Inaugural Dissertation
to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)

The consequences of DNA methylation maintenance deficiency in human Embryonic Stem Cells

submitted to the Department of Biology, Chemistry, Pharmacy
of Freie Universität Berlin

by

Aleksandra Alicja Sieńska

2020

The thesis summarizes the work was done from 10/2015 to 01/2020
at the Harvard Stem Cell Institute and Max Planck Institute for Molecular Genetics
under the supervision of Prof. Dr. Alexander Meissner

1st Reviewer: Prof. Dr. Alexander Meissner
2nd Reviewer: Prof. Dr. Katja Nowick

date of defense: 21.09.2020

I declare to Freie Universität Berlin that I have prepared this doctoral thesis independently and without using any other than the specified sources and aids. The present work is free from plagiarism. This thesis has not been submitted in the same or a similar form as an examination at any other university.

ACKNOWLEDGEMENTS

I would like to thank Prof. Dr. Alexander Meissner for giving me the opportunity to work in his laboratory.

***

My kindest thanks to Toshiyuki Ushijima for performing the DNMT3A/B KO cell methylcytosine ELISA, Simon Lauer for kindly sharing his degron-DNMT1 tagged cell lines with me and performing the HEK293T cell western blot, and to Raha Weigert for performing the Scorecard qPCR.

***

This work would not have been possible without the support of Dr. Andreas Gnirke, Dr. Hongcang Gu and Arman Mohammad, who performed the dual protocol single cell sequencing library preparation, as well as Dr. Martin Aryee, Dr. Ayush Raman and Caleb Laureau, who performed the Quality Control and genome alignment of the single-cell data. I would also like to thank Dr. Jocelyn Charlton and Dr. Sudhir Thakurela for their help with the analysis of scRRBS and scRNAseq data, respectively.

***

I would like to thank the Joslin FACS Core and the HMS Imaging core for their support.

***

A warm thank you to all the members of the Meissner Lab, former and present, for the amazing atmosphere and all the inspiring discussions on scientific and non-scientific topics alike. I am especially thankful to my students, whom I have learned from even more than they have learned from me. Working with you has been a privilege.

***

Thank you to my family for their support and faith, I am truly blessed to be surrounded by so much kindness and love.

***

I am eternally grateful to my husband, who supported me through thick and thin, from Boston to Berlin. It is no understatement to say that I would have never done this without you. Thank you for the long hours of scientific discussions, reading papers and always challenging me to think like a biologist.

Table of contents

1. ABSTRACT

A methyl group deposited on cytosines incorporated into the sequence of the DNA, so called DNA methylation, decorates the genomes of a large number of species, from archaea to man. Over the last two decades, a large body of research discovered that this small chemical moiety elicits a profound effect on the gene expression program. In particular, DNA methylation restricts transcriptionally active regions of the genome, therefore ensuring a faithful interpretation of the regulatory information encoded in the DNA sequence. This fundamental role played by the methylation of DNA helps define cell identity at a molecular level, thus it enables a biologically complex transition such as from a zygote to an organism to occur in a unidirectional and orchestrated manner. Perturbations in the pattern of DNA methylation have been frequently found in pathological processes such as tumorigenesis.

The pattern of DNA methylation decorating the genome of a cell is precisely copied during cell division by maintenance machinery composed of the DNMT1 enzyme and its associated proteins. The absence of DNMT1 elicits a wide-range of deleterious effects, from loss of cell fitness of *in vitro* cultured cells to embryonic lethality and loss of homeostasis of somatic tissues. Previous studies reported pleiotropic effects and mutually exclusive phenotypes of DNMT1 knockout depending on the design of the study – from apoptosis and genomic instability to accelerated cell cycle and trans-differentiation. How exactly these phenotypes arise in a response to DNMT1 deficiency is unknown.

We employed the state-of-the-art next generation sequencing technologies and coupled them with molecular and cell biology techniques to elucidate the causes for the loss of fitness of DNMT1-deficient human embryonic stem cells. In contrast to previous studies, we did not observe the proposed DNA damage or genomic instability. Our work demonstrated that an acute depletion of DNMT1 results in a uniform decay of DNA methylation that we characterized in depth at a single cell level. Interestingly, our transcriptome profiling in single cells followed by functional validations revealed a change in the way how the transcriptional machinery interprets the genome in the absence of DNMT1. The loss of global DNA methylation without its maintenance machinery resulted in transcriptional changes mainly related to some gonad-specific genes and also a few genes encoding key players of a signaling transduction pathway. This finding inspired us to discover that the cells deficient for DNMT1 display a lower threshold for activating transcription once challenged with external stimuli. Our findings therefore provide new insights into how genome deficient for cytosine methylation becomes transcriptionally

amenable, thus capable to integrate and respond to new signals from the environment. Our work lays a foundation for future studies on how such process leads to developmental defects and disease states.

2. ZUSAMMENFASSUNG

Methylgruppen an in die DNA eingebauten Cytosinbasen, die sogenannte DNA-Methylierung, zieren die Genome einer großen Anzahl von Arten, von den Archaea bis zu den Menschen. In den letzten zwei Jahrzehnten hat eine Vielzahl von Forschungsarbeiten ergeben, dass diese kleine chemische Einheit eine tiefgreifende Auswirkung auf das Genexpressionsprogramm hat. Insbesondere beschränkt die DNA-Methylierung die transkriptionell aktiven Regionen des Genoms und gewährleistet so eine genaue Interpretation der in der DNA-Sequenz kodierten regulatorischen Informationen. Diese grundlegende Rolle der Methylierung von DNA hilft dabei, die Zellidentität auf molekularer Ebene zu definieren und ermöglicht so, dass biologische komplexe Übergänge, wie z.B. von der Zygote zum Organismus, auf unidirektionaler und orchestrierter Art und Weise stattfinden können. Störungen im Muster der DNA-Methylierung werden häufig bei pathologischen Prozessen wie der Tumorentstehung gefunden.

Das Muster der DNA-Methylierung auf dem Genom einer Zelle wird während der Zellteilung durch eine Wartungsmaschinerie, die aus dem DNMT1-Enzym und damit verbundenen Proteinen bestehen, präzise kopiert. Die Abwesenheit von DNMT1 hat eine Vielzahl von schädlichen Auswirkungen, von einem Verlust der Tauglichkeit von *in vitro* kultivierten Zellen bis zur embryonalen Letalität und einem Verlust der Homöostase von somatischen Geweben. Frühere Studien wurden pleiotrope Effekte und sich gegenseitig ausschließende Phänotypen des DNMT1-Knockouts in Abhängigkeit vom Studiendesign beschrieben–von Apoptose und genomischer Instabilität bis hin zu einem beschleunigten Zellzyklus und der Transdifferenzierung von Zellen. Wie genau diese Phänotypen bei einer Reaktion auf einen DNMT1-Mangel auftreten, ist nicht bekannt.

Wir setzten die neuesten Sequenzierungstechnologien ein und kombinierten diese mit molekularen und zellbiologischen Techniken, um die Ursachen für den Fitnessverlust von menschlichen embryonalen Stammzellen mit DNMT1-Mangel aufzuklären. Im Gegensatz zu bisherigen Studien haben wir keine DNA-Schädigung oder genomische Instabilität beobachtet. Unsere Arbeit zeigte, dass eine akute Reduktion von DNMT1 zu einem gleichmäßigen Abbau der DNA-Methylierung führt, den wir auf Einzelzellebene eingehend charakterisierten. Interessanterweise ergab unsere Auswertung des Transkriptoms einzelner Zellen, gefolgt von funktionellen Validierungen, dass sich die Art und Weise, wie die Transkriptionsmaschinerie das Genom interpretiert, in Abwesenheit von DNMT1 verändert. Der Verlust der globalen DNA-Methylierung ohne ihre Erhaltungsmaschinerie führte zu Veränderungen der Transkription einiger

gonadenspezifischer Gene und einigen wenigen Genen, die die Hauptakteure eines Signaltransduktionswegs kodieren. Dieser Befund führte zu unserer Entdeckung, dass die Zellen, denen DNMT1 fehlt, eine niedrigere Schwelle für die Aktivierung der Transkription aufweisen, sobald sie mit externen Stimuli in Kontakt gebracht werden. Unsere Ergebnisse liefern daher neue Erkenntnisse darüber, wie Genome, denen die Cytosinmethylierung fehlt, transkriptionszugänglich werden und somit in der Lage sind, neue Signale aus der Umwelt zu integrieren und auf diese zu reagieren. Unsere Arbeit bildet die Grundlage für zukünftige Studien darüber, wie ein solcher Prozess zu Entwicklungsstörungen und Krankheitszuständen führt.

3. INTRODUCTION

Through the process of cell division and subsequent differentiation, a single cell can form all the tissues of a multicellular organism. This is possible, because all the cells share the same genome – a deoxyribonucleic acid (DNA) entity, which encodes genes. Intrinsic and extrinsic signals cooperate to shape the expression of genes resulting in morphologically and functionally diverse cells. Inability to secure the proper gene expression profile can be detrimental to a cell by changing its properties, restricting differentiation potential or altering its interaction with the surrounding environment. This simple principle lays a foundation for each process in biology such as development, tissue homeostasis or aging. How this complex process gene expression is executed and regulated has been one of the fundamental questions in biology.

3.1 Regulation of gene expression

The genes are expressed in the process of transcription, where a specialized protein machinery uses DNA as a template to produce molecules of messenger RNA (mRNA), which then serve to instruct the synthesis of proteins. The decision which DNA sequence is transcribed depends mainly on two factors: (1) the presence and engagement of transcription factors (TFs), instructed *in cis* by the so-called TF binding motifs encoded in the DNA and (2) an additional layer of information independent on the DNA sequence that influences the accessibility of the DNA. The latter is largely defined by chromatin – the assembly of proteins and DNA, which serves to compact the genetic information in the cell nucleus. The basic unit of chromatin is the nucleosome – a composite of DNA wound around an octamer of histone proteins. The N-terminal tails of histones are substrates for post-translational modification by enzymes, which in turn recruits or inhibits the binding of large chromatin-modifying complexes that control the accessibility of the chromatin and thus impact transcription[1].

Not only the nucleosomes, but also the nitrogenous bases of the DNA can be modified by enzymes in a manner that does not change the DNA base pairing. Such modification of may be instructive for DNA-binding proteins and, indirectly, regulate the transcriptional potential of underlying genes. Collectively, heritable modifications affecting gene expression without modifying the DNA sequence are called “epigenetic modifications”.

Among many possible modifications of the DNA, methylation of the 5th carbon in cytosine ring embedded in the DNA (referred thereafter as the DNA methylation) was identified in 1925[2] and has been extensively studied for decades. This small modification is highly conserved, present in bacteria, through the earliest eukaryotes[3], plants[4], fungi[5] and animals[6]. Interestingly, DNA methylation is absent in some model organisms, such as yeast, *C. elegans* or the fruit fly[7]. Why some organisms do not possess DNA methylation remains unknown.

Technological advances over the last decades enabled both quantitative (mass spectrometry[8]) and qualitative (DNA sequencing[9]) studies of modification of nucleotides present in the DNA. Mass spectrometry experiments performed on mammalian genomic DNA identified methylation of about 1-2% cytosines in the genome of mammalian cells, which makes the cytosine methylation the predominant modification found in the mammalian DNA. The DNA methylation occurs in the context of CpG dinucleotides and, to a lesser extent, CpA dinucleotides[10]. The methyl group is deposited on cytosines in the DNA by specialized enzymes, called the DNA methyltransferases (DNMTs)[11].

3.2 The establishment of DNA methylation

To date, three enzymes were identified to deposit methyl groups onto previously unmodified cytosines (*de novo* methylation), thus establishing a genome-wide DNA methylation pattern. DNA methyltransferase 3A and 3B (DNMT3A and DNMT3B)[12] are expressed across various cell types, while the recently-identified DNMT3C[13] is restricted to the male germline of mice and rats, and is not in the scope of this study.

In order to secure the specificity and fidelity of their enzymatic activity, DNMTs are composed of multiple domains.

DNA methyltransferase domain. Cytosine methylation is performed by a highly conserved methyltransferase (MTase) domain present at the C-terminus of all catalytically active DNMTs[14]. The catalytic activity of DNMT3A and B is enhanced by DNMT3L[15] (an inactive DNA methyltransferase) in the germline. Two chromatin-recognizing domains restrict the substrate specificity of de novo methyltransferases.

ADD domain. The ATRX-DNMT3-DNMT3L (ADD) domain binds unmodified N-terminal tail of histone H3 and it is repelled by modification on lysine 4 (H3K4)[16]. Unless the binding to H3K4 occured, the ADD domain has autoinhibitory properties and hinders the MTase domain[17].

PWWP domain. The Proline-Tryptophan-Tryptophan-Proline (PWWP) domain specifically binds the methylated histone H3 at the lysine 36 (H3K36me3), which is sufficient for DNMT3B recruitment[18]. As the methylation of H3K36 is a co-transcriptional process[19], it is believed that the PWWP domain confers specificity for methylation of actively-transcribed genes by DNMT3A and B.

3.3 Maintenance of DNA methylation

Cytosines are methylated in the context of a CpG dinucleotide, the symmetry of which allows maintenance of this mark over DNA replication. Specifically, when nascent the strand is synthesized, by the virtue of DNA complementarity, a CpG dinucleotide on the parental strand will be paired with another CpG on the new strand. The existence of an enzyme that can recognize mCpG:CpG pairing (so-called “hemimethylated DNA”) and methylate the nascent strand CpG using the parental strand as a template had been predicted[20] and, eventually, the responsible protein was identified. The DNA methyltransferase 1 (DNMT1) faithfully maintains the genome-wide DNA methylation over cell divisions[21].

DNMT1 is structurally different from the *de novo* DNMTs even though they all share the MTase domain. In addition, DNMT1 contains different N-terminal domains and also distinct interacting proteins that altogether dictate its specificity to hemimethylated DNA, as well as its stability[22].

BAH domains. In contrast to DNMT3A and 3B, the isolated MTase domain of DNMT1 is catalytically inactive and its proper folding depends on two bromo-adjacent homology (BAH) domains, which adopt a dumbbell-like shape in close contact with the catalytic domain[23].

RFTS domain. Similarly to the ADD domain in DNMT3 enzymes, DNMT1 is auto-inhibited by an replication foci targeting sequence (RFTS) domain[24]. In order to release the MTase domain, RFTS binds the ubiquitinated N-terminal tail of histone H3. This mark is bestowed upon histones by DNMT1 binding partner – Ubiquitin-like with PHD and Ring Finger Domains 1 (UHRF1)[25,26].

DNMT1-interacting protein, UHRF1. This E3 ubiquitin-protein ligase recognizes hemimethylated DNA through its SET- and RING- associated (SRA) domain[25,26]. Furthermore, its tandem TUDOR-PHD (TTD- PHD) domain allows for binding to methylated lysine 9 on histone H3 (H3K9me2 and me3)[27], facilitating the recruitment of

DNMT1 to heterochromatin. In agreement with UHRF1 being required for DNMT1 function, UHRF1 knockout phenocopies the DNMT1 knockout *in vivo*[26].

CxxC domain. Although DNMT1 possesses a CXXC domain, which is known to bind unmethylated CpG-rich DNA regions[28], it is unclear whether this domain regulates the DNMT1 enzyme specificity[29].

PCNA-interacting domain. The faithful maintenance of the methylated cytosine pattern on both strands of DNA occurs during DNA replication in the S-phase of the cell cycle. Indeed, DNMT1 interacts with the DNA polymerase complex through Proliferating Cell Nuclear Antigen (PCNA), which ensures that the entire genome becomes a substrate for copying the methyl mark[30].

PTMs regulating DNMT1. Besides the protein-protein interactions dictating the spatial distribution of DNMT1, DNMT1 is also restricted temporally based on its abundance via post-translational modifications (PTMs). A number of PTMs on DNMT1 were described so far. For example, the ubiquitination of DNMT1 on lysine K142 occurs in late S-phase and marks the protein for proteasomal degradation in a cell-cycle dependent manner, so that it is not present outside of the S-phase[31]. In contrary, a phosphorylation of serine S143 stabilizes DNMT1[32]. Another mode of the proteasomal degradation of DNMT1 may be proceeded and promoted by acetylation of lysine residues followed by the UHRF1-mediated DNMT1 ubiquitination[33]. This mode of DNMT1 degradation is antagonized by a specific set of deacetylases and deubiquitinases[34,35].

Altogether, multiple enzymes and mechanisms act on DNMT1 to restrict its function in a spatial and temporal manner, and to secure the process of maintaining the pattern of DNA methylation.

3.4 Genome-wide distribution of DNA methylation

The distribution of DNA methylation across the genome is non-random. This is because the distribution of substrate cytosines in CpG dinucleotide is not uniform across the genome. Less than 10% of CpGs occur in the highly CpG-dense regions of high cytosine and guanine content termed the CpG islands (CGIs)[36]. This DNA sequence feature is enriched in the promoters of two thirds of genes and virtually all housekeeping genes[37]. In contrast, the “open sea” CpG dinucleotides (outside of the CGIs) are mostly methylated. Overall, out of about 28 million CpGs in the human genome, 60-80% are methylated in somatic cells[38,39]. In the following section, we will further discuss the

function of DNA methylation across the genomes and focus on its role in (1) controlling transcription and (2) as a source of mutagenesis.

DNA methylation controls gene expression on sex chromosomes.

Male and female genomes differ with respect to the so-called sex chromosomes. While females have two X chromosomes, males have an X and a Y chromosome. This poses a challenge for the female genome, which has to reduce the dose of the gene products present on the X chromosomes, so that it equals the transcriptional output of the X chromosome in the male genome. This so-called dosage compensation process occurs during early embryonic development via a transcriptional shutoff of one of the X chromosomes in the process termed X chromosome inactivation (XCI)[40]. The inactivated X becomes coated in the *Xist* lnc-RNA, which leads to transcriptional silencing, followed by chromatin remodeling[41]. The transcriptional silencing process requires the CGIs of the inactive X chromosome to become methylated by DNMT3B[42].

DNA methylation controls of gene expression on autosomes.

Imprinted and germline-specific genes. A number of studies reported the function of DNA methylation in suppressing the expression of imprinted genes[43]. A selective DNA methylation of either the maternal or the paternal allele of certain genomic regions, known as the imprint control regions (ICR), leads to monoallelic transcription of neighboring genes[44]. The exact mechanism of imprinting is not known, however, loss of imprinting is associated with multiple human diseases[45,46]. In addition, high levels of DNA methylation are found on the promoters of germline-specific genes[47]. It is thought that this mark restricts the expression of such genes outside the gonads.

Transposable elements. About a half of the human genome consists of mobile selfish genetic elements, termed transposable elements (TEs). Such genetic elements have been debated as a potential source of new genes and cis-regulatory elements therefore contributing to genetic variation[48]. However, an uncontrolled expression and transposition of these genetic elements leads to a random mutagenesis, thus potentially impairing cell fitness[49]. Therefore, multiple mechanisms are deployed in gonads and somatic cells to restrict the expression of these genetic elements.

In gonads, the PIWI-interacting small RNAs (piRNAs) pathway acts to recognize transcriptionally active transposons and repress their transcription of transposons using chromatin modification, such as histone H3 methylation and DNA methylation[50]. Specific details of the piRNA pathway-induced TE silencing in mammals have not been yet elucidated.

While the piRNA pathway operates on evolutionary young TEs, another TE silencing pathway operates on evolutionary old TEs in the somatic cells. Here, the hundreds of Krüppel-associated box domain zinc finger proteins (KRAB-ZFPs) evolved to recognize DNA sequence motifs of TEs[51]. ZFPs recruit KRAB-associated protein 1 (KAP1), which acts as a scaffold for assembling gene-silencing complexes operating on chromatin. This includes the chromatin-modifying NuRD complex, as well as SETDB1 histone methyltransferase and HP1 chromodomain protein[52]. Upon the removal of active histone marks by the NuRD, SETDB1 methylates histone 3 lysine 9, which is recognized by HP1, a protein which compacts the chromatin. Subsequently, the DNMTs are recruited to the site, which results in hypermethylation of transposable element sites[53].

Genomic repeats. DNA methylation is enriched in tandem repeat regions and is believed to act to restrict their latent transcriptional activity[54]. An example of that are pericentromeric repeats, which consist of thousands of tandemly repeated copies, stretching from either side of the centromere to ensure the proper centromeric assembly[55]. Uncontrolled transcription from these loci is thought to interfere with the centromere assembly[54,56].

Other genes. In contrast to germline-specific genes, imprints or repetitive elements, the relationship between DNA methylation and most coding gene expression is not linear. Most coding gene promoters are in fact unmethylated. Conversely, the DNA methylation of active gene bodies is mediated through H3K36me3-dependent recruitment of de novo methyltransferases and thus positively correlates with gene expression[18]. Roles in inhibiting cryptic, intragenic promoters and instructing splicing have both been proposed for DNA methylation at gene bodies[57,58]. Also, a recent study analyzing the DNMT3B-depleted cells found aberrant transcription initiation sites in subtelomeric regions[59]. These findings demonstrate that although DNA methylation is often associated with transcriptionally inert genomic regions, it has also roles in regulating actively transcribed genes.

3.5 How DNA methylation exerts its function

Several modes of action have been proposed to date. DNA methylation is recognized specifically by methylated DNA binding domain (MBD) family of proteins[60]. As MBD proteins are known to interact with histone deacetylase and chromatin remodeling complexes[61], the transcriptionally repressive character of DNA methylation could be achieved by these 5mC-specific readers. High redundancy has obscured the studies of

the role of MBD proteins in gene silencing, as single knockouts do not exhibit severe phenotypes[62].

Recently, multiple lines of research have focused on how DNA methylation impacts the ability of TF to bind its cognate sequence in the DNA. Exhaustive studies of TF specificity revealed that the binding of over 170 transcription factors was impaired by DNA methylation[63]. Presence of DNA methylation at distal regulatory elements could, in principle, impede transcriptional activity even of unmethylated CGI-promoter genes[64]. Strikingly, there are transcription factors with preference for methylated DNA adding a further layer of complexity to the interpretation of the roles of DNA methylation in transcriptional regulation[63].

3.6 Methylated cytosine as a mutagen

The high levels of DNA methylation on the TEs and other repeats in the mammalian genome not only confer the transcriptional silencing, but could also introduce to a permanent alteration in the underlying DNA sequence through the process of deamination[65]. Unmodified cytosines undergo spontaneous deamination, which results in a conversion of the base to uracil. Such mutation is identified by the cell and corrected by base excision repair (BER)[66]. However, when methylcytosine is deaminated, it is converted to thymine, therefore introducing mutation. There is a dedicated enzyme called thymine DNA glycosylase (TDG) that corrects the aberrant G:T pairing[67]. However, if the mismatch was not recognized and repaired, it can be passed on to the daughter cell. Alteration in the DNA sequence of regulatory elements could, for example, impair transcription.

Recent study found that DNA methyltransferases introduced the toxic lesion 3-methylcytosine at a low rate both *in vitro* and *in vivo*. Comparative studies of nematode species that possess or lost DNMTs demonstrated that DNA methylation co-evolved with Alpha-Ketoglutarate Dependent Dioxygenase (ALKB2), an enzyme which repairs alkylation DNA damage[68]. Altogether, mutations are a byproduct of DNA methylation.

3.7 The removal of DNA methylation

There are two ways in which DNA methylation can be removed from its locus.

First, emerging directly from the mechanism of DNA methylation, is the lack of the maintenance for example by disrupting its machinery. Then, DNA methylation will become diluted over consecutive replication cycles (passive demethylation). Alternatively, in the process of active demethylation, specialized enzymes from the ten to eleven transferase (TET) family can catalyze the hydroxylation of methylcytosine to 5-hydroxymethylcytosine (5hmC)[69], 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC)[70]. Not only are these modifications not recognized by DNMT1 and are not maintained, they also engage thymine DNA glycosylase (TDG), which excises the base and the abasic site is repaired by downstream DNA damage repair[71]. TET dioxygenases and de novo DNMTs were found to competitively bind and regulate the level of DNA methylation in some loci[72]. This active demethylation acts independently of DNA replication and has more specificity than ablating DNA methylation maintenance. It is therefore well-suited for dynamically regulating the levels of this modification. In fact, TETs are indispensable for mouse embryonic development[73].

3.8 DNA methylation is dynamic in the early development

Although, the CpG methylation reaches 60-80% in the mammalian somatic cells, the levels of global DNA methylation fluctuate throughout the early embryonic development (FIG 1)[74]. During the early embryonic development, the DNA methylation pattern is erased from specialized cells (the egg and the sperm) in the so-called first wave of global DNA demethylation[75]. A subsequent re-establishment of DNA methylation is essential for the development of the newly forming organism[76,77].

Both passive demethylation (by excluding DNMT1 from the nucleus) and active demethylation mediated by TET oxidases contribute to the demethylation wave that result overall in about 20% of DNA methylation being retained. This includes the maternal and paternal imprints, as well as DNA methylation at the evolutionarily young and more potent transposable elements[78]. During embryo implantation, the genome becomes re-methylated to reach the ~80% DNA methylation maintained throughout the somatic lineages of the organism with a few modifications thereafter[79].

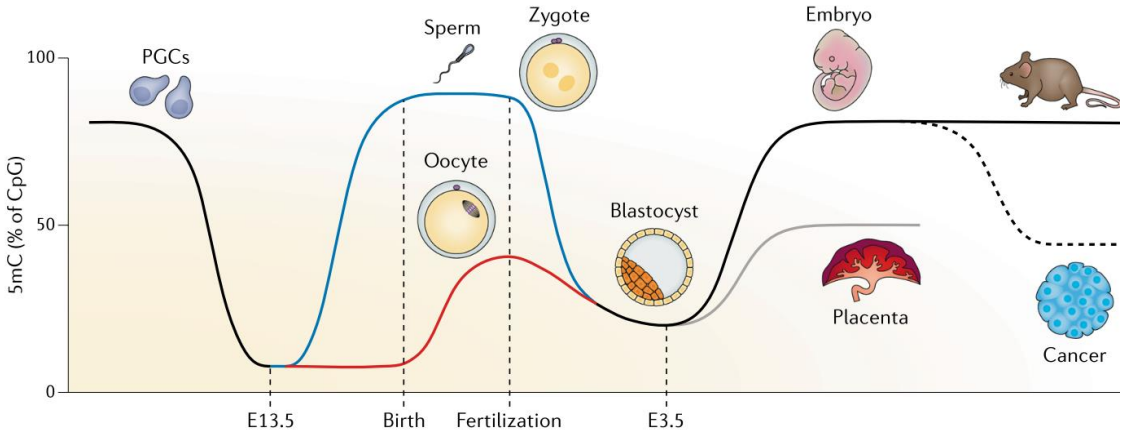


Figure 1: Global DNA methylation dynamics during development. Adapted from[80].

The second demethylation wave occurs during the development of primordial germ cells (PGCs) in a two-step manner. First, passive demethylation by the exclusion of UHRF1 from the nucleus occurs. In the second step[81], TET dioxygenases (specifically TET1 and TET2) remove DNA methylation mostly from imprinted loci, preparing the germline for establishing ICRs anew[82,83]. After the passive and active demethylation, only about 7% of CpGs in the primordial germ cells remain methylated[84]. The global DNA methylation is re-established in a sex-specific manner, with sperm having a significantly higher levels of methylation than oocytes[85,86]. While both waves of global DNA demethylation result in cell types with very low levels of methylcytosine, the state of hypomethylation occurs only transiently and even then the methylation of several loci is strictly maintained.

3.9 The impact of loss of DNMT on embryonic development

Function of DNA methylation in regulating gene expression, as well as its high conservation throughout the tree of life implies that it may play an important role at organismal level. Indeed, the previous genetic studies of the knockouts of DNA methyltransferases DNMT1 and DNMT3B demonstrated their early embryonic lethality, while the mice devoid of DNMT3A died shortly after birth[76,77]. Mouse embryos deficient for DNMT1 show a number of pathologies that emphasize the role of DNA methylation in a developing organism. In these DNMT1-deficient embryos resorption occurs around the neurulation stage[76]. Also, the knockout embryos were distinguished from littermates by their stunted growth at the E9 stage of development. The specimens fail to close the neural tube with an incomplete somitogenesis[76]. Likewise, the development of limb buds is retarded as well as allantois structures were abnormal in most of these embryos.

Moreover, histological analysis of several knockout embryo-derived tissues showed an increase in apoptosis. Lastly, the embryos do not produce blood and no blood vessels have been observed in the developing yolk sac[76].

In order to bypass the embryonic lethality of DNMT1KO embryos, mice containing one copy of a hypomorphic DNMT1 allele were engineered to produce only ~10% of DNMT1 protein compared to wild type (WT) cells[87]. Although these mice were viable, a marked loss of DNA methylation was observed on repetitive elements. These mice developed fatal T cell lymphomas[87].

The pleiotropic KO phenotype and link between DNA hypomethylation and cancer upon DNMT1 depletion prompted multiple *in vitro* studies to identify the exact cause of DNMT1KO embryo lethality and identify which of the functions of DNA methylation specifically is imperative for the survival of an organism.

3.10 DNMT1 deficiency in embryonic stem cells

In order to model the early embryonic development, cells from the inner cell mass of the blastocyst are cultured *in vitro*, as the so-called embryonic stem cells (ESCs). These cells retain their ability to self-renew and differentiate into the three germ layers that form the embryo: ecto-, meso- and endoderm[88]. Previous efforts established both mouse and human embryonic stem cells in cultures[89,90]. Interestingly, both cultures differ by their morphology, growth conditions and also molecular features such as gene expression and DNA methylation level. Comparative studies between the two cell types characterized mouse ESCs (mESCs) as closely resembling the inner cell mass[91]. In contrast, the human ESCs (hESCs) display the molecular markers of the epiblast stage of development, which proceeds the blastocyst implantation[92]. Reflective of these developmental differences, hESCs have higher levels of DNA methylation than mESCs[93].

Genetic studies depleting the DNA methyltransferases in hESCs and mESCs were vastly different. While the knockout of de novo methyltransferases (DNMT3A and 3B) was tolerated in both cell types[94,95], the removal of the maintenance methyltransferase (DNMT1) resulted in a loss of fitness in hESCs exclusively, although global DNA hypomethylation was observed in both mouse and human cells. The removal of Dnmt1 alone or in combination of Dnmt3A and 3B led to a complete loss of DNA methylation in mESCs without affecting their self-renewal capability[94]. However, the differentiation of

Dnmt1-deficient mouse ESCs was impaired[96]. For example, in methylcellulose hematopoietic progenitor assay (used to assess the potential of ES cells to differentiate into hematopoietic lineage cells) had a low level of erythroid cell differentiation and no myeloid cells were formed[96].

3.11 DNMT1 deficiency in somatic stem cells

Because (1) a direct differentiation of Dnmt1KO mESCs to cell type of interest is impossible and (2) knockout embryos do not develop to the point to form many of the tissues of interest, alternative means were used to obtain somatic cells deficient for Dnmt1.

The commonly used Cre recombinase-based system for inducing gene knockout is bases on the ability of the CRE protein to guide recombination between small DNA sequences called loxP[97]. When two loxP sites are inserted into the genome in the same orientation, the recombination will result in the removal of the sequence between them, which is a potent way to engineer knockouts. Crosses between Dnmt1$^{fl/fl}$ (loxP flanked exons 4 and 5)[98] mice and ones that expressed CRE in a tissue-specific manner (usually by placing it under the control of a tissue-specific gene promoter) enabled studies of Dnmt1KO cells in embryonic fibroblasts[98], pancreas[99], developing intestine[100], myoblasts[101], hematopoietic stem cells (HSCs)[102], neural progenitor cells (NPCs)[103] and both post-mitotic neurons and neuronal precursors[104]. The results of these tissue-specific knockouts of DNMT1 are confounding and often mutually exclusive, therefore only a few general, common themes emerge that will be described below.

Cell cycle arrest followed by apoptosis. Mouse embryonic fibroblasts (MEFs) were reported to undergo cell cycle arrest and apoptosis in culture[98]. Furthermore, this loss of fitness phenotype was accompanied by gene expression deregulation and activation of transposable elements.

Biased differentiation. In contrast, DNMT1 knockout in the more committed myoblasts (progenitors of muscle cells) in mice resulted in the preferential transdifferentiation of these somatic stem cells into osteocyte-like cells (progenitors of bone cells), both *in vivo* and *in vitro*. Trowbridge and colleagues examined the effect of the Dnmt1 ablation on hematopoiesis, by crossing the Dnmt1fl/fl and the Mx-Cre mice[102]. Interestingly, using the Mx-Cre system lead to a knockout in both HSCs and their bone marrow microenvironment, however, no defect in lineage composition or cellularity of the bone

marrow was observed. Hematopoietic stem cell knockout of Dnmt1 followed by transplantation into irradiated wild-type host (normal microenvironment) resulted in a preferential differentiation down the lymphoid lineage instead of myeloid lineage. At the same time, the overall number of HSCs was reduced upon transplantation. This event was not caused by apoptosis, but rather an increased rate of cell division and terminal differentiation. Of note, transcriptional analysis reported the expression of the same transposable elements as those observed in MEFs[98,102].

An analogous situation was observed in neural progenitor cells, which also had a biased differentiation potential and preferentially formed glial, rather than neuronal cells[103]. A more detailed analysis of Dnmt1 knockout in the neural lineage was performed by Fan and colleagues[104]. In agreement with the role of Dnmt1 as a maintenance methyltransferase acting during DNA replication, the depletion of this enzyme in the postmitotic neurons did not have a detrimental effect either *in vitro* or *in vivo*. Using the Nestin-Cre system to specifically knock out Dnmt1 in the brain resulted in either the embryonic lethality or the postnatal death, depending on how efficient the Dnmt11 knockout was performing (95% versus 30%, respectively)[104]. The differentiated neurons did not display apoptotic markers, however, in the mice that survived due to the Dnmt1-KO mosaicism (Cre recombinase worked in only 30% of cells), the Dnmt1-deficient neurons were eliminated from the postnatal brain within 3 weeks of birth[104].

Failure in differentiation. In contrast to the reports from HSC and NPC-specific knockouts, pancreatic progenitor cells fail to differentiate, rather than differentiating precociously. The p53-dependent apoptosis was attributed to the loss of progenitors and haploinsufficiency of trp53 rescued the ability of the cells to form an organ, albeit of a smaller size than in the wild-type littermates[99]. Both premature differentiation and apoptosis were the cause of reduced proliferation of the perinatal intestinal epithelium and consequently the death of most pups before weaning. The surviving crypt cells exhibited differences in the transcriptional program, notably upregulating DNA damage response and cell cycle checkpoint genes[100].

DNMT1 knockdown approach was in turn used to study the effect of maintenance methyltransferase depletion on the keratinocyte proliferation[105]. By expressing a short hairpin (sh)RNA with the sequence complementary to the DNMT1 mRNA, RNA interference (RNAi) machinery is directed to the targeted mRNA for its subsequent degradation that eventually leads to either partial or complete absence of a protein product. This strategy was employed to deplete DNMT1 in human keratinocytes (skin progenitors), followed by engraftment into immunodeficient mice[105]. Loss of the

maintenance methyltransferase resulted in premature differentiation of these progenitors and eventual tissue loss, similarly to what was reported for intestinal cells.

3.12 DNMT1 deficiency in cancer cells

Cancer cells divide rapidly and thus were utilized as a model to study the consequences of DNMT1 depletion. With the RNAi-based methods, previous studies generated cancer cell lines[106,107] displaying higher or lower reduction of DNMT1 protein. These knockdowns were reported to trigger replication stress checkpoint[106], or mismatch repair (MMR) deficiency-like phenotype[107]. A complete knockout of DNMT1 in HCT116 colon cancer cell line resulted in G2/M phase arrest[108].

Overall, these results present a complex repertoire of *in vivo* phenotypes that have been associated with the loss of DNMT1. Molecular mechanism of how a cell responds to and tolerates the deficiency in DNA methylation would likely dissect the underlying causes of such a broad range of observations in the literature.

4. AIMS OF THE STUDY

The aim of this study is to consolidate our understanding of how loss-of-function of the DNA maintenance methyltransferase affects cell homeostasis. Through a combination of next generation sequencing and molecular biology methods, we aim to characterize the general mechanisms of response to the absence of DNMT1. We concentrate on verifying phenotypes reported in the previous studies of DNMT1KO cell lines – both cancer and non-transformed.

Specifically, we aim to address the following:

(1) Characterize how DNMT1-deficiency affects cell growth, by assessing the number of growing cells, the cell cycle length and potential checkpoints, as well as the number of apoptotic cells;

(2) Describe the effect of DNMT1-deficiency on genomic stability, by investigating DNA damage markers, expression of transposable elements and possible defects occurring during mitosis;

(3) Investigate the effects of DNMT1 depletion on the level of DNA methylation and its distribution across the genome and how it affects the transcriptional output of the globally hypomethylated cells.

5. MATERIALS AND METHODS

Cell culture

Human Embryonic Stem Cells HUES64[109] or Human Induced Pluripotent Stem Cells ZIP13K2[110] were cultured in a 5% CO_2, 37°C incubator. The medium consisted of mTeSR1 (STEMCELL Technologies), with a 1:100 supplementation of Revitacell (Thermo scientific) during passaging and the first 24h after plating. The cells were plated on Geltrex (Thermo Scientific) basement membrane matrix-coated plates. The Geltrex stock was diluted 1/10 in Dulbecco Phosphate Buffered Saline (DPBS; Thermo Scientific) and the plates were stored in the incubator for 1h for coating.

The media was changed daily and the cultures were passaged once every 4 days. For passaging, the cells were washed once with DPBS, followed by 3-5minute incubation in 0.5nM EDTA in DPBS solution in the incubator. Cells were collected and washed once with mTeSR1 media and re-plated in small clumps at a 1:8 dilution. For freezing, cells were collected in the same way, re-suspended in mTeSR1, 1:100 Revitacell, 10% dimethyl sulfoxide (DMSO; Sigma-Aldrich), transferred to cryo vials (Thermo scientific) and stored overnight at -80°C in a Mr. Frosty (Thermo) ispopropanol freezing box. The vials were stored in liquid nitrogen tank.

Cell counting

For cell count experiments, the cells were collected using Accutase (STEMCELL Technologies), according to manufacturer's instructions. Single cells were plated at 40 000/well in a 12-well plate in triplicates and collected daily with the same method and for subsequent passage, another 40 000cells/well in a 12well plate were seeded. For counting, 10µL of cell suspension was mixed with Countess Trypan Blue Stain (0.4%, Thermo Scientific), loaded onto countess slides (Thermo Scientific) and counted on a Countess Automated Cell counter. The number of living cells was noted.

The number of cells were plotted as "surviving cells with respect to control", where at each time point the percentage of cells was calculated according to the formula:

$$surviving\ cells(y) = \frac{\#cells\ in\ the\ treated\ condition \times 100\%}{\#of\ cells\ in\ the\ control\ condition}$$

Western Blot

Nuclear protein extraction protocol was performed on live cells, briefly, we rinsed a confluent 10cm plate 2x with PBS and scraped into a 15mL Falcon. Spun @1200rpm for 3min. Re-suspended in 7mL ice cold buffer A (+fresh PI and DTT). Spun @1500rpm for 5min. Re-suspended the pellet in 500µL ice cold buffer C (+fresh PI and DTT). Spun @3000g for 5min. For WB re-suspended in 500µL RIPA (+PI and DTT, also DNase I and RNase A). Vortex 20min @4°C. Spun @max speed for 15min, moved the supernatant to a new tube and discarded the pellet. Added 83.3uL 4x LDS sample buffer (Thermo) to the supernatant. Denatured the samples @70°C for 10min and stored them at -20 to -80°C, or used immediately. If frozen, we first thawed the sample and heated it up to 40°C. Vortexed briefly before running on a NuPAGE gel (Thermo). For immunoblotting, we resolved 7.5 – 15uL sample in the XCell SureLock Mini-Cell (Thermo) using 4-12% BisTris gels (Thermo) and MOPS buffer (Thermo) according to the manufacturer's instructions. The protein was then transferred into nitrocellulose membrane using the iBlot 2 blotting system with pre-made transfer sandwiches (Thermo). The membrane was then blocked in 5% blotting grade blocker (Biorad) w/v solution in 1x Tris Buffered Saline – Tween-20 (TBS-T, Thermo). The same solution was used for staining with antibodies (overnight at 4°C): anti-DNMT1 (Chromotek E8; 1:500), anti-tubulin (Hybridoma bank 12G10; 1:10 000). The blots were washed 5 times with TBS-T and then incubated with secondary antibodies: anti-Rat Horse-Raddish Peroxidase (HRP; Thermo 1:10 000) and anti-mouse HRP (Thermo; 1:10 000). The blots were visualized using SuperSignal kit as per manufacturer's instructions (BioRad), followed by imaging in BioRad ChemiDoc on automatic settings. HiMark Protein standard (Thermo) was used to estimate the height of the bands.

Buffers used consisted of:

Buffer A	
25mM HEPES pH 7.6	**Buffer C**
5mM MgCl2	**10mM HEPES pH 7.6**
25mM KCl	**3mM MgCl2**
0.05mM EDTA	**100mM KCl**
10% Glycerol	**0.01mM EDTA**
0.1% IGEPAL	**10% Glycerol**

RIPA
50mM Tris pH 8
200mM NaCl
5% Glycerol
0.5% Triton X-100
0.1% Sodium Deoxycholate

Table 1. Nuclear extraction buffer composition.

The ready buffers were filtered through 0.2µm filter and stored at 4°C

Added to all buffers right before use: 1xRoche PI (25x stock in dH2O; Sigma), 1mM DTT (1M stock in dH2O; Sigma) and optionally 1mM PMSF (100mM stock in isopropanol; Sigma). Added to RIPA buffer right before use: 1:1000 DNaseI (NEB, optional), 1uL RNase A (Qiagen).

Genomic DNA extraction

Genomic DNA was extracted using the Qiagen blood and tissue sample kit (Qiagen), according to manufacturer's instructions.

5mC ELISA

Purified genomic DNA was diluted to 20ng/µL and 100µL of each sample was used per replicate in 5mC ELISA (Zymo). The procedure was performed according to manufacturer's instructions, with one modification: the DNA was denatured for 20 minutes, instead of 5 minutes. The plates were resolved in CyTation5 Plate Reader (Biotek).

EdU stain

The cells were stained using the Click-iT EdU Stain Kit, (protocol for attached cells; Thermo) according to manufacturer's instructions, with modifications: the growing cells were incubated with half of the recommended EdU dose dissolved in basal DMEM (Gibco) medium for 30 minutes, followed by immediate fixation with 4% Paraformaldehyde (PFA; Thermo) in PBS. Stained cells were imaged using Cell

Discoverer (Zeiss) microscope and the number of EdU positive and total cells were automatically counted using ZEN software (Zeiss).

TUNEL stain

The cells were stained using the In Situ Cell Death Detection Kit, Fluorescein (Sigma), according to manufacturer's instructions. Stained cells were imaged using Cell Discoverer (Zeiss) microscope and the number of TUNEL-positive and total cells were automatically counted using ZEN software (Zeiss).

CFSE stain

The cells were detached using Accutase (described above) and incubated for 30 minutes with the CellTrace CFSE reagent dissolved in basal DMEM medium in the incubator, according to manufacturer's instructions. The cells were then washed and plated normally, followed by detachment using Accutase and fixation with RNAprotect Cell Reagent (Qiagen). The cells were stored in the fridge until all the samples were collected. The cells were then stained with Hoechst 33342. The green and blue fluorescent cells were gated for and counted. The absolute fluorescence of CFSE was plotted using FlowJo software (FlowJo LLC).

Hoechst 33342/DAPI stain

Either DAPI (1mg/mL; Thermo) or Hoechst solution (10mg/mL; Thermo) were dissolved at 1:10 000 in PBS. A 10-minute incubation on a rotating platform in the darkness was followed by FACS analysis on FACS Aria II (BD Biosciences); or a single DPBS wash and fluorescence microscopy.

Immunofluorescence staining

The cells were imaged on the plates they were grown in. For confocal microscopy, we used ibidi polymer coated 60mm thin bottom dishes (Ibidi). The cells were fixed with 4% PFA, followed by 2 washes in DPBS. The cells were then permeabilized using 0.5% Triton X in PBS solution for 10 minutes, followed by blocking in a 5% Bovine Serum Albumin (BSA; Sigma), 0.2% Triton X in PBS for 30 minutes on a rotating platform, at room temperature. The same blocking solution was used to immunostain the cells with the following antibodies overnight, at 4°C:

Antigen	Concentration	Company	Catalog number
Mre11	1:500	NOVUS	NB100-1425s
BLM	1:250	Santa Cruz	365753
53BP1	1:250	Bethyl	A300-272A
T	1:500	R&D	AF2085
SOX17	1:500	R&D	MAB1924
Nestin	1:1000	R&D	MAB1259
tubulin	1:500	Sigma	T9026

Table 2. Primary antibodies used in immunofluorescence microscopy.

The samples were washed using 5% Bovine Serum Albumin (BSA; Sigma), 0.2% Triton X in PBS, 3 times each. The same buffer was used to incubate the cells with secondary antibodies:

Antigen	Concentration	Company
Mouse IgG; Alexa488 conjugated	1:1000	Thermo
Rabbit IgG; Alexa488 conjugated	1:1000	Thermo
Goat IgG; Alexa488 conjugated	1:1000	Thermo
Goat IgG; Alexa594 conjugated	1:1000	Jackson ImmunoResearch

Table 3. Secondary antibodies used in the study.

After 3 additional washed, the cells were stained with DAPI (described above) and imaged using CellDiscoverer microscope (Zeiss) for fluorescence microscopy or LSM800 with Airyscan (Zeiss) for confocal microscopy. Representative images were pseudocolored, merged and cropped using ImageJ (NIH).

Regular brightfield images were obtained using Nikon Eclipse Ts2 benchtop microscope (Nikon).

Karyotyping

G-band karyotyping was performed by Cell Line Genetics (CLG). 20 cells per line were imaged and representative karyotypes were visualized in a karyogram.

Cloning

Plasmids were constructed by restriction enzyme digestion of a donor plasmid. 10nM each forward and reverse primers were used for PCR-amplifying desired DNA fragments (Phusion High-fidelity Master Mix GC, NEB), purifying both DNA fragments in Tris-Acetate-EDTA (TAE) agarose gel. Bands of the desired size were excised from the gel and purified using Qiagen Gel extraction kit (Qiagen). The fragments were recombined together using Infusion cloning kit (Takara) according to manufacturer's instructions. Stellar competent bacteria (Takara) were used for plasmid transformation, according to manufacturer's instructions. In order to confirm correct insertion of the fragments, the transformed bacteria were plated on agarose plates consisting of LB and Ampicillin (100ng/mL). Resistant colonies were incubated in LB broth overnight on a bacteria shaker at 37°C and subsequently purified using Qiagen Miniprep kit. Purified recombinant plasmids were Sanger-sequenced by Eurofins genomics. The correct sequence of the plasmid was verified using Benchling.

For transfection-quality purification, 50mL cultures of bacteria were grown overnight on a bacterial shaker at 37°C and purified using the Midiprep kit (Qiagen) according to manufacturer's instructions.

The knockout lines presented in this study were performed using CRISPR/Cas9 technology[111]. Guide RNAs: lentiCRISPRv2 was a gift from Feng Zhang (Addgene plasmid # 52961 ; http://n2t.net/addgene:52961 ; RRID:Addgene_52961)[112]. We modified the plasmid by substituting Puromycin resistance gene with a Blasticidin resistance gene (BLA) sequence ordered from IDT.

The design of guide RNAs used in this study was performed in Benchling (www.benchling.com) and the cloning was performed according to the instructions from the GeCKO protocol (http://genome-engineering.org/gecko/wp-content/uploads/2013/12/lentiCRISPRv2-and-lentiGuide-oligo-cloning-protocol.pdf)[112].

Target	gRNA sequence
TP53	GGATGATTTGATGCTGTCCC
LEFTY1	CTGTGGCTCTGCTGGGCACTC

Table 4. Guide RNAs used in this study.

The pLV-TRE3G-UCOE-GFP-WPRE plasmid was a kind gift from the Lindquist lab. NODAL or LEFTY1 cDNA sequences were ordered from IDT and inserted into the plasmid downstream of the promoter site as described above.

Viral transduction

HEK 293T cells were grown in a Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12 (DMEM/F12) basal medium (Gibco), supplemented with 10% Fetal Bovine Serum (FBS; Gibco), 1x Penicillin-Streptomycin (STEMCELL Technologies) and 1x GlutaMax (Thermo). To make lentiviruses, these HEK cells were plated at $3.2x10^6$ cells in a 10cm dish, and next day transfected with 30µL Mirus LT1 reagent (Mirus), 5µg donor plasmid, 3.75µg psPAX2 plasmid (psPAX2 was a gift from Didier Trono (Addgene plasmid # 12260 ; http://n2t.net/addgene:12260 ; RRID:Addgene_12260), 1.5µg VSV-G (pCMV-VSV-G was a gift from Bob Weinberg (Addgene plasmid # 8454 ; http://n2t.net/addgene:8454 ; RRID:Addgene_8454) in 1mL DMEM basal medium.

The next day the cells were fed 20% FBS medium (otherwise the same recipe as above) and incubated for 48-72h. The medium from the cells was harvested, centrifuged to get rid of the pellet of dead cells and mixed with LentiX concentrator (Takara) in a 3:1 ratio. The mix was incubated at 4°C overnight and centrifuged for 45min at 1500g in cooled centrifuge. The pellet containing lentivirus was re-suspended in 2mL DMEM, aliquoted and either used immediately or stored at -80°C.

For transduction, hESCs were seeded in 6well plates at a regular density. On the following day they were incubated with 1mL mTeSR1, followed by the addition of 5µg/mL Polybrene (Santa Cruz Biotechnology), between 5 and 25µL concentrated virus re-suspended in another 1mL of mTeSR1. After 24h of incubation with the virus mix, the media was changed to regular mTeSR and, following 2-4days of recovery, the cells were selected for virus integration.

Antibiotic selection

Cells recovered after viral transduction were plated sparsely on 10cm dishes in the presence of 4µg/mL Blasticidin. The drug was supplied until a control plate of WT cells treated in the same way has perished (usually within 4 days). Surviving colonies were grown for 10 days and manually scraped off and moved to a well in a 96well plate with a P200 pipette. The content of each well was then divided into two 96-well plates and grown until confluency. One plate was expanded and maintained, while the other served for genotyping.

Genotyping

Confluent cells were lysed using Direct PCR Lysis Reagent (Viagen) according to manufacturer’s instructions. 1µL of lysate was used for a genotyping PCR ran with the

Phusion GC master mix (NEB), the PCR products were resolved on a 1% TAE agarose gel, excised and purified using Qiagen gel extraction kit and cloned into TOPO vector using a Blunt-end TOPO kit (Thermo), according to manufacturer's instructions. The bacteria were selected and genotyped as described above.

Genotyping primer	Sequence
TP53_forward	CGTCGAGCCCCCTCTGAGTCAGG
TP53_reverse	GGCCAGGCATTGAAGTCTCATGGAAGC
Lefty1_forward	AGGCTATAAAGCTGCCCAGGCTTG
Lefty1_reverse	CTTAGACCGTGGCCCTCACTCAGC

Table 5. Genotyping primers used in this study.

RNA extraction

One well of a 6well plate was used per extraction. The cells were lysed in 1mL TRIzol reagent (Thermo), each milliliter of Trizol was mixed with 200µL of Chloroform (Sigma) and transferred into a heavy phase lock tube (VWR). After 15 seconds shaking, the mixture was incubated for 5min at room temperature to allow phase separation. The tubes were centrifuged at 12000g 4°C centrifuge for 15 minutes. The top aqueous phase was mixed with 1mL isopropanol, incubated at -20°C for an hour and centrifuged at maximum speed in a 4°C centrifuge for 30min. The supernatant was discarded and the pellet was washed once with 75% ethanol, followed by a wash with pure ethanol. After air-drying the pellet, the RNA was re-suspended in pure water (RNase free) and the concentration was measured using Nanodrop (Thermo).

Reverse transcription

1µg RNA was first subjected to a DNaseI digest using amplification grade DNaseI (Invitrogen), according to the manufacturer's instructions. The reaction was then used directly for first strand synthesis using RevertAid kit (Thermo), according to the manufacturer's instructions. Random hexamers were used for non-exon spanning primers (such as repeats), or oligo dT for exon-spanning primers (coding genes). The lack of DNA product was confirmed with a –RT (no reverse transcriptase enzyme) negative control reaction for each sample.

qRT-PCR

1µL RT reaction was mixed with 10nM each forward and reverse primer, 2x SYBR Green Master Mix (Thermo) and water, up to a total volume of 11µL. The reactions were pipetted into 384-well qPCR plates and ran on a ViiA 7 System machine (Thermo), using default settings. The results were analyzed in Excel normalized to the internal housekeeping gene control $x = average\ value\ of\ \frac{2\wedge(-Ct\ sample)}{2\wedge(-Ct\ housekeeping)}$ for n=2 technical replicates.

