, A., Herrmann, B. G. & Veenvliet, J. V. (2021). Generation of mouse pluripotent stem cell-derived trunk-like structures: an in vitro model of post-implantation embryogenesis. *Bio-protocol*, *11*(11), e4042-e4042.

Sampath Kumar, A.*, Tian, L.*, Bolondi, A.*, Hernández, A. A., Stickels, R., Kretzmer, H., Murray, E., Wittler, L., Walther, M., Barakat G., **Haut, L**., Elkabetz, Y., Macosko E. Z.**, Guignard L.**, Chen F.** & Meissner, A.** (2023). Spatiotemporal transcriptomic maps of whole mouse embryos at the onset of organogenesis. *Nature Genetics*, 1-10.

*These authors contributed equally to this study

**These authors contributed equally to this study