Norwegian University
of Life Sciences

Master's Thesis 2020 120 ECTS
Faculty of Chemistry, Biotechnology and Food Science

Contribution of epigenetic and epitranscriptomic factors in gene regulation of embryonic stem cells

I. Generation of *Ythdf1-3* knock-outs and characterization of their phenotypes

II. Investigating the role of a chromatin modifier, MMSET in R-loop biology

Gina Henriette Østby & Katrine Råkjær Aarseth
Biotechnology (M.Sc.) – Molecular biology

Acknowledgement

The work presented in this thesis was performed for the Department of Microbiology at Oslo University Hospital, as a part of a Master program in Biotechnology, at the Faculty of Chemistry, Biotechnology and Food Science (KBM) at the Norwegian University of Life Sciences (NMBU).

First and foremost, we would like to thank our supervisor Deo Prakash Pandey, for an outstanding job supervising us throughout this thesis. We are very grateful for being included in the research group, and that he gave us the opportunity to perform a handful of exciting experiments in the laboratory. His support, encouragement, and feedback has been invaluable to this thesis and has greatly contributed to our understanding of the field of molecular biology. We would also like to express a note of gratitude to the whole department for their warm welcome, and for including us in both social and work-related gatherings.

We are truly grateful for the guidance and assistance from Anna Lång, regarding immunofluorescence microscopy and analyses, and for valuable help in the laboratory by Gaute Nesse. Furthermore, we would like to thank the rest of Deo Prakash Pandey´s research group, including Preeti Cute, Silje Lier, Sigrid Berg, and Kasturi Raorane for valuable advice and support, and for contributing to an exceptionally pleasant time in the laboratory. Additionally, our supervisor at NMBU, Professor Harald Carlsen, deserves a great deal of gratitude for valuable feedback and guidance.

Lastly, we are grateful for the everlasting support and encouragement of our families and friends.

Abstract

Both epigenetic and epitranscriptomic modifications have been revealed to impact gene regulation and are linked to numerous diseases and oncogenic reprogramming. Reversible alterations in gene expression with no effects on the nucleotide sequence are defined as epigenetic regulation, while modifications occurring on RNA, the transcriptome, is known as epitranscriptomic regulation. The study aimed to investigate the contribution of epigenetic and epitranscriptomic factors in gene regulation of mouse embryonic stem (mES) cells by looking at two different aspects: the impact of the reader proteins YTHDF1-3 (DF1-3) on cellular processes in mES cells, and the role of a chromatin modifier, MMSET in R-loop biology.

The first aspect in this study examined the cytoplasmic reader proteins DF1-3, which recognize and bind N^6-methyladenosine (m^6A) modifications. The DF1-3 proteins are reported to promote translation and affect mRNA stability. Some studies suggest that DF1-3 possesses similar functions, due to their high sequence similarity, however, the majority of studies supports dissimilar functions. Hence, it remains to be elucidated how the DF proteins mediate different functions, and if they bind to the same or different m^6A sites.

We wanted to generate single- and double *Df1-3* knock-outs (KOs) in mES cells to determine how DF1-3 affect proliferation and differentiation phenotypes. By creERT2 recombinase and CRISPR/Cas9 technology, we generated *Df1*, *Df3*, *Df1/2*, and *Df1/3* KO cells. The observed growth capacity alterations in the *Df* knock-out cells were negligible. Furthermore, we tested the differentiation capacity, and the ability to differentiate was affected in the generated knock-out cells, especially toward the neuroectoderm lineage. These findings coincide with literature reporting that epigenetic and epitranscriptomic alterations do not impact proliferation, but the differentiation ability in ES cells.

The second aspect reviewed the role of a chromatin modifier, MMSET in R-loop biology. MMSET is a histone methyltransferase, that additionally plays a role in pluripotency exit and mesendoderm specification. Moreover, MMSET is an interaction partner of R-loops, and several proteins have been revealed as common interaction partners of MMSET and R-loops. R-loop is a three-stranded structures that comprises a DNA-RNA hybrid and displaced single-stranded DNA. R-loops play a physiological role in cellular processes, in addition to a pathological role with a negative impact on transcription and replication, and are linked to DNA damage. Furthermore, R-loops have been reported to be enriched in polycomb group repressed genes. Hence, we wanted to validate MMSET as an interaction partner of R-loops, and further determine effects upon MMSET deletion.

By performing immunoprecipitation western blotting, we validated MMSET and R-loops as interaction partners with each other, and with the RNA helicases DHX9, DDX5, and DDX3. Three differentiation assays performed on heterozygous MMSET and MMSET null cells revealed no delayed pluripotency exit, but an altered ability to differentiate into the three germ layers upon MMSET removal. During immunofluorescence staining, we detected alterations in R-loop accumulation, and in foci formation of the DNA damage response proteins, γH2AX and 53BP1, in MMSET deleted cells upon oxidative stress. Moreover, we validated R-loop enrichment in the PcG repressed gene *Gata4,* independent of MMSET. Despite the observed altered differentiation ability, the overall effects upon removal of MMSET in R-loop biology were small. Overall, future studies may provide additional knowledge of how MMSET affects R-loop accumulation and distribution along the genome.

Sammendrag

Epigenetikk og epitranskriptomikk er vist å påvirke genregulering, og er koblet til mange ulike sykdommer og utvikling av kreft. Reversible endringer i genuttrykk som ikke påvirker nukleinsyresekvensen, er definert som epigenetisk regulering, mens modifikasjoner på RNA er kjent som epitranskriptomisk regulering. Målet med denne studien var å undersøke hvordan epigenetiske og epitranskriptomiske modifikasjoner påvirker genregulering i embryonale stamceller fra mus, ved å se på to ulike aspekter: hvordan «leserproteinene» YTHDF1-3 (DF1-3) påvirker cellulære prosesser, og rollen til et kromatinmodifiserende protein, MMSET, i biologien til R-loops.

YTHDF1-3-proteinene er cytoplasmiske lesere, som gjenkjenner og binder N^6-metyladenosin (m^6A) modifikasjoner. DF1-3-proteinene har blitt vist å både promotere translasjon og påvirke mRNA-stabiliteten. Enkelte studier rapporterer at DF1-3 har like funksjoner på grunn av høy sekvenslikhet, likevel påpeker de fleste andre studier ulike funksjoner. Derfor gjenstår det fortsatt å finne ut hvordan DF1-3-proteinene utfører deres potensielt ulike oppgaver, og om de binder de samme eller ulike m^6A seter.

Vi ønsket å slå ut *Df1-3* genene hver for seg og i ulike kombinasjoner i embryonale stamceller fra mus, og endelig undersøke hvordan DF1-3 proteinene påvirker proliferasjon og differensiering, samt samspillet mellom dem. Vi brukte creERT2 rekombinase og CRISPR/Cas9 teknologi for å slå ut *Df1*, *Df3*, *Df1/2,* og *Df1/3* genene. Vi observerte ubetydelige forandringer med tanke på vekst i embryonale stamceller fra mus som manglet en eller to ulike *Df1-3* gener. Deretter testet vi muligheten for differensiering i de genererte cellelinjene, og vi observerte da en forandring for differensiering, spesielt mot den neuroektoderme cellelinjen. De funnene som ble gjort samsvarer med teorien om at endringer i epigenetikk og epitranskriptomikk ikke påvirker celledeling, men heller differensiering av embryonale stamceller.

Den andre delene av denne oppgaven handler om rollen til et histonmodifiserende protein, Multiple-Myeloma SET domain, MMSET, som er viktig for R-loop-biologi. MMSET-proteinet metylerer histoner, og har vist seg å være viktig i overgangen fra pluripotent til differensiert tilstand, samt i dannelse av det mesendoderme kimlaget. I tillegg har MMSET blitt koblet til R-loop-biologi som en interaksjonspartner med R-loops. R-loop består av tre nukleinsyre tråder: ett dobbelttrådet DNA-RNA hybrid og et enkelttrådet DNA. R-loops har en fysiologisk rolle i cellulære prosesser, men spiller også ofte en patologisk rolle med negativ påvirkning på transkripsjon og replikasjon, i tillegg til å være forbundet med DNA skade. I

tillegg er R-loops rapporter å være beriket i polycomb nedregulerte gener. Derfor ville vi validere MMSET som en interaksjonspartner med R-loops i embryonale stamceller fra mus, og videre se på effektene i R-loop biologien ved å slå ut MMSET genet.

Ved å utføre immunpresipitering kombinert med western blotting ønsket vi å validere MMSET og R-loops som interaksjonspartnere med RNA helikasene DHX9, DDX5, og DDX3, i tillegg til med hverandre. Differensieringsforsøk ble utført på $MMSET^{+/-}$ og $MMSET^{-/-}$ celler. Resultatene viste at det ikke var noen forsinket utgang fra en pluripotent tilstand ved å slå ut MMSET. På den andre siden, var evnen for differensiering til de tre ulike kimlagene endret. Ved å bruke immunfluorescence forsøk oppdaget vi endringer i akkumulering av R-loops, samt økt foci-dannelse av γH2AX og 53BP1 ved å indusere oksidativt stress i $MMSET^{-/-}$ celler. Vi validerte at R-loops er beriket i det polycomb nedregulerte genet *Gata4* uavhengig av MMSET. Med unntak av den reduserte evnen for å nå en differensiert tilstand, observerte vi generelt små endringer med tanke på R-loop-biologien ved å fjerne MMSET. Fremtidige studier vil forhåpentligvis gi en dypere forståelse av hvordan MMSET påvirker akkumulering og distribusjon av R-loop i genomet.

Table of Contents

List of Abbreviations

2i	GSK3 and MEK1 inhibitor
5mC	5-methyl-cytosine
ATRA	all-trans retinoic acid
Cas9	CRISPR-associated protein 9
CRISPR	clustered regularly interspaced short palindromic repeats
creERT2	cyclization recombinase (Cre), fused with estrogen receptor (ER)
DNA	deoxyribonucleic acid
DDR	DNA-damage response
DRIP	DNA-RNA immunoprecipitation
DSB	double-stranded break
EB	embryonic body
EGF	epidermal growth factor
ES cell	embryonic stem cell
FACS	fluorescence-activated cell sorting
FGF	fibroblast growth factor
gDNA	genomic DNA
IgG	immunoglobin G
IP	immunoprecipitation
KD	knock-down
KO	knock-out
LIF	leukemia inhibitory factor
mES cell	mouse embryonic stem cell
mRNA	messenger RNA
m^6A	N^6-methyladenosine
NSC	neuronal stem cell
ORF	open reading frame
PBS	phosphate-buffered saline
PcG	polycomb group proteins
PCR	polymerase chain reaction
qPCR	quantitative polymerase chain reaction
RNA	ribonucleic acid

RNAPII	RNA polymerase II
SD	standard deviation
sgRNA	single guide RNA
SSB	single-stranded breaks
TF	transcription factor
UTR	untranslated region

1. Introduction

1.1 Embryonic stem cells

Embryonic stem (ES) cells are characterized as pluripotent cells since they can differentiate into the three germ layers, ectoderm, endoderm, and mesoderm, and grow indefinitely while maintaining pluripotency, also referred to as "self-renewal". ES cells are derived from the inner cell mass (ICM) of the mammalian blastocyst, prior to implantation (Fig. 1). Under appropriate culture conditions, ES cells can re-enter normal development when introduced into morula- or blastocyst stage embryos (Bradley et al., 1984; Yagi et al., 2017). Some of the unique advantages with ES cells *in vitro* is their ability to undergo genetic modifications and remain pluripotent, and that genetically altered ER cells easily can be clonally expanded while in culture (Yagi et al., 2017). Furthermore, today's knowledge of gene editing techniques like CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)/Cas9 (CRISPR associated protein 9), and other genome editing methods, make a vast research potential using ES cells (Jinek et al., 2012). Altogether, ES cells provide a great model system for investigation of early development, disease modeling, and drug screening, and comprise a unique potential for usage in regenerative therapy (Yagi et al., 2017).

1.1.1 Embryonic stem cell properties

There are two stable but epigenetically different states of pluripotent ES cells, termed naïve and primed ES cells. Naïve ES cells are isolated from pre-implantation embryos, while primed ES cells, also named epiblast stem cells (EpiSCs), are commonly isolated from post-implantation epiblast (Brons et al., 2007; Tesar et al., 2007) (Fig. 1). Naïve ES cells have increased clonogenicity, diverse growth factor requirements, and altered morphology and energy metabolism compared to EpiSCs. In contrast to naïve ES cells, primed cells can not contribute substantially to blastocyst chimera formation, since they are in an advanced developmental stage and "primed" for differentiation (Nichols & Smith, 2009).

Some of the transcription factors required (TFs) during the resetting of primed ES cells are dispensable during self-renewal, however, a common biological program that both induce and maintain naïve pluripotency has been suggested (Dunn et al., 2019). Whether naïve pluripotency is acquired in a random fashion, where individual cells follow different trajectories, or in an ordered process with a precise sequence of events, was undetermined for

a long time. First, it was believed that ES cell fate was decided through a complex and vast regulatory system, however, later research suggested that ES cell fate can be decided through a relatively simple process of molecular computation (Dunn et al., 2014). Recently, it has been reported that ES cells progress through defined stages, before stabilization and hierarchical activation of the naïve state and its requiring pluripotency transcription network (Dunn et al., 2019).

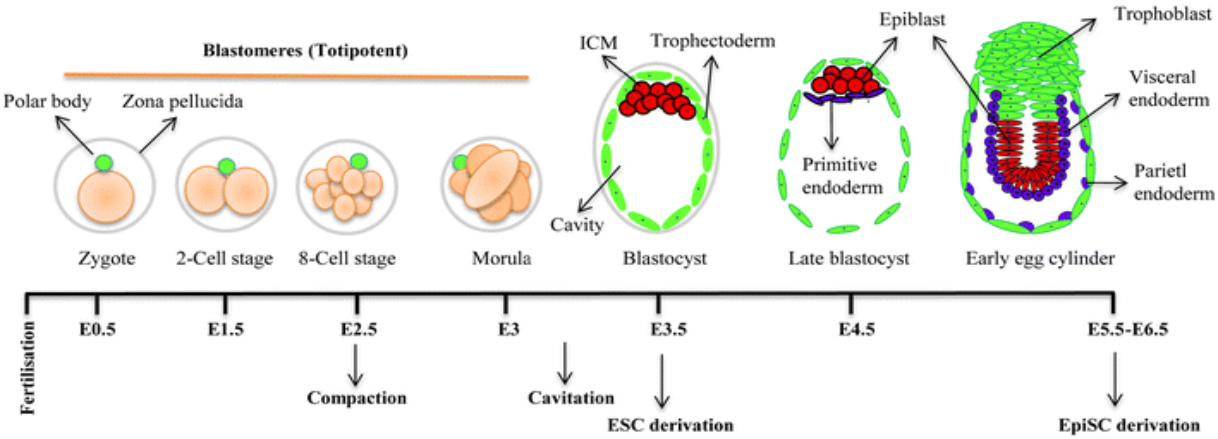


Figure 1. Origin of mouse embryonic stem cells. A single cell zygote develops after fertilization. At this stage the cells are totipotent, thus developing to all three germ layers, placenta, and extra-embryonic tissue. The zygote proceeds through a series of cleavage divisions, and compaction occurs at E2.5 when the cells of the outer part of the morula gets bound tight together. The cells of the outside and inside starts to differentiate into progenitors of the trophoblasts (outside) and inner cell mass (inside). A cavity is formed inside the morula and creates the blastocyst, where the outer cells make up the trophectoderm, while the cells that are left in the interior make up the inner cell mass. The trophectoderm turns into the placenta and the extra-embryonic tissue. Moreover, the late blastocyst will develop at E4.5, and at E5.5-6.5 the blastocyst will further develop to an early egg cylinder and prepares for germ layer specification. Naïve and primed ES cells are derived from the ICM at E3.5, and the epiblast at E5.5-6.5, respectively. "E" refers to embryonic day. Figure from Huang et al., 2015.

1.1.2 Pluripotency is maintained through the core transcription factors

Numerous transcription factors contribute to maintaining a pluripotency state in embryos and ES cells. Some of the main pluripotency factors include OCT4, NANOG, SOX2, KLF4, and c-MYC (Avilion et al., 2003; Cartwright et al., 2005; Chambers et al., 2003; Godini et al., 2018; Mitsui et al., 2003; Nichols et al., 1998). The three factors NANOG, OCT4, and SOX2 have later been described as the core transcription factors in maintaining pluripotency. These core TFs inhibit the transcription of genes associated with differentiation lineages while promoting the expression of those associated with self-renewal (Boyer et al., 2005; Loh et al., 2006).

NANOG is a homeodomain-containing protein, which is expressed in pluripotent ES cells, ES cell derivatives, and in the developing germline of mammals (Boyer et al., 2005; Loh et al., 2006; Mitsui et al., 2003). In humans, NANOG facilitates molecular reprogramming, and it is reported that *Nanog* null embryos do not develop beyond implantation (Mitsui et al., 2003). However, NANOG is dispensable for pluripotency *ex vivo* (Chambers et al.,

2007). Reprogramming studies have reported that NANOG is decisive in the last phase for reaching a pluripotent ground state. At this point in the reprogramming, the other significant factors are already present, and studies suggest that activation of endogenous *Nanog* is efficient to complete the reprogramming together with other key factors present, like OCT4 and SOX2 (Silva et al., 2009).

OCT4 is a member of the POU (PIT/OCT/UNC) class of homeodomain proteins, encoded by *Pou5f1* (Godini et al., 2018; Nichols et al., 1998). OCT4 interacts with TFs to activate or repress gene expression in mouse-derived ES (mES) cells, and inactivation of *Oct4* have been shown to prevents self-renewal and induce the differentiation of ES cells (Nichols et al., 1998; Niwa et al., 2000; Rizzino & Wuebben, 2016). OCT4 forms a heterodimer with the SRY-related high-mobility group (HMG) box (SOX) transcription factor 2, SOX2, and together they regulate gene expression (Masui et al., 2007).

OCT4 and SOX2 operate synergistically to activate the Oct-Sox enhancer, which regulates the expression of genes like *Nanog*, in addition to *Oct4* and *Sox2* themselves. SOX2 is dispensable in this activation but required for regulation of TFs associated with *Oct4* expression. The essential role of SOX2 in pluripotency is therefore to stabilize ES cells in a pluripotent state by maintaining the required levels of *Oct4* expression (Masui et al., 2007).

Additionally, there are a handful of transcription factors reported in the ES cell network, and many of these work together to maintain pluripotency. For example, ESRRB and SALL4 interacts with NANOG, thus co-occupying genomic sites of NANOG in mES cells, and function as hubs between intrinsic pluripotency determinants and external signaling pathways (Huang et al., 2015). For a full description of transcriptions factors influencing pluripotency, see review Huang et al., 2015. Additionally, many of these TFs influencing the ES cell network are also reported to be associated with several types of human cancers (Rizzino & Wuebben, 2016).

Many of the core pluripotency transcription factors are linked to oncogenic reprogramming. For example, the amount of OCT4 is important for ES cell fate, but a dependency on the dose of OCT4 is also linked to oncogenic potential, as a higher dose of OCT4 increases the malignant potential (Gidekel et al., 2003). *Oct4* expression is recently found to be associated with tumor size and number, as well as during differentiation in hepatocellular carcinoma, and in glioma tumor cells of children (Liang et al., 2018; Zheng et al., 2018). For example, the *Oct4* expression in tumors of glioma tumor cells are significantly increased at both messenger RNA (mRNA) and protein level, compared to healthy children (Zheng et al., 2018). The *Sox* family and *Nanog* are also related to tumorigenesis and metastasis.

SOX2 is suggested to be required for maintaining cancer stem cells in ovarian cancer, and NANOG helps cancer cells escape attacks from natural killer cells in the immune system during tumorigenesis, as some examples (Saga et al., 2019; Wen et al., 2017). Overall, this indicates the importance of *Nanog*, *Oct4*, and *Sox2*, not only for maintenance and induction of pluripotency, but also for understanding oncogenic reprogramming and as possible drug targets.

1.1.3 Induced pluripotent stem cells

Pluripotent stem cells have a vast potential in research and regenerative medicine, but because there are a lot of ethical difficulties regarding the use of human embryonic stem cells in research new approaches have been sought for a long time. By using stem cells in transplantation another problem is tissue rejection in patients, but if cells from the patient are manipulated and then used in regenerative therapy, this rejection is avoided (Godini et al., 2018). A major breakthrough came in 2006 when differentiated embryonic and adult fibroblasts were reprogrammed to pluripotent cells, termed induced pluripotent stem (iPS) cells. iPS cells were generated by introducing OCT3/4, SOX2, c-MYC, and KLF4 under ES cell culture conditions, later termed "Yamanaka factors" (Takahashi & Yamanaka, 2006) (Fig. 2A).

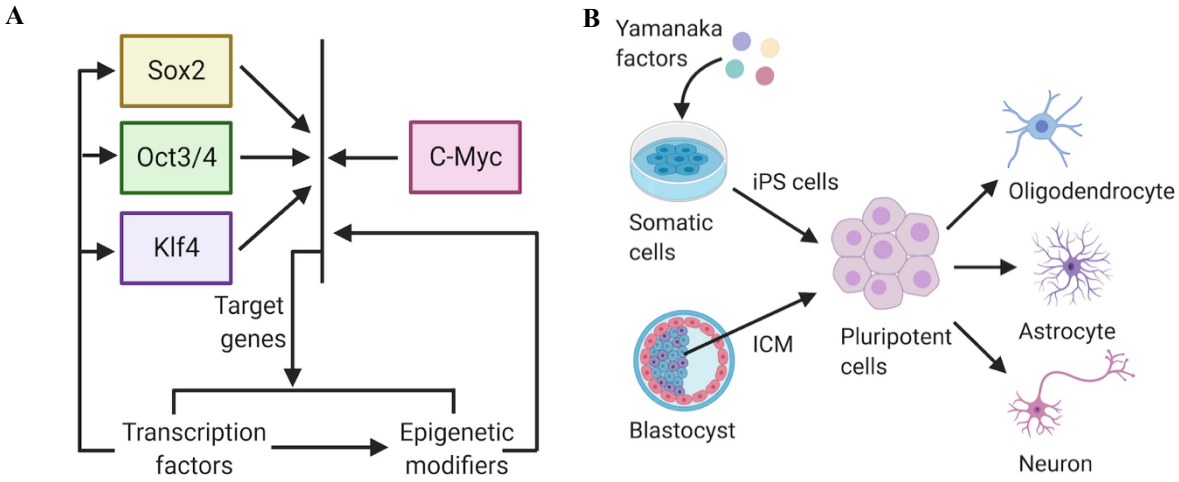


Figure 2. Induction of iPS cells and differentiation of pluripotent cells.
(A) The "Yamanaka factors" OCT3/4, SOX2, c-MYC, and KLF4 are essential for generation of iPS cells. Through regulation of transcription factors and epigenetic modifiers, the "Yamanaka factors" induces pluripotency in differentiated cells. Figure inspired by Niwa, 2007.
(B) Pluripotent cells derived from the inner cell mass or somatic cells, can be differentiated into all cell lines, including cells in the neuronal lineage.

The "Yamanaka factors" are essential during reprogramming. When the "Yamanaka factors" are introduced the global gene expression patterns change, and subsequently erase epigenetics marks. This is essential since all epigenetic marks that are related to cell identity need to be removed for stem cells to gain pluripotency characteristics (Godini et al., 2018).

Additionally, other transcription factors and microRNAs are also identified with essential regulatory roles in reprogramming and maintaining pluripotency in stem cells (Kuppusamy et al., 2013). Since the "Yamanaka factors" were adequate for the reprogramming, it indicates that endogenous levels of the other key transcription factors were sufficient. iPS cells express ES cell marker genes, have equal growth properties, and identical morphology to ES cells (Takahashi & Yamanaka, 2006). Additionally, both iPS cells and ES cells can differentiate into the three germ layers (Niwa, 2007) (Fig. 2B). With this knowledge, an emerging possibility to use cells from patients in stem cell therapy evolved.

1.1.4 MEKi, GSKi, and LIF – essential components for *in vitro* culture

It is well established that the combination of mitogen-activated protein kinase kinase (MEK) inhibitor (MEKi) and glycogen synthase kinase-3 (GSK-3) inhibitor (GSKi), also termed 2i, in addition to the leukemia inhibitory factor (LIF), provide optimal cell culture conditions of mES cells *in vitro* (Martello & Smith, 2014; Sim et al., 2017) (Fig. 3). Together 2i/LIF preserve mES cells in a naïve ground state, with maintained pluripotency and self-renewal. Through two distinct mechanisms, 2i maintain the ES cells´ ground state by downregulating global DNA methylations (see section 1.2.1 and 1.2.2 for further information about DNA methylations, DNMTs and TET).

The MEK enzyme MEK1 phosphorylates JMJD2C, and marks JMJD2C for ubiquitin-mediated protein degradation (Sim et al., 2017). In contrast MEKi, increases JMJD2C protein levels, which subsequently promotes ten-eleven translocation (TET) 1 activity. Hence, TET1 (a DNA demethylase) increases 5-hydroxymethylcytosine (5hmC) levels. Additionally, MEKi reduces 5mC levels as a result of reduced DNA methyltransferase (DNMT) 3 expression. Also, GSKi suppresses *Dnmt3* transcript expression and subsequently decreases both DNMT3A/B levels and DNA methylations. GSKi decreases 5mC levels without altering 5hmC levels, which indicates that GSKi promotes DNA demethylation with another mechanism than MEKi, and that GSKi demethylation do not involve active demethylation by the TET family. 2i also promotes DNMT3A/B protein degradation by inhibiting DNMT3A/B protein expression. Altogether, 2i maintain ES cells in a naïve ground state by DNMT3A/B protein degradation and TET1 activation through JMJD2C (Sim et al., 2017).

In addition to 2i, ES cells were earlier cultured with a feeder layer of mitotically arrested embryonic fibroblast, and fetal calf serum, but this complex system was altered after the discovery of LIF (Mulas et al., 2019; Pease et al., 1990). LIF strengthens ES cells in culture by

supporting derivation and propagation, and is essential for self-renewal pathways. The LIF receptor consists of two signaling chains, gp130 and LIFRβ (Fig. 3A). LIF activates JAK1 that phosphorylates both gp130 and LIFRβ, which then act as scaffolds to recruit, among other, STAT3. Further, JAK1 also phosphorylates STAT3, and STAT3 is required for maintaining ES cell pluripotency. Moreover, overexpression of a dominant-negative STAT3 construct in ES cells has been reported to promote differentiation and loss of self-renewal (Nicola & Babon, 2015). All in all, the discovery of LIF strengthened the culture conditions of ES cells.

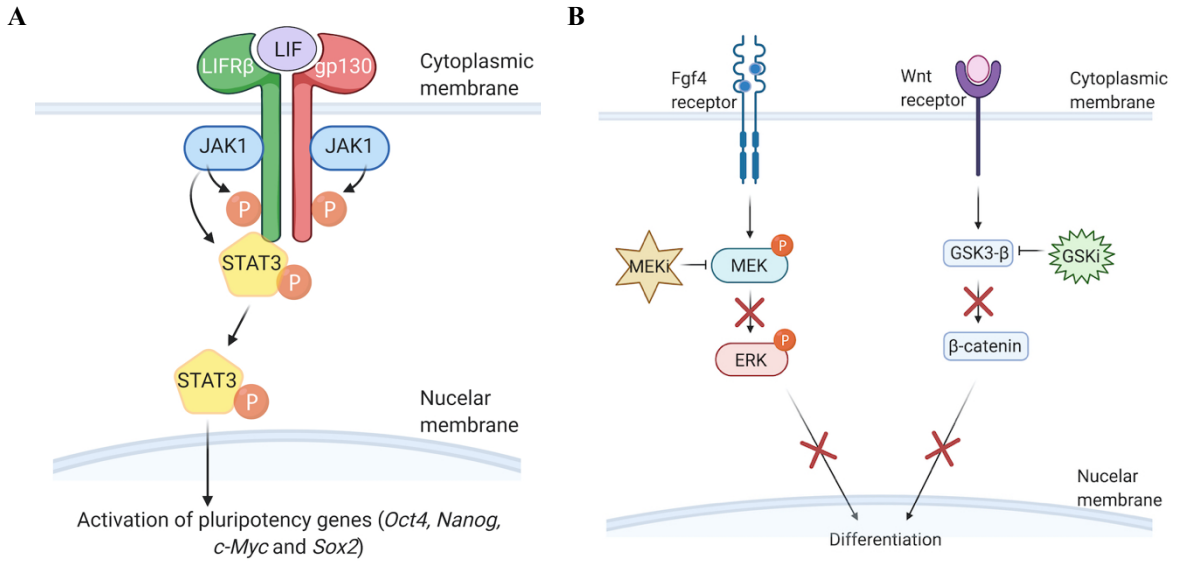


Figure 3. 2i and LIF signaling in mES cells.
(A) LIF signaling. Gp130 and LIFRβ are two side chains of the LIF receptor. When LIF is bound to the receptor, JAK1 will be activated and phosphorylate these two side chains, which again function as scaffolds to recruit STAT3. JAK1 will then phosphorylate STAT3, which results in activation of several pluripotency genes, including *Oct4*, *Nanog*, *c-Myc* and *Sox2*.
(B) 2i signaling. Presence of LIF and inhibition of GSK3 and MEK/ERK by GSKi and MEKi, respectively, allows efficient ES cell self-renewal and suppresses differentiation.

1.1.5 ES cells harbor highly conserved m^{6}A modifications

m^{6}A modifications are widespread in the mammalian ES cell genome (Batista et al., 2014). See section 1.3.1 for further details regarding m^{6}A modifications. Most key regulators of ES cell pluripotency and germ lineage determinations exhibit m^{6}A modifications. The m^{6}A modification pattern in ES and somatic cells is more or less identical, which indicates one mechanism that deposits m^{6}A modifications in early embryonic states. This is in high contrast to 5-methyl-cytosine (5mC) in DNA and histone lysine methylations, which follow distinct rules during the reprogramming in pluripotent versus somatic cells. m^{6}A modifications are suggested to mark RNA for turnover in both mES and human ES (hES) cells. A significant overlap between mES and hES cells proposes a functional significance of m^{6}A patterns in ES

cells. At the same time, there are numerous species-specific m^6A patterns, that might contribute to specific aspects of ES cell biology (Batista et al., 2014)

Several pluripotency and differentiation factors in both mES and hES cells are m^6A modified. Many of the main regulators of pluripotency, like *Nanog*, *Klf4*, and *Myc* are found to be m^6A modified in mES cells. In hES *Nanog* and *Sox2* are m^6A modified, while *Oct4* does not harbor this modification in neither mES nor hES cells (Batista et al., 2014). Hence, m^6A plays an important role between pluripotency exit and differentiation, thus ensures orderly differentiation of ES cells (Geula et al., 2015).

1.1.6 Differentiation capacity towards the three germ layers

The application of small molecules has approached as a new perspective for target differentiation, and for culture-expansion of stem cells *in vitro*. There are a handful of molecules that potentially can affect pluripotency, like 2i/LIF. Recently, also other molecules have been discovered, like the phytoestrogen molecule Icaritin, that enhances mES cell self-renewal through upregulating the core pluripotency factors. This upregulation contributes to the long-term expansion of mES cells, while the cells maintain pluripotency (Tsang et al., 2017). Under defined conditions, mES cells can be differentiated into lineages of the three germ layers, ectoderm, endoderm, and mesoderm (Fig. 2B). These three germ layers are progenitors for all fetal tissues, and the specification to the different layers is affected by cellular interactions and signaling (Tam & Behringer, 1997; Tam et al., 1999; Yagi et al., 2017). The *in vitro* capacity of mES cells to differentiate is essential in the iPS disease model, for therapeutic applications and for studying early development. This emphasizes the importance of small molecules in maintaining pluripotency and for target differentiation of ES cells in culture (Takahashi & Yamanaka, 2006; Yagi et al., 2017).

Upon LIF withdrawal during cultivation of mES cells, the mES cells express genes representing the three germ lineages. Eventually, mES cells has a definite expression pattern of the primitive ectoderm lineages, since they have a tendency to differentiate into neuronal stem cells (NSCs) upon LIF removal (Ying et al., 2003). At this stage, the ES cells would still express OCT4 and initiate expression of the fibroblast growth factor 5 (FGF5), a primitive ectoderm marker. In LIF free media, the pluripotency markers *Nanog* and *Oct4* will be downregulated first. After the downregulation of *Nanog* and *Oct4*, the ES cells will gradually increase the expression of mesendoderm markers, including *Gata4, Gata6, Foxa2*, and *Brachyury*. Simultaneously, the differentiation will further be directed into ectodermal, and subsequently

neuroectodermal fates, where *Nestin, Pax6,* and *Sox1* are some of the very early expressed neuronal genes, and they are well characterized at different stages of neurogenesis (Cai & Grabel, 2007; Desai & Pethe, 2020).