PRIMER TARGET	SEQUENCE	REFERENCE
GAPDH	GCACCGTCAAGGCTGAGAAC	[113]
	AGGGATCTCGCTCCTGGAA	
NODAL	agacatcatccgcagccta	[114]
	caaaagcaaacgtccagttct	
LEFTY1	ctgcacaccctggacctt	[114]
	atcccctgcaggtcaatgta	
LEFTY2	cctggacctcagggactatg	[114]
	atcccctgcaggtcaatgta	
CER1	gccatgaagtacattgggaga	[114]
	cacagccttcgtgggttatag	
NEUROG2	GAC ATT CCC GGA CAC ACA C	[115]
	TAC CTC CTC TTC CTC CTT CA	
NF1B	GACATGAACTCGGGGGTCAATCTT	[113]
	GTAGTCGGAGAAGACATATCTTGATC	
SOX9	CGAGCACTCGGGGCAATCC	[113]
	CTGCCCCCCTCTGGCAAG	
GFAP	GGTTGAGAGGGACAATCTGGCACA	[116]
	CTATAGGCAGCCAGGTTGTTCTCGG	
18S RRNA	GTAACCCGTTGAACCCCATT	[117]
	CCATCCAATCGGTAGTAGCG	
ALPHA SATELLITE	AAGGTCAATGGCAGAAAAGAA	[118]
	CAACGAAGGCCACAAGATGTC	
L1	GCTGGATATGAAATTCTGGGTTGA	[118]
	AGGAAATACAGAGAACGCCACAA	
HERVK	AAATAAGACCCAACCGCCAGTAGC	[117]

	GAATTGCCATGCCTCAGTATCTCC

Table 6. qRT-PCR primers used in this study.

ScoreCard

The Scorecard (Thermo) experiment was performed as per manufacturer's instructions, with modifications in the Reverse Transcription stage. We used 1.5µg RNA reverse-transcribed according to the RevertAid kit (Thermo) instructions, in double the recommended volume. The reactions were then diluted with miliQ water and processed according to the ScoreCard manual. The score was calculated by the SCORE web browser app (Thermo).

NPC differentiation

80% confluent hESCs were detached from the plate using Dispase (Thermo) for about 20min. The colonies were washed in DMEM/F12 basal medium and carefully transferred to a low attachment 6-well plate (Corning) and cultured in media containing N2 supplement (Thermo), B-27-RA (no retinoic acid) supplement (Thermo) in DMEM/F12 medium, supplemented freshly with 100nM LDN193189 (Sigma) and 10µM SB431542. Cells were cultured in suspension to form Embryoid Bodies and their media was carefully changed every second day. After 7 days, the EBs were plated on regular culture plates coated for 3h with matrix consisting of Poly-ornithine (20µg/mL; Sigma) and Laminin (5µg/mL; Thermo) and cultured in the same medium for another 7-10 days until neural rosettes formed. These were collected with neural rosette medium (STEMCELL Technologies). The rosettes were seeded again on Polyornithine-Laminin coated plates in NPC medium: the same DMEM/F12; N2; B27-RA base, supplemented with 20ng/mL basal FGF (STEMCELL) and 1µg/mL laminin. After a week, the cells differentiated into NPCs and were grown in the NPC medium on 83.3µg/mL Matrigel (Corning) re-suspended in ice cold DMEM/F12 matrix (coating 1h, double washing before re-plating cells). The NPCs were passaged every week at a 1:3 ratio.

Small molecules, antibiotics, cytokines and morphogens

The cells were incubated with the following chemicals, which were diluted as per manufacturer's instructions to make stock solutions stored at -20 to -80°C (the dose and the treatment length for each experiment are indicated in the Methods and Results sections):

Name	Company
Doxycycline	Sigma
dTAG-13	Torcis
5-AzaC	Sigma
Camptothecin	Selleckchem
Zeocin	Thermo
Blasticidin	Thermo
Doxorubicin	Sigma
SB431542	STEMCELL
LEFTYB	Millipore
NODAL	R&D
WNT3A	R&D
LIF (mouse)	Thermo
BMP4	Sigma
LDN193189	Sigma
bFGF	STEMCELL

Table 7. Small molecules, morphogens and cytokins used in the study.

Figures

Figures were generated and assembled using GraphPad Prism 7 (Graphpad), Adobe Illustrator 2017 (Adobe) and ImageJ (NIH). The tables were generated in Excel (Microsoft).

Dual protocol scRNAseq+RRBS

The protocol was designed by Hongcang Gu, Andreas Gnirke and Arman Mohammad, who also prepared the libraries. It is based on G&Tseq with minor modifications[119]. Full protocol can be found in supplementary material, along with expression and methylation data. Briefly:

The cells were detached from the plate using accutase, followed by fixation in RNA Protect buffer (Qiagen). The cells were then stored in the fridge and FACS sorted together into 96well plates containing 15 µl of RLT plus buffer (Qiagen) and 1 U μl^{-1} of SUPERase·In RNase inhibitor (Thermo). After sorting and freezing β-mercaptoethanol (Sigma), was added to the sample and the suspension was transferred into 96-well DNA

LoBind plate (Eppendorf). M-280 streptavidin beads conjugated to oligo dT reverse transcription primer were added to each sample, followed by incubation at 72 °C for 3 min and at room temperature for 25 min on a rotating platform. The mRNA was separated by a DynaMag-96 Side Magnet (Thermo) and to reverse transcribed as described in the Smart-seq2 protocol[119]. The genomic DNA was stored in a fresh LoBind plate. Following reverse transcription, the cDNA was amplified and an RNA-seq library was generated as described before[120]. The libraries were pooled and sequenced by an Illumina Hiseq2500 sequencer.

Agencourt AMPure beads (Beckman Coulter) were used to purify genomic DNA, which was eluted with 15 µl of low Tris–EDTA buffer. For the RRBS library generation, we used the CutSmart buffer (NEB) for the enzymatic reactions: MspI digestion, end-repair/A-tailing and DNA ligation using T4 DNA ligase (NEB), without changing the solution between steps to avoid sample loss.

The genomic DNA was set up for an 80min digestion reaction using 16U MspI (New England Biolabs) at 37 °C, followed by incubation at 65 °C for 15 min to inactivate the enzyme. The DNA ends were repaired and A-tailed using Klenow fragment (3’ to 5’ exonuclease) (New England Biolabs) reaction, supplemented with 0.3 mM dATP and 0.03 mM each dCTP and dGTP at 30 °C for 25 min, followed by incubation at 37 °C for 25 min, and heat inactivation at 70 °C for 10 min.

The A-tailed DNA fragments were then ligated with 7nM indexed adapters overnight at 16 °C, using T4 DNA ligase, followed by heat inactivation at 65 °C for 15 min. The libraries were pooled, and the adaptor dimers were removed using AMPure beads. The tagged DNA was eluted in 30µl of low Tris-EDTA buffer.

Qiagen EpiTect Fast Bisulfite Conversion Kit was used according to manufacturer’s instructions (protocol marked with *), with a minor modification. We extended the bisulfite conversion time to 2 cycles of 20 min each, maximize the bisulfite conversion rates to over 99%. The bisulfite-converted DNA fragments were PCR amplified and the library DNA was purified using AMPure beads. The RRBS libraries were sequenced for 2 × 100 cycles (paired-end).

Quality control

The QC and read alignment were performed by Ayush Raman, Caleb Laureau and Martin Aryee. The scRNAseq data was processed as described before[120]. The scRRBS data was processed using a cloud-based computational pipeline[121]. Briefly, the quality of raw reads was assessed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc ; Andrews, 2010). The average per-base quality score across the sequencing samples was satisfactory to pass the QC. Trim galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) was used to trim low-quality bases from the ends of the reads. The reads were aligned to the human hg38 reference genome using Bismark2[122]. Duplicated reads were removed, and subsequently the number of covered CpGs was calculated.

The estimate of methylation from the ratio between methylated and unmethylated alleles calculated across all the CpGs in the genome was summarized in a so-called ß-matrix. In this method, a ß-value per CpG is between 0 (unmethylated) and 1 (fully methylated).

The data set was then tested for biases using *scmeth* package in R environment[121]. Low-quality samples were identified based on alignment and discarded, average methylation per base across the reads (M-bias, Citation: Vermunt et al. Cell Reports 2014) was estimated, followed by calculating CpG and mean methylation across samples. Sufficient sequencing depth was ensured by plotting the downsampling saturation curve (number of CpGs as a function of read number), indicating that the probability of sequencing a new CpG plateaued with the number of reads we covered.

The replicates of the cells from the same time point (i.e., same day) were pooled together by taking the mean of the ß-values. Next, the genome was divided into 1-kb non-overlapping tiles and averaged ß-value per tile was calculated. All the tiles with no signal (i.e., ß-value and coverage equal to 0) were removed.

scRRBS data was analysed by Dr Jocelyn Charlton using R studio and Bioconductor packages[123], while the scRNAseq and combined analyses were performed by Dr Sudhir Thakurela using R studio and Seurat package[124]. The DEG analysis, GO term analysis[125,126], gene pools overlap analysis (using Venny: https://bioinfogp.cnb.csic.es/tools/venny/) and plotting the expression of particular genes (using Graphpad Prism), as well as data interpretation was performed by myself.

Bulk RNA-seq

The libraries were prepared by the MPI-MG Berlin sequencing facility, briefly: RNA libraries were prepared using the TrueSeq RNA Sample Prep v2 HS Protocol (Illumina), followed by 50 bp paired-end sequencing on the NextSeq500 Sequencing System (Illumina). The QC and alignment were performed by Dr Jocelyn Charlton. FastQ raw reads were trimmed using Cutadapt[127]. The reads were then aligned using STAR aligner[128] (to the hg19 human genome assembly) and Stringtie v1.3[129] was used for transcript assembly.

6. RESULTS

6.1 Genetic approach to study the consequences of the disruption of the maintenance of DNA methylation.

We set out to study DNMT1 using a genetic approach considering the following criteria:

1. As DNMT1 acts alongside replication, its loss-of-function phenotype was almost exclusively reported in mitotic cells. We thus reasoned that the culture needed to be replicating to allow the passive loss of methylation from the DNA, to recapitulate the most severe phenotypes of DNMT1KO studied before. In fact, many of the cases of DNMT1-deficiency published up to date resulted in a loss of fitness[76,98,102,106].
2. To study the precise cause for this, we would also employ a model with a strong and penetrant phenotype arising from the depletion of DNMT1.
3. Lastly, we were interested in cells that are genetically stable. Previous, molecular biology studies often utilize cancer cell lines, such as U2OS or HCT116, due to their easy culturing, rapid growth and suitability for nucleic acid delivery, which renders them permissive for genetic manipulation[130]. However, due to their genomic instability and abnormal global DNA methylation pattern, cancer cell lines could potentially yield confounding results in our study. Therefore, we decided to utilize a non-tumor-derived line instead. This will enable the distinction between an intrinsic cell line instability and the consequences of DNMT1 absence as has been reported for some of the cases of DNMT1 depletion *in vitro*[87,106,108,131].

Given the listed criteria, an attractive alternative to cancer cells were embryonic stem cells (ESCs). This cell type, derived from the inner cell mass (ICM) of a developing mammalian embryo, is characterized by a virtually indefinite self-renewal potential. Importantly for our studies, ESCs remain genetically stable and euploid in cell culture conditions over multiple passages. Previous studies demonstrated that mouse ESCs sustain normal cell proliferation in the absence of DNMT1 and *de novo* methyltransferases DNMT3A and 3B[94]. Unlike the mouse model, DNMT1 is essential for the proliferation of human ESCs[95]. Although both cell lines were derived from the blastocyst inner cell mass of the respective species, human ESCs are believed to represent a later stage of development called the epiblast, based on their growth condition requirements, X chromosome inactivation and gene expression pattern[91,132,133].

Because of these differences, we chose human embryonic stem cells as a suitable model for studying DNMT1.

Since the growth of hESCs is incompatible with the absence of DNMT1[95], we employed an inducible depletion strategy instead of conventional knockout strategies. In this study, we used two complementary approaches to dissect the consequences of DNMT1. Either strategy eliminated technical biases that were confounding factors in previous studies. In order to ensure consistency in our methodology, we have cultured the both cell lines in identical conditions in a complete human pluripotent stem cell medium (mTeSR) on membrane matrix (Geltrex).

CHAPTER 1. Characterization of tools to acutely remove DNMT1.

6.2 DNMT1 depletion on the transcript level with tetracycline-inducible system

The generation of the cell line has been described in ref. 7, where our group employed a Tet-off system[134] to control the expression of DNMT1. Briefly, the cell line was created using lentiviruses to ectopically integrate a wild-type DNMT1 cDNA under the control of Tet-response element (TRE) and a tetracycline transactivatior (tTA). The binding of tTA to TRE allows for the expression of DNMT1 from a separately delivered exogenous allele in the absence of tetracycline analog doxycycline (dox). The endogenous alleles of DNMT1 were subsequently knocked out (Fig. 2A). The resulting cell line expresses a functional DNMT1 only from the transgenic allele, which is not transcribed in the presence of doxycycline (Fig. 2B). We will further refer to this line as “TetOFF DNMT1”.

To independently validate the system for this study, we first assayed its performance. To this end, we checked the kinetics of the DNMT1 protein depletion using western blot (Fig 4A, left panel). We extracted the nuclear proteins from the control and DNMT1-depleted cells. The protein lysates were separated on an SDS-polyacrylamide gel, transferred onto a nitrocellulose membrane and incubated with a monoclonal antibody specific to human DNMT1. We observed a nearly complete depletion of DNMT1 already after 48h post transcriptional inhibition and the protein is not detectable after 72h (FIG 4A).

Due to the differences in bioethics regulations between the USA and Germany, the originally used HUES64-based line had to be substituted with a human induced pluripotent cell (hiPSC) line. These cells are derived by reprogramming of human

fibroblasts[135] and were found nearly identical to hESCs in terms of morphology, gene expression, DNA methylation and histone modification landscape[136,137](Fig 4B).

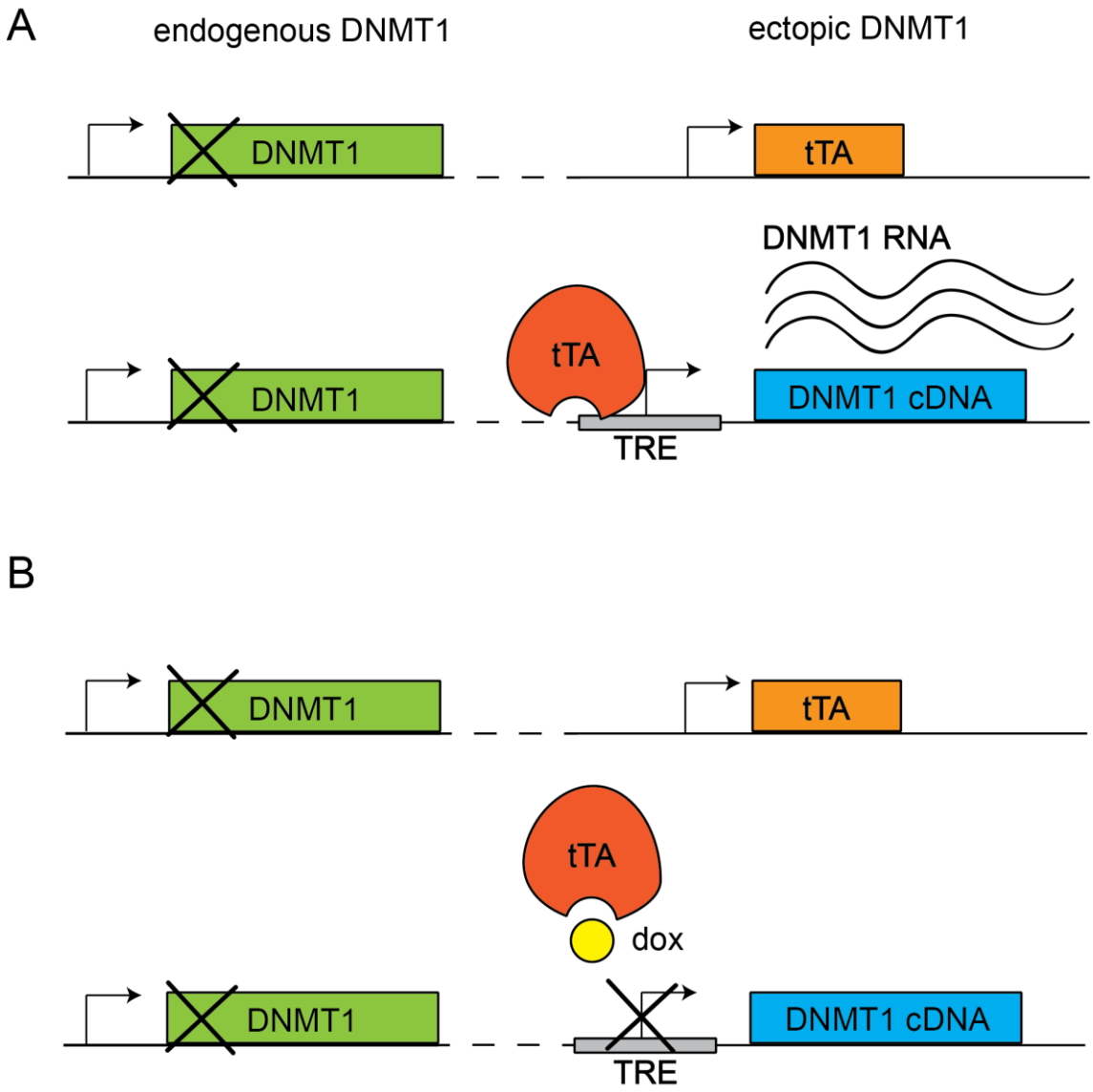


Figure 2. TetOFF DNMT1 inducible depletion system. ***A.*** *In the absence of endogenous DNMT1 (crossed out green bar), the source of functional DNMT1 transcript (black lines) in the cells is an exogenous allele (DNMT1 cDNA – blue box), under the control of TRE promoter (grey bar). It is activated by tTA (red shape), the product of exogenous tTA gene (orange bar).* ***B.*** *In the presence of doxycycline (yellow circle), tTA is unable to the TRE promoter and the expression of transgenic DNMT1 is shut down.*

6.3 DNMT1 depletion on the protein level with PROteolysis TArgeting Chimera (PROTAC)

In the final stages of this work, we switched to using a hiPSC cell line (created by Simon Lauer), which utilizes an alternative approach to deplete DNMT1 at the protein level (FIG 3). We employed a proteolysis targeting chimera (PROTAC) based technology[138], whereby a heterofunctional, small molecule is used to recruit the ubiquitin ligase complex to the protein of interest. The resulting ubiquitinated product is subjected to proteasomal degradation. Because there was no molecule that targets DNMT1 specifically, we used a strategy described by Nabet et al[139]. Briefly, the gene of interest is endogenously tagged with a variant of a chaperone protein (mutFKBP), which is recognized by a highly specific PROTAC molecule called dTAG-13. In the absence of dTAG-13, the stability of the mutFKBP-tagged protein is not affected. However, upon addition of dTAG-13 to the media, both mutFKBP and the Cereblon (CRBN) E3 ligase complex directly bind, thereby facilitating ubiquitination of the tagged protein. Due to its high specificity for mutFKBP (and not the ubiquitously expressed wild-type FKBP), the system does not have off-target effects and has been validated for a variety of targets[139,140].
In our study, this protocol was further modified to knock in the mutFKBP linked to mCerulean fluorescent protein into both endogenous DNMT1 alleles. In our study, this protocol was further modified to knock in the cDNA encoding the mutFKBP linked to mCerulean into the both alleles of the DNMT1 gene.

By default, the TetOFF DNMT1 was used for the majority of the expreiments presented here. The experiments involving degDNMT1 have a suitable annotation in the figure legends.

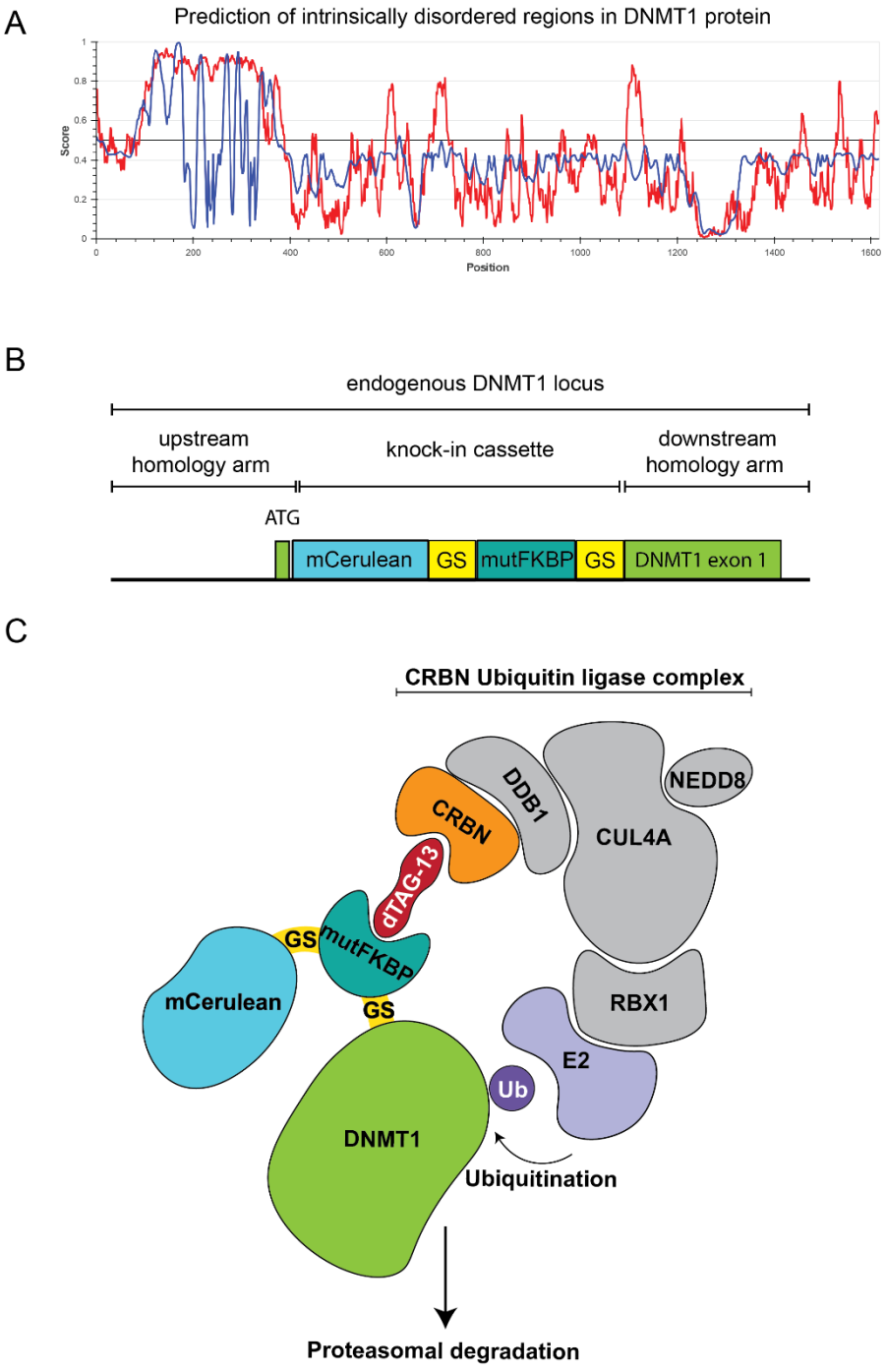


Figure 3. PROTAC-based DNMT1 inducible depletion system. ***A****. XY plot summarizing a bioinformatic analysis of the DNMT1 protein predicted structural organization using two programs: IUpred2 (red line) and ANCHOR (blue line). Higher score represents less intrinsic structure.* ***B****. Schematic representation of donor sequence used to endogenously target the DNMT1 gene for the creation of the degDNMT1 fusion gene. mCerulean (blue), GS linker (yellow), mutFKBP (dark green), endogenous DNMT1 (light green).* ***C.*** *The DNMT1 protein (light green) has been endogenously tagged with mutFKBP (dark green) and mCerulean (blue). The domains are linked by a flexible GS linker (yellow). The addition of dTAG-13 small molecule (red) facilitates the recruitment of Cereblon E3 ligase complex (scaffold shown in grey) by directly binding CRBN (orange). The E2 ligase (violet) ligates ubiquitin (purple) to the DNMT1 chimeric protein, thus targeting it for proteasomal degradation.*

Bioinformatic analysis of amino acid sequence composition of DNMT1. In order to choose a suitable terminus to tag DNMT1, we performed a bioinformatic analysis of the amino acids composition and the domain organization of the DNMT1 protein (FIG 3A), to ensure that the tag does not impair its function. The most suitable place for the tag on a protein is at the terminus which is intrinsically unstructured with no functionally important fold or domain in its vicinity. Our analysis using the Prediction of Intrinsically Unstructured Proteins tool[141,142] revealed that these criteria are met for the N terminus of DNMT1 (FIG 3A).

Indeed, based on previous studies, which tagged either the endogenous wild-type or ectopically expressed DNMT1 with GFP, the N-terminal fusions do not affect DNMT1 stability or localization[143]. This is consistent with the structure of DNMT1 protein, which is highly organized towards the C-terminal end, due to the presence of the structured methyltransferase domain.

Validation of the PROTAC-based strategy. Our N-terminal tagging approach resulted in the transcription of a fusion mutFKBP-mCerulean-DNMT1 protein referred to here as “degDNMT1” (FIG 3A,B). Such mutFKBP-mCerulean-DNMT1 chimeric protein enables depletion at the protein level after the addition of small molecule ligand dTAG-13. We validated the depletion system by monitoring protein level using a western blot assay. In contrast to the TetOFF system, the PROTAC-based depletion cell line underwent a rapid DNMT1 degradation within 24h post 125nM dTAG-13 addition and the protein was not detectable after 48h (Fig 4A, right panel). Consistent with our expectations, the direct degradation of DNMT1 at the protein level elicited faster depletion of the protein than controlling the transcription of the DNMT1 gene.

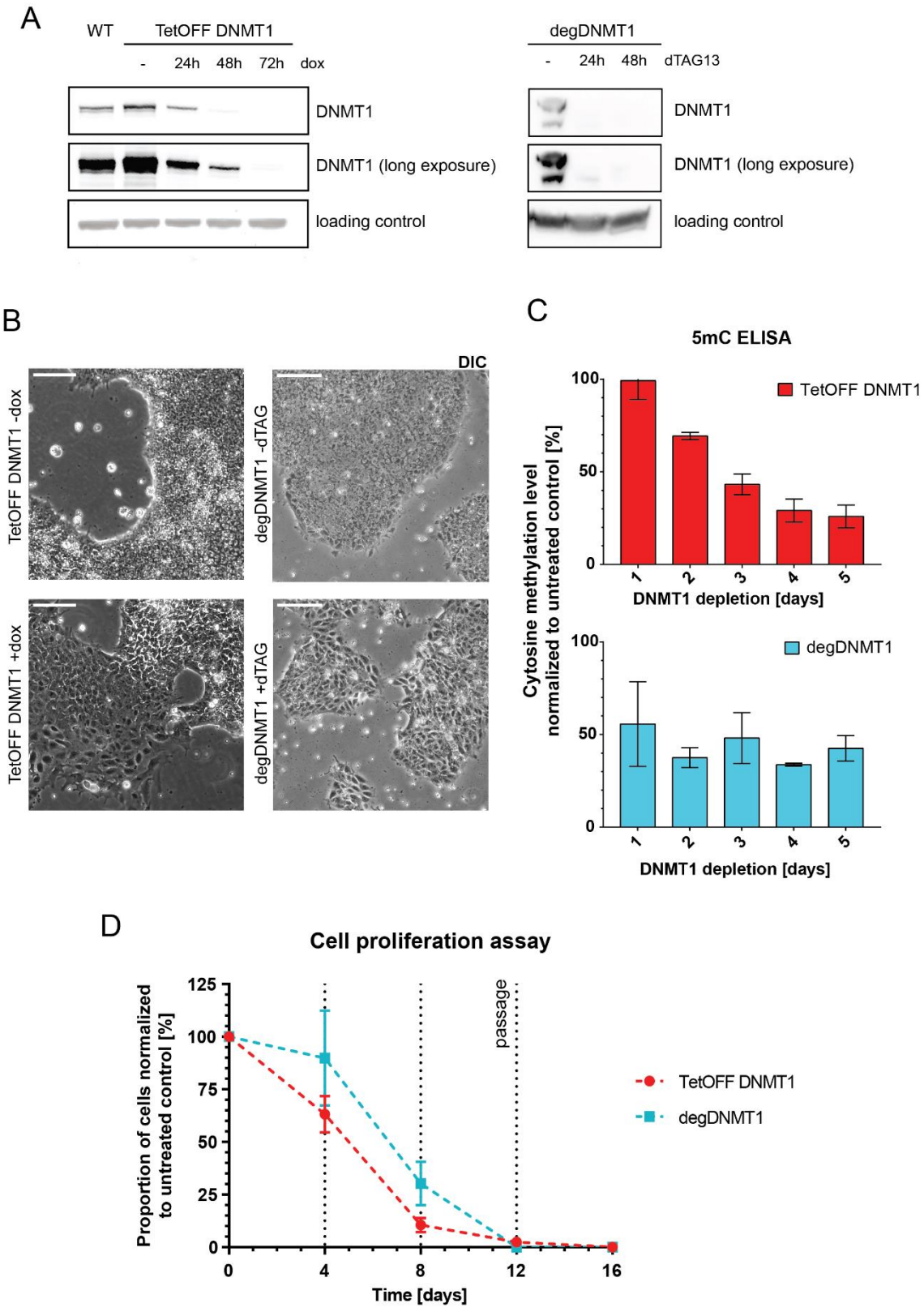


Figure 4. Molecular characterization of the inducible DNMT1 depletion systems. ***A.*** *Western blot showing the presence of DNMT1 under normal and long exposure. TetOFF DNMT1 (left panel) and degDNMT1 (right panel) were treated with doxycycline (2 µg/mL) and dTAG-13 (125 nM), respectively, until DNMT1 protein was no longer detectable.* ***B.*** *Representative brightfield images of the control (top panels) and small molecule treated (bottom panels) TetOFF DNMT1 and degDNMT1 cell lines. DNMT1 was depleted in*

each of the lines for 5days. Scale bar: 100 µm. ***C.*** *5mC ELISA results depicting the DNA methylation retained by doxycycline-treated TetOFF DNMT1 (red, top panel) or dTAG-13 treated degDNMT1 (blue, bottom panel) cell lines. The value is expressed as the proportion of 5mC compared to the untreated, control line. Error bars: SD. N=3* ***D.*** *XY plot depicting the survival of small-molecule treated depletion cell lines, calculated as the percentage of live cells with respect to the untreated controls. TetOFF DNMT1 is represented by red line and degDNMT1 by blue line. Passages are indicated by vertical dashed lines. Error bars: SD. N=3.*

6.4 Systematic analysis of downstream effects caused by the loss of DNMT1.

After successful validation of the DNMT1 depletion systems, we turned to study the downstream consequences triggered by the lack of DNMT1. Because of DNMT1's function and mode of action, we focused on DNA methylation levels, cell cycle and cell fitness.

Global levels of DNA methylation. Given that DNMT1 maintains DNA methylation throughout replication, as expected, the depletion of this enzyme in hESCs leads to a decrease in global levels of DNA methylation[3]. We validated these observations in our study using an ELISA assay that provides a quantitative measure of 5-methylcytosine (5mC) abundance (FIG 4C). Briefly, we purified the genomic DNA and hybridized it to the assay plates for a subsequent incubation with a 5mC-specific antibody. After rigorous washing of the unbound antibody, we detected the specific signal, which reports on the amount of the 5mC in the sample. Depleting DNMT1 on the transcriptional or protein level led to a 75% reduction in global DNA methylation as compared to the control cells with intact DNMT1. In agreement with the technical differences between the TetOFF and degron systems, the loss of global levels of DNA methylation in the two varies, yet shows the consistent downtrend. In the TetOFF system the methylation is gradually lost over the course of 5 days (equal to 5 cell doublings) post dox addition. This DNMT1 depletion leaves the cells with only a quarter of their initial DNA methylation content. This is in line with our western blot results showing complete protein depletion after 3 days. In contrast, the degDNMT1 cells reach 50% of wild-type global DNA methylation levels already 24h after depletion of DNMT1, followed by modest losses of methylation over the course of the next few days. Neither of the cell lines lost 100% of DNA methylation, despite the absence of DNMT1. The incomplete demethylation regardless of the DNMT1 depletion strategy could be explained by the activity of remaining DNA methyltransferases DNMT3A and 3B, which function independent of DNMT1.

Cell growth. Since DNMT1 is essential for the survival of hESC[95], we assessed the fitness of the DNMT1-deficient cells and primarily focused on the viability of the population. To this end, we adapted the hESCs and hiPSCs to single-cell growth. Next, we plated the DNMT1-deficient and non-induced control cells in equal numbers and counted them during each passage to assay their proliferation potential. The alive cells were subsequently plated in equal numbers for the next passage until no live cells were present. The cells deficient for DNMT1 grew slower than their respective control lines (FIG 4D). The absence of DNMT1 led to a progressive reduction in the number of live cells in three to four passages. Regardless of the method used and its efficiency in depleting DNMT1, we observed the most striking effect on population growth during the second passage of the experiment. Between days 4 and 8 of the experiment, the proportion of live cells decreased the most, concurrently to the changes in cell morphology and the reduction in global DNA methylation level (FIG 4A and C). Therefore, we concluded that the ability to proliferate is inversely correlated with the number of passages following DNMT1 depletion.

6.5 Cell cycle progression

The reduction in cell proliferation prompted us to investigate the cell cycle of the DNMT1-depleted cells. Since DNMT1 acts concurrent to replication, we hypothesized that their attenuated growth could be attributed to a cell cycle arrest, specifically the S-phase disruption.

EdU incorporation. To measure the proportion of dividing cells, we employed a thymine homolog 5-Ethynyl-2´-deoxyuridine (EdU) incorporation followed by fluorescent microscopy. EdU is integrated into the newly synthesized DNA and can be fluorescently stained using click chemistry[144] allowed us to mark cells actively undergoing replication[145]. The proportion of EdU positive cells did not change immediately after DNMT1 was depleted, however, the proportion of cells actively undergoing S-phase was reduced after the first passage (Fig. 4A). Furthermore, although DNA replication occurred in around 20% of cells even after 3 passages, it was significantly diminished with 40% of EdU positive cells, compared to the control population. The loss of DNA methylation may therefore affect the ability of cells to undergo DNA replication. We have a few explanations for this observation. First, we noted that not all cells behaved uniformly (i.e. some remained replicating while others did not), which might be explained by the amount of retained DNA methylation that could affect the severity of the phenotype between cells. Alternatively, the DNMT1-deficiency could lead to terminal differentiation

in a fraction of cells, which in turn decreases the proportion of cells entering the S-phase. Also, we cannot exclude that the DNMT1-depleted cells have a defect in the incorporation of the EdU during replication. The cell cycle and the EdU result specifically will be discussed later in this thesis.

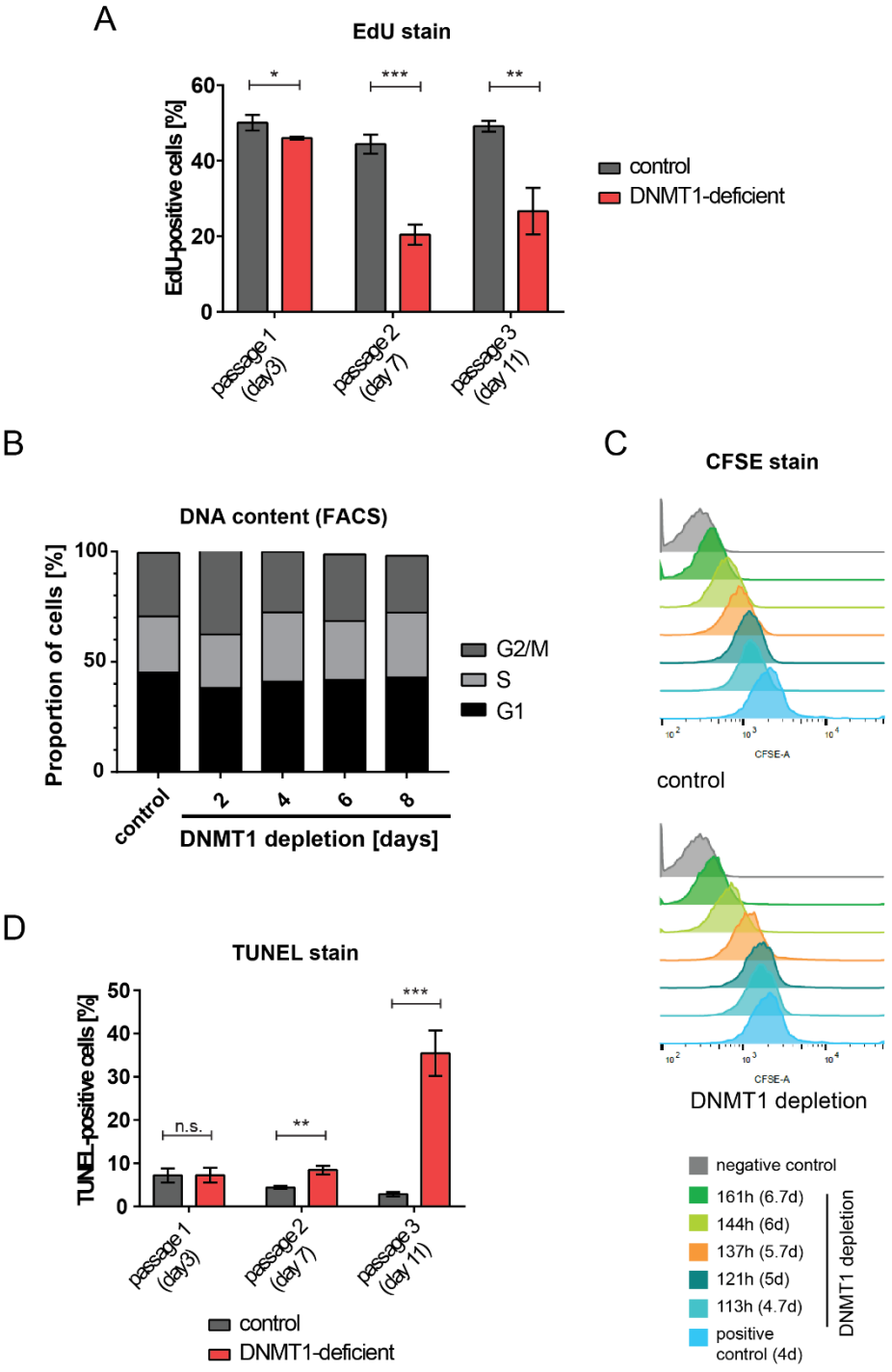


Figure 5. Characterization of the cell cycle progression in the TetOFF DNMT1 cell line. ***A.*** *Bar graphs summarizing the proportion EdU-. DNMT1-depleted cells (red bars) and control cells (grey bars) were stained with respective stains and imaged using Zeiss Cell Discoverer. Positive cells in the images, as well as total number of cells counterstained with DAPI, were automatically counted by Zeiss ZEN software. Error bars: SD. **

*pval≤0.05; ** pval≤0.005; *** pval≤0.0005; n.s. – not significant. N=3* ***B****. Bar plot summarizing the results of FACS analysis on Hoechst-stained cells. N=2* ***C.*** *Histograms of CFSE-positive cells in control (top panel) and DNMT1-depleted (bottom panel) cells. Histogram colors represent time-matched cells, defined in the legend. Light blue – freshly stained positive control. Grey – unstained negative control. N=2* ***D*** *Bar graph summarizing the number of TUNEL- positive cells. N=3 The imaging, counting and plotting strategy employed are described in* ***A****.*

Cell cycle arrest. Next, we checked if the proportion of cells in respective phases of the cell cycle was perturbed by the loss of DNA methylation. To this end, we stained the cells with DNA-binding chemical Hoechst 33342, which provides an estimate the total amount of DNA in the cells[146].

We used fluorescence-activated cell sorting (FACS) approach to estimate the abundance of cells in G1 (2n DNA), G2/M (4n DNA) and S phase (2-4n DNA). In order to avoid technical bias in the staining efficiency, we have fixed TetOFF DNMT1 cells at several time points (Fig 5B) and stained them together. We observed an enrichment of cells in the G2/M phase after 2 days of DNMT1 depletion, but did not detect other signatures of cell cycle arrest at any other time points.

The length of the cell cycle. The decreased proportion of actively replicating cells could result from the prolongation of the entire cell cycle that would have been missed when applying the previous methods. In order to examine whether the cells in the critical time point between day 4 and day 8 of DNMT1 depletion divide slower than their wild type counterparts, we employed Carboxyfluorescein succinimidyl ester (CFSE) staining with the cell-penetrable dye that covalently attaches to lysine residues, thus fluorescently tags all cellular proteins and becomes progressively diluted over cell divisions[147]. After a pulse stain with the dye followed by a wash out, we quantified the fluorescence of protein-incorporated dye by FACS (Fig 5C). Comparing the DNMT1-deficient line to untreated control demonstrated no changes in the rate of dye dilution indicating no change in the duration of the cell cycle. Taken together, we did not observe any defect in the length of or an arrest in the cell cycle in DNMT1-deficient hESCs despite a defect in proliferation. Although the EdU staining showed fewer DNA hypomethylated cells underwent DNA replication, we did not observe arrest at any of the cell cycle phases or decreased rate of the cell cycle progression in the TetOFF DNMT1-depleted cells compared to control. We thus hypothesized that the DNMT1 absence could trigger apoptosis, rather than cell cycle arrest or cell cycle prolongation.

Apoptosis. Programmed cell death can result from a failure to respond to stress subsequently leading to loss of fitness of cultured cells[148]. As DNA fragmentation is the final step and a hallmark of cell death, we used Terminal deoxynucleotidyl transferase dUTP Nick End Labeling (TUNEL), a method to stain the free DNA ends resulting from genome fragmentation in the cells undergoing apoptosis[149]. We observed an increase in the number of TUNEL-positive DNMT1-deficient cells (compared to control) in the second passage, which was elevated further in the third passage (Fig 5D). However, we find that this minor difference in the apoptotic cells cannot explain the strong defect in cell proliferation at the corresponding time point (passage 2; FIG 4D). We speculate that a different mechanism of cell death, such as necrosis, could be responsible for the loss of fitness in the DNMT1-deficient cells. Further work will be required to dissect this mode of death.

6.6 Uncoupling DNA hypomethylation from the presence of DNMT1

Although the three main catalytically active DNA methyltransferases DNMT1, DNMT3A and DNMT3B are essential for development and survival of mouse pups, the double knockout of de novo methyltransferases (DNMT3A-/-; DNMT3B-/-), hence referred to as the DKO line, is compatible with human ESC survival[95]. The reason for that could be that DKO cells retain more global DNA methylation than DNMT1-depleted cells. Overall, this may suggest a limit to how DNA hypomethylated a viable hESC can become.

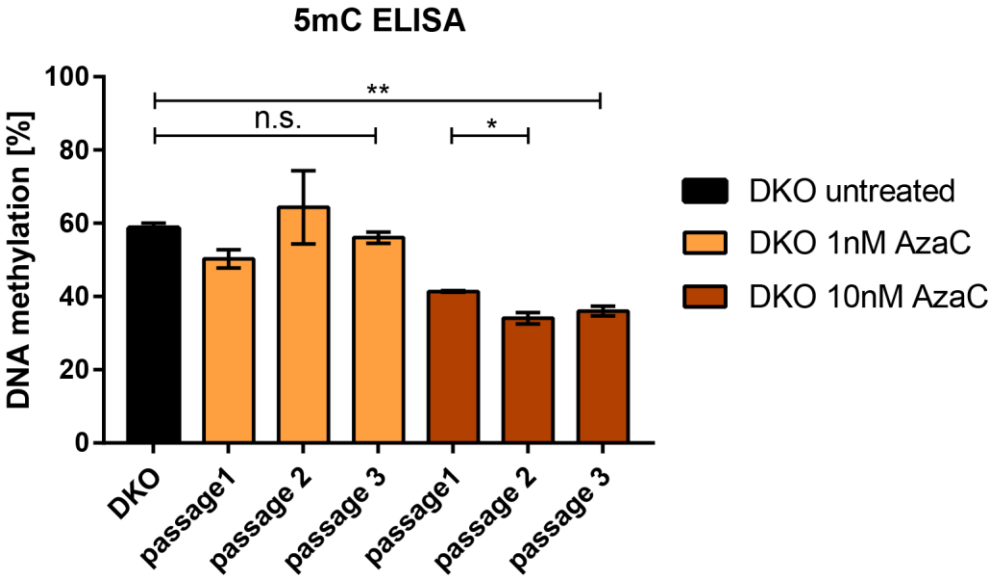


*Figure 6. Global DNA hypomethylation using 5-azacytydine. 5mC ELISA results depicting the global cytosine methylation amount in the control 3A/B double knockout cells and 5-AzaC treated counterparts. Error bars: SD. * pval≤0.05; ** pval≤0.005; *** pval≤0.0005; n.s. – not significant. N=2.*

Small molecule-induced DNA hypomethylation. In order to assess whether the total amount of DNA methylation is correlated with cell viability, we induced global DNA demethylation in a DNMT1-independent way with a small molecule. Cytosine homolog, 5-azacytydine (AzaC), is incorporated into the DNA and mimics a substrate for DNA methylation. However, unlike the regular cytosine, it entraps the methyltransferases covalently linked to the DNA. Because the DNMTs are no longer able to catalyze methyl group transfer, this results in a global DNA hypomethylation. In addition, the crosslinking of DNMTs to the DNA across the entire genome leads to DNA damage[150]. In order to reduce the global DNA methylation levels while avoiding cytotoxicity caused by AzaC, we utilized the DKO line. We reasoned that any methylation lost in the DKO cells by trapping the remaining DNMT1 enzyme would be permanently lost, thus the DNA methylation can no longer be fully replenished. Therefore, we considered to treat DKO cells with AzaC for over two cell divisions (3 days). This would deplete DNA methylation so we would then stop treatment to allow recovery from any DNA damage while maintaining the new, less methylated state. The cycle of AzaC treatment and recovery could be repeated in consecutive passages, until a minimal level of DNA methylation compatible with cell survival is reached. In this experiment (performed by Toshiyuki Ushijima), the DKO cells were treated with low doses (1 or 10nM) of AzaC for 3 days, followed by a day of recovery, passaging (in the absence of AzaC) and resumed treatment two days after attachment. We assayed the amount of methylated DNA after each treatment using 5mC ELISA (Fig 5). While 1nM AzaC did not change level of DNA methylation, 10nM AzaC treatment reduced the DNA methylation to 41%. The treatments in the two following passages rendered cell populations retained a surprisingly high 35% global DNA methylation level. Higher concentrations of the drug or subsequent treatments resulted in cell death. Notably, the retained DNA methylation level is higher than in the TetOFF DNMT1-depleted cell line. Although we cannot completely rule out the impact of DNA damage on the survival of the AzaC-treated DKO cells, we note that the cells were still able to proliferate between the second and third treatment, but the drug no longer reduced their DNA methylation levels. It is therefore likely, that complete DNA demethylation is incompatible with hESC survival. The level of 35% DNA methylation could represent a threshold, below which cell fitness is affected.

6.7 Summary

In this chapter, we confirmed that the depletion of DNMT1 in hESCs results in diminished levels of DNA methylation and a perturbed cell fitness. Specifically, we observed a reduction in cell number during consecutive passaging, accompanied by alterations in cell morphology. Even though we did not detect changes in the overall duration of the cell cycle or the cell cycle profiles, loss of DNMT1 correlated with a decrease in EdU-incorporating cells. Moreover, the loss of cell fitness correlated with an increased number of apoptotic cells after each consecutive passage. Interestingly, we noticed that the DNA methylation level is reduced and maintained at the low level (about 25% of the control line methylation in the TetOFF system and about 35% in degDNMT1 cells), but not directly after the loss of DNMT1 protein. Furthermore, when using AzaC to reduce the global levels of DNA methylation in hESCs, we also observed a retention of a similar level of global DNA methylation, despite repetitive attempts with prolonged treatment. We conclude that the loss of methylated DNA, rather than the absence of DNMT1 is the cause of the loss-of-fitness phenotype described in this chapter. In fact, our group previously showed that DNMT1KO-associated cell death cannot be rescued with a catalytically inactive DNMT1[95].

The lethality phenotype we observed appears fundamentally different than what was reported in cancer lines or mouse embryonic fibroblasts devoid of DNMT1[98]. For example, we did not observe an accumulation of G2/M arrested cells, as was shown by Chen and colleagues for DNMT1KO HCT116 cells[108]. Altogether, the DNMT1-deficient hESCs display a severe loss-of-fitness phenotype, distinct from previously published studies[98,102,105,108].

CHAPTER 2. The impact of global DNA hypomethylation on genome stability.

Because ablation of DNMT1 from cells elicits a vast reduction in the methylated cytosines, we wondered if this effect has consequences on genomic integrity, as previously suggested

6.8 DNA damage

Could then the hypomethylation cause a damage in the DNA? In fact, multiple studies have put forward this hypothesis studying various systems. For example, the DNMT1KO cancer model cell lines[106,108], non-transformed cells[106] and a tissue-specific DNMT1KO in mouse liver[151].

All in all, the conclusions drawn from depleting DNMT1 from cancer cells are not coherent and depend on the specific design of a study, therefore considered controversial. For example, DNMT1-deficient hepatocytes were reported to present markers of DNA damage that in turn induced the hepatocyte senescence; measured by the G1 phase arrest and an upregulation of a senescence marker, p21[151]. Another study found that the DNMT1KO in HCT116 cells resulted in the G2/M phase arrest, due to the accumulated DNA damage and mitotic defects[108]. Lastly, Unterberger and colleagues reported DNMT1 knockdown with siRNA to cause ataxia telangiectasia and Rad3-related (ATR)-dependent activation of DNA damage signaling in human bladder transitional carcinoma-derived cells and also in untransformed human fibroblasts. The authors proposed that the DNMT1 deficiency triggered an activation of the S-phase checkpoint that stopped further progression through the cell cycle[106]. If these inconsistencies stem from technical differences or heterogeneity of cancers remains a burning question in the field.

Because our system for depleting DNMT1 from cells has proven efficient and tight, and presents a technological advance over the previously used approaches, we explored if DNMT1-deficient iPSCs show hallmarks of genomic instability. Notably, our cells do not show defects in all aspects of the cell cycle studied, such as length and arrest, therefore we do not anticipate an excessive DNA damage that would perturb these parameters. We set out to confirm the observation with another technique and asses the signature of DNA damage in a more direct way. To this end, we employed fluorescence microscopy to monitor a highly conserved sensor for DNA double-stranded breaks (DSB) called Mre11[152]. As a part of the MRN complex, Mre11 binds the sites of damage and orchestrates the response to the insult[153]. Importantly, Mre11 is present at the sites of DNA damage occurring due to an error, such as replication fork collapse[154] or exposure

to genotoxic agents, but also when DNA damage is caused by sister chromatid exchange (SCE)[155]. It promotes both the repair via homologous recombination[156] and non-homologous end joining[157]. We, therefore, anticipated that regardless of the mechanism leading to double-strand break formation, MRE11 foci accumulation will be a reliable readout. Furthermore, it marks the DSBs very specifically, without creating a wide domain around the damage site[158].

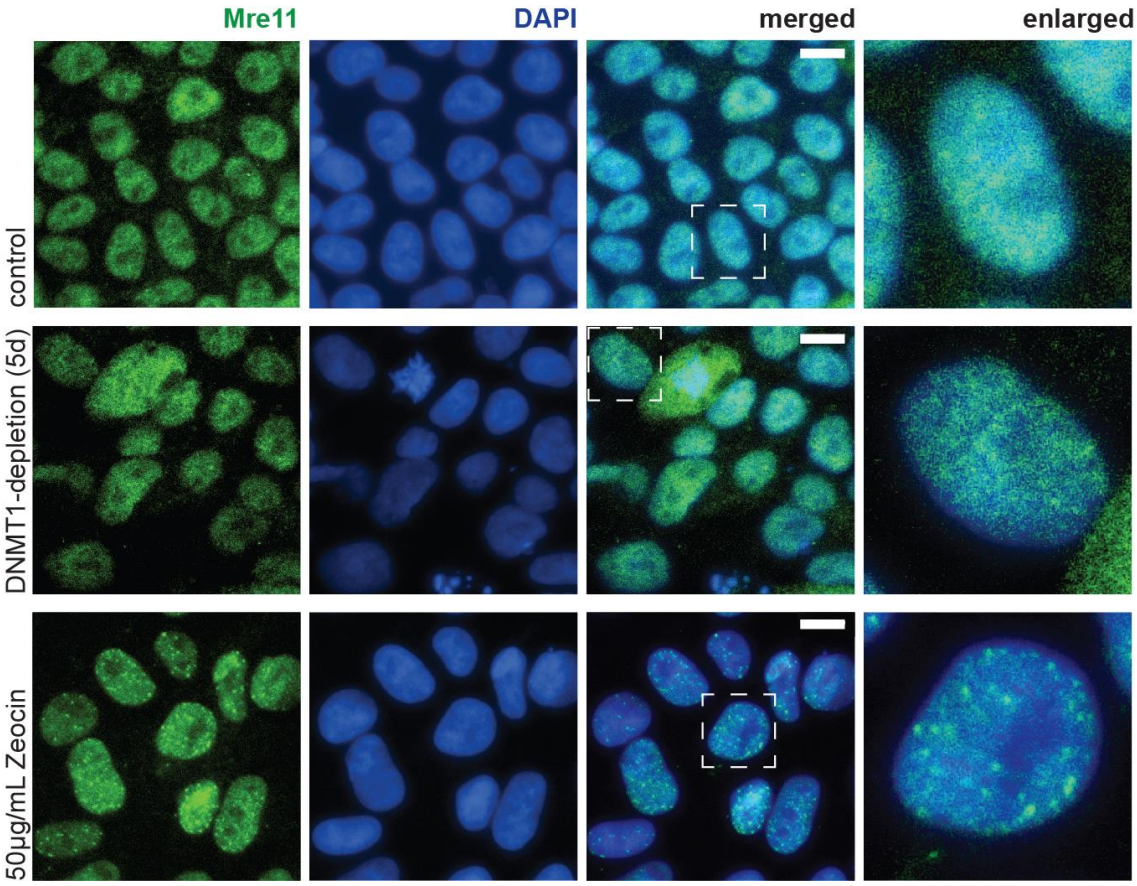


Figure 7. DNMT1 depletion does not cause Mre11 foci formation. Fluorescent microscopy images of degDNMT1 cell line stained with antibody against Mre11 (green) and counterstained with DAPI (blue). Cells in white dashed-line squares are shown enlarged in the last column. Top row: control degDNMT1 cells. Middle row: hypomethylated degDNMT1 cells, treated with dTAG13 for 5 days. Bottom row: control positive for DNA damage – cells treated with 50µg/ml zeocin. Scale bars: 10µm.