Further neuronal differentiation can be induced by adding EGF/FGF, and expression of transcription factors crucial for mesoderm and endoderm formation decreases (Desai & Pethe, 2020). Finally, the ES cells will differentiate into mature NSCs, which can be observed by morphological alterations of the cells, since the cells elongate into a radial glia-like phenotype. When the cells are differentiated into mature NSCs, *Nestin,* and *Sox1* are highly up-regulated and the cells respond to EGF and FGFb (Cai & Grabel, 2007).

1.2 Epigenetic gene regulation in embryonic stem cells

Epigenetic regulation is defined as a reversible alteration in gene expression with no effects on the nucleotide sequence of DNA. By DNA methylation, transcription factor assembly, and through interfering in chromatin structures, epigenetic modifications exhibit its impact on gene regulation (Godini et al., 2018). After each cell division, the cells may or may not preserve the epigenetic pattern of their parental cells, depending upon their fate. The identity of each cell type is defined through its epigenetic signature, and the epigenetic pattern of the daughter cells will be changed to a new cell identity during differentiation. In stem cells, all epigenetic signatures related to cell-identity need to be removed to reach a pluripotent ground state. Furthermore, the generation of iPS cells from somatic cells is dependent on resetting the gene expression pattern to a ground state (Godini et al., 2018; Takahashi & Yamanaka, 2006). Epigenetic modifications are reversible in contrast to genetic modifications, but they can both trigger cancer formation. The fact that epigenetic modifications are reversible makes it possible to revert tumors to more benign phenotypes, and an enormous potential is emerging with possibilities for novel therapeutic targets (Yagi et al., 2017). This reflects the power of epigenetics modulation, made possible by experimental manipulations of epigenetic patterns that are known (Godini et al., 2018).

1.2.1 Epigenetic modification of DNA

DNA cytosine methylation is the major DNA modification, and is associated with gene silencing and plays an important part in developmental processes, by inactivation of X-chromosomes and genomic imprinting. DNA methylations are stably maintained in somatic

tissues, but dynamic changes in patterns and levels have been observed during embryonic development. Many DNA-binding proteins attach to the major groove of the DNA helix, where the methyl moiety of the methylcytosine exists. Therefore, the DNA cytosine methylations are likely to perform their activity by attracting or repelling various DNA-binding proteins (Li & Zhang, 2014).

The formation of 5mC methylation at CpG sites is the most abundant epigenetic modification (Jeltsch & Jurkowska, 2016; Siegfried & Cedar, 1997). CpG sites are palindromic sites with a very high density of C•G base pairs, and have a unique potential since methylation occurs at both DNA strands and can be transferred from parent to daughter strands. This is why CpGs are the most studied DNA methylation. CpHs (H=A, T, C) may also be methylated, but in somatic cells, more than 98% of DNA methylations occur in a CpG dinucleotide context. CpHs are rare, but in special cell types like oocytes, neurons, and ES cells, CpHs are more common (Haines et al., 2001; Imamura et al., 2005; Lister et al., 2009; Ramsahoye et al., 2000; Tomizawa et al., 2011; Ziller et al., 2011). In ES cells, as much as 25% of all methylations appear in a non-CpG context (Ramsahoye et al., 2000).

DNA methyltransferases (DNMTs) catalyzes DNA methylations, which can be lost spontaneously or actively. Spontaneous demethylation happens when cytosines are not methylated during replication, or from DNA repair when regions in the genome are repaired due to some errors, while active demethylation of 5mC is a process performed by ten-eleven translocator (TET) and thymine DNA N-glycosylation (TDG) enzyme actions (Li & Zhang, 2014; Okano et al., 1999).

DNMT´s – de novo and maintenance transferases

The DNMT family are enzymes catalyzing the transfer of a methyl group to DNA, and this family is divided into three classes: DNMT1, DNMT2, and DNMT3A/B/L (Fig. 4A). Unlike the other enzymes in the DNMT family, DNMT3L does not possess any enzymatic activity, but works as a coactivator for DNMT3A and DNMT3B (Okano et al., 1999; Zeng & Chen, 2019). Both DNMT1, 2, and 3A/B transfer a methyl group from the universal methyl donor, *S*-adenosyl-L-methionine (SAM), to the 5-position of cytosine residues. Even though DNMT2 contains all the 10 conserved sequence motives as canonical DNA methyltransferases, including the SAM motif, sequence comparisons have shown that DNMT2 lacks a putative nucleic acid binding cleft, which makes it difficult to accommodate duplex DNA (Goll et al., 2006; Jeltsch & Jurkowska, 2016; Van den Wyngaert et al., 1998). The function of DNMT1

and 3 are more investigated than DNMT2´s function, but it appears that DNMT2 neither methylate single- nor double-stranded DNA, but RNA (Goll et al., 2006).

DNA methylations can be divided into two types, and the first is *de novo* DNA methylation activities by DNMT3A/B, while the second type maintains DNA methylations during cell division. DNMT3A/B catalyze the reaction when methyl groups are added to cytosine at unmethylated DNA during embryonic development (Jeltsch & Jurkowska, 2016; Okano et al., 1999). Unmethylated cytosine residues of hemimethylated DNA (a single CpG that is methylated on one of the complementary strands), are methylated on newly replicated DNA in the S-phase. This ensures that the methylation pattern of the parental DNA is preserved (Bestor & Verdine, 1994; Jeltsch & Jurkowska, 2016). In contrast to CpG methylations, there is no evidence of maintaining methylations at CpH sites, thus these methylations needs to be reestablished *de novo* after each cell division (Ramsahoye et al., 2000).

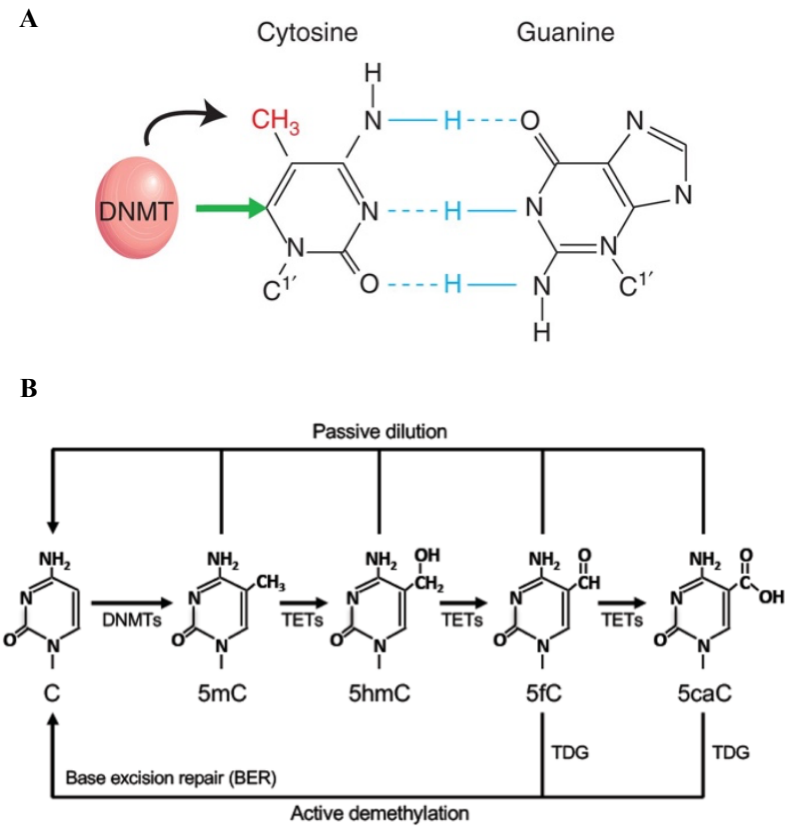


Figure 4. DNA cytosine methylation. (A) DNA cytosine methylation is mediated by DNMTs. The DNMT bind covalently to the 6th carbon on the pyrimidine ring (shown with green arrow) and transfer a methyl (CH3) group (red) to the 5th carbon on cytosine. This methylation does not affect the binding capacity of cytosine to guanine (blue). Figure from Li & Zhang, 2014.
(B) Methylation and demethylation of cytosine by DNMT, TET, and TDG. DNMTs work to methylate cytosine to 5mC, and TETs and TDGs work together to active demethylate 5mC to 5fC, or to 5caC via 5hmC, to regular cytosine. Figure from Zeng & Chen, 2019, with some modifications.

DNA demethylases – TET and TDG

Loss of methylation can occur passively or actively, which is dependent of DNA replication, in addition to the TET and TDG enzymes (Li & Zhang, 2014). Active loss of methylation happens by indirectly removing methylated cytosines, a mechanism involving the DNA hydroxylase family known as the TET enzymes. This family consists of three enzymes: TET1, TET2, and

TET3 (Gu et al., 2011; Ito et al., 2010; Tahiliani et al., 2009; Williams et al., 2011). Dioxygenase activity implies the ability to convert 5mC to the 5mC derivatives: 5-hydroxymethyl C (5hmC), 5-formyl cytosine (5fC), and 5-carboxylcytosine (5caC) (Ito et al., 2010; Tahiliani et al., 2009) (Fig. 4B). This is a stepwise process that results in the demethylation of cytosines. The derivatives 5fC and 5caC will be recognized and excised by another DNA demethylase termed TDG, thus, reverting the derivatives to unmodified cytosines (He et al., 2011). TDG works as a base-excision repair system (BER), and acts as both a thymine and uracil DNA N-glycosylase, on T and U in T/U·G mismatches, to keep stability in CpG sites in the genome. TDG demethylates down-stream of the TET dioxygenases, by recognizing and removing 5fC and 5caC, and then base-excision repair occurs to finally revert 5fC and 5caC to unmodified cytosine (He et al., 2011) (Fig. 4B). Finally, the activity of TET and TDG demethylate 5mC to regular cytosine.

1.2.2 Epigenetic modification of histones

In eukaryotic cells, DNA is tightly packed around four core histones (H3, H4, H2A, and H2B) in the nucleus. DNA and histones exist in repeating units termed nucleosomes, which form the structure of the chromatin (Oudet et al., 1975; Zhao & He, 2015; Zubay & Wilkins, 1962). This structure affects gene regulation, by controlling the accessibility of the transcription machinery to the DNA. Histone modifications and the enzymes implementing them, can affect transcription as well as nucleosome dynamics and chromatin compaction. Both external and

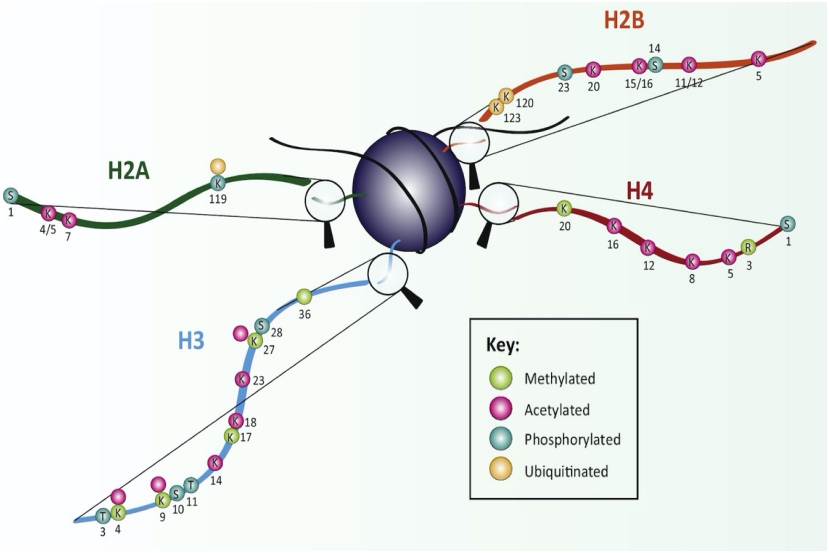


Figure 5. Schematic overview of post-translational modifications of the histone tails.
The nucleosome with the four core histones tails, H3, H4, H2A, and H2B. Each residue is marked with a color, that indicates which modification that occurs. Amino acids are marked with K, R, S, and T, for lysine, arginine, serine, and threonine, respectively.
Methylations are marked with green, acetylation with pink, blue for phosphorylation, and yellow indicates ubiquitination. Figure from Lawrence et al., 2016.

intrinsic signals can stimulate histone modification implementations, and a disrupted balance of these processes is often observed in human cancers (Zhao & Shilatifard, 2019).

Post-translational modifications (PTMs) can affect the folding of chromatin and the recruitment of proteins to the chromatins (Allshire & Madhani, 2018). The histone can be divided into globular regions and N-terminal tails. The globular regions form the core of the nucleosome, while the N-terminal tails protrude from the nucleosome. The N-terminal tails makes contact with other nucleosomes, and are highly enriched with PTMs (Godini et al., 2018) (Fig. 5). The globular regions of the histone that are in contact with DNA can also undergo PTMs (Lawrence et al., 2016). PTMs occurring on N-terminal tails protruding from the nucleosome affects inter-nucleosomal interactions, hence, affecting the chromatin structure (Allshire & Madhani, 2018).

The most dominant post-translational modifications histones can undergo include methylation, acetylation, phosphorylation, and ubiquitination (Godini et al., 2018). Histone acetylases (HAT) catalyzes the transfer of an acetyl group to a positively charged lysine side chain and neutralizes it. This can weaken the interaction between histones and DNA, and HAT is associated with active genes. Histone deacetylases (HDAC) perform the opposite reaction and are predominantly associated with transcription repression (Bannister & Kouzarides, 2011). Histone phosphorylation occurs on serines, threonines, and tyrosines, and adds a negative charge on mainly the N-terminal of the histone tails. There are different events associated with phosphorylation of histone tails, including DNA repair, where phosphorylation of Serine 139 of H2AX (γH2AX) is induced by DNA damage, an early response in DNA double-strand break (DSB) signaling (Zhao & Shilatifard, 2019). Histone methyltransferases (HMT) are histone-modifying enzymes that mono-, di-, or trimethylates histones, and are removed by histone demethylases (HDM). Details regarding HMTs and HDMs, and their functions, will further be explained in the next two paragraphs.

Histone methylation

Histone methylations are associated with transcriptional regulation by affecting the chromatin, transcription factors recruitment, interaction with elongation, and initiation factors, and by affecting RNA processing. These methylations have key roles in differentiation and development, while in tumorigenesis, studies indicate that abnormal levels of histone methylations play a causal role (Zhao & Shilatifard, 2019). Histone methylation occurs at lysine (K) and arginine (R) residue sites, predominantly on H3 followed by H4 (Zhao & Shilatifard, 2019) (Fig. 5). We will further focus on lysine methylations, due to the interest of MMSET, a

lysine methyltransferase. Six classes of methyltransferase complexes (KMT1-6) have been identified to mono-, di-, and trimethylate histone lysines (Mohan et al., 2012). KMT family 1, 2, 4, 5, and 6 methylate H3K9, H3K4, H3K79, H4K20, and H3K27, respectively (Jørgensen et al., 2013; Lehnertz et al., 2003; Margueron & Reinberg, 2011; Nguyen & Zhang, 2011; Shilatifard, 2012; Steger et al., 2008; Zhao & Shilatifard, 2019). The KMT3 family consists of the nuclear receptor-binding SET domain (NSD) proteins 1, 2, and 3, which methylate predominantly H3K36, and are explained in detail in paragraph 1.2.3 (Li et al., 2009; Rahman et al., 2011; Rayasam et al., 2003). Methylations do not alter the charge of the side chains but elicit different outcomes of transcriptional regulation. H3K4, H3K36, and H3K79 are the best-characterized methylations marks on lysine residues linked to transcriptional activation, while H3K9, H4K20, and H3K27 are the main marks linked to transcriptional repression (Zhao & Shilatifard, 2019). No histone demethylase has been identified to remove H3K79 marks yet, but all the other methylation marks can be removed by histone demethylases (KDMs) (Chory et al., 2019).

Histone demethylation

Six families of histone lysine demethylases have been identified, which have overlapping and unique functions (Zhao & Shilatifard, 2019). Lysine demethylase (LSD) 1 was the first lysine demethylase identified, which utilizes FAD as a co-factor. A protonated nitrogen is required for this demethylation process and it is therefore only compatible with lysine substrates that are mono- and dimethylated (Shi et al., 2004). The Jumonji (JmjC) domain is present in all other family members of lysine demethylases, which makes these members capable of removing trimethyl marks. The six families of KDMs, in addition to the JMJD2 subfamily, are reviewed in Zhao & Shilatifard, 2019. Like the histone methyltransferases, all KDMs have been associated with the development of different cancers, and elucidation of their cellular functions are important in the development of KDMs as a potential drug target (Zhao & Shilatifard, 2019).

1.2.3 The nuclear receptor-binding SET domain proteins

The NSD family catalyze and recognize methylation of histone lysine marks, thus, regulate both gene expression and chromatin integrity. The NSD proteins mono- and dimethylate H3K36, which in general is a hallmark of active transcription and are associated with transcription of active euchromatin (Li et al., 2009; Rahman et al., 2011; Rayasam et al., 2003; Wagner & Carpenter, 2012). NSD methyltransferases have been shown to be translocated,

overexpressed, or somatically mutated in multiple types of human cancer. Further, alteration and translocation of NSDs are reported to affect cell proliferation and differentiation, which is linked to different developmental defects. This implicates the important role of NSD proteins in human diseases, and as a possible novel therapeutic target (Bennett et al., 2017). NSDs are a phylogenetically separate subfamily of lysine HMTs and comprise NSD1, NSD2, and NSD3 (Morishita & di Luccio, 2011) (Fig. 6). NSD2 are also termed Multiple-Myeloma SET domain (MMSET) or Wolf-Hirschhorn Syndrome Candidate 1 (WHSC1) (we refer to this protein as MMSET from here). The NSD family are large multidomain proteins with a conserved SET domain, which can be subdivided into pre-SET, SET, and post-SET domains. Further, the NSD family comprises two PWWP (proline-tryptophan-tryptophan-proline) domains which are important in the binding to methylated histone H3, and their plant homeodomain (PHD) zinc fingers are critical for interactions with other methylated histones (Baker et al., 2008; Lucio-Eterovic & Carpenter, 2011; Pasillas et al., 2011; Sankaran et al., 2016).

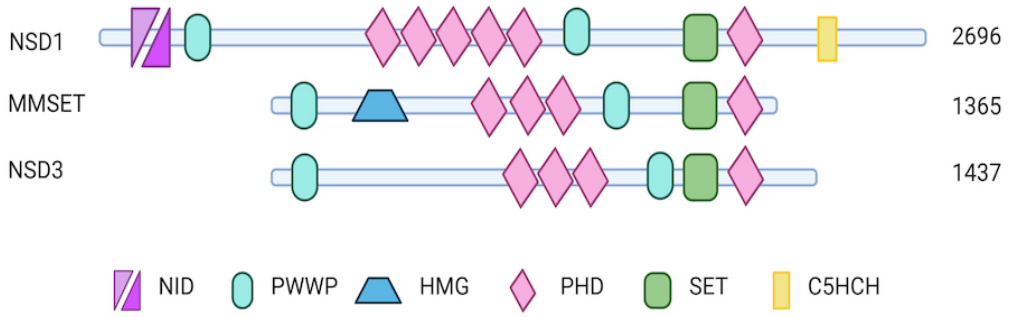


Figure 6. Structural overview of the NDS family proteins, NSD1, MMSET (NSD2/Whsc1), and NSD3. The major domains of the NSD family are: nuclear receptor binding domains (NID) (light purple indicates ligand independent and dark purple ligand dependent), PWWP domain (turquoise), high mobility group (HMG) (blue), plant homeodomain (PHD) (pink), SET domain (green) and CH5CH (indicates chromatin associated zing fingers) (yellow). Figure inspired by Lucio-Eterovic & Carpenter, 2011, and not drawn to scale.

MMSET

MMSET produces two transcripts, one full-length protein of ~180 kDa, and an isoform at ~100 kDa that corresponds to the amino (N)-terminal portion (Tian et al., 2019). The catalytic SET domain on the carboxy (C) terminus is required for dimethylation of H3K36, the principal chromatin regulatory activity of MMSET (Kuo et al., 2011; Tian et al., 2019) (Fig. 6). Studies have reported that ES cells derived from mice with excised SET domain of MMSET, behaved equal to the wild type, and that these ES cells still expressed two protein isoforms, a large protein at ~130 kDa, which corresponds to a truncated C-terminal, and the ~100 kDa isoform. A complete MMSET knock-out leads to the deletion of a fragment encoding both protein

isoforms, resulting in ES cells lacking both the long and short isoform of MMSET. After differentiation induction of MMSET knock-out cells, a delayed downregulation of *Oct4* and *Nanog* were observed. These findings indicate that the N-terminus of MMSET has a role in the induction of pluripotency exit in absence of the catalytic SET domain, but the N-terminus has also been found to be essential for mesendoderm specification, by binding to enhancers of mesendoderm regulators (Tian et al., 2019).

It has been found that MMSET is either overexpressed or hyper-activated in certain forms of cancer, including acute lymphoblastic leukemia and multiple myeloma. In multiple myelomas, up to 20% of patients have translocation of MMSET to IgH enhancer locus leading to a massive overexpression of MMSET, and concomitant high levels of global H3K36me2 (Chesi et al., 1998; Kuo et al., 2011). Additionally, the methyltransferase activity of the SET domain mediates increased methylation of H3K36 on promoters of oncogenes that drive the disease, as a result of this overexpression/hyper-activation (Kuo et al., 2011; Nimura et al., 2009; Popovic et al., 2014; Tian et al., 2019).

MMSET has also been linked to R-loop biology as an interactor of DNA-RNA hybrids, in addition, MMSET and R-loops have many common interaction partners. More information about R-loops in section 1.4. However, there are few studies highlighting the function of MMSET in R-loop biology, but it is reported that knock-down of MMSET decrease the global levels of R-loops (Cristini et al., 2018).

NSD1 and NSD3

In addition to MMSET, the KMT3 family also consists of NSD1 and NSD3. NSD1 knock-out mice are embryonic lethal, showing that NSD1 is required for normal growth and development (Rayasam et al., 2003). Depletion of NSD1 has in addition, revealed a reduction in levels of H3K36me1/2/3, indicating that NSD1s products can be further trimethylated by SETD2, a mammalian methyltransferase that preferentially performs trimethylations. Furthermore, the reduction of NSD1 have also been reported to reduce RNA polymerase II (RNAPII) promoter occupancy, and the transition of RNAPII from an initiation to elongation-component state is hindered (Lucio-Eterovic et al., 2010). Different diseases are linked to abnormal expression of NSD1, including Sotos syndrome and tumorigenesis, but it still requires further studies to fully elucidate what cellular context or mechanism that controls the oncogenic properties of NSD1 (Bennett et al., 2017).

NSD3 interacts with TET3 (see section 1.2.1) to get activated and stimulates H3K36 trimethylation (Perera et al., 2015). NSD3 is an important methyltransferase for, among other,

neural crest (gives rise to most of the peripheral nervous system) gene expression during specification, and during transcription of neuronal genes in retinal cells (light-sensing cells) (Jacques-Fricke & Gammill, 2014). Overexpression of NSD3 has additionally been linked to many types of cancer, and frequent NSD3 amplicons are found in cancer samples from, among other, breast, lung, head, and neck (Chen, Y. et al., 2014). Recently, the oncogenic potential regarding H3K36 has been further revealed, as discoveries that mutations of H3K36 can drive oncogenesis in chondroblastoma and undifferentiated sarcoma (Yang et al., 2016). Altogether, this indicates the importance for further elucidating the cellular functions of NSD3.

1.3 Epitranscriptomic gene regulation in embryonic stem cells

Modifications in the RNA, transcriptome, are known as epitranscriptomic regulation. These modifications are chemically diverse and together they regulate the RNA metabolism. The most abundant internal modification is N^6-methyladenosine (m^6A) methylation of mRNA, which regulates gene expression in many physiological processes and has a widespread regulatory mechanism. In addition, this modification is linked to cellular differentiation and cancer progression. This indicates the importance of epitranscriptomic gene regulation (Zaccara et al., 2019).

1.3.1 The most prevalent modification in mRNA – m^6A methylation

The presence of m^6A in mRNAs was first established in the 1970s and account for the most prevalent modifications of mRNAs in eukaryotes (Desrosiers et al., 1974; Zaccara et al., 2019) (Fig. 7). m^6A modifications also exist in RNA in organisms including viruses and yeast (Beemon & Keith, 1977; Bodi et al., 2010). In 1997, cloning of what's later shown to be the major enzyme synthesizing m^6A, the METTL3 enzyme, was performed and in the 2000s, and it was observed that m^6A modifications were essential for specific developmental processes (Bokar, 1997; Clancy et al., 2002; Zaccara et al., 2019; Zhong et al., 2008). m^6A modified mRNA often only contain a single m^6A site, with some exceptions where mRNAs contain 20 or more m^6A sites. The m^6A sites are described as constitutive, as they are very similarly distributed among diverse cell lines and different tissues (Schwartz et al., 2014). m^6A modifications tend to occur at the consensus motif DRACH; D = G, A or U; R = G or A; and H = A, C, or U (Zaccara et al., 2019). m^6A methylation happens co-transcriptionally, hence only nuclear mRNA and not cytosolic mRNA, undergoes this methylation (Huang et al., 2019).

In addition to mRNA, m^6A modifications also exist in other forms of RNA, including tRNA and rRNA (Iwanami & Brown, 1968; Meyer et al., 2012; Saneyoshi et al., 1969).

Multiple functions of m^6A in mRNA metabolism have been identified in studies performed *in vitro*, like processing in the nucleus, translation, and decay in the cytoplasm (Zhao et al., 2017). m^6A has been reported to regulate gene expression through mRNA splicing, degradation, and localization, in addition to having an important role in maintenance of the genome stability (Xiang et al., 2017; Xiao et al., 2016). Methylated transcripts have a significantly shorter half-life compared to unmethylated transcripts (Geula et al., 2015). In humans, dysregulation of m^6A modification has been linked to different diseases, including obesity and cancer (Sibbritt et al., 2013). Overall, m^6A modifications are involved in several aspects of RNA metabolism, including mRNA stability and translation efficiency.

In 2012, the first transcriptome-wide profile of m^6A localization in RNA was published. It was revealed that m^6A sites are enriched near stop codons, in 3´untranslated regions (3´UTRs), and in association with miRNA binding sites (Meyer et al., 2012). In the 3´UTRs of protein-coding genes, m^6A has been shown to mark unstable transcripts (Batista et al., 2014; Dominissini et al., 2012). In the absence of m^6A, a direct increase of mRNA stability of m^6A-containing transcripts is observed, and an indirect increase in the translation efficiency of GC-rich transcripts (Geula et al., 2015).

Further, it has been reported that miRNAs regulate m^6A modifications in mRNA. This occurs through a sequence pairing mechanism. When miRNA expression or sequences are manipulated, the binding of METTL3 to mRNA containing miRNA targeting sites is modulated. Finally, this leads to altered m^6A modification levels (Chen, T. et al., 2015).

1.3.2 m^6A life cycle

The m^6A modifications are installed, removed, and read by numerous of proteins (Fig. 7). m^6A methylation is installed during transcription in the nucleus by the methyltransferase complex, which comprises the N^6-adenosine core methyltransferase-like protein 3 and 14 (METTL3/14) and Wilms´ tumor 1-associating protein (WTAP) (Bokar, 1997; Ping et al., 2014; Wang, Y. et al., 2014; Zaccara et al., 2019). These adenosine methyltransferases are known as writers and add m^6A to mRNA co-transcriptionally (Wang, Y. et al., 2014; Zaccara et al., 2019). m^6A-erasers function as demethylases, where AlkB homolog 5 (ALBH5) is the main eraser that oxidatively removes m^6A modifications in nuclear RNA, mostly mRNA (Zheng et al., 2013). The fat mass and obesity-associated protein (FTO) is another protein revealed to erase m^6A

modifications (Jia et al., 2011). m^{6}A-binding proteins are known as the reader proteins, and it has been documented five YTH domain-containing proteins in mammals; YTHDC1-2 (DC1-2), localized in both nucleus and cytoplasm, and YTHDF1-3 (DF1-3), localized in the cytoplasm (Dominissini et al., 2012; Liao et al., 2018) (Fig. 7). The eukaryotic initiation factor 3 (eIF3) is another m^{6}A reader, also localized in the cytoplasm, while the heterogeneous nuclear ribonucleoproteins A2B1 (HNRNPA2B1) has been found to be a nuclear RNA-binding reader protein, that binds to m^{6}A-containing RNAs (Meyer et al., 2015; Zaccara et al., 2019). Overall, these proteins work together to fulfill the m^{6}A life cycle.

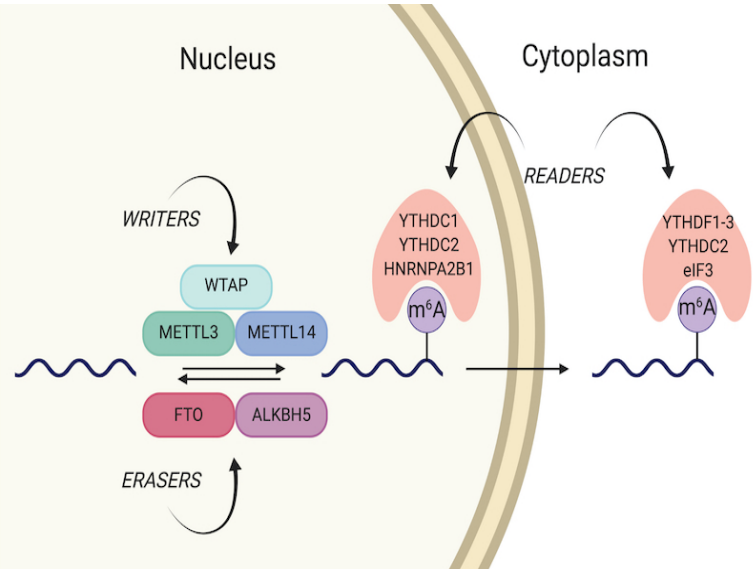


Figure 7. The m^{6}A life cycle.
A schematic overview of the m^{6}A life cycle. This life cycle starts with METTL3, and its adaptors, adding m^{6}A to the mRNA co-transcriptionally. The m^{6}A sites are targeted by the erasers ALKBH5 and FTO, which removes m^{6}A. When m^{6}A is located in the nucleus, it can be bound by DC2 or HNRNPA2B1, but m^{6}A is preferentially bound by DC1. Upon mRNA export to the cytoplasm, m^{6}A are bound by DF1/3, eIF3, and METTL3, which promote translation, while binding of DF2/3 mediate degradation.

1.3.3 m^{6}A writers – METTL3, METTL14, and WTAP

The first protein catalyzing m^{6}A methylations was isolated in the 1990s (Bokar et al., 1994; Bokar, 1997; Zaccara et al., 2019). There are two separate protein factors, introducing m^{6}A modifications, termed methyltransferase A and B (MT-A/B). MT-A is a smaller complex which comprises two complexes, METTL3, the catalytic subunit, and METTL14, the allosteric adapter of METTL3, which forms a stable heterodimer (Bokar, 1997; Liu et al., 2014; Ping et al., 2014; Zaccara et al., 2019). Both METTL3 and 14 consist of an MTase domain, which is a consensus fold in the methyltransferase family (Bujnicki et al., 2002). There is a sequence homology greater than 35% for the two MTase domains of METTL3 and 14 (Wang, Y. et al., 2014). MT-B is a larger complex and includes adaptor proteins that recruit METTL3 to RNA (Bokar et al., 1994; Bokar, 1997; Schöller et al., 2018). The installation of m^{6}A in mRNA can be possessed by both MT-A and MT-B. The regulatory subunit WTAP lacks methylation activity, but has been shown to be required for the formation of the core catalytic complex

METTL3/14, thereby affecting the deposition of cellular m^{6}A (Wang, Y. et al., 2014; Zaccara et al., 2019). All in all, METTL3/14 and WTAP are essential for installing m^{6}A modifications.

1.3.4 m^{6}A erasers – ALKBH5 and FTO

ALKBH5 and FTO act as erasers, by demethylating m^{6}A in RNA (Jia et al., 2011; Zheng et al., 2013). Both ALKBH5 and FTO belong to the 2-oxoglutarate-(2OG)-dependent oxygenase family (Aik et al., 2014; Han et al., 2010). The demethylation activity of ALKBH5 plays an important role in the export of mRNA and in RNA metabolism. ALKBH5 is found to localize with nuclear speckles, while FTO has been shown to shuttle between both the nucleus and cytoplasm (Gulati et al., 2014; Zheng et al., 2013). The substrate preference for these two enzymes is similar, but the reaction pathways appear to be different, additionally, it is reported that the preferred cellular substrate of FTO is m^{6}A$_m$, and not m^{6}A (Chen, W. et al., 2014; Mauer et al., 2017) (Fig. 8). It has been reported that m^{6}A is directly converted to adenosine, without any observed intermediates when removed by ALKBH5. Conversely, the two intermediates hm^{6}A and fm^{6}A can be observed when FTO mediate m^{6}A demethylation (Chen, W. et al., 2014).

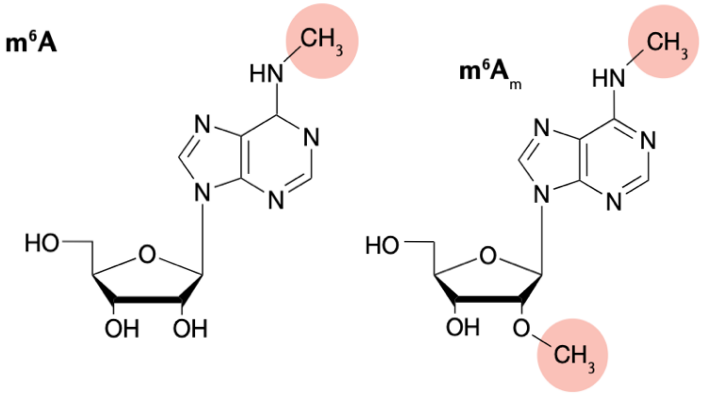


Figure 8. The structure of m^{6}A and m^{6}A$_m$. The difference between m^{6}A and m^{6}A$_m$ is the 2′-*O*-methyladenosine (A$_m$) modification. Figure from Zaccara et al., 2019.

1.3.5 m^{6}A readers – YTHDF1-3 and YTHDC1-2

The YTH domain was identified in 2002, and later it has been identified five YTH containing reader proteins (Dominissini et al., 2012; Stoilov et al., 2002). The YTH family proteins recognize RNA containing m^{6}A through their conserved YTH domain (Wang, X. et al., 2014). The YTH domain consists of ~150 amino acids that bind RNA in an m^{6}A-dependent manner (Li et al., 2014; Zhu et al., 2014) (Fig. 9). There are five YTH domain-containing reader proteins in the mammalian genome; DF1-3 and DC1-2 (Dominissini et al., 2012; Liao et al., 2018). Additionally, the nuclear RNA binding protein HNRNPA2B1 has also been categorized as a reader protein, thath regulates alternative splicing of nuclear transcripts, similar to the m^{6}A

writer METTL3 (Alarcón et al., 2015). Binding of HNRNPA2B1 is regulated by structural changes induced by m^{6}A and it will preferentially bind noncoding RNA. eIF3 is a cytosolic reader protein that recognize m^{6}A in the 5′UTR of mRNAs, and promotes, among other, DF1-independent translation (Zaccara et al., 2019). The DF1-3 reader proteins and DC1 have been shown to be ubiquitously expressed in mRNA, while DC2 is predominantly enriched in the testes (Hsu et al., 2017; Jain et al., 2018; Wojtas et al., 2017).

YTHDF1, YTHDF2, and YTHDF3

The YTHDF1-3 family are cytoplasmic reader proteins (Liao et al., 2018; Wang, X. et al., 2014). DF1-3 are three very similar paralogues and their sequences share high amino acid similarity over their entire lengths. The YTH-domain resists in the C-terminal, and the rest of their ~350 amino acids sequence length comprises a low complexity region, which includes several prion-like Q/N (glutamine/asparagine)- rich domains, but no modular protein domains (Patil et al., 2018) (Fig. 9).

These domains make DF1-3 undergo phase separation into liquid droplets within the cytosol, polymers, or gel (Zaccara et al., 2019). m^{6}A regulates the fate of cytosolic mRNA since they scaffold DF proteins, which leads to the formation of phase-separated DF-m^{6}A-mRNA complexes. Since polymethylated mRNA can scaffold more DF proteins than monomethylated mRNAs, this effect is seen particularly in phase separation (Ries et al., 2019).

DF1 interacts with the translation machinery and promotes protein synthesis of m^{6}A-containing mRNAs. eIF3 and ribosomes are interaction partners of DF1, and eIF3 binds to mRNA by m^{6}A sites in the 5′UTR, thus, promotes DF1-independent and cap-dependent translation. A knock-down of the *Df1* gene has been shown to reduce the translation efficiency of target transcripts, which supports the functional role of DF1 (Wang et al., 2015).

Several thousand m^{6}A-containing mRNAs are distributed in the cytoplasm by DF2. DF2 controls RNA decay, where it selectively binds to m^{6}A-methylated mRNA. The codon region and stop codon region in the 3′UTR are the targets of DF2. Knockdown of DF2 resulted in about 30% prolonged lifetime of mRNA targets, and these findings suggest that the DF2 protein might play a role in mRNA translation and/or stability and that m^{6}A will mark the exported mRNA for a shorter half-life in the cytosol (Wang, X. et al., 2014). The deadenylase complex CCR4-NOT has later been shown to be recruited by DF2, for initiation of re-adenylation and decay of m^{6}A-containing mRNAs (Du et al., 2016). This validates the suggestion made by Wang, X. et al., 2014, that DF2 affects mRNA stability and translation.

DF3 has been characterized as a partner of both DF1 and DF2, to control the metabolism of m^{6}A containing mRNAs in the cytoplasm. After the export of m^{6}A-containing mRNA from the nucleus to the cytoplasm, the translation might be facilitated by the DF3 in cooperation with DF1, which occurs through binding of m^{6}A-containing mRNAs and interactions with the ribosome subunits 40S and 60S (Li et al., 2017; Shi et al., 2017).

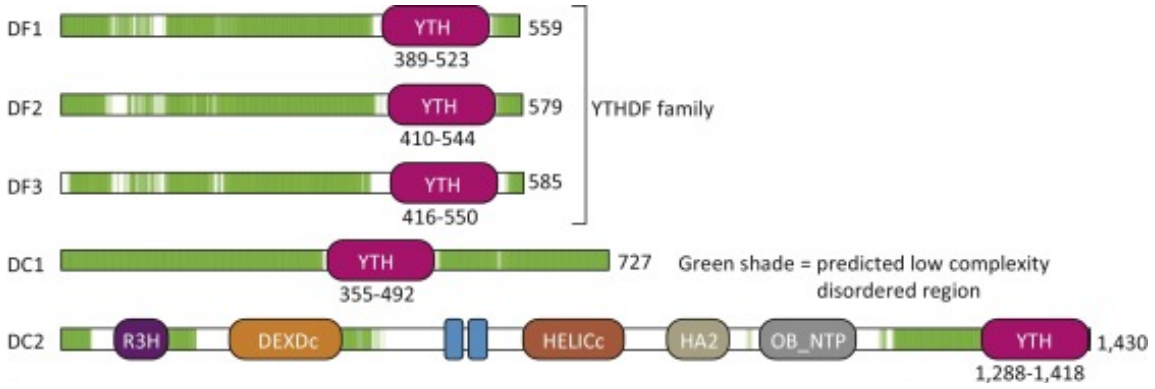


Figure 9. Domain structure of human YTH proteins.
Schematic figure of the DF1-3 and DC1-2 domain. The YTH domain is marked with pink, and as the figure show, DC1 have an internal YTH domain, while the DF family proteins and DC2 have a C-terminal YTH domain. The low-complexity disordered regions are marked with green. Figure from Patil et al., 2016.

The side chain of DF1 and DF2 are important for binding to m^{6}A sites in mRNA. The tryptophan (W) side chains of W411, W465, and W470 forms a positively charged pocket in DF1, where m^{6}A will be positioned. These tryptophan residues establish the methylation-dependent recognition (Liao et al., 2018; Xu et al., 2015). The hydrophobic pocket of DF2 consist of W432 and W486, which are important for recognizing m^{6}A. The binding affinity of the YTH domain will decrease if mutations occur in this aromatic cage, however, the binding affinity to RNA which is unmethylated will rarely be affected (Zhu et al., 2014).

A long pondered upon question is whether DF1-3 possesses redundant or different functions. For long DF1-3 were reported to have distinct functions on m^{6}A mRNA, where DF1 enhances translation, DF2 promotes degradation of m^{6}A-modified mRNA, and DF3 as a protein performing both functions (Shi et al., 2017; Wang et al., 2015; Wang, Y. et al., 2014). Later, it has been reported that DF1-3 have similar functions, where they recruit CCR4-NOT to m^{6}A modified mRNAs (Du et al., 2016). During deadenylation assays, it has been observed that all three proteins show similar roles in mRNA degradation (Du et al., 2016; Kennedy et al., 2016). Some studies propose that m^{6}A residues can be bound by all three DF paralogues, while other suggests that they only bind one of the three DF proteins. Their highly similar sequences support that the DF proteins have similar functions, but since this is unclear, it requires further investigation on how DF proteins bind m^{6}A sites in mRNA and their specific functions (Patil et al., 2016; Zaccara et al., 2019).

YTHDC1 and YTHDC2

The DC1 protein family is located in the nucleus, and mediate the export of methylated mRNA, affects mRNA splicing, and epigenetic silencing (Patil et al., 2016; Roundtree et al., 2017; Xiao et al., 2016). It is thought that DC1 functions as an m^6A reader by recruiting pre-mRNA splicing factors to the target RNA elements, thereby disposing dissimilar outcomes of the target mRNA splicing (Xiao et al., 2016). In general, studies have revealed that DC1 preferentially binds m^6A in noncoding RNA, like the noncoding RNA X-inactive specific transcript (XIST). XIST is involved in transcriptional inactivation of genes on the X-chromosome, and it is reported that DC1 recognizes m^6A modifications on XIST, in addition to being required for its function. (Patil et al., 2016). So far, it is unclear if functions of other noncoding RNAs are affected by DC1. As opposed to DC1, DC2 exists in both the nucleus and cytosol of cells.

Unlike the other YTH-domain family proteins, DC2 binds m^6A with a much lower affinity, and it is possible that DC2 uses another binding method (Zaccara et al., 2019). Studies indicate that DC2 primarily functions in spermatogenesis, and DC2 knock-out mice show defects in spermatogenesis, without other developmental defects (Hsu et al., 2017; Wojtas et al., 2017; Xu et al., 2015). In addition to the highly conserved YTH domain, DC2 contains a helicase domain, like DHX29, which indicates a function in mRNA translation (Dhote et al., 2012; Zaccara et al., 2019). Other researchers have reported that DC2 mediate mRNA degradation through recruitment of the exoribonuclease Xrn1 (Kretschmer et al., 2018; Wojtas et al., 2017). Nevertheless, many of these revealed effects are very small, indicating that the function of DC2 has not been fully elucidated yet (Zaccara et al., 2019).

1.4 R-loop biology

R-loops are formed when the template DNA strand anneals with the nascent RNA resulting in a displacement of the non-template DNA strand, forming a three-stranded structure (Crossley et al., 2019; Thomas et al., 1976) (Fig. 10A). The DNA-RNA hybrid and the displaced ssDNA are termed R-loop, but commonly both the terms “DNA-RNA hybrid” and “R-loops” refers to this three-stranded structure (García-Muse & Aguilera, 2019). R-loops are formed during transcription behind the elongating RNA polymerase and consists of 100 to 2,000 base pairs (Santos-Pereira & Aguilera, 2015). For a long time, it was believed that R-loops were byproducts of transcription, with little or no impact on the cells, however, this has been disproved during the last decade (Aguilera & García-Muse, 2012).

R-loops are made in a transient and reversible manner, but R-loops may also act as epigenetic marks and affect longer-lived changes in chromatin state, read by chromatin remodelers and other proteins. This is highly supported by R-loops presence across the genome, as well as R-loops being formed at a much higher frequency than first anticipated (Chédin, 2016). It has also been revealed that R-loops can be a great threat to genome integrity (Aguilera & García-Muse, 2012). Therefore, R-loops have sparked a new and interesting research field within, among other, cancer research.

R-loops display two different roles, physiological and pathological roles (García-Muse & Aguilera, 2019). Programmed processes that require specific factors form physiological R-loops, while R-loops occurring accidentally in a non-scheduled manner are termed pathological. It still remains to investigate what distinguishes physiological from pathological R-loops. To understand the role of R-loops it is important to have insight into how often unscheduled R-loops form along the genome, how cells protect themselves against the pathological R-loops, and how R-loops affect chromatin integrity, thus affecting genome integrity (García-Muse & Aguilera, 2019).

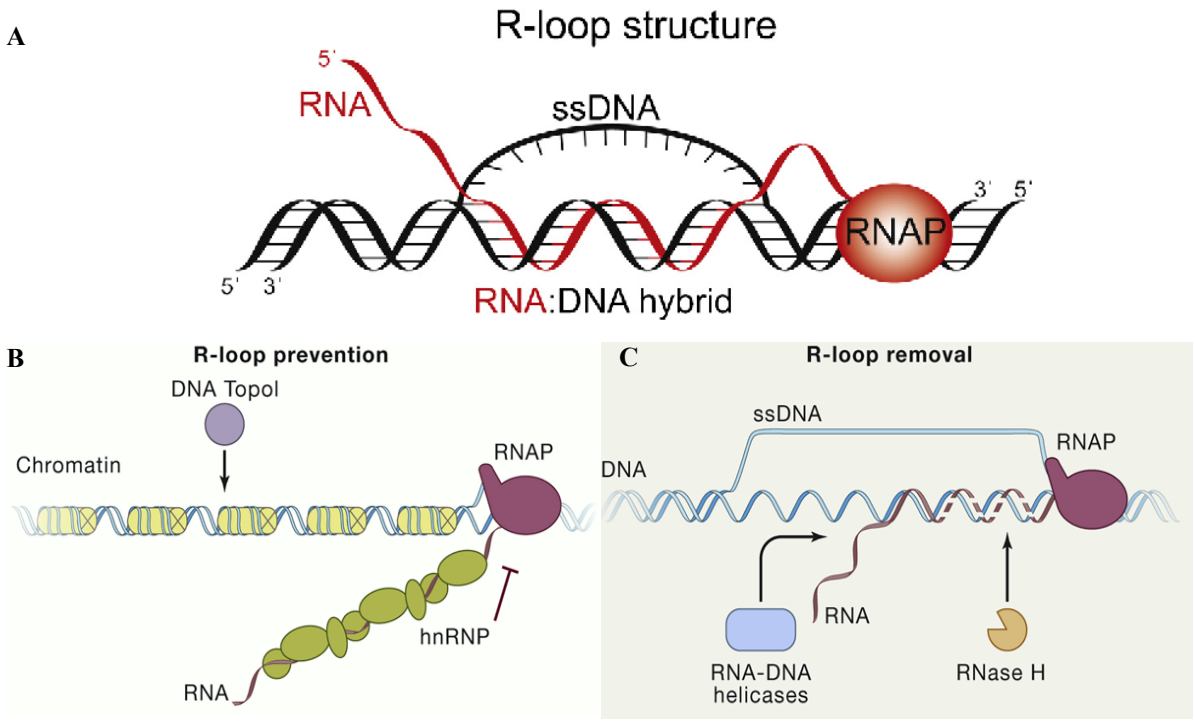


Figure 10. R-loop structure, prevention, and removal.
(A) R-loop structure: the DNA-RNA hybrid and the displaced ssDNA. Figure is from Hamperl & Gimprich, 2014.
(B) R-loops are prevented by RNA-binding proteins involved in RNA biogenesis (hnRNP), chromatin, and topoisomerase 1 (Topo I), which resolve negative supercoiling behind RNAPII during elongation.
(C) The RNase H enzyme degrades the RNA strand in the R-loop, which results in removal of DNA-RNA hybrids. DNA-RNA helicases are able to unwind R-loops and can remove these three-stranded structures.
Figure B/C from García-Muse & Aguilera, 2019.

In addition to being formed during transcription, R-loops can cause DNA damage in some contexts, and therefore provides a possible link between the two hallmarks of cancer: DNA damage and mutagenesis (Crossley et al., 2019; Hamperl & Cimprich, 2014). R-loops are found to accumulate in, among other, breast cancer cells, and to drive DNA damage induced by estrogen (Crossley et al., 2019; Stork et al., 2016). Moreover, R-loops have been related to different diseases, including autoimmune and neurological diseases, highlighting the importance of investigating the cellular function of R-loops and its interaction partners (Crossley et al., 2019).

1.4.1 The role of R-loops in cellular processes

R-loops have been detected all over the genome of yeast, bacterial, and eukaryotic cells and enhanced R-loop formation has been shown in regions where they play a physiological function (Crossley et al., 2019; Lang et al., 2017; Stork et al., 2016). Such regions include CRISPR/Cas9 activity regions, in which the guide-RNA forms a DNA-RNA hybrid to identify the Cas9 cleavage target, in addition to regions for initiation of DNA replications in mitochondrial DNA or bacterial plasmids (Aguilera & García-Muse, 2012; Jinek et al., 2012). R-loops are also enriched at unmethylated CpG-island promoters and termination sites, where they enhance activation. On the other hand, R-loops are also linked to histone marks of active transcription, such as mono- and trimethylation of H3K4 and acetylation of H3 (Crossley et al., 2019; Skourti-Stathaki et al., 2019). Furthermore, R-loops can affect genome function and integrity by affecting cellular processes like DNA transcription, replication, and repair, in addition to being a source of DNA damage. Therefore, mechanisms to prevent or resolve R-loops have been developed in the cells, to regulate R-loop formation and prevent R-loops from becoming a source of cellular pathologies, by threatening cell proliferation and genome integrity (García-Muse & Aguilera, 2019) (Fig. 10B).

Helicases, nucleases, and signaling pathways are some factors that have been proposed to prevent or resolve R-loops. It is still not clear whether these factors work in a direct or indirect manner (Crossley et al., 2019). Interactome studies may distinguish factors acting directly at R-loops, from those factors that act indirectly. The discovery of the highly specific DNA-RNA hybrid monoclonal antibody S9.6, and that R-loops are degraded by RNase H, has the last decade resulted in highly improved research regarding R-loop biology (García-Muse & Aguilera, 2019) (Fig. 10C). This has recently resulted in novel R-loop interactome studies, showing the interaction partners of R-loops.

Cristini et al., 2018, identified 469 candidate factors enriched in the DNA-RNA hybrid interactome (Cristini et al., 2018). Helicases acting on RNA, like DDX5 and DHX9, and DNA, like MCM3, were found to have a high enrichment of binding domains for R-loops (Cristini et al., 2018; García-Muse & Aguilera, 2019). Overall, the identified factors almost exclusively contained nuclear proteins.

R-loops in transcription

R-loops are enriched at both gene termination and promoter regions, which supports the evidence of R-loops playing a role in gene expression (García-Muse & Aguilera, 2019). Also, R-loops are enriched over unmethylated CpG-islands, and experiments have suggested that these R-loops will protect such regions against DNA methylation since DNA methyltransferases bind poorly to DNA-RNA hybrids. Thus, R-loops facilitate transcription by suppressing methylation-associated silencing (Ginno et al., 2012; Grunseich et al., 2018). Furthermore, a connection between R-loops at promoters and chromatin-remodeling has been reported, since R-loops can both inhibit and promote chromatin remodelers. However, it is unknown what distinguishes R-loops that inhibit protein binding from those who promote it. By recruitment of activated chromatin-remodeling complexes, and inhibition of repressive chromatin-modifying enzymes, R-loops facilitate a poised chromatin state and promote differentiation genes, in mES cells (Chen, P. B. et al., 2015).

When RNA polymerases are stacked in front of DNA damage sites, due to roadblocks caused by R-loops, transcription-coupled repair is triggered (García-Muse & Aguilera, 2019). Transcription-coupled repair is defined as the preferential repair of DNA lesion of the transcribed strand of actively transcribed genes (Strick & Portman, 2019). R-loop-transcription conflicts could be removed by transcription-coupled repair-like mechanisms for transcription to resume, but so far the existence of such a function, or transcription-coupled repair-like mechanisms have not been reported (García-Muse & Aguilera, 2019).

Additionally, R-loops are linked to transcription termination. This role in termination was proposed due to R-loop enrichment at the 3´end of some mammalian genes, especially over G-rich terminator elements. Furthermore, studies have shown that R-loops are linked to termination by facilitating efficient transcription termination (Skourti-Stathaki et al., 2011; Skourti-Stathaki et al., 2014; Skourti-Stathaki & Proudfoot, 2014). RNAPII is hindered by R-loops downstream of the polyadenylation sequence, further, the nascent RNA is degraded by exonucleases when RNA helicases, like DHX9 resolve the R-loops. This finally leads to transcriptional termination. Experiments have shown that loss of DHX9 results in the

accumulation of R-loops and defective termination, supporting the aforementioned statements (Cristini et al., 2018). Overall, R-loops have been shown to play an important role in transcription termination, but if the turnover of R-loops is deregulated it might lead to DNA damage, which is strongly associated with a block of replication fork progression (Crossley et al., 2019; Sollier & Cimprich, 2015).

R-loops in replication

If the replication fork is stalled it needs to be protected to avoid collapsing, otherwise, it will result in DNA breaks. R-loops are likely to cause genome instability in cycling S-G2 cells since it stalls the replication fork (García-Muse & Aguilera, 2019). Lethal DNA damage can be caused by transcription-replication collisions (TRC), which is possible since the DNA template is common during transcription and replication. TRCs occur when the replication fork faces the transcription machinery, and R-loops are associated with a higher occurrence of TRCs (Crossley et al., 2019). Overexpression of RNase H reduces DNA damage and causes the replication fork to slow down under hormone or oncogene-induced replication stress, supporting the role of R-loops in replication (Kotsantis et al., 2016). Several replication fork pathway factors suppress R-loops, thereby preventing damage from transcription-replication collisions (García-Muse & Aguilera, 2019).

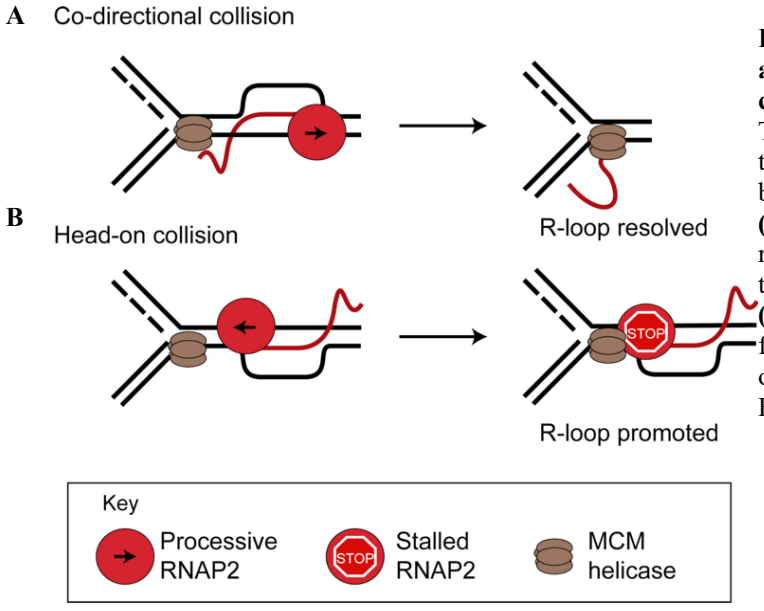


Figure 11. R-loops in co-directionally and head-on transcription-replication collisions.
The DNA helicase, MCM, has the ability to unwind R-loops during CD orientation, but not in HO orientation.
(A) Co-directional collision of the replication fork with R-loop can resolve the R-loop.
(B) Head-on collision of the replication fork with R-loops can result in promotion of R-loop formation and stabilization.
Figure from Crossley et al., 2019.

R-loop levels are highly affected by the orientation of the TRCs (Crossley et al., 2019). Due to the movement of the replication machinery, the transcription machinery either moves on lagging strand genes, head-on (HO), or on leading strand genes, co-directionally (CD). This

direction depends on the coding strand of the given gene, as observed in the engineered bacterial genome and the mammalian episomal system (Hamperl et al., 2017; Lang et al., 2017) (Fig. 11). When the replication fork encounters R-loops in the HO orientation, R-loops have been shown to completely block replication, inhibit gene expression, and elevate mutagenesis (Fig. 11A). The effects are more tolerable when replication and transcription occur CD (Crossley et al., 2019; Hamperl et al., 2017; Lang et al., 2017) (Fig. 11B). Hence, it has been demonstrated that transcription in HO TRCs promotes R-loop formation, while on the other side, CD TRCs reduces R-loop levels, since the R-loop then can be resolved (Crossley et al., 2019; Lang et al., 2017).

R-loops in DNA-damage repair

If R-loops occur unscheduled, it will give rise to DNA damage and ultimately genome instability (García-Muse & Aguilera, 2019). The ssDNA fiber of R-loops, is the most vulnerable part of the action of genotoxics and nucleases (García-Muse & Aguilera, 2019; Lindahl, 1993). R-loops can cause DNA damage, but studies have also shown that R-loops formation may take place after DNA damage. Therefore, R-loops are widely considered to have a dual function in the DNA damage response, by both promoting and inhibiting DNA repair (Crossley et al., 2019).

It is reported that single-stranded DNA breaks (SSBs) or double-stranded DNA breaks (DSBs) functions as a driving force in R-loop formation (Aguilera & García-Muse, 2012; Aguilera & Gómez-González, 2017). Transcription at resected DSBs is required for efficient DSB repair, followed by the formation of R-loops and subsequently RNase H removal (García-Muse & Aguilera, 2019). Therefore, R-loops can promote DSB repair, since higher levels of RNAPII are observed at DSB with R-loop accumulation (Ohle et al., 2016).

There are multiple ways that persistent R-loops compromise DNA repair. R-loops can affect DNA repair by disturbing the chromatin structure flanking DSBs, and hamper the binding of DNA repair factors to DSBs (Aguilera & Gómez-González, 2017; Cohen et al., 2018). R-loops influence DSB repair to alter efficiency or resection, which determines whether the repair will be proceeded by non-homologous end joining (NHEJ) or homologous recombination (HR) (Crossley et al., 2019). Removal of R-loops has been shown to reduce the efficiency of the two main DSB repair pathways, NHEJ and HR (Lu et al., 2018). Overall, both SSBs and DSBs favor R-loop formation, but it requires extensive research to understand whether R-loops function in repair as an intermediate, or as a secondary consequence of breaks arising at DNA undergoing transcription (García-Muse & Aguilera, 2019).

1.5 Embryonic stem cells as a model system

Mouse embryonic stem cells are easy to handle, and can easily be clonally expanded and become genetically modified while maintaining pluripotency *in vitro*. Additionally, they have the ability to differentiate into the three germ layers: ectoderm, endoderm, and mesoderm. Altogether, these properties emphasize ES cells as a great model system for studying epigenetic and epitranscriptomic changes (Niwa, 2010; Yagi et al., 2017).

1.5.1 Differentiation assays of mouse embryonic stem cells

There are numerous methods to induce differentiation of ES cells. In this thesis, three different protocols for differentiation have been tested: differentiation against neuronal stem cells, differentiation upon induction of all-trans retinoic acid (ATRA), and embryoid body (EB).

NSC differentiation was included because of the group's interest in the neuronal cell lineage. NSC differentiation was induced by LIF withdrawal since mES cells have a tendency towards differentiating to the neuronal lineage upon LIF withdrawal (Ying et al., 2003). Further, there will be a forced differentiation when EGF and FGFb are added, which promotes differentiation toward an ectoderm lineage. Cells will lose their differentiated phenotype, and only cells differentiating toward the ectoderm lineage will survive.

ATRA is an active form of vitamin A, that regulates gene expression through the retinoic acid receptor. ATRA is made in the body and helps cells grow and develop, especially in the embryo. When added to ES cells *in vitro* ATRA induces cells to differentiate. ATRA differentiation will induce neuronal differentiation, as well as other lineages, like towards adipocytes and ventricular cardiomyocytes (Simandi et al., 2010).

EBs are three-dimensional aggregates of pluripotent cells that were cultured in hanging drops and induced to differentiate upon induction of FBS. This method makes it possible to study complex developmental processes, since EBs represent cells in an embryo-like state, thus comprising a combination of all three developmental lineages. The disadvantages of the hanging drop method is that media exchange is impossible and microscopy examination of the forming EBs is difficult during cultivation. It is also time-consuming to handle the EBs, compared to the cells during NSC and ATRA protocols.

1.5.2 Genome editing methods for generation of knock-outs

creERT2 recombinase and CRISPR/Cas9 are two genome editing methods which can be used to generate knock-outs in mES cells. CreERT2 is a second generation of a cyclization recombinase (Cre) fused with a mutant form of a estrogen receptor (ER) (Feil et al., 1997; Kristianto et al., 2017). CreERT2 is activated by a nanomolar concentration of 4-hydroxytamoxifen, later referred to as tamoxifen. Cre function by binding to loxP/loxP sites in the genome, and can cause knock-outs of target genes. Cells that are homozygote for the loxP/loxP target allele can be transduced with lentivirus with an inserted open reading frame (ORF) for creERT2. Further, when these creERT2 cells are treated with tamoxifen, the heat-shock protein 90 (HSP90) is removed from the creERT2 complex (Fig. 12A). The creERT2 complex translocates from the cytoplasm to the nucleus, and leads to a knock-out of the target gene by recombination, when the loxP-sites are directed the same way (Feil et al., 1997; Kristianto et al., 2017).

Furthermore, CRISPR/Cas9 genome editing can be used for generation of knock-outs. CRISPR/Cas9 single- guide RNA (sgRNA) pairs are complementary to the target sequence directly before the protospacer adjacent motif (PAM) and guide the Cas9 nuclease to the target (Fig. 12B). The Cas9-sgRNA complex simultaneously mediate blunt-ended DSBs at both strands of the target locus, and the DNA repair, NHEJ, will result in a knock-out (Jinek et al., 2012; Tian et al., 2017).

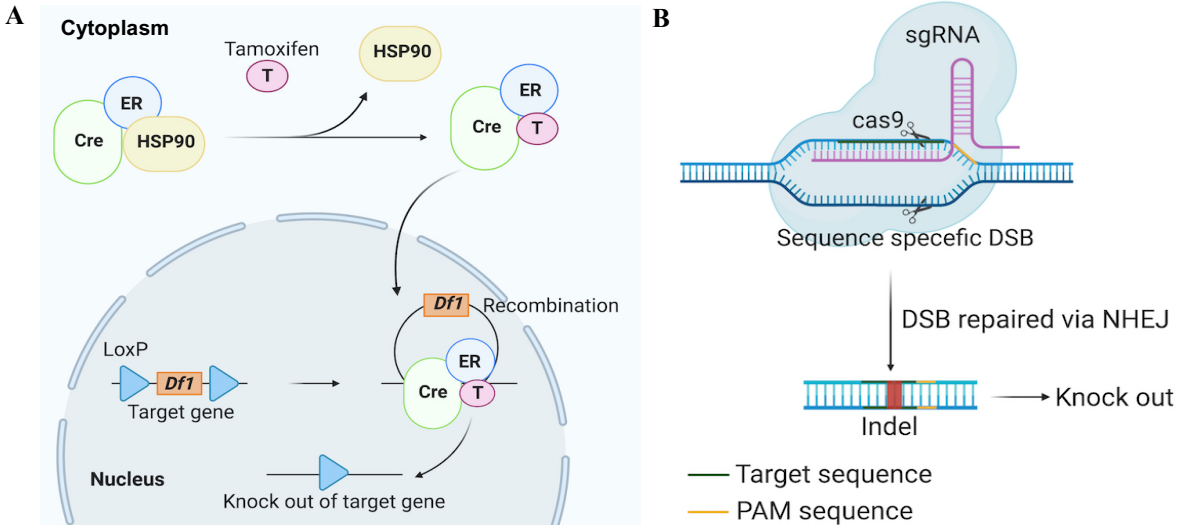


Figure 12. Schematic overview of methods for generation of knock-outs by creERT2 recombination and CRISPR/Cas9 technology.
(A) Tamoxifen induced recombination to generate KO in loxP/loxP cells. When the loxP/loxP cells are treated with tamoxifen, the heat-shock protein 90 (HSP90) will be separated from the creERT2 complex. Further, this allows the creERT2 complex to translocate from the cytoplasm to the nucleus. Since the loxP sites are going in the same direction, it results in a knock-out of the target gene by recombination.
(B) CRISPR/Cas9 induced knock-out a target sequence. The Cas9 nuclease induces double-stranded breaks, specified by target specific single-guide RNAs (sgRNAs). This DSB will then be repaired by the non-homologous end joining repair (NHEJ) pathway, and a site-specific insertion or deletion (indel) will result in a knock-out.