We stained the hypomethylated degDNMT1 hiPSC line with an antibody specifically recognizing Mre11 and found that while Mre11 accumulated in nuclear foci after treatment with a genotoxic agent zeocin[159] (positive control, FIG 7). After 5 days of continuous DNMT1 depletion, we did not observe any accumulation of Mre11 in the DNMT1-depleted cells as compared to controls (FIG 7). We attempted to corroborate these results with an alternative method, that relies on monitoring DNA damage-

dependent phosphorylation of histone H2A.X, which is a widely used marked of DNA damage[160]. To this end, we employed fluorescence microscopy with all commercially available antibodies. However, we found that none of them succeeded in reporting a specific signal in our positive control samples (data not shown), unlike the staining with the antibody against Mre11.

Altogether, we concluded that the depletion of DNMT1 and the subsequent DNA hypomethylation does not lead to the Mre11 foci accumulation, and therefore likely not causing double strand DNA breaks in human pluripotent stem cells.

6.9 Chromosomal instability

The link between DNA methylation and chromosomal integrity is an area of active research. One of the reasons is a rare, autosomal recessive disease called the Immunodeficiency, Chromosomal instability and Facial anomaly (ICF) syndrome[161] caused primarily by mutations in de novo methyltransferase DNMT3B[162]. On the molecular level, the defect is characterized by the loss of DNA methylation on classical satellites of chromosomes 1, 9 and 16 and decondensation of juxtacentromeric regions[163]. This leads to the formation of abnormally condensed chromosomes characterized by multiradial shape. If global hypomethylation of DNA elicits similar abnormalities has not yet been studied. Similarly, there is no known mechanism that would link the loss of DNMT1 to increased genomic instability.

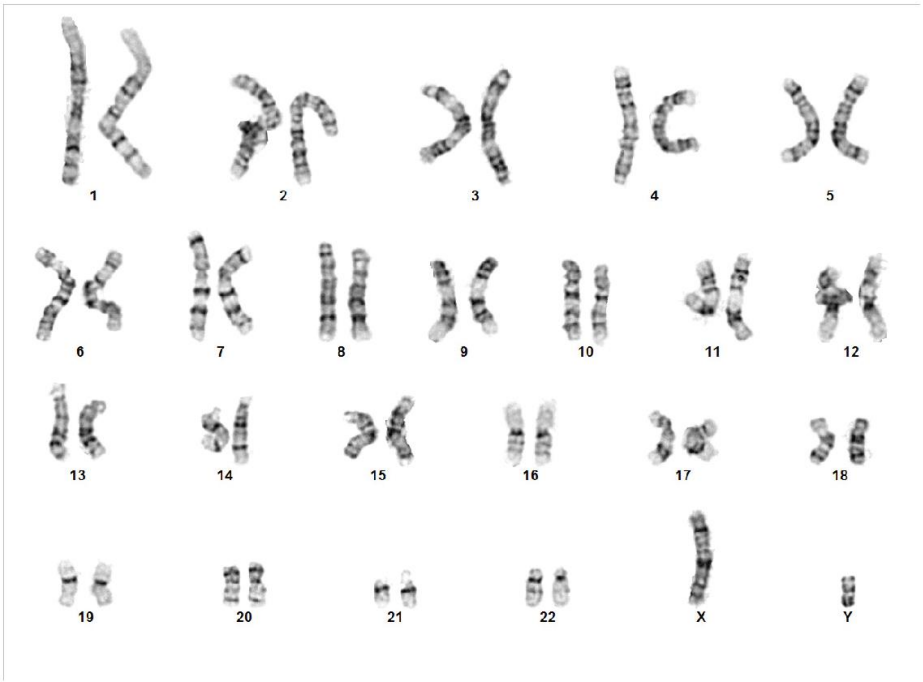
Karyotyping of DNMT1-deficient cells. To address if the DNA hypomethylation impacts centromere stability, we employed a widely-used method, Giemsa (G-band) karyotyping of TetOFF DNMT1 cells. We looked for hallmarks of ICF such as centromere decondensation, however, we did not observe differences between cells with a normal DNA methylation level and the cells affected by the loss of DNMT1 for 5 consecutive doublings (Representative karyotypying results for two genotypes; FIG 8). Unexpectedly, 3 out of 20 karyotyped, deficient for DNMT1 cells exhibited loss of chromosomes 3, 15 or Y, respectively. Because of that, the cells have undergone loss of heterozygosity (LOH) on these chromosomes.

A similar observation was previously made in the hypomethylated mouse lymphomas[87]. Because DNMT1 is a haplosufficient gene, the Jaenisch lab created a “chip” hypomorphic allele of DNMT1, which led to a decreased expression of DNMT1 in the derived from it mESCs[164]. Mice bearing only the chip DNMT1 allele (the so-called chip/- mice[87]) expressed about 10% of the wild-type DNMT1 levels, which resulted in DNA

hypomethylation of the tested IAP retrotransposable elements and centromeric repeats. These specimens developed aggressive T-cell lymphomas that led to death by reaching 8 months old in 80% of the pups. While retrotransposable elements were not upregulated in the chip/- mice, the lymphomas exhibited gains or partial losses of chromosomes. However, no karyotype analysis was performed in other mouse tissues in this study. The genomic rearrangements in the T-cells (such as gain of chromosome 15) are well-known to result in a growth advantage and frequently occur in lymphomas[87] suggesting that this particular mutation could show a selective advantage in the course of the disease development. Despite the reduction of DNA methylation and accumulation of chromosomal abnormalities, both ICF syndrome patients and model chip/- mice undergo complete embryonic development. Therefore, it is highly unlikely, that the occasional chromosomal loss in DNMT1-depleted hESCs directly leads to the dramatic loss of fitness observed in the TetOFF DNMT1 cells or the DNMT1KO mouse embryos.

A

Result: 46,XY



B

Result: 46,XY

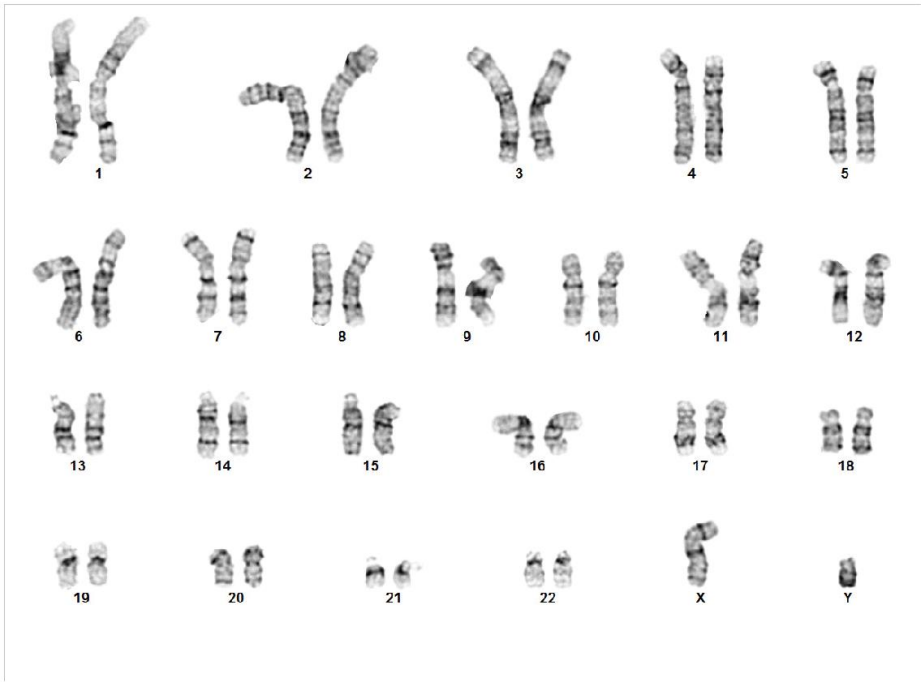


*Figure 8. Representative karyotypes of control and DNMT1-deficient cells. Representative images of G-band karyotyping in control TetOFF DNMT1 cell line **A**. and doxycycline-treated **B.** performed by Cell Line Genetics.*

6.10 Mitotic defects

Anaphase bridges. We wondered if the LOH observed in a few cells depleted for DNMT1 could point to a defect in mitosis. Interestingly, a previous study demonstrated that the knockdown of DNMT3B in HCT116 cells yielded accumulation of anaphase bridges and misaligned metaphase chromosomes[165] that are prime examples of abnormal mitosis[166]. Therefore, we investigated similar phenotypes in the TetOFF DNMT1 cells in the control and after the DNMT1-depletion (FIG 9). We utilized 4′,6-diamidino-2-phenylindole (DAPI) to image mitotic chromosomes during the anaphase of mitosis. Regardless of the presence or absence of DNMT1 (5 cell doublings - minimal time required to hypomethylate the genome), we observed identical early anaphase bridges and resolved late anaphases without lagging chromosomes (FIG 9).

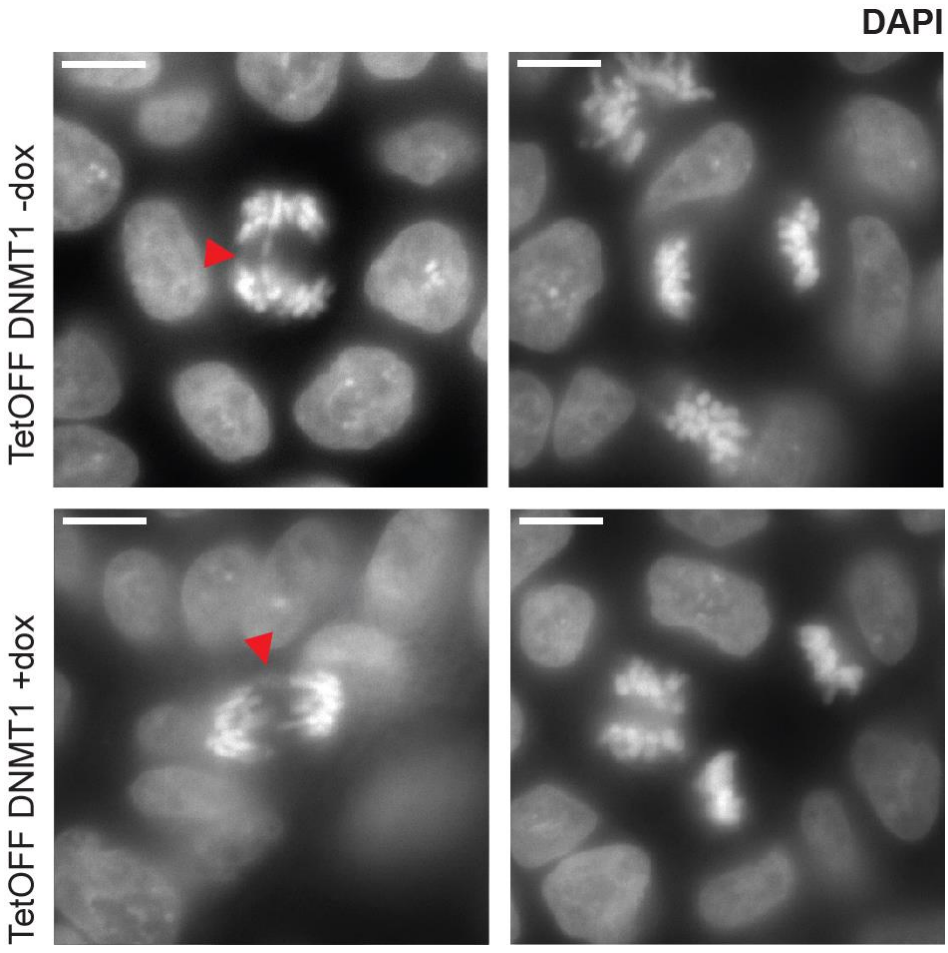


Figure 9. Anaphase bridge formation occurs in early anaphases during hESC mitosis. Representative pictures of anaphases stained with DAPI in control (top row) and DNMT1-deficient (bottom row) hESCs. Early (<10μm distance between daughter cells' chromosomes) and late (>10μm distance between daughter cells' chromosomes) anaphases are shown in left and right column, respectively. Red arrow – anaphase bridge. Scale bars: 10μm.

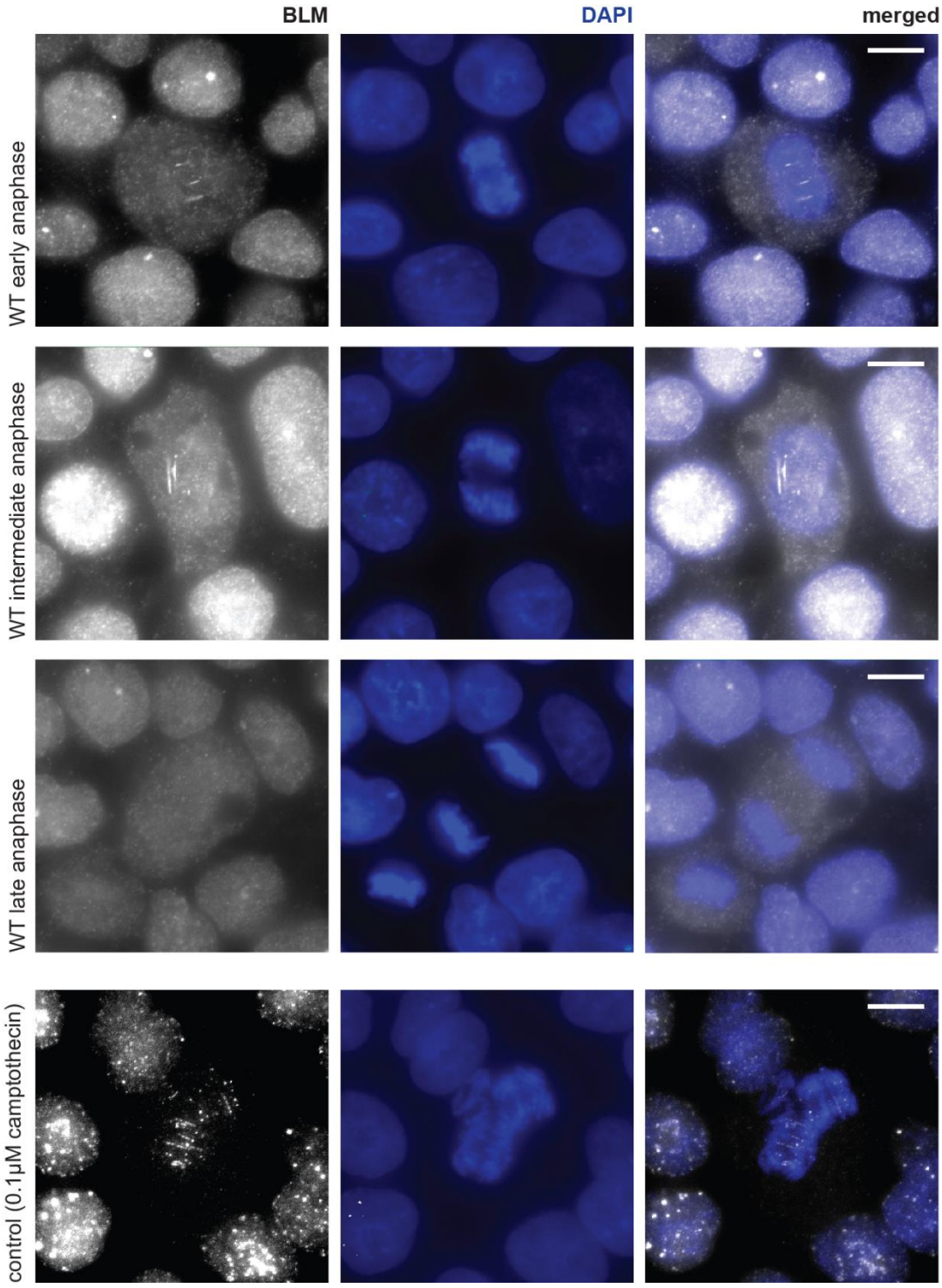


Figure 10. Ultra-fine BLM bridges are resolved during anaphase. Fluorescent microscopy images depicting early (first row; <3µm distance between daughter cells' chromosomes), intermediate (second row; <10µm distance between daughter cells' chromosomes) and late (third row; >10µm distance between daughter cells' chromosomes) anaphases in HUES64 hESCs, stained with antibodies against BLM (white) and counterstained with DAPI (blue). Bottom row: positive control for BLM bridge formation depicting anaphase in cells treated with topoisomerase inhibitor camptothecin. Scale bars: 10 µm.

Chromosome decatenation. Next, we further examined alternative anaphase-related phenotypes and focused on chromosome decatenation, which is a mechanism to untangle sister chromatids for faithful chromosome segregation[167]. This process is performed on anaphase DNA structures called PICH-bridges, also known as Ultra-Fine Bridges (UFBs)[168], proposed to be the remnants of centromeric chromatin catenates that hold sister chromatids together and provide tension between the kinetochores to clear the spindle assembly checkpoint (SAC). PICH and BLM DNA-dependent ATPases resolve such catenates with topoisomerase, Topo IIα[169]. Due to these DNA bridges being very thin, they do not accumulate enough DAPI signal to be visible under a fluorescent microscope unlike the intact chromosomes; however, the thin DNA bridges accumulate the Bloom helicase that could be stained with an antibody to indirectly monitor them. Consistently with previous studies[168], we observed a decline in the number of UFBs over the course of anaphase in wild-type cells (FIG 10). Cells in early anaphase, with mitotic chromosomes of daughter cells only beginning separate, display multiple BLM-positive bridges. The bridges are stretched out in intermediate anaphase (less than 10μm distance between chromosomes of daughter cells) and are completely resolved before late anaphase (over 10μm distance between chromosomes of daughter cells). The wild-type cells treated with a topoisomerase inhibitor, camptothecin[170] (positive control) display an increased formation of the DNA bridges (FIG 10). When we examined cells deficient for DNMT1 (for 2, 4, 6 or 8 doublings, respectively), we observed no difference as compared to wild-type cells; the vast majority of observed anaphases were completely resolved before or during intermediate anaphase, regardless of condition (Fig 11). These results indicate that the hypomethylated cells are capable of decatenating UFBs. Therefore, we concluded that DNMT1-deficient cells do not display decatenation defects in anaphase.

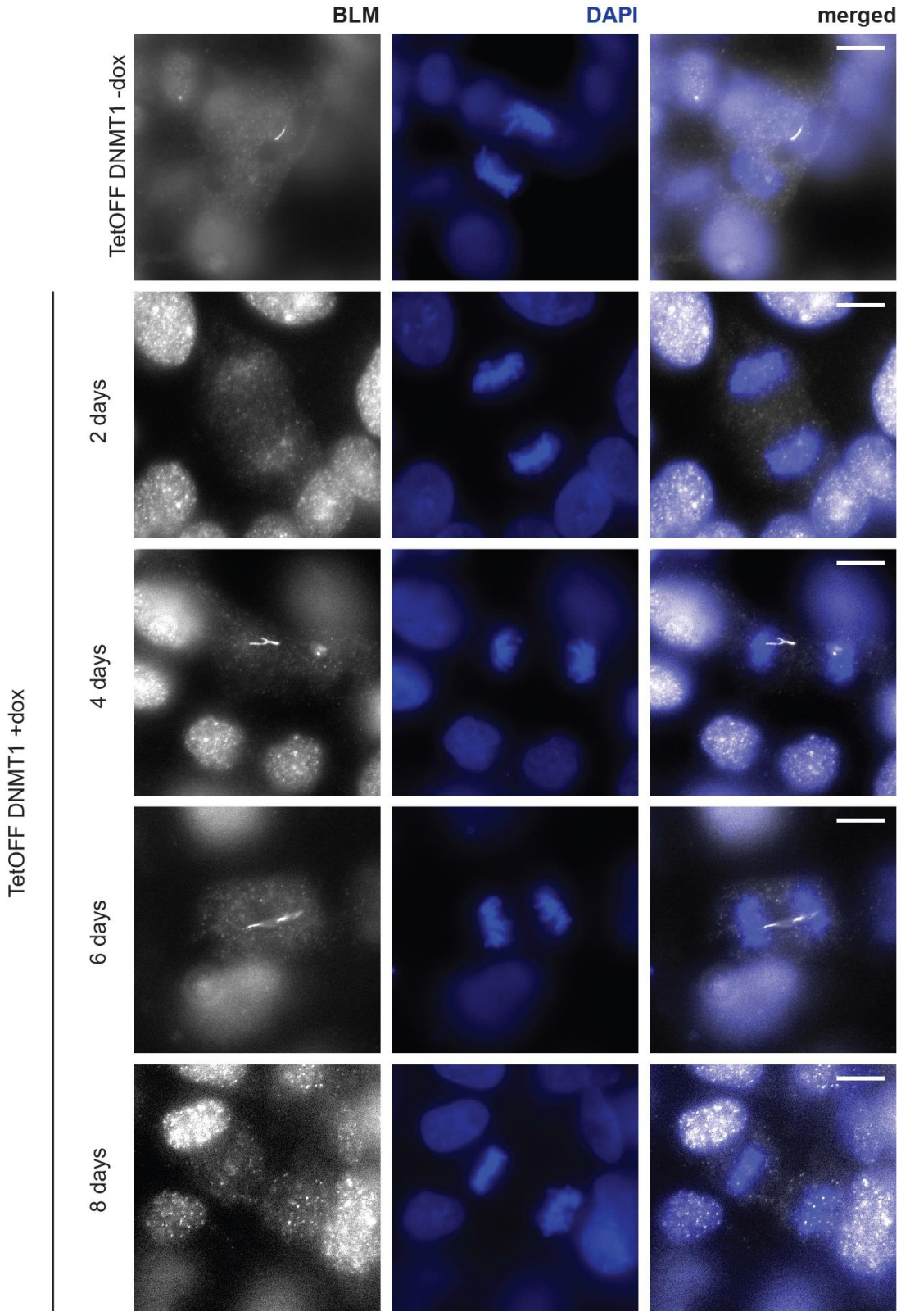


Figure 11. DNMT1-depleted cells are able to resolve anaphase UFBs. . Fluorescent microscopy images depicting late anaphase with bridge resolution in TetOFF DNMT1 cells stained with antibodies against BLM (white) and counterstained with DAPI (blue). Top row: control cells; subsequent rows: cells treated with dox for 2, 4, 6 or 8 days, respectively. Scale bars: 10 μm.

Because DNMT1 acts co-replicatively, we hypothesized that its depletion could lead to a defect in the S-phase of the cell cycle. However, this type of defect would not manifest itself by arresting the cells in S-phase, as we did not observe any accumulation of the S-phase-arrested cells after removal of DNMT1 (FIG 5B). In such case, we considered an accumulation of under-replicated DNA as a possible mechanism. Discrete regions in the genome called Chromosomal Fragile Sites (CFS) are difficult to replicate due to, for example, their repetitive nature[171]. Thus, if replication of these regions is incomplete, the partially synthesized DNA can be carried on through the cell cycle[172]. Previous studies identified 53BP1 as a protein that creates protective foci around the under-replicated DNA in the subsequent G1 phase to prevent further damage and was used as a sensitive marker of such loci[173]. In order to elucidate whether the under-replication phenomenon occurs in the hypomethylated hESCs, we quantified the 53BP1 foci. Lukas and colleagues found that wild type cells accumulate fewer than 6 foci per nucleus[173], whereas a challenge with doxorubicin (a drug that intercalates in the DNA causing damage) vastly increased the number of foci (FIG 12A).

In the cells depleted for DNMT1 (for 2, 4, 6 or 8 days, respectively) we observed a minor, but not statistically significant increase in the 53BP1 foci (day 4, 6 and 8), as compared to matched, untreated controls (FIG 12B). This is an interesting observation given the hypomethylated hESCs incorporate less EdU during replication, which may implicate a change in kinetics or fidelity of DNA synthesis (FIG 5A). Future experiments will be required to address this discrepancy with alternative approaches, such as DNA combing, and also examine, for example, if the cells depleted for DNMT1 do not display a compromised uptake of the EdU nucleotide.

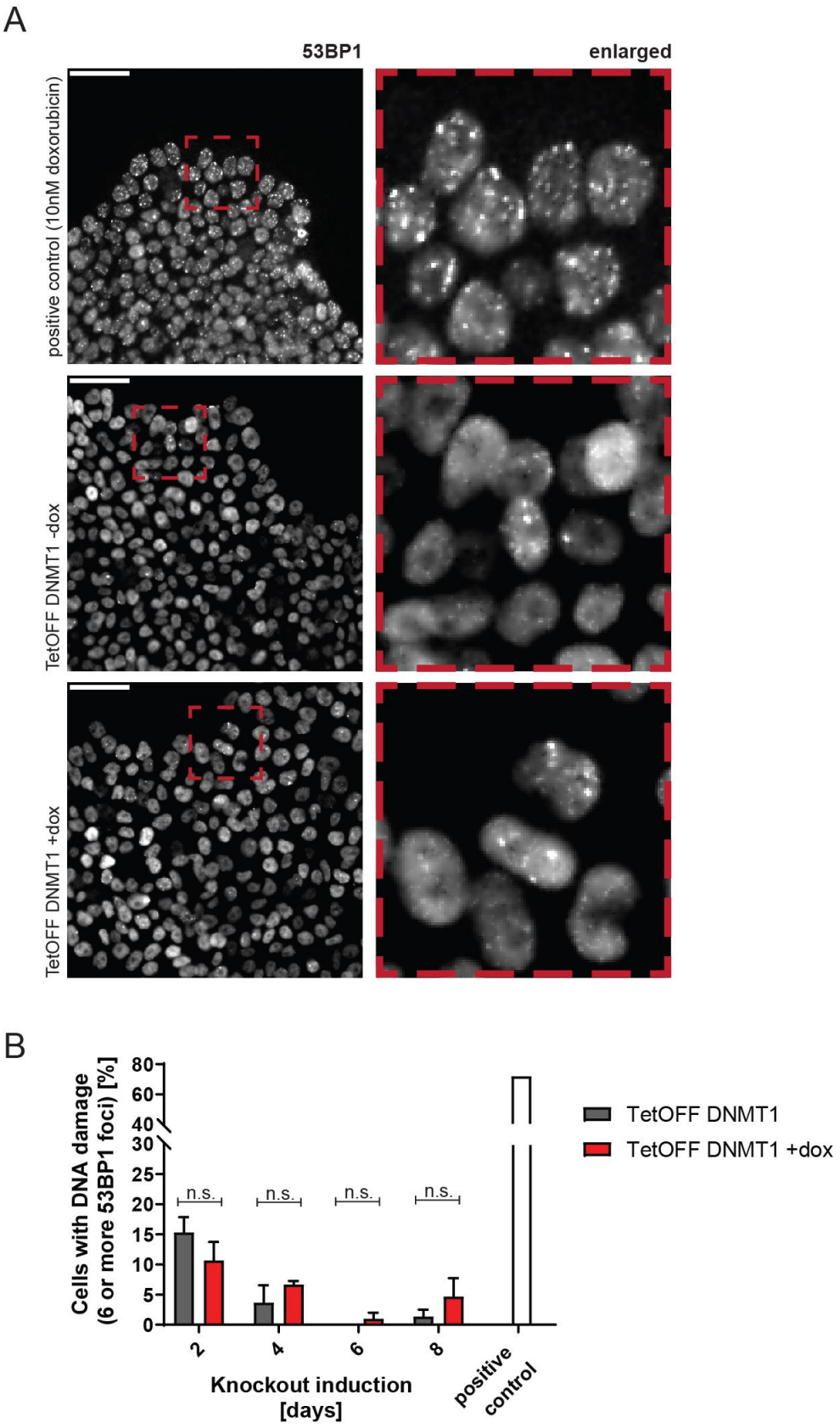


Figure 12. Loss of DNMT1 does not promote the accumulation of 53BP1 foci. ***A.*** *Representative pictures of fluorescence microscopy imaging of TetOFF DNMT1 cells stained with an antibody against 53BP1. Top row: positive control of cells treated with 10nM doxorubicin. Middle row: TetOFF DNMT1 cells. Bottom row: DNMT1-deficient*

TetOFF DNMT1 line treated with doxycycline for four days. Right column: enlarged view of cells shown in the red dotted line squares. Scale bar: 50 µm. 100 cells were screened per condition. ***B.*** *Quantification of 53BP1 staining in TetOFF DNMT1 cells. Bars represent the percentage of cells with 6 or more 53BP1 foci.*

Mechanisms masking mitotic defects. Although our data demonstrates only occasional loss of chromosomes in a very few cells, we considered the possibility that the major defects could have already caused a cell death and, because of that, were missed in our analysis. Alternatively, the potential deleterious effect of global hypomethylation of DNA could be masked by a cellular buffering system that compensates for the insults to genomic stability. In fact, a very thorough study of cell-cycle regulators demonstrated that a single knockout of genes involved in replication, such as *CDT1* or *ORC1*, did not impact the mitosis in these cells, however a double knockout with the *TP53* gene had a synergistic effect, with nearly all mitoses being erroneous[174]. Therefore, the product of the *TP53* gene, the p53 protein may have a modulatory effect on several knockouts and buffer their loss. p53 is a tumor suppressor transcription factor, often referred to as “the guardian of the genome” due to its central role in responding to cellular stress. Some of its many roles include integrating signals from DNA damage or triggered cell cycle checkpoints and regulates genes involved in DNA repair, senescence and apoptosis to elicit a suitable response[175]. Interestingly, knockout of the *trp53* gene in mouse embryonic fibroblasts was shown to partially rescue the DNMT1KO-associated growth defect[98]. We sought to test these findings in the DNMT1-deficient hESCs. To this end, we knocked out the *TP53* gene in the TetOFF DNMT1 cell line using lentivirally-delivered Cas9 together with a guide RNA that targets the *TP53* genomic locus (FIG 13). To select the most efficient gRNA with the lowest number of off-targets, directing the Cas9 enzyme to an exon shared by all TP53 transcript isoforms, we used an integrated gRNA prediction platform provided by Benchling. We confirmed homozygous knockout in clonally derived cell line using a PCR-based genotyping strategy followed by Sanger sequencing. The selected clone contained an out-of-frame deletion in one allele and an in-frame insertion of 51 nucleotides containing three in-frame stop codons that would result in the loss of protein product. The resulting TetOFF DNMT1; *TP53*-/- line was characterized by significantly increased growth rate compared to the parent line (FIG 14A). We then compared the cell number of DNMT1- and DNMT1;TP53-depleted cells after one passage. In line with the findings of Jackson-Grusby and colleagues[98], the hypomethylated cells devoid of *TP53*

show an increased survival as compared to the hypomethylated cells (FIG 14B). Next, we monitored the kinetics of the loss of fitness for both cell lines.

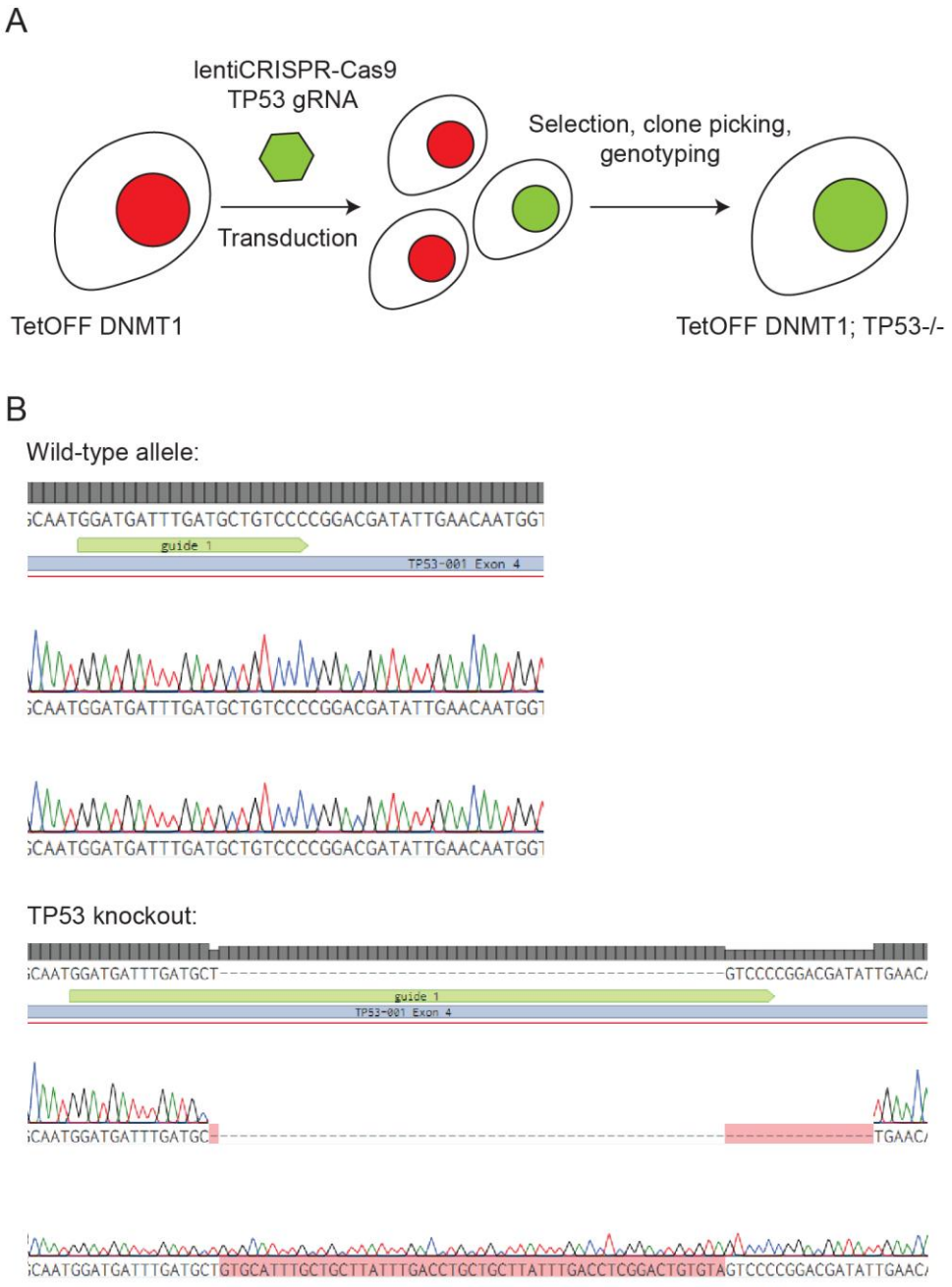


Figure 13. The creation of TetOFF DNMT1; TP53KO cell line. ***A.*** *Parental TetOFF DNMT1 was transduced with a lentiviral vector containing wild-type Cas9 cDNA, a guide against human TP53 and a drug resistance cassette. The untransduced cells were selected against with blasticidin. Clonal colonies of surviving cells were picked and genotyped.* ***B.*** *Chromatogram results of genotyping of the TP53-exon 4 in WT hESCs*

(top panel) and the TetOFF DNMT1; TP53KO line. Consensus sequence is shown on top of each panel as a grey bar. Green arrows: guide sequence. Chromatograms: sequences of the alleles of TP53.

Interestingly, the *TP53KO* suppressed the cell death of DNMT1KO cells at passage 2, hence delaying severe loss of cells by one passage. However, the TP53KO was not able to rescue the initial drop in cell viability within the first passage (4 doublings) of DNMT1 depletion (FIG 14C). These results suggest that the DNMT1-deficiency-associated growth phenotype in hESCs is at least partially dependent on TP53.

To check if the presence of p53 in the hypomethylated cells masks otherwise-occurring genomic instability, we have employed confocal microscopy to score metaphase defects. Previous study[174] employed DAPI and tubulin staining to image mitotic chromosomes and the mitotic spindle and distinguished three abnormal metaphase phenotypes. We used the same method to score the mitotic abnormalities occurring in the DNMT1-deficient hESCs. Representative examples of each defect are shown in (FIG 15A). Off-axis chromosomes (FIG 15 A, white arrow) do not assemble with the other chromosomes in the mitotic plate. Spindle multipolarity results from more than two centrosomes creating additional mitotic poles. Major plate defects manifest themselves by the lack of mitotic plate formation.

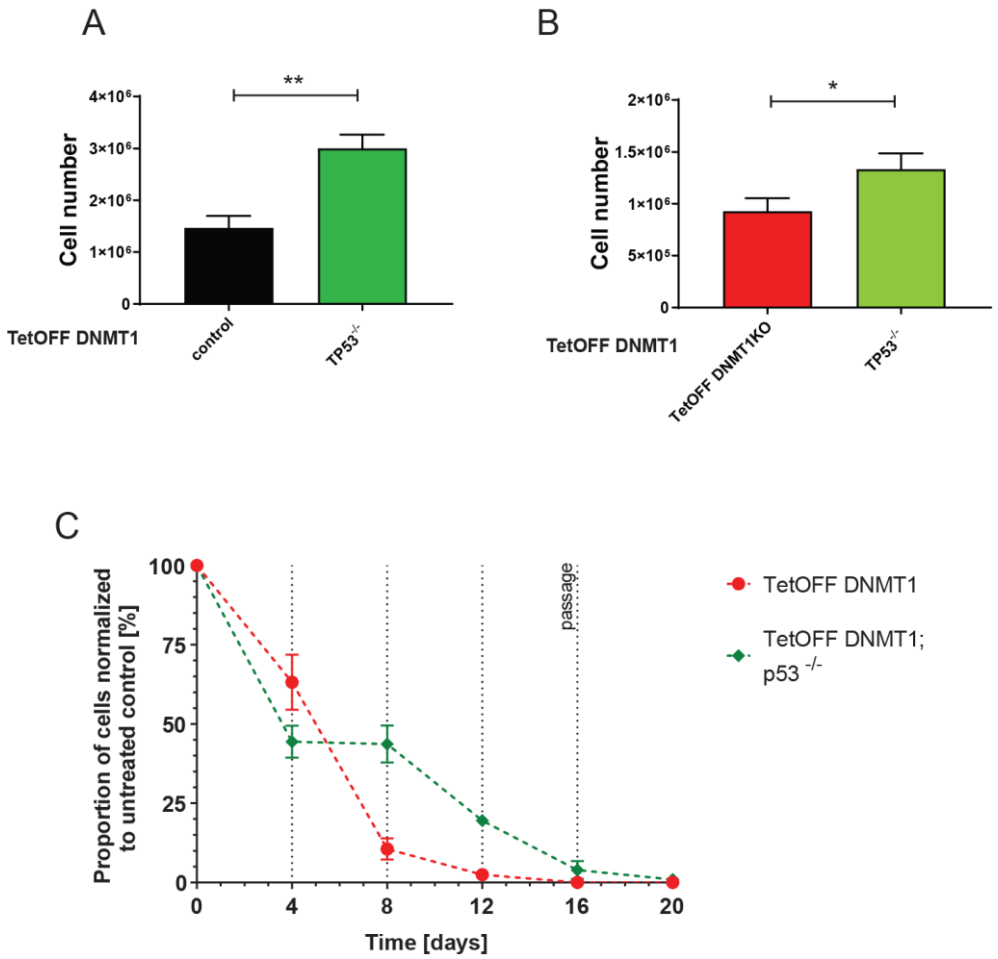


Figure 14. TP53 knockout impacts the growth of control and DNMT1-deficient hESCs. ***A.*** *and* ***B.*** *Bar graphs depicting the number of cells grown one passage (4days) after single-cell plating of TetOFF DNMT1 control cells (black bar) in comparison to TetOFF DNMT1; TP53KO lineage (A), or TetOFF DNMT1-depleted cells compared to cells deficient in both DNMT1 and TP53 (B).* ***C.*** *Scatter plot depicting the DNMT1-deficient TetOFF DNMT1 or TetOFF DNMT1; TP53KO cells that survived each passage. The values are expressed as a fraction of cell number achieved by control dox untreated cells of respective lineage. Vertical dashed lines represent re-plating for a new passage. * pval < 0.05; ** pval < 0.005; *** pval < 0.0005. Error bars: SD.*

We scored the mitoses in DNMT1-depleted cells and observed whether the absence of p53 can modulate the phenotype (FIG 15B). While DNMT1-depleted cells displayed a minor increase in the proportion of off-axis chromosomes, we observed similar ratios of unaffected mitoses in the control and DNMT1-depleted cells (87% versus 86%, respectively). In stark contrast, TP53 deletion resulted in an increase in all mitotic defects compared to control cells, with only 70% normal mitoses. Lastly, depletion of both DNMT1 and p53 further exacerbated the number of mitotic defects seen in the TP53KO cell line. For example, a major increase in off-axis chromosome defect cell arose from

19% in TP53KO to 31% in the double knockout. Our data suggests a synergistic effect between DNMT1 and TP53 with respect to mitotic defects, thus the double knockout presents an additive phenotype that is stronger than each of them separately Therefore, this genetic experiment implies that DNMT1 and TP53 act in separate pathways in regards to mitosis defects. Interestingly, despite the accumulated genomic insults in the double knockout line, these cells show a growth advantage over the DNA hypomethylated cells. If this is due to the impaired checkpoints in the TP53 knockout or whether genomic instability does not affect the fitness of DNMT1-depleted cells will be an exciting area of further research. Future studies will be required to address the mechanism behind the increased survival of the DNMT1KO-TP53KO cells. In addition, to examine specificity of TP53, it will be interesting to examine if a deletion of any other tumor suppressor gene would yield a similar result.

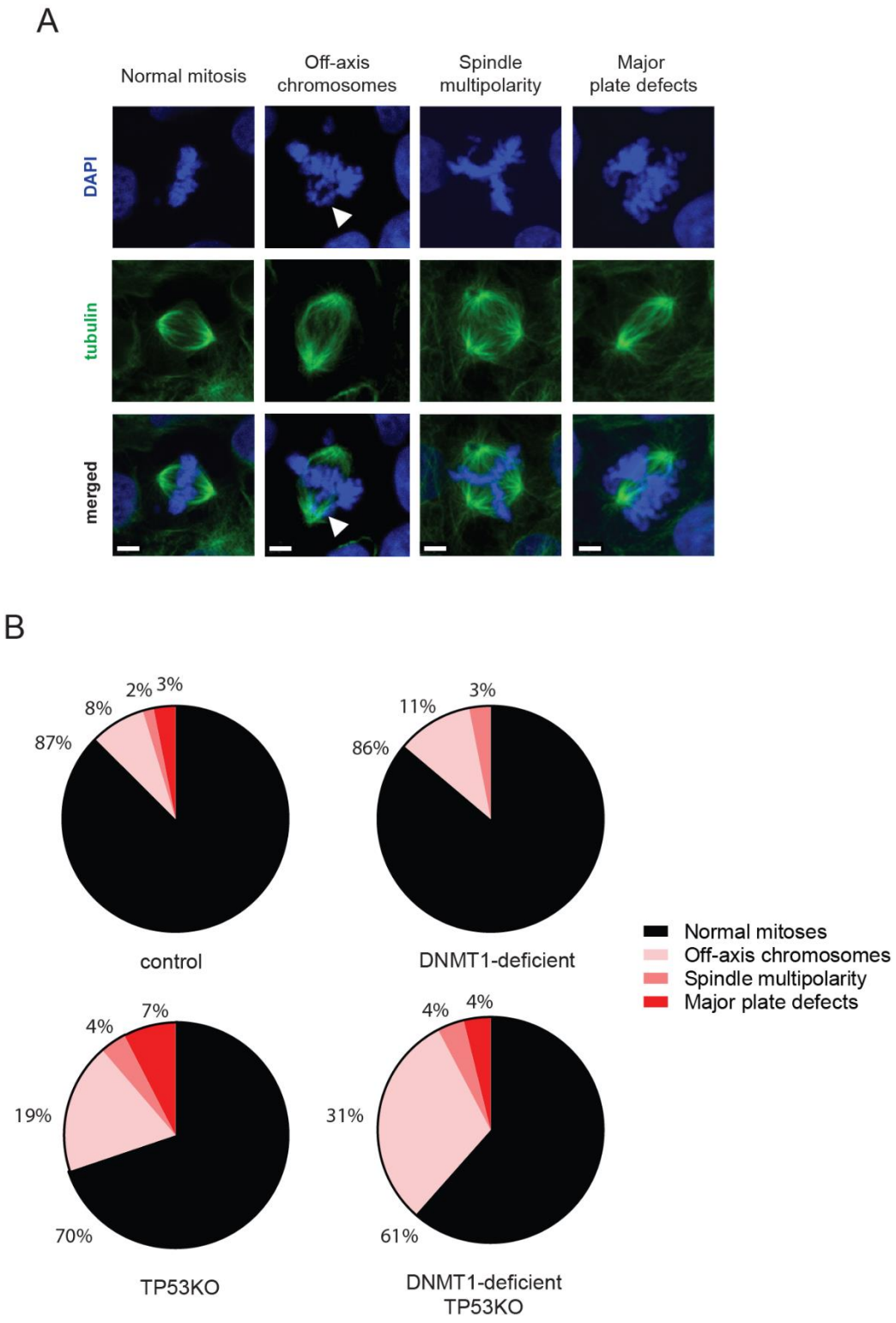


Figure 15. TP53KO does not act synergistically with DNMT1-depletion in causing metaphase errors. ***A.*** *Representative images of cells undergoing normal (first column) or erroneous metaphases (second to fourth columns). Cells were stained with an antibody against tubulin (green), counterstained with DAPI (blue) and imaged using confocal microscopy. White arrows – off-axis chromosomes. Scale bars: 5 μm.* ***B.*** *Pie charts depicting the quantification of the proportion of metaphases scored according to the key shown in* ***A.*** *Each genotype is presented in a separate circle; colors represent mitotic phenotypes. 20<N<40 metaphases per condition.*

6.11 Expression of transposable elements

Repetitive sequences such as transposable elements (TEs) typically bear highly methylated DNA at their genomic loci. This was proposed to be a mechanism by which the cells prevent the expression and transposition of these elements. Multiple studies reported the demethylation of transposable elements in DNMT1 knockout cell lines and embryos[87,98,102,176]. The global loss of DNA methylation resulted in transcriptional activation of TEs[98,102]. Jackson-Grusby and colleagues hypothesized that such expression of transposable elements, which belonged to the intracisternal A particle (IAP) family, could cause of DNMT1KO MEF cell death via increased transposition and resulting DNA damage[98].

Although transposons make up nearly half of the human genome[80], very few are able to transpose. This is ascribed to the degeneration of TE sequences, caused by accumulation of mutations or truncations due to an incomplete transposition. DNA transposons are believed to be virtually extinct in the context of the human genome[177]. RNA transposons, however, contain transposition-competent elements. Human LINE elements, which occupy the most of the human genome out of all the TE classes, only an estimated 60-100 are capable of transposition[178]. Another transposition-competent class, analogous to the mouse IAP, are human endogenous retroviruses (HERVs). Particularly the HERVk subtype has been shown to transpose into new locations in the human genome[179].

Although the expression of different classes of transposons has resulted in mutations[180], however, it also occurs during normal embryonic development and transition from primed to naïve state in hESC[117]. We decided to investigate if the transcription of transposable elements occurs in the TetOFF DNMT1-deficient hESCs. To this end we used quantitative reverse transcription polymerase chain reaction (qRT-PCR) to directly compare the level of TE transcripts between control and the globally hypomethylated cells (FIG 16). Specifically, we chose primers amplifying the HERVk *gag* sequence[117] and to the consensus sequence among all L1 transposons, including the most recent L1_Hs class[181].

No significant upregulation of the L1 transcript level accompanied the loss of DNMT1 (FIG 16A). In contrast, HERVk elements were increased 4 fold (FIG 16B). This is comparable to HERVk upregulation during *in vitro* reprogramming of primed hESCs to naïve state, a process that does not lead to cell death[117] We therefore conclude that minor upregulation of the TEs is not likely the underlying cause of the DNMT1 depletion-mediated loss of fitness in disagreement with the observations by Jackson-Grusby and colleagues[98].

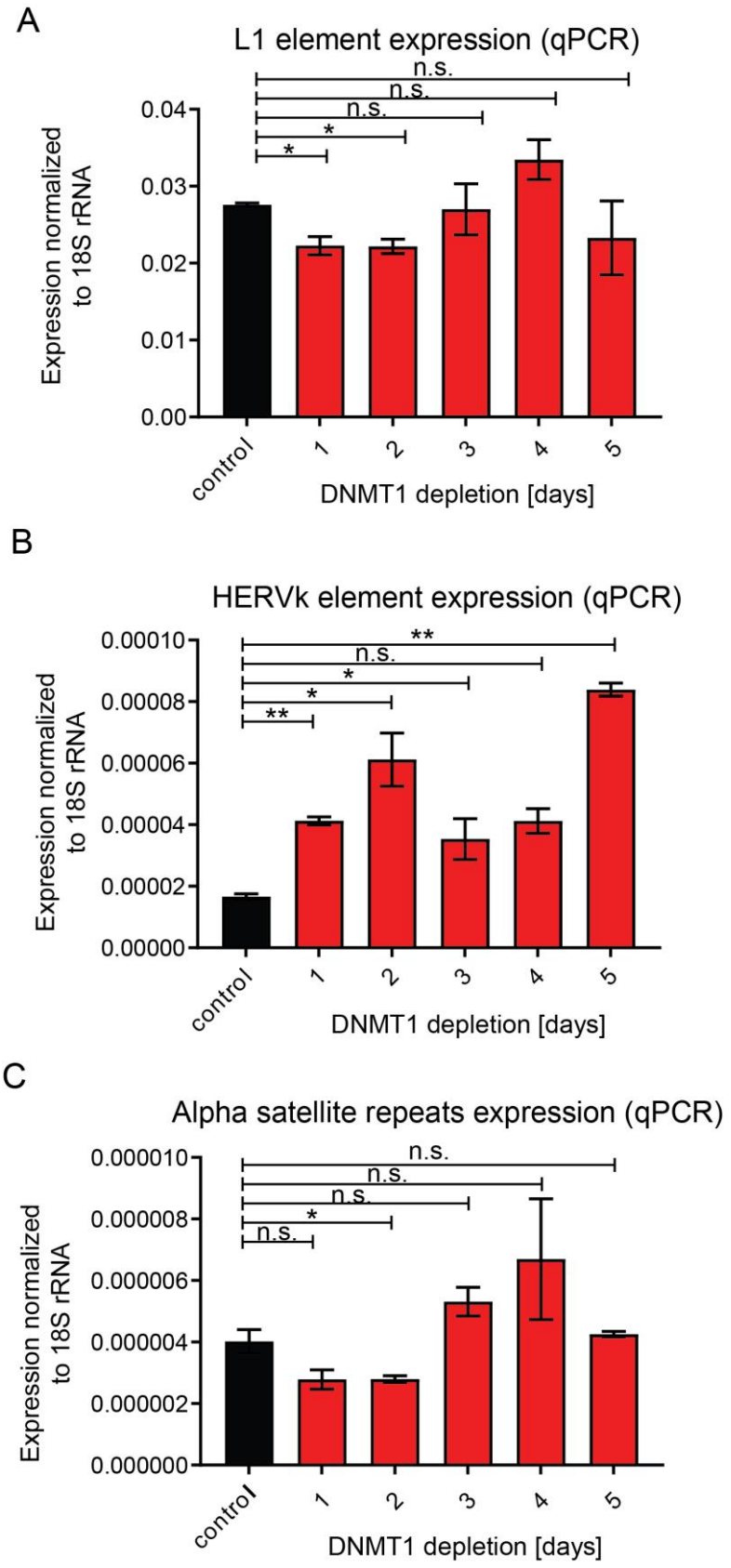


*Figure 16. The effect of DNMT1 depletion on repetitive element expression. Bar graphs depicting the expression level of **A.** L1, **B.** HERVk, **C.** Alpha satellite repetitive elements. * pval < 0.05; ** pval < 0.005; *** pval < 0.0005. n.s – not significant. Error bars: SD. N=3.*

Centromeric repeats. Centromeric DNA hypomethylation following the mutations in DNMT3B is one of the hallmarks of ICF syndrome[161,162]. Although we did not find any evidence for chromosome decondensation in our TetOFF DNMT1-depletion cell line, we explored whether global DNA hypomethylation could impact the transcription of centromeric repeats.

The change in transcription level of satellite repeats has been observed in various cancer types. For example, pancreatic cancer cells upregulated alpha satellites over 40-fold, compared to healthy tissue[182]. Nevertheless, its role in disease or genomic instability is not well understood.

We performed qRT-PCR in the TetOFF DNMT1-deficient hESCs using primers mapping to the consensus sequence of human alpha satellite 171bp repeats, which are present in the higher order repeats in centromeres of all the human chromosomes[183]. Interestingly, we only observed a modest, less than 2-fold upregulation of ALRs following DNMT1 depletion (FIG 16C). This result indicates that the hypomethylated DNA in DNMT1-deficient cells is not likely a trigger for centromeric transcription.

6.12 Summary

We have employed fluorescence and confocal microscopy approaches to investigate if a hypomethylated genomic DNA poses an insult to genomic stability. We showed that the DNA damage markers were not enriched in the acutely hypomethylated human ESCs compared to control, even though the DNA damage was previously reported in other types of DNMT1-deficient cells. If this is due to the unique biology of embryonic stem cells as compared to cancer cells or simply an indirect effect remains to be studied. Even though the hypomethylated ESCs lack the known DNA damage markers, we detected a loss of single chromosomes in a few cells. In the future, it will be interesting to track chromosomal aberrations in hESCs, examine how these affect the cell survival and when they occur with respect to the loss of DNA methylation.

We also considered if the hypomethylated hESCs would display markers of genomic instability more prominently in the absence of p53. While loss of TP53 increased the survival of DNMT1-deficient cells, it also increased the occurrence of mitotic abnormalities. This observation uncouples the impaired proliferation from the presence of mitotic defects, therefore suggesting that both processes are rather independent. Whether the absence of p53 uncovered genomic insults posed by the hypomethylated

genome or if it creates a sensitive, genetic background more prone to defects, remains to be studied in the future and likely will bring new mechanistic insights.

Our results demonstrated that global DNA hypomethylation did not significantly impact the expression of transposition-competent elements or the major satellite repeat sequences. Although DNA methylation levels are significantly downregulated in the absence of DNMT1, DNMT-independent mechanisms in the cells work to repress transposable elements. Notably, the KRAB-associated protein (KAP1), directed by Zinc Finger Proteins, is able to recruit the H3K9 methyltransferase SET domain bifurcated 1 (SETDB1) to the transposon loci and thus facilitate their silencing even after the depletion of DNMT1[176].

Other non-coding sequences in the genome, such as microsatellites, are too repetitive to be examined using qRT-PCR or conventional RNA sequencing techniques. Further studies using nanopore sequencing methods, which are adapted for long reads, could be applied to elucidate the changes in expression of these sequences in DNMT1-deficient cells. Although demethylated transposable elements were found to be recombination hotspots in testis[49], our DNA damage staining and karyotyping results argue against the recombination or chromosomal translocation events occurring in the DNMT1-deficient hPSCs.

Taken together, we conclude that hypomethylation of DNA does not likely trigger a genomic instability directly, albeit we cannot exclude the presence of DNA abnormalities as secondary, indirect consequence. These observations prompted us to consider and investigate other mechanisms that are responsible for the loss of fitness observed upon the DNMT1 removal in human embryonic stem cells.

CHAPTER 3. Simultaneous profiling of cytosine methylation and transcripts levels in individual cells.

Our 5mC ELISA-based measurements in the DNMT1-depleted hESCs highlighted that about 30% of DNA methylation was retained in the cell population. This result could be achieved in two ways: (1) either the majority of cells retain a similar low levels of the global DNA methylation or (2) the population consists of a heterogeneous mixture of cells containing both high and low levels of DNA methylation. In order to elucidate which possibility is correct, we employed a technique that would allow us to quantify the methylation level and genome-wide distribution of this epigenetic mark in individual cells.

Bisulfite conversion. Sequencing-based methods assaying DNA methylation rely on bisulfite conversion[9] reaction, in which unmethylated cytosines are deaminated and thereby converted to uracil. The methylation status of each cytosine is monitored by the subsequent DNA sequencing, where deaminated cytosines pair with adenines and are thus interpreted as thymines, while methylated cytosines are not. By comparing the DNA sequence of a reference human genome with the sequencing results after bisulfite conversion, the converted cytosines (previously methylated) are distinguished from endogenous thymines. Because of the introduced nucleotide alteration, the sequencing reads carry the information about the DNA methylation status and can be subjected to quantitative and qualitative analyses.

Genome-wide profiling of DNA methylation. Due to the large information content and complexity of genomes, mapping DNA methylation sites across the entire genome requires next generation sequencing (NGS) approaches. Several methods implementing bisulfite conversion have been developed to date. For example, the Whole-Genome Bisulfite Sequencing (WGBS)[184], can be used for quantifying the cytosine methylation level across all the CpGs in the genome. Because CpGs are underrepresented and unevenly distributed across the human genome, sequencing of the entire genome is cost inefficient. Therefore, several strategies have been developed to enrich for the reads of interest. One of such approaches is an immunoprecipitation of methylated DNA (MeDIP), which uses an antibody against 5-methylcytosine to enrich for DNA fragments containing methylated cytosines[185]. Another strategy, called Reduced Representation Bisulfite Sequencing (RRBS), focuses sequencing on the relevant, CpG-containing loci, enriching for the gene-proximal regions and lowering the coverage of numerous repetitive sequences or gene deserts [186]. RRBS utilizes the DNA restriction enzymes that contain the CG dinucleotide in their recognition sites to fragment the DNA, producing fragments containing CpG-enriched sequences.

Single-cell sequencing. We recognized that the RRBS technology presents an attractive and powerful approach to assess the cytosine methylation status in our studies on the DNMT1-dependent DNA methylation. Therefore, we leveraged the power of DNA methylation analysis and combined it with the sequencing of single cells. Single cell sequencing methods are the state-of-the-art approach to address heterogeneity in a cell population. For example, single cell sequencing of transcriptomes facilitated the identification of cell subpopulations in complex tissues, such as the brain[187] and characterization of dynamically changing processes, such as gastrulation[188]. Protocols assaying transcription, DNA methylation, chromatin accessibility, histone modifications or combinations of several readouts have been developed and implemented in a number of recent studies[189–191].

Because the amount of DNA material for all single cell analyses is the limiting factor, we considered the variant of RRBS, the single-cell reduced representation bisulfite sequencing (sc-RRBS)[192] approach as the most suitable for our studies. The protocol for the sc-RRBS includes all the steps of a regular RRBS in a single tube to minimize loss of DNA material from the single cell. This technique succeeded at quantifying the demethylation of maternal and paternal pronuclei in mouse zygotes, as well as, providing qualitative data on the regions that were losing DNA methylation with a different kinetics than others[192].

DNA methylation profiling combined with transcriptome profiling in individual cells. Because one of the roles of methylated cytosines is to restrict transcription[193], ideally one would combine information of the cytosine methylation status together with transcript profiles in every cell. Therefore, we decided to combine the sc-RRBS with another technique: single cell RNA sequencing (scRNAseq)[120]. RNA sequencing is a method for assaying the mRNA transcripts level of all the genes; and its single cells version has been broadly used recently.

The development of a new technology (FIG X) that combines single-cell methylome and transcriptome analysis was performed by Hongcang Gu, Arman Mohammad and Andreas Gnirke of the Broad Institute. To obtain single-cell libraries, the polyadenylated RNA was first isolated by hybridysation of biotinylated oligo(dT) followed by immunopurification of poly(A)-containing mRNAs with streptavidin beads. After the subsequent reverse-transcription reaction, the cDNA was converted into SMART-Seq2 sequencing library and processed as described before[120]. Concomitantly, the genomic DNA of each sample was purified, digested with enzymes, ligated to adapters and bisulfite-converted. Our collaborators Caleb Laureau, Ayush Raman and Martin Aryee

have processed the raw sequencing data and aligned the reads to the human genome using a custom firecloud-based pipeline for scRRBS[121]. The DNA methylation and transcription data were analyzed by Jocelyn Charlton and Sudhir Thakurela, respectively. Altogether, the resulting technology of dual sequencing - scRNAseq+RRBS (described in more details in the Methods section) empowered us to simultaneously study the distribution of DNA methylation and the transcriptional output originating from each individual cell.

6.13 The combined scRNAseq and RRBS dual protocol dataset

We followed the decay of DNA methylation and its impact on the steady state RNA abundance after transcriptional shutoff of DNMT1 in the TetOFF DNMT1 hESCs. Our study included 9 timepoints (day 0 to day 8 with or without the transcriptional shutoff of the DNMT1 transgene) where cells were fixed daily and FACS-sorted simultaneously. The libraries were prepared as described and sequenced on the Illumina HiSeq 4000 machine. With 78% read mapping to the human genome a median of 1,031,026 CpGs were covered at least once, which is very similar to the efficiency obtained by Guo and colleagues in their sc-RRBS protocol[194]. By comparison, an alternative sc-WGBS[195] protocol yielded more CpGs mapped ($3.7x10^6$ on average), however, (1) it covers fewer CpGs than the RRBS methods and (2) there are fewer matched regions covered between cells, which would make that method less optimal for our purposes.

The scRNAseq protocol yielded a median of 963,842 transcripts uniquely mapping to the genome. This covers a median of 5,626 genes expressed per cell with an expression level of at least 0.1 transcripts per million (TPM). Overall, these statistics from the both sc-RRBS and sc-RNAseq indicate that the experiment and the general approach were technically successful and thus will be suitable for further analyses.

In order to avoid biases and false data, we filtered out cells where the protocol yielded suboptimal results, for example, we excluded cells with low number of mapped reads in either protocol. In the analysis of scRNAseq, we also excluded the cells with read number higher than 1 standard deviation from the average, as these were highly likely to represent cell doublets. 96% and 76% of the cells passed the scRBBS and RNAseq, respectively. Taken together, 72% of the cells passed both the RRBS and RNAseq criteria and were used in our analysis.

While mapping the cells to the genome, we noticed that cells collected after day 4 show expression profiles enriched in ribosomal RNAs (rRNAs). Because the transcript capture

method used in our protocol is based on purifying the polyadenylated RNAs and rRNAs are not polyadenylated, we suspected that this scRNAseq library was contaminated with the genomic DNA. Thus, the samples representing this time point were excluded from the transcriptome analyses.

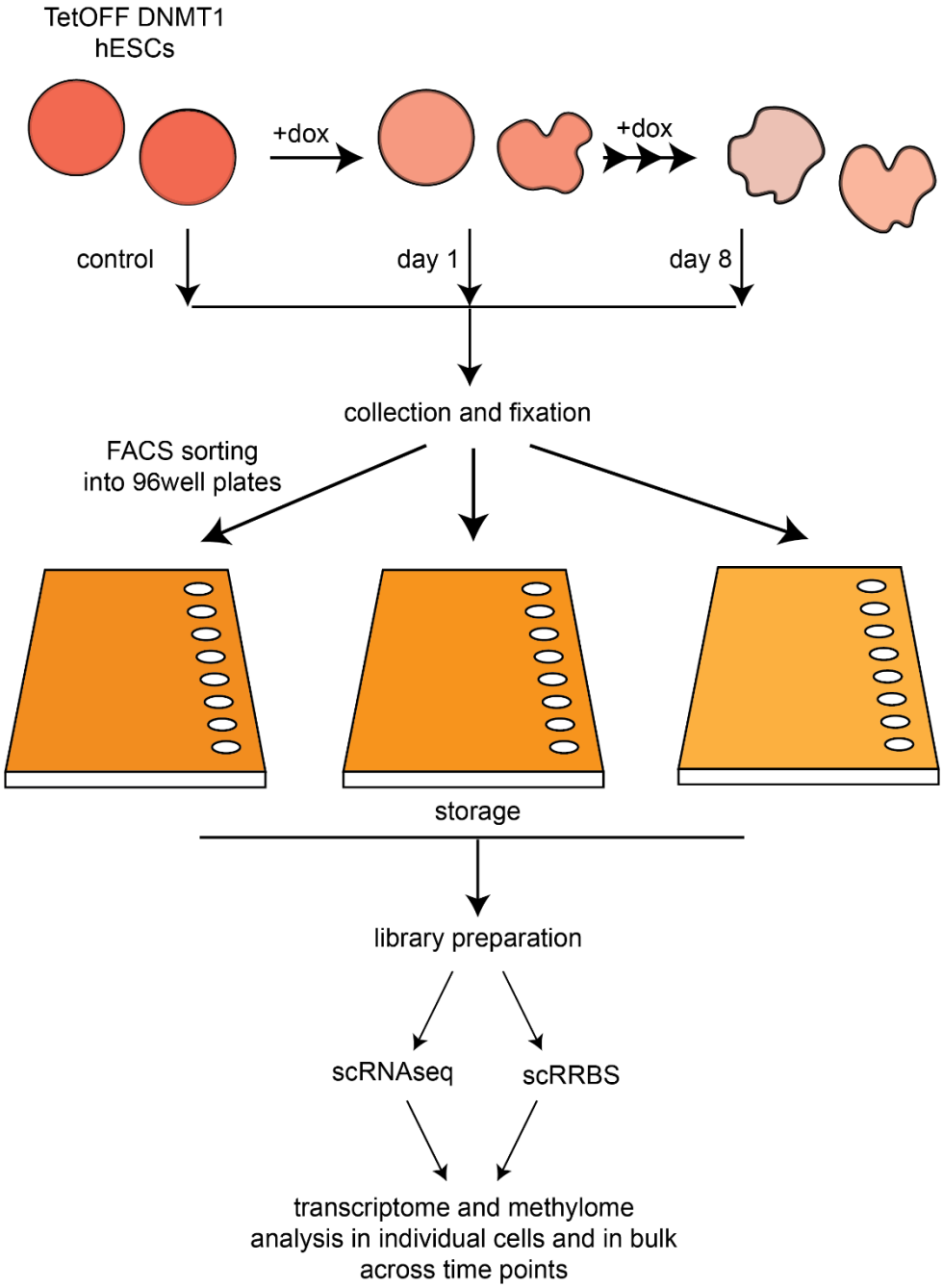


Figure 17. Dual protocol overview.

6.14 Quantitative and qualitative study of DNA methylation in the individual DNMT1-depleted cells

We used the resulting data sets to analyze the amount and distribution of the methylated cytosines in the individual cells, as well as in the pools of cells from the matching time point with the goal of finding a common characteristic for the regions retaining DNA methylation in the DNMT1-deficient cells. To this end, we investigate the methylation on the scale of chromosomes, as well as at specific sets of loci, stratified by their function. First, we plotted the mean DNA methylation of cells at each time point, as well as the methylation average of individual cells (FIG 18). Consistently with our 5mC ELISA results (FIG 4C), the global levels of methylated cytosines decrease gradually over time and plateau in samples collected five days after the transcriptional shutdown of DNMT1. Interestingly, the DNA methylation profile shows highly variable patterns between the cells before this time point (for example, at day 3) and thus, after averaging, gives an intermediate value. This variability could stem from (1) a technical reason, the heterogeneous response to blocking transcription in the TET-OFF system; (2) a biological reason, intrinsically heterogeneous response to the reduction of the DNMT1 protein levels; or (3) the number of cell division the cells went through since the reduction of DNMT1 protein.

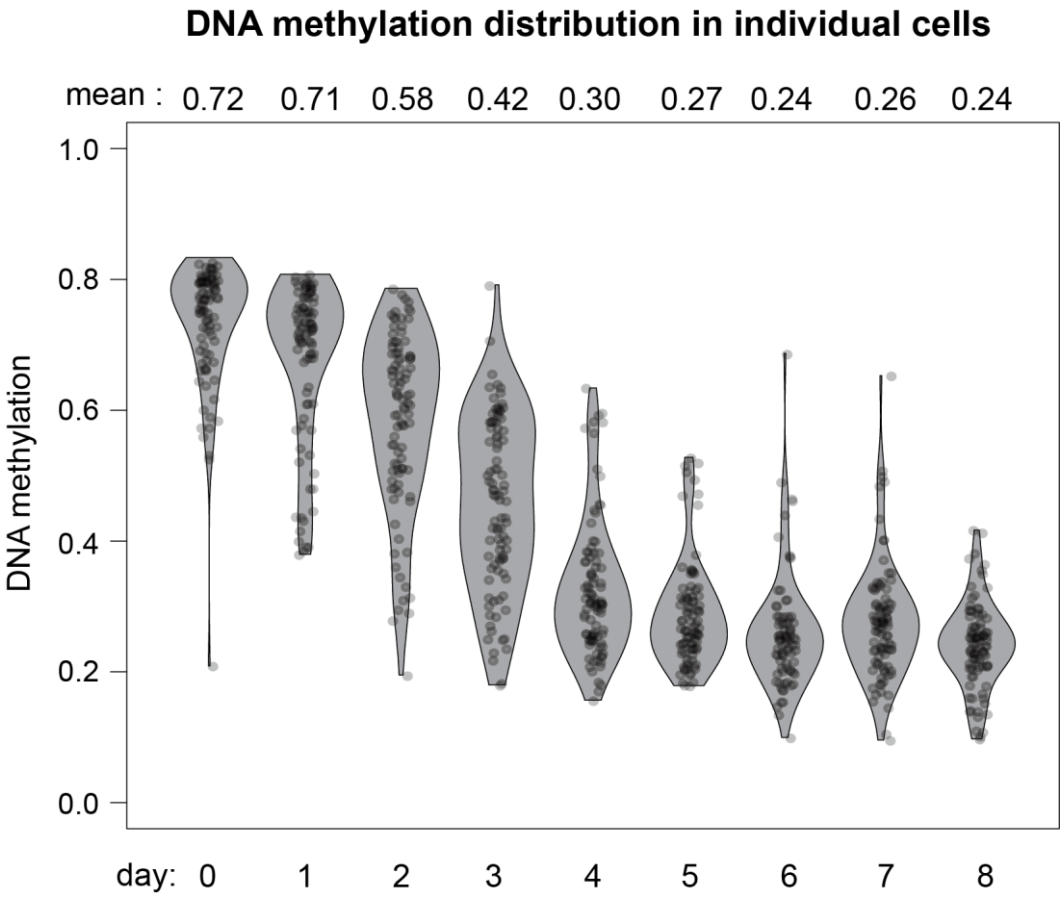


Figure 18. DNMT1 depletion results in loss of DNA methylation. Violin plot representing the mean methylation (indicated on top) across hESCs depleted for DNMT1 for the indicated amount of time. Each dot represents the mean DNA methylation of a single cell.

Interestingly, even at the final time point (after 8 days of DNMT1 depletion), we did not find cells that would retain less than 10% global methylation, suggesting that further reduction in global DNA methylation could be incompatible with survival. On the contrary, only a few cells retained the methylation levels above 60% beyond day 5, which could be a result of their early, spontaneous terminal differentiation into a non-mitotically engaged cell type that would prevent dilution of the methylated cytosines in the genome even in the absence of the maintenance DNA methyltransferase, DNMT1.

6.15 Investigating the DNA methylation on chromosomal level

We wondered if any retention of methylated cytosines could occur by the virtue of inheriting highly methylated chromosomes from the parental cell. Therefore, we sought

to characterize the global DNA methylation decay on the chromosomal level in more details.

Even though cytosine in the CpG dinucleotide is the substrate for DNA methylation, there are specific sequence contexts that are preferably unmethylated. These are the CpG-dense regions of high guanine-cytosine richness called the CpG islands (CGIs)[36], genomic features overlapping with 70% of gene promoters in the human genome[37].

In contrast, in the other parts of the genome, CpG dinucleotides are typically underrepresented (averages at less than 1% while expected is 4.4%)[38]. The sequence of the human chromosome 19 is enriched for the CpGs, as it contains high gene density. In fact, its total cytosine methylation content is reduced 3-6% as compared to the average in the human genome. The same trend was maintained in the DNMT1-depleted cells while monitoring the loss of DNA methylation (FIG 19A, B).

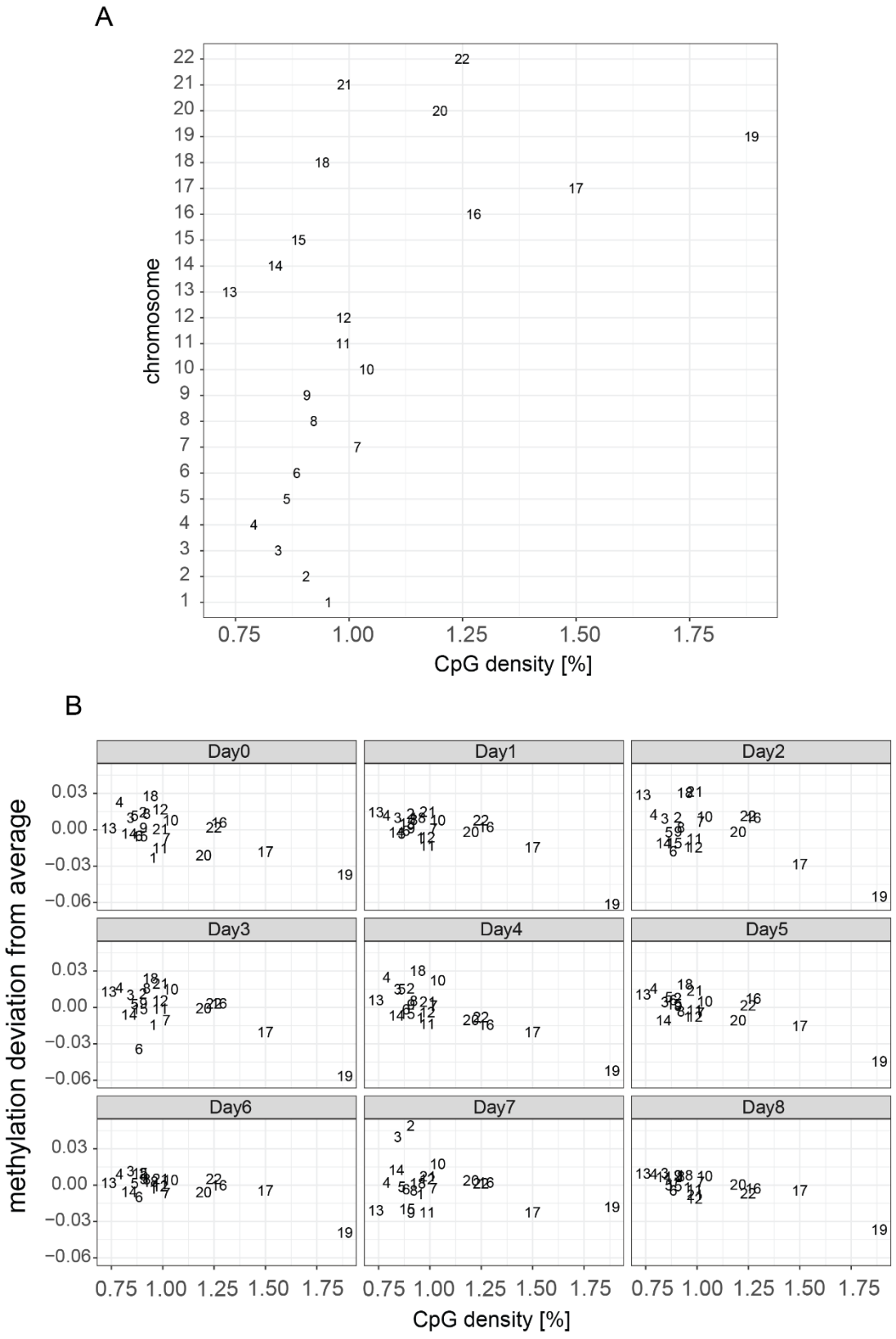


Figure 19. Mean chromosome CpG density does not predict DNA methylation retention in DNMT1-depleted hESCs. ***A.*** *Plot depicting the CpG density of each of the human autosomes.* ***B.*** *Plot depicting the deviation of DNA methylation from the average of all the cells belonging to the indicated time point (Y axis), and the chromosome CpG density (X axis).*

We plotted the mean methylation of each chromosome in each individual cells (FIG 20). The control cells (day 0) display a high level of DNA methylation across all chromosomes (60-80% of CpGs). We suspect that this variation is likely due to some cells actively undergoing replication. In the control cells, the DNA methylation of any given chromosome closely correlates with the mean DNA methylation level of the cell.

As the cells lose DNMT1, the pattern of mean DNA methylation on chromosomes changes profoundly. After three days of DNMT1 depletion, the cells retain a consistent methylation average on their chromosomes, however, we observed variability between cells (FIG 20). In contrast, the cells collected at day 4 consisted of both highly and lowly methylated chromosomes, despite their similar methylation means. We hypothesize that these variable patterns at day 3 could likely result from a kinetics of DNMT1 protein depletion, for example the cells could retain DNA methylation across the genome thanks to the residual DNMT1 (that would be below the western-blot detection limit; see FIG 5A). At day 4, however, the DNMT1 protein is no longer present, thus the retained methylation level will depend solely on how much DNA methylation was inherited from the parental cell. At the latest time point, day 8 after the transcriptional shutoff of DNMT1, the analyzed cells show a homogenous and consistently low DNA methylation pattern on chromosomes. We did not find evidence of specific retention of DNA methylation on any particular chromosome. The DNA methylation level we found in the cells at day 8 likely marks the minimal level of DNA methylation compatible with survival.

Overall, our results indicate that the survival of the DNMT1-deficient cells is not linked to the DNA methylation level at any particular chromosome, but rather the retention of this epigenetic mark across the entire genome. We hypothesize that the retained DNA methylation reflects the level of the mark required for survival. It remains to be studied if, in the absence of DNMT1, the de novo methylation obtained through overexpressing DNMT3A/B could substitute it.

Next, we asked if the position of CpGs along the chromosome determines their maintenance in the absence of DNMT1. FIG 21 shows a representative plot of CpG methylation at chromosome 5 in 30 randomly chosen cells at day 0, 5 and 8 (due to the limiting number of reads, these CpGs were not matched). We found the DNA methylation seems to be scattered across the chromosome length and does not show any preferential maintenance towards the telomeric or centromeric end of either arm (FIG 21).

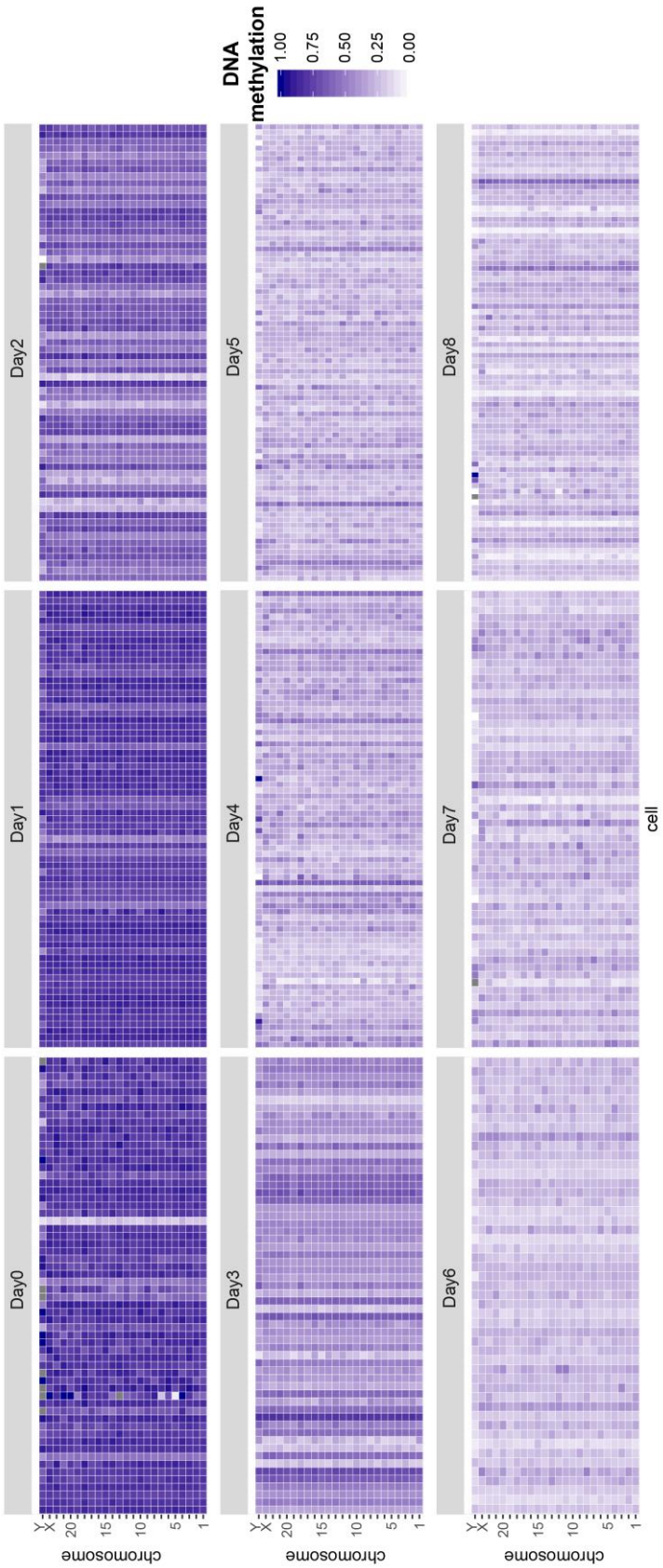


Figure 20. Mean chromosome methylation of globally DNA hypomethylated DNMT1-deficient hESCs. Heatmap representing the mean DNA methylation of each chromosome in the cells at the indicated time point. Gray square – no data.

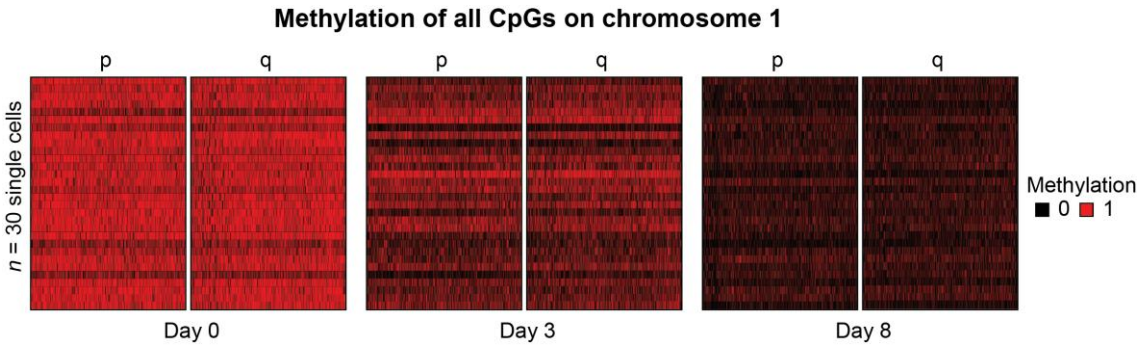


Figure 21. Distrubution of DNA methylation across the chromosome in DNMT1-depleted cells. Heatmap of DNA methylation of each CpG mapped across the short (p) and long (q) arms of chromosome 1 in 30 randomly selected cells and day 0, 3 and 8. Red bar represents methylated cytosines and black bar represents unmethylated cytosines.

6.16 Qualitative analysis of DNA methylation

Because downstream in-depth analyses required a more substantial read coverage, we pooled the scRRBS reads from the cells belonging to a single time point. For the subsequent analysis, we stratified the human genome based on the function of each region. Our features of interest were gene promoters, introns, exons, enhancers, CpG islands and transposable elements. Intrinsically, these different regions vary in the DNA methylation level. The CGIs had the lowest average methylation level, averaging at less than 15% methylation. As these coincide with most of gene promoters in human genome, the promoter regions showed lower DNA methylation than the genome average (FIG 22A). In contrast, gene bodies are methylated, evident from the mean methylation levels of both introns and exons. We identified intermediate levels of DNA methylation at annotated enhancers. The DNA methylation level of transposable elements, such as LINEs, was about 10% higher than the genome average (FIG 22B). This is consistent with the conserved role of DNA methylation in transcriptional silencing of these regions[196]. Again, once we analyzed the genome DNA methylation profile of DNMT1-depleted cells, the genomic features showed neither preferential depletion nor preferential retention of the mark. Therefore, we considered to stratify the genome based on histone modifications and monitored the methylation level at regions enriched in specific histone marks based on the ENCODE ChIP-seq data[197]. Of our interest were marks found at enhancers (H3K27ac, H3K4me3), promoters (H3K4me3) and active genes (H3K36me3) as well as the repressive mark H3K9me3[198] (FIG 22C). Similarly to genomic features, none of the histone marks was associated with preferential retention

of DNA methylation. For example, H3K9 trimethylation, a histone mark associated with heterochromatin was enriched with high DNA methylation, as previously observed[199]. Similar to the pattern in the entire genome, the loss of DNMT1 elicits a progressive reduction in DNA methylation in the H3K9me3-associated regions.

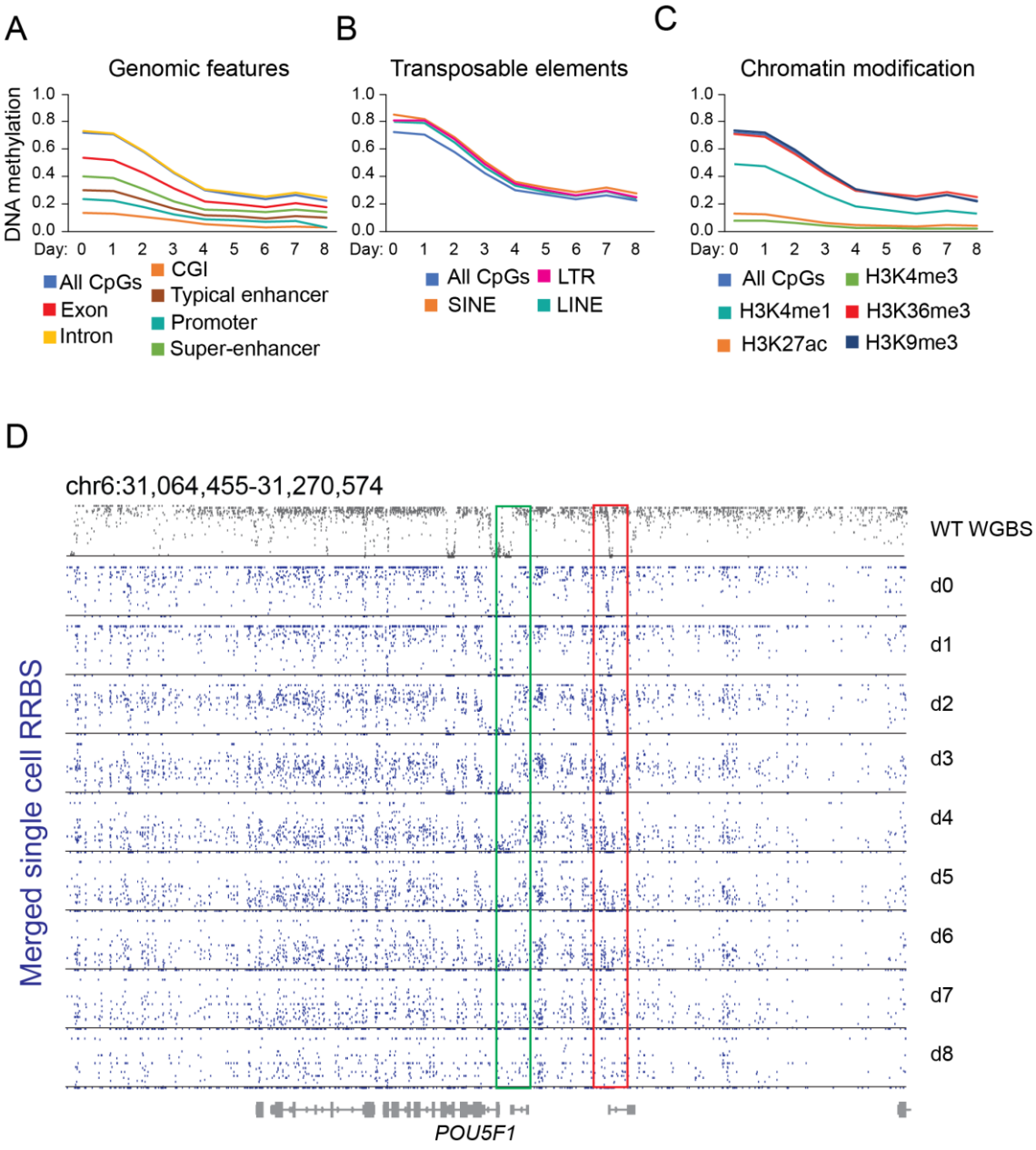


Figure 22. DNA methylation is lost proportionally across genomic features in DNMT1 deficient cells. ***A., B.,*** *and* ***C.*** *Mean DNA methylation level from pooled scRRBS data for each time point for listed genomic features, repetitive elements and DNA underlying chromatin marks, respectively.* ***D****. Representative snapshot of DNA methylation plotted across 200kb of the human genome (location listed above), including POU5F1 locus. The data was captured in WGBS*[200] *(in gray) or scRRBS (cells from each time point were pooled together). Each dot represents a single cytosine. Red box encircles an enhancer element, while the green box encircled the promoter.*

We used the pooled the scRRBS data to visualize the extend of the global DNA methylation decay (FIG 22D). A representative snapshot of POU5F1 locus shows the high global levels of DNA methylation in wild-type hESCs (gray; published WGBS data[200]) and untreated TetOFF DNMT1 line, punctuated by loci of low DNA methylation (promoter). Upon the depletion of DNMT1, previously highly methylated regions become intermediately methylated. Sharp dip in enhancer methylation (red box) became wider due to the methylation around being lost. The POU5F1 promoter (green box) maintained its unmethylated state (FIG 22D).

Altogether, consistent with previously published results[95], we observed that DNA methylation was lost proportionally at all the listed features, following the trend line of the bulk genome average. While the features with low DNA methylation lost less of the mark as compared to transposons, all of them retained the methylated cytosines at about 50% their respective initial levels. Indeed, only 22% of the autosomal CpGs are dynamically regulated throughout normal development[200], suggesting that the remaining CpG methylation could likely be maintained by DNMT1. This agrees with the DNMT1 role as a global maintenance methyltransferase[201].

Gene promoters. Subsequently, we studied the DNA methylation decay in the promoters of coding genes (FIG 23A). We hypothesized that since the very few promoters are highly methylated in hESCs[202], they could preferentially retain DNA methylation, but not be well-represented when plotting the average methylation of all promoters. Instead, we plotted the frequency of promoters bearing certain DNA methylation levels at each time point of the experiment. As previously reported, the majority of promoters were unmethylated, with very few promoters methylated above 80% in control cells (red line). Contrary to our hypothesis, the highly methylated promoters did not retain their methylation when DNMT1 was absent. Instead, the population of highly methylated promoters (>80%) was gradually reduced after day 5 (<20%; FIG 23A).

Because the majority of promoters for coding gene consist of CpG Islands, they display low levels of DNA methylation. In fact, the difference is so striking that they create dips of a local hypomethylation in the otherwise highly methylated genome[200]. The edges of these CpG Islands were named CGI shores and these are thought to be dynamically methylated by competition between de novo DNA methyltransferases (DNMT3A/B) and TET demethylases [72]. In order to examine if the distribution of DNA methylation within the promoter region was altered, we performed a metagene analysis centered around the transcriptional start sites (TSS) of all genes surrounded by 5kb upstream and 5kb

downstream sequences (FIG 23B). We found that the DNA methylation in TetOFF DNMT1 cells was reduced evenly throughout the experiment. We observed no preferential retention of the epigenetic mark up- or downstream of the TSS. A similar metagene analysis of all the exons and their surrounding regions displayed the identical trend of a gradual loss of DNA methylation with no qualitative change in its pattern (FIG 23C).

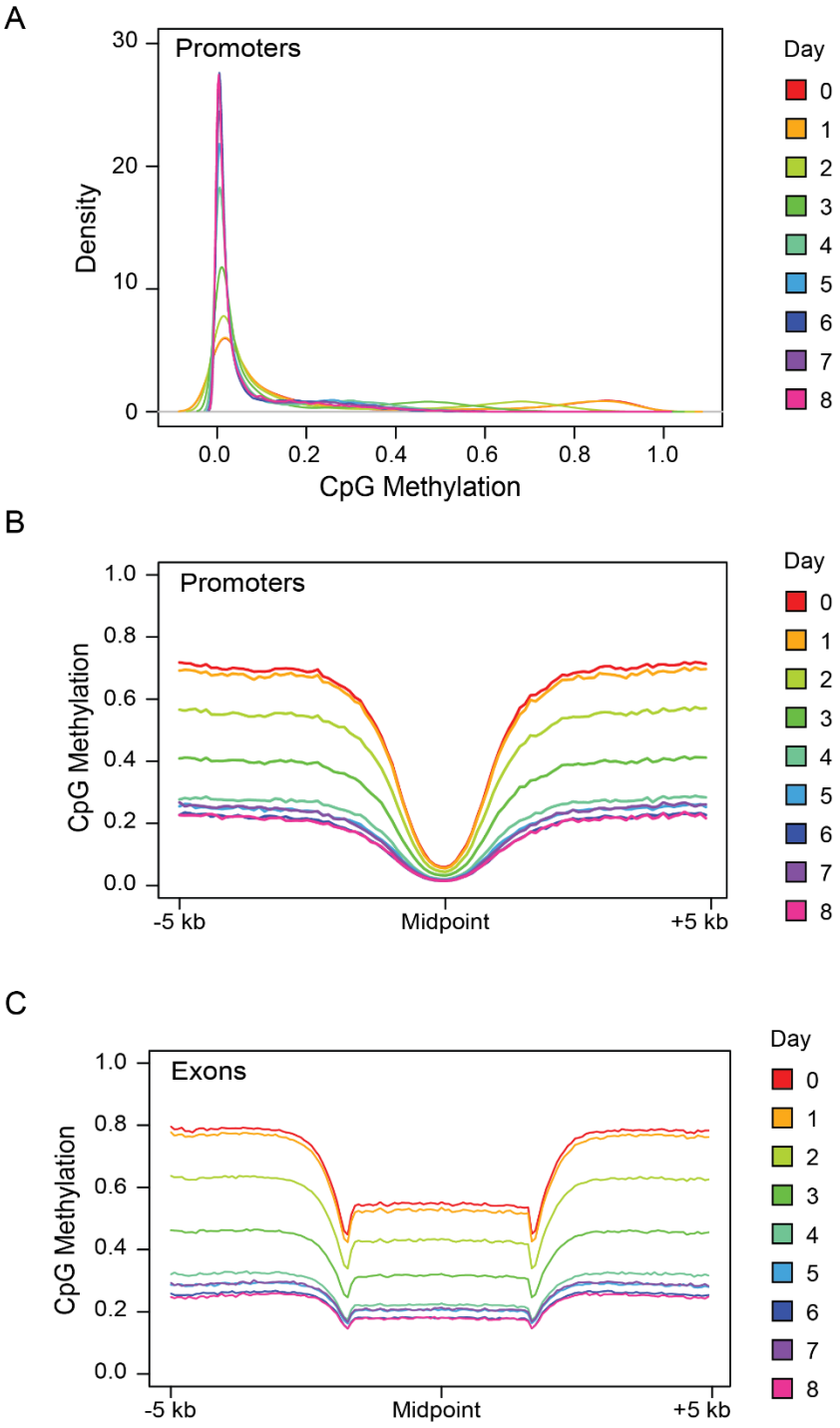
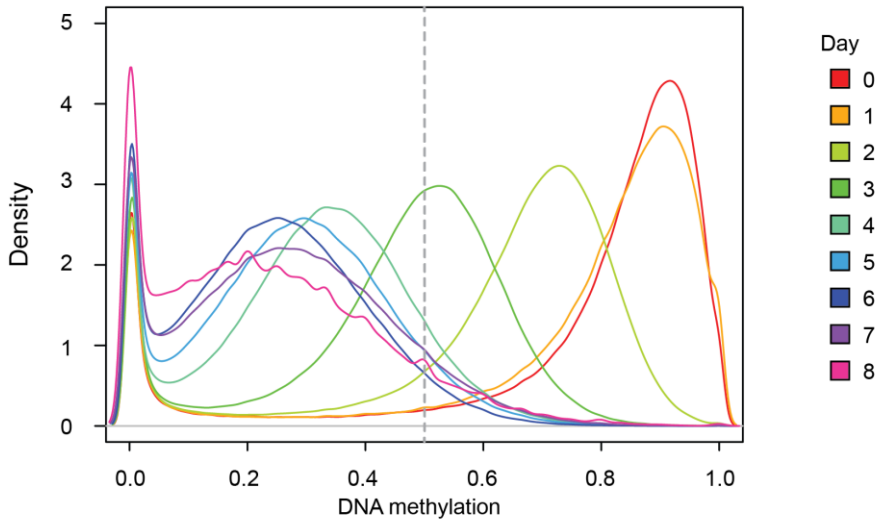


*Figure 23. The distribution of DNA methylation across promoters and exons in DNMT1-deficient hESCs**. A.** Density plot depicting DNA methylation across the promoters*

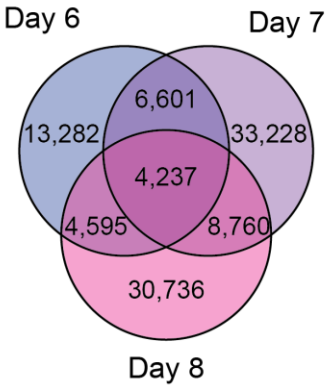
(± 1kb from the TSS, with at least 5 CpGs covered; n = 22,555) in pooled cells from each time point. ***B.*** *and* ***C.*** *Metagene analysis of DNA methylation of all promoters and all exons, respectively, including ±5kb region around.*

Overall, our in-depth analysis demonstrates the DNMT1-depleted cells did not preferentially maintain DNA methylation at any genomic features. It is possible, however, that there are certain genomic regions that are (1) not annotated or (2) covered by multiple annotations. To gain an unbiased view on the genome, we tested if DNA methylation was retained at any loci using a differential analysis. To this end we divided the human genome into 500bp tiles and kept only the tiles with five or more CpGs present (FIG 24A). Next, we asked if any of the tiles contained over 50% DNA methylation at the given time point. As expected, the number of such tiles decreased over time, but there were still some present even at late time points, between days 6 and 8 of DNMT1-depletion (FIG 24A). We asked if any of these methylated tiles are present in multiple final time points (FIG 24B). We have found 4,237 regions maintained (0.7% of all tiles) at days 6, 7 and 8, and 19,956 (3.5% of all tiles) that were maintained in at least 2 out of 3 time points. We asked what genomic features the overlapping tiles belonged to (FIG 24C). The majority of tiles was located in introns and they were mostly mapped to repeat elements (75%). The main repeat element was assigned to short interspersed nuclear elements (SINEs) (over 50% of tiles). Closer inspection revealed that most of these SINEs were Alu elements, the most abundant class of SINEs in the human genome(FIG 24D)[80]. Curiously, both evolutionary young Alu elements[203] (such as AluY) and more conserved (AluSx) were found with high DNA methylation levels despite the loss of DNMT1 (FIG 24E). Of note, our analysis used reads uniquely mapped to the human genome to assign their identity and genomic coordinates with high confidence. As most of the repeat elements are present in multiple copies in the human genome, we might only capture a fraction of highly methylated Alu elements. Future analysis using the newest sequencing technologies applying long sequencing reads for long single DNA molecule analysis will be required to determine the methylation at other repetitive elements.

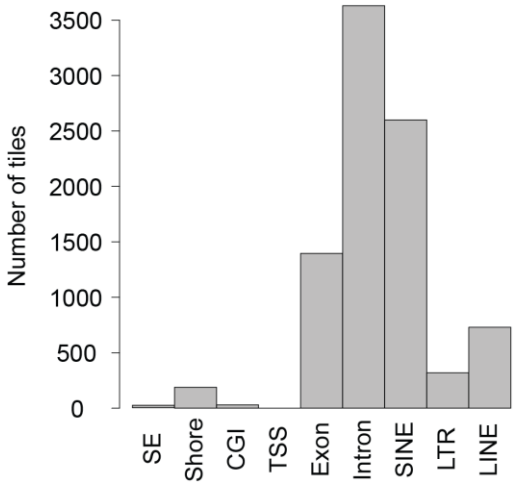
A



B



C



D

Family	Freq (retain tiles)	%
5S-Deu-L2	1	0.03
acro	0	0.00
Alu	1855	58.44
centr	1	0.03
CR1	16	0.50

Family	Freq (whole genome)	%
5S-Deu-L2	2550	0.05
acro	86	0.00
Alu	1238897	22.44
centr	2882	0.05
CR1	68182	1.24

E

Name	Element	Family	Freq
AluSx1	SINE	Alu	226
AluSx	SINE	Alu	216
AluSz	SINE	Alu	195
AluY	SINE	Alu	159
AluJb	SINE	Alu	134
AluSp	SINE	Alu	115
AluSq2	SINE	Alu	111
MIRb	SINE	MIR	91
AluJr	SINE	Alu	89

Figure 24. Preferential retention of CpG methylation in DNMT1-deficient hESCs. ***A.*** *Density plot depicting the methylation of the genome divided into 500bp tiles (n = 571,303 500bp tiles with at least 5 CpGs covered) in cells pooled from each time point of the scRRBS experiment. Bray dotted line represents 50% DNA methylation cutoff.* ***B.*** *Venn diagram depicting the overlap of 500bp genomic tiles from* ***A.****, which retained at least 50% methylation at designated time points. Number of overlapping tiles is listed. Overlap significance pval ≤10^{-270}* ***C.*** *The identity of genomic features that include the 50%≤ DNA methylated tiles overlapping at time points 6, 7 and 8. SE – super enhancers, CGI – CpG islands, TSS – transcriptional start site.* ***D****. Table depicting the frequency of top 5 classes of transposable elements overlapping with the DNA methylation-retaining tiles in the experiment (left side) and in the genome (right side).* ***E.*** *Table depicting the identity of top10 SINE elements most frequently overlapping with the tiles preferentially retaining DNA methylation.*

Why is the DNA methylation retained at these elements? One possible explanation could be that these particular elements were recognized by the TRIM28/ZFP machinery specialized in silencing transposons[51]. The TRIM28-based silencing system facilitates specific recruitment of de novo DNA methyltransferases. The odds of these elements being actively methylated could be affected by the fact that they lie mostly in the intragenic regions. Alternative a TRIM28-independent mechanism would involve a chromatin marker of transcribed genes, a histone modification H3K36me3, which is recognized by PWWP domain-containing proteins[204]. It was shown that DNMT3A and B both contain the PWWP domain that are necessary and sufficient to recruit the enzymes to DNA for its subsequent action[18]. Further experiment will need to be conducted to address the role of TRIM28 and the transcription process in the DNMT1-independent DNA methylation maintenance of SINEs.

6.17 Summary

In this study, we employed a single-cell sequencing method to measure the abundance and genomic distribution of DNA methylation in the DNMT1-deficient hESCs. In agreement with our previous 5mC ELISA results, we found that depletion of DNMT1 on the transcriptional level leads to a gradual, but not complete, loss of DNA methylation. By examining methylation levels in individual cells, we demonstrated that in the final time point of the experiment the low average levels (around 25%) of this mark in DNMT1-depleted cells are maintained across the population. The DNMT1-depletion phenotype is therefore robust, as we did not see cells that would retain their DNA methylation despite the absence of the maintenance methyltransferase.

Curiously, we have not captured any cell that would retain less than 10% of global DNA methylation, suggesting that, unlike mESCs, a complete demethylation of hESCs is deleterious. The alternative explanation could be that de novo methyltransferases DNMT3A/B are able to maintain this level of DNA methylation on their own. Single-cell sequencing data of even later time points than examined here (such as the day 12th), could further corroborate this hypothesis.

Through the analyses of (1) chromosomes, (2) regions or (3) tile-level DNA methylation, we were able to demonstrate that the distribution of retained DNA methylation was largely stochastic. The only exception was a small subset of elements, mostly belonging to the Alu TE family, which remained highly methylated (over 50% mean methylation), even up to 8 days post DNMT1 depletion. These regions were continuously methylated across examined time points suggesting the ongoing maintenance process in a DNMT1-independent manner. In such case, the maintenance could only be performed by the de novo methyltransferase enzymes DNMT3A and DNMT3B.

Our lab demonstrated that DNMT3A and 3B maintain DNA methylation on a set of genomic loci[95]. However, we did not observe changes on these loci in the absence of DNMT1. Our data could indicate that DNMT3A and 3B are only preferentially recruited to a few of their substrate loci for methylation, while the rest of their targets are the genomic regions maintained by DNMT1. In the absence of DNMT1, de novo DNA methyltransferases are presented with thousands of new substrate loci for DNA methylation, nevertheless the DNA methylation drops to low levels. We hypothesize that the de novo DNMTs abundance could be the limiting factor for maintaining global DNA methylation in the absence of DNMT1. An alternative explanation could be a lack of additional factor, which would be needed to recruit DNMT3A or 3B to the hypomethylated loci. Finally, the loss of DNA methylation could be followed by transcriptional activation which, in turn, prevents the establishment of *de novo* DNA methylation. Further studies will offer some other possible interpretations of this paradigm. Indeed, the overexpression of DNMT3B in mouse cells was found to cause DNA hypermethylation[205], reinforcing the hypothesis that the abundance of *de novo* methyltransferases in a cell could be a mechanism to maintain global DNA methylation in the absence of DNMT1. It would be interesting to study, whether overexpression of DNMT3B could substitute for DNMT1 by maintaining higher levels of global DNA methylation.