2. Aims of the study

The overall goal of this study was to investigate the contribution of epigenetic and epitranscriptomic modifications in gene regulation of mouse embryonic stem cells. We wanted to achieve this by approaching two different aspects: how the reader proteins, YTHDF1-3, impact cellular processes in embryonic stem cells (I), and the role of a chromatin modifier, MMSET in R-loop biology (II).

I. The first aim of this study was to generate single-, double-, and triple *Ythdf1-3* knock-outs in mES cells, and subsequently characterize their phenotype. We expected that phenotype characterization of the *Ythdf1-3* knock-outs would provide further insight into how these genes affect cellular processes. The study therefore aimed to:
    - Generate the knock-outs for *Ythdf1-3,* individual and in combinations, using creERT2 recombinase and CRISPR/Cas9 technology.
    - Analyze the differentiation phenotype upon deletion of *Ythdf1-3* by differentiation assays.
    - Investigate the proliferation capacity of the *Ythdf1-3* knock-out cells by proliferation assay.

II. The second aim of this study was to characterize the role of MMSET in R-loop biology. We wanted to investigate MMSET as an interaction partner with R-loops, and their overlapping interaction partners, in addition to MMSETs role during differentiation. The study therefore aimed to:
    - Validate the role of MMSET in pluripotency exit and during differentiation, by carrying out differentiation assays.
    - Ascertain the RNA helicases DHX9, DDX3, and DDX5, as interaction partners of MMSET by immunoprecipitation with MMSET antibody.
    - Validate DHX9, DDX3, and DDX5 as interaction partners of R-loops, by R-loop immunoprecipitation with the S9.6 antibody, both in presence and absence of MMSET. At the same time, confirm MMSET as an overlapping interaction partner of R-loops.
    - Analyze R-loop accumulation in heterozygous MMSET and MMSET null cells with immunofluorescence.
    - Perform R-loop immunoprecipitation qPCR to confirm R-loop enrichment at PcG repressed mesendodermal genes, in addition to elucidate possible effects upon MMSET deletion.

3. Materials and methods

3.1 Cell lines and culture conditions

Cell lines

E14 is one of the most standard mES cell lines and it was present in the laboratory before we arrived. The heterozygous ($MMSET^{+/-}$) and MMSET null ($MMSET^{-/-}$) mES cells were acquired from Kaneda lab, as described in Nimuera et al., 2009. MMSET-YFP-Ty1 cells, termed MM10, are cells derived from E14 and they stably express BAC-tagged MMSET with YFP-Ty1, to ensure a near endogenous expression of the tagged MMSET protein. This line was generated earlier before we arrived in the laboratory, and we confirmed through western blotting (WB) and immunoprecipitation (IP) that they were as advertised. 35.6 (wild type) and 37.4 loxP/loxP are mES cells derived by the transgenic core facility, Danstem, Copenhagen. 35.6 indicates the 6^{th} ES cell line from the mouse number 35 which carries the wild type alleles for YTHDF1 and 37.4 indicates the 4^{th} ES cell line from the mouse number 37 which was homozygous for YTHDF1 loxP/loxP alleles. These ES cells were verified by PCR using the primers in supplemental table S3 that they were as advertised. Wild type (35.6) will provide a band at 171 bp and 37.4 loxP/loxP a band at 344 bp.

Cell culture conditions

All media compositions are listed in supplementary "7.1 Media compositions".

Coating for embryonic stem cell culturing

Dishes were coated with 0.1% gelatin (Sigma) in 1x phosphate-buffered saline (PBS) (Oslo University Hospital) for at least 30 minutes at 37°C, according to table 1.

Coating for neuronal stem cell culturing

Dishes were coated according to table 1, with 500 ng/mL poly-D-lysine (PDL) (Merckmillipore A-003-E) in 1x PBS and incubated for at least 1 hour. PDL was aspirated and replaced with 200 ng/mL laminin (RD systems) in 1x PBS, and incubated for a minimum of 3 hours, to a maximum of 5 days. Incubations were performed at 37°C.

Table 1. Volumes and cell density for culturing ES cells in different plates sizes. K, $\times 10^3$, M, $\times 10^6$.

Size	To culture cells	To coat	TrypLE	PBS+TI	Cells seeded
Tray	100 (80-120) mL	50 mL	8 mL	40 mL	10-20 M
15 cm	20 (16-25) mL	10 (8-12) mL	1.5 mL	8-12 mL	3 (2-10) M
10 cm	10 (7-12) mL	5 (4-6) mL	1 mL	4-9 mL	2.5 (1-5) M
6 cm	4 (3-5) mL	2 (1.5-3) mL	350 µL	1-4 mL	0.5 (0.2-1) M
6-well	2 (1.5-3) mL	1 (0.8-1.5) mL	200 µL	1 mL	200K (50K-500K)
12-well	1 (0.5-2) mL	0.5 (1) mL	100 µL	1 mL	100K (20K-200K)
24-well	1 (0.5-1.2) mL	0.5 (0.3-1) mL	50 µL	1 mL	50K (10K-100K)

Passage of cells

All cells were cultured in tissue culture dishes (Falcon, A Corning Brand) and incubated in humidified incubators at 37°C and 5% CO_2. All mES cells were generally split every second or third day when grown in ES cell media. Cells were split by gently aspirating media and adding appropriate amount of TrypLE™ (Gibco) to the dish (Table 1). To decompact and detach the cells, the dish was tapped, and if necessary, incubated at 37°C for ~30 seconds. Appropriate volume, according to table 1, of PBS+TI (500 mL 1x PBS and 100 mg Trypsin inhibitor (Glycine max soybean, Sigma)) was added and pipetted up and down to obtain detached and dispersed cells, without disturbing the bottom of the dish. The cell suspension was centrifuged to a pellet at 1,200 rpm for 3-5 minutes, before the supernatant was aspirated gently. Next, the pellet was resuspended in 1-3 mL media by pipetting up and down to create single cell suspension. By counting with Countess II FL (Invitrogen), the appropriate number of cells were transferred to a laminin- or gelatin-coated dish containing room tempered media (table 1).

3.2 Transducing with lentivirus

mES cells carrying the loxP/loxP *Df1* allele (37.4 loxP/loxP) were transfected with lentivirus for generation of 37.4-creERT2. 37.4-creERT cells were used as base cells for generation of all *Ythdf1-3* knock-out cells. P2a-tagged lentivirus was made by VectorBuilder and had an inserted

ORF for creERT2, in addition to an enhanced green fluorescent protein (EGFP), and a hygromycin resistance gene which allows cells to be resistant to hygromycin B (Fig. 13).

Vector ID	VB180602-1231mat
Vector Name	pLV[Exp]-Hygro-EF1A>CreERT2(ns):P2A:EGFP
Date Created (Pacific Time)	2018-06-01
Vector Size	11687 bp
Viral Genome Size	8212 bp
Vector Type	Mammalian Gene Expression Lentiviral Vector
Inserted Promoter	EF1A
Inserted ORF	CreERT2(ns), EGFP
Inserted Linker	P2A
Inserted Marker	Hygro
Plasmid Copy Number	High
Antibiotic Resistance	Ampicillin
Cloning Host	Stbl3 (or alternative strain)

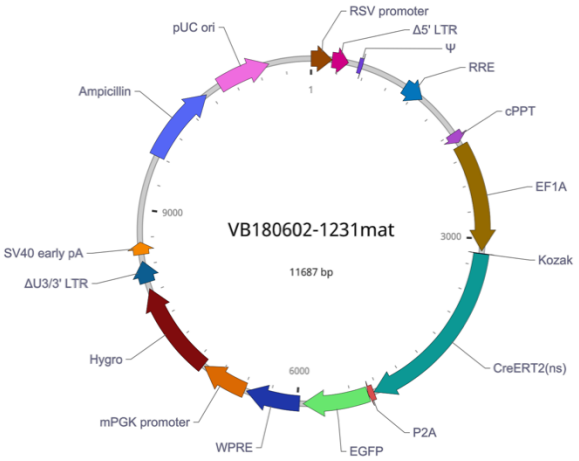


Figure 13. Information about the vector used to transduce cells for generation of the 37.4-creERT2 mES cells.
(A) A table showing vector information.
(B) Vector map of pLV[Exp]-Hygro-EF1A>CreERT2(ns):P2A:EGFP.

Lentiviruses with inserted ORF for creERT2 were thawed. In a 6 cm dish, 0.5 million 37.4 loxP/loxP mES cells were seeded in 2 mL ES cell media. 8 µL of polybrene (at 8 mg/mL concentration) and 2 mL of the virus was added to the cells. The next day, 2 mL ES cell media was added to the plate. At day 3 (d3) and d6, cells were split and 2 million cells seeded in a 15 cm dish with ES cell media. At d10 cells were sorted for GFP-positive cells by fluorescence-activated cell sorting (FACS), and the GFP-positive cells were seeded in a 6 cm dish.

Genotyping of 37.4-creERT2 cells

All buffers used for genotyping are listed in supplementary "7.2.1 Genotyping".

Genotyping was performed on the 37.4-creERT2 cells to validate the transduction with lentivirus. DNA was extracted from ~1-2 million cells by adding 75 µL alkaline lysis reagent at 95°C for 20 minutes, followed by a cool down to 4°C, before adding 75 µL neutralization buffer. The Master Mix for genotyping consisted of 1.3 µL Cre-F and Cre-R primers (Table S3: List of primers), 34 µL H_2O, 11.1 µL 5x MyTaq Reaction buffer (Bioline), 0.3 µL MyTaq™ DNA Polymerase (Bioline), and 2 µL of the extracted DNA, to a final volume of 50 µL in each PCR tube.

The following FOX1J PCR program was used: initial denaturation at 94°C for 3 minutes, followed by 35 cycles of denaturation at 94°C for 30 seconds, annealing at 60°C for

40 seconds, and extension at 72°C for 50 seconds, ending with a completion step at 72°C for 5 minutes.

The PCR product with a final concentration of 5% glycerol (Oslo University Hospital) was loaded on a 1.25% TAE agarose gel (Invitrogen) containing 1:20,000 SYBR Safe DNA stain (Invitrogen). Gene Ruler DNA ladder mix (Thermo Scientific) was applied as ladder. The gel ran at 90V for 30-40 minutes, and was imaged using ChemiDoc™ MP system (Bio-Rad). 37.4-creERT2 cells provides a band at a size of 327 bp.

3.3 Generation and validation of knock-outs

Generation of Ythdf1 knock-out with tamoxifen treatment

4-hydroxytamoxifen (4-OHT) (Sigma) was used to mediate creERT2 knock-out of *Df1* in 37.4-creERT2 mES cells. 2 million cells were seeded in a 10 cm dish with 12 mL ES cell media and treated with 0.5 μM 4-OHT. After two days, the cells were passaged, and again treated with 0.5 μM 4-OHT in 12 mL ES cell media. At d4 the cells were washed with 1x PBS before fresh ES cell media was added. After 2-3 days, the cells were sorted single as cells in a 96-well plate by FACS. Genotyping and western blotting were performed to validate the knock-outs, as described in section 3.3 and 3.4.

Generation of Ythdf2 or Ythdf3 knock-outs by CRISPR/Cas9 genome editing

CRISPR/Cas9 technology was used to generate *Df2* and *Df3* KOs, by using sgRNAs specific for each gene. Furthermore, this experiment was carried out using 37.4-creERT2 mES cells. In the morning, 1 million cells were seeded to a final volume of 4 mL ES cell media in a 6 cm dish. The sgRNAs targeting *Ythdf2* (DLP4 and DLP5) and *Ythdf3* (DLP7 and DLP8) were designed and cloned earlier, and their sequences are in the supplemental Table S5. For targeting *Ythdf2*, 0.5 μg of Blast-Cas9 and 1 μg of the sgRNAs DLP5 and DLP4 were added to 200 μL Opti-MEM® I 1X (Gibco) (tube A1). For targeting *Ythdf3*, 2 μg of Blast-Cas9 and 4 μg of the sgRNAs DLP7 and DLP8 were added to 200 μL Opti-MEM® I 1X (Gibco) (tube A2). In a separate tube, 8 μL lipofectamine 2000 (Invitrogen) and 200 μL Opti-MEM® I 1X (Gibco) were mixed (tube B). 200 μL from tube B was mixed with 200 μL from tube A1 or A2 and incubated for ~25 minutes at 37°C. After incubation, 400 μL of this mix (A1/A2 + B) was added on top of the cells, and 6-8 hours later, cells were washed with 1x PBS and fresh ES cell media was added. After two days of recovering, the cells were sorted as single cells in a 96-well plate by FACS. 10-14 days later, colonies were picked up and transferred to a 12-well

plate. Western blotting was performed to validate the generated knock-outs, as described in section 3.4.

Genotyping of Ythdf1 knock-out cells

gDNA extraction was carried out as described for genotyping of 37.4-creERT2 cells, see section 3.3. The Master Mix (see below) was made to a total of 48 µL per reaction, and 2 µL of the extracted gDNA was added, to a final volume of 50 µL in each PCR tube. The following program for PCR was used: initial denaturation at 94°C for 2 minutes, followed by 35 cycles of denaturation at 94°C for 30 seconds, annealing at 65°C for 30 seconds, and extension at 68°C for 1 minutes, ending with a completion step at 68°C for 8 minutes.

The PCR product with a final concentration of 5% glycerol (Oslo University Hospital) was directly ran on a 1.25% TAE agarose gel (Invitrogen) containing 1:20,000 SYBR Safe DNA stain (Invitrogen). Gene Ruler DNA ladder mix (Thermo Scientific) was applied as ladder. The gel ran at 90V for 30-40 minutes. The gel was imaged using ChemiDoc™ MP system (Bio-Rad).

Master Mix for genotyping of Ythdf1 knock-out cells

The Master Mix for genotyping consisted of the following components; 34 µL H_2O, 11.1 µL 5x MyTaq Reaction buffer (Bioline), 0.27 µM forward and reverse primer (182262cre-KLU13 and 182263cre-KLU13) (Table S3: Primers for genotyping) and 0.3 µL MyTaq™ DNA Polymerase (Bioline).

3.4 Western blotting

All buffer compositions are listed in supplementary "7.2.2 Western blotting".

Pellets were resuspended with 3-4 times pellet size lysis buffer containing 1 mM dithiothreitol (DTT) (Saween & Werner), 1 µg aprotinin (AP) (Sigma-Aldrich), 0.5 µg leuprotein (LP) (Sigma-Aldrich) and 2 mM phenylmethylsulfonyl fluoride (PMSF) (Sigma-Aldrich) per mL. Cell lysis was performed for 10 minutes on ice, before centrifugation for 20-30 minutes, 20,000 g at 4°C. The protein-containing supernatant was sonicated for seven cycles at 4°C for 30 seconds ON/OFF (Diagenode Bioruptor), and protein extracts were then incubated on ice for 10 minutes. Protein concentrations were measured by Bradford assay, at 595 A with Ultrospec® 2000 (Pharmacia Biotech), by mixing 800 µL milli-Q water, 200 µL Protein Assay Dye Reagent

(Bio-Rad) and 2μL of the protein extract. 2 μL lysis buffer was used to set reference. The standard equation "y= 6,2783*x+2,0083" was generated and commonly used by the group, and further utilized for calculating the protein concentrations. The preferred concentration of the protein extracts, milli-Q water, 50 μL 4x LDS Sample Buffer (Novex by life technologies) and 100 mM DTT (Saween & Werner), were mixed to a final volume of 200 μL. The proteins were incubated at 95°C for 5 minutes, before loading on a Bolt™ Bis-Tris plus SDS-gel (Invitrogen), for protein separation. Precision Plus Protein™ Dual Color Standards (Bio-Rad) was used as a ladder.

After transferring the gel to a 0.2 μm nitrocellulose Single Application membrane (Bio-Rad) with Trans-Blot® Turbo™, Ponceau S solution (Sigma) was used to confirm successful blotting by staining the proteins on the membrane. The membrane was washed with water, before incubation in 5% blocking milk (Sigma-Aldrich) for at least 1 hour or overnight (ON). After blocking, the membrane was washed with 1% blocking milk for 10 minutes. Thereafter, the membrane was incubated with primary antibody diluted in 5% blocking milk for 2 hours or ON. Excess primary antibody was washed off with 1% blocking milk, 3 x 5-10 minutes, before incubating with secondary antibody diluted in 5% blocking milk for 30-45 minutes. Excess secondary antibody was washed off with TBS-T, 1 x 5-10 minutes, and 1% blocking milk, 2 x 5-10 minutes. SuperSignal™ West Pico PLUS Chemiluminescent Substrate (Thermo Scientific) was used to develop the membranes. The mix was wiped off before pictures were conducted with the ChemiDoc™ MP system (Bio-Rad). The housekeeping genes vinculin and GAPDH (glyceraldehyde 3-phosphate dehydrogenase) were used as loading controls.

3.5 Differentiation protocols

Differentiation into embryoid bodies

To test the differentiation ability, mES cells were differentiated into embryoid bodies by hanging-drop culturing. 160 hanging drops of 20 μL, with 400 cells in each droplet, were hanging cultured in the lid of a 15 cm dish. 20 mL 1x PBS was added to the bottom of the dish to avoid drying out the droplets (Fig. 17A).

Harvested pellets at d3 by adding 160 μL 1x PBS and the solution was transferred to a 15 mL falcon tube. When the EBs formed a pellet, the supernatant was removed with care before the EBs were centrifuged and snap-frozen in liquid nitrogen. For the remaining plates, d5, d7, and d10, the EBs were transferred to a 10 cm bacterial dish with 10 mL differentiation

media. d5, d7, and d10 were harvested according to the timepoints, by transferring the solution from the bacterial plate to a falcon tube, and then pelleted as described for d3.

cDNA was generated by isolation of total RNA (described in 3.5.1) and reverse transcription PCR. qPCR was then performed with cDNA (described in 3.5.2) for the markers of interest, including: *RPLPO/RPO* (Ribosomal protein) (we refer to this protein as *RPO* from her)*, Nanog, Oct4, Sox1, Nestin, Tubb3, Pax3, Pax6, Gata4, Gata6, Brachyury,* and *Foxa2* (Table S1: List of oligonucleotides). All values were normalized to d0 *RPO*.

Monolayer differentiation with all-trans retinoic acid

This experiment was carried out using heterozygous MMSET and MMSET null cells, seeded on 6 cm laminin-coated dishes. Pellets were harvested at four different time points, d0, d2 (48 hours), d4 (96 hours), and d6 (144 hours). 1 million, 0.25 million, and 0.1 million cells were seeded for d2, d4, and d6, respectively. Cells were seeded in 5 mL differentiation media applied with 2 μM ATRA. Pellets were harvested and snap-frozen according to the timepoints, while differentiation media with 2 μM ATRA was changed in the other dishes every second day.

cDNA was generated and qPCR performed as described in 3.5.1 and 3.5.2. The following markers were used: *RPO, Nanog, Oct4, Sox1, Nestin, Tubb3, Pax3, Pax6, Gata4, Gata6, Brachyury,* and *Foxa2* (Table S1: List of oligonucleotides). All values were normalized to d0 *RPO*. For values that were undetermined as a consequence of no gene expression, the CT-values were set to 40.

Differentiation of mES cells into neuronal stem cells

This experiment was carried out on mES cells. 2 million cells were seeded on 10 cm laminin-coated dishes with N2B27 media (table 1). The cells were split after two days and 2 million cells were seeded in fresh N2B27 media. Cells were split at d4, and 2 million cells were seeded in NSC media with EGF/FGFb. Cells were split again at d6, as described for d4.

Isolation of RNA and generation of cDNA was carried out as described in 3.5.1, and qPCR was then performed, as described in 3.5.2 with the following markers: *RPO, Nanog, Sox1, Nestin, Pax6, Gata4, Gata6,* and *Foxa2* (Table S1: List of oligonucleotides). All values were normalized to d0 *RPO*.

Spontaneous differentiation of mES cells upon LIF withdrawal

This spontaneous differentiation was performed for the following knock-out cells; *Df1*, *Df3*, *Df1/2,* and *Df1/3*. Seeded 2 million cells for d2 and d4, and 1 million for d6, on laminin-coated 10 cm dishes with N2B27 media (Table 1). Pellets were harvested according to the timepoints, and cells were washed with 1x PBS before fresh media was added in the other dishes.

Isolation of RNA and generation of cDNA was done as described in 3.5.1, and qPCR was then performed, as described in 3.5.2 using the following markers: *RPO, Nanog, Sox1, Nestin, Pax6, Gata4, Gata6,* and *Foxa2* (Table S1: List of oligonucleotides). All values were normalized to d0 *RPO*.

3.5.1 Isolation of RNA and cDNA synthesis

RNA was isolated using RNeasy® Plus Mini Kit (50) kit (Qiagen). The RNA concentration was measured with NanoDrop One (Thermo scientific). 500 ng RNA was mixed to a total volume of 10 µL with RNase free water. cDNA was synthesized with The High Capacity cDNA Reverse Transcription Kit (Thermo Fisher) without RNase inhibitor, mixed with 10 µL RNA to a total volume of 20 µL per PCR tube.

The following PCR program was used: initial denaturation at 25°C for 10 minutes, denaturation at 37°C for 1 hour, annealing at 37°C for 1 hour, extension at 85°C for 5 minutes, ending with a completion step at 4°C ∞. cDNA was diluted 1:10.

3.5.2 Quantitative polymerase chain reaction

11.1 pmol/µL forward and reverse primers, were made according to the "Oligonucleotide synthesis report" from Eurofins (Table S1: List of oligonucleotides). 2 µL of the primer mix was mixed with 10 µL PowerUp™ SYBR™ Green (Thermo Fisher Scientific), and 6 µL RNase free water. 18 µL of the Master Mix and 2 µL cDNA/gDNA were used to perform qPCR, with StepOnePlus™ system from Applied Biosystems using following conditions: initial denaturation at 95°C for 10 minutes, 40 cycles of denaturation at 95°C for 15 seconds and annealing at 60°C for 25 seconds, ending with a melting curve at 95°C for 15 seconds, 60°C for 1 minute and 95°C for 15 seconds.

3.6 Immunoprecipitation

DNA-RNA hybrid immunoprecipitation (DRIP) western blot

All buffer compositions are listed in supplementary "7.2.3 Immunoprecipitation".

This experiment was carried out using ~10 million heterozygous MMSET and MMSET null cells. Pellets were lysed in 3-4 times pellet size lysis buffer, resuspended by pipetting, and incubated on ice for at least 10 minutes, or until the pellets were lysed. The extracts were centrifuged at 1,000 g for 10 minutes at 4°C before 200 µL nuclear lysis buffer was added to resuspend the nuclear pellet. Extracts were then sonicated for ten cycles at 4°C for 30 seconds ON/OFF (Diagenode Bioruptor), and centrifuged at 20,000 g for 20 minutes at 4°C. The supernatant of the extracts was diluted 1:4 with RSB-T. 50 µL of the extracts was kept for input, and 950 µL was subjected to IPs with S9.6 or IgG antibodies (Table S4: List of antibodies).

30-50 µL protein G dynabeads (Life Technologies) per IP were conjugated with 1 mL 1x PBS and either with 1 µg/µL S9.6 or IgG. The beads were incubated at 4°C, 40 rpm tumbling (Rocker & Rotator PTR 35, cmscientific) ON, or at least for 3 hours before IP.

IP was performed for 2 hours at 4°C on 40 rpm tumbling. After IP, the beads were washed 4x with RSB-T and 2x with RSB. Beads were resuspended between each wash, and if clumping occurred the beads were severely resuspended by pipetting. Elution was performed with 50 µL 2x LSD (diluted in RSB), 100 mM DTT, for 10 minutes at 70°C. Magnetic beads were separated, and the supernatant kept after elution.

Protein extracts were loaded on a 6% Bolt™ Bis-Tris plus SDS-gel (Invitrogen), for protein separation. 1.25% was loaded as input. The same protocol as described for western blotting (3.4) was followed. The following primary antibodies were used: DHX9, DDX3, DDX5, MMSET, and vinculin (loading control) (Table S4: List of antibodies).

MMSET immunoprecipitation

Performed as described in 3.6. Used beads conjugated with 1 mL of internally laboratory made mouse monoclonal antibody supernatant 8C3 for MMSET (table S4).

DNA-RNA hybrid immunoprecipitation qPCR

Immunoprecipitation was performed as described in 3.6. During IP 1 µL benzonase (EMD Millipore, Novagen) was added where indicated. Elution was performed using 120 µL de-crosslinking buffer and 6.67 U/mL proteinase K (BioLabs), this was also added in the input.

Input and IPs were incubated at 65°C for 3 hours on shaking (1,000 rpm). DNA was purified using QIAquick® PCR purification kit (QIAGEN) with 100 µL elution buffer. 2 µL was used in qPCR (described in 3.5.2) with the following markers; *β-actin P* and *C*, *CyclinB1 P* and *C*, *Gata4 P,* and *Myf5 P* (Table S2: List of oligonucleotides).

3.7 Immunofluorescence staining

This experiment was carried out using heterozygous MMSET and MMSET deleted cells. 150,000 cells were seeded in a 24-well plate, which was gelatin-coated over coverslips, to a final volume of 500 µL. Three hours later, 200 µM H_2O_2 (Sigma) in 500 µL ES cell media was added, while the media was gently changed to fresh ES cell media for non-treated cells. After 1 hour of treatment, the media was gently removed, and the coverslips were washed twice with ice-cold 1x PBS. All cells were then fixed with 4% paraformaldehyde (Sigma) for 10-15 minutes. The fixative was washed off 3 x 5 minutes with 1x PBS. Coverslips with antibodies diluted in blocking buffer (ice-cold 1x PBS with 10% fetal bovine serum (FBS)) were incubated on parafilm in a humid chamber at 4°C, overnight (Table S4: List of antibodies). Excess primary antibodies were washed off 3 x 5 minutes with PBS containing 0.1% Triton X-100 (Sigma) at room temperature. 100 µL secondary antibody diluted in blocking buffer, was added for 1 hour at room temperature (Table S4: List of antibodies). Excess bound secondary antibody was washed off 3 x 5 minutes with 1x PBS containing 0.1 % Tween-20 (Sigma-Aldrich). Cells were then mounted with 12 µL Vectashield, a Mounting medium for fluorescence with DAPI (4′,6-diamidino-2-phenylindole) (Vector). Coverslips dried ON before imaging with Leica TCS SP8 STED, super-resolution microscope.

3.8 Statistical analysis

One-way analysis of variance (ANOVA)

The statistical analysis one-way ANOVA can be used if more than two independent groups are being compared, and the *interval*-scale variable is almost normally distributed (McCrum-Gardner, 2008). One-way ANOVA were used for all static analysis with more than two groups, hence all experiments including mRNA expression of the *Ythdf1-3* knock-out cells were analyzed with one-way ANOVA. The one-way ANOVA analysis were calculated using GraphPad Prism 8 software. Unless otherwise specified the “default” parameters were used. The followup tests “compare the mean of each column with the mean of a control column” was

chosen. Error bars indicate ± SD (standard deviation) and n=2. Asterisks indicate statistical significance (*p <0.05, **p <0.01,***p <0.001).

Student´s t-test

Student´s unpaired t-test is used for two independent groups when the distribution is almost normal, to compare the sample means (McCrum-Gardner, 2008). Hence, the two-tailed unpaired t-test was performed for all given gene expression values of $MMSET^{+/-}$ and $MMSET^{-/-}$ cells, where indicated in legends. The two-tailed t-tests were calculated using GraphPad Prism 8 software. We performed unpaired t-test, with the "default" settings, unless otherwise specified. Error bars indicate ± SD and n=2. Asterisks indicate statistical significance (*p <0.05, **p <0.01,***p <0.001).

4. Results

4.1 Choice of model system

Mouse embryonic stem cells are an excellent model system to study genetic, epigenetic, and epitranscriptomic changes. ES cells are derived from the inner cell mass of developing blastocyst and are pluripotent. Furthermore, ES cells have the unique ability to indefinitely self-renew and to differentiate into all germ lineages, both *in vitro* and *in vivo* (Niwa, 2010). ES cells grow fast and are overall easy to handle, but their greatest advantage is that they can easily be clonally expanded and become genetically modified while remaining pluripotent. This makes it relatively easy to genetically modify ES cells.

ES cells´ ability to differentiate *in vitro*, and the fact that ES cells can re-enter normal development when introduced into morula- or blastocyst stage embryos, provides a great model system to study early developmental processes (Evans & Kaufman, 1981; Martin, 1981; Yagi et al., 2017). Since cell identity or loss of it is a hallmark of cancer and many other diseases, pluripotent stem cells make a powerful tool to investigate this. Both human and mouse ES cells exhibit these properties, but because mES cell lines were available at the laboratory, we chose them as the model system for all experiments presented in this thesis.

4.2 Generation of *Ythdf1-3* KO cells and characterization of their phenotypes

A need for further investigation of the YTHDF1-3 protein functions

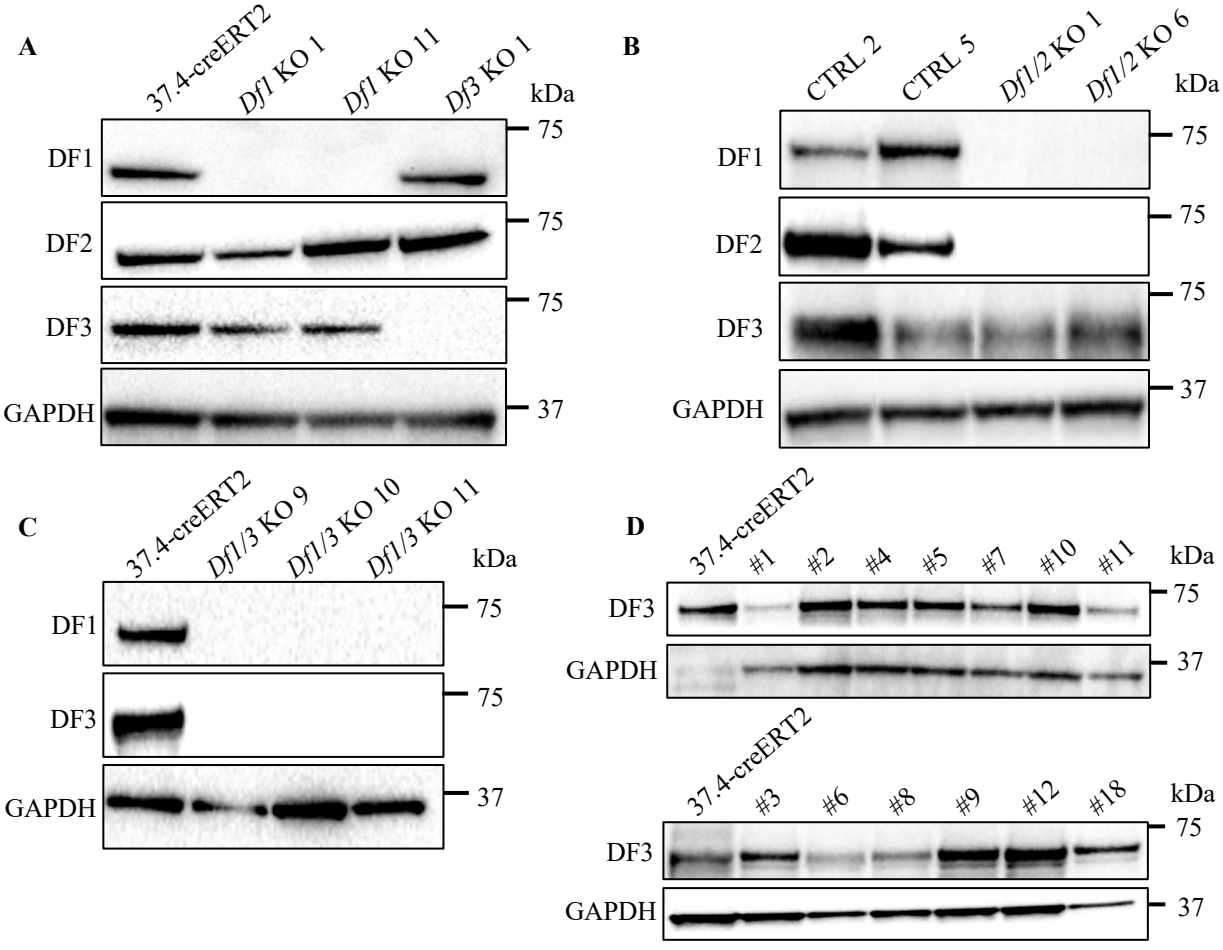
For a long time, studies have shown dissimilar results whether DF1-3 possesses the same or different functions (Zaccara et al., 2019). Some studies have suggested that they all recruit mRNA deadenylation complexes and have similar functions in mRNA degradation, while most studies report dissimilar functions (Du et al., 2016; Kennedy et al., 2016) (Fig. 9).