6.18 Single-cell transcriptome

The transcriptional state determines the identity of the cell and reacts to intracellular and extracellular signals. Because of the role of DNA methylation in repressing transcription, we hypothesized that a genome-wide loss of DNA methylation will impact the transcriptome of the cells. Previous studies of the expression changes in DNMT1KO cells[98,102,176] reported different extent to which the transcriptomes were affected by the global DNA hypomethylation. This could be ascribed to many factors, such as differences between cell types and sequencing methods used. With our dual-sequencing protocol, we analyzed the transcriptome of single cells, as well as the relationship between gene methylation and expression. We have also pooled scRNA-seq data from cells to compare changes in gene expression across time points.

With our dual-sequencing protocol, we had the unique ability to analyze not only DNA methylation but also the transcriptome of single cells that gave us a relationship between the gene methylation and expression. Therefore, we have the pooled scRNA-seq data from cells to compare changes in gene expression across time points.

The tSNE clustering. In order to visualize the differences between transcriptional states of single cells, we utilized tSNE clustering, which is an analysis method that groups cells based on the similarities between their transcriptomes (FIG 25). While, in this method the composition of the transcriptomes is the most relevant to assign single cells into clusters, the distances between clusters is less crucial. We then colored the plot based on different criteria, such as the time point of collection (FIG 25, 26), number of mapped reads, number of expressed genes or mean DNA methylation. We did not observe technical bias introduced by these variables that could affect the clustering.

Next, we examined how biological factors, such as the time point of collection post-DNMT1 depletion or the mean level of DNA methylation correlate with the clustering. We observed distinct clustering of cells from time points 0 and 3, which formed one cluster. The cells from time points 1 and 2 formed a separate cluster. In contrast, the cells from later time points (5 through 8) formed three distinct clusters. These results suggest that later time point hypomethylated cells are more similar to one another than to the cells at earlier time points (days 1 and 2), or the control. We did not observe clustering of cells according to the mean level of DNA methylation (FIG 25B). For example, cells from the highly methylated time point 0 cluster with the cells from time point 3, which exhibit a variable amount of DNA methylation. These results suggest that the length of DNMT1

deficiency period is a better predictor of transcriptional cell-to-cell similarity than the amount of methylation retained by the individual cells.

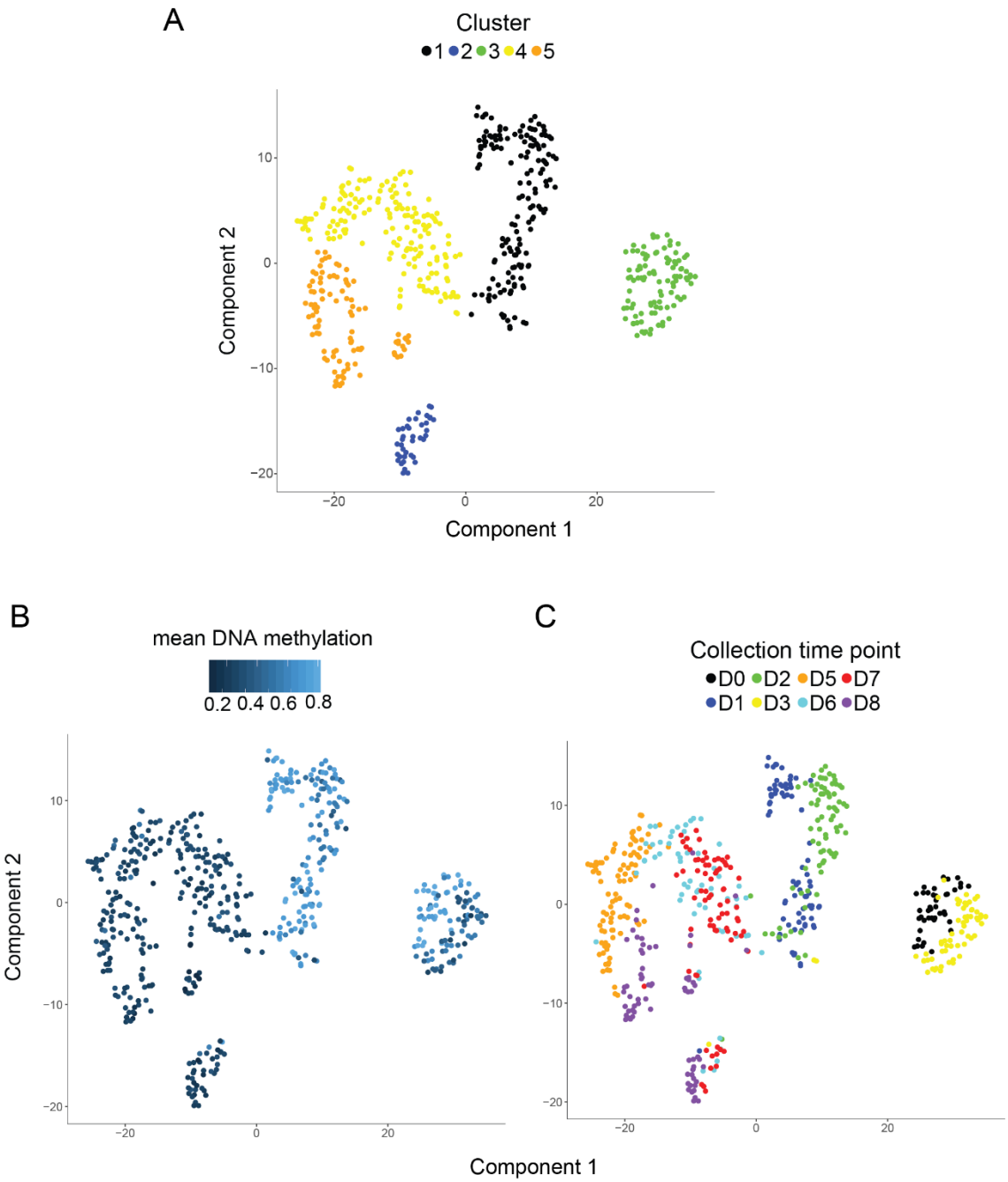


Figure 25. tSNE clustering of DNMT1-deficient hESCs based on their transcriptomes. ***A.*** *Unsupervised clustering of TetOFF DNMT1 hESCs collected in the scRNAseq experiment.* ***B.*** *tSNE clustering colored by mean DNA methylation of each cell.* ***C.*** *tSNE clustering colored by the collection time point.*

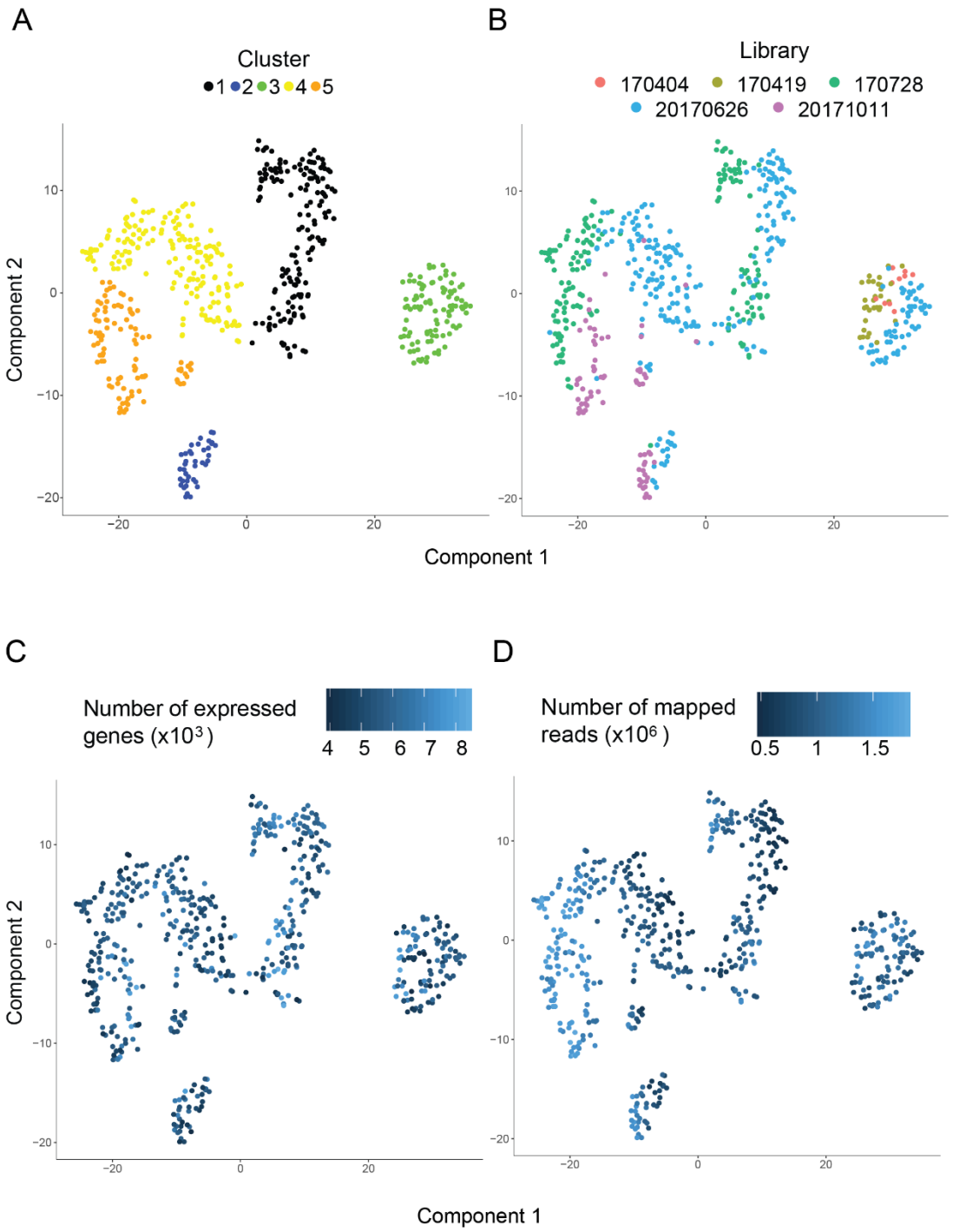
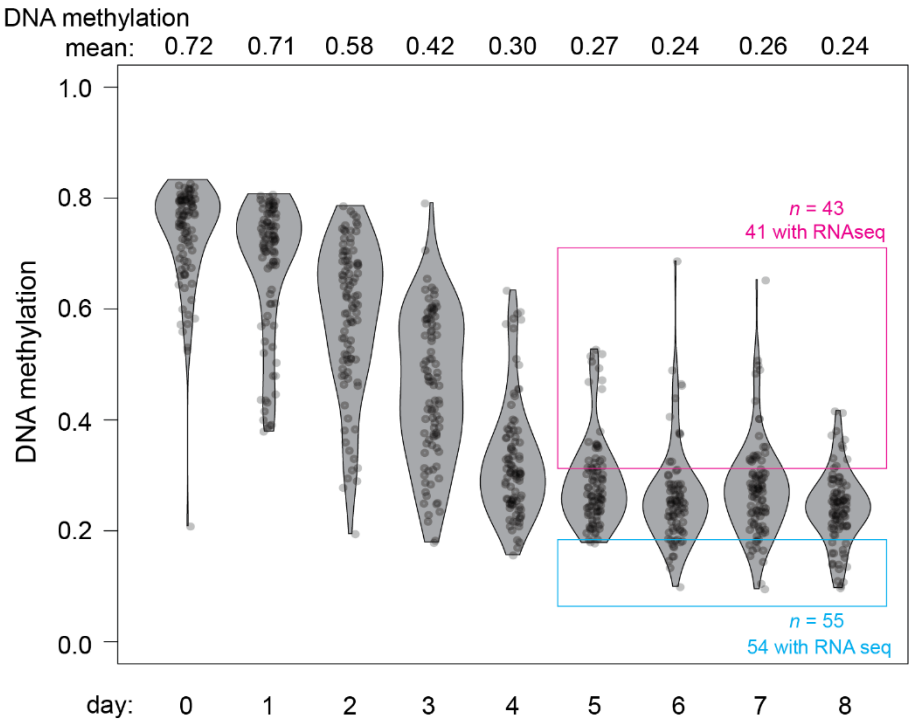


Figure 26. tSNE clustering of DNMT1-deficient hESCs does not depend on technical variability. ***A.*** *Unsupervised clustering of TetOFF DNMT1 hESCs collected in the scRNAseq experiment.* ***B.*** *tSNE clustering colored by library run in which the cells were prepared and sequenced.* ***C.*** *tSNE clustering colored by number of genes expressed per cell.* ***D.*** *tSNE clustering colored by the number of mapped reads.*

6.19 Identifying transcriptional changes occurring in response to the loss of DNMT1

In order to identify any DNA methylation level-driven changes, we analyzed differentially expressed genes between the highly and lowly methylated cells from time points 5-8. We compared the transcriptomes of cells retaining the highest mean DNA methylation (more than 35%) to the least methylated ones (less than 20% average DNA methylation; FIG 27). Interestingly, no genes differed significantly in their expression between the two groups (representative examples shown in FIG 27B). We then concentrated on the transcriptomic differences between time points to define their possible effect on cell fitness. To this end, we explored the differences between the transcriptomes of pooled cells at each time point of the experiment (1-8) to the control cells (time point 0). The number of differentially expressed genes (DEGs; at least 2-fold difference in the average expression between the time points and a qval≤0.05) is shown in FIG 28. We observed that the number of upregulated genes is higher than the number of genes that become downregulated for each of the time points. Both day 1 and day 2 exhibited around 300 DEGs each. While we only saw 24 DEGs at day 3, consistent with cells from that time point clustering closely with control cells, cells at day 5 again exhibited an increased transcriptional deregulation. Moreover, the longer the cells were hypomethylated, as seen in the case of time points 5 through 8, the more DEGs were identified. The lists of top 100 differentially expressed genes can be found in Table 8.

A



B

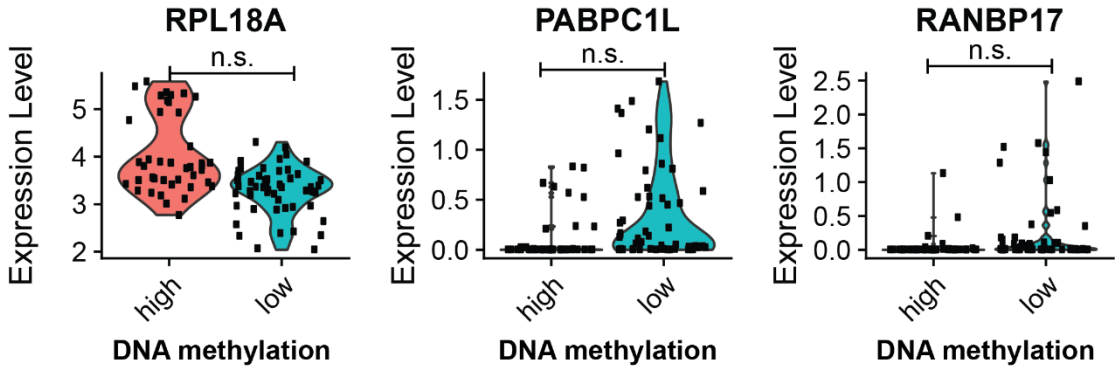


Figure 27. Expression in most and least DNA methylated cells from DNA hypomethylated time points. ***A.*** *Violin plot depicting mean DNA methylation of cells at the experimental time points post DNMT1 depletion. Pink rectangle encircles the top highly methylated cells (mean methylation >0.35). Blue rectangle encircles the most hypomethylated cells (mean methylation <0.2). The number of cells and cells also covered in the scRNAseq analysis is listed.* ***B.*** *Violin plots depicting representative examples of differentially expressed genes.*

In order to explore the function of the differentially expressed genes, we sub-divided them into classes of functionally related genes. To this end, we performed gene ontology (GO) analysis[125,126,206] on the separated groups of up- and downregulated genes at each

time point. Top 10 GO terms with the lowest FDR for the gene sets are shown in Table 9. Since day 3 only presented 24 DEG in total, it was not significantly enriched for any of the terms and is not represented in the figure.

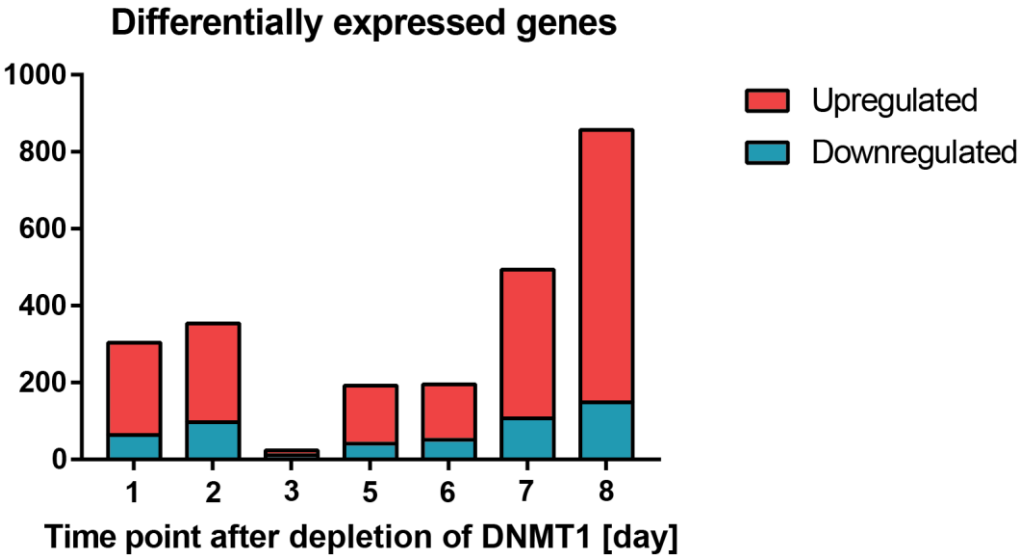


Figure 28. Differentially expressed genes. Bar graph representing the number of differentially expressed genes between pooled scRNAseq data from control cells and each experimental time point. Up- and downregulated genes are colored in orange and blue, respectively.

Transcriptional downregulation in the DNMT1-deficient cells. We first focused on the analysis of downregulated genes that came out unexpectedly given the negative regulator role of DNA methylation. We found the analyzed time points were enriched in several categories, such as Signal Recognition Particle (SRP)-dependent co-translational protein targeting to membrane, viral process, or nonsense-mediated decay (NMD). It turned out that all these processes were actually significantly enriched because of the same group of genes ascribed to multiple categories. The inspected genes belonged to Ribosomal Protein Genes (RPGs), a family of highly conserved proteins that, together with ribosomal RNAs (rRNA), make up the ribosome. There are nearly 80 RPGs in the human genome[207] and a recent publication has sub-divided RPGs based on profiles of expression in different tissues and in cancer cells, suggesting that subsets of the ribosomal protein genes have a common regulatory mechanism[208]. We have compared the RPGs downregulated across early (days 1 and 2) and late (days 6 – 8) time points with the published RPG classification. The RPGs downregulated in our

experiment did not belong to tissue-specific or cancer-downregulated classes defined in the previous study[208] (FIG 29).

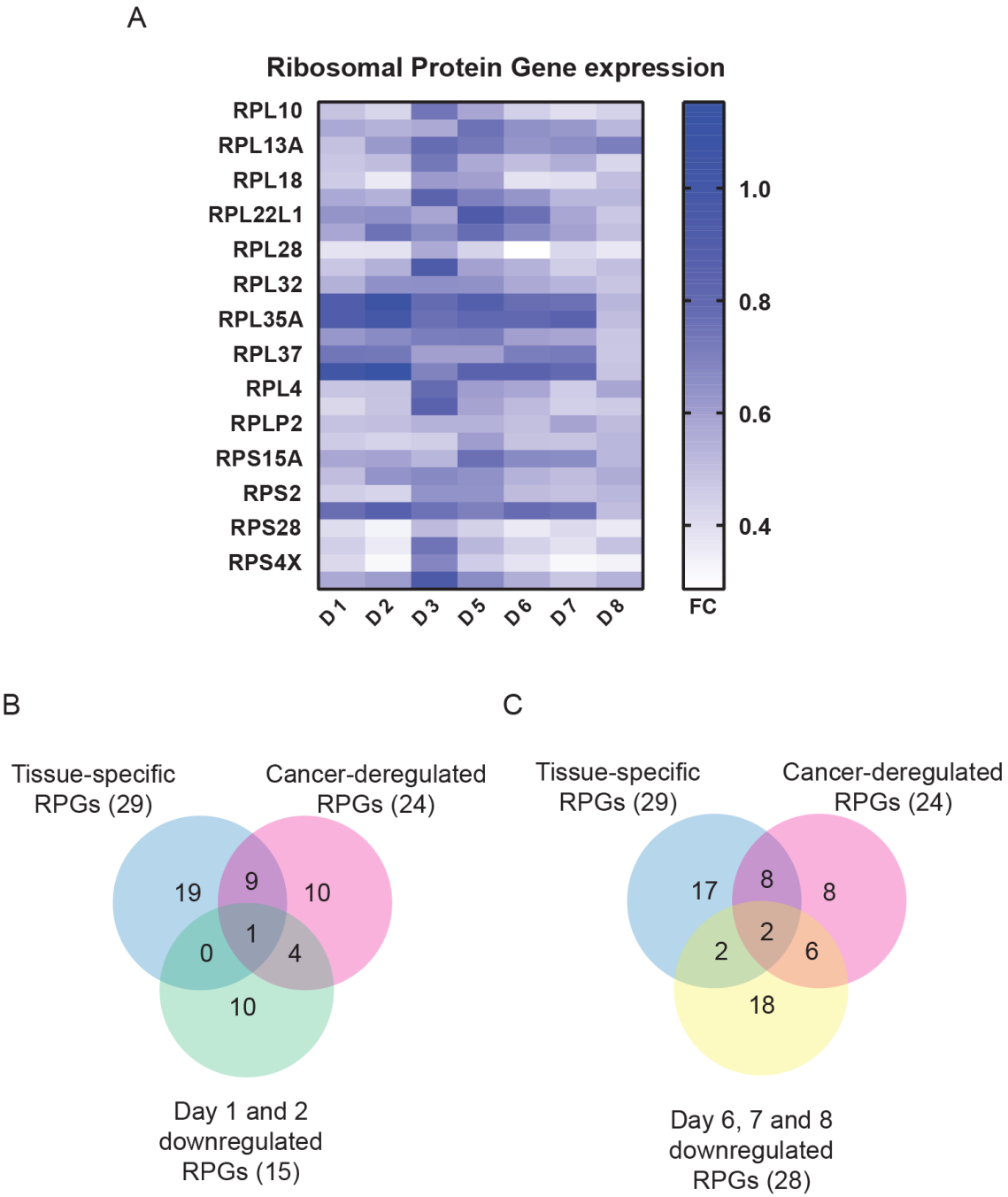


Figure 29. Downregulated Ribosomal Protein Genes do not belong to a single expression group. ***A.*** *Heat map representing the expression of downregulated RPGs.* ***B.*** *and* ***C.*** *Venn diagrams depicting the overlap between previously distinguished[208] subgroups of ribosomal protein genes (in blue and pink) and RPGs downregulated at day 2 (B.) or day 8 (C.) in the DNMT1-deficient hESCs.*

Upon the removal of RPGs from the downregulated gene lists, a few general metabolic terms were found to be statistically significant. We investigated the roles of these genes in metabolism and found several housekeeping genes[209] that became downregulated, however, they did not belong to a single metabolic pathway. For example, we observed a downregulation of PGK1 (a glycolytic enzyme), tubulin (TUBA1B, which codes for a component of microtubules) or PPIA, a peptidyl prolyl *cis/trans* isomerase, a chaperone. It is worth noting that none of the housekeeping genes was completely silenced, however, they underwent a statistically significant 2- to 3-fold reduction in their expression.

6.20 Transcriptional upregulation following depletion of DNMT1.

Pro-apoptotic BAX gene. We next moved to study the genes upregulated in the DNMT1-deficient cells that are conceptually in line with the repressive role of DNA methylation in transcription. This pro-apoptotic protein is able to dimerize, open channels in the mitochondrial membrane to cause cytochrome C release into the cytosol and, in consequence, trigger the canonical apoptosis pathway[210]. The process is dependent on BH3-only family of proteins, which activate BAX. Another mitochondrial membrane protein called Bcl-2 acts as a safeguard. Upon BAX activation, Bcl-2 can inhibit mitochondrial membrane perforation by directly binding BAX[211]. Only after Bcl-2 is completely bound, excess of active BAX can lead to the activation of apoptosis[212]. BH3-only family of proteins are activated by p53[213]. In the previous chapter we were able to demonstrate that TP53 knockout is able to partially rescue the DNMT1 absence-caused loss of fitness (FIG 14). If spontaneous BAX activation occurred in the DNMT1-depleted hESCs, the knockout of upstream factor (p53) would not confer rescue.

The common upregulated genes in DNA hypomethylated cells. We continued the analysis of upregulated genes at other time points, again relying on the gene ontology analysis. Both day 1 and day 2 upregulated genes were enriched for GO terms related to metabolism. The fact that the transcriptome at day 3 was again more resembling the control argues against a differentiation program being triggered in response to the loss of DNMT1 protein. We hypothesize that this transient gene deregulation could be a response to the addition of doxycycline to medium or the cell plating conditions which, in order to maintain good attachment efficiency and survival, included treatment with Rho kinase inhibitor (ROCKi) for 24h after plating. Ideally, a second control consisting of dox-untreated cells plated in the same way as the treated cells could confirm this hypothesis.

We then compared the differentially expressed genes conserved across the time points and asked what were the constantly upregulated genes. Out of 151, 144, 357 and 707 genes upregulated with respect to control for days 5,6,7 and 8; 64 were upregulated in all time points (FIG 30).

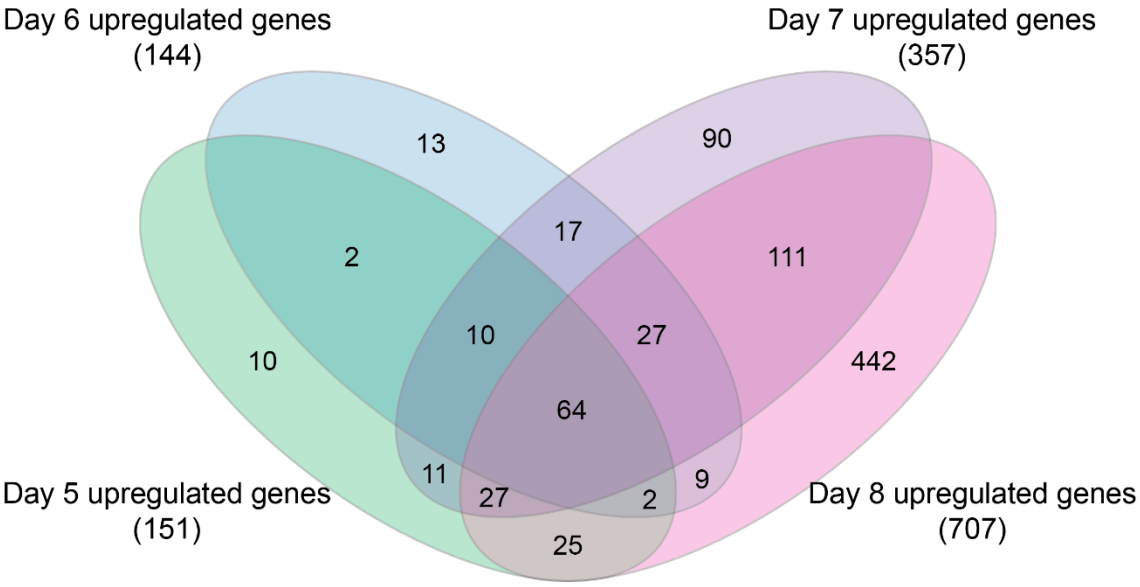


Figure 30. Overlap of Differentially Expressed Genes in globally DNA hypomethylated hESCs. Venn diagram depicting the overlap between DEGs from time point 5 (green), 6 (blue), 7 (purple) and 8 (pink). The number of genes in each overlap is indicated.

We have also explored genes upregulated at earlier time points (d5 and d6) and later time points (d7 and d8). The lists of top 100 upregulated genes common for designated time points can be found in Table 10. The upregulated genes common for all time points between day 5 and 8 included transcripts present only in a variety highly specialized tissues, such as MYL7 – myosin light chain specifically expressed only in the heart muscle. Another example is the upregulation of neurotensin (NTS) that encodes for a neurotransmitter. A different category found in the upregulated genes included metabolic genes which are highly expressed in most tissues, such as the carbonyl reductase 1 (CBR1), an NADH-dependent enzyme with wide substrate specificity. Interestingly, a common group of genes significantly upregulated in the examined time points were morphogens that belong to the Activin/Nodal signal transduction pathway – LEFTY1, CER1 and FST. The expression of LEFTY1 lies downstream of NODAL-mediated signal transduction and we were able to confirm NODAL upregulation at day 3, preceding LEFTY1 upregulation. The analysis of genes upregulated at only at the earlier time points (day 5 and day 6) showed similar upregulation of a few widely expressed and specialized

tissue genes, such as leucine tRNA synthetase (LARS1) and cranio-facial E3 ubiquitin ligase NOSIP, respectively. The latest time points (day 7 and day 8) were also enriched in metabolic and specialized tissue-specific genes, like cystatin (CST; a protease inhibitor present in saliva and other secretions). We also observed an upregulation of gonad-specific genes, such as the transposon silencer MAEL or a meiotic protein, called SYCP3. In line with the germline-specific coding genes becoming upregulated, we observed an upregulation of germline-specific long, non-coding RNAs (lncRNAs) from the XAGE, MAGE and GAGE families. Although numerous germline genes were derepressed, we did not observe the expression of the canonical germline markers (such as DPPA3 or DDX4), which suggests that the cells did not undergo differentiation into the germline lineage.

Previous studies conducted in diverse cell types reported that DNMT1KO leads to the deregulation of germline-specific genes, since the expression of some gonad-specific genes is inversely correlated with the promoter DNA methylation, which is a negative regulator of gene expression[91]. We aimed to find out if more genes in the DNMT1-deficient hESCs display similar expression dynamics. To explore this possibility, we identified genes, which lose DNA methylation at their promoter regions (d0 methylation is higher than d8 methylation by at least 40%) and belong to the genes upregulated at the d8 category. Both mean methylation and expression of such genes is plotted in FIG 31. Indeed, for most germline-specific genes (including MAEL and SYCP3) the expression increases after promoter methylation has decreased. The upregulation of a few specialized tissue-specific genes (such as MYL7) also correlated with loss of the repressive DNA methylation from their promoter region. Nevertheless, less than 70 genes identified in our analysis exhibited the trend of transcriptional upregulation following promoter hypomethylation. This is consistent with the idea that not only the absence of the repressive chromatin mark but also a positive signal, such as the presence of correct transcription factors is necessary for transcriptional activation.

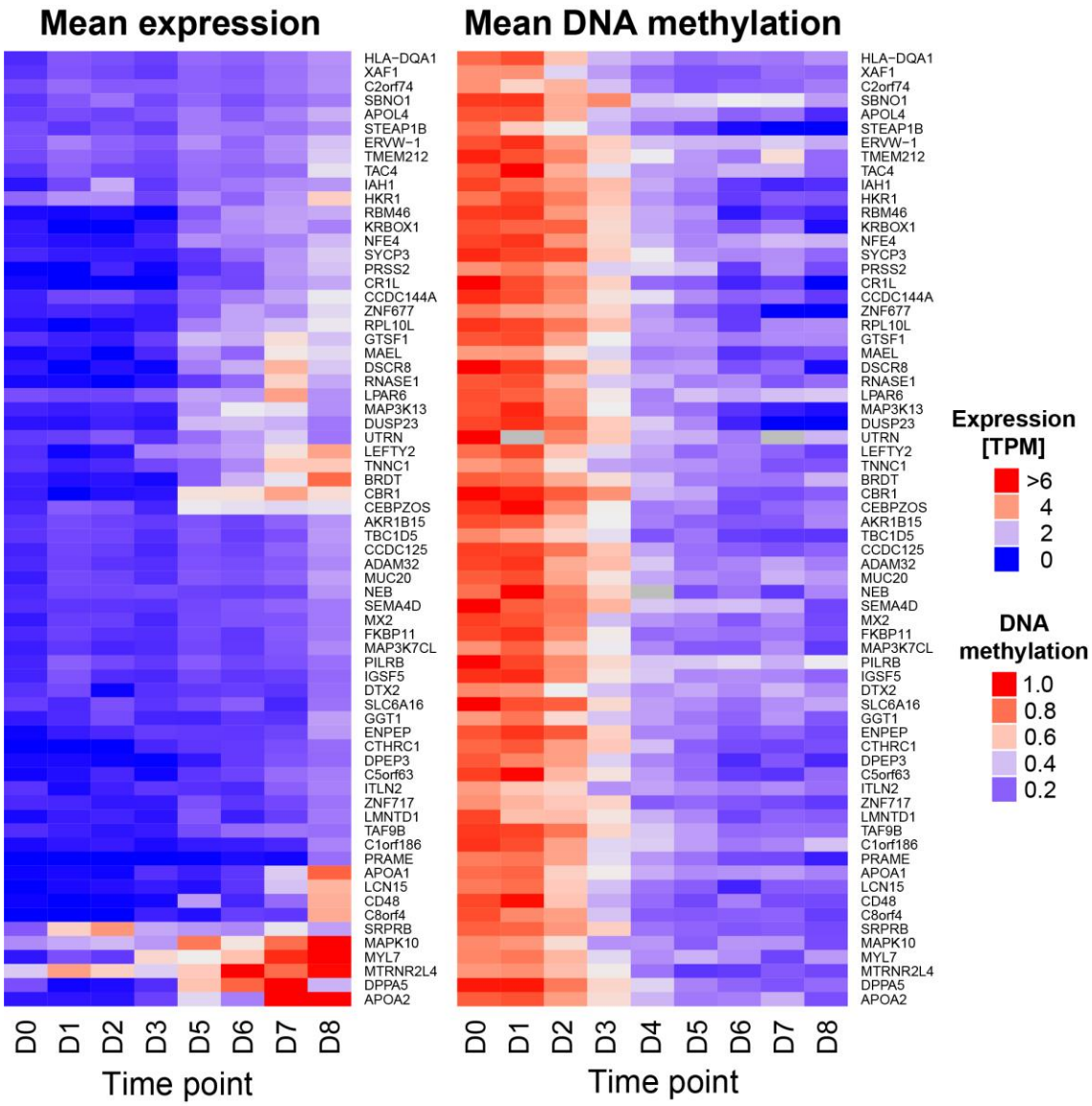


Figure 31. Genes in which promoter methylation negatively correlates with expression. Heat maps representing the expression of indicated gene in all cells in the designated time point (left) or mean DNA methylation of their promoter (right).

6.21 Summary

We purified polyadenylated RNA from the single cell-sorted hESCs and processed it concurrently to the genomic DNA to obtain both transcriptome and DNA methylome information from single cells. We clustered the cells based on their gene expression profiles and found out that our cells clustered into three main groups: (1) Day 3 cells showed few genes that were differentially expressed with regards to control cells and thus clustered with time point 0. (2) Co-clustering of day 1 and day 2 cells, which was most likely caused by a response to passaging, ROCK inhibitor and doxycycline common

to these time points. (3) Cells collected at day 5 and onwards clustered separately and showed a most severe difference in gene expression compared to the control cells out of all the experimental time points.

We analyzed the differentially expressed genes between cells belonging to different time points and control cells to identify how the loss of DNMT1 affects the transcriptional output of the cells. The DNA hypomethylated cells from time points 5- 8 demonstrated a progressively increased number of DEGs, suggesting that even when the DNA methylation is brought down to a minimum (a mean of about 25%), the changes in transcriptome are progressing. Upon investigation of the upregulated gene lists, we were able to identify certain gene expression patterns. The upregulated genes in DNA hypomethylated cells (time points 5-8) could generally be assigned to one of the three categories: (1) widely expressed metabolic genes; (2) specialized tissue-specific genes, which belong to diverse lineages; (3) gonad-specific transcripts. The deregulation of a large number of genes seemed to be mostly stochastic; the loss of DNA methylation at the promoter region directly preceded the activation of only a few genes.

Taken together, our scRNAseq results identified some of the known genes upregulated in response to DNA hypomethylation, as well as other genes, the upregulation of which could be a hESC-specific response to global DNA hypomethylation. In the next chapter, we follow up on the changes in a few of these newly-identified DEGs, which belong to the Activin/Nodal signal transduction pathway.

CHAPTER 4. Loss of DNMT1 increases sensitivity to the stimulus-dependent activation of transcription.

Our transcriptomic analysis revealed that DNMT1 deficiency resulted in deregulation of gene expression. Furthermore, each consecutive day after ablation of DNMT1 yielded more differentially expressed genes with respect to controls. Among the upregulated genes, we found the Activin/Nodal signal transduction pathway. Since the activity of this pathway is essential for the pluripotency of hESCs[214,215], we decided to explore how the imbalance in the expression of the Activin/Nodal pathway impacts the growth of hESC, and whether it contributes to the DNMT1-deficiency growth defect.

6.22 Characterization of the TGFβ superfamily Activin/Nodal signal transduction pathway.

Morphogens are molecules which act proximally and distally to orchestrate embryo patterning[216]. Because of their mode of action, they are typically secreted and act through binding to an extracellular receptor. Some of the most studied morphogens belong to the Transforming Growth Factor Beta (TGFβ) superfamily, which includes the Activin/Nodal, BMP and GDF signal transduction pathways[217].

In mammals, the TGFβ/Activin/Nodal signal transduction pathway is activated by the following morphogens: TGFβ, Activin A, Activin B or NODAL. Activin A/B are recognized by heterotetrameric receptors consisting of two subunits of each ActRIIB and Alk4[218]. NODAL requires a co-receptor CRIPTO to enable signal transduction via ActRIIB and Alk4[219]. Therefore, both Activin A/B and NODAL converge on the same ActRIIB and Alk4 receptors. On the other hand, TGFβ is recognized by a distinct heterotetrameric receptor consisting of TβRII and Alk5[220].

The binding of morphogens to their respective receptors results in the phosphorylation of receptor-associated Small Mothers Against Decapentaplegic (SMAD) proteins SMAD2 and SMAD3[221]. Upon phosphorylation, these factors bind a common mediator-SMAD (co-SMAD), called SMAD4, and translocate to the nucleus to activate downstream target genes[222]. The target specificity is confined by distinct DNA sequence-specific transcription factors modulating the binding of SMADs, such as p53, FoxH1 or OTX2[223,224]. In this way, the Activin/Nodal signal transduction pathway activates gene expression in a cell non-autonomous manner. One of the downstream targets of the pathway is the Nodal gene itself, which creates a positive feedback loop to amplify the initial input signal in this signaling pathway. This mechanism exists in the early embryo to maintain the high levels of NODAL protein in the epiblast and sustain its growth and identity[225]. Other genes activated downstream in a response to the NODAL stimulus are genes, the protein products of which act as direct inhibitors of NODAL: Lefty homologs (LEFTYA and B)[226,227] and Cerberus (CER1). CER1 functions as a broader inhibitor of Nodal, BMP and Wnt[228] signal transduction pathways. This creates a negative feedback loop, which in turn restricts the spatial activity of NODAL.

In response to TGFβ/Activin/Nodal stimuli, a wide range of genes change their expression pattern. This transcriptional response downstream of the stimuli controls the viability of the epiblast cells and primitive streak formation (*in vivo*), or hESCs survival and meso- and endoderm differentiation (*in vitro*)[229–231]. We therefore explored if the

perturbations in morphogen expression could contribute to the loss of fitness phenotype in the DNMT1-deficient cells.

6.23 Nodal signal transduction pathway components in cells deficient for DNMT1

We re-visited our results of scRNAseq to investigate the expression of genes acting in the Activin/Nodal signal transduction pathway. We observed that loss of DNA methylation coincided with the deregulation of nearly all the morphogens in the Nodal signal transduction pathway (*NODAL, FST, LEFTY1, LEFTY2, CER1*) at different time points of the scRNAseq analysis. The only factor not deregulated was the gene encoding for Activin A, which is not expressed in hESCs (Fig 32A). The expression level of Follistatin (*FST*), a direct inhibitor of Activin[232], increased as soon as day 1 after the transcriptional shut off of DNMT1 and remained upregulated throughout the experiment (FIG 32A). In contrast, *NODAL* was upregulated as soon as day 3 post DNMT1 depletion. Indeed, *NODAL* was one of the few differentially expressed genes between control and day 3 cells (Table 8). In agreement with *LEFTY1* being a direct target of the NODAL-induced signal transduction, the expression level *LEFTY1* increased at day 5 and onwards, following the surge in *NODAL* expression (FIG 32A).

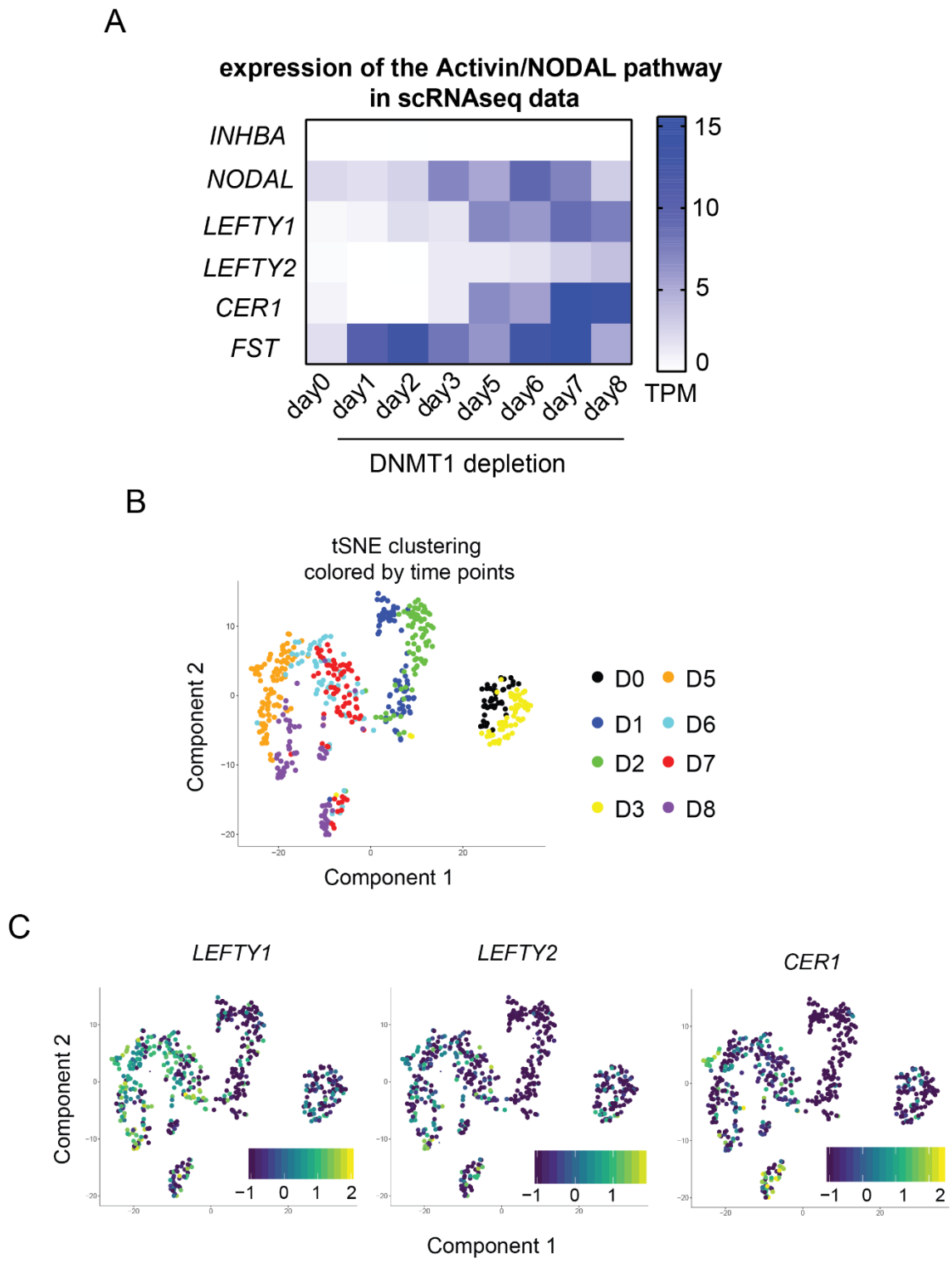


Figure 32. scRNAseq expression of Activin/Nodal signal transduction pathway morphogens. ***A.*** *Heat map representing the expression value [TPM] of chosen genes of the Activin/Nodal signal transduction pathway. INHBA – Activin A.* ***B.*** *tSNE clustering of cells used in the DNMT1-depletion scRNAseq colored by time point, for reference to* ***C.*** *tSNE plot of scRNAseq results colored by the expression of chosen morphogens. Scale bar: $log_{10}(TPM+0.1)$.*

To elucidate if the changes in expression were present in the most DNA hypomethylated cells, or if a small number of cells highly upregulated these factors, we examined the expression of NODAL inhibitors in the individual cells (FIG 32B and C). While nearly all cells from time points 5-8 expressed *LEFTY1*, the two remaining inhibitors of NODAL - *LEFTY2* and *CER1* became expressed as late as at days 7 and 8 and in smaller subset of cells. Nevertheless, the upregulation in these few cells significantly raised the average expression for the respective time point (Fig 32C and 33A).

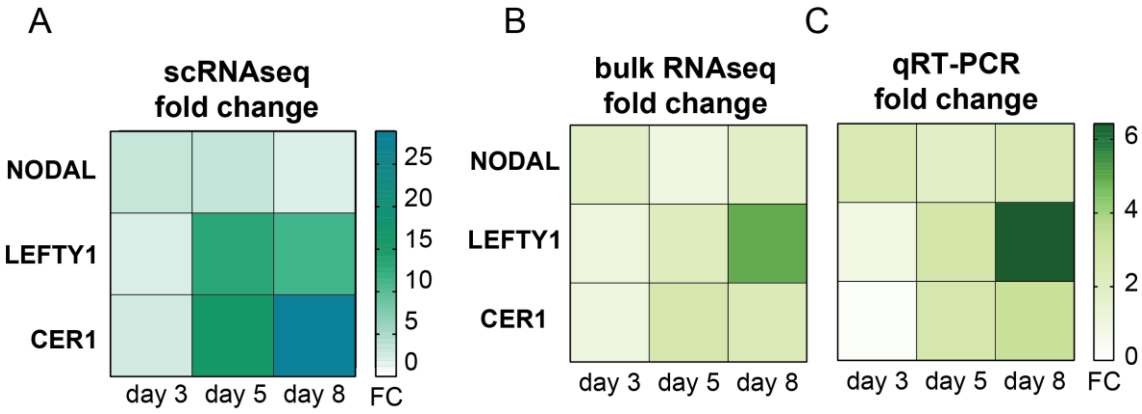


Figure 33. Expression of chosen morphogenes in DNMT1-deficient hESCs. ***A.*** *Heat map representing the fold change in expression of NODAL, LEFTY1 and CER1 in scRNAseq hESC,* ***B.*** *bulk RNAseq and* ***C.*** *qRT-PCR (N=3).*

In order to corroborate these observations and gain further insights in gene expression, we performed RNAseq on the bulk population of DNMT1-deficient hESCs at the crucial time points of day 3, day 5 and day8 post DNMT1 depletion. To avoid naming confusion with the single-cell study, this will be referred to as “bulk RNAseq” (FIG 33B).

We observed a similar trend of changes in the two sequencing analyses, when looking at the expression patterns of *NODAL* and *LEFTY1.* For example, we found ~2-fold *NODAL* upregulation at day 3, followed by subsequent increase in the expression of *LEFTY1* at day 5, which was elevated further (5-6 fold) at day 8. Conversely to our scRNAseq results, which showed a 25-fold upregulation of *CER1* at day 8, we observed a twofold upregulation of that gene in the bulk RNAseq study (FIG 33B). This analysis was further confirmed by qPCR measuring the transcription of the morphogens of interest (FIG 33C). As *CER1* was highly expressed in a small subset of cells in scRNAseq, it could have been underrepresented in the bulk sequencing. Alternatively, the variation between the two approaches could stem from normalization methods during data processing. Nevertheless, both data sets agree on the trend of the change.

Altogether, our analyses show that the upregulation of *NODAL* is followed by the upregulation of *LEFTY1* after depletion of DNMT1 in hESCs.

6.24 Morphogen deregulation in the absence of Activin/Nodal signal transduction

We next examined, whether the global DNA hypomethylation is (1) sufficient for the deregulation of NODAL and its antagonists and (2) if the signaling via SMADs is necessary in this process. Since hESCs rely on the SMAD2/3-mediated signaling for survival[215], we genetically engineered the degron into the DNMT1 locus in Human Embryonic Kidney (HEK293T) cells, (developed by Simon Lauer); and utilized the PROTAC-based degDNMT1 strategy to induce small-molecule mediated depletion of DNMT1 (Described in chapter 2.2; further referred to as HEKdegDNMT1 line). Addition of dTAG13 small molecule to these engineered HEKdegDNMT1 cells resulted in depletion of the DNMT1 protein in the western blot assay (fig 34A, clone #3).

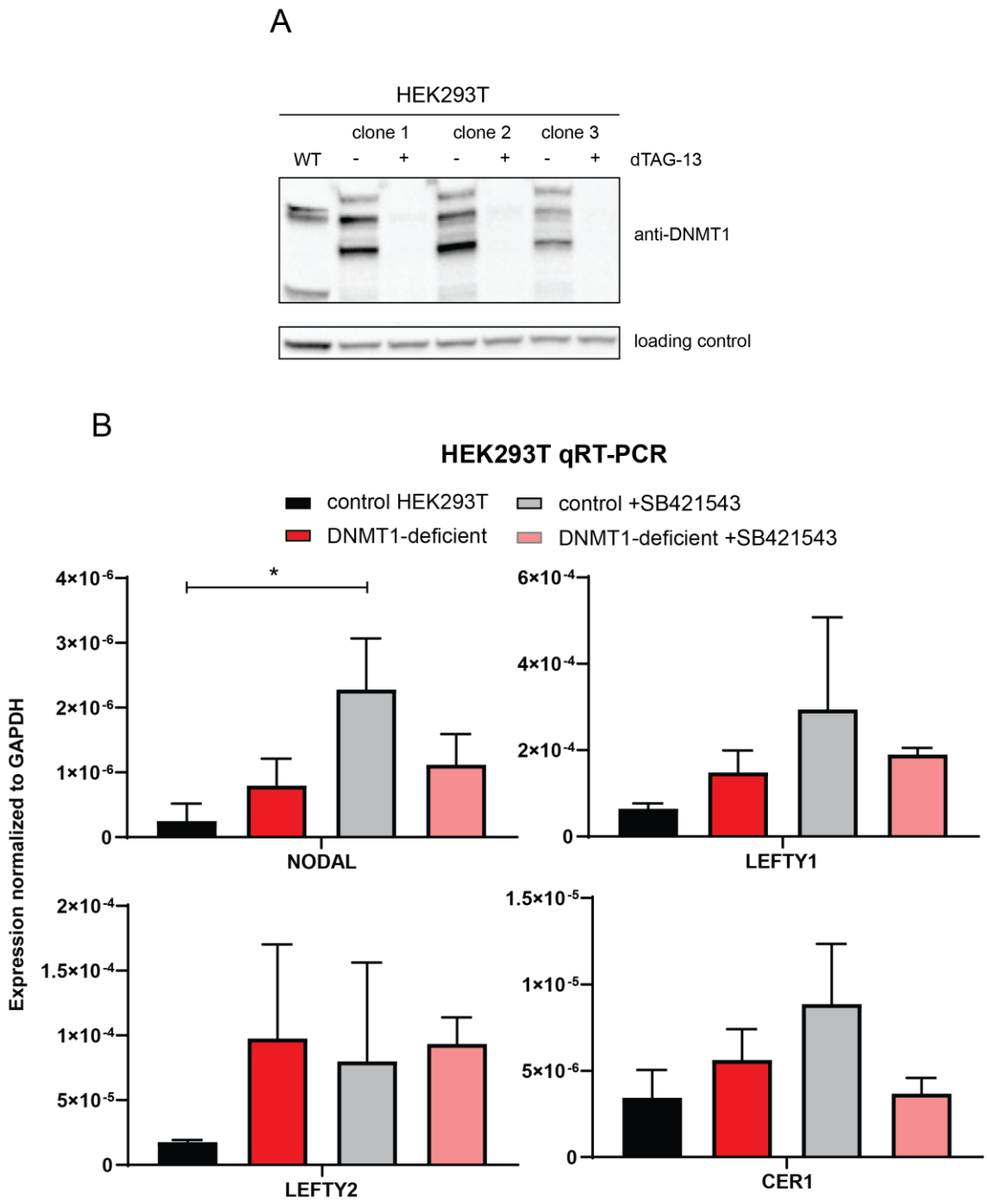
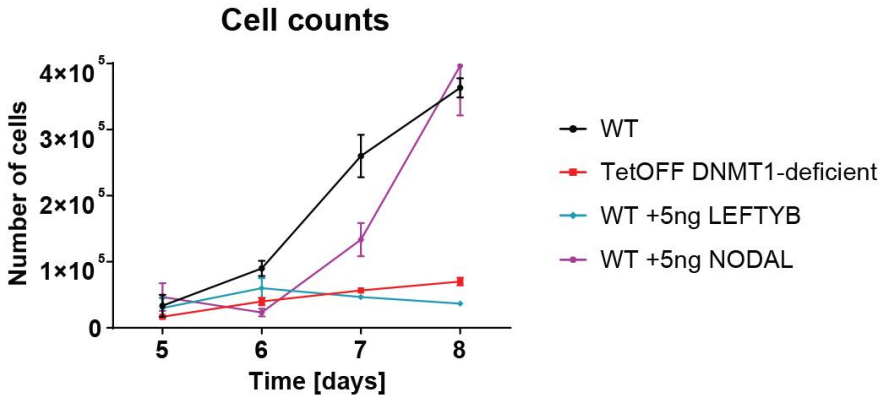


Figure 34. Absence of Activin/Nodal morphogen upregulation in response to DNMT1 depletion in HEK263T cells. ***A.*** *Western blot results depicting the degDNMT1 HEK cell line inducible depletion of DNMT1 in the presence of dTAG13. Loading control – tubulin. Clone #3 was used in the subsequent analysis. Performed by Simon Lauer.* ***B.*** *qRT-PCR analysis of the expression of chosen morphogens (normalized to GAPDH) in cells deficient in DNMT1 and treated for 48h with 2μM SB421543 (TGFβ/Activin/Nodal receptor inhibitor). Error bars: SD. * pval≤0.05; ** pval≤0.005; *** pval≤0.0005; no comparison – not significant. N=3*

To confirm if the SMAD2/3 phosphorylation does not occur in the DNMT1-depleted cells, we used a small molecule inhibitor of Alk4/5/7 receptor, called SB431542[233]. Next, we performed a qPCR analysis of the expression of genes that function in the Nodal signaling pathway in the control- and DNMT1-depleted HEKdegDNMT1cells, as well as the SB431542-treated cells (FIG 34B). As expected, none of the studied targets (*NODAL, LEFTY1, LEFTY2* or *CER1*) was expressed in the HEK293T cells in the presence or absence of SB431542. Depleting DNMT1 using the degron system also did not lead to any significant changes in the expression of the examined morphogens (FIG 34B; note very low expression levels). The result did not change upon the depletion of DNMT1 combined with SMAD2/3 inhibition by SB431542 (FIG 34B). The result did not change upon the depletion of DNMT1 combined with SMAD2/3 inhibition (FIG 34B). This is consistent with previous studies demonstrating that the expression of NODAL and its antagonists dependents on the Activin/Nodal signal transduction pathway activity[227]. Inducing the DNA hypomethylation is therefore insufficient to activate the expression of *NODAL*, *LEFTY1/2* and *CER1* in HEK293T cells probably due to the absence of TGFβ/Activin/Nodal receptor stimulation.

While these results suggest that changes in the expression of Activin/Nodal morphogens are not a universal response to global DNA hypomethylation, we reasoned that it could contribute to the loss-of-fitness phenotype we observe in the DNMT1-deficient hESCs. For example, LEFTY1 could negatively impact the fitness of the DNMT1-deficient hESCs if NODAL-dependent signal transduction is required for hESC growth. In fact, corroborating observations were reported recently[234]. In the study by Fiorenzano and colleagues, a CRIPTO-blocking peptide was used to treat hESCs, which specifically disrupted the interaction between NODAL and its receptor (without affecting TGFβ1 binding to TβRII and Alk5)[235]. Despite culturing in TGFβ-supplemented medium, the cells performed poorly in proliferation assay compared to control controls. Moreover, reducing the amount of CRIPTO protein using shRNAs led to lower growth rate and a spontaneous differentiation in the *CRIPTO* knock-down hESC lines[234]. Inspired by previous studies, we perturbed the NODAL-induced receptor stimulation in hESCs by ectopic expression of *LEFTY1* or its protein product LEFTYB (FIG 35).

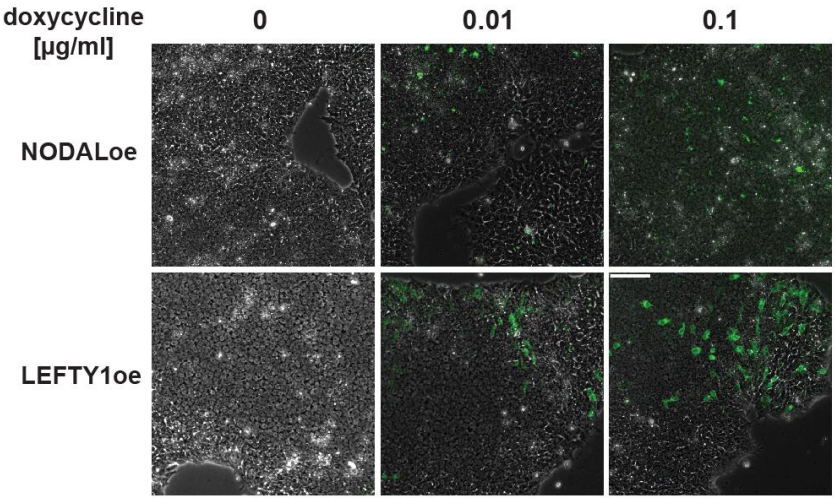
A


Cell counts
Number of cells
4×10^5
3×10^5
2×10^5
1×10^5
0
5
6
7
8
Time [days]
WT
TetOFF DNMT1-deficient
WT +5ng LEFTYB
WT +5ng NODAL

B


LTR
UCOE
TRE
NODAL
GFP
rtTA
P2A
Blasticidin
LTR
LTR
UCOE
TRE
LEFTY1
GFP
rtTA
P2A
Blasticidin
LTR

C


doxycycline
[µg/ml]
0
0.01
0.1
NODALoe
LEFTY1oe

D

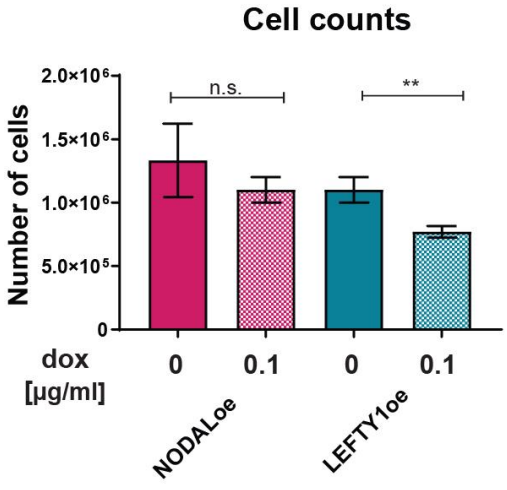

Cell counts
Number of cells
2.0×10^6
1.5×10^6
1.0×10^6
5.0×10^5
0
n.s.
**
dox
[µg/ml]
0
0.1
0
0.1
NODALoe
LEFTY1oe

Figure 35. LEFTY1 is a negative regulator of hESC growth. ***A.*** *XY graph depicting the growth curve of TetOFF DNMT1 hESCs (black), treated with 5ng/mL recombinant LEFTYB (green), 5ng/mL recombinant NODAL (purple) or DNMT1-deficient TetOFF DNMT1 cells, counted at passage 2, days 5-8 post DNMT1 depletion (red). N=2.* ***B.*** *Schematic representation of lentiviral vector construct used for NODAL or LEFTY1 overexpression in hESCs. LTR – Long Terminal Repeats, UCOE – Ubiquitous Chromatin Opening Element, TRE – Tet Response Element, rtTA – recombinant Tet transactivator.* ***C.*** *Representative images of NODALoe and LEFTY1oe cell lines exposed to varying concentrations of doxycycline for 24h. Phase view of live colonies (gray) is overlaid with morphogen expression visualized via GFP fusion (green).* ***D.*** *Bar graph representing cell counts to compare growth of NODALoe and LEFTY1oe lines exposed to indicated amount of doxycycline for a single passage (4 days). Error bars: SD. * pval≤0.05; ** pval≤0.005; *** pval≤0.0005; n.s. – not significant. N=3*

6.25 LEFTY1 impacting the proliferation of hESC

Gain of function assessment. To examine whether LEFTYB (the protein product of the gene *LEFTY1*) could affect the growth of hESCs, we compared the growth of DNMT1-deficient cells in passage 2 (day 5-day 8 after DNMT1 depletion) with the growth of control cells treated with recombinant NODAL and LEFTYB proteins (Fig 35A). Treatment with a recombinant NODAL resulted in a similar growth as the control cells, despite an initial lag (FIG 35A). In contrast, treatment with LEFTYB inhibited the growth of hESCs. We thus concluded that LEFTYB, but not NODAL, negatively impacts proliferation of hESCs in a manner comparable to the depletion of DNMT1 (FIG 35A).

Supplementation of a recombinant protein does not recapitulate the steps of processing and secretion that naturally follow the expression of a signaling molecule. In order to examine the impact of ectopic expression of *NODAL* and *LEFTY1* on hESC growth, we engineered a lentiviral delivery system that would enable an inducible expression of a gene in wild-type HUES64 cell line (FIG 35B). Briefly, the expression cassette sequence was flanked by LTR repeats[236] to mediate random integration into the genome. Next, an upstream chromatin element opening sequence (UCOE) was introduced to prevent silencing of a TRE promoter[237]. We PCR-amplify the cDNA of either *NODAL* or *LEFTY1* and cloned them into the donor vector. Each gene was C-terminally fused to super-folded GFP (sfGFP) to provide reliable readout of the doxycycline-induced transgene expression. The second transgene consisted of an rtTA transactivator linked by cleavable liner (P2A)[238] to the blasticidin-resistance cassette. In the presence of doxycycline, the rtTA binds the TRE promoter and activates the transcription of the *NODAL* or *LEFTY1* gene. The blasticidin resistance gene served as a selection method for donor integration and activity. The clonal picking of the blasticidin-resistant colonies

led to the generation of the morphogen-overexpression hESC lines call thereafter: NODALoe and LEFTY1oe. After doxycycline supplementation (24h), the respective transgenes were activated in hESCs in a dose-dependent manner (FIG 35C). To measure the effect of morphogen overexpression on cell growth, we performed a cell count on the NODALoe/LEFTY1oe lines induced with 0.1µg/ml doxycycline (FIG 35D). Again, the results demonstrated a deleterious effect of *LEFTY1*, but not *NODAL*, expression on the hESC growth. Taken together, we concluded that the ectopic expression of *LEFTY1* has a negative effect on the growth of hESCs. We suspect that this is linked to its role as an inhibitor of the Activin/Nodal signal transduction pathway.

Loss of function. If the increased expression of *LEFTY1* negatively affects the hESC growth, we hypothesized that knockout of *LEFTY1* could alleviate the DNMT1-deficient cell growth defect. In order to explore this possibility, we turned to lentiviral strategy presented in FIG 13. to deliver the Cas9 protein and an sgRNA against *LEFTY1* to disrupt the gene in the TetOFF DNMT1-inducible depletion cell line. After transduction with such virus and subsequent selection for its expression, we individually collected single cell-derived colonies and genotyped the *LEFTY1* locus to identify a *LEFTY1* heterozygote (+/-) and a *LEFTY1* homozygous knockout (-/-) clones (FIG 36A and B). We noted that while the *LEFTY1*$^{+/-}$ cell line did not affect the growth of hESCs, the complete *LEFTY1* knockout cell line had a significant growth advantage over wild-type cells, consistent with our observations of *LEFTY1* being a negative regulator of hESC growth (FIG 36C).

We next compared the growth of the TetOFF DNMT1-depleted parental line with TetOFF DNMT1; *LEFTY1*$^{+/-}$ and TetOFF DNMT1; *LEFTY1*$^{-/-}$ lines (FIG 36D). We observed that the initial loss of fitness of the DNMT1-deficient cells in the first passage was not affected by of the absence of *LEFTY1*. This is consistent with our RNA sequencing data, which demonstrated that *LEFTY1* is not significantly upregulated immediately after DNA methylation loss (FIG 33). In contrast, the proportion of surviving cells at day 8 was significantly higher in the TetOFF; *LEFTY1*$^{-/-}$ compared to just the DNMT1-depleted cells. On average, the growth of DNMT1-depleted; *LEFTY1*$^{-/-}$ double-deficient line was a quarter of that of reached by the TetOFF; *LEFTY1*$^{-/-}$ single knockout line. In contrast, after two passages the growth of DNMT1-deficient cells was 10% of that reached by control cells. The heterozygous cell line TetOFF; *LEFTY1*$^{+/-}$ did not show a statistically significant rescue of the TetOFF DNMT1 parental line.

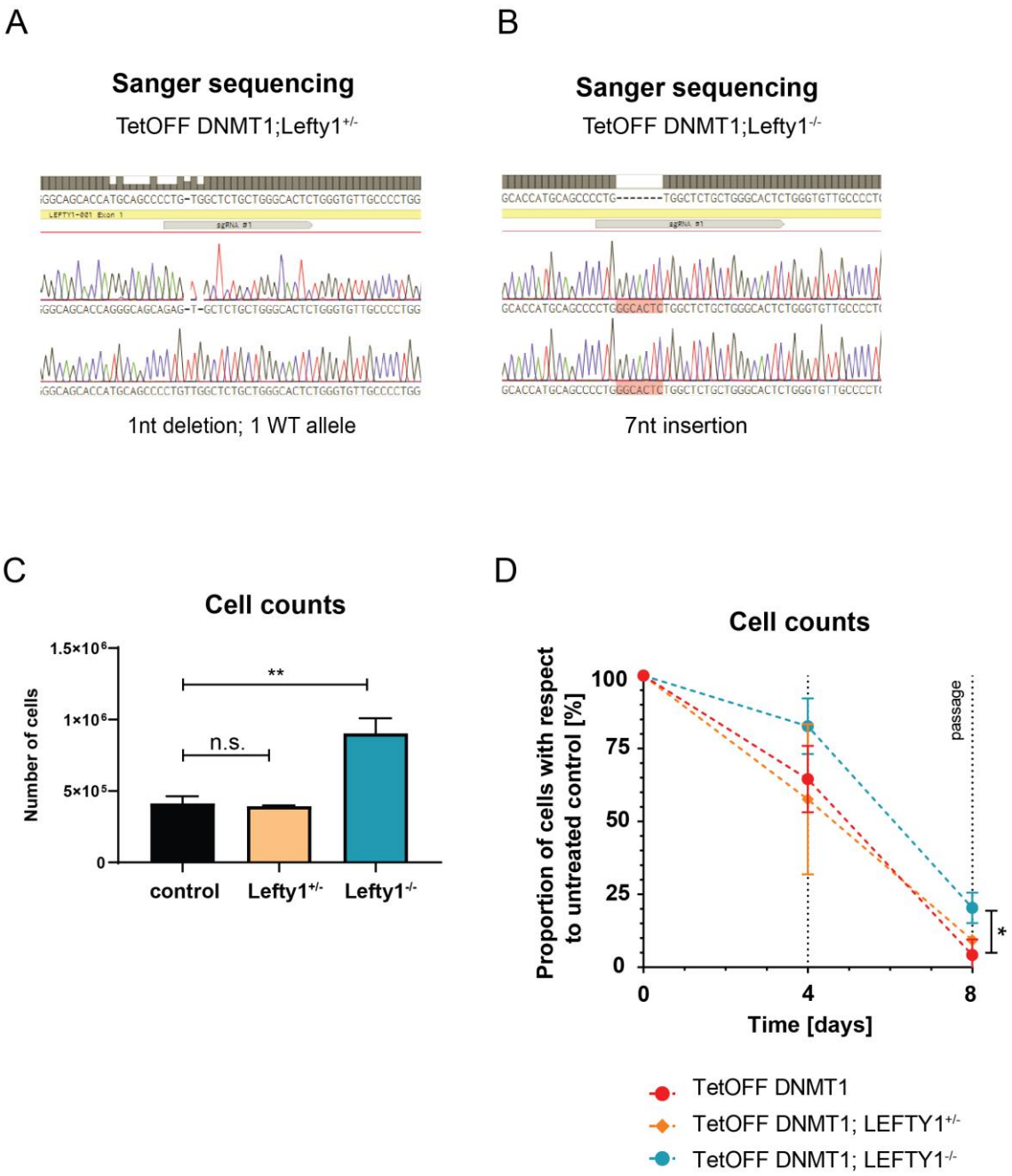


Figure 36. LEFTY1 knockout can partially rescue DNMT1 depletion-induced growth defect. ***A.*** *and* ***B.*** *Sanger sequencing results visualizing the mutations in LEFTY1 exon 1 (yellow bar) in TetOFF DNMT1; LEFTY1$^{+/-}$ cell line and TetOFF DNMT1; LEFTY1$^{-/-}$ line, respectively. The sequence of guide RNA used is shown in gray.* ***C.*** *Bar graph depicting the number of cells of the respective TetOFF DNMT1 cell lines after a single passage (4 days) in the absence of doxycycline.* ***D.*** *XY graph representing the cell growth of DNMT1-deficient cell lines as a proportion of live cells with respect to DNMT1-present controls, calculated after every passage. Error bars: SD. * pval≤0.05; ** pval≤0.005; *** pval≤0.0005; n.s. – not significant. N=3*

We conclude that *LEFTY1* upregulation is one of the consequences of global loss of DNA methylation in hESCs, which negatively impacts the growth of the cells. Since the rescue provided by *LEFTY1* knockout was partial, additional mechanisms must be involved in reducing the fitness of the DNMT1-deficient cells.

6.26 DNMT1-deficiency affects transcriptional response in hESC

The pluripotency of hPSCs is maintained in the cell culture conditions by supplementing the medium with TGFβ1 and the Fibroblast Growth Factor (FGF)[133]. Balance between these two signal transduction pathways dictate the propensity of hPSCs to differentiate into ectoderm (FGF) versus meso- and endoderm (Activin/Nodal). This process is coordinated by the transcription factor NANOG[215]. Interestingly, NANOG is one of the direct downstream targets of the Activin/Nodal signal transduction, as well as a co-factor acting to fine-tune this pathway and restrict its endoderm-inducing potential. On the other hand, NANOG acts to inhibit the expression of the FGF2-mediated neuroectoderm factor. The multiple functions of NANOG are regulated by inputs from extracellular stimuli that activate the downstream signaling cascade that converge on transcription.

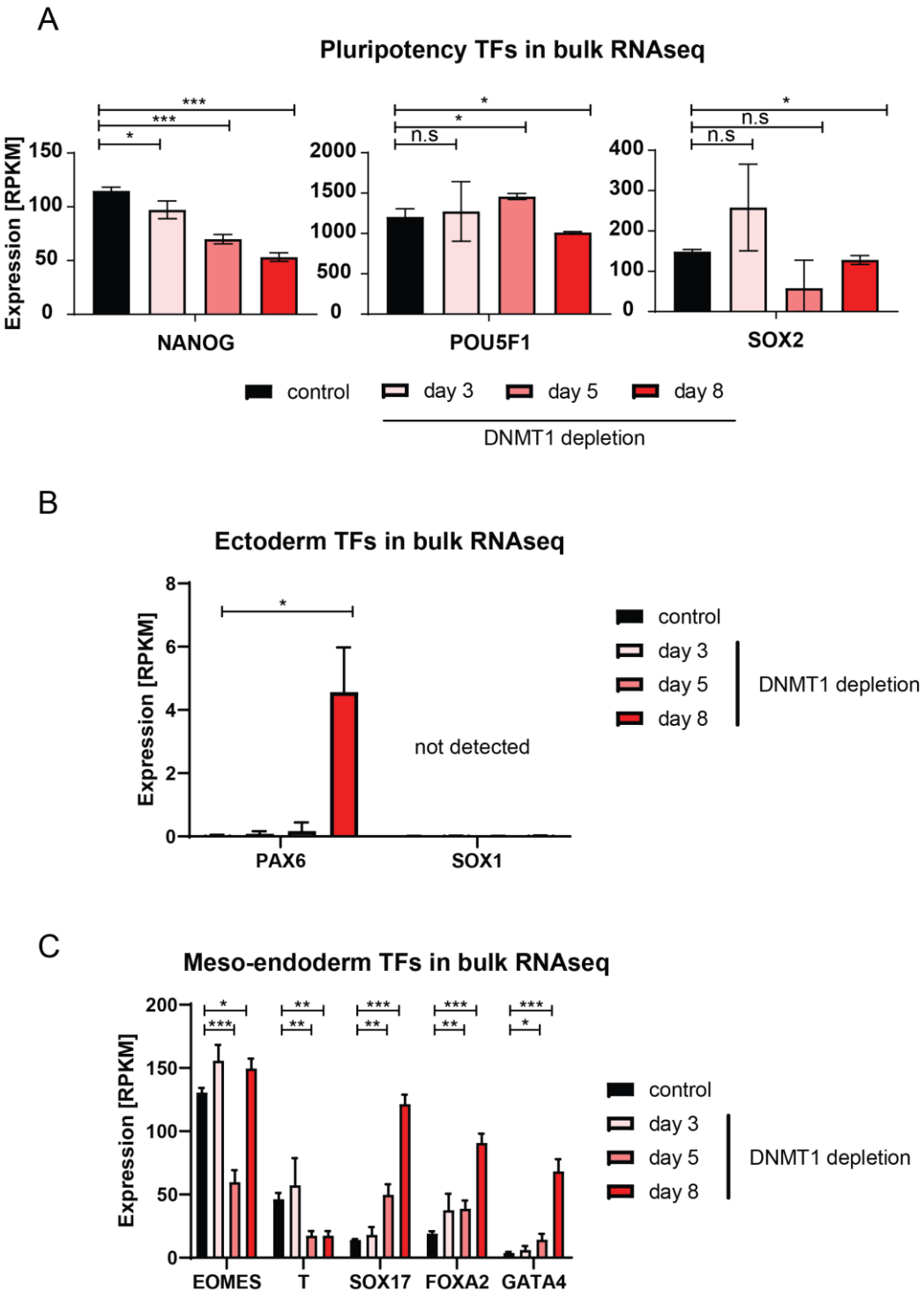


Figure 37. DNMT1 deficiency results in changes in expression of transcription factors. ***A.*** *Bar graphs representing the bulk RNAseq results of expression of the pluripotency-associated transcription factors (NANOG, SOX2 and POU5F1, respectively) in control cells and TetOFF DNMT1-depleted cells for 3, 5 or 8 days.* ***B.*** *Bar graph of bulk RNAseq expression of ectoderm-associated TFs. SOX1 has not been detected at any of the time points (RPKM<0.1).* ***C.*** *Bar graph representing the bulk RNAseq measurement of the expression of meso- and endoderm-associated transcription factors in TetOFF DNMT1 cells. Error bars: SD. * pval≤0.05; ** pval≤0.005; *** pval≤0.0005; n.s. – not significant.*

Since DNA hypomethylation influences the expression of the Activin/Nodal signal transduction pathway morphogens, we hypothesized that this could influence the transcriptional balance maintained by NANOG and in turn lead to the differentiation of the DNMT1-deficient hESC. To test this hypothesis, we turned to our bulk RNAseq data to compare the expression level of transcription factors characteristic for either the pluripotent hESCs (*NANOG, POU5F1, SOX2*; FIG 37A), ectoderm (*PAX6, SOX1*; FIG 37B) or combined meso- and endoderm stages of germ layer differentiation (*EOMES, T, SOX17, FOXA2, GATA4*; FIG 37C). Although we did not observe a transcriptional shut off of *POU5F1* in the DNA hypomethylated cells, *NANOG* was significantly downregulated. The decrease in *NANOG* expression could further result in differentiation of hESCs in agreement with its role as a balance between Activin/Nodal and FGF signal transduction pathways. Our analysis showed the low level of *PAX6* expression at day 8 of DNMT1 depletion and the absence of *SOX1* expression that overall indicate a lack of ectoderm differentiation. In contrast, the expression pattern of the meso- and endoderm markers after DNMT1 depletion gradually resembled the *in vitro* differentiation of hESCs towards mesendoderm and endoderm, respectively[239]. For example, we observed an initial upregulation of BRACHYURY (encoded by the *T* gene), which was subsequently downregulated, as well as upregulation of *FOXA2* and *SOX17* (FIG 38). In addition, we observed an upregulation of PRDM1, a transcription factor found in mesoderm-derived primordial germ cells, indicating that the hESCs are not strictly following endoderm differentiation program.

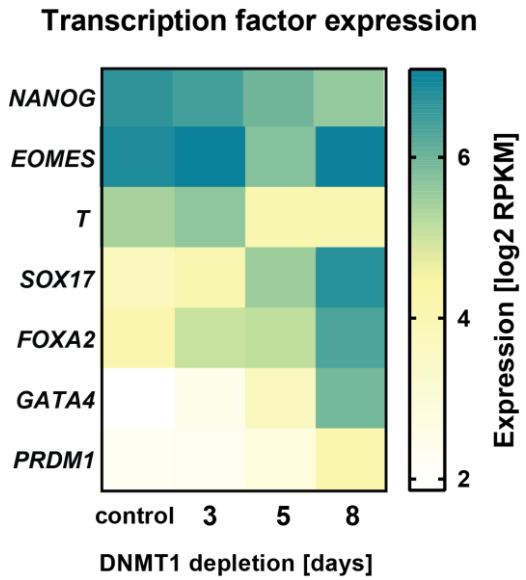


Figure 38. Changes in expression of key TFs suggest mesendodermal differentiation of DNMT1-deficient hESCs. Heat map representing the expression (log_2RPKM) of key TFs measured in bulk RNAseq in the presence (control) or absence of DNMT1 for 3, 5 or 8 days, respectively.

In order to validate these observations, we performed immunofluorescence microscopy on control and DNMT1-deficient hESCs to estimate the protein level of BRACHYURY and SOX17. Consistent with our RNAseq results, we noted higher amount of SOX17 present in the DNMT1-deficient cells compared to control. The levels of BRACHYURY protein were also increased even at day 8 of DNMT1 depletion, despite lowered abundance of the *T* transcript with respect to the control in the RNAseq results (FIG 39).

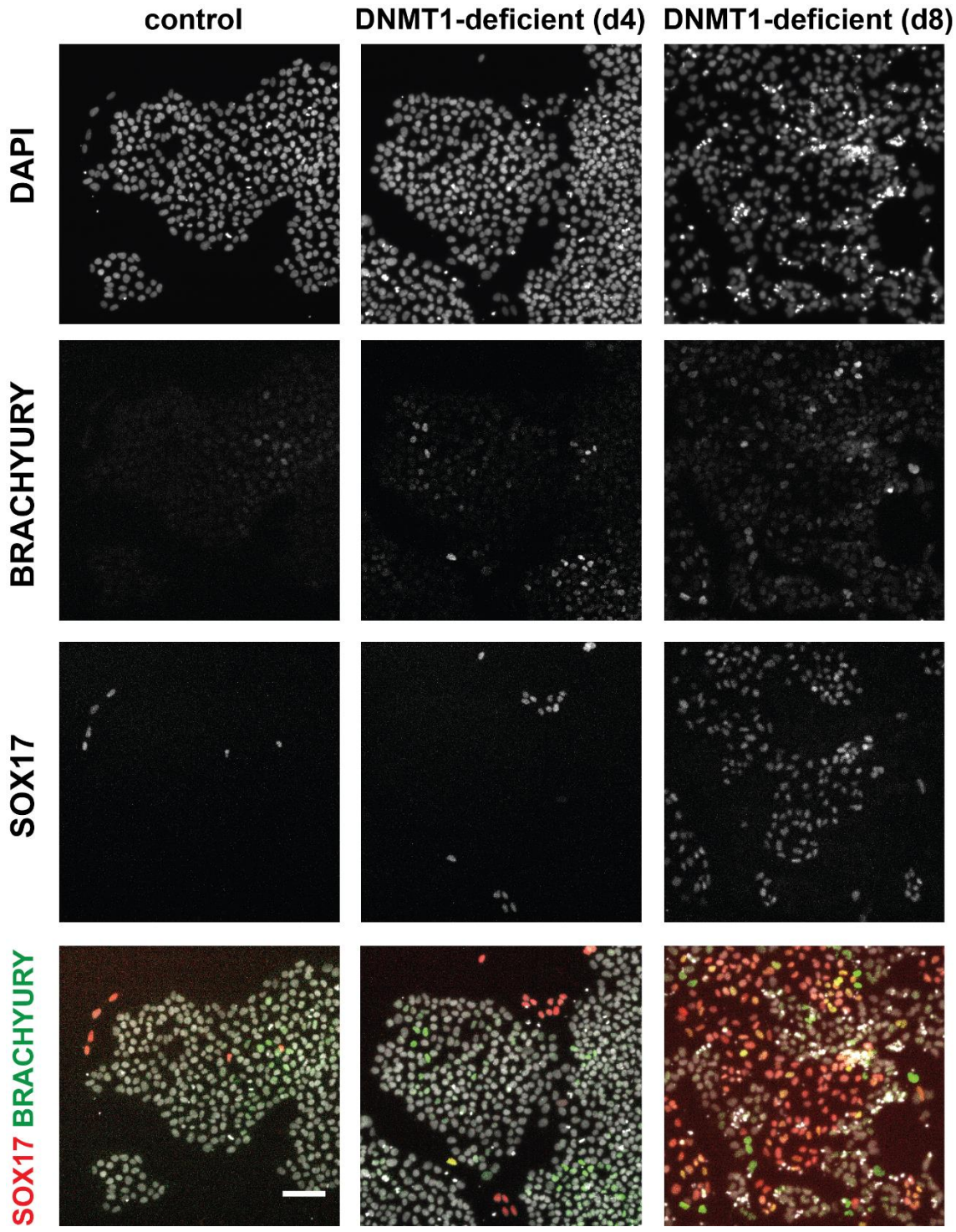


Figure 39. The expression of T and SOX17 changes in DNMT1-deficient hESCs over time. Representative images depicting the DNA staining using DAPI (top row) and immunofluorescent staining against T (second row) and SOX17 (third row). The images were pseudocolored and merged (bottom row). Scale bar: 100μm.

The differentiation genes are targets of SMADs. Since the mesendoderm differentiation program is induced in the presence of Activin *in vitro*[240], we explored if the upregulation of genes in response to global DNA hypomethylation could be ascribed to the mesendodermal gene activation by SMAD2/3, the signal transducers of the TGFβ/Activin/Nodal receptors (FIG 40). We employed Genomic Regions Enrichment of Annotations Tool (GREAT)[241] to identify genes with a TSS within 10kb of a SMAD2/3 binding site in undifferentiated hESCs, or cells subjected to mesendoderm differentiation regimes. This approach identified 583 and 3326 SMAD2/3target genes in ESCs and mesendoderm, respectively. We then investigated the overlap between these targets and genes upregulated at the final time point in DNMT1-depleted cells (scRNAseq; 707 genes total).

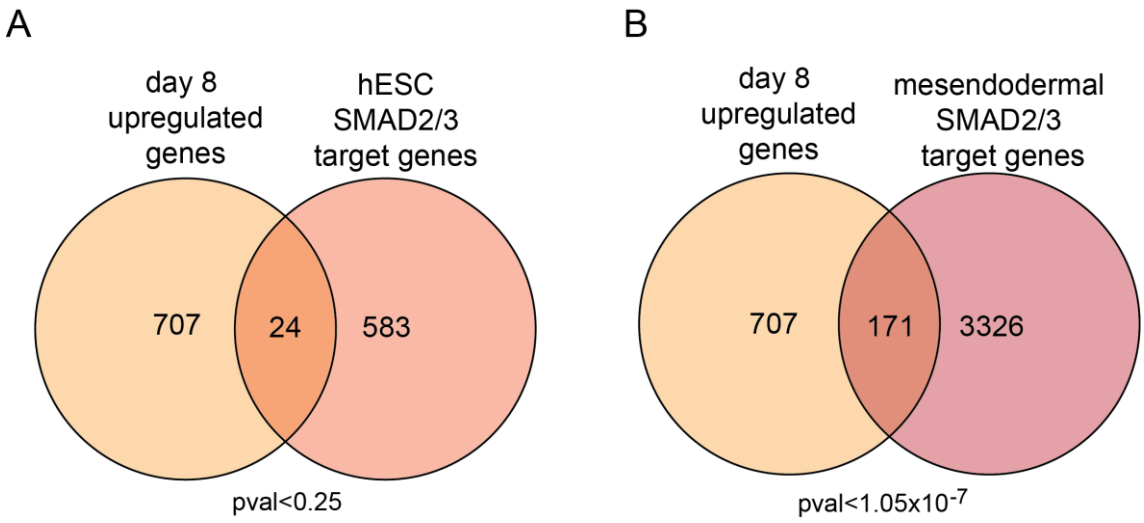


Figure 40. DNMT1-deficient hESCs upregulate mesendodermal targets of SMAD2/3. ***A.*** *Venn diagram depicting the overlap between genes upregulated at day 8 of DNMT1-depletion (scRNAseq) and hESC SMAD2/3 binding target genes.* ***B.*** *Venn diagram depicting the overlap between genes upregulated at day 8 of DNMT1-depletion (scRNAseq) and human mesendoderm SMAD2/3 binding target genes. Significance of each overlap is listed beneath the graphs.*

Out of 583 hESC SMAD target genes, only 24 overlapped with the day 8 upregulated genes (FIG 40A). We expected this result, as the genes regulated by SMAD2/3 are likely already active in the control hESCs. Strikingly, 171 genes overlapped between the target genes of SMAD 2/3 binding and the day 8 upregulated genes (24% of day 8 upregulated genes), which bore high statistical significance ($p < 1.05 \times 10^{-7}$; FIG 40B). As a control, we assessed the overlap between day 8 upregulated genes and a randomly generated 3326-gene list. Only 84 genes were shared by the control lists, therefore, the overlap

between SMAD2/3 targets and day 8 upregulated genes is non-random. This suggests that the changes in the transcriptional program resulting from DNA hypomethylation could, to some extent, be caused by the abnormal engagement of the already active Activin/Nodal signal transduction pathway. Instead of being restricted to the pluripotency genes, this signaling pathway activates additionally the gene expression program of the mesendoderm. In agreement with these results, we observed an increased expression of meso- and endoderm marker transcription factors as early as 3 days post DNMT1 depletion, followed by the upregulation of mesendoderm targets of the Activin/Nodal signal transduction pathway – of note, the mesendoderm-upregulated transcription factor genes, such as *T* and *SOX17* also are the SMAD2/3 targets in this lineage (FIG 31; FIG 38).

We next examined the differentiation potential of the DNMT1-deficient hESCs using a Scorecard experiment (FIG 41). Scorecard is a qRT-PCR based assay, which uses an algorithm established based on differentiation profiles of differentiated hESC lines[136,242] to estimate the propensity of hESCs to differentiate into ecto-, meso- or endoderm. The result is presented as a "score" - an auxiliary unit summarizing the expression of key lineage markers, used to directly compare the "differentiation state" of cells within one line or across different experiments. Consistent with our observations of mesoderm program activation in the absence of DNMT1 in hESCs, the DNMT1-deficient cells have a higher mesoderm score than the control cell line. Our hypothesis was that the treatment with the meso- or endoderm differentiation-promoting signaling molecules will affect the score of the DNMT1-deficient and control cells. To this end, we cultured the cells for 6 days in the absence of DNMT1 to induce DNA hypomethylation, followed by 2 days treatment with LIF (as a control, as this cytokine is not associated with hESC differentiation), WNT3A (endoderm-promoting factor) or BMP4 (mesoderm-promoting factor). Both the control and DNMT1-deficient cells responded transcriptionally to the treatment with each factor, as reported by the change in score. Curiously, although the response always displayed the same trend, i.e, both control and DNMT1-deficient hESCs had an increased endoderm or mesoderm score in response to WNMT3A and BMP4, respectively, the score was always higher for the DNMT1-deficient cells. These results suggest that the DNMT1-deficient hESCs are more receptive to respond when exposed to extracellular cues, compared to control cells. The implications of this finding will be further elaborated on in the discussion section.

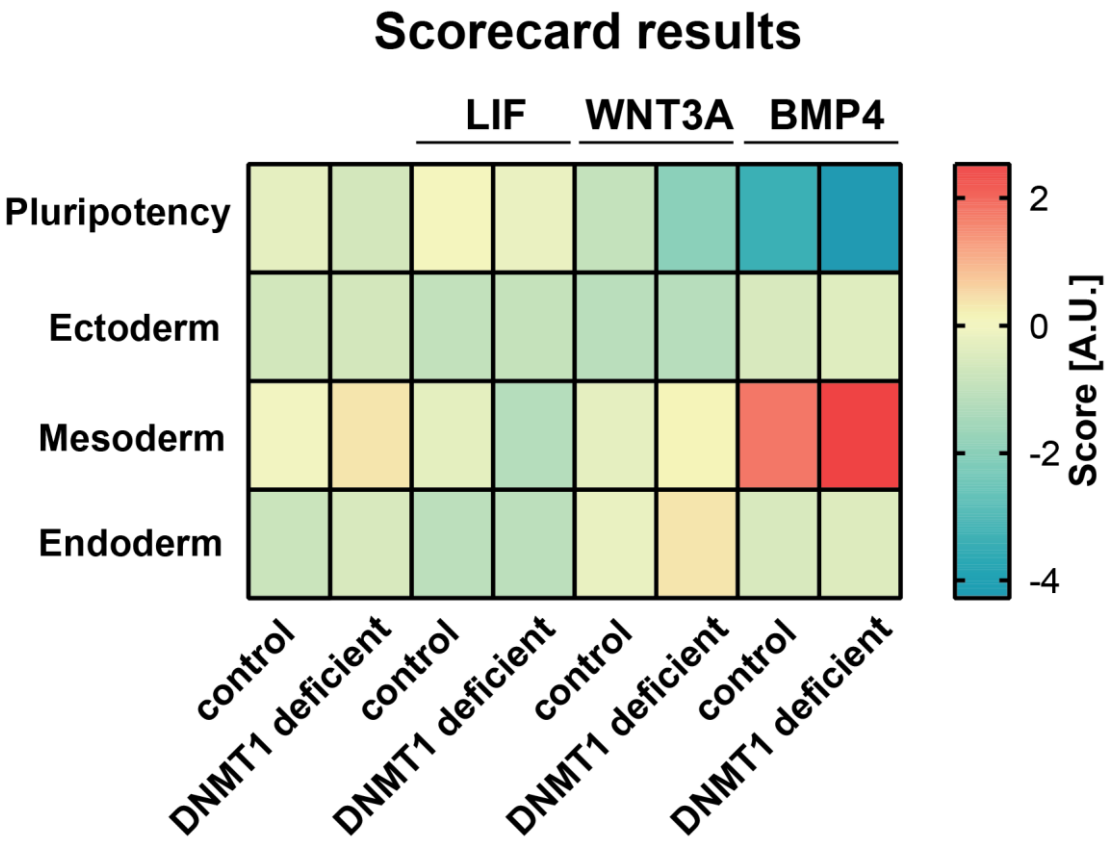


Figure 41. hESC response to morphogen induction changes upon DNMT1 depletion. Heatmap depicting the score (auxiliary unit) of control and degDNMT1 hiPSCs, derived from their expression of a panel of Pluripotency, Ecto-, Meso- and Endoderm-associated genes. Mock control untreated cells (8d – 2 passages culture), DNMT1-deficient cells (degDNMT1; 8d of 125nM dTAG-13 – 2 passages culture) or either cell line grown for 6 days, followed by 2 days of treatment with 500U ml^{-1} LIF, 100 ng ml^{-1} BMP4 or 50nM ml^{-1} WNT3A, respectively were used (total 8d – 2 passages growth).

Whether the DNA methylation-dependent, spontaneous differentiation is a cause of loss of fitness is unclear. Prompted by the recent report that DNMT1KO hNPCs are viable[243], we next investigated if alternative model using different stimulus such as differentiation towards the ectoderm lineage could alleviate the loss-of-fitness phenotype of the TetOFF DNMT1-deficient cells.

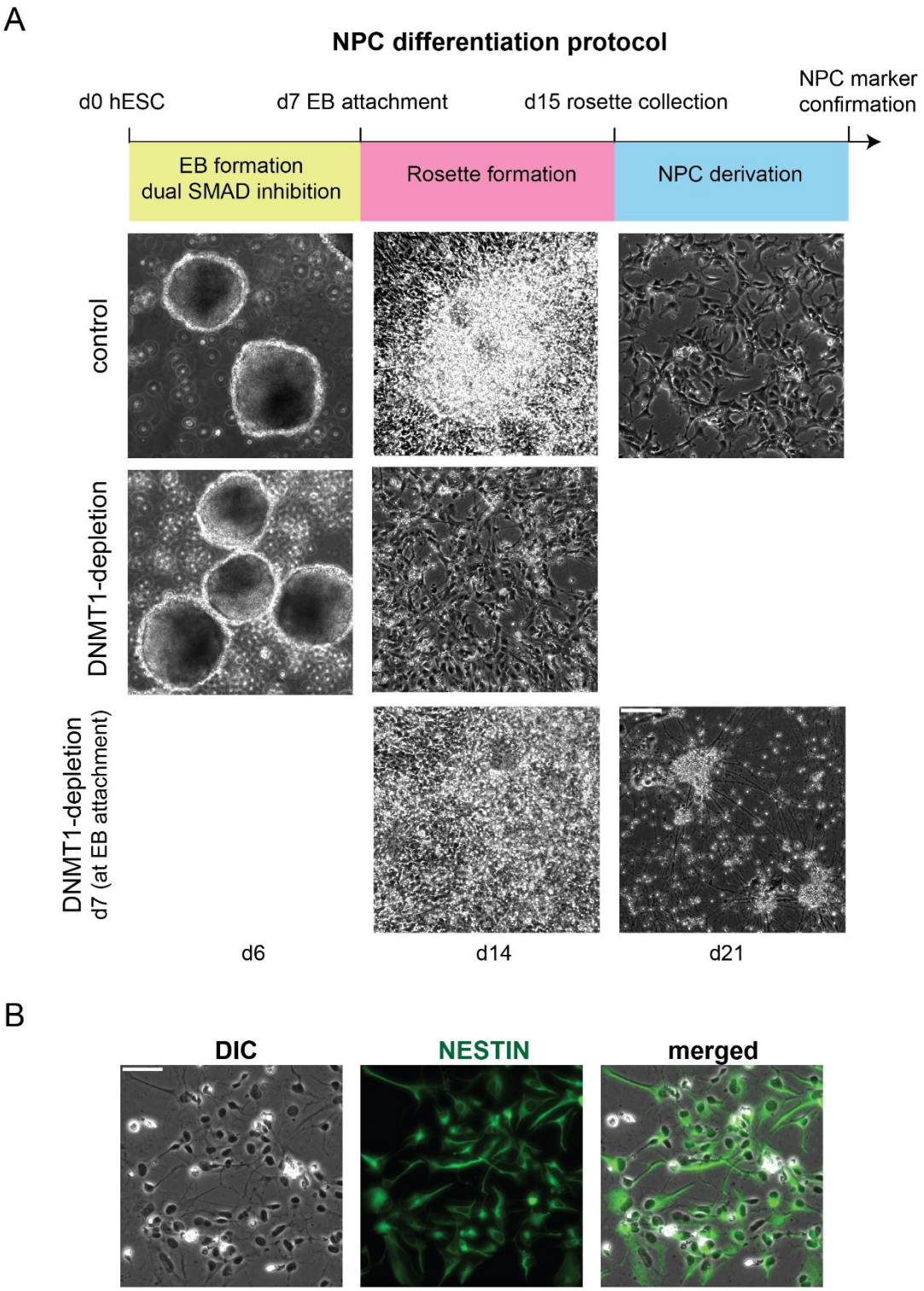


Figure 42. DNMT1-deficient hESCs cannot undergo NPC-directed differentiation. ***A.*** *Schematic of the hESC to NPC differentiation protocol steps, with representative images of each stage depicted below. First row – differentiation of control cells. Middle row – differentiation of cells depleted of DNMT1 at the beginning (time point 0) of the protocol. Bottom row – differentiation of cells depleted of DNMT1 after EB plating (time point 7). Scale bar 50μm.* ***B.*** *Representative image of phase microscopy of wild-type NPCs and fluorescent microscopy to visualize the immunofluorescence of the NPC marker Nestin (green). Scale bar 50μm.*

6.27 Neural Progenitor Cell (NPC) differentiation of DNMT1-deficient hESCs

We subjected the TetOFF DNMT1 cell line to a three-step differentiation regime towards NPC (FIG 42A). Briefly, the undifferentiated colonies were aggregated into Embryoid Bodies (EBs) and grown in the presence of inhibitors for the Activin/Nodal and BMP4 signal transduction pathways to promote the ectoderm differentiation over the meso- and endoderm lineages. The EBs were then seeded on laminin-covered plated until they formed neural rosettes. Selectively passaging the rosettes into the bFGF-supplemented medium yielded the differentiation into NPCs, which show presence for the NPC marker Nestin over the course of 21 days (FIG 42B). First, we wanted to learn if the DNMT1 depletion interferes with the ability of hESCs to undergo and complete this differentiation procedure. To this end, we induced the transcriptional shutdown of DNMT1 in our cells at the beginning of the NPC-derivation. While the DNMT1-depleted cells were able to form EBs, they failed in producing rosettes (FIG 42A; middle row). When DNMT1 was depleted later in the grown rosettes, the cells still failed to form NPCs when the conditions were switched from Nodal/BMP inhibition to bFGF (FIG 42A; bottom row). Altogether, these results demonstrated that global DNA hypomethylation interferes with differentiation of hESCs into NPCs. This is likely due to an interference of the global DNA hypomethylation with the ectoderm differentiation program or with coordinating the transcriptional shutdown of the pluripotency factors necessary to commit to a new lineage.

As embryogenesis in mice deficient for DNMT1 does not progress to the NPC-differentiation stage, we could not compare our results to the *in vivo* model. However, a study by Fan and colleagues engineered mice bearing the NPC-specific DNMT1 knockouts to study the impact of DNA hypomethylation on NPC differentiation potential[103]. NPCs give rise to both neurons and glial cells (such as astrocytes), however, their propensity to differentiate into either lineage was affected without DNMT1. The DNMT1-depleted NPCs are more likely to differentiate towards the glial lineage, rather than neural lineage *in vivo*[103]. Furthermore, global DNA hypomethylation accelerated the differentiation towards glia in *in vitro* cultured DNMT1KO primary NPCs compared to controls, when both were exposed to the glia-inducing Leukemia Inhibitory Factor (LIF)[103]. LIF is a small a cytokine which activates the Janus Kinase – Signal Transducers and Activators of Transcription (JAK-STAT) signal transduction pathway[244]. This pathway is essential for neural progenitor cell differentiation into astrocytes. Fan and colleagues demonstrated that 2 days of LIF supplementation was sufficient to induce the

expression of Glial Fibrillary Acidic Protein (GFAP; an astrocyte marker) in DNMT1KO, but not in the control NPCs.

Inspired by the previous findings, we tested them in our system. We depleted DNMT1 in the TetOFF DNMT1-derived NPCs and investigated the expression levels of *GFAP*, along with the transcription factors found to induce astrocyte differentiation (*NF1b, SOX9*)[113] and, as a control, the master transcription factor regulating neural differentiation – Neurogenin2 (*NEUROG2*) ([245]; FIG 43A). Remarkably, we observed that the expression of *SOX9* was elevated in the DNMT1-depleted NPCs compared to the control cells. Even though the expression levels of *NEUROG2* changed in the DNA hypomethylated cells as well, the difference was not statistically significant. We did not detect the expression of *GFAP* in either condition, thus, DNMT1-depletion is insufficient to trigger and complete the entire glial differentiation in human NPCs *in vitro* (FIG 42A).

When we induced the differentiation towards glia via the addition of LIF for 2 days, we saw that the expression levels of both *SOX9* and *NF1b* increased in both wild-type and DNMT1-depleted NPCs, with regard to respective controls (FIG 43A). Despite the presence of these glia-specific transcription factor, we did not detect *GFAP* mRNA in either of the conditions (FIG 43A). We hypothesize that additional environmental stimuli might be required to trigger expression of this glial differentiation marker in hESCs, even in the DNMT1-deficient state. Indeed, a special astrocyte medium is used to differentiate human NPCs into astrocytes[246]. Similarly to the LIF treatment experiment, culturing the DNMT1-deficient NPCs in the astrocyte medium did not result in an increased expression of *GFAP* (data not shown).

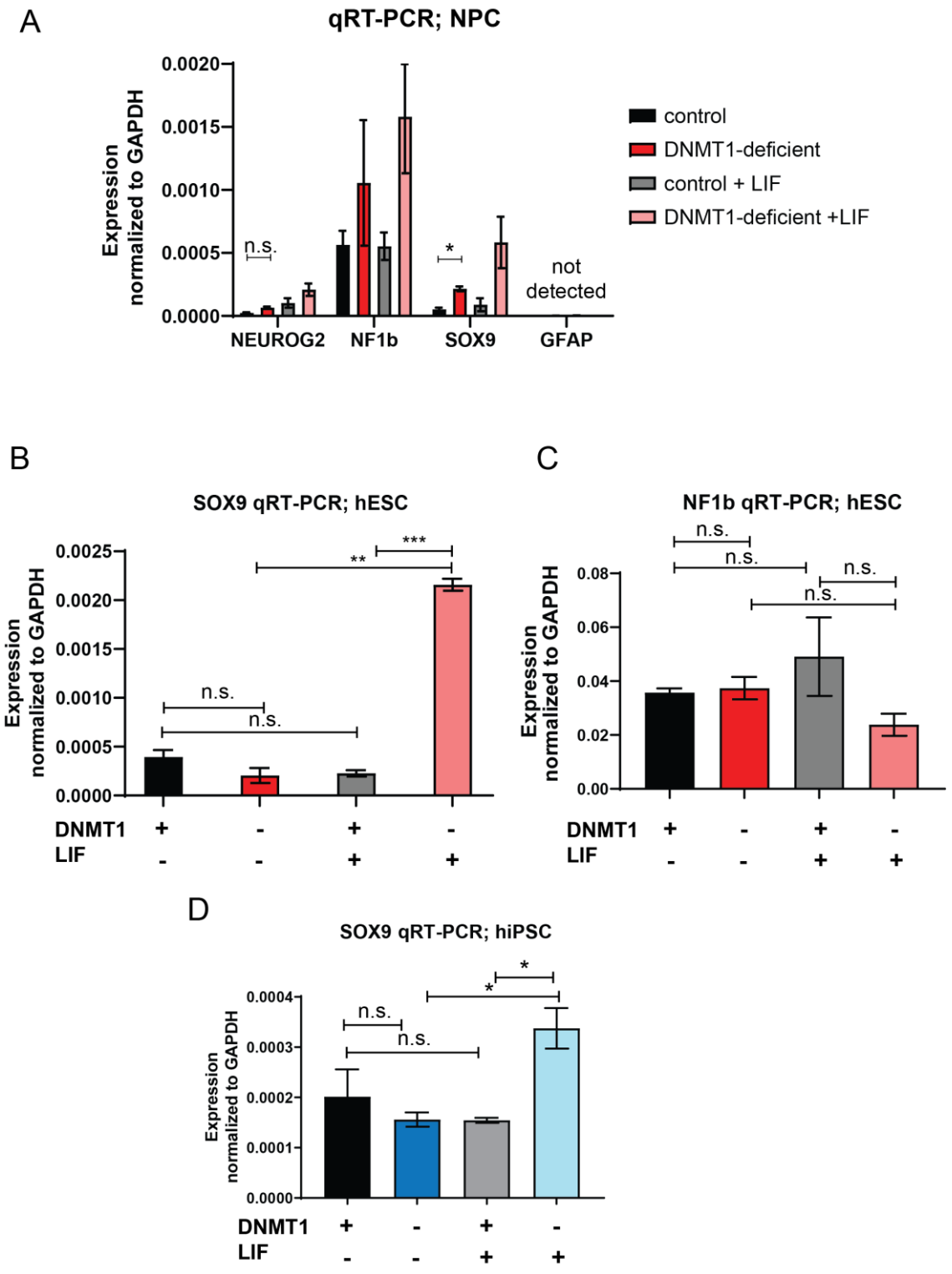


Figure 43. DNMT1 depletion leads to an increase in SOX9 in the presence of JAK-STAT agonist. ***A.*** *Bar graph depicting qRT-PCR analysis of NPC expression of neuronal and glial genes.* ***B.*** *Bar graph of representative qRT-PCR analysis of the expression of SOX9, an astrocyte TF that is a known JAK-STAT signal transduction pathway target in TetOFF DNMT1 hESCs. N=5.* ***C.*** *Bar graph depicting qRT-PCR analysis of the expression of NF1b, a glial TF that is not a known JAK-STAT signaling pathway target, in TetOFF DNMT1 hESC line. Cells in A., B. and C underwent DNMT1-depletion for a total of 8 days in the presence of 2μg/mL doxycycline. LIF treatment:*

48h, 1000U. ***D.*** *Bar graph representing the results of qRT-PCR analysis of SOX9 depletion in degDNMT1 hiPSCs. The cells were DNMT1-depleted for a total of 8 days and the LIF treatment was 48h long in the presence of 500U LIF. Error bars: SD. * pval≤0.05; ** pval≤0.005; *** pval≤0.0005; n.s. – not significant. N=2.*

Taken together, the human NPC deficient for DNMT1 did not spontaneously differentiate into astrocytes, nor was the glial differentiation induced shortly after exposing cells to astrogenesis-promoting conditions, as this was the case for mouse DNMT1KO NPCs. This could be an inherent difference between mouse and human NPCs, or a technical difference in the propensity to differentiate between primary and in-vitro derived NPCs. Nevertheless, our results demonstrated that *SOX9* – one of the main transcription factor driving glial development, was significantly upregulated in human DNMT1-deficient NPCs compared to controls. Furthermore, LIF treatment for two days was sufficient to increase *SOX9* expression further in control and DNMT1-deficient NPCs. This is consistent with *SOX9* being a direct target of JAK-STAT signal transduction pathway[247] We hypothesize that cells with the reduced levels of DNA methylation are more prone to mount an unrestricted response to a signaling cue.