To elucidate whether the DF proteins possesses similar or dissimilar functions, it is essential to reveal how the DF proteins bind m^6A sites in mRNA. Additionally, their target specificity needs to be determined, and whether the m^6A sites are bound by one, two, or all three reader proteins. So far, it has been reported that DF1 promotes protein synthesis, while DF2 affects mRNA translation and stability, and accelerates mRNA decay. DF3 is shown to be a partner of DF1 and DF2, therefore promoting both translation and degradation of m^6A containing mRNAs (Liao et al., 2018; Shi et al., 2017; Wang, X. et al., 2014; Wang et al., 2015).

Based on this knowledge, we wanted to investigate DF1-3 functions. So far (May 2020), there are no published studies on single, double, or triple *Df1-3* knock-down or knock-out in mES cells. Therefore, we designed an experimental plan to generate single (*Df1-3),* double (*Df1/2*, *Df2/3*, *Df1/3*) and triple (*Df1-3*) KO cells, and further characterize the phenotype with proliferation and differentiation assays.

4.2.1 Validation of the generated *Ythdf1-3* knock-out cells

After treatment with tamoxifen and transfection with sgRNAs to generate *Df1*, *Df2,* and *Df3* KOs*,* they were sorted as single cells by FACS (Fig. S4), and colonies were picked up for characterization with western blot analysis. Nine (*Df2*), ten (*Df3),* and twelve (*Df1*) clones were analyzed, and four *Df1* KOs, two *Df2* KOs, and one *Df3* KO were identified (Fig. 14A/B and S2). As shown in figure 14B, the two *Df2* KOs were also identified as *Df1* knock-outs. Further, clone 2 (#2) and clone 5 (#5) were kept as control cells, since they had been through the same conditions as the validated *Df1/2* KO cells (Fig. 14B and S2B). Moreover, three clones out of ten analyzed were positive for *Df1/3* KO (Fig. 14C). To summarize, we generated *Df1*, *Df3*, *Df1/2,* and *Df1/3* KO cells.



(legend on next page)

Figure 14. Validation of *Df1, Df3*, *Df1/2,* and *Df1/3* KOs by western blot analysis.
Western blots were probed with the indicated antibodies and GAPDH was used as a loading control.
(A) Two *Df1* KO cells were generated by treating 37.4-creERT2 cells with tamoxifen. By transfecting 37.4-creERT2 cells with sgRNAs specific for *Df3,* one *Df3* single KO was generated.
(B) Transfection of 37.4-creERT2 cells with sgRNA specific for *Df2,* resulted in two *Df2* KO cells, which later were validated as a double *Df1/2* KOs.
(C) After treating the generated *Df3* KO 1 with tamoxifen, three clones were validated as double *Df1/3* KO cells.
(D) The *Df1/2* KO 6 was transfected with sgRNA specific for *Df3,* to generate a triple *Df1-3* KO. 36 clones were picked up from the 96-well plate after a single-cell sorting by FACS, and only 13 were analyzed with western blot.

Further, the validated *Df1/2* KO 6 cells were transfected with sgRNAs for generation of the triple *Df1-3* knock-out (Fig. 14D). The transfected cells were sorted as single cells into two 96-well plates. 36 clones were picked up, whereas 13 of the colonies obtained sufficient cell density for western blot analysis. Finally, results from the western blot analysis revealed that it was not generated any triple *Df* knock-out, since the DF3 protein could be detected in all the tested clones (Fig. 14B).

4.2.2 Phenotype characterization

During cultivation of the *Df* knock-out cells, different morphologies were observed, hence, microscopy imaging of the cells were conducted (Fig. 15A). The *Df1/3* KO was not generated at the time the further assays were performed.

The control cell lines, 35.6 and 37.4-creERT2, exhibit a round shape with tight borders, a morphology typical for mES cells. The morphology of *Df1* KO cells consists of mostly round shaped cells, with colonies appearing to be bigger than the control cells. The cells with *Df3* and *Df1/2* KO, seem to have lost this distinct border integrity, and the cells appear to spread out. Overall, it looks like the phenotype is affected by knock-out of the *Df* genes.

Proliferation assay was performed over twelve days for the *Df1*, *Df3,* and *Df1/2* KO cells (Fig. 15C). Cumulative growth was calculated based on 100,000 cells seeded at d0 and describes a percentage of increase during the twelve days. No major differences in growth ability was observed upon *Ythdf1-3* single and double knock-out. At this stage of understanding, all knock-out cells appear to have preserved pluripotency, due to the similar mRNA expression of *Nanog* in comparison to the control cells (Fig. 15B). The preserved pluripotency and altered proliferation phenotype, collectively made it interesting to further investigate the differentiation capacity of the knock-out cells.

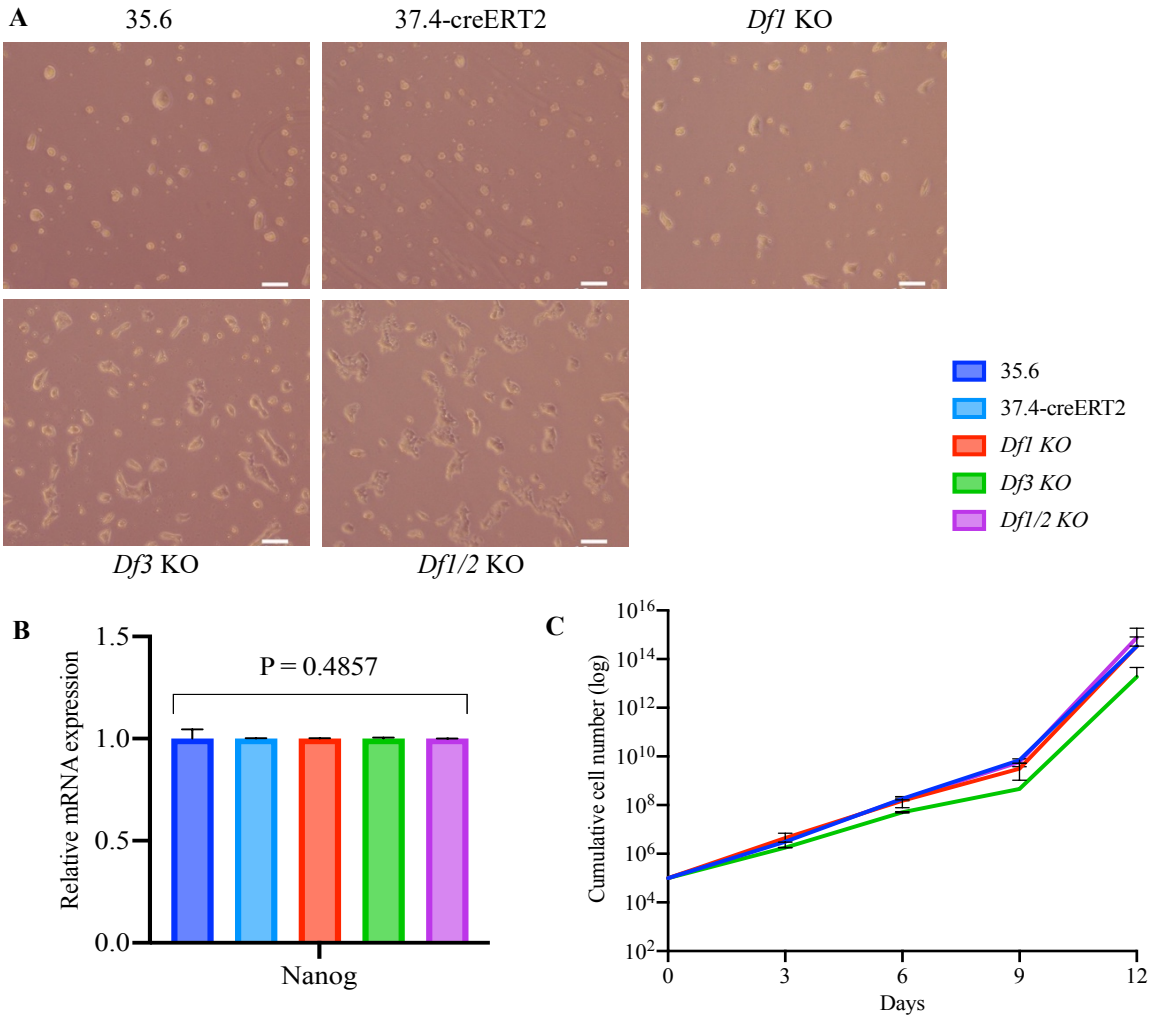


Figure 15. Characterization of *Df1*, *Df3,* and *Df1/2* knock-out cells. 35.6 and 37.4-creERT2 were included as control cell lines. When the characterization assays were performed, the *Df1/3* KO was not generated.
(A) Microscopy images of the *Df1*, *Df3,* and *Df1/2* KO cell morphology, taken with 4x objective. One million cells where cultured for two days on gelatin-coated dishes in ES cell media. Scale bar: 10 µm.
(B) Relative mRNA expression of the pluripotency gene *Nanog*. Normalized to d0 *RPO*-values. Error bars indicate ±SD; n=2, and P value were calculated by one-way ANOVA.
(C) Cells were seeded on gelatin-coated dishes in ES cell media and counted every 3rd day. Cumulative growth was calculated based on 100,000 cells seeded at d0. Only 35.6 was included as a control. Error bars indicate ±SD; n=2, and P value were calculated by one-way ANOVA, and p>0.05.

4.2.3 Differentiation capacity of mES cells upon deletion of the YTHDF proteins

To investigate the differentiation capacity, we set up an NSC differentiation experiment. Upon 2i/LIF withdrawal, the growth seemed hampered in the knock-out cells, and a high cell death was observed after splitting. When EGF/FGF were added at d4 the majority of knock-out cells died (Fig. 16A). As a consequence of this experience, we only relied on the data from d2. Following, this experiment was analyzed as spontaneous differentiation upon 2i/LIF withdrawal (Fig. 16B).

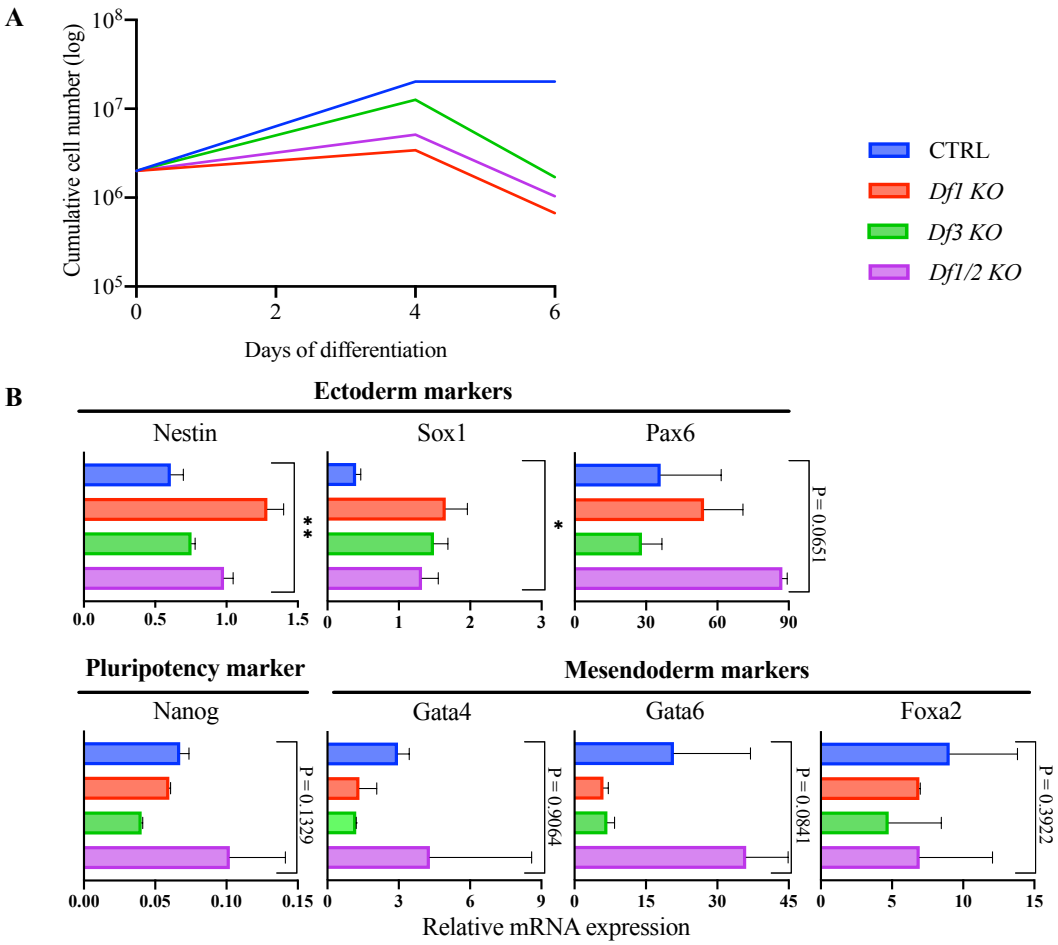


Figure 16. Differentiation capacity of *Df*1, *Df*3 and *Df*1/2 KOs upon 2i/LIF withdrawal.
(A) Growth curve showing cumulative growth of the different knock-out cells during differentiation towards the ectoderm lineage. Cells were cultured without 2i/LIF for four days, before EGF/FGF were added. Control 5 was used as a control.
(B) Relative mRNA expression of pluripotency, ectoderm, and mesendoderm genes, two days after 2i/LIF withdrawal. 35.6 was used as a control. Values are normalized to d0 *RPO*. Error bars indicate ±SD; n=2, and P value were calculated by one-way ANOVA, *p <0.05, **p <0.01.

The mRNA expression of the pluripotency gene *Nanog*, are ranging from 0.04-0.10 in the control and knock-out cells two days upon 2i/LIF withdrawal. It is important to assess these observed values in comparison to *Nanog* mRNA expression at d0 (Fig 15C). These results collectively predict that the downregulation of *Nanog* is not impaired upon 2i/LIF withdrawal and that the knock-out cells can exit pluripotency (Fig. 16B).

Additionally, both ectoderm and mesendoderm gene expression seem to be affected upon 2i/LIF withdrawal. The results demonstrated that the mRNA expression of the ectoderm markers, *Nestin* and *Sox1*, is slightly higher in the knock-out cells compared to the control. Furthermore, a lower expression pattern of both *Df1* and *Df3* KO cells is observed for all the tested mesendoderm markers. Moreover, by performing EB differentiation, the objective was to avoid the great cell death observed during NSC differentiation and further investigate the differentiation capacity of the knock-out cells.

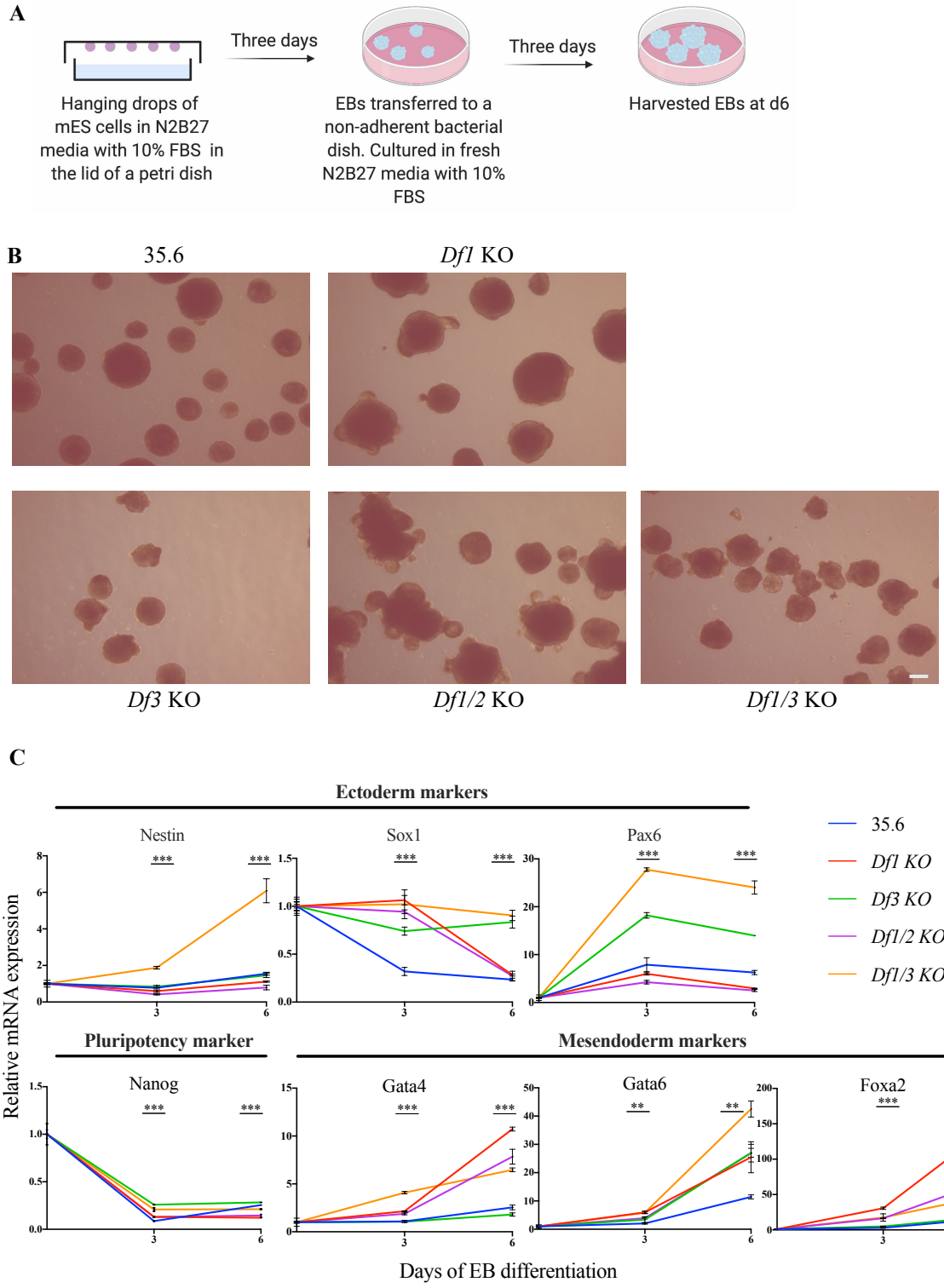


Figure 17. EB differentiation of *Df1*, *Df3*, *Df1/2*, and *Df1/3* KOs.
(A) Schematic overview of the EB differentiation method. Hanging drops with 400 cells in N2B27 media with 10% FBS were cultured in the lid of a petri dish. After three days, the EBs were transferred to a non-adherent bacterial dish. At d6 the EBs were harvested.
(**B**) Morphology of the EBs at d6. Microscopy images conducted with 4x objective. Scale bar: 20µm.
(C) Relative mRNA expression of pluripotency, ectoderm, and mesendoderm genes. 35.6 was used as a control. Normalized to d0 *RPO*-values. Error bars indicate ±SD; n=2, and P values were calculated by one-way ANOVA. **p <0.01,***p <0.001.

By the time EB differentiation was carried out, the *Df1/3* KO was generated and included in the assay (Fig. 17A). The microscopy images of EBs at d6 display alterations in morphology (Fig. 17B). Moreover, the morphology of the control cell line, 35.6, appears to be round with a smooth surface, while the knock-out EBs exhibit a rougher surface. From the microscopy images it is clear that the EBs differ in size. The EBs of *Df1* and *Df1/2 KO* cells are largest, while the *Df3* and *Df1/3* are smallest.

The relative mRNA expression is fluctuating for all ectoderm and mesendoderm genes during EB differentiation (Fig. 17C). The *Df1/3* KO has an increased expression of the ectoderm genes *Nestin* and *Pax6*. *Sox1* is downregulated in the control at d3, contrary the mRNA expression in the knock-out cells are more or less unaltered. Furthermore, cells knocked out for *Df1* exhibit increased expression of the mesendoderm genes *Gata4* and *Foxa2*, additionally all knock-out cells display higher mRNA expression of *Gata6,* in comparison with the control. The generated knock-out cells exit pluripotency successfully, due to downregulation of *Nanog*. Overall, it is hard to distinguish a clear phenotype for any of the knock-out cells, but it appears that the ability to induce differentiation toward the three germ lineages varies.

4.3 The role of a chromatin modifier, MMSET in R-loop biology

4.3.1 MMSET in pluripotency exit and mesendoderm specification

Tian et al., 2019, reported that MMSET has a dual role in pluripotency exit and germ layer specification of mES cells. This role is independent of the methyltransferase activity of MMSET, which is mediated by the SET-domain. During the induction of differentiation, they observed that a proportion of MMSET-depleted ES cells remained entrapped in a pluripotent state. Hence, MMSET does not appear to play a major role in generating neuroectoderm, but the downregulated MMSET cells failed to generate mesendoderm. This was explained by MMSETs capability of activating different mesendoderm regulators. MMSET together with Brd4, can bind to enhancers of the mesendoderm activators, and thereby activate the expression of the mesendoderm genes, including *Gata4, Gata6, Brachyury,* and *Foxa2*. The catalytic SET domain of MMSET was redundant for both pluripotency exit and mesendoderm differentiation (Tian et al., 2019).

We designed experiments to reproduce results reported by Tian and colleagues with our heterozygous MMSET and MMSET null mES cells, which were also used in Tian et al., 2019. These cells were acquired from Kaneda laboratory as described in Nimura et al., 2009. Before investigating differentiation ability upon deletion of MMSET, we characterized the heterozygous MMSET and MMSET null cell lines by analyzing morphology, proliferation capacity, and the presence of MMSET proteins.

4.3.2 Validation and characterization of the MMSET$^{+/-}$ and MMSET$^{-/-}$ mES cell lines

For validation of the cell lines, western blot analysis was carried out. The western blot revealed that the MMSET protein expression was absent in the MMSET null cells (Fig. 18C). In comparison to the E14 cell line, the heterozygous MMSET exhibit a lower expression of the MMSET protein (Fig. 18C and 21B). Microscopy examination of the cells revealed no major morphological difference between the MMSET$^{+/-}$ and MMSET$^{-/-}$ cells (Fig. 18A). In an attempt to address a growth defect in MMSET deleted cells, a proliferation assay was performed. Nevertheless, it is interesting to note that the growth density of MMSET$^{-/-}$ is marginally higher as opposed to the observed density of MMSET$^{+/-}$, which is consistent with observations in the laboratory (Fig.18B). However, it is difficult to state any definite growth effects upon MMSET deletion.

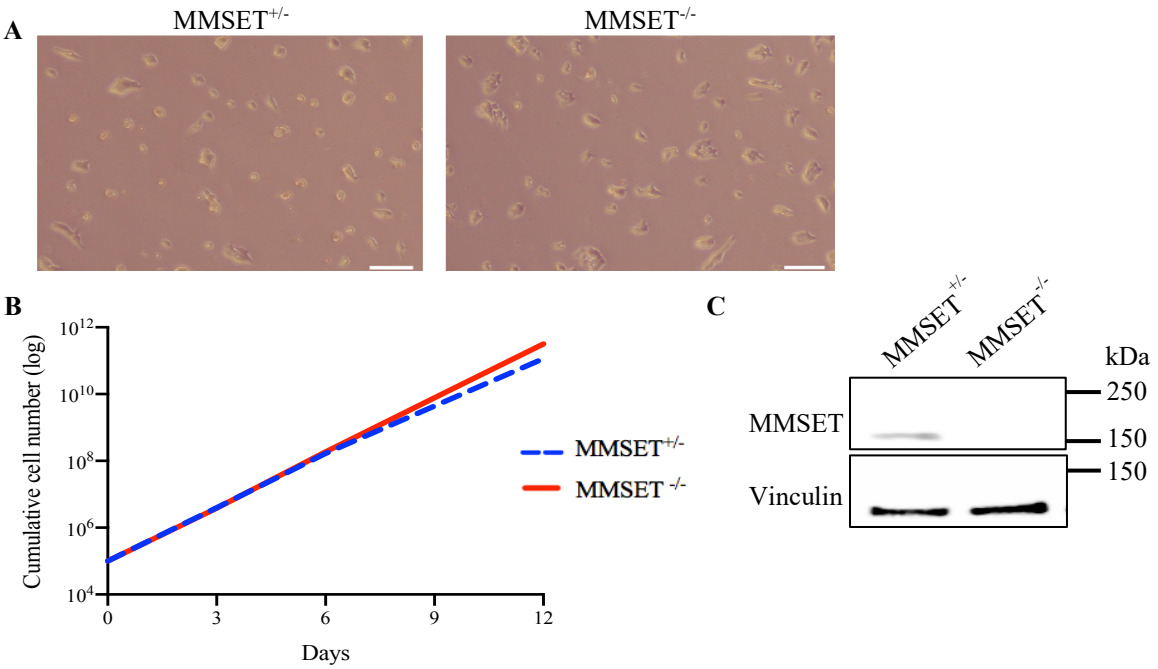
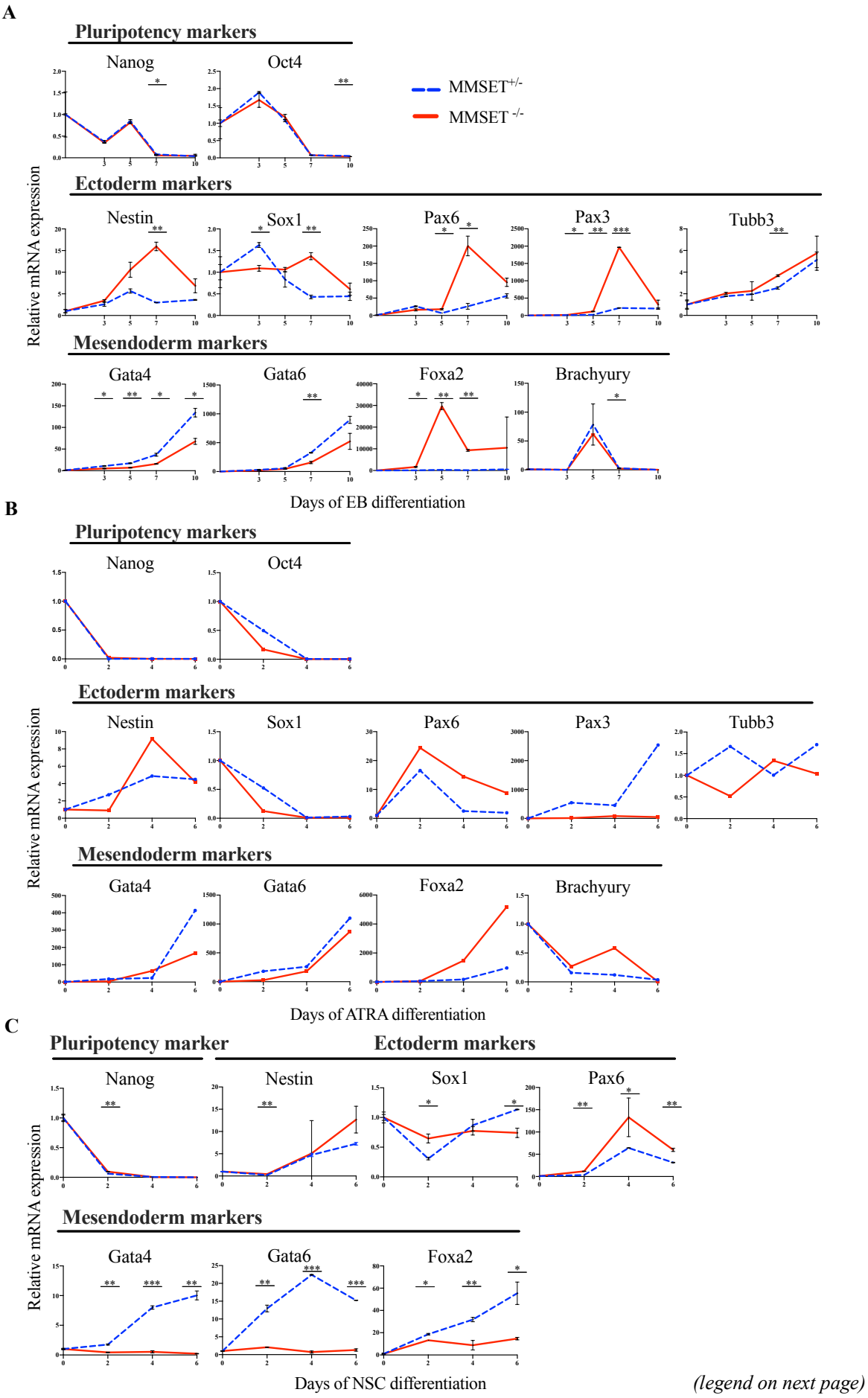


Figure 18. Characterization and validation of the MMSET$^{+/-}$ and MMSET$^{-/-}$ mES cell lines.
(A) Microscopy images of the MMSET$^{+/-}$ and MMSET$^{-/-}$ cells, acquired with 4x objective. Scale bar: 10µm.
(B) Proliferation assay of MMSET$^{+/-}$ and MMSET$^{-/-}$ cells showing cumulative cell growth. 100,000 cells were seeded in ES cell media and cell density was monitored evert 3rd day.
(**C**) Western blot analysis validating the MMSET$^{+/-}$ and MMSET$^{-/-}$ cell lines, probed with MMSET antibody. Vinculin was used as a loading control.

4.3.3 Differentiation ability upon MMSET deletion

To recapitulate the results from Tian et al., 2019, revealing delayed pluripotency exit and impaired mesendoderm formation upon MMSET removal, we designed an EB differentiation assay. In addition, to determinate the ability to exit pluripotency and to differentiate towards mesendoderm, we wanted to examine ectoderm lineage formation as well. We observed that the mRNA expression of *Nanog* and *Oct4* were downregulated simultaneously at d7, both for heterozygous MMSET and MMSET null cells (Fig. 19A). Contrary to the findings by Tian and collagenous, our results did not show delayed downregulation of the pluripotency markers. In MMSET null cells, a reduced expression of *Gata4* and *Gata6* was observed, on the other hand, the mesendoderm gene *Foxa2* exhibited a higher expression. Moreover, all ectoderm genes appear to normalize since they are seeking a common mRNA expression level. In MMSET null cells, the bulk of the ectoderm genes show an increased mRNA expression at d7, prior to normalization. Overall, pluripotency exit seems to be preserved in both heterozygous MMSET and MMSET null cells. Additionally, fluctuating mRNA expression was observed for the mesendoderm genes, while the effects of MMSET deletion was predominantly detected for the ectoderm genes at d7.



(legend on next page)

Figure 19. The ability of MMSET$^{+/-}$ and MMSET$^{-/-}$ mES cells to exit pluripotency and differentiate.
Relative mRNA expression of pluripotency, ectoderm, and mesendoderm genes upon differentiation. Values are normalized to d0 *RPO*.
(A) EB differentiation assay. EBs derived by hanging drops of 400 cells in N2B27 media with 10% FBS. p <0.001.
(B) ATRA differentiation assay. Cells were cultured on gelatin-coated dishes with N2B27 media and ATRA for differentiation. Genes that were not expressed gave undetermined values, hence, the Ct values were then set to 40. This applies to *Pax3* d0 and *Nanog* d2 in MMSET$^{+/-}$ cells, and for *Foxa2* d0 and *Brachyury* d6 in MMSET$^{-/-}$ cells. No specified error bars or P values due to n=1.
(C) NSC differentiation assay. Cells were cultured on laminin-coated dishes with N2B27 media for two days, before EGF/FGF were added at d4, for further differentiation towards ectoderm.
For A/C: Error bars indicate ±SD; n=2, and P values were calculated by two-tailed unpaired t-test. *p <0.05, **p <0.01,***p <0.001.

To further evaluate the differentiation capacity, ATRA and NSC differentiation assays were carried out. A complete pluripotency exit was observed simultaneously between the MMSET$^{+/-}$ and MMSET$^{-/-}$ cells during the three differentiation assays (Fig. 19). When comparing our results to those of Tian and colleagues´, it is surprising that none of the differentiation assays show impaired pluripotency exit upon MMSET deletion. Of the mesendoderm genes, *Gata4* and *Gata6* seem to be most consistently reduced in MMSET null cells throughout the EB and ATRA differentiation assays, while the mesendoderm expression was impaired for all genes during NSC differentiation (Fig. 19). Furthermore, mRNA expression of the ectoderm genes was affected by MMSET deletion during ATRA and NSC differentiation, however, a consistent mRNA expression pattern was not observed (Fig. 19B/C). To summarize, the three tested differentiation assays showed altered ability to generate mesendoderm and ectoderm lineages, while the ability to exit pluripotency was preserved, upon deletion of MMSET.

4.3.4 DHX9, DDX3, and DDX5 – interaction partners of MMSET and R-loops

An earlier study performed in the group to identify the protein-protein interaction partners of MMSET revealed a large overlap between MMSET interaction partners and R-loop binding proteins (Fig. 21A). The group´s study showed 113 overlapping proteins out of 629 tested. Many RNA helicases, including DHX9 and DDX5 in addition to chromatin-modifying proteins, like MMSET, were reported as interaction partners of R-loops (data not shown). Interestingly, DHX9, DDX5, and MMSET were also validated as interaction partners of R-loops by Cristini et al., 2018. Overall, there are several common interaction partners between R-loops and MMSET, but how MMSET deletion influence the interactions between R-loops and the RNA helicases still remains undetermined.

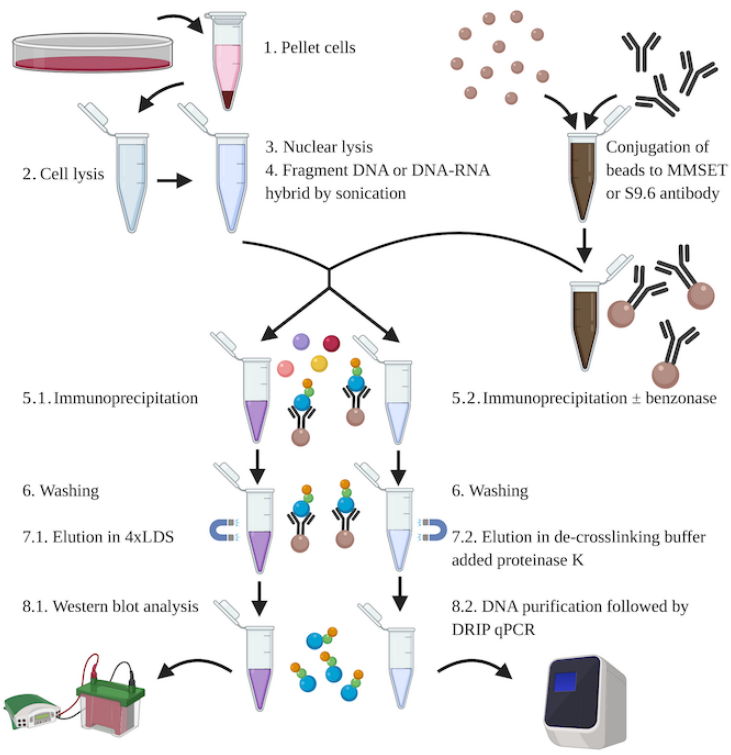


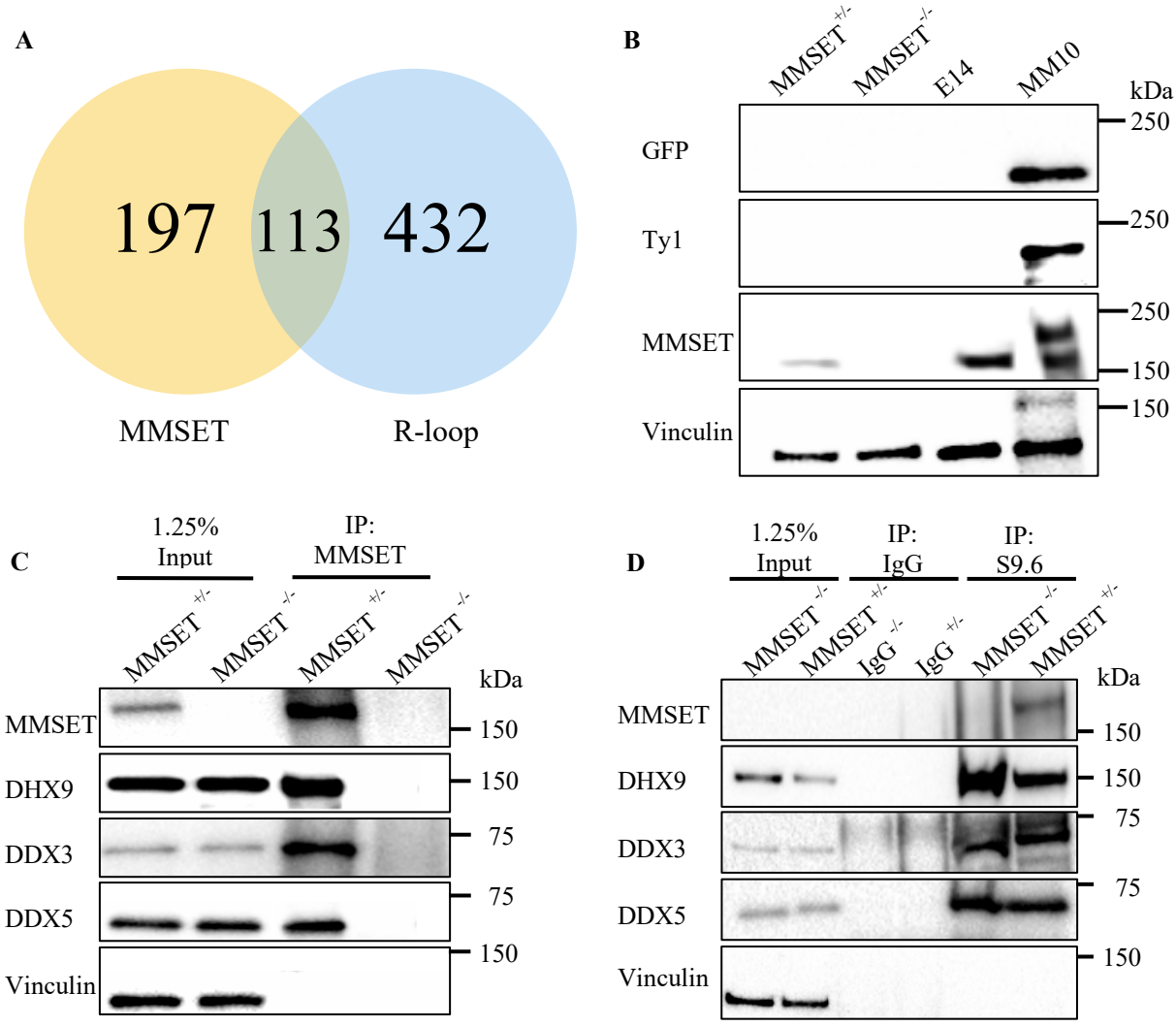
Figure 20. Schematic overview of the IP and DRIP workflow.
mES cells were harvested by pelleting, before cell lysis and nuclear lysis were carried out. DNA or DNA-RNA hybrids were fragmented by sonication before IP/DRIP, to minimize the probability of copurification of unspecific proteins (1-4). In the upper right corner, the figure shows conjugation of antibodies to dynabeads. Dynabeads used for DRIP WB and qPCR were conjugated to the R-loop specific S9.6 antibody, while dynabeads for MMSET IP were conjugated to MMSET (8C3) antibody. Then, the nuclear extracts were immunoprecipitated (5.1 and 5.2). Benzonase was added during step 5.2 as a control, since it degrades all forms of nucleic acids, including R-loops. After several washing steps (6), proteins were eluted in 4x LDS (7.1) and gDNA was eluted in de-crosslinking buffer (7.2). Proteinase K was added to the DRIP qPCR extracts to degrade nucleases present during gDNA extraction (7.2). MMSET and DRIP WB was performed directly (8.1), while the gDNA was purified before DRIP qPCR was carried out (8.2).

We wanted to examine the interaction partners of MMSET by MMSET immunoprecipitation and western blot analysis (Fig. 20). We used dynabeads to pull down MMSET in both MMSET$^{+/-}$ and MMSET$^{-/-}$ cell. Further, by performing western blot analysis, we could easily probe for the proteins of interest, hence, we included the RNA helicases DHX9, DDX3, and DDX5. These helicases have been reported as interaction partners of both MMSET and R-loops, and had the strongest enrichment in the MMSET pulldown (data from the group study in 2006 is not shown). In conclusion, MMSET, DHX9, DDX3, and DDX5 were detected by MMSET IP in heterozygous MMSET cells, while none of the proteins were pulled down in MMSET deleted cells (Fig. 21C).

It was interesting to further validate the interaction partners of R-loops and the effects upon MMSET deletion. This was examined by R-loop immunoprecipitation in MMSET$^{+/-}$ and MMSET$^{-/-}$ cells. We only detected the MMSET protein in the heterozygous cell line (Fig. 21D). Moreover, DHX9, DDX3, and DDX5 were pulled down in both heterozygous MMSET and MMSET null cells, and importantly, no proteins were detected in the IgG pulldowns (Fig. 21D and S5). A limitation in these results is that MMSET was not detected in the input for heterozygous MMSET cells. Finally, DHX9, DDX3, and DDX5 were validated as interaction partners with R-loops, independent of MMSET.

4.3.5 Validation of the Ty1-YFP-MMSET insert in the MM10 cell line

The MM10 cells stably express BAC-tagged MMSET with YFP-Ty1 to ensure a near endogenous expression of the tagged MMSET protein. Before the use of this cell line in the planned experiments, including MMSET chromatin immunoprecipitation (ChIP), we confirm the inserts through western blotting and immunoprecipitation. As figure 21B shows, Ty1, YFP, and MMSET inserts were detected in the MM10 cells. YFP could be detected with anti-GFP antibody, due to high amino acid sequence similarity between YFP and GFP (Kaltwasser et al., 2002; Veening et al., 2004). Collectively, the inserts were detected as bands at ~200 kDa for Ty1, YFP, and MMSET, and were importantly not detected in E14, MMSET$^{+/-}$, and MMSET$^{-/-}$ cells.



(legend on next page)

Figure 21. Validation of some MMSET and R-loop interactome candidates using the MMSET$^{+/-}$ and MMSET$^{-/-}$ mES cell lines. Vinculin was included as a loading control.
(A) Venn diagram showing overlapping interaction partners of MMSET and R-loop. Study performed by the group in 2006, presented in accordance with the group.
(B) Western blot analysis validating the Ty1-YFP-MMSET insert in the MM10 cell lines. Probed with indicated antibodies. Loaded ~2x protein amount of E14 and MM10, due to expected lower signals.
(C) MMSET IP western blot showing proteins interacting with MMSET. Probed with indicated antibodies.
(D) Enrichment of R-loop interaction partners by DRIP. IgG was included as a negative control, and the western blot was probed with indicated antibodies. Input could not be detected for heterozygous MMSET (top blot). See also figure S4.

4.3.6 R-loop enrichment in PcG repressed genes upon removal of MMSET

A recent study reported that R-loops are enriched in genes repressed by Polycomb group proteins (Skourti-Stathaki et al., 2019). The PcG proteins are major epigenetic regulators of transcriptional repression, assembled by two major complexes: PRC1 and PRC2. They are, among others, required to silence CpG rich developmental regulator genes in ES cells. Moreover, R-loops are enriched at both promotor (P) and coding (C) regions over five different PcG-repressed genes, revealed by DRIP analysis of mES cells (Skourti-Stathaki et al., 2019). Furthermore, the five PcG-repressed genes that R-loops are enriched at, includes *Gata4,* and these genes are GC-rich throughout both the P and C region, have well-annotated CpG island promoters, and are co-occupied by PRC1/2 (Skourti-Stathaki et al., 2019).

Based on results from Skourti-Stathaki et al., 2019, we wanted to investigate if R-loop formation over PcG-repressed genes was affected upon MMSET deletion. We selected some loci reported to be enriched in PcG-repressed genes and carried out DRIP-qPCR analyses. *Gata4* was included because of the already observed alteration in phenotype of *Gata4* in MMSET null cells (Fig. 19). Moreover, *CyclinB1-C* does not form R-loops, hence included as a negative control, while *β-actin-C*, a highly expressed active gene, was included as a positive control, because it forms R-loops over both P and C region (Skourti-Stathaki et al., 2011; Skourti-Stathaki et al., 2014). Additionally, benzonase treatment was performed since it degrades all forms of nucleic acids, including R-loops. *Myf5* was included since it is not associated with Pol II or PcG in mES cells, as illustrated in figure S6 (Brookes et al., 2012; Stock et al., 2007). It appears that a few of the primer pairs did not work optimally, therefore, the main results are shown in figure 22, while all primer pairs tested are shown in figure S6.

The DRIP analyses revealed an enrichment of R-loops at the *Gata4* promotor compared to the negative control *Cyclin-B1* in the coding region. A slightly lower enrichment at *Gata4* is observed in MMSET deleted cells (Fig. 22). *CyclinB1-C* exhibit an expression close to the benzonase treated cells, however *β-actin-C* shows a higher expression than *CyclinB1-C*. The

considerably lower expression of *β-actin-C* compared to *Gata4,* indicates that *β-actin-C* does not work optimally. As expected, benzonase treatment strongly reduced the R-loop signal in *Gata4*, but surprisingly this was not observed for *β-actin-C.* Conclusively, our results are broadly in line with earlier studies, showing enrichment of R-loops in the PcG repressed gene *Gata4.*

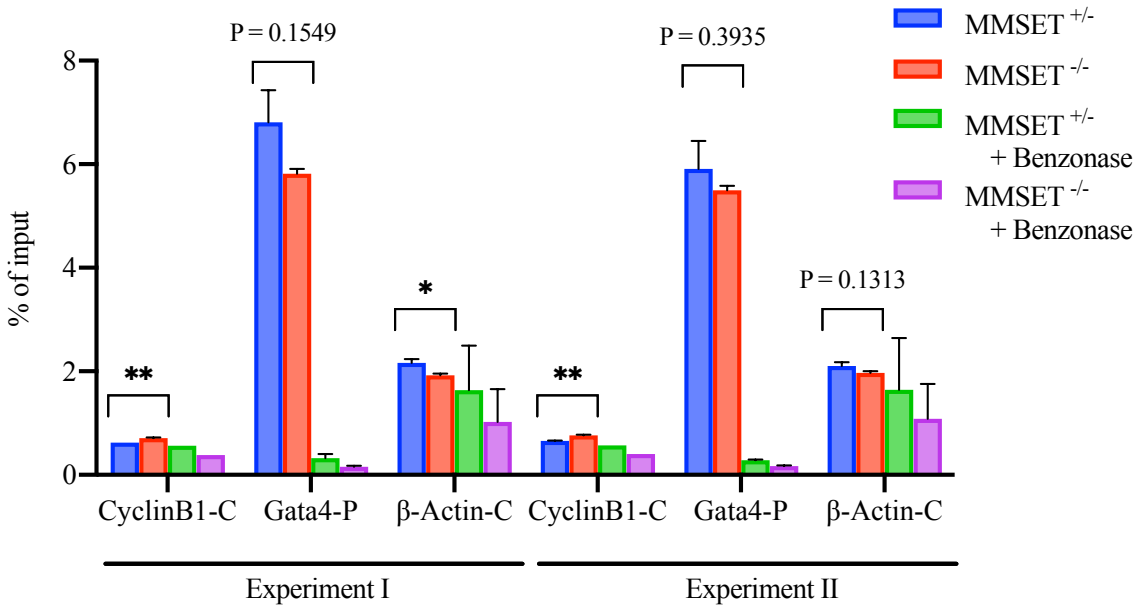


Figure 22. DRIP qPCR analyses of Polycomb-repressed and active genes.
The active genes *β-actin-C* and *Cyclin B1-C* were used as positive and negative controls, respectively. P and C represents promoter and coding regions. Two separate experiments were performed with similar results. Values are normalized to input values and represent the means. Where indicated, extracts were treated with 1 µL benzonase. Error bars indicate ±SD; n=2, and P values were calculated by two-tailed unpaired t-test. *p <0.05, **p <0.01. See also Figure S6.

4.3.7 Effects of MMSET deletion on R-loop accumulation

Several papers report that MMSET and R-loops are involved in DNA-damage responses (DDRs) (Hajdu et al., 2011; Pei et al., 2011). MMSET has been shown to mediate methylation of H4K20 in mammals, which increases locally upon induction of DSBs. p53 binding protein 1 (53BP1) is an important mediator in DDR and H4K20me2 function as 53BP1´s binding site at DSBs. R-loops and factors regulating R-loops are also involved in DNA-damage responses, while knock-down of such factors, including MMSET and DHX9, have been shown to decrease the global levels of R-loops in HeLa cells (Cristini et al., 2018; Cristini et al., 2019). Generally, perturbation in factors leading to a subsequently altered DDR, like MMSET, can be measured by γH2AX foci formation. Overall, MMSET seems to play an important role during induced DNA damage, by affecting several aspects of the response, like methylation of H4K20 and R-loop accumulation.

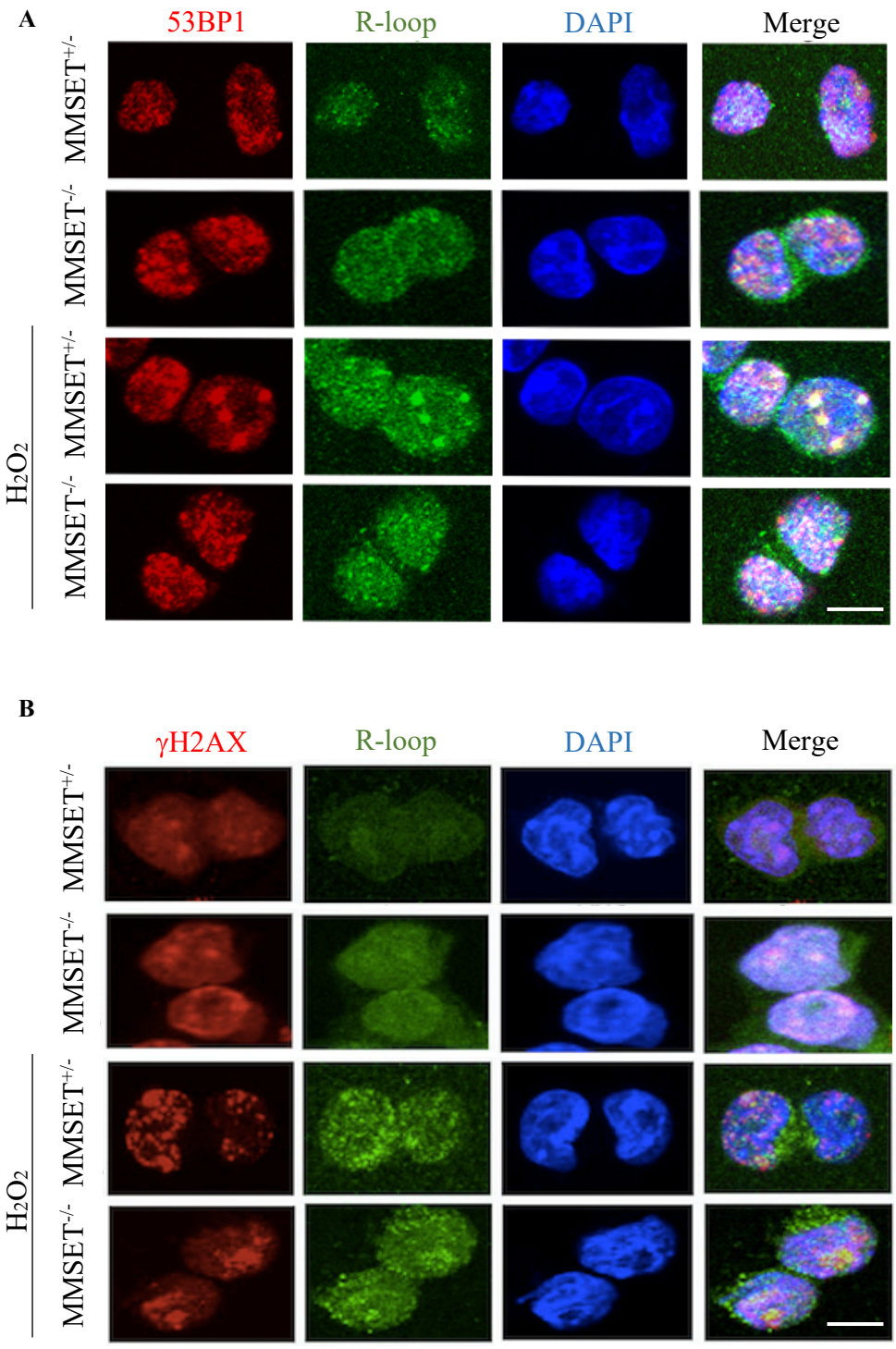


Figure 23. Immunostaining of MMSET$^{+/-}$ and MMSET$^{-/-}$ mES cells. Where indicated, the cells were treated with 200μM H_2O_2 for 1 hour to induce double-stranded breaks. The cells were cultured on gelatin-coated cover slips for 4 hours, before they were fixated with formaldehyde. Finally, the cells were stained with the indicated antibodies. DAPI staining (blue) was used as a control. Scale bar: 10μm.
(A) MMSET$^{+/-}$ and MMSET$^{-/-}$ cells were stained with 53BP1 (red) and R-loop (green) antibodies.
(B) MMSET$^{+/-}$ and MMSET$^{-/-}$ cells were stained with γH2AX (red) and R-loop (green) antibodies.

We wanted to validate the reported decrease in R-loop levels upon MMSET knock-down, with our heterozygous MMSET and MMSET null cell lines. We set up an immunofluorescence experiment and stained $MMSET^{+/-}$ and $MMSET^{-/-}$ mES cells with antibodies detecting R-loops, 53BP1, and γH2AX. As figure 23 shows, we observed that R-loop, 53BP1, and γH2AX signals were slightly stronger in MMSET deleted cells, but no major differences were observed. Based on the relatively unchanged signal upon MMSET deletion, oxidative stress was applied to further determine MMSETs role in DDR.

The cells were treated with 200μM hydrogen peroxide to induce single- and double-stranded breaks without great impact on the viability. Generally, this led to a higher foci formation of both γH2AX and R-loops, compared to non-treated cells (Fig. 23). Upon oxidative stress we detected a slightly lower foci formation of γH2AX and R-loops in MMSET null cells as compared to heterozygous MMSET cells. Furthermore, it is difficult to determine alterations in 53BP1 foci formation following H_2O_2 treatment, but it appeared to be a change in localization of the 53BP1 foci in MMSET null cells (Fig. 23A). An apparent limitation of these results is the lack of measurement regarding the fluorescent signal. Thus, no definite conclusion regarding foci formation, can be drawn based on these immunofluorescence results. However, the overall differences in foci formation of 53BP1 and γH2AX in $MMSET^{+/-}$ and $MMSET^{-/-}$ cells seems to be less visible upon oxidative stress, while the R-loop accumulation appears to be reduced in MMSET deleted cells upon oxidative stress.

5. Discussion

5.1 Phenotype characterization of the generated *Ythdf1-3* single- and double KO cells

Several studies report that YTHDF1-3 have dissimilar roles in cellular processes, despite their high sequence similarity. However, it is unclear how DF1-3 proteins mediate different functions, and if they bind the same or different m^6A sites (Zaccara et al., 2019). No study to date has examined how mES cells are affected upon single-, double-, or triple knock-outs of *Df1-3*, and their phenotypes. Hence, we wanted to generate single-, double-, and triple knock-out cells of *Ythdf1-3* in mES cells. To further investigate the phenotype of the generated *Df* knock-out cells, we performed proliferation and differentiation assays.

Single knock-outs of *Df1* and *Df3*, and double knock-outs of *Df1/2* and *Df1/3* were generated by creRT2 recombinase and CRISPR/Cas9 technology . An attempt to generate the triple *Df1-3* was performed, but it seemed like it was not possible to generate this triple knock-out with our approach. It appeared that the differentiation phenotype was highly affected by knock-out of the *Df* genes, while the proliferation phenotype was unaffected by the removal of the *Df* genes.

5.1.1 The strategy to generate *Ythdf1-3* knock-outs

Generation and validation of single- and double knock-outs

Studies have reported that spontaneous creERT2 activity can occur, thus resulting in knock-out without tamoxifen treatment (Kristianto et al., 2017). Upon western blot analysis of the *Df2* KO cells, a *Df1* KO was also detected even though these cells were not treated with tamoxifen (Fig. 14B). This implies that the knock-out of the *Df1* allele was due to leaky creERT2. Hence, all 37.4-creERT2 cells needed to be sorted as single cells, to ensure a homogenous population. Because of the leaky creERT2, it would most likely have been less time consuming and more convenient to generate all three *Df* KO cells by CRISPR/Cas9 genome editing. As a consequence of the leaky creERT2, the future experiments were performed in *Df1* and *Df3* single KOs, and *Df1/2* and *Df1/3* double KOs.

Western blot analysis was carried out to confirm that the tamoxifen treatment and CRISPR/Cas9 editing were executed successfully, while genotyping further validated the

knock-out of the *Df1* allele (Fig. 14 and S1C). Since western blot analysis detects protein expression, this method indicates that the gene is knocked out or non-functional. Further, genotyping and/or sequencing should be performed for all knock-outs to verify that the intended region has been deleted. Therefore, genotyping was performed on the *Df1* KO cells.

However, the chosen method for genotyping included whole-cell lysate and no step for DNA purification, before amplification with PCR. As a result, inconsistent signals were observed in the agarose gel pictures for a bulk of the genotyping analyses. Since these analyses provide important information, another approach that includes DNA purification would have been beneficial. Following, the project´s next priority would be to perform an improved genotyping for all generated single- and double knock-out cells.

An attempt to generate the triple Ythdf1-3 knock-out

We treated the *Df1/2* KO cells with sgRNAs for generation of a triple knock-out. Only a few colonies grew after single cell sorting of the transfected *Df1/2* KO, compared to previous single cells sorting of the *Df* knock-out cells. Additionally, the picked colonies appeared to have hampered proliferation due to reduced growth ability. Moreover, as shown in figure 14D, the triple *Df1-3* knock-out was not generated. This demonstrates that we were not able to generate the triple knock-out. So far, no studies have reported any attempts to generate the *Df1-3* KO. Hence, more investigation is needed to study whether the triple knock-out is lethal, or if it possibly can be generated with another approach. However, our results suggest that the DF proteins are compensating each other, since we were not able to generate the triple knock-out.

5.1.2 Preserved proliferation, but altered differentiation capacity

We wanted to examine alterations in phenotype upon deletion of *Df1-3* in mES cells. This proceeded in two stages: first with proliferation assay, and second with differentiation assay. The proliferation assay was carried out to determine possible growth defects in the generated *Df1, Df3*, and *Df1/2* KO cells. Since we did not observe any major alterations in the proliferation phenotype upon deletion of the *Df* genes, we wanted to further examine the phenotype by performing differentiation assays (Fig. 15). This could give valuable information about how or if the genes affect the ability to exit pluripotency and to generate the three germ layers. Since there are no existing studies of *Df1-3* deletions and their phenotype in mES cells, *Mettl3* knock-out studies provided comparative data.

Negligible alterations in proliferation capacity of the Ythdf1-3 knock-outs

During the proliferation assay no substantial differences was detected between the *Df* knock-outs compared to the control (Fig. 15C). A dramatically impaired proliferation phenotype, upon removal of two reader proteins, compared to the deletion of one protein was not observed either. There are no studies showing the effects upon *Df1-3* deletion in mES cells, however, there are studies revealing the effects upon depletion or overexpression in other cells than ES cells. Recently, the first papers revealing how DF proteins affect proliferation in HeLa and cancer cells have been published (Berlivet et al., 2019; Nishizawa et al., 2018; Shi et al., 2017; Zhong et al., 2019). These studies have either overexpressed, knocked down, or knocked out the *Df* genes. To summarize, these studies revealed alterations in the proliferation phenotype upon deviant DF protein expression.

Also, there are published data regarding the phenotype of ES cells upon the deletion of the m^{6}A writer, *Mettl3.* Reports of *Mettl3* KD or KO in ES cells have shown both slightly reduced and increased growth ability. Despite the fact that researchers have observed slightly reduced growth upon deletion of *Mettl3*, they all conclude that *Mettl3* is not required for the growth of mES cells (Batista et al., 2014; Geula et al., 2015). While removal of *Mettl3* will subsequently deplete m^{6}A marks, removal of the DF reader proteins will affect the stability and/or translation efficiency of m^{6}A marked transcripts (Geula et al., 2015). Based on *Mettl3* knock-out resulting in a slightly reduced growth, an altered growth rate upon deletion of the *Df* genes could be expected. Nevertheless, epigenetic and epitranscriptomic regulators are not often found to be required for the proliferation of ES cells, but rather, they are often found to be involved in the regulation of other factors, like the differentiation capacity toward different lineages. In that regard, the negligible growth defects upon *Df* deletion coincide with current literature regarding epitranscriptomic regulation in proliferation (Batista et al., 2014).

Altered differentiation capacity upon Ythdf1-3 deletion

To investigate the maintenance of pluripotency upon deletion of the *Ythdf* genes, we measured the mRNA expression of the pluripotency gene *Nanog,* and conducted images of the *Df* KO cell morphology when cultured in ES cell media. The pluripotency seemed to be preserved in all the *Df* KO cells as shown from unaltered *Nanog* expression compared to the control (Fig. 15B). However, different morphology was observed for the control and knock-out cells during cultivation, as figure 15A depicts, which can be an indication of lost pluripotency phenotype. Furthermore, it should be mentioned that the altered morphology in the *Df* knock-out cells needs

more research to draw any conclusion. The cells were stored in a busy incubator, and the disruptions following this could affect their morphology appearance. In that regard, the mRNA expression of *Nanog* is more reliable. There is no published data regarding how *Df1-3* knock-outs affects the ability to express pluripotency markers, but it is reported that Mettl3$^{-/-}$ cells retain normal expression of the pluripotency markers upon differentiation. Hence, our results regarding *Nanog* mRNA expression levels are in line with previous studies about *Mettl3* knock-out.

To further look into the morphology or arguably their pluripotency changes, the cells need to be cultured in a separate incubator to remove the disturbing factors. It would also be valuable to analyze several m^{6}A marked pluripotency genes in addition to *Nanog*, to further analyze the effects on pluripotency upon DF protein deletion. Conclusively, our results imply that despite the deletion of the different *Df* genes, pluripotency is preserved in the mES.

In an attempt to examine the differentiation capacity, NSC differentiation was carried out. We observed high cell death of the *Df* knock-out cells, which indicated that they could not differentiate efficiently (Fig. 16A). By the addition of EGF/FGFb the vast cell death observed for the *Df* KO cells upon 2i/LIF removal, was further enhanced. Simultaneously the control cells were able to adapt to the conditions. As a consequence of adding EGF/FGFb, further differentiation toward the ectoderm lineage occurs in the cells, and the cells that have initiated differentiation toward mesendoderm or endoderm lineages will die. This expected cell death is in line with the observations of the control, illustrated in figure 16A, and the knock-out cells show a considerably lower ability to differentiate into the neuronal lineage. It has earlier been reported that Mettl3$^{-/-}$ ES cells did not efficiently differentiate into mature neurons (Geula et al., 2015), which is consistent with our observations in the *Df* deleted cells during differentiation towards NSCs.

To further address the differentiation capacity, we performed EB differentiation to avoid passage of the cells, because of the subsequently high cell death (Fig. 17A). Upon EB differentiation, we detected exit from pluripotency after three days, regardless of the *Df1-3* KOs (Fig. 17C). EB differentiation studies with *Mettl3* knock-outs have shown that exit from pluripotency is blocked even after 21 days (Geula et al., 2015). Even though we did not replicate the previously reported blocked pluripotency exit upon *Mettl3* knock-out, our results suggest that *Df* KO cells are able to differentiate efficiently.

Images were conducted of the EBs at d6 and show morphology varieties (Fig. 17B). The *Df* KO EBs appears to undergo cavitation and maturation, while the control EBs remain densely packed. Based on these observations, the *Df* KOs appears to be further differentiated compared

to the control cells. On the other hand, only a slightly altered expression pattern of some differentiation markers was observed. All in all, our results indicate that the differentiation capacity is altered upon the deletion of all tested *Df* KOs. This correlates with the knowledge that epigenetic and epitranscriptomic regulators often are involved in the regulation of the differentiation capacity in ES cells, and that many regulators of pluripotency and lineage commitment are m^6A marked (Batista et al., 2014).