Strikingly, when instead of the NPCs, we repeated the experiment in the TetOFF DNMT1 hESCs (where the cells were depleted of DNMT1 for 6 days, followed by 2 days of LIF treatment) we also saw a statistically significant increase in the expression of *SOX9*, but not the negative control *NF1b* in the DNA hypomethylated cells (FIG 43B and C). Next, we sought to reproduce these results in the degDNMT1 cell line (FIG 43D). Exposure to LIF elicited a statistically significant upregulation of this transcription factor in DNMT1-deficient, but not control cells. Of note, unlike the TetOFF DNMT1, the degDNMT1 cells expressed a basal level of *SOX9* in the control and the DNMT1-deficient state.

Why was *SOX9* upregulated in the presence of LIF in the DNMT1-deficient, but not control hESCs? We hypothesized that the global loss of DNA methylation could affect the chromatin accessibility of *SOX9* enhancer regions and enable transcription of this gene exclusively in the DNA hypomethylated cells. To further explore this possibility, we investigated the changes in DNA methylation levels on the known *SOX9* enhancers in control and the DNMT1-deficient hESCs using our scRRBS data (CHAPTER 3).

SOX9 is a transcription factor important for the development and function of diverse tissues. Aside from glial differentiation, SOX9 is indispensable for chondrocyte and testis gene regulation[248,249]. Mutations in a single allele of *SOX9* results in a severe developmental defect called campomelic dysplasia (CD)[250], which results in a pleiotropic phenotype including skeletal malformations and sex reversal in male patients. Multiple

enhancers of *SOX9* have been identified through studies of CD cases resulting from mutations in non-coding loci[251] and mutagenesis approaches[252]. We sought to compare the DNA methylation status of *SOX9* enhancer loci in the control and DNMT1-depleted cells.

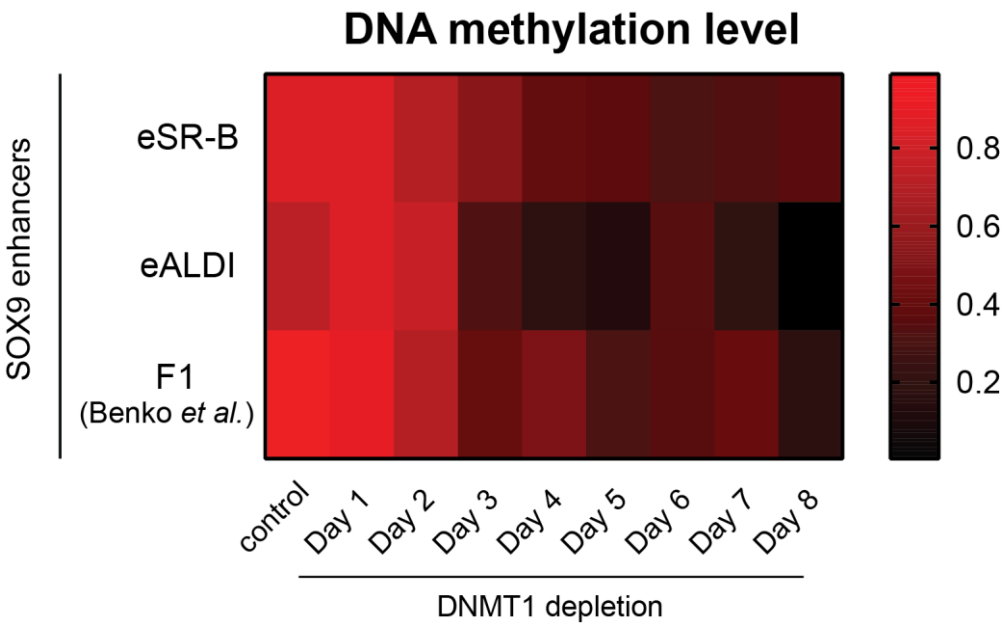


Figure 44 SOX9 enhancers lose DNA methylation in the absence of DNMT1. Heat map depicting scRRBS DNA methylation averages of three SOX9 enhancer regions in hESCs under control conditions or after DNMT1 depletion for indicated period of time.

We identified three enhancer regions that were represented in our sequencing data with enough coverage for further analyses. Interestingly, these regions were all highly methylated in control cells (mean methylation of 82%, 75% and 99%, respectively.; FIG 44) which is consistent with no detectable gene product. In turn, after 8 days of DNMT1 depletion, their methylation was reduced to 35%, 0.4% and 17%, respectively. Further experiments will be required to identify the role of DNA methylation in indirectly controlling *SOX9* expression. Deeper coverage of DNA methylation data combined with TF binding motif analysis, will be instrumental to elucidate the requirement for the JAK-STAT signaling-dependent in the *SOX9* activation in the normal development (such as germline) and the DNMT1-deficiency models of the global DNA hypomethylation.

6.28 Summary

Among genes upregulated in response to global DNA hypomethylation, the Activin/Nodal signaling pathway drew our attention in particular. Not only is that pathway essential for the survival of hESCs, but also the deregulation of its components (especially *LEFTY1*) was consistent across single-cell and bulk RNAseq measurements and further confirmed by qPCR. We therefore examined the impact of their protein products, morphogens: NODAL and LEFTY1, on the hESCs proliferation. Both the addition of the recombinant LEFTYB, as well as its ectopic overexpression had a negative impact on the hESC growth. In contrast, the NODAL treatment or overexpression did not exert any visible effect on these stem cells. In order to directly probe the impact *LEFTY1* upregulation could elicit on the loss of fitness of DNMT1-deficient hESCs, we knocked out either one or both alleles of *LEFTY1* in the TetOFF DNMT1 background. The depletion of *LEFTY1* partially restored the stunted growth of the DNMT1-depleted cells in the late (passage 2), but not the early (passage 1) time point, consistent with the timing of *LEFTY1* expression in the DNMT1-deficient hESCs.

While LEFTY1 is a negative regulator of hESC growth, our study of NODAL and LEFTY1 expression in HEK293 cells demonstrated that the loss of DNMT1 is not sufficient to cause the expression of these morphogens. NODAL and its inhibitors could have been upregulated in the hypomethylated hESCs by the virtue of being regulated by SMAD2/3, rather than directly orchestrating the expression of downstream genes. We therefore investigated if the global DNA hypomethylation-induced changes in gene expression could be explained by activity of the Activin/Nodal-stimulated signal transduction. Our bioinformatics analysis demonstrated that nearly a quarter of genes upregulated in response to DNA hypomethylation are previously described targets of the Activin/Nodal signal transduction pathway. Surprisingly, the targets were not specific to hESCs, but rather to mesendoderm, a lineage derived from pluripotent cells in response to the WNT and Activin signaling pathways, even though no WNT was supplemented in the medium. This change in expression could have been executed significant increase in the expression meso- and endodermal transcription factors, which followed the loss of DNMT1 in hESCs.

Depletion of DNMT1 in hNPCs in the presence of LIF resulted in an increased expression of glial marker and JAK-STAT target SOX9. Strikingly, loss of DNMT1 facilitated the activation of SOX9 in the presence of LIF in hESCs as well. These results suggest that the changes in gene expression upon global DNA hypomethylation are influenced by the environment of the cells. Comparison of gene expression between hESCs and DNMT1-deficient cells demonstrates necessity of DNA methylation as a licensing factor in

response to all the environmental cues that otherwise would be instructive to elicit an aberrant transcriptional program.

The fact that the DNMT1-deficient hESCs did not uniformly differentiate into endoderm could be explained by multiple reasons: (1) the lack of WNT, another signal transduction pathway crucial for mesendodermal development, (2) the inhibitory effect of FGF present in the medium or (3) the inability of the DNA hypomethylated cells to inhibit the expression of pluripotency factors, which are incompatible with the new identity (such as *POU5F1*), which has been reported before[96].

The aberrant gene expression downstream of the Activin/Nodal signal transduction in the absence of DNMT1 prompted us to explore if changing the culture conditions to exclude extracellular signal cues could result in a rescue the loss of fitness of the DNMT1-depleted hESCs. Temporal ablation of signaling during NPC derivation did not improve the growth of DNA hypomethylated cells in comparison to controls under the same conditions, suggesting other, cell-autonomous mechanisms are responsible for the loss-of-fitness phenotype of the DNMT1-deficicent hESCs.

7. DISCUSSION

7.1 The epigenetic landscape

Over the course of development, a totipotent cell undergoes differentiation into all the cell types of an organism. Totipotent cells give rise to pluripotent cells, which have a restricted potential, because they can become one of the three germ layers, and can no longer commit to the extraembryonic lineage. Each subsequent commitment to a new cell type restricts the number of possible cell states that can be achieved, until the cell commits to a final function, thus becoming terminally differentiated.

Conrad Waddington described that process in the metaphor of a ball (representing a cell) rolling down a rugged hillside (FIG 45A)[253]. The ball falls into one of the valleys and continues down this specific route, until it meets an inflection in the landscape, where it can again roll in one of several possible directions. At the same time, the slope and the hills prevent the cells from rolling over to the neighboring valley. In the same way, cells cannot abruptly change their identity by switching to a parallel differentiation route. Altogether, the route taken by the cell determines where it eventually lands (which terminally differentiated cell fate it assumes).

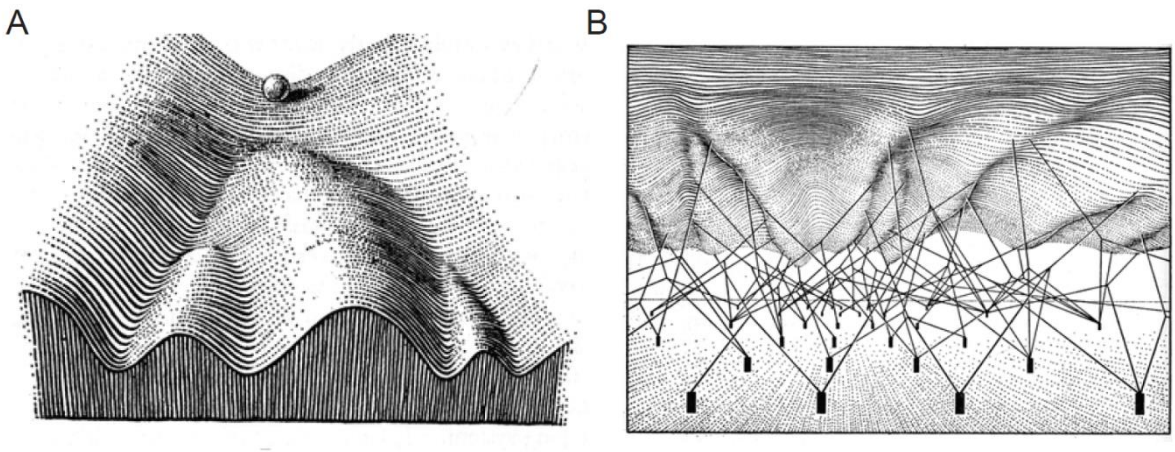


Figure 45. The Waddington landscape.Adapted from[253].

The topology of this landscape is underpinned by ropes anchored to the ground (FIG 45B). Each anchor represents a gene exerting its influence on the shape of the landscape above through a set of interconnected ropes. Although genes are the ultimate anchor points, their action is modulated through the chemical properties of gene products (ropes), which can interact with one another and the environment, changing the ultimate

shape of the landscape above the genes – an epigenetic landscape (greek “epi”, meaning “above”)[253].

Decades of research identified a vast repertoire of complex processes that regulate gene expression in a DNA sequence-independent manner. This is orchestrated by chromatin – an entity composed of the DNA and nucleosomes, which carry posttranslational modifications, that can alter the binding capabilities of proteins initiating gene expression. Such modes of regulation offer a platform for genome interpretation.

Transcription factors take advantage of permissive environment to bind their recognition motif. While the short binding motifs are found in multiple loci across the genome for each transcription factor, only a fraction are actually bound[254]. The specificity of transcription factors is also controlled by other means, such as combinatorial binding of two or more factors. Furthermore, DNA and chromatin modifications can act refractory to transcription factors by blocking the binding. For example, the methylation of cytosines in the DNA can inhibit DNA-binding proteins directly by altering their binding motif, or indirectly via recruiting methyl-binding proteins (MBDs), which occupy methylcytosine-marked loci[255].

Chromatin modifications, DNA methylation and transcription factors can integrate extracellular cues and, altogether, these variables shape the Waddington epigenetic landscape. How exactly the cell-autonomous and non-cell-autonomous cues are reconciled and integrate at chromatin to instruct a cell for a lineage commitment has been an area of intensive research.

7.2 What happens in the absence of instructive signals?

Pluripotent stem cells can exit their self-renewing state to differentiate into ecto-, meso- or endoderm. *In vivo* genetic experiments elucidated multiple signal transduction pathways, which are essential for these differentiation processes[216]. However, can a cell autonomously commit down a differentiation pathway in the absence of environmental stimuli that supports either self-renewal or lineage commitment? Or, to put it in the metaphor of Waddington pathway, does a single “valley” constitute the preferred direction of differentiation for a cell? This concept of a “default pathway” was explored in elegant experiments by Tropepe and colleagues[256]. The culture of mouse embryonic stem cells in the absence of any signaling molecules resulted in a spontaneous differentiation towards the ectoderm. The explanation to this finding was proposed by

Argelaguet and colleagues, who found that the DNA methylation and chromatin accessibility profile of pluripotent and ectoderm cells are more similar to one another than to the primitive streak-derived meso- and endoderm[191]. Active remodeling of chromatin is, therefore, required to enable the differentiation towards the latter two germ layers. In line with these findings, active remodeling of the pattern of DNA methylation occurs during the gastrulation and is guided by the DNA demethylating TET enzymes. Perturbation of their function in the mouse embryo result in defects in gastrulation and the formation of mesoderm-derived tissues[73].

7.3 How can DNA methylation contribute to maintaining hESC pluripotency?

The work presented here explored how the withdrawal of one factor controlling the shape of Waddington landscape, namely DNA methylation, affects the self-renewal of human embryonic stem cells. We found that, unlike the results presented in the “default pathway” model, cells with reduced levels of DNA methylation did not differentiate towards the ectoderm. We believe these results are consistent with the presence of TGFß in the hESC medium, as the early observations of the “default pathway” claimed that the TGFß-superfamily of signal transduction factors act refractory to the default commitment[256]. The human ESCs are cultured in a medium which contains TGFß and FGF - these signaling cues promote self-renewal, but can also induce a lineage commitment. These contradicting roles are reconciled through the action of NANOG, which imposes stasis on lineage commitment[215].

Could the withdrawal of the extracellular signaling cues permit the “default pathway” differentiation of the DNMT1-deficient hESCs? We tested this possibility during the hESC differentiation toward the neural progenitor cell lineage. We observed that these cells were not viable. Whether DNA methylation was necessary for the differentiation towards the ectoderm or for sustaining the ectoderm lineage is a topic for future studies.

We describe how the depletion of DNMT1 in hESCs under normal culturing conditions resulted in an upregulation of the meso- and endoderm-specific transcription factors followed by the expression of the TGFß targets specific for these lineages. Together with our previous study, this work demonstrated that the DNMT1 depletion in hESCs led to the upregulation of genes associated with the embryo posterior fates (meso- and endoderm), such as NODAL or LEFTY1[136,226,257].

Importantly, this effect is not a general response to the DNMT1 depletion, as it was not present in the DNMT1-deficient cells devoid of the TGFß signaling. We hypothesize that the global DNA hypomethylation led to a misinterpretation of the signaling cues – the TGFß mediated signal transduction prompted the incorrect activation of the mesendoderm-specific targets of the SMAD proteins (effectors of the TGFß signaling)[258], instead of exclusively to loci promoting self-renewal.

Can the loss of DNA methylation alter TF binding? *In vitro* and *in vivo* studies investigating the TF specificity to either methylated or unmethylated DNA reported that various TFs show preference to methylated or unmethylated DNA, or not exhibit a preference at all[63,259,260]. The answer is, therefore, highly context-specific. Whether the SMAD family of transcription factors are DNA methylation sensitive is not known. Due to the SMAD binding as obligate heterotrimers, they were not thoroughly characterized in systematic *in vitro* assays. Testing the specificity of the SMAD binding in response to the different signal transduction pathways and with different co-factors is, therefore, an interesting direction for future study.

Curiously, when we introduced additional molecules capable of inducing meso- and endoderm lineage commitment, we observed a more robust transcriptional response in the DNMT1-deficient hESCs compared to control. We hypothesize that the loss of DNA methylation does not activate genes, but rather renders their loci permissive for TF binding, such that newly induced transcription factors (effectors of signal transduction pathways) can easily activate their targets in the more accessible chromatin environment. Investigating whether the global loss of DNA methylation changed the binding specificity of signal transduction pathway effectors or other transcription factors is necessary to better understand the transcriptional changes underlying the DNMT1 depletion in hESCs. Combining our data on the DNA methylation level at the enhancer elements with the information on transcription factor binding and the nascent transcription (for example, after metabolic labelling of RNA with the SLAM-seq method[261]), could provide a comprehensive picture of the stepwise induction of differentiation by environmental factors in the presence and absence of DNA methylation.

DNMT1 depletion has profound effects on the proliferation of hESC, which manifests itself shortly after DNA methylation is depleted throughout the genome. Transcriptional deregulation could be one of the factors that contributes to such loss of fitness. It could act both through the upregulation of negative regulators of the hESC growth, such as LEFTY1, and conflicting transcriptional programs being active simultaneously (for

example, the pluripotency program versus the mesendodermal program). The rapid onset of the loss of fitness of the DNMT1-deficient hESCs could also result from a deregulation of other, transcription-independent mechanisms of maintaining cell homeostasis. Investigating the immediate consequences of DNMT1 depletion in more detail is another area open to future studies.

7.4 DNMT1KO in mouse embryonic stem cells

Reduction of the global level of DNA methylation is an essential step during early embryonic development *in vivo*. Both the cells of the inner cell mass of the blastocyst and the primordial germ cells contain significantly less DNA methylation than somatic cell types. Whether there is a common mechanism guarding these cells from the negative effects of reduced DNA methylation is not known. Surprisingly, the knockout of DNMT1 has a completely different phenotype in both the cell types. Loss of DNMT1 in the PGCs led to a precocious terminal differentiation and, in the long term, infertility[262]. In contrast, the depletion of the maintenance DNA methyltransferase, or even all three DNA methyltransferases, did not affect mouse embryonic stem cell self-renewal, however, it disabled them from differentiating into the three germ layers[94]. In agreement with these observations, DNMT1KO mESCs are unable to form teratomas when injected into immunocompromised mice[263]. Of note, previous studies demonstrated that several factors associated with chromatin repression, such as Polycomb Repressive Complex 2 or Mbd3 subunit of the NuRD deacetylase complex are also dispensable for mouse embryonic stem cell self-renewal, but essential for differentiation and in the somatic lineages[264]. The differences in how chromatin is assembled and maintained in embryonic stem cells compared to somatic cells to tolerate such deep changes in chromatin landscape is an exciting question for future studies.

How mouse embryonic stem cells specifically cope with the loss of DNA? It is possible that even if new transcription factor-binding motifs become accessible in the DNMT1-deficient mESCs, these cells simply lack the transcription factors to bind them and prompt spontaneous differentiation. A study of DNaseI hypersensitive binding sites that arise upon DNA demethylation in mESCs found that few transcription factors expressed in mESCs created novel binding sites[259]. How many TFs in somatic cell lines modify their binding *in vivo* in response to global loss of DNA methylation is not known.

Alternatively, other mechanisms of transcriptional repression, such as by the H3K27me3-mediated silencing (a mark mediated by the Polycomb repressive complex)

could repress hypomethylated loci in the absence of DNMT1[265]. The relationship between DNA methylation, chromatin accessibility and gene expression is complex. DNA methylation and the Polycomb-mediated H3K27me3 chromatin mark are mutually exclusive and, in the absence of DNMTs, the Polycomb loci were found to spread during the transition from primed (more DNA methylation) to naïve (less DNA methylation) stem cell states[265]. Recent study demonstrated that the "confinement" of the Polycomb domains by DNA methylation is necessary for the local compaction and maintaining the higher-order chromatin structure[266]. Strikingly, the depletion of DNMTs led to a loss of some higher order chromatin contacts, however, the impact of these changes on gene expression was very limited[266]. Whether the spreading of H3K27me3 occurs in other cell types with reduced levels of DNA methylation (for example, DNMT1-deficient somatic stem cells), and how that affects both the chromatin structure and gene expression is an interesting direction for future studies.

7.5 DNMT1KO in somatic stem cells

Our findings have pronounced implications for the previous studies of DNMT1KO in somatic stem cells. The propensity of the DNMT1-deficient cells to transdifferentiate or spontaneously differentiate into one of the possible lineages has been reported in multiple lineages (described in detail in the Introduction section). For example, the preference towards lymphoid versus myeloid differentiation in the DNMT1KO HSCs could be dictated by the change in the interpretation of extracellular cues. The interaction with bone marrow stromal cells[267] and the cytokine-inducted transcription mediated by interleukin-7 (IL-7)[268] are both important for the differentiation of lymphoid cells. Curiously, the DNMT1KO phenotype manifested itself only after knockout HSCs were transplanted into wild type donors, but not in the DNMT1KO bone marrow of mutant mice. Whether the DNMT1KO HSCs misinterpret the stimuli from the wild type environment and prematurely differentiate towards the lymphoid lineage remains to be elucidated.

Could the loss of DNA methylation enable signal transduction that does not specifically lead to stem cell self-renewal or differentiation? A study of DNMT1-deficient myoblasts demonstrated that myogenic differentiation is impaired in the DNA hypomethylated cells. Liu and colleagues observed trans-differentiation of the mutant myoblast cells to osteoblasts, which was exacerbated by the BMP signaling[101].

We previously reported that the knock-out of de novo methyltransferase DNMT3A in hESCs led to an aberrant expression of transcription factors (including SOX9) during attempts to differentiate these cells into ecto- or mesoderm[257]. In the experiment described here, the LIF treatment of the DNMT1-deficient hESCs caused an activation of SOX9 – a downstream JAK-STAT pathway target, which is not normally activated during human embryonic stem cell differentiation towards any of the three germ layers. DNA methylation could act as a barrier not only to differentiation, but also other cell fate changes. Whether an aberrant activation of transcription factors in DNMT1 or DNMT3A-deficient cells led to the subsequent binding and regulation of their target loci is an important follow-up question for future consideration.

Altogether, our study presents a possible mechanism by which DNMT1 depletion induces a range of seemingly dissimilar phenotypes across various stem cell types. The knockout of DNA maintenance methyltransferase leads to global DNA hypomethylation which, in turn, could render the chromatin more permissive to transcription factor binding. This would lead to various gene activation depending on the cell type and extracellular environment, thus producing different outcomes.

7.6 Maintaining cell identity

The imperative character of transcription factors in the process of determining cellular identity was demonstrated by Yamanaka and colleagues in the Nobel prize-winning discovery, that ectopic expression of four transcription factors can reprogram fibroblasts into pluripotent stem cells, essentially allowing the cells to “roll upwards” the epigenetic landscape of Waddington[135].

Nevertheless, reprogramming to pluripotency happens with a low efficiency. The process is inhibited by the barriers, which safeguard somatic cell identity. As a byproduct of reprogramming induction, some cells stop at an intermediate stage and fail to reprogram. DNA methylation inhibition using AzaC resulted in increased reprogramming efficiency[269]. Moreover, partially reprogrammed cells which failed to form iPSCs were characterized by high DNA methylation levels at key pluripotency loci (such as POU5F1) and were able to reprogram completely in the presence of DNA methylation inhibitor[269].

Recent RNAi screens identified several other factors, which are refractory to iPSC generation. Chromatin assembly factor (CAF-1) is a histone chaperone complex, the suppression of which loosened the chromatin structure, reduced H3K9me3 levels at

regions normally resistant to reprogramming and thus increased both speed and efficiency of reprogramming. Maintaining proper chromatin structure is also refractory to changing the cell fate, the equivalent of “rolling sideways” in the epigenetic landscape of Waddington. Increasing chromatin accessibility by the means of knocking down the subunits of CAF-1 complex enhanced transdifferentiation from the B cells to macrophages and from the MEFs to neurons[270].

Multiple studies reported a negative correlation between DNA methylation and chromatin accessibility[271,272]. Notably, when the two variables were compared in pluripotent stem cells induced to differentiate, the correlation between them decreased[272]. This suggests that as the cells commit to a somatic lineage, the methylated DNA becomes more prevalent. We hypothesize that DNA demethylation during reprogramming induced by Mikkelsen and colleagues, as well as by the DNMT1 depletion in this study, promotes chromatin relaxation and increased binding of transcription factors[269]. In agreement with our hypothesis, loss of nearly all (95%) total DNA methylation in a DNMT hypomorphic cell line increased the number of accessible sites[272].

7.7 DNA hypomethylation and cancer

The increased accessibility of unmethylated DNA to transcription factors could not only contribute to reprogramming or transdifferentiation, but generally increase transcriptional promiscuity, which could be a potential mechanism of tumorigenesis.

The DNA methylation landscape in cancer cells is characteristic – high local hypermethylation of CpG islands combined with global hypomethylation[273,274]. In contrast, the bulk of a healthy methylome is hypermethylated, with puncta of unmethylated CpG islands. How that change in the epigenome of cancer cells is achieved is not yet understood, however, mutations in DNMTs, as well as their overexpression, have been identified in multiple cancer types[275–277]. Furthermore, previous studies in patients with Chronic Lymphocytic Leukemia (CLL) reported that changes in the global DNA methylation pattern precede cancer[278]. A different study of tumorigenesis in the haematopoietic lineage identified mutations in DNMT3A which confer repopulation advantage to the hematopoietic stem cells and lead to Acute Myeloid Leukemia (AML)[279]. Whether the changes in the DNA methylation landscape are causative to the tumorigenic transformation and by what mechanism is not known. However, our previous study identified that transcriptional variability is increased in DNMT3A and 3B/DKO knockout hESCs providing a potential explanation[257]. The global

DNA hypomethylation of pre-cancer cell could result in an increased transcription of various genes followed by positive selection for cells with (1) high expression of self-renewal genes or (2) silencing of tumor suppressor genes via promoter hypermethylation. As both the tissue vascularization and the cell growth are dependent on extracellular signaling, investigation into whether a reduction of global DNA methylation levels could lead to an increased response to these signals is an interesting direction for future studies.

Despite mutations in DNA methyltransferases reported in cancer, DNA methylation is indispensable for transformed cells. In fact, the DNMT1-depletion in cancer cell lines results in DNA damage and apoptosis. Curiously, no apoptosis or DNA damage were reported in most of DNMT1KO somatic stem cells[102,103,105]. Likewise, we did not observe such a phenotype in the hESCs, despite the extensive investigation of the DNA damage markers, as well as the mitotic abnormalities. This difference in the phenotype could stem from the use of different models. Cancer cells are known to selectively hypermethylate the CGI promoters of tumor suppressor genes[280], such as at the Ink4a-ARF locus[281]. The loss of DNMT1 could lead to the de-repression of these factors that would eventually mediate senescence and apoptosis via imposing cell cycle arrest.

7.8 The role of DNA methylation maintenance in embryogenesis

Nearly three decades ago, Li and colleagues demonstrated that DNMT1 is essential for the mouse embryonic development[76]. Crossing the DNMT1 heterozygous mice yielded no homozygous pups brought to term. Further investigation revealed an embryonic lethality before the embryonic day 12.5 (E12.5). Compared to their wild-type littermates, the E10 DNMT1-knockout embryos were smaller and developmentally delayed. Abnormalities observed in these embryos included also defects in somitogenesis, neural tube closure, allantois and no blood formation, accompanied by an increase in apoptotic cells[76]. How the loss of DNA methylation resulted in this broad range of developmental phenotypes is not known.

The authors pointed out that the allantois abnormal structures were observed before in parthenogenetic and gynogenetic embryos[282] and were thus likely caused by loss of imprinting. One of the crucial roles of DNA methylation is maintaining monoallelic expression of some genes known as imprinted genes. Maternal contribution of oocyte-specific isoform of Dnmt1 (Dnmt1o) mRNA is essential for the imprint preservation during the pre-implantation reprogramming of the global DNA methylation patterns[283]. Lack of

maternal Dnmt1o resulted in loss of imprinting (LOI) on some imprint control regions and a variety of phenotypes. The study reported embryo resorption in the last trimester of pregnancy or stunted growth of rare pups delivered to term, despite maintaining a high level of global DNA methylation and the somatic version of Dnmt1[283]. In agreement with these results, a knockout of Zinc finger protein 57 (Zfp57), which is responsible for maintaining imprints on many loci, phenocopies the Dnmt1o depletion[284]. Altogether, the maintenance of imprinted genes is essential for embryonic development, however, the LOI does not fully explain the lethality around E10 in the full Dnmt1KO embryos.

Although the Dnmt1KO embryos were noted to develop normally from E1-7, the analysis of E9.5 noted no blood or blood vessel formation, a process which begins at E7, with the migration of mesoderm cells to the yolk sac to form hemangioblasts[285]. These progenitors of blood and epithelium develop outside of the embryo in so-called blood islands and form structures essential for the transport of nutrients to the developing organism. A knockout of Flk1, a gene encoding vascular endothelial growth factor (VEGF) receptor is essential for the formation of hemangioblasts. Homozygous Flk1 knockout embryos fail to form blood vessels and blood and die in utero around E9.5, mid-somitogenesis[286]. The Dnmt1KO embryo lethality could, therefore, also result from a lack of blood and vasculature. Whether such lack of blood formation in the Dnmt1KO embryo occurs due to the disruption of the VEGF-mediated gene expression or via another mechanism remains to be determined. Further studies on the role of DNA methylation in the VEGF signaling and early hematopoiesis, for example using *in vitro* cultured hemangioblasts, could shed light on the developmental lethality and delay of Dnmt1-deficient embryos.

The indispensability of DNA methylation for embryo development and tissue homeostasis is well-documented, yet the precise mechanisms by which DNA methylation exerts its functions in these processes is still unknown. Our work presented here opens new study trails for *in vitro* and *in vivo* research, which could bring insight into how chromatin integrates extracellular signals and helps maintain cell identity, which has great consequences for embryonic development and tumorigenesis.

8. REFERENCES

1. Kouzarides, T. Chromatin Modifications and Their Function. *Cell* **128**, 693–705 (2007).
2. Johnson, T. B. & Coghill, R. D. Researches on pyrimidines. C111. The discovery of 5-methyl-cytosine in tuberculinic acid, the nucleic acid of the tubercle bacillus. *J. Am. Chem. Soc.* **47**, 2838–2844 (1925).
3. Zemach, A., McDaniel, I. E., Silva, P. & Zilberman, D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science (80-.).* **328**, 916–919 (2010).
4. Zhang, X. *et al.* Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in Arabidopsis. *Cell* **126**, 1189–1201 (2006).
5. Selker, E. U. *et al.* The methylated component of the Neurospora crassa genome. *Nature* **422**, 893–897 (2003).
6. Eckhardt, F. *et al.* DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat. Genet.* **38**, 1378–1385 (2006).
7. Zemach, A. & Zilberman, D. Evolution of eukaryotic DNA methylation and the pursuit of safer sex. *Curr. Biol.* **20**, R780–R785 (2010).
8. Ehrich, M. *et al.* Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15785–15790 (2005).
9. Frommer, M. *et al.* A genomic sequencing protocol that yields a positive display of 5- methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 1827–1831 (1992).
10. Doskočil, J. & Šorm, F. Distribution of 5-methylcytosine in pyrimidine sequences of deoxyribonucleic acids. *Biochim. Biophys. Acta - Spec. Sect. Nucleic Acids Relat. Subj.* **55**, 953–959 (1962).
11. Lyko, F. The DNA methyltransferase family: A versatile toolkit for epigenetic regulation. *Nature Reviews Genetics* **19**, 81–92 (2018).
12. Okano, M., Xie, S. & Li, E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases Non-invasive sexing of preimplantation stage mammalian embryos. *Nat. Am. Inc.* **19**, 219–220 (1998).
13. Barau, J., Zamudio, N. & Guillou, F. The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. **2**,
14. Cheng, X. AdoMet-dependent methylation, DNA methyltransferases and base flipping. *Nucleic Acids Res.* **29**, 3784–3795 (2001).
15. Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B. & Bestor, T. H. Dnmt3L and the establishment of maternal genomic imprints. *Science (80-.).* **294**, 2536–2539 (2001).
16. Zhang, Y. *et al.* Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. *Nucleic Acids Res.* **38**, 4246–4253 (2010).
17. Guo, X. *et al.* Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* **517**, 640–644 (2015).

18. Morselli, M. *et al.* In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse. *Elife* **2015**, 1–21 (2015).

19. Strahl, B. D. *et al.* Set2 Is a Nucleosomal Histone H3-Selective Methyltransferase That Mediates Transcriptional Repression. *Mol. Cell. Biol.* **22**, 1298–1306 (2002).

20. Holliday, R. & Pugh, J. E. DNA modification mechanisms and gene activity during development. *Science (80-.).* **187**, 226–232 (1975).

21. Yen, R.-W. C. *et al.* Isolation and characterization of the cDNA encoding human DNA methyltransferase. *Nucleic Acids Res.* **20**, 2287–2291 (1992).

22. Jeltsch, A. & Jurkowska, R. Z. Allosteric control of mammalian DNA methyltransferases – a new regulatory paradigm. 1–20 (2016). doi:10.1093/nar/gkw723

23. Song, J., Teplova, M., Ishibe-Murakami, S. & Patel, D. J. Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. *Science (80-.).* **335**, 709–712 (2012).

24. Syeda, F. *et al.* The replication focus targeting sequence (RFTS) domain is a DNA-competitive inhibitor of Dnmt1. *J. Biol. Chem.* **286**, 15344–15351 (2011).

25. Bostick, M. *et al.* UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. *Science (80-.).* **317**, 1760–1764 (2007).

26. Sharif, J. *et al.* The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* **450**, 908–912 (2007).

27. Nady, N. *et al.* Recognition of multivalent histone states associated with heterochromatin by UHRF1 protein. *J. Biol. Chem.* **286**, 24300–24311 (2011).

28. Lee, J. H., Voo, K. S. & Skalnik, D. G. Identification and Characterization of the DNA Binding Domain of CpG-binding Protein. *J. Biol. Chem.* **276**, 44669–44676 (2001).

29. Bashtrykov, P. *et al.* Specificity of dnmt1 for methylation of hemimethylated CpG sites resides in its catalytic domain. *Chem. Biol.* **19**, 572–578 (2012).

30. Iida, T. *et al.* PCNA clamp facilitates action of DNA cytosine methyltransferase 1 on hemimethylated DNA. *Genes to Cells* **7**, 997–1007 (2002).

31. Estève, P. O. *et al.* Regulation of DNMT1 stability through SET7-mediated lysine methylation in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 5076–5081 (2009).

32. Estéve, P. O. *et al.* A methylation and phosphorylation switch between an adjacent lysine and serine determines human DNMT1 stability. *Nat. Struct. Mol. Biol.* **18**, 42–49 (2011).

33. Qin, W., Leonhardt, H. & Spada, F. Usp7 and Uhrf1 control ubiquitination and stability of the maintenance DNA methyltransferase Dnmt1. *J. Cell. Biochem.* **112**, 439–444 (2011).

34. Cheng, J. *et al.* Molecular mechanism for USP7-mediated DNMT1 stabilization by acetylation. *Nat. Commun.* **6**, 1–11 (2015).

35. Du, Z. *et al.* DNMT1 stability is regulated by proteins coordinating deubiquitination and acetylation-driven ubiquitination. *Sci. Signal.* **3**, 1–11 (2010).

36. Bird, A., Taggart, M., Frommer, M., Miller, O. J. & Macleod, D. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell* **40**, 91–99 (1985).

37. Saxonov, S., Berg, P. & Brutlag, D. L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 1412–1417 (2006).

38. Craig Venter, J. *et al.* The sequence of the human genome. *Science (80-.).* **291**, 1304–1351 (2001).

39. Li, E. & Zhang, Y. DNA methylation in mammals. *Cold Spring Harb. Perspect. Biol.* **6**, (2014).

40. Lyon, M. F. Gene Action in the X-chromosom (Mus musculus L.). *Nature* **190**, 372–373 (1961).

41. Chaumeil, J., Baccon, P. Le, Wutz, A. & Heard, E. A novel role for Xist RNA in the formation of a repressive nuclear. *Genes Dev.* **20**, 2223–2237 (2006).

42. Gendrel, A. V. *et al.* Smchd1-Dependent and -Independent Pathways Determine Developmental Dynamics of CpG Island Methylation on the Inactive X Chromosome. *Dev. Cell* **23**, 265–279 (2012).

43. John, R. M. & Surani, M. A. Imprinted genes and regulation of gene expression by epigenetic inheritance. *Curr. Opin. Cell Biol.* **8**, 348–353 (1996).

44. Thorvaldsen, J. L., Duran, K. L. & Bartolomei, M. S. Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2. *Genes Dev.* **12**, 3693–3702 (1998).

45. Cassidy, S. B. & Schwartz, S. Prader-Willi and Angelman syndromes: Disorders of genomic imprinting. *Medicine* **77**, 140–151 (1998).

46. SILVER, H. K., KIYASU, W., GEORGE, J. & DEAMER, W. C. Syndrome of congenital hemihypertrophy, shortness of stature, and elevated urinary gonadotropins. *Pediatrics* **12**, 368–76 (1953).

47. Borgel, J. *et al.* Targets and dynamics of promoter DNA methylation during early mouse development. *Nat. Genet.* **42**, 1093–1100 (2010).

48. Kazazian, H. H. Mobile Elements: Drivers of Genome Evolution. *Science (80-.).* **303**, 1626–1632 (2004).

49. Zamudio, N. *et al.* DNA methylation restrains transposons from adopting a chromatin signature permissive for meiotic recombination. *Genes Dev.* **29**, 1256–1270 (2015).

50. Aravin, A. A., Hannon, G. J. & Brennecke, J. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science (80-.).* **318**, 761–764 (2007).

51. Rowe, H. M. *et al.* KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**, 237–240 (2010).

52. Sripathy, S. P., Stevens, J. & Schultz, D. C. The KAP1 Corepressor Functions To Coordinate the Assembly of De Novo HP1-Demarcated Microenvironments of Heterochromatin Required for KRAB Zinc Finger Protein-Mediated Transcriptional Repression. *Mol. Cell. Biol.* **26**, 8623–8638 (2006).

53. Rowe, H. M. *et al.* De novo DNA methylation of endogenous retroviruses is

shaped by KRAB-ZFPs/KAP1 and ESET. *Development* **140**, 519–529 (2013).

54. Bouzinba-Segard, H., Guais, A. & Francastel, C. Accumulation of small murine minor satellite transcripts leads to impaired centromeric architecture and function. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 8709–8714 (2006).

55. Saksouk, N., Simboeck, E. & Déjardin, J. Constitutive heterochromatin formation and transcription in mammals. *Epigenetics and Chromatin* **8**, 1–17 (2015).

56. Bersani, F. *et al.* Pericentromeric satellite repeat expansions through RNA-derived DNA intermediates in cancer. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 15148–15153 (2015).

57. Maunakea, A. K. *et al.* Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **466**, 253–257 (2010).

58. Maunakea, A. K., Chepelev, I., Cui, K. & Zhao, K. Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res.* **23**, 1256–1269 (2013).

59. Toubiana, S. *et al.* Persistent epigenetic memory impedes rescue of the telomeric phenotype in human ICF iPSCs following DNMT3B correction. *Elife* **8**, 1–28 (2019).

60. Meehan, R. R., Lewis, J. D., McKay, S., Kleiner, E. L. & Bird, A. P. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* **58**, 499–507 (1989).

61. Nan, X. *et al.* Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**, 386–9 (1998).

62. Zhao, X. *et al.* Mice lacking methyl-CpG binding protein 1 have deficits in adult neurogenesis and hippocampal function. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 6777–6782 (2003).

63. Yin, Y. *et al.* Impact of cytosine methylation on DNA binding specificities of human transcription factors. **2239**, (2017).

64. Stadler, M. B. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**, 490–495 (2011).

65. Holliday, R. & Grigg, G. W. DNA methylation and mutation. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **285**, 61–67 (1993).

66. Santos, F. *et al.* Active demethylation in mouse zygotes involves cytosine deamination and base excision repair. *Epigenetics and Chromatin* **6**, 1–12 (2013).

67. Cortellino, S. *et al.* Demethylation by Linked Deamination-Base Excision Repair. *Cell* **146**, 67–79 (2011).

68. Rošić, S. *et al.* Evolutionary analysis indicates that DNA alkylation damage is a byproduct of cytosine DNA methyltransferase activity. *Nat. Genet.* **50**, 452–459 (2018).

69. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science (80-.).* **324**, 930–935 (2009).

70. Ito, S. *et al.* Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science (80-.).* **333**, 1300–1303 (2011).

71. He, Y.-F. *et al.* Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science (80-.).* **333**, 1303–1307 (2011).

72. Wiehle, L. *et al.* Tet1 and Tet2 Protect DNA Methylation Canyons against Hypermethylation. *Mol. Cell. Biol.* **36**, 452–461 (2016).

73. Dai, H. Q. *et al.* TET-mediated DNA demethylation controls gastrulation by regulating Lefty-Nodal signalling. *Nature* **538**, 528–532 (2016).

74. Smith, Z. D. *et al.* A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484**, 339–344 (2012).

75. Guo, F. *et al.* Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. *Cell Stem Cell* **15**, 447–459 (2014).

76. Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).

77. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. **99**, 247–257 (1999).

78. Smith, Z. D. *et al.* DNA methylation dynamics of the human preimplantation embryo. *Nature* **511**, 611–615 (2014).

79. Zhang, Y. *et al.* Dynamic epigenomic landscapes during early lineage specification in mouse embryos. *Nat. Genet.* **50**, 96–105 (2018).

80. Deniz, Ö., Frost, J. M. & Branco, M. R. Regulation of transposable elements by DNA modifications. *Nat. Rev. Genet.* **20**, (2019).

81. Kagiwada, S., Kurimoto, K., Hirota, T., Yamaji, M. & Saitou, M. Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. *EMBO J.* **32**, 340–353 (2013).

82. Yamaguchi, S., Shen, L., Liu, Y., Sendler, D. & Zhang, Y. Role of Tet1 in erasure of genomic imprinting. *Nature* **504**, 460–464 (2013).

83. Hackett, J. A. *et al.* Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science* **339**, 448–52 (2013).

84. Zhu, P. *et al.* Single-cell DNA methylome sequencing of human preimplantation embryos. *Nat. Genet.* **50**, 12–19 (2018).

85. Smallwood, S. A. *et al.* Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat. Genet.* **43**, 811–814 (2011).

86. Wang, L. *et al.* Programming and inheritance of parental DNA methylomes in mammals. *Cell* **157**, 979–991 (2014).

87. Gaudet, F. *et al.* Induction of tumors in mice by genomic hypomethylation. *Science* **300**, 489–92 (2003).

88. Martello, G. & Smith, A. The Nature of Embryonic Stem Cells. *Annu. Rev. Cell Dev. Biol.* **30**, 647–675 (2014).

89. Bryja, V., Bonilla, S. & Arenas, E. Derivation of mouse embryonic stem cells. *Nat. Protoc.* **1**, 2082–2087 (2006).

90. Ludwig, T. E. *et al.* Derivation of human embryonic stem cells in defined conditions. *Nat. Biotechnol.* **24**, 185–187 (2006).

91. Ginis, I. *et al.* Differences between human and mouse embryonic stem cells. *Dev. Biol.* **269**, 360–380 (2004).

92. Adewumi, O. *et al.* Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. *Nat. Biotechnol.* **25**, 803–816 (2007).

93. Altun, G., Loring, J. F. & Laurent, L. C. DNA methylation in embryonic stem cells. *J. Cell. Biochem.* **23**, n/a-n/a (2009).

94. Tsumura, A. *et al.* Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes to Cells* **11**, 805–814 (2006).

95. Liao, J. *et al.* Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. *Nat. Genet.* **47**, 469–478 (2015).

96. Jackson, M. *et al.* Severe Global DNA Hypomethylation Blocks Differentiation and Induces Histone Hyperacetylation in Embryonic Stem Cells. **24**, 8862–8871 (2004).

97. Sauer, B. & Henderson, N. Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 5166–5170 (1988).

98. Jackson-Grusby, L. *et al.* Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nat. Genet.* **27**, 31–9 (2001).

99. Georgia, S. & Kanji, M. DNMT1 represses p53 to maintain progenitor cell survival during pancreatic organogenesis. 372–377 (2013). doi:10.1101/gad.207001.112.sing

100. Elliott, E. N., Sheaffer, K. L., Schug, J., Stappenbeck, T. S. & Kaestner, K. H. Dnmt1 is essential to maintain progenitors in the perinatal intestinal epithelium. 2163–2172 (2015). doi:10.1242/dev.117341

101. Liu, R., Kim, K., Jung, Y. & Park, I. Dnmt1 regulates the myogenic lineage specification of muscle stem cells. *Nat. Publ. Gr.* 1–9 (2016). doi:10.1038/srep35355

102. Trowbridge, J. J., Snow, J. W., Kim, J. & Orkin, S. H. DNA Methyltransferase 1 Is Essential for and Uniquely Regulates Hematopoietic Stem and Progenitor Cells. *Cell Stem Cell* **5**, 442–449 (2009).

103. Fan, G. *et al.* DNA methylation controls the timing of astrogliogenesis through regulation of JAK-STAT signaling. 3345–3356 (2002). doi:10.1242/dev.01912

104. Fan, G. *et al.* DNA hypomethylation perturbs the function and survival of CNS neurons in postnatal animals. *J. Neurosci.* **21**, 788–797 (2001).

105. Sen, G. L., Reuter, J. A., Webster, D. E., Zhu, L. & Khavari, P. A. DNMT1 maintains progenitor function in self-renewing somatic tissue. *Nature* **463**, 563–567 (2010).

106. Unterberger, A., Andrews, S. D., Weaver, I. C. G. & Szyf, M. DNA Methyltransferase 1 Knockdown Activates a Replication Stress Checkpoint. **26**, 7575–7586 (2006).

107. Loughery, J. E. P. *et al.* DNMT1 deficiency triggers mismatch repair defects in human cells through depletion of repair protein levels in a process involving the DNA damage response. **20**, 3241–3255 (2011).

108. Chen, T. *et al.* Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells. *Nat. Genet.* **39**, 391–6 (2007).

109. Chen, A. E. *et al.* Optimal Timing of Inner Cell Mass Isolation Increases the Efficiency of Human Embryonic Stem Cell Derivation and Allows Generation of Sibling Cell Lines. *Cell Stem Cell* **4**, 103–106 (2009).

110. Tandon, R. *et al.* Generation of two human isogenic iPSC lines from fetal dermal fibroblasts. *Stem Cell Res.* **33**, 120–124 (2018).

111. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).

112. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).

113. Canals, I. *et al.* Rapid and efficient induction of functional astrocytes from human pluripotent stem cells. *Nat. Methods* **15**, 693–696 (2018).

114. Martyn, I., Kanno, T. Y., Ruzo, A., Siggia, E. D. & Brivanlou, A. H. Self-organization of a human organizer by combined Wnt and Nodal signaling. *Nature* **558**, 132–135 (2018).

115. Zhang, Y. *et al.* Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* **78**, 785–798 (2013).

116. Ho, S.-M. *et al.* Rapid Ngn2-induction of excitatory neurons from hiPSC-derived neural progenitor cells. *Methods* **101**, 113–124 (2016).

117. Grow, E. J. *et al.* Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. (2015). doi:10.1038/nature14308

118. Bogerd, H. P. *et al.* Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 8780–8785 (2006).

119. Macaulay, I. C. *et al.* G&T-seq: Parallel sequencing of single-cell genomes and transcriptomes. *Nat. Methods* **12**, 519–522 (2015).

120. Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1100 (2013).

121. Kangeyan, D. *et al.* A (fire)cloud-based DNA methylation data preprocessing and quality control platform. *BMC Bioinformatics* **20**, 1–5 (2019).

122. Krueger, F. & Andrews, S. R. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).

123. Smyth, G. K. limma: Linear Models for Microarray Data. in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* (eds. Gentleman, R., Carey, V. J., Huber, W., Irizarry, R. A. & Dudoit, S.) 397–420 (Springer New York, 2005). doi:10.1007/0-387-29362-0_23

124. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

125. Carbon, S. *et al.* The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–D338 (2019).

126. Ashburner, M. *et al.* The Gene Ontology Consortium, Michael Ashburner1, Catherine A. Ball3, Judith A. Blake4, David Botstein3, Heather Butler1, J.

Michael Cherry3, Allan P. Davis4, Kara Dolinski3, Selina S. Dwight3, Janan T. Eppig4, Midori A. Harris3, David P. Hill4, Laurie Is. *Nat. Genet.* **25**, 25–29 (2000).

127. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10 (2011).

128. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

129. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).

130. Landan, G. *et al.* Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. *Nat. Genet.* **44**, 1207–1214 (2012).

131. Eden, A., Gaudet, F., Waghmare, A. & Jaenisch, R. Chromosomal instability and tumors promoted by DNA hypomethylation. *Science* **300**, 455 (2003).

132. Thomson, J. A. Embryonic stem cell lines derived from human blastocysts. *Science (80-.).* **282**, 1145–1147 (1998).

133. Tesar, P. J. *et al.* New cell lines from mouse epiblast share defining features with human embryonic stem cells. **448**, 2–8 (2007).

134. Gossen, M. & Bujard, H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 5547–5551 (1992).

135. Takahashi, K. *et al.* Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell* **131**, 861–872 (2007).

136. Bock, C. *et al.* Reference maps of human es and ips cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* **144**, 439–452 (2011).

137. Guenther, M. G. *et al.* Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. *Cell Stem Cell* **7**, 249–257 (2010).

138. Sakamoto, K. M. *et al.* Protacs: Chimeric molecules that target proteins to the Skp1-Cullin-F box complex for ubiquitination and degradation. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 8554–8559 (2001).

139. Nabet, B. *et al.* The dTAG system for immediate and target-specific protein degradation. *Nat. Chem. Biol.* **14**, 431–441 (2018).

140. Weintraub, A. S. *et al.* YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell* **171**, 1573-1588.e28 (2017).

141. Mészáros, B., Erdös, G. & Dosztányi, Z. IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **46**, W329–W337 (2018).

142. Mészáros, B., Simon, I. & Dosztányi, Z. Prediction of protein binding regions in disordered proteins. *PLoS Comput. Biol.* **5**, e1000376 (2009).

143. Schermelleh, L. *et al.* Trapped in action: Direct visualization of DNA methyltransferase activity in living cells. *Nat. Methods* **2**, 751–756 (2005).

144. Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click Chemistry: Diverse Chemical

Function from a Few Good Reactions. *Angew. Chemie - Int. Ed.* **40**, 2004–2021 (2001).

145. Buck, S. B. *et al.* Detection of S-phase cell cycle progression using 5-ethynyl-2′-deoxyuridine incorporation with click chemistry, an alternative to using 5-bromo-2′-deoxyuridine antibodies. *Biotechniques* **44**, 927–929 (2008).

146. Latt, S. & Gail, S. SPECTRAL STUDIES ON FOR FLUORESCENT ACID AND SYNTHESIS RELATED DETECTION OF Absorption , fluorescence and circular dichroism deoxyribonucleic acid (DNA) complexes are consistent of dye-binding One type , which persists is highly specific for adenine-thy. *J. Histochem. Cytochem.* **24**, 24–33 (1976).

147. Parish, C. R. Fluorescent dyes for lymphocyte migration and proliferation studies. *Immunol. Cell Biol.* **77**, 499–508 (1999).

148. Kerr, J. F., Wyllie, A. H. & Currie, A. R. Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. *Br. J. Cancer* **26**, 239–57 (1972).

149. Gorczyca, W., Traganos, F., Jesionowska, H. & Darzynkiewicz, Z. Presence of DNA Strand Breaks and Increased Sensitivity of DNA in Situ to Denaturation in Abnormal Human Sperm Cells: Analogy to Apoptosis of Somatic Cells. *Exp. Cell Res.* **207**, 202–205 (1993).

150. Juttermann, R., Li, E. & Jaenisch, R. Toxicity of 5-aza-2'-deoxycytidine to mammalian cells is mediated primarily by covalent trapping of DNA methyltransferase rather than DNA demethylation. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 11797–11801 (1994).

151. Kaji, K. *et al.* DNMT1 is a required genomic regulator for murine liver histogenesis and regeneration. *Hepatology* **64**, 582–598 (2016).

152. D'Amours, D. & Jackson, S. P. The Mre11 complex: At the crossroads of DNA repair and checkpoint signalling. *Nat. Rev. Mol. Cell Biol.* **3**, 317–327 (2002).

153. Grenon, M., Gilbert, C. & Lowndes, N. F. Checkpoint activation in response to double-strand breaks requires the Mre11/Rad50/Xrs2 complex. *Nat. Cell Biol.* **3**, 844–847 (2001).

154. Zhu, M., Zhao, H., Limbo, O. & Russell, P. Mre11 complex links sister chromatids to promote repair of a collapsed replication fork. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 8793–8798 (2018).

155. González-Barrera, S., Cortés-Ledesma, F., Wellinger, R. E. & Aguilera, A. Equal sister chromatid exchange is a major mechanism of double-strand break repair in yeast. *Mol. Cell* **11**, 1661–1671 (2003).

156. Lee, J. H. & Paull, T. T. Direct Activation of the ATM Protein Kinase by the Mre11/Rad50/Nbs1 Complex. *Science (80-.).* **304**, 93–96 (2004).

157. Xie, A., Kwok, A. & Scully, R. Role of mammalian Mre11 in classical and alternative nonhomologous end joining. *Nat. Struct. Mol. Biol.* **16**, 814–818 (2009).

158. Wienert, B. *et al.* Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq. *Science (80-.).* **364**, 286–289 (2019).

159. Ehrenfeld, G. M. *et al.* Copper-Dependent Cleavage of DNA by Bleomycin. *Biochemistry* **26**, 931–942 (1987).

160. Sharma, A., Singh, K. & Almasan, A. Histone H2AX Phosphorylation: A Marker for DNA Damage. in *DNA Repair Protocols* (ed. Bjergbæk, L.) 613–626 (Humana Press, 2012). doi:10.1007/978-1-61779-998-3_40

161. Maraschio, P., Zuffardi, O., Dalla Fior, T. & Tiepolo, L. Immunodeficiency, centromeric heterochromatin instability of chromosomes 1, 9, and 16, and facial anomalies: The ICF syndrome. *J. Med. Genet.* **25**, 173–180 (1988).

162. Wijmenga, C. *et al.* Localization of the ICF syndrome to chromosome 20 by homozygosity mapping. *Am. J. Hum. Genet.* **63**, 803–809 (1998).

163. Tiepolo, L. *et al.* Multibranched chromosomes 1, 9, and 16 in a patient with combined IgA and IgE deficiency. *Hum. Genet.* **51**, 127–137 (1979).

164. Tucker, K. L. *et al.* Germ-line passage is required for establishment of methylation and expression patterns of imprinted but not of nonimprinted genes. *Genes Dev.* **10**, 1008–1020 (1996).

165. Gopalakrishnan, S., Sullivan, B. A., Trazzi, S., Della Valle, G. & Robertson, K. D. DNMT3B interacts with constitutive centromere protein CENP-C to modulate DNA methylation and the histone code at centromeric regions. *Hum. Mol. Genet.* **18**, 3178–3193 (2009).

166. Levine, M. S. & Holland, A. J. The impact of mitotic errors on cell proliferation and tumorigenesis. *Genes Dev.* **32**, 620–638 (2018).

167. Chan, Y. W., Fugger, K. & West, S. C. Unresolved recombination intermediates lead to ultra-fine anaphase bridges, chromosome breaks and aberrations. *Nat. Cell Biol.* **20**, 92–103 (2018).

168. Chan, K. L., North, P. S. & Hickson, I. D. BLM is required for faithful chromosome segregation and its localization defines a class of ultrafine anaphase bridges. *EMBO J.* **26**, 3397–3409 (2007).

169. Rouzeau, S. *et al.* Bloom's syndrome and PICH helicases cooperate with topoisomerase IIα in centromere disjunction before anaphase. *PLoS One* **7**, (2012).

170. Hsiang, Y. H., Hertzberg, R., Hecht, S. & Liu, L. F. Camptothecin induces protein-linked DNA breaks via mammalian DNA topoisomerase I. *J. Biol. Chem.* **260**, 14873–14878 (1985).

171. Glover, T. W., Berger, C., Coyle, J. & Echo, B. DNA polymerase α inhibition by aphidicolin induces gaps and breaks at common fragile sites in human chromosomes. *Hum. Genet.* **67**, 136–142 (1984).

172. Chan, K. L., Palmai-Pallag, T., Ying, S. & Hickson, I. D. Replication stress induces sister-chromatid bridging at fragile site loci in mitosis. *Nat. Cell Biol.* **11**, 753–60 (2009).

173. Lukas, C. *et al.* 53BP1 nuclear bodies form around DNA lesions generated by mitotic transmission of chromosomes under replication stress. *Nat. Cell Biol.* **13**, 243–53 (2011).

174. Mckinley, K. L. & Cheeseman, I. M. Knockouts Reveals the Diversity of p53-Dependent Responses to Cell-Cycle Defects Resource Large-Scale Analysis of CRISPR / Cas9 Cell-Cycle Knockouts Reveals the Diversity of p53-Dependent Responses to Cell-Cycle Defects. *Dev. Cell* **40**, 405-420.e2 (2017).

175. Hafner, A. The multiple mechanisms that regulate p53 activity and cell fate. *Nat.*

Rev. Mol. Cell Biol. **20**, (2019).

176. Sharif, J. *et al.* Activation of Endogenous Retroviruses in Dnmt1−/− ESCs Involves Disruption of SETDB1-Mediated Repression by NP95 Binding to Hemimethylated DNA. *Cell Stem Cell* **19**, 81–94 (2016).

177. Mills, R. E., Bennett, E. A., Iskow, R. C. & Devine, S. E. Which transposable elements are active in the human genome ? **23**, (2007).

178. Brouha, B. *et al.* Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 5280–5285 (2003).

179. Belshaw, R. *et al.* Genomewide screening reveals high levels of insertional polymorphism in the human endogenous retrovirus family HERV-K(HML2): implications for present-day activity. *J. Virol.* **79**, 12507–14 (2005).

180. Chen, J. M., Chuzhanova, N., Stenson, P. D., Férec, C. & Cooper, D. N. Meta-analysis of gross insertions causing human genetic disease: Novel mutational mechanisms and the role of replication slippage. *Hum. Mutat.* **25**, 207–221 (2005).

181. Marchetto, M. C. N. *et al.* Differential L1 regulation in pluripotent stem cells of humans and apes. *Nature* **503**, 525–529 (2013).

182. Ting, D. T. *et al.* Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. *Science* **331**, 593–6 (2011).

183. Nakano, M., Okamoto, Y., Ohzeki, J. I. & Masumoto, H. Epigenetic assembly of centromeric chromatin at ectopic α-satellite sites on human chromosomes. *J. Cell Sci.* **116**, 4021–4034 (2003).

184. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).

185. Mohn, F., Weber, M., Schübeler, D. & Roloff, T.-C. Methylated DNA Immunoprecipitation (MeDIP). in *DNA Methylation: Methods and Protocols* (ed. Tost, J.) 55–64 (Humana Press, 2009). doi:10.1007/978-1-59745-522-0_5

186. Meissner, A. *et al.* Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* **33**, 5868–5877 (2005).

187. Tasic, B. *et al.* Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**, 72–78 (2018).

188. Pijuan-Sala, B. *et al.* A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490–495 (2019).

189. Hou, Y. *et al.* Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Res.* **26**, 304–319 (2016).

190. Rooijers, K. *et al.* Simultaneous quantification of protein–DNA contacts and transcriptomes in single cells. *Nat. Biotechnol.* **37**, 766–772 (2019).

191. Argelaguet, R. *et al.* Multi-omics profiling of mouse gastrulation at single-cell resolution. *Nature* (2019). doi:10.1038/s41586-019-1825-8

192. Guo, H. *et al.* Single-Cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. *Genome Res.* **23**, 2126–2135 (2013).

193. Weber, M. *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* **39**, 457–466 (2007).

194. Guo, H. *et al.* Profiling DNA methylome landscapes of mammalian cells with single-cell reduced-representation bisulfite sequencing. *Nat. Protoc.* **10**, 645–59 (2015).

195. Smallwood, S. A. *et al.* Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* **11**, 817–820 (2014).

196. Walsh, C. P., Chaillet, J. R. & Bestor, T. H. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat. Genet.* **20**, 116–7 (1998).

197. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

198. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **21**, 381–395 (2011).

199. Lehnertz, B. *et al.* Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr. Biol.* **13**, 1192–200 (2003).

200. Ziller, M. J. *et al.* Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500**, 477–481 (2013).

201. Law, J. A. & Jacobsen, S. E. Patterns in Plants and Animals. *Nat. Rev. Genet.* **11**, 204–220 (2010).

202. Maatouk, D. M. *et al.* DNA methylation is a primary mechanism for silencing postmigratory primordial germ cell genes in both germ cell and somatic cell lineages. **3418**, 3411–3418 (2006).

203. Price, A. L., Eskin, E. & Pevzner, P. A. Whole-genome analysis of Alu repeat elements reveals complex evolutionary history. *Genome Res.* **14**, 2245–2252 (2004).

204. Dhayalan, A. *et al.* The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J. Biol. Chem.* **285**, 26114–26120 (2010).

205. Zhang, Y. *et al.* Targets and genomic constraints of ectopic Dnmt3b expression. *Elife* **7**, 1–25 (2018).

206. Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* **47**, D419–D426 (2019).

207. Yoshihama, M. *et al.* The human ribosomal protein genes: Sequencing and comparative analysis of 73 genes. *Genome Res.* **12**, 379–390 (2002).

208. Guimaraes, J. C. & Zavolan, M. Patterns of ribosomal protein expression specify normal and malignant human cells. *Genome Biol.* **17**, 1–13 (2016).

209. Eisenberg, E. & Levanon, E. Y. Human housekeeping genes, revisited. *Trends Genet.* **29**, 569–574 (2013).

210. Gross, A., Jockel, J., Wei, M. C. & Korsmeyer, S. J. Enforced dimerization of BAX results in its translocation, mitochondrial dysfunction and apoptosis. *EMBO*

J. **17**, 3878–3885 (1998).

211. Gross, A., McDonnell, J. M. & Korsmeyer, S. J. BCL-2 family members and the mitochondria in apoptosis. *Genes Dev.* **13**, 1899–1911 (1999).

212. Pastorino, J. G., Chen, S. T., Tafani, M., Snyder, J. W. & Farber, J. L. The overexpression of Bax produces cell death upon induction of the mitochondrial permeability transition. *J. Biol. Chem.* **273**, 7770–7775 (1998).

213. Villunger, A. *et al.* p53- and Drug-Induced Apoptotic Responses Mediated by BH3-Only Proteins Puma and Noxa. *Science (80-.).* **302**, 1036–1038 (2003).

214. Vallier, L., Reynolds, D. & Pedersen, R. A. Nodal inhibits differentiation of human embryonic stem cells along the neuroectodermal default pathway. **275**, 403–421 (2004).

215. Vallier, L., Alexander, M. & Pedersen, R. A. Activin / Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. (2005). doi:10.1242/jcs.02553

216. Rogers, K. W. & Schier, A. F. Morphogen Gradients: From Generation to Interpretation. *Annu. Rev. Cell Dev. Biol.* **27**, 377–407 (2011).

217. De Caestecker, M. The transforming growth factor-β superfamily of receptors. *Cytokine Growth Factor Rev.* **15**, 1–11 (2004).

218. Harrison, C. A., Gray, P. C., Koerber, S. C., Fischer, W. & Vale, W. Identification of a functional binding site for activin on the type I receptor ALK4. *J. Biol. Chem.* **278**, 21129–21135 (2003).

219. Yeo, C.-Y. & Whitman, M. Nodal Signals to Smads through Cripto-Dependent and Cripto-Independent Mechanisms TGF-Beta ligands signal by binding to transmembrane. *Mol. Cell* **7**, 949–957 (2001).

220. Kitisin, K. *et al.* TGF-beta Signaling in Development. *Sci. STKE* **2007**, cm1–cm1 (2007).

221. Heldin, C. H., Miyazono, K. & Ten Dijke, P. TGF-β signalling from cell membrane to nucleus through SMAD proteins. *Nature* **390**, 465–471 (1997).

222. Zhang, Y., Feng, X., We, R. & Derynck, R. Receptor-associated Mad homologues synergize as effectors of the TGF-beta response. *Nature* **383**, 168–72 (1996).

223. Randall, R. A. *et al.* Recognition of Phosphorylated-Smad2-Containing Complexes by a Novel Smad Interaction Motif. *Mol. Cell. Biol.* **24**, 1106–1121 (2004).

224. Cordenonsi, M. *et al.* Links between tumor suppressors: p53 is required for TGF-β gene responses by cooperating with Smads. *Cell* **113**, 301–314 (2003).

225. Camus, A., Perea-Gomez, A., Moreau, A. & Collignon, J. Absence of Nodal signaling promotes precocious neural differentiation in the mouse embryo. *Dev. Biol.* **295**, 743–755 (2006).

226. Bisgrove, B. W., Essner, J. J. & Yost, H. J. Regulation of midline development by antagonism of lefty and nodal signaling. *Development* **126**, 3253–3262 (1999).

227. Meno, C. *et al.* Mouse lefty2 and zebrafish antivin are feedback inhibitors of nodal signaling during vertebrate gastrulation. *Mol. Cell* **4**, 287–298 (1999).

228. Piccolo, S. *et al.* The head inducer cerberus is a multifunctional antagonist of Nodal, BMP and Wnt signals. *Nature* **397**, 707–710 (1999).

229. Brennan, J. *et al.* Nodal signalling in the epiblast patterns the early mouse embryo. *Nature* **411**, 965–969 (2001).

230. Bertero, A. *et al.* Activin / Nodal signaling and NANOG orchestrate human embryonic stem cell fate decisions by controlling the H3K4me3 chromatin mark. 702–717 (2015). doi:10.1101/gad.255984.114.

231. Schuldiner, M., Yanuka, O., Itskovitz-Eldor, J., Melton, D. A. & Benvenisty, N. Effects of eight growth factors on the differentiation of cells derived from human embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 11307–11312 (2000).

232. Thompson, T. B., Lerch, T. F., Cook, R. W., Woodruff, T. K. & Jardetzky, T. S. The structure of the follistatin: Activin complex reveals antagonism of both type I and type II receptor binding. *Dev. Cell* **9**, 535–543 (2005).

233. Inman, G. J. *et al.* SB-431542 is a potent and specific inhibitor of transforming growth factor-β superfamily type I activin receptor-like kinase (ALK) receptors ALK4, ALK5, and ALK7. *Mol. Pharmacol.* **62**, 65–74 (2002).

234. Fiorenzano, A. *et al.* Cripto is essential to capture mouse epiblast stem cell and human embryonic stem cell pluripotency. *Nat. Commun.* **7**, (2016).

235. Lonardo, E. *et al.* A small synthetic cripto blocking peptide improves neural induction, dopaminergic differentiation, and functional integration of mouse embryonic stem cells in a rat model of Parkinson's disease. *Stem Cells* **28**, 1326–1337 (2010).

236. Zufferey, R. *et al.* Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. *J. Virol.* **72**, 9873–80 (1998).

237. Müller-Kuller, U. *et al.* A minimal ubiquitous chromatin opening element (UCOE) effectively prevents silencing of juxtaposed heterologous promoters by epigenetic remodeling in multipotent and pluripotent stem cells. *Nucleic Acids Res.* **43**, 1577–1592 (2015).

238. Ryan, M. D., King, A. M. Q. & Thomas, G. P. Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. *J. Gen. Virol.* **72**, 2727–2732 (1991).

239. Gifford, C. A. *et al.* Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell* **153**, 1149–1163 (2013).

240. Hay, D. C. *et al.* Highly efficient differentiation of hESCs to functional hepatic endoderm requires ActivinA and Wnt3a signaling. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 12301–12306 (2008).

241. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).

242. Tsankov, A. M. *et al.* A qPCR ScoreCard quantifies the differentiation potential of human pluripotent stem cells. *Nat. Biotechnol.* **33**, 1182–1192 (2015).

243. Jönsson, M. E. *et al.* Activation of neuronal genes via LINE-1 elements upon global DNA demethylation in human neural progenitors. *Nat. Commun.* **10**, 1–11 (2019).

244. Boeuf, H., Hauss, C., De Graeve, F., Baran, N. & Kedinger, C. Leukemia inhibitory factor-dependent transcriptional activation in embryonic stem cells. *J.*

Cell Biol. **138**, 1207–1217 (1997).

245. Bertrand, N., Castro, D. S. & Guillemot, F. Proneural genes and the specification of neural cell types. *Nat. Rev. Neurosci.* **3**, 517–530 (2002).

246. TCW, J. *et al.* An Efficient Platform for Astrocyte Differentiation from Human Induced Pluripotent Stem Cells. *Stem Cell Reports* **9**, 600–614 (2017).

247. Hall, M. D., Murray, C. A., Valdez, M. J. & Perantoni, A. O. Mesoderm-specific Stat3 deletion affects expression of Sox9 yielding Sox9-dependent phenotypes. *PLoS Genet.* **13**, 1–30 (2017).

248. Mochizuki, Y. *et al.* Combinatorial CRISPR / Cas9 Approach to Elucidate a Far-Upstream Enhancer Complex for Tissue- Technology Combinatorial CRISPR / Cas9 Approach to Elucidate a Far-Upstream Enhancer Complex for Tissue-Specific Sox9 Expression. *Dev. Cell* **46**, 794-806.e6 (2018).

249. Sekido, R. & Lovell-Badge, R. Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. *Nature* **453**, 930–934 (2008).

250. Wagner, T. *et al.* Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9. *Cell* **79**, 1111–1120 (1994).

251. Benko, S. *et al.* Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. *Nat. Genet.* **41**, 359–364 (2009).

252. Bagheri-Fam, S. *et al.* Long-range upstream and downstream enhancers control distinct subsets of the complex spatiotemporal Sox9 expression pattern. *Dev. Biol.* **291**, 382–397 (2006).

253. WADDINGTON, C. H. *The strategy of the genes. A discussion of some aspects of theoretical biology. With an appendix by H. Kacser.* (London: George Allen & Unwin, Ltd., 1957).

254. Héberlé, É. & Bardet, A. F. Sensitivity of transcription factors to DNA methylation. *Essays Biochem.* **0**, 727–741 (2019).

255. Baubec, T., Ivánek, R., Lienert, F. & Schübeler, D. Methylation-dependent and -independent genomic targeting principles of the mbd protein family. *Cell* **153**, 480–492 (2013).

256. Tropepe, V. *et al.* Direct neural fate specification from embryonic stem cells: A primitive mammalian neural stem cell stage acquired through a default mechanism. *Neuron* **30**, 65–78 (2001).

257. Tsankov, A. M. *et al.* Loss of DNA methyltransferase activity in primed human ES cells triggers increased cell-cell variability and transcriptional repression. *Development* **146**, dev174722 (2019).

258. Massagué, J., Seoane, J. & Wotton, D. Smad transcription factors. *Genes Dev.* **19**, 2783–2810 (2005).

259. Domcke, S. *et al.* transcription factors determines binding of NRF1. *Nature* (2015). doi:10.1038/nature16462

260. Zhu, H., Wang, G. & Qian, J. Transcription factors as readers and effectors of DNA methylation. *Nat. Rev. Genet.* **17**, 551–565 (2016).

261. Muhar, M. *et al.* SLAM-seq defines direct gene-regulatory functions of the BRD4-MYC axis. *Science (80-.).* **360**, 800–805 (2018).

262. Hargan-Calvopina, J. *et al.* Stage-Specific Demethylation in Primordial Germ Cells Safeguards against Precocious Differentiation. *Dev. Cell* **39**, 75–86 (2016).

263. Biniszkiewicz, D. *et al.* Dnmt1 Overexpression Causes Genomic Hypermethylation, Loss of Imprinting, and Embryonic Lethality. *Mol. Cell. Biol.* **22**, 2124–2135 (2002).

264. Geula, S. *et al.* m 6 A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science (80-.).* **347**, 1002–1006 (2015).

265. Walter, M., Teissandier, A., Pérez-Palacios, R. & Bourc'his, D. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *Elife* **5**, 1–6 (2016).

266. McLaughlin, K. A. *et al.* DNA methylation directs polycomb-dependent 3D genome re- organisation in naive pluripotency. *bioRxiv* **29**, 527309 (2019).

267. Tokoyoda, K., Egawa, T., Sugiyama, T., Choi, B. Il & Nagasawa, T. Cellular niches controlling B lymphocyte behavior within bone marrow during development. *Immunity* **20**, 707–718 (2004).

268. Von Freeden-Jeffry, U. *et al.* Lymphopenia in interleukin (IL)-7 gene-deleted mice identifies IL-7 as a nonredundant cytokine. *J. Exp. Med.* **181**, 1519–1526 (1995).

269. Mikkelsen, T. S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**, 49–55 (2008).

270. Cheloufi, S. *et al.* The histone chaperone CAF-1 safeguards somatic cell identity. *Nature* **528**, 218–224 (2015).

271. Kelly, T. K. *et al.* Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res.* **22**, 2497–2506 (2012).

272. Spektor, R., Tippens, N. D., Mimoso, C. A. & Soloway, P. D. Methyl-ATAC-seq measures DNA methylation at accessible chromatin. *Genome Res.* **29**, 969–977 (2019).

273. Ohm, J. E. *et al.* A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nat. Genet.* **39**, 237–242 (2007).

274. Widschwendter, M. *et al.* Epigenetic stem cell signature in cancer. *Nat. Genet.* **39**, 157–158 (2007).

275. Walter, M. J. *et al.* Recurrent DNMT3A mutations in patients with myelodysplastic syndromes. *Leukemia* **25**, 1153–1158 (2011).

276. Linhart, H. G. *et al.* Dnmt3b promotes tumorigenesis in vivo by gene-specific de novo methylation and transcriptional silencing. *Genes Dev.* **21**, 3110–3122 (2007).

277. Saito, Y. *et al.* Increased protein expression of DNA methyltransferase (DNMT) 1 is significantly correlated with the malignant potential and poor prognosis of human hepatocellular carcinomas. *Int. J. Cancer* **105**, 527–532 (2003).

278. Gaiti, F. *et al.* Epigenetic evolution and lineage histories of chronic lymphocytic leukaemia. *Nature* **569**, 576–580 (2019).

279. Shlush, L. I. *et al.* Identification of pre-leukaemic haematopoietic stem cells in

acute leukaemia. *Nature* **506**, 328–333 (2014).

280. Herman, J. G. Hypermethylation of tumor suppressor genes in cancer. *Semin. Cancer Biol.* **9**, 359–367 (1999).

281. Burri, N. *et al.* Methylation silencing and mutations of the p14ARF and p16INK4a genes in colon cancer. *Lab. Investig.* **81**, 217–229 (2001).

282. Mann, J. R. & Lovell-Badge, R. H. Two maternally derived X chromosomes contribute to parthenogenetic inviability. *Development* **104**, 129–136 (1988).

283. Howell, C. Y. *et al.* Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. *Cell* **104**, 829–838 (2001).

284. Li, X. *et al.* A Maternal-Zygotic Effect Gene, Zfp57, Maintains Both Maternal and Paternal Imprints. *Dev. Cell* **15**, 547–557 (2008).

285. Xiong, J. W. Molecular and developmental biology of the hemangioblast. *Dev. Dyn.* **237**, 1218–1231 (2008).

286. Shalaby, F. *et al.* Failure of blood-island formation and vasculogenesis in Flk-1-deficient mice. *Nature* **376**, 62–6 (1995).

APPENDIX

Table 8. Top 100 up- and downregulated differentially expressed genes

Upregulated genes								Downregulated genes						
d1	d2	d3	d5	d6	d7	d8		d1	d2	d3	d5	d6	d7	d8
B3GNT7	B3GNT7	NPPB	MYL7	NPPB	NPPB	C8orf4		RPS16	GAPDH	GLIPR1L1	MT-CO3	HNRNPA1	MT-ND3	MT-ND2
SYTL2	ARRDC3	MYL7	APOA2	DKK1	LCN15	NPPB		GABRP	NACA	IFITM2	CLDN11	AMFR	RPL19	COX7C
ALG1	SLC35F6	CST1	DUSP23	GAGE12F	DKK1	LCN15		MT-ND4	CALB1	AMFR	ALDOA	RPS12	NPM1	GAPDH
HIST1H4C	AL109811.4	FABP3	RHBDD3	MYL7	APOA1	CDH6		RPL13A	RPLP2	MTRNR2L10	CFL1	MT-ND1	NME2	NME2
MVD	ARHGEF12	HAS2	CBR1	SYTL2	XAGE1B	APOA1		HSPD1	RPL4	AXL	LDHB	KPNA2	RPS16	RPS5
TMEM222	SYTL2	FN1	FAM110D	PLSCR2	XAGE1A	PRSS2		SLC25A3	TUBA1B	ZNF688	HSPD1	RPL10	MT-ND4	UBA52
AKAP9	DCLRE1C	SPP1	KEL	FAM110D	RNASE1	CD48		RPLP2	C14orf166	CLDN11	RPS4X	CFL1	RPS2	MT-CYB
SLCO1C1	HIST1H4C	FST	LEFTY1	DUSP23	APOA2	XAGE1B		C14orf166	PPIA	KIN	RACK1	ALDOA	MT-ND5	COX5A
CCND1	AKAP9	PON2	SYTL2	RHBDD3	MAEL	XAGE1A		RPL10	RPL7A	MT1G	PPIA	RPS3	GABRP	GSTP1
DHDDS	FABP3	TMEM97	CEBPZOS	CBR1	GAGE2A	APOA2		UBA52	UBE2C	TNFSF13B	MT-ND2	CD24	ATP5A1	CCT5
NSMAF	ARID4B	HMGCS1	STAG3	DDIT3	FAM110D	LGI1		RPL4	CTSC	PRRT2	RPS28	PPIA	KRT18	RPL35
APC	BRD8	NODAL	VCY	MAP3K13	GAGE12F	CST1		RPL14	MT-CYB		RPL28	HSPD1	MT-ND4L	RPS12
RAF1	FST	ACAT2	VCY1B	FAM184A	MYL7	GAGE2A		RPL29	ATP1B3		MT-ND5	RACK1	RPS12	TPM3
C19orf48	MYDGF		CER1	EPHA4	CST1	MAEL		NACA	EDF1		C14orf166	AXL	RPS5	IFITM1
STAG3	VSIG10		HAS2	CEBPZOS	SYTL2	GAGE12F		RPL18	RPS12		KPNA2	HHLA1	UBA52	EIF3M

SPP1	ANKRD12		LSP1	LEFTY1	RPL10L	MYL7		RPS12	RPS2		EPCAM	RPL18	MT-CYB	RPL11
DCLRE1C	NR6A1		AGAP4	EOMES	PLSCR2	RPL10L		MT-CO3	SLC2A3		MT-CYB	LDHB	UBB	RPS2
RNF121	AGAP4		CCND1	SMARCA2	DSCR8	CXCR4		MT-ND6	RPL10		HIST1H2BK	RPS4X	ENO1	RPS15A
PEF1	C19orf48		GTSF1	INPP5F	FAM184A	CROT		RPS2	MT-CO3		DYNLRB1	HIST1H2BK	HDAC2	ATP5I
CCNL2	TNFRSF12A		TAGLN	HAS2	DDIT3	GDF15		RPS3	MT-ND4		MT-ND1	OAT	PGAM1	RPL19
FABP3	CD58		TNFRSF12A	TNFRSF12A	CBR1	AC270107.8		LDHB	PRDX1		MT-ND4L	C14orf166	HSPA8	UBE2C
WDR6	CCND1		RASEF	AGAP4	RHBDD3	SHISA2		RACK1	EIF1		S100A10	DYNLRB1	MDH2	NDUFA13
FST	MUM1		BAX	C12orf4	HAS2	FAM184A		MT-CYB	CFL1		FH	RPS28	RPL4	HSPE1
SLC5A6	BCL11A		AKAP9	DPPA5	KEL	BRDT		MT-ND5	DDX39B		VDAC1	EDF1	PSME2	ATP5A1
ZNF75A	ZNF625		CDK10	FST	CER1	TTN		RHOA	ATF7IP2		KIN	SYPL1	PRDX6	CLDN11
TNFRSF12A	SPP1		DPPA5	CADM1	BRDT	PRTG		HHLA1	LDHB		CSNK2B	S100A10	GPX1	NPM1
WDR74	RNF34		C19orf48	MAGEA6	LEFTY1	PLSCR2		RPL7A	IFITM3		AXL	ATP6V1F	RPL29	RPLP2
KANSL2	PAN3		DDX55	TAGLN	TNFRSF12A	AMPD2		MT-ND1	MPV17		GDF3	MT-ATP8	PRDX1	SLC25A3
THOC6	TIMP4		TSPAN4	DCLRE1C	MAP3K13	PLK2		RPS4X	HSPD1		CDA	EPCAM	EIF2S1	NDUFS8
NSD1	GOLIM4		TRDN	STAG3	EPHA4	RAB17		TUBA1C	CLDN11		LDHA	VIM	HNRNPDL	PGAM1
SLC3A2	SRSF6		CCNL2	CER1	P3H4	FAM110D		IFITM3	NADSYN1		PGK1	PIGU	RPL7A	RPS25
ING5	C6orf62		LPP	C19orf48	BBS9	VCX		RPS28	MDH2		SH3BGRL3	LDHA	NACA	ATP5G2
MCM5	STAG3		CYP3A5	CDK10	SMARCA2	BBS9		FXYD5	C17orf49		NABP1	MAPK1IP1L	HHLA1	CUTA

VSIG10	CENPX		SLFN13	CD58	NSMAF	VCX3A		MT-ND2	HNRNPA1		FXYD5	EDNRB	TUBA1C	RPL35A
IFRD2	EBP		LUZP1	SLC3A2	CEBPZOS	CBR1		IFITM2	RPL28		MT-CO1	DCXR	EIF3M	EIF3C
RABGAP1L	CTTN		TMEM106C	LPP	LMO7	SYTL2		MT-ND3	PRDX6		IFITM2	RPL28	PDHB	RPL18
SURF2	ATRX		CADM1	BAX	CCND1	DSCR8		MDH2	SLC25A3		MT-ATP8	GDF3	FKBP1A	RPL29
MYDGF	SRPRB		ACADVL	CCND1	RCOR3	SAXO2		RPL28	RACK1		ZNF688	CSNK2B	ILF2	RPL24
HM13	MCRS1		FSTL4	RASEF	STAG3	RHBDD3		CCT8	RPL18		PRRT2	KDELR1	HNRNPA1	COX8A
ZDHHC12	ZNF654		SLC25A6	NFKBIA	LEFTY2	NFE4		S100A11	VIM		AMFR	SH3BGRL3	UBE2C	HSP90AB1
TOX4	SYF2		DCLRE1C	AKAP9	DCLRE1C	CER1		LDHA	RPS3		TNFSF13B	MT-CO1	ATP5G2	BTF3
MCRS1	HNRNPH1		ANKRD12	SPINK9	TNNC1	FN1		MT-ND4L	RHOA			VDAC1	RPL28	GSTO1
DDX55	ANKRD10		SLCO1A2	HUWE1	RUNX1T1	ZNF677		CSNK2B	MT-ND5			PGK1	PSMB2	PGRMC1
ZWINT	KANSL2		SPINK9	DCAF8	FN1	TMEM63A		KIN	KDELR1			ZNF688	RHOA	GPX1
BAX	METTL1		CHD9	MYDGF	TXNRD2	CCDC144A		CSTB	SYPL1			KIN	MT-ND6	RAN
GPATCH8	MAP1B		INTS11	MCM5	GTSF1	HAS2		HHLA3	S100A13			CAPNS1	ATP5G3	WDR83OS
IGF2BP1	POLR3E		MCM5	AC015813.2	DPPA5	NFRKB		CRABP1	RPS28			CSTB	TPM3	SNRPD3
AUP1	CCNL2		PDE4DIP	ANKRD12	LRIG3	SPATA7		MT-ATP8	BEX4			FXYD5	MT-ND2	RPS3
BRD8	SURF2		ARID4B	GLS	TAGLN	LEFTY1		VDAC1	TIMP1			CDA	TPI1	PSENEN
WDR73	MRPL1		C19orf60	HIST1H4C	APC	KEL		DUSP6	AMFR			PRRT2	RPS28	TUBA1B
NELFCD	HMGCS1		INPP5F	NODAL	FST	NDRG4		ZNF688	SUCLA2			TNFSF13B	BRK1	TUBA1A

DCP1A	DGCR6L		SLC25A37	TRDN	INPP5F	SHC4		SH3BGRL3	FXYD5				BANF1	LSM4
DGCR6L	NELFCD		MRPL28	MTRR	APOBEC2	ITGA5		CTSK	OAT				CCT8	PSMD6
MAP1B	BAX		FANCA	MCRS1	COMMD3	LEFTY2		PGK1	CSNK2B				RPL18	RPL38
TCF7L2	KPNA6		SUOX	BCL11A	AGAP4	LRIG3		DYNLRB1	RPS4X				RPL10	TXNDC17
PDCD11	RNF121		LRRTM4	RNF34	GAD1	ACAA1		MT-CO1	IFITM2				RPS3	RPL32
SERINC5	PIN1		BCL11A	DCP1A	EOMES	SLC38A4		GDF3	PPIH				GANAB	ATP5H
DHCR24	SLC25A37		ZNF124	FSTL4	TSPAN4	CSNK1G1		NUDT15	LRRK1				NANS	RPL36
QTRT1	IQCF3		IGF2BP1	YTHDF3	CADM1	RCOR3		HIST1H2BK	CRABP1				VIM	RPL22L1
TAGLN	CTNND1		HMGXB4	WDR74	C19orf48	RHOBTB3		CLDN11	MT-ND4L				CFL1	TRAPPC2L
FADS1	TMEM106C		SERPINB9	TMEM106C	PYCR1	CEBPZOS		MT1G	CSTB				TIMP1	RPL37
BRCC3	KANSL1L		NSD1	IFRD2	B3GNT7	STAG3		CDA	SNRPA				TUBA1A	UBB
MTRR	CDC42BPA		PHACTR2	GTF3C3	RASEF	PHYKPL		PRRT2	SLC7A3				SLC25A3	HNRNPA1
TMEM208	QTRT1		MCRIP2	SERPINB9	AKAP9	LSP1		AMFR	TTC9C				GAPDH	RPL7A
PHAX	SON		MARK3	GPATCH8	MCRS1	POLE		TNFSF13B	DCXR				C14orf119	PHB2
BCL11A	TGFBR1		PWWP2A	MRPL28	ANKRD12	CYP26A1			DUSP6				C9orf135	CD24
GPRC5B	KNOP1		DCAF8	DCTN5	LPP	SYCP3			HIST1H2BK				EDF1	TPI1
FGFR1	PUS7L		NOL8	SRSF6	SLC3A2	TNNC1			HHLA1				AMFR	EPCAM
SRPRB	BAZ2B		FUOM	KANSL2	SRSF6	CADM1			EDNRB				C14orf166	RHOA

ATP2B1	PSMC3		FKBP10	GAD1	CDK10	TNPO2			S100A10				JPT1	PRDX1
POLR3E	TLK1		KIFC3	CBY1	TCF7L2	LAMA1			VDAC1				PPIA	HSPA8
MTG1	ANKIB1		PHAX	BRD8	RAF1	CRIM1			GABRP				PCLAF	NQO2
DCTN5	TMEM208		SLC3A2	ESRP1	LPAR6	ANKRD1			MT-ND2				KPNA2	NDUFB11
LPP	NUCB2		GLS	NSD1	ZNF84	CDK10			S100A11				TUBA1B	PCLAF
FADS2	FRG1		PDXDC1	ZNF518A	SOCS2	EPHA4			PTBP1				RACK1	C17orf49
CTNND1	TMED1		DGCR6L	PHAX	SCAF11	LPP			GANAB				MT-ATP8	DECR1
RSF1	SMC2		NOSIP	KNOP1	BRD8	DCAF8			LDHA				RPS4X	UBE2I
ANKRD12	FADS1		FGF8	PDE4DIP	RNF34	INPP5F			ZNF688				PTBP1	RPL10
MAP4K5	CDK10		MYDGF	SERINC5	SPP1	AGAP4			MT-ATP8				AC087632.1	GAL
SON	NKTR		BET1L	SON	CCNT1	EXOG			MT-ND3				HSPD1	RACK1
HMGXB4	NUBP2		SPG7	C19orf60	DCAF8	PDXDC1			SH3BGRL3				MT-ND1	H3F3A
CHD9	GPATCH8		HYOU1	SLCO1A2	MYDGF	TAGLN			MT-ND1				LDHB	SSBP1
PSMC3	PHAX		CTNND1	SPG7	CYP26A1	RAF1			HHLA3				CDK2AP1	PSMB2
KNOP1	OTX2		GRB2	BCAS3	SLFN13	LY6E			TUBA1C				DPCD	HIST1H4J
GPATCH4	NOL8		NUBP2	TTC3	PTGR1	HOOK2			MT-ND6				MAPK1IP1L	RPL14
ZMYND8	RABGAP1L		MED7	DPY19L4	PAN3	TXNRD2			CTSK				DECR1	POLR3G
NR6A1	PPP2R2B		C11orf49	SNUPN	ZNF518A	SUOX			DYNLRB1				ZNF688	PIGU

MAGEA2	FABP7		FRG1	FRG1	CASP8AP2	LIFR			KIN				ALDOA	SNU13
RNF34	MTG1		MYO19	SLC11A2	UTRN	CCND1			PGK1				DYNLRB1	CFL1
SCAF11	PWWP2A		MANBAL	PWWP2A	ERBIN	ANKRD36C			NUDT15				PIGU	HSD17B10
SGPL1	CRBN		FST	MED7	BAX	PTGR1			GDF3				VSNL1	SUMO2
ARID4B	MCM7		NDUFV1	BET1L	PHACTR2	UBE2D4			MT-CO1				SLC25A5	MRPS23
TMED1	SERINC5		FN1	EXOSC5	RNPC3	SZT2			CAPNS1				OAT	NACA
NDUFV1	THUMPD1		BCAS3	AC007326.4	SPINK9	AKAP9			CDA				LDHA	ALDOA
PDE5A	DDX55		ZNF808	CTNND1	B2M	NR2C2			MT1G				VDAC1	ATP5G3
MANBAL	CCDC66		SCAF11	B9D1	C6orf62	F3			TNFSF13B				CSTB	MT-ND1
DKC1	SGPL1		AC007326.4	LAMTOR3	MCM5	CFLAR			PRRT2				SH3BGRL3	NME1
PLEKHA5	HMGXB4		HUWE1	PHACTR2	FSTL4	NFKBIA							KIN	CNMD
LMAN2	TIMM10B		ZNF518A	ARL4A	ESRP1	BAX							GDF3	KPNA2
TAX1BP1	RBBP8		MAPK10	C21orf91	DDX55	CYP3A5							FXYD5	CAPNS1

TABLE 9. Top 10 GO terms for differentially expressed genes

	Upregulated genes			Downregulated genes	
	Top GO terms	FDR		Top GO terms	FDR
d1	primary metabolic process	6.43E-09	d1	SRP-dependent cotranslational protein targeting to membrane	9.82E-19
	cellular metabolic process	5.36E-09		cotranslational protein targeting to membrane	1.02E-18
	metabolic process	1.64E-08		protein targeting to ER	2.05E-18
	organic cyclic compound metabolic process	1.46E-08		viral transcription	2.36E-18
	organic substance metabolic process	1.26E-08		establishment of protein localization to endoplasmic reticulum	2.59E-18
	cellular nitrogen compound metabolic process	1.85E-08		nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	3.67E-18
	gene expression	3.68E-08		viral gene expression	1.28E-17
	nitrogen compound metabolic process	1.54E-07		protein localization to endoplasmic reticulum	1.94E-17
	heterocycle metabolic process	7.77E-07		translational initiation	3.61E-17
	cellular aromatic compound metabolic process	1.50E-07		protein targeting to membrane	6.40E-16
d2	gene expression	1.06E-09	d2	ATP metabolic process	6.68E-09
	cellular nitrogen compound metabolic process	6.36E-09		viral gene expression	7.84E-09
	cellular metabolic process	1.05E-08		cotranslational protein targeting to membrane	8.69E-09
	organic cyclic compound metabolic process	2.27E-08		symbiotic process	8.79E-09
	metabolic process	2.72E-08		SRP-dependent cotranslational protein targeting to membrane	9.19E-09
	organic substance metabolic process	1.14E-07		protein localization to endoplasmic reticulum	9.82E-09
	cellular aromatic compound metabolic process	1.50E-07		translational initiation	1.27E-08
	nucleic acid metabolic process	3.40E-07		interspecies interaction between organisms	1.32E-08
	heterocycle metabolic process	4.78E-07		protein targeting to ER	1.44E-08
	primary metabolic process	6.49E-07		viral transcription	1.69E-08
d5	no statistically significant results		d5	ATP metabolic process	2.93E-09
				generation of precursor metabolites and energy	5.59E-09
d6	regulation of execution phase of apoptosis	1.77E-02		oxidative phosphorylation	4.04E-07
	negative regulation of execution phase of apoptosis	1.49E-02		ATP synthesis coupled electron transport	2.00E-06

	negative regulation of molecular function	1.26E-02		cellular respiration	2.30E-06
				mitochondrial ATP synthesis coupled electron transport	2.33E-06
d7	nitrogen compound metabolic process	8.14E-08		electron transport chain	4.75E-06
	cellular metabolic process	9.07E-08		respiratory electron transport chain	5.00E-06
	macromolecule metabolic process	1.02E-07		energy derivation by oxidation of organic compounds	2.37E-05
	primary metabolic process	1.22E-07		oxidation-reduction process	2.57E-05
	organic substance metabolic process	1.41E-07			
	metabolic process	7.49E-07	d6	interspecies interaction between organisms	9.13E-09
	cellular macromolecule metabolic process	1.69E-05		symbiotic process	1.47E-08
	nucleic acid metabolic process	3.58E-04		viral process	2.53E-08
	organic cyclic compound metabolic process	1.03E-03		protein localization to endoplasmic reticulum	6.99E-06
	cellular component organization	1.20E-03		SRP-dependent cotranslational protein targeting to membrane	1.56E-05
				cotranslational protein targeting to membrane	1.81E-05
d8	cellular component organization	2.06E-11		protein targeting to ER	2.55E-05
	cellular component organization or biogenesis	2.89E-10		viral transcription	2.65E-05
	cellular metabolic process	2.30E-09		establishment of protein localization to endoplasmic reticulum	2.82E-05
	primary metabolic process	4.35E-09		nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	3.16E-05
	organelle organization	1.25E-08			
	cellular macromolecule metabolic process	1.26E-08	d7	symbiotic process	6.70E-20
	macromolecule metabolic process	1.32E-08		interspecies interaction between organisms	5.41E-19
	nitrogen compound metabolic process	1.34E-08		viral process	1.77E-18
	organic substance metabolic process	3.82E-08		translational initiation	9.31E-15
	metabolic process	2.19E-07		SRP-dependent cotranslational protein targeting to membrane	2.15E-14
				cotranslational protein targeting to membrane	3.53E-14
				viral gene expression	3.86E-14
				ATP metabolic process	5.58E-14
				protein targeting to ER	6.56E-14
				establishment of protein localization to endoplasmic reticulum	9.59E-14
			d8	translational initiation	2.77E-26

SRP-dependent cotranslational protein targeting to membrane	4.83E-26
cotranslational protein targeting to membrane	9.55E-26
protein targeting to ER	3.70E-25
establishment of protein localization to endoplasmic reticulum	6.46E-25
viral transcription	6.52E-25
symbiotic process	1.40E-24
nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	1.42E-24
viral gene expression	9.00E-24
protein localization to endoplasmic reticulum	1.84E-23

Table 10: Top 100 upregulated genes overlapping between the listed time points

Unique				Shared				
d5	d6	d7	d8	d5d6	d5d6d7d8	d5d6d7	d6d7d8	d7d8
VCY	C12orf4	RNASE1	C8orf4	DUSP23	MYL7	DPPA5	NPPB	LCN15
VCY1B	MAGEA6	P3H4	CDH6	DPPA5	RHBDD3	BCL11A	GAGE12F	APOA1
INTS11	CD58	NSMAF	PRSS2	TMEM106C	CBR1	SERPINB9	PLSCR2	XAGE1B
MCRIP2	YTHDF3	LMO7	CD48	BCL11A	FAM110D	PHAX	FAM184A	XAGE1A
MARK3	IFRD2	RUNX1T1	LGI1	SERPINB9	LEFTY1	NOSIP	EPHA4	MAEL
FUOM	DCTN5	APC	CXCR4	PHAX	SYTL2	LARP7	EOMES	GAGE2A
DGCR6L	KNOP1	APOBEC2	CROT	NOSIP	CEBPZOS	GNL3	MTRR	CST1
HYOU1	DPY19L4	COMMD3	GDF15	HUWE1	STAG3	PSMC3	MCRS1	RPL10L
PRR11	SNUPN	PYCR1	AC270107.8	ETFB	CER1	LARS	RNF34	DSCR8
CENPX	LAMTOR3	B3GNT7	SHISA2	LARP7	HAS2	SSB	DCP1A	BRDT
	C21orf91	TCF7L2	TTN	GNL3	AGAP4		GPATCH8	BBS9
	MPHOSPH6	LPAR6	PRTG	PSMC3	CCND1		KANSL2	RCOR3
	TRA2A	ZNF84	AMPD2	LARS	TAGLN		SERINC5	LEFTY2
		SOCS2	PLK2	SSB	TNFRSF12A		SON	TNNC1
		CCNT1	RAB17		RASEF		TTC3	TXNRD2
		CASP8AP2	VCX		BAX		SLC11A2	LRIG3
		UTRN	VCX3A		AKAP9		ANKRD18B	RAF1
		ERBIN	SAXO2		CDK10		TAX1BP1	SPP1
		RNPC3	NFE4		C19orf48		ANXA3	CYP26A1
		RNF170	ZNF677		TRDN		PCSK5	PTGR1
		MMS22L	TMEM63A		CCNL2		MCM7	PAN3
		MGST2	CCDC144A		LPP		MTRNR2L8	C6orf62
		GNS	NFRKB		CADM1		FSTL1	FNDC3B
		PDCD11	SPATA7		FSTL4		ATP2B1	CFLAR
		ATP1B1	NDRG4		DCLRE1C		MCFD2	TOX4
		PUS7L	SHC4		ANKRD12		MTRNR2L12	NUCB2
		TIMM23B	ITGA5		SLCO1A2		PWP1	USP24
		RABGAP1L	ACAA1		SPINK9			UPP1
		THAP5	SLC38A4		CHD9			WDR73
		ERGIC2	CSNK1G1		MCM5			PLCB1
		TOP3A	RHOBTB3		PDE4DIP			CHD4
		PPIC	PHYKPL		C19orf60			ENOSF1
		CAV1	POLE		INPP5F			KANSL1L
		RBBP8	SYCP3		MRPL28			MFSD8
		CENPE	TNPO2		LRRTM4			CCDC84
		AC138811.2	LAMA1		NSD1			NKTR
		FKBP14	CRIM1		PHACTR2			OTX2
		SECISBP2L	ANKRD1		PWWP2A			FBXO38
		RIC8B	EXOG		DCAF8			CCDC66
		GAR1	LY6E		SLC3A2			ZNF649
		ADM	HOOK2		GLS			MAP4K5
		FAM111B	LIFR		MYDGF			POLR3E
		PBRM1	ANKRD36C		BET1L			XPO6

		IFRD1	UBE2D4		SPG7			SAR1B
		RND3	SZT2		CTNND1			EIF2AK4
		PIGP	NR2C2		MED7			ZNF420
		TMEM138	F3		FRG1			TMED1
		SRPRB	RNF213		FST			REV3L
		CCNDBP1	PTK7		BCAS3			C2orf68
		ZMYND8	ZKSCAN1		AC007326.4			ALG11
		CLDN7	TOP3B		ZNF518A			SPINT2
		TIPIN	TAC4		GAD1			AGO3
		SALL4	ANKRD36		GABRG2			HIST2H2AA4
		ALG5	ST5		DTNA			ODF2L
		PPP4R3A	PPFIBP1		ANKRD10			ZFC3H1
		HYAL2	NELFCD		GTF3C3			SPATS2L
		PSMG3	ZNF160		BRD8			NSFL1C
		REST	FBXO16		LMAN2			ADAMTSL3
		TMEM208	KDM5A		PIGN			TMEM243
		MLEC	TIGD1		KIAA1328			NPC1
		NSRP1	ZNF571		CALD1			GPBP1
		DNMT3A	WDR72		BLOC1S6			APOC1
		KIAA1551	ZNF611		NTS			ZC2HC1A
		RARS2	SELENOP		B2M			FANCD2
		CHID1	MOSPD2					LCLAT1
		OSBPL8	PRMT7					SLC20A1
		COMMD2	ANKFY1					PON2
		TTC14	GABBR1					CRBN
		CCDC59	NAT9					TNRC6B
		CHD1	ZNF331					LRP6
		MGEA5	PEAK1					MDK
		CCL2	MFAP5					MRPS25
		QPRT	PRKAA1					MAP1B
		ICE2	RAD51B					PLEKHA5
		TLK1	DDX11					ABCD4
		SELENOT	CENPC					NEPRO
		SELENOS	ZNF614					CLN5
		BST2	RANBP17					NR6A1
		TMED3	MGA					ATM
		CD320	ANK2					FAM133B
		MAGOHB	PSMD5					UBE2B
		ARL6IP1	GGA1					KTN1
		ITM2C	SIN3A					NEMF
		DERL2	CD99					TPR
		CNIH4	DRAM1					NCOR1
		CAPZA1	ASH1L					TMEM87A
		ATP6V0B	PI4KA					RIMKLB
		PRDX4	CYR61					PLPP5
		U2SURP	PAAF1					YIF1B
		CXADR	AC256236.1					POLR2E

			BMP4					ZCCHC11
			FSIP2					SERINC3
			FZD5					USP34
			ATG13					POGZ
			BACH1					TMEM41B
			ENTPD4					LGALS1
			KIAA0100					ZNF195
			PABPC1L					CELF1
			MIPOL1					TBL1XR1
			PDE7A					CBWD5

Table 11. Abbreviations used in this thesis.

Abbreviation	Meaning
5-azaC	5-azacytydine
A	adenine
ac	acetylation
ADD	ATRX-DNMT3-DNMT3L
ALKB2	Alpha-Ketoglutarate Dependent Dioxygenase
AML	Acute Myeloid Leukemia
Amp	ampicillin
APC/C	Anaphase Promoting Complex / Cyclosome
ATP	adenosine 5'triphosphate
BAH	bromo adjacent homology
BER	base excision repair
BMP	bone morphogenesis factor
bp	base pairs
C	cytosine
C.elegans	Caernohabditis elegans
CGI	CpG island
CLL	Chronic Lymphatic Leukemia
CpG	cytosine phosphate guanine
CXXC	cytosine-x-x-cytosine
d	day
DEG	differentially expressed gene
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DNAase	deoxyribonuclease
DNMT	DNA methyltransferase
DNMT3L	DNA methyltransferase 3-like
DSB	double strand break
DTT	dithiothreitol
E1	ubiquitin activation enzyme
E2	ubiquitin conjugation enzyme
E3	ubiquitin ligase
E4	ubiquitin chain elongation enzyme
EDTA	ethylenediaminetetraacidic acid
ESC	embryonic stem cell
FACS	Fluorescence Activated Cell Sorting
G	guanine
GFAP	Glial Fibrillary Acidic Protein
h	hour
H3K27	Histone 3 Lysine 27
H3K36	Histone 3 Lysine 36
H3K4	Histone 3 Lysine 4
H3K9	Histone 3 Lysine 9
HR	homologous recombination
HRP	Horse Radish Peroxidase
IAP	intracisternal A-particle

Ig	immunoglobulin
iPSC	induced pluripotent stem cell
JAK-STAT	Janus Kinase – Signal Transducers and Activators of Transcription
KAP1	KRAB-associated protein 1
kb	kilo base pairs
kDa	kilo Daltons
KRAB-ZFP	Krüppel-associated box domain zinc finger proteins
LB	Luria Bertani
LIF	leukemia inhibotory factor
LINE	long interspersed element
LTR	long terminal repeat
MBD	methyl binding domain
mC	methylcytosine
me2	dimethylation
me3	trimethylation
MEF	mouse embryonic fibroblast
min	minute
MOPS	3-N-Morpholinopropane sulfonic acid
NEUROG 2	Neurogenin2
NPC	Neural Progenitor Cells
ORF	Open Reading Frame
PAGE	Polyacrylamide Gel Electrophoresis
PBS	phosphate buffered saline
PCNA	Proliferating Cell Nuclear Antigen
PCNA	Proliferating Cell Nuclear Antigen
PCR	polymerase chain reaction
PHD	Plant Homeodomain
PIWI	P-element Induced WImpy testis
PMSF	phenylmethylsulfonyl fluoride
POL	polymerase
PROTAC	PROteolysis TArgeting Chimera
PTM	posttranslational modification
PWWP	Proline-Tryptophan-Tryptophan-Proline
RAD	RADiation sensitive
RFTS	replication foci targeting sequence
RING	Really Interesting New Gene
RNA	ribonucleic acid
RNase	ribonuclease
rpm	rounds per minute
RRBS	Reduced Representation Bisulfite Sequencing
RT	room temperature
sc	single cell
SDS	sodium dodecylsulfate
seq	sequencing
SINE	interspersed element
SMAD	Small Mothers Against Decapentaplegic

SOX	SRY-related HMG-box genes
SRA	SET- and RING- associated
TBS	tris buffered saline
TDG	thymine DNA glycosylase
TE	transposable element
TET	ten-eleven-transferase
TF	transcription factor
Tris	Tris (hydroxymethyl) aminomethane
Ub	ubiquitin
UHRF1	Ubiquitin-like with PHD and Ring Finger Domains 1
UV	ultraviolet light
v/v	volume per volume
VEGF	Vascular endothelial growth factor
w/v	weight per volume
WNT	Wingless Int-1
WT	wild type
XCI	X chromosome inactivation