5.2 Gene regulation of MMSET in R-loop biology

We validated MMSET and R-loops as interaction partners, and DHX9, DDX3, and DDX5 as common interaction partners of MMSET and R-loops. MMSET deletion showed no delayed pluripotency exit or no major alterations in the growth capacity. However, we detected alterations in the commitment ability into the different lineages. Furthermore, differences in foci formation of the DNA damage proteins, 53BP1 and γH2AX appeared to be reduced upon oxidative stress in $MMSET^{+/-}$ and $MMSET^{-/-}$. On the contrary, there seemed to be less accumulation of R-loops in MMSET null cells upon oxidative stress. Additionally, R-loop enrichment in *Gata4* was confirmed in heterozygous MMSET, and further also detected in MMSET deleted cells.

5.2.1 Dysregulation in differentiation markers upon MMSET deletion

Recently, the first papers revealing the differentiation capacity upon MMSET deletion have been published. It has been shown that removal of MMSET will entrap mES cells in a pluripotent state, due to a detected delayed pluripotency exit and impaired ability to upregulate the mesendoderm genes (Tian et al., 2019). Hence, we performed three differentiation assays with our two mES cell lines, heterozygous MMSET and MMSET null cells, to determine whether the deletion of MMSET affects the mES cells´ ability to differentiate.

Upon induction of differentiation, no delayed exit from pluripotency was observed in MMSET deleted cells, while the mesendodermal genes *Gata4* and *Gata6* were consistently downregulated in MMSET null cells during all differentiation assays (Fig. 19). The impaired ability of mesendoderm specification Tian et al., 2019 described, was successfully recapitulated for *Gata4* and *Gata6*, and coincide with the theory that MMSET is crucial for activating mesendodermal regulator genes. MMSET has been shown to interact with the histone acetyltransferase BRD4, and together they activate the expression of the mesendodermal

regulator genes by binding to their enhancers (Sarai et al., 2013; Tian et al., 2019). Furthermore, the downregulation of mesendodermal genes has been shown to delay exit from pluripotency. As opposed to their results, we did not detect any delayed exit from pluripotency in MMSET null cells. Based on the current knowledge, the preserved pluripotency exit observed in our experiments, might be explained by the maintained expression of the mesendodermal genes *Foxa2* and *Brachyury* in MMSET null cells (Fig. 19A/B).

The ectodermal regulator genes showed fluctuating expression kinetics in heterozygous MMSET and MMSET null cells (Fig. 19). Differentiation of mES cells into EBs revealed no alterations in the expression pattern of ectodermal genes post MMSET deletion (Tian et al., 2019). In general, our observation regarding ectodermal expression patterns did not coincide with their findings. During the differentiation of normal neural cells, a decreased MMSET expression has been reported (Hudlebusch et al., 2011). Therefore, the influence of MMSET in ectoderm differentiation will be mainly at the beginning of the differentiation, and thereafter decrease alongside with further differentiation. This might explain why we observe a fluctuating expression pattern before the tendency of normalization. However, this field is barely investigated and needs further research to determine the role of MMSET in ectodermal differentiation.

We observed a dysregulation of all differentiation markers upon MMSET deletion in the three differentiation assays. We were able to reproduce impaired mesendoderm mRNA expression of *Gata4* and *Gata6* in MMSET null cells during the three differentiation assays, while this was not the case for the other tested mesendodermal genes (Fig. 19). A possible reason why downregulation was observed for *Gata4* and *Gata6,* and not the other mesendodermal genes, might be explained by differences in the culture conditions. It is well known that even small variations in culture condition and environment can alter gene expression patterns in ES cells (Mulas et al., 2019). Finally, we recapitulated the downregulation of *Gata4* and *Gata6*, while mRNA expression of the other mesendodermal and pluripotency markers do not coincide with the predicted results in this study.

5.2.2 Confirmation of some common interaction partners of R-loops and MMSET

A mass spectrometry study performed by the group in 2006 showed MMSET and R-loops as interaction partners, in addition to many other overlapping interaction partners between MMSET and R-loops. Among the common interaction partners of MMSET and R-loops, the

top interaction partners of MMSET are the RNA helicases DHX9, DDX3, and DDX5, which later were validated by other researchers with DRIP (R-loop IP) western blot analysis (Cristini et al., 2018). We wanted to validate MMSET and R-loops as interaction partners, as well as DHX9, DDX3, and DDX5 as overlapping partners of both MMSET and R-loops. We performed DRIP western blotting and MMSET immunoprecipitation WB, with our MMSET$^{+/-}$ and MMSET$^{-/-}$ mES cells, to validate these interaction partners.

R-loop and MMSET immunoprecipitation western blotting confirmed DHX9, DDX3, and DDX5 as interaction partners of both MMSET and R-loops, and additionally, MMSET and R-loops were validated as interaction partners with each other (Fig. 21C/D). These results coincides with the group´s findings in 2006, and the current literature (Cristini et al., 2018). Furthermore, our western blot analysis shows that DHX9, DDX3, and DDX5 were pulled down by DRIP in both heterozygous MMSET and MMSET null cells (Fig. 21D). Our results imply that the presence of MMSET is not required for the interaction between either of the validated RNA helicases and R-loops, since all of them were pulled down by DRIP in the MMSET null cells. These results indicate that the binding of the RNA helicases to R-loops are independent of MMSET. Nevertheless, these results were not surprising since it is uncommon that the removal of one protein largely alter a whole protein complex (D. P. Pandey, personal communication, May 2nd, 2020). Moreover, the R-loop immunoprecipitation was observed to easily be affected by human bias. A limitation in these results, is the lack of bands in the MMSET input for the DRIP western blot, but due to time limitations this experiment was not repeated (Fig. 21D). Overall, we observed that DHX9, DDX3, and DDX5 binds to R-loops regardless of whether MMSET is present or not. Additionally, our results confirmed MMSET and R-loop as interaction partners, and DHX9, DDX3, and DDX5 as overlapping interaction partners of both MMSET and R-loops.

5.2.3 MMSET affects R-loop accumulation upon oxidative stress

Immunofluorescence was performed with the two mES cell lines, MMSET$^{+/-}$ and MMSET$^{-/-}$, to investigate how MMSET affects foci formation of R-loops, in addition to the DNA damage results 53BP1 and γH2AX. The observed differences in foci formation of R-loops, 53BP1, and γH2AX upon MMSET deletion were small (Fig. 23). Since it has been reported that both MMSET and R-loops are involved in DNA-damage responses, we decided to give the cells oxidative stress (Hajdu et al., 2011; Pei et al., 2011). This was performed to further investigate changes upon MMSET deletion, due to small differences in R-loop accumulation in MMSET$^{+/-}$

and MMSET$^{-/-}$ before treatment, which might indicate that MMSET does not have an essential role in R-loop accumulation at steady-state conditions.

Based on the MMSET protein interaction with R-loops, we wanted to examine the foci formation of R-loops upon the deletion of MMSET. Before induction of DNA damage, the R-loop signal was similar in heterozygous MMSET and MMSET null cells (Fig. 23). However, after treatment with H_2O_2, we observe a stronger signal in both heterozygous and MMSET null cells. To some extent, the signal appears to be stronger in heterozygous MMSET cells (Fig. 23). As recently reviewed, SSB and DSB can be a driving force in the formation of R-loops (Crossley et al., 2019; García-Muse & Aguilera, 2019). This is highly correlated to our findings. Overall, it appears that MMSET deletion affects R-loops foci formation upon oxidative stress. Further, IF staining for MMSET could provide important information regarding colocalization between R-loops and MMSET in the cells.

The observed foci formation of γH2AX was slightly lower in MMSET null cells compared to heterozygous MMSET cells upon induced DNA damage (Fig. 23B). It has been shown that downregulation of MMSET does not decrease DNA-damage induced foci formation of γH2AX significantly. This is because mainly the BRCT domain of MDC1, and not MMSET, is required for binding of γH2AX to DNA damage sites (Pei et al., 2011). Hence, our results are in agreement with the literature in this field.

The difference in foci formation of 53BP1 in MMSET$^{+/-}$ and MMSET$^{-/-}$ cells after H_2O_2 treatment was negligible (Fig. 23A). Earlier, Pei et al., 2011, have reported that 53BP1 mediate DNA-damage response by binding to H4K20me2. Methylation of H4K20 increases when double-stranded breaks are introduced, and this methylation is conducted by MMSET. Since the removal of MMSET alters the H4K20 methylation pattern, and we did not observe any differences upon MMSET deletion, our results are not in agreement with what we expected. On the other hand, 53BP1 foci formation is also regulated by another protein termed RNF8 (Huen et al., 2007; Kolas et al., 2007; Mailand et al., 2007). The recruitment of RNF8 to double-stranded breaks is not affected by the downregulation of MMSET, and this might explain why we observe more or less equal foci formation of 53BP1 in the two cell lines (Pei et al., 2011).

When performing immunofluorescence staining, single cells are needed to properly visualize the signal of interest. The difficulties by using mES cell in immunofluorescence is their high cell proliferation and ES cells´ typical ability to form ball structures. Further, an improvement when performing immunofluorescence with ES cells could be to use different conditions while seeding the ES cells for example, fibronectin or other commercially available

reagents. This could improve the ability to ensure single cell populations during immunofluorescence, while still maintaining a high cell number.

We observed altered foci formation of R-loops, γH2AX, and 53BP1 upon oxidative stress in MMSET$^{+/-}$ and MMSET$^{-/-}$ mES cells (Fig. 23). Overall, by comparison of the foci formation of R-loops, γH2AX, and 53BP1 in non-treated and treated heterozygous MMSET and MMSET null cells, the effects seem smaller between the cell lines for γH2AX and 53BP1 foci formation, while R-loop accumulation is reduced in MMSET null cells upon oxidative stress. This indicates that removal of MMSET affects the foci formation of R-loops upon oxidative stress. Furthermore, immunofluorescence staining for MMSET could provide important and interesting information regarding the colocalization between R-loops and MMSET in the cells.

5.2.4 MMSET appears to not affect R-loop enrichment in *Gata4*

R-loops are revealed to be enriched at several polycomb group repressed genes, including *Gata4* (Skourti-Stathaki et al., 2019). We observed altered differentiation abilities of *Gata4* during all differentiation assays upon MMSET deletion (Fig. 19). Thus, we wanted to validate the enrichment of R-loops at the PcG repressed gene *Gata4*, and additionally investigate the effects of MMSET deletion upon this enrichment, by carrying out DRIP qPCR.

We observed R-loop enrichment in the PcG repressed gene *Gata4*, independent of MMSET (Fig. 22). Our results indicate that there are little or negligible effects upon the deletion of MMSET. Despite the previously reported binding of MMSET to enhancers of *Gata4* for activation of its expression, no such regulation of *Gata4* is observed during R-loop immunoprecipitation. However, our results are in agreement with current literature showing that R-loops are enriched in *Gata4.* To our knowledge, reports implying if MMSET influence this enrichment are lacking. Finally, it would provide interesting results to examine the enrichment of R-loops in *Gata4*, upon induced differentiation of both heterozygous MMSET and MMSET deleted cells.

5.3 Concluding remarks and future perspectives

Concluding remarks

The study aimed to investigate the contribution of epigenetic and epitranscriptomic modifications in gene regulation of mES cells, by approaching two different aspects. The first aim was to investigate how the reader proteins, YTHDF1-3, impact cellular processes in embryonic stem cells. We approached this by generating *Ythdf1-3* knock-outs and characterized their phenotypes. So far, no studies show the effects upon *Ythdf1-3* deletion in embryonic stem cells. By generation and characterization of single- and double *Ythdf1-3* knock-out cells, our results revealed that DF1-3 have no major impact on proliferation but plays a role in the differentiation ability of embryonic stem cells. Hence, our results contributes to a better understanding of how *Ythdf1-3* impact cellular processes in embryonic stem cells.

The second aim of this study was to characterize the role of MMSET in R-loop biology. We confirmed MMSET as an interaction partner of R-loops, and DHX9, DDX3, and DXX5 as common interaction partners of MMSET and R-loops. Furthermore, we wanted to investigate the role of MMSET in R-loop accumulation and in DNA-damage responses, by performing immunofluorescence staining upon oxidative stress. The overall differences in foci formation of the DNA damage proteins 53BP1 and γH2AX in $MMSET^{+/-}$ and $MMSET^{-/-}$ cells seemed to be reduced upon oxidative stress, while less R-loop accumulation was observed in MMSET deleted cells upon oxidative stress. Finally, we examined the R-loop enrichment in PcG active and repressed genes in presence and absence of MMSET. We did not detect any major differences in R-loop enrichment upon MMSET deletion in neither PcG active nor repressed genes, but we confirmed R-loop enrichment in the PcG repressed gene *Gata4*.

Future YTHDF experiments

The generated *Df1-3* KO provide a valuable model system for further in-depth investigation of DF1-3 target specificity. As the first step for further investigation, the *Df2* and *Df2/3* knock-outs should be generated and validated by CRISPR/Cas9 technology and western blot analyses or genotyping, respectively. Furthermore, RNA-sequencing could be carried out to investigate changes in mRNA upon *Df1-3* KO, while m^6A-sequencing could provide information about m^6A distribution and sites in mRNA. m^6A-sequencing is a unbiased localization analysis, where RNA immunoprecipitation (RIP) is performed on fragmented poly(A) RNA with anti-m^6A antibodies, followed by RNA-seq. The RNA-seq will give information about enriched m^6A positive RNA fragments, while the input consists of non-immunoprecipitated RNA fragments

(Molinie & Giallourakis, 2017). Research with this method could elucidate m^6A distribution and levels. Since we generated the single- and double *Df* KOs, it would be possible to investigate how the m^6A levels are affected by double knock-outs compared to single knock-outs. Furthermore, this method could determine if the deletion of one DF protein is more critical than others, due to alterations in m^6A pattern and further impact regarding cellular functions.

To gain information about the target mRNA bound by the DF1-3 proteins, an appropriate method is crosslinking immunoprecipitation sequencing, followed by sequencing using the antibodies against the individual proteins. If the protein-protein interaction partners of DF1-3 are of interest, a GFP-pulldown on endogenously tagged DF1-3 mES cells may provide valuable information. A further continuation of this study could be to investigate the success of chimera formation following injection of the different knock-out cells into the blastocyst. Finally, this could provide insights regarding the roles of the different *Df* genes in developmental processes, and if they are essential for early embryonic development.

Future MMSET projects

To further determine the role of MMSET in R-loop biology, several experiments could give valuable information. It would be interesting to further gain information regarding the colocalization between MMSET and R-loops. This could be carried out by immunofluorescence staining with MMSET and R-loop specific antibodies. In addition to the $MMSET^{+/-}$ and $MMSET^{-/-}$ mES cells, the E14 mES cell line could be included as a wild type, to look at the effects of knocking out both one and two alleles. Furthermore, it would be interesting to elucidate whether MMSET affects R-loop enrichment in *Gata4* during differentiation of heterozygous MMSET and MMSET null cells. This could be performed by R-loop immunoprecipitation qPCR on differentiating cells, with the highly specific R-loop antibody, S9.6. Moreover, it would be interesting to reveal H3K36me2/3 and H4K20me2 abundance in heterozygous MMSET and MMSET null cells compared to the wild type E14 mES cell line, by MMSET chromatin immunoprecipitation. Overall, R-loop immunoprecipitation sequencing with the S9.6 antibody could further provide information about R-loops site distribution in the genome, and MMSETs possible impact on cellular functions and gene expression.

6. References

Aguilera, A. & García-Muse, T. (2012). R loops: from transcription byproducts to threats to genome stability. *Molecular cell*, 46 (2): 115-124.

Aguilera, A. & Gómez-González, B. (2017). DNA–RNA hybrids: the risks of DNA breakage during transcription. *Nature structural & molecular biology*, 24 (5): 439.

Aik, W., Scotti, J. S., Choi, H., Gong, L., Demetriades, M., Schofield, C. J. & McDonough, M. A. (2014). Structure of human RNA N6-methyladenine demethylase ALKBH5 provides insights into its mechanisms of nucleic acid recognition and demethylation. *Nucleic Acids Research*, 42 (7): 4741-4754. doi: 10.1093/nar/gku085.

Alarcón, C. R., Goodarzi, H., Lee, H., Liu, X., Tavazoie, S. & Tavazoie, S. F. (2015). HNRNPA2B1 Is a Mediator of m(6)A-Dependent Nuclear RNA Processing Events. *Cell*, 162 (6): 1299-308. doi: 10.1016/j.cell.2015.08.011.

Allshire, R. C. & Madhani, H. D. (2018). Ten principles of heterochromatin formation and function. *Nat Rev Mol Cell Biol*, 19 (4): 229-244. doi: 10.1038/nrm.2017.119.

Avilion, A. A., Nicolis, S. K., Pevny, L. H., Perez, L., Vivian, N. & Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes & development*, 17 (1): 126-140.

Baker, L. A., Allis, C. D. & Wang, G. G. (2008). PHD fingers in human diseases: Disorders arising from misinterpreting epigenetic marks. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 647 (1): 3-12. doi: https://doi.org/10.1016/j.mrfmmm.2008.07.004.

Bannister, A. J. & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell research*, 21 (3): 381-395.

Batista, P. J., Molinie, B., Wang, J., Qu, K., Zhang, J., Li, L., Bouley, D. M., Lujan, E., Haddad, B., Daneshvar, K., et al. (2014). m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell*, 15 (6): 707-19. doi: 10.1016/j.stem.2014.09.019.

Beemon, K. & Keith, J. (1977). Localization of N6-methyladenosine in the Rous sarcoma virus genome. *Journal of Molecular Biology*, 113 (1): 165-179. doi: https://doi.org/10.1016/0022-2836(77)90047-X.

Bennett, R. L., Swaroop, A., Troche, C. & Licht, J. D. (2017). The Role of Nuclear Receptor–Binding SET Domain Family Histone Lysine Methyltransferases in Cancer. *Cold Spring Harbor Perspectives in Medicine*, 7 (6). doi: 10.1101/cshperspect.a026708.

Berlivet, S., Scutenaire, J., Deragon, J.-M. & Bousquet-Antonelli, C. (2019). Readers of the m6A epitranscriptomic code. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1862 (3): 329-342.

Bestor, T. H. & Verdine, G. L. (1994). DNA methyltransferases. *Current Opinion in Cell Biology*, 6 (3): 380-389. doi: https://doi.org/10.1016/0955-0674(94)90030-2.

Bodi, Z., Button, J. D., Grierson, D. & Fray, R. G. (2010). Yeast targets for mRNA methylation. *Nucleic Acids Res*, 38 (16): 5327-35. doi: 10.1093/nar/gkq266.

Bokar, J. A., Rath-Shambaugh, M. E., Ludwiczak, R., Narayan, P. & Rottman, F. (1994). Characterization and partial purification of mRNA N6-adenosine methyltransferase from HeLa cell nuclei. Internal mRNA methylation requires a multisubunit complex. *J Biol Chem*, 269 (26): 17697-704.

Bokar, J. A., Shambaugh, M. E, Polayes, D., Matera, A. G., Rottman, F. M. (1997). Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. *Rna*, 3 (11): 1233-47.

Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., Guenther, M. G., Kumar, R. M., Murray, H. L. & Jenner, R. G. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. *cell*, 122 (6): 947-956.

Bradley, A., Evans, M., Kaufman, M. H. & Robertson, E. (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature*, 309 (5965): 255-256.

Brons, I. G., Smithers, L. E., Trotter, M. W., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S. M., Howlett, S. K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R. A., et al. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature*, 448 (7150): 191-5. doi: 10.1038/nature05950.

Brookes, E., de Santiago, I., Hebenstreit, D., Morris, K. J., Carroll, T., Xie, S. Q., Stock, J. K., Heidemann, M., Eick, D., Nozaki, N., et al. (2012). Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. *Cell Stem Cell*, 10 (2): 157-70. doi: 10.1016/j.stem.2011.12.017.

Bujnicki, J. M., Feder, M., Radlinska, M. & Blumenthal, R. M. (2002). Structure prediction and phylogenetic analysis of a functionally diverse family of proteins homologous to the MT-A70 subunit of the human mRNA:m(6)A methyltransferase. *J Mol Evol*, 55 (4): 431-44. doi: 10.1007/s00239-002-2339-8.

Cai, C. & Grabel, L. (2007). Directing the differentiation of embryonic stem cells to neural stem cells. *Dev Dyn*, 236 (12): 3255-66. doi: 10.1002/dvdy.21306.

Cartwright, P., McLean, C., Sheppard, A., Rivett, D., Jones, K. & Dalton, S. (2005). LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. *Development*, 132 (5): 885-896.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S. & Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell*, 113 (5): 643-55.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L. & Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. *Nature*, 450 (7173): 1230-4. doi: 10.1038/nature06403.

Chédin, F. (2016). Nascent connections: R-loops and chromatin patterning. *Trends in Genetics*, 32 (12): 828-838.

Chen, P. B., Chen, H. V., Acharya, D., Rando, O. J. & Fazzio, T. G. (2015). R loops regulate promoter-proximal chromatin architecture and cellular differentiation. *Nature structural & molecular biology*, 22 (12): 999.

Chen, T., Hao, Y.-J., Zhang, Y., Li, M.-M., Wang, M., Han, W., Wu, Y., Lv, Y., Hao, J., Wang, L., et al. (2015). m6A RNA Methylation Is Regulated by MicroRNAs and Promotes Reprogramming to Pluripotency. *Cell Stem Cell*, 16 (3): 289-301. doi: https://doi.org/10.1016/j.stem.2015.01.016.

Chen, W., Zhang, L., Zheng, G., Fu, Y., Ji, Q., Liu, F., Chen, H. & He, C. (2014). Crystal structure of the RNA demethylase ALKBH5 from zebrafish. *FEBS Letters*, 588 (6): 892-898. doi: https://doi.org/10.1016/j.febslet.2014.02.021.

Chen, Y., McGee, J., Chen, X., Doman, T. N., Gong, X., Zhang, Y., Hamm, N., Ma, X., Higgs, R. E., Bhagwat, S. V., et al. (2014). Identification of Druggable Cancer Driver Genes Amplified across TCGA Datasets. *PLOS ONE*, 9 (5): e98293. doi: 10.1371/journal.pone.0098293.

Chesi, M., Nardini, E., Lim, R. S., Smith, K. D., Kuehl, W. M. & Bergsagel, P. L. (1998). The t (4; 14) translocation in myeloma dysregulates both FGFR3 and a novel gene, MMSET, resulting in IgH/MMSET hybrid transcripts. *Blood, The Journal of the American Society of Hematology*, 92 (9): 3025-3034.

Chory, E. J., Calarco, J. P., Hathaway, N. A., Bell, O., Neel, D. S. & Crabtree, G. R. (2019). Nucleosome turnover regulates histone methylation patterns over the genome. *Molecular cell*, 73 (1): 61-72. e3.

Clancy, M. J., Shambaugh, M. E., Timpte, C. S. & Bokar, J. A. (2002). Induction of sporulation in Saccharomyces cerevisiae leads to the formation of N6-methyladenosine in mRNA: a potential mechanism for the activity of the IME4 gene. *Nucleic Acids Research*, 30 (20): 4509-4518. doi: 10.1093/nar/gkf573.

Cohen, S., Puget, N., Lin, Y.-L., Clouaire, T., Aguirrebengoa, M., Rocher, V., Pasero, P., Canitrot, Y. & Legube, G. (2018). Senataxin resolves RNA: DNA hybrids forming at DNA double-strand breaks to prevent translocations. *Nature communications*, 9 (1): 1-14.

Cristini, A., Groh, M., Kristiansen, M. S. & Gromak, N. (2018). RNA/DNA hybrid interactome identifies DXH9 as a molecular player in transcriptional termination and R-loop-associated DNA damage. *Cell reports*, 23 (6): 1891-1905.
Cristini, A., Ricci, G., Britton, S., Salimbeni, S., Huang, S.-y. N., Marinello, J., Calsou, P., Pommier, Y., Favre, G., Capranico, G., et al. (2019). Dual Processing of R-Loops and Topoisomerase I Induces Transcription-Dependent DNA Double-Strand Breaks. *Cell Reports*, 28 (12): 3167-3181.e6. doi: https://doi.org/10.1016/j.celrep.2019.08.041.
Crossley, M. P., Bocek, M. & Cimprich, K. A. (2019). R-loops as cellular regulators and genomic threats. *Molecular cell*, 73 (3): 398-411.
Desai, D. & Pethe, P. (2020). Polycomb repressive complex 1: Regulators of neurogenesis from embryonic to adult stage. *Journal of cellular physiology*, 235 (5): 4031-4045.
Desrosiers, R., Friderici, K. & Rottman, F. (1974). Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *Proceedings of the National Academy of Sciences*, 71 (10): 3971-3975.
Dhote, V., Sweeney, T. R., Kim, N., Hellen, C. U. & Pestova, T. V. (2012). Roles of individual domains in the function of DHX29, an essential factor required for translation of structured mammalian mRNAs. *Proceedings of the National Academy of Sciences*, 109 (46): E3150-E3159.
Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., et al. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature*, 485 (7397): 201-206. doi: 10.1038/nature11112.
Du, H., Zhao, Y., He, J., Zhang, Y., Xi, H., Liu, M., Ma, J. & Wu, L. (2016). YTHDF2 destabilizes m(6)A-containing RNA through direct recruitment of the CCR4-NOT deadenylase complex. *Nat Commun*, 7: 12626. doi: 10.1038/ncomms12626.
Dunn, S.-J., Martello, G., Yordanov, B., Emmott, S. & Smith, A. (2014). Defining an essential transcription factor program for naive pluripotency. *Science*, 344 (6188): 1156-1160.
Dunn, S. J., Li, M. A., Carbognin, E., Smith, A. & Martello, G. (2019). A common molecular logic determines embryonic stem cell self-renewal and reprogramming. *The EMBO journal*, 38 (1).
Evans, M. J. & Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature*, 292 (5819): 154-156. doi: 10.1038/292154a0.
Feil, R., Wagner, J., Metzger, D. & Chambon, P. (1997). Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. *Biochemical and biophysical research communications*, 237 (3): 752-757.
García-Muse, T. & Aguilera, A. (2019). R loops: from physiological to pathological roles. *Cell*.
Geula, S., Moshitch-Moshkovitz, S., Dominissini, D., Mansour, A. A., Kol, N., Salmon-Divon, M., Hershkovitz, V., Peer, E., Mor, N., Manor, Y. S., et al. (2015). m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science*, 347 (6225): 1002-1006. doi: 10.1126/science.1261417.
Gidekel, S., Pizov, G., Bergman, Y. & Pikarsky, E. (2003). Oct-3/4 is a dose-dependent oncogenic fate determinant. *Cancer cell*, 4 (5): 361-370.
Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I. & Chédin, F. (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Molecular cell*, 45 (6): 814-825.
Godini, R., Lafta, H. Y. & Fallahi, H. (2018). Epigenetic modifications in the embryonic and induced pluripotent stem cells. *Gene Expr Patterns*, 29: 1-9. doi: 10.1016/j.gep.2018.04.001.
Goll, M. G., Kirpekar, F., Maggert, K. A., Yoder, J. A., Hsieh, C.-L., Zhang, X., Golic, K. G., Jacobsen, S. E. & Bestor, T. H. (2006). Methylation of tRNA^{Asp} by the DNA Methyltransferase Homolog Dnmt2. *Science*, 311 (5759): 395-398. doi: 10.1126/science.1120976.
Grunseich, C., Wang, I. X., Watts, J. A., Burdick, J. T., Guber, R. D., Zhu, Z., Bruzel, A., Lanman, T., Chen, K. & Schindler, A. B. (2018). Senataxin mutation reveals how R-

loops promote transcription by blocking DNA methylation at gene promoters. *Molecular cell*, 69 (3): 426-437. e7.
Gu, T.-P., Guo, F., Yang, H., Wu, H.-P., Xu, G.-F., Liu, W., Xie, Z.-G., Shi, L., He, X., Jin, S.-g., et al. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature*, 477 (7366): 606-610. doi: 10.1038/nature10443.
Gulati, P., Avezov, E., Ma, M., Antrobus, R., Lehner, P., O'Rahilly, S. & Yeo, G. S. (2014). Fat mass and obesity-related (FTO) shuttles between the nucleus and cytoplasm. *Biosci Rep*, 34 (5). doi: 10.1042/bsr20140111.
Haines, T. R., Rodenhiser, D. I. & Ainsworth, P. J. (2001). Allele-Specific Non-CpG Methylation of the Nf1 Gene during Early Mouse Development. *Developmental Biology*, 240 (2): 585-598. doi: https://doi.org/10.1006/dbio.2001.0504.
Hajdu, I., Ciccia, A., Lewis, S. M. & Elledge, S. J. (2011). Wolf–Hirschhorn syndrome candidate 1 is involved in the cellular response to DNA damage. *Proceedings of the National Academy of Sciences*, 108 (32): 13130-13134. doi: 10.1073/pnas.1110081108.
Hamperl, S. & Cimprich, K. A. (2014). The contribution of co-transcriptional RNA: DNA hybrid structures to DNA damage and genome instability. *DNA repair*, 19: 84-94.
Hamperl, S., Bocek, M. J., Saldivar, J. C., Swigut, T. & Cimprich, K. A. (2017). Transcription-replication conflict orientation modulates R-loop levels and activates distinct DNA damage responses. *Cell*, 170 (4): 774-786. e19.
Han, Z., Niu, T., Chang, J., Lei, X., Zhao, M., Wang, Q., Cheng, W., Wang, J., Feng, Y. & Chai, J. (2010). Crystal structure of the FTO protein reveals basis for its substrate specificity. *Nature*, 464 (7292): 1205-1209. doi: 10.1038/nature08921.
He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., et al. (2011). Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science*, 333 (6047): 1303-1307. doi: 10.1126/science.1210944.
Hsu, P. J., Zhu, Y., Ma, H., Guo, Y., Shi, X., Liu, Y., Qi, M., Lu, Z., Shi, H., Wang, J., et al. (2017). Ythdc2 is an N(6)-methyladenosine binding protein that regulates mammalian spermatogenesis. *Cell Res*, 27 (9): 1115-1127. doi: 10.1038/cr.2017.99.
Huang, G., Ye, S., Zhou, X., Liu, D. & Ying, Q.-L. (2015). Molecular basis of embryonic stem cell self-renewal: from signaling pathways to pluripotency network. *Cellular and molecular life sciences*, 72 (9): 1741-1757.
Huang, H., Weng, H., Zhou, K., Wu, T., Zhao, B. S., Sun, M., Chen, Z., Deng, X., Xiao, G., Auer, F., et al. (2019). Histone H3 trimethylation at lysine 36 guides m(6)A RNA modification co-transcriptionally. *Nature*, 567 (7748): 414-419. doi: 10.1038/s41586-019-1016-7.
Hudlebusch, H. R., Skotte, J., Santoni-Rugiu, E., Zimling, Z. G., Lees, M. J., Simon, R., Sauter, G., Rota, R., De Ioris, M. A. & Quarto, M. (2011). MMSET is highly expressed and associated with aggressiveness in neuroblastoma. *Cancer research*, 71 (12): 4226-4235.
Huen, M. S., Grant, R., Manke, I., Minn, K., Yu, X., Yaffe, M. B. & Chen, J. (2007). RNF8 transduces the DNA-damage signal via histone ubiquitylation and checkpoint protein assembly. *Cell*, 131 (5): 901-914.
Imamura, T., Kerjean, A., Heams, T., Kupiec, J.-J., Thenevin, C. & Pàldi, A. (2005). Dynamic CpG and Non-CpG Methylation of the Peg1/Mest Gene in the Mouse Oocyte and Preimplantation Embryo. *Journal of Biological Chemistry*, 280 (20): 20171-20175. doi: 10.1074/jbc.M501749200.
Ito, S., D'Alessio, A. C., Taranova, O. V., Hong, K., Sowers, L. C. & Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, 466 (7310): 1129-1133. doi: 10.1038/nature09303.
Iwanami, Y. & Brown, G. M. (1968). Methylated bases of ribosomal ribonucleic acid from HeLa cells. *Archives of Biochemistry and Biophysics*, 126 (1): 8-15. doi: https://doi.org/10.1016/0003-9861(68)90553-5.
Jacques-Fricke, B. T. & Gammill, L. S. (2014). Neural crest specification and migration independently require NSD3-related lysine methyltransferase activity. *Molecular Biology of the Cell*, 25 (25): 4174-4186. doi: 10.1091/mbc.e13-12-0744.

Jain, D., Puno, M. R., Meydan, C., Lailler, N., Mason, C. E., Lima, C. D., Anderson, K. V. & Keeney, S. (2018). ketu mutant mice uncover an essential meiotic function for the ancient RNA helicase YTHDC2. *eLife*, 7: e30919. doi: 10.7554/eLife.30919.

Jeltsch, A. & Jurkowska, R. Z. (2016). Allosteric control of mammalian DNA methyltransferases–a new regulatory paradigm. *Nucleic acids research*, 44 (18): 8556-8575.

Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., Yi, C., Lindahl, T., Pan, T., Yang, Y.-G., et al. (2011). N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nature Chemical Biology*, 7 (12): 885-887. doi: 10.1038/nchembio.687.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. & Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *science*, 337 (6096): 816-821.

Jørgensen, S., Schotta, G. & Sørensen, C. S. (2013). Histone H4 lysine 20 methylation: key player in epigenetic regulation of genomic integrity. *Nucleic acids research*, 41 (5): 2797-2806.

Kaltwasser, M., Wiegert, T. & Schumann, W. (2002). Construction and application of epitope- and green fluorescent protein-tagging integration vectors for Bacillus subtilis. *Appl. Environ. Microbiol.*, 68 (5): 2624-2628.

Kennedy, Edward M., Bogerd, Hal P., Kornepati, Anand V. R., Kang, D., Ghoshal, D., Marshall, Joy B., Poling, Brigid C., Tsai, K., Gokhale, Nandan S., Horner, Stacy M., et al. (2016). Posttranscriptional m6A Editing of HIV-1 mRNAs Enhances Viral Gene Expression. *Cell Host & Microbe*, 19 (5): 675-685. doi: https://doi.org/10.1016/j.chom.2016.04.002.

Kolas, N. K., Chapman, J. R., Nakada, S., Ylanko, J., Chahwan, R., Sweeney, F. D., Panier, S., Mendez, M., Wildenhain, J. & Thomson, T. M. (2007). Orchestration of the DNA-damage response by the RNF8 ubiquitin ligase. *Science*, 318 (5856): 1637-1640.

Kotsantis, P., Silva, L. M., Irmscher, S., Jones, R. M., Folkes, L., Gromak, N. & Petermann, E. (2016). Increased global transcription activity as a mechanism of replication stress in cancer. *Nature communications*, 7 (1): 1-13.

Kretschmer, J., Rao, H., Hackert, P., Sloan, K. E., Höbartner, C. & Bohnsack, M. T. (2018). The m6A reader protein YTHDC2 interacts with the small ribosomal subunit and the 5′–3′ exoribonuclease XRN1. *Rna*, 24 (10): 1339-1350.

Kristianto, J., Johnson, M. G., Zastrow, R. K., Radcliff, A. B. & Blank, R. D. (2017). Spontaneous recombinase activity of Cre–ERT2 in vivo. *Transgenic research*, 26 (3): 411-417.

Kuo, Alex J., Cheung, P., Chen, K., Zee, Barry M., Kioi, M., Lauring, J., Xi, Y., Park, Ben H., Shi, X., Garcia, Benjamin A., et al. (2011). NSD2 Links Dimethylation of Histone H3 at Lysine 36 to Oncogenic Programming. *Molecular Cell*, 44 (4): 609-620. doi: https://doi.org/10.1016/j.molcel.2011.08.042.

Kuppusamy, K., Sperber, H. & Ruohola-Baker, H. (2013). MicroRNA regulation and role in stem cell maintenance, cardiac differentiation and hypertrophy. *Current molecular medicine*, 13 (5): 757-764.

Lang, K. S., Hall, A. N., Merrikh, C. N., Ragheb, M., Tabakh, H., Pollock, A. J., Woodward, J. J., Dreifus, J. E. & Merrikh, H. (2017). Replication-transcription conflicts generate R-loops that orchestrate bacterial stress survival and pathogenesis. *Cell*, 170 (4): 787-799. e18.

Lawrence, M., Daujat, S. & Schneider, R. (2016). Lateral thinking: how histone modifications regulate gene expression. *Trends in Genetics*, 32 (1): 42-56. doi: 10.1016/j.tig.2015.10.007.

Lehnertz, B., Ueda, Y., Derijck, A. A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T. & Peters, A. H. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Current Biology*, 13 (14): 1192-1200.

Li, A., Chen, Y.-S., Ping, X.-L., Yang, X., Xiao, W., Yang, Y., Sun, H.-Y., Zhu, Q., Baidya, P., Wang, X., et al. (2017). Cytoplasmic m6A reader YTHDF3 promotes mRNA translation. *Cell Research*, 27 (3): 444-447. doi: 10.1038/cr.2017.10.

Li, E. & Zhang, Y. (2014). DNA methylation in mammals. *Cold Spring Harb Perspect Biol*, 6 (5): a019133. doi: 10.1101/cshperspect.a019133.

Li, F., Zhao, D., Wu, J. & Shi, Y. (2014). Structure of the YTH domain of human YTHDF2 in complex with an m6A mononucleotide reveals an aromatic cage for m6A recognition. *Cell Research*, 24 (12): 1490-1492. doi: 10.1038/cr.2014.153.

Li, Y., Trojer, P., Xu, C. F., Cheung, P., Kuo, A., Drury, W. J., 3rd, Qiao, Q., Neubert, T. A., Xu, R. M., Gozani, O., et al. (2009). The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate. *J Biol Chem*, 284 (49): 34283-95. doi: 10.1074/jbc.M109.034462.

Liang, C., Xu, Y., Ge, H., Li, G. & Wu, J. (2018). Clinicopathological and prognostic significance of OCT4 in patients with hepatocellular carcinoma: a meta-analysis. *OncoTargets and therapy*, 11: 47.

Liao, S., Sun, H. & Xu, C. (2018). YTH Domain: A Family of N(6)-methyladenosine (m(6)A) Readers. *Genomics Proteomics Bioinformatics*, 16 (2): 99-107. doi: 10.1016/j.gpb.2018.04.002.

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *nature*, 362 (6422): 709-715.

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q.-M., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, 462 (7271): 315-322. doi: 10.1038/nature08514.

Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., Jia, G., Yu, M., Lu, Z., Deng, X., et al. (2014). A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat Chem Biol*, 10 (2): 93-5. doi: 10.1038/nchembio.1432.

Loh, Y.-H., Wu, Q., Chew, J.-L., Vega, V. B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B. & Liu, J. (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature genetics*, 38 (4): 431-440.

Lu, W.-T., Hawley, B. R., Skalka, G. L., Baldock, R. A., Smith, E. M., Bader, A. S., Malewicz, M., Watts, F. Z., Wilczynska, A. & Bushell, M. (2018). Drosha drives the formation of DNA: RNA hybrids around DNA break sites to facilitate DNA repair. *Nature communications*, 9 (1): 1-13.

Lucio-Eterovic, A. K., Singh, M. M., Gardner, J. E., Veerappan, C. S., Rice, J. C. & Carpenter, P. B. (2010). Role for the nuclear receptor-binding SET domain protein 1 (NSD1) methyltransferase in coordinating lysine 36 methylation at histone 3 with RNA polymerase II function. *Proceedings of the National Academy of Sciences*, 107 (39): 16952-16957. doi: 10.1073/pnas.1002653107.

Lucio-Eterovic, A. K. & Carpenter, P. B. (2011). An open and shut case for the role of NSD proteins as oncogenes. *Transcription*, 2 (4): 158-161.

Mailand, N., Bekker-Jensen, S., Faustrup, H., Melander, F., Bartek, J., Lukas, C. & Lukas, J. (2007). RNF8 ubiquitylates histones at DNA double-strand breaks and promotes assembly of repair proteins. *Cell*, 131 (5): 887-900.

Margueron, R. & Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. *Nature*, 469 (7330): 343-349.

Martello, G. & Smith, A. (2014). The nature of embryonic stem cells. *Annual review of cell and developmental biology*, 30: 647-675.

Martin, G. R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences*, 78 (12): 7634-7638. doi: 10.1073/pnas.78.12.7634.

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A. A., et al. (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol*, 9 (6): 625-35. doi: 10.1038/ncb1589.

Mauer, J., Luo, X., Blanjoie, A., Jiao, X., Grozhik, A. V., Patil, D. P., Linder, B., Pickering, B. F., Vasseur, J. J., Chen, Q., et al. (2017). Reversible methylation of m(6)Am in the 5' cap controls mRNA stability. *Nature*, 541 (7637): 371-375. doi: 10.1038/nature21022.
McCrum-Gardner, E. (2008). Which is the correct statistical test to use? *British Journal of Oral and Maxillofacial Surgery*, 46 (1): 38-41.
Meyer, K. D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C. E. & Jaffrey, S. R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell*, 149 (7): 1635-46. doi: 10.1016/j.cell.2012.05.003.
Meyer, K. D., Patil, D. P., Zhou, J., Zinoviev, A., Skabkin, M. A., Elemento, O., Pestova, T. V., Qian, S. B. & Jaffrey, S. R. (2015). 5' UTR m(6)A Promotes Cap-Independent Translation. *Cell*, 163 (4): 999-1010. doi: 10.1016/j.cell.2015.10.012.
Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M. & Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell*, 113 (5): 631-642.
Mohan, M., Herz, H. M. & Shilatifard, A. (2012). SnapShot: Histone lysine methylase complexes. *Cell*, 149 (2): 498-498.e1. doi: 10.1016/j.cell.2012.03.025.
Molinie, B. & Giallourakis, C. C. (2017). Genome-Wide Location Analyses of N6-Methyladenosine Modifications (m 6 A-Seq). In *RNA Methylation*, pp. 45-53: Springer.
Morishita, M. & di Luccio, E. (2011). Cancers and the NSD family of histone lysine methyltransferases. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1816 (2): 158-163. doi: https://doi.org/10.1016/j.bbcan.2011.05.004.
Mulas, C., Kalkan, T., von Meyenn, F., Leitch, H. G., Nichols, J. & Smith, A. (2019). Defined conditions for propagation and manipulation of mouse embryonic stem cells. *Development*, 146 (6): dev173146.
Nguyen, A. T. & Zhang, Y. (2011). The diverse functions of Dot1 and H3K79 methylation. *Genes Dev*, 25 (13): 1345-58. doi: 10.1101/gad.2057811.
Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H. & Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell*, 95 (3): 379-91.
Nichols, J. & Smith, A. (2009). Naive and primed pluripotent states. *Cell stem cell*, 4 (6): 487-492.
Nicola, N. A. & Babon, J. J. (2015). Leukemia inhibitory factor (LIF). *Cytokine & growth factor reviews*, 26 (5): 533-544.
Nimura, K., Ura, K., Shiratori, H., Ikawa, M., Okabe, M., Schwartz, R. J. & Kaneda, Y. (2009). A histone H3 lysine 36 trimethyltransferase links Nkx2-5 to Wolf–Hirschhorn syndrome. *Nature*, 460 (7252): 287-291.
Nishizawa, Y., Konno, M., Asai, A., Koseki, J., Kawamoto, K., Miyoshi, N., Takahashi, H., Nishida, N., Haraguchi, N. & Sakai, D. (2018). Oncogene c-Myc promotes epitranscriptome m6A reader YTHDF1 expression in colorectal cancer. *Oncotarget*, 9 (7): 7476.
Niwa, H., Miyazaki, J. & Smith, A. G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet*, 24 (4): 372-6. doi: 10.1038/74199.
Niwa, H. (2007). How is pluripotency determined and maintained? *Development*, 134 (4): 635-646.
Niwa, H. (2010). Mouse ES cell culture system as a model of development. *Development, Growth & Differentiation*, 52 (3): 275-283. doi: 10.1111/j.1440-169X.2009.01166.x.
Ohle, C., Tesorero, R., Schermann, G., Dobrev, N., Sinning, I. & Fischer, T. (2016). Transient RNA-DNA hybrids are required for efficient double-strand break repair. *Cell*, 167 (4): 1001-1013. e7.
Okano, M., Bell, D. W., Haber, D. A. & Li, E. (1999). DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell*, 99 (3): 247-257. doi: https://doi.org/10.1016/S0092-8674(00)81656-6.
Oudet, P., Gross-Bellard, M. & Chambon, P. (1975). Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell*, 4 (4): 281-300.

Pasillas, M. P., Shah, M. & Kamps, M. P. (2011). NSD1 PHD domains bind methylated H3K4 and H3K9 using interactions disrupted by point mutations in human sotos syndrome. *Human Mutation*, 32 (3): 292-298. doi: 10.1002/humu.21424.

Patil, D. P., Chen, C.-K., Pickering, B. F., Chow, A., Jackson, C., Guttman, M. & Jaffrey, S. R. (2016). m6A RNA methylation promotes XIST-mediated transcriptional repression. *Nature*, 537 (7620): 369-373. doi: 10.1038/nature19342.

Patil, D. P., Pickering, B. F. & Jaffrey, S. R. (2018). Reading m6A in the transcriptome: m6A-binding proteins. *Trends in cell biology*, 28 (2): 113-127.

Pease, S., Braghetta, P., Gearing, D., Grail, D. & Williams, R. L. (1990). Isolation of embryonic stem (ES) cells in media supplemented with recombinant leukemia inhibitory factor (LIF). *Developmental biology*, 141 (2): 344-352.

Pei, H., Zhang, L., Luo, K., Qin, Y., Chesi, M., Fei, F., Bergsagel, P. L., Wang, L., You, Z. & Lou, Z. (2011). MMSET regulates histone H4K20 methylation and 53BP1 accumulation at DNA damage sites. *Nature*, 470 (7332): 124-128.

Perera, A., Eisen, D., Wagner, M., Laube, Silvia K., Künzel, Andrea F., Koch, S., Steinbacher, J., Schulze, E., Splith, V., Mittermeier, N., et al. (2015). TET3 Is Recruited by REST for Context-Specific Hydroxymethylation and Induction of Gene Expression. *Cell Reports*, 11 (2): 283-294. doi: https://doi.org/10.1016/j.celrep.2015.03.020.

Ping, X. L., Sun, B. F., Wang, L., Xiao, W., Yang, X., Wang, W. J., Adhikari, S., Shi, Y., Lv, Y., Chen, Y. S., et al. (2014). Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Res*, 24 (2): 177-89. doi: 10.1038/cr.2014.3.

Popovic, R., Martinez-Garcia, E., Giannopoulou, E. G., Zhang, Q., Zhang, Q., Ezponda, T., Shah, M. Y., Zheng, Y., Will, C. M. & Small, E. C. (2014). Histone methyltransferase MMSET/NSD2 alters EZH2 binding and reprograms the myeloma epigenome through global and focal changes in H3K36 and H3K27 methylation. *PLoS genetics*, 10 (9).

Rahman, S., Sowa, M. E., Ottinger, M., Smith, J. A., Shi, Y., Harper, J. W. & Howley, P. M. (2011). The Brd4 extraterminal domain confers transcription activation independent of pTEFb by recruiting multiple proteins, including NSD3. *Mol Cell Biol*, 31 (13): 2641-52. doi: 10.1128/mcb.01341-10.

Ramsahoye, B. H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A. P. & Jaenisch, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc Natl Acad Sci U S A*, 97 (10): 5237-42. doi: 10.1073/pnas.97.10.5237.

Rayasam, G. V., Wendling, O., Angrand, P. O., Mark, M., Niederreither, K., Song, L., Lerouge, T., Hager, G. L., Chambon, P. & Losson, R. (2003). NSD1 is essential for early post-implantation development and has a catalytically active SET domain. *Embo j*, 22 (12): 3153-63. doi: 10.1093/emboj/cdg288.

Ries, R. J., Zaccara, S., Klein, P., Olarerin-George, A., Namkoong, S., Pickering, B. F., Patil, D. P., Kwak, H., Lee, J. H. & Jaffrey, S. R. (2019). m6A enhances the phase separation potential of mRNA. *Nature*, 571 (7765): 424-428. doi: 10.1038/s41586-019-1374-1.

Rizzino, A. & Wuebben, E. L. (2016). Sox2/Oct4: A delicately balanced partnership in pluripotent stem cells and embryogenesis. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1859 (6): 780-791.

Roundtree, I. A., Luo, G. Z., Zhang, Z., Wang, X., Zhou, T., Cui, Y., Sha, J., Huang, X., Guerrero, L., Xie, P., et al. (2017). YTHDC1 mediates nuclear export of N(6)-methyladenosine methylated mRNAs. *Elife*, 6. doi: 10.7554/eLife.31311.

Saga, K., Park, J., Nimura, K., Kawamura, N., Ishibashi, A., Nonomura, N. & Kaneda, Y. (2019). NANOG helps cancer cells escape NK cell attack by downregulating ICAM1 during tumorigenesis. *Journal of Experimental & Clinical Cancer Research*, 38 (1): 1-13.

Saneyoshi, M., Harada, F. & Nishimura, S. (1969). Isolation and characterization of N6-methyladenosine from Escherichia coli valine transfer RNA. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis*, 190 (2): 264-273. doi: https://doi.org/10.1016/0005-2787(69)90078-1.

Sankaran, S. M., Wilkinson, A. W., Elias, J. E. & Gozani, O. (2016). A PWWP Domain of Histone-Lysine N-Methyltransferase NSD2 Binds to Dimethylated Lys-36 of Histone H3 and Regulates NSD2 Function at Chromatin. *Journal of Biological Chemistry*, 291 (16): 8465-8474. doi: 10.1074/jbc.M116.720748.

Santos-Pereira, J. M. & Aguilera, A. (2015). R loops: new modulators of genome dynamics and function. *Nature reviews Genetics*, 16 (10): 583-597.

Sarai, N., Nimura, K., Tamura, T., Kanno, T., Patel, M. C., Heightman, T. D., Ura, K. & Ozato, K. (2013). WHSC1 links transcription elongation to HIRA-mediated histone H3. 3 deposition. *The EMBO journal*, 32 (17): 2392-2406.

Schwartz, S., Mumbach, Maxwell R., Jovanovic, M., Wang, T., Maciag, K., Bushkin, G. G., Mertins, P., Ter-Ovanesyan, D., Habib, N., Cacchiarelli, D., et al. (2014). Perturbation of m6A Writers Reveals Two Distinct Classes of mRNA Methylation at Internal and 5′ Sites. *Cell Reports*, 8 (1): 284-296. doi: https://doi.org/10.1016/j.celrep.2014.05.048.

Schöller, E., Weichmann, F., Treiber, T., Ringle, S., Treiber, N., Flatley, A., Feederle, R., Bruckmann, A. & Meister, G. (2018). Interactions, localization, and phosphorylation of the m(6)A generating METTL3-METTL14-WTAP complex. *Rna*, 24 (4): 499-512. doi: 10.1261/rna.064063.117.

Science, W. I. o. (2016). *B27 Suppplement (AKA B22)*. Ver. 2 ed. Available at: https://hannalabweb.weizmann.ac.il/wp-content/uploads/2016/02/HANNA-LAB-B22-B27-PROTOCOL-V3.pdf (accessed: 01.06.20).

Shi, H., Wang, X., Lu, Z., Zhao, B. S., Ma, H., Hsu, P. J., Liu, C. & He, C. (2017). YTHDF3 facilitates translation and decay of N6-methyladenosine-modified RNA. *Cell Research*, 27 (3): 315-328. doi: 10.1038/cr.2017.15.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A., Casero, R. A. & Shi, Y. (2004). Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. *Cell*, 119 (7): 941-953. doi: https://doi.org/10.1016/j.cell.2004.12.012.

Shilatifard, A. (2012). The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annual review of biochemistry*, 81: 65-95.

Sibbritt, T., Patel, H. R. & Preiss, T. (2013). Mapping and significance of the mRNA methylome. *Wiley Interdisciplinary Reviews: RNA*, 4 (4): 397-422.

Siegfried, Z. & Cedar, H. (1997). DNA methylation: A molecular lock. *Current Biology*, 7 (5): R305-R307. doi: https://doi.org/10.1016/S0960-9822(06)00144-8.

Silva, J., Nichols, J., Theunissen, T. W., Guo, G., van Oosten, A. L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I. & Smith, A. (2009). Nanog is the gateway to the pluripotent ground state. *Cell*, 138 (4): 722-737.

Sim, Y.-J., Kim, M.-S., Nayfeh, A., Yun, Y.-J., Kim, S.-J., Park, K.-T., Kim, C.-H. & Kim, K.-S. (2017). 2i maintains a naive ground state in ESCs through two distinct epigenetic mechanisms. *Stem cell reports*, 8 (5): 1312-1328.

Simandi, Z., Balint, B. L., Poliska, S., Ruhl, R. & Nagy, L. (2010). Activation of retinoic acid receptor signaling coordinates lineage commitment of spontaneously differentiating mouse embryonic stem cells in embryoid bodies. *FEBS Letters*, 584 (14): 3123-3130. doi: 10.1016/j.febslet.2010.05.052.

Skourti-Stathaki, K., Proudfoot, N. J. & Gromak, N. (2011). Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Molecular cell*, 42 (6): 794-805.

Skourti-Stathaki, K., Kamieniarz-Gdula, K. & Proudfoot, N. J. (2014). R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature*, 516 (7531): 436-439.

Skourti-Stathaki, K. & Proudfoot, N. J. (2014). A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes & development*, 28 (13): 1384-1396.

Skourti-Stathaki, K., Torlai Triglia, E., Warburton, M., Voigt, P., Bird, A. & Pombo, A. (2019). R-Loops Enhance Polycomb Repression at a Subset of Developmental Regulator Genes. *Mol Cell*, 73 (5): 930-945.e4. doi: 10.1016/j.molcel.2018.12.016.

Sollier, J. & Cimprich, K. A. (2015). Breaking bad: R-loops and genome integrity. *Trends in cell biology*, 25 (9): 514-522.
Steger, D. J., Lefterova, M. I., Ying, L., Stonestrom, A. J., Schupp, M., Zhuo, D., Vakoc, A. L., Kim, J. E., Chen, J., Lazar, M. A., et al. (2008). DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. *Mol Cell Biol*, 28 (8): 2825-39. doi: 10.1128/mcb.02076-07.
Stock, J. K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A. G. & Pombo, A. (2007). Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat Cell Biol*, 9 (12): 1428-35. doi: 10.1038/ncb1663.
Stoilov, P., Rafalska, I. & Stamm, S. (2002). YTH: a new domain in nuclear proteins. *Trends in Biochemical Sciences*, 27 (10): 495-497. doi: https://doi.org/10.1016/S0968-0004(02)02189-8.
Stork, C. T., Bocek, M., Crossley, M. P., Sollier, J., Sanz, L. A., Chedin, F., Swigut, T. & Cimprich, K. A. (2016). Co-transcriptional R-loops are the main cause of estrogen-induced DNA damage. *Elife*, 5: e17548.
Strick, T. & Portman, J. (2019). Transcription-coupled repair: from cells to single molecules and back again. *Journal of molecular biology*, 431 (20): 4093-4102.
Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., et al. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, 324 (5929): 930-5. doi: 10.1126/science.1170116.
Takahashi, K. & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *cell*, 126 (4): 663-676.
Tam, P. P. & Behringer, R. R. (1997). Mouse gastrulation: the formation of a mammalian body plan. *Mech Dev*, 68 (1-2): 3-25.
Tam, P. P., Goldman, D., Camus, A. & Schoenwolf, G. C. (1999). 1 Early Events of Somitogenesis in Higher Vertebrates: Allocation of Precursor Cells During Gastrulation and the Organization of a Meristic Pattern in the Paraxial Mesoderm. In vol. 47 *Current topics in developmental biology*, pp. 1-32: Elsevier.
Tesar, P. J., Chenoweth, J. G., Brook, F. A., Davies, T. J., Evans, E. P., Mack, D. L., Gardner, R. L. & McKay, R. D. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature*, 448 (7150): 196-9. doi: 10.1038/nature05972.
Thomas, M., White, R. L. & Davis, R. W. (1976). Hybridization of RNA to double-stranded DNA: formation of R-loops. *Proceedings of the National Academy of Sciences*, 73 (7): 2294-2298.
Tian, S., Jiang, L., Gao, Q., Zhang, J., Zong, M., Zhang, H., Ren, Y., Guo, S., Gong, G. & Liu, F. (2017). Efficient CRISPR/Cas9-based gene knockout in watermelon. *Plant Cell Reports*, 36 (3): 399-406.
Tian, T. V., Di Stefano, B., Stik, G., Vila-Casadesus, M., Sardina, J. L., Vidal, E., Dasti, A., Segura-Morales, C., De Andres-Aguayo, L., Gomez, A., et al. (2019). Whsc1 links pluripotency exit with mesendoderm specification. *Nat Cell Biol*, 21 (7): 824-834. doi: 10.1038/s41556-019-0342-1.
Tomizawa, S.-i., Kobayashi, H., Watanabe, T., Andrews, S., Hata, K., Kelsey, G. & Sasaki, H. (2011). Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. *Development*, 138 (5): 811-820. doi: 10.1242/dev.061416.
Tsang, W. P., Zhang, F., He, Q., Cai, W., Huang, J., Chan, W. Y., Shen, Z. & Wan, C. (2017). Icaritin enhances mESC self-renewal through upregulating core pluripotency transcription factors mediated by ERα. *Scientific reports*, 7: 40894.
Van den Wyngaert, I., Sprengel, J., Kass, S. U. & Luyten, W. H. M. L. (1998). Cloning and analysis of a novel human putative DNA methyltransferase. *FEBS Letters*, 426 (2): 283-289. doi: 10.1016/s0014-5793(98)00362-7.

Veening, J.-W., Smits, W. K., Hamoen, L. W., Jongbloed, J. D. & Kuipers, O. P. (2004). Visualization of differential gene expression by improved cyan fluorescent protein and yellow fluorescent protein production in Bacillus subtilis. *Appl. Environ. Microbiol.*, 70 (11): 6809-6815.

Wagner, E. J. & Carpenter, P. B. (2012). Understanding the language of Lys36 methylation at histone H3. *Nature Reviews Molecular Cell Biology*, 13 (2): 115-126. doi: 10.1038/nrm3274.

Wang, X., Lu, Z., Gomez, A., Hon, G. C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G., et al. (2014). N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature*, 505 (7481): 117-20. doi: 10.1038/nature12730.

Wang, X., Zhao, Boxuan S., Roundtree, Ian A., Lu, Z., Han, D., Ma, H., Weng, X., Chen, K., Shi, H. & He, C. (2015). N6-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell*, 161 (6): 1388-1399. doi: https://doi.org/10.1016/j.cell.2015.05.014.

Wang, Y., Li, Y., Toth, J. I., Petroski, M. D., Zhang, Z. & Zhao, J. C. (2014). N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat Cell Biol*, 16 (2): 191-8. doi: 10.1038/ncb2902.

Wen, Y., Hou, Y., Huang, Z., Cai, J. & Wang, Z. (2017). SOX 2 is required to maintain cancer stem cells in ovarian cancer. *Cancer science*, 108 (4): 719-731.

Williams, K., Christensen, J., Pedersen, M. T., Johansen, J. V., Cloos, P. A. C., Rappsilber, J. & Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature*, 473 (7347): 343-348. doi: 10.1038/nature10066.

Wojtas, M. N., Pandey, R. R., Mendel, M., Homolka, D., Sachidanandam, R. & Pillai, R. S. (2017). Regulation of m6A Transcripts by the 3'→5' RNA Helicase YTHDC2 Is Essential for a Successful Meiotic Program in the Mammalian Germline. *Molecular Cell*, 68 (2): 374-387.e12. doi: https://doi.org/10.1016/j.molcel.2017.09.021.

Xiang, Y., Laurent, B., Hsu, C. H., Nachtergaele, S., Lu, Z., Sheng, W., Xu, C., Chen, H., Ouyang, J., Wang, S., et al. (2017). RNA m(6)A methylation regulates the ultraviolet-induced DNA damage response. *Nature*, 543 (7646): 573-576. doi: 10.1038/nature21671.

Xiao, W., Adhikari, S., Dahal, U., Chen, Y.-S., Hao, Y.-J., Sun, B.-F., Sun, H.-Y., Li, A., Ping, X.-L., Lai, W.-Y., et al. (2016). Nuclear m6A Reader YTHDC1 Regulates mRNA Splicing. *Molecular Cell*, 61 (4): 507-519. doi: https://doi.org/10.1016/j.molcel.2016.01.012.

Xu, C., Wang, X., Liu, K., Roundtree, I. A., Tempel, W., Li, Y., Lu, Z., He, C. & Min, J. (2015). Correction: Corrigendum: Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. *Nature Chemical Biology*, 11 (10): 815-815. doi: 10.1038/nchembio1015-815c.

Yagi, M., Yamanaka, S. & Yamada, Y. (2017). Epigenetic foundations of pluripotent stem cells that recapitulate in vivo pluripotency. *Laboratory Investigation*, 97 (10): 1133-1141.

Yang, S., Zheng, X., Lu, C., Li, G.-M., Allis, C. D. & Li, H. (2016). Molecular basis for oncohistone H3 recognition by SETD2 methyltransferase. *Genes & development*, 30 (14): 1611-1616.

Ying, Q.-L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. (2003). Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nature biotechnology*, 21 (2): 183-186.

Zaccara, S., Ries, R. J. & Jaffrey, S. R. (2019). Reading, writing and erasing mRNA methylation. *Nature Reviews Molecular Cell Biology*, 20 (10): 608-624. doi: 10.1038/s41580-019-0168-5.

Zeng, Y. & Chen, T. (2019). DNA Methylation Reprogramming during Mammalian Development. *Genes (Basel)*, 10 (4). doi: 10.3390/genes10040257.

Zhao, B. S. & He, C. (2015). Fate by RNA methylation: m6A steers stem cell pluripotency. *Genome Biol*, 16 (1): 43. doi: 10.1186/s13059-015-0609-1.

Zhao, B. S., Roundtree, I. A. & He, C. (2017). Post-transcriptional gene regulation by mRNA modifications. *Nature Reviews Molecular Cell Biology*, 18 (1): 31-42. doi: 10.1038/nrm.2016.132.

Zhao, Z. & Shilatifard, A. (2019). Epigenetic modifications of histones in cancer. *Genome Biol*, 20 (1): 245. doi: 10.1186/s13059-019-1870-5.
Zheng, G., Dahl, J. A., Niu, Y., Fedorcsak, P., Huang, C. M., Li, C. J., Vågbø, C. B., Shi, Y., Wang, W. L., Song, S. H., et al. (2013). ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol Cell*, 49 (1): 18-29. doi: 10.1016/j.molcel.2012.10.015.
Zheng, N., Zhang, D. & Liu, X. (2018). Prognostic significance of expression of Myogenin and OCT4 in glioma. *Minerva pediatrica*.
Zhong, L., Liao, D., Zhang, M., Zeng, C., Li, X., Zhang, R., Ma, H. & Kang, T. (2019). YTHDF2 suppresses cell proliferation and growth via destabilizing the EGFR mRNA in hepatocellular carcinoma. *Cancer letters*, 442: 252-261.
Zhong, S., Li, H., Bodi, Z., Button, J., Vespa, L., Herzog, M. & Fray, R. G. (2008). MTA Is an Arabidopsis Messenger RNA Adenosine Methylase and Interacts with a Homolog of a Sex-Specific Splicing Factor. *The Plant Cell*, 20 (5): 1278-1288. doi: 10.1105/tpc.108.058883.
Zhu, T., Roundtree, I. A., Wang, P., Wang, X., Wang, L., Sun, C., Tian, Y., Li, J., He, C. & Xu, Y. (2014). Crystal structure of the YTH domain of YTHDF2 reveals mechanism for recognition of N6-methyladenosine. *Cell Res*, 24 (12): 1493-6. doi: 10.1038/cr.2014.152.
Ziller, M. J., Müller, F., Liao, J., Zhang, Y., Gu, H., Bock, C., Boyle, P., Epstein, C. B., Bernstein, B. E., Lengauer, T., et al. (2011). Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. *PLoS Genet*, 7 (12): e1002389. doi: 10.1371/journal.pgen.1002389.
Zubay, G. & Wilkins, M. H. (1962). An X-ray diffraction study of histone and protamine in isolation and in combination with DNA. *Journal of Molecular Biology*, 4 (6): 444-IN6.

7. Supplementary

7.1 Media compositions

N2B27 media

250 mL Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12 (DMEM/F-12)(1X), 250 mL Neurobasal Medium 1X (Gibco), N2 supplement (16 µg putricine dihydrochloride, 6,25 µg insulin, 50 µg apotransferrin, 21.6 ng progesterone in EtOH, 15 nM sodium selenite), 1 vial B27 supplement, 5 mL Penicillin Streptomycin Glutamine (100X) (Gibco), 5 mL glutaMAX (100X) (Gibco), 5 mL Minimum Essential Medium – Non-essential Amino Acids Solution (MEM-NEAA) (100X) (Gibco), 5 mL 100 mM Sodium Pyruvat (100X) (Gibco), 50 µM β-mercaptoethanol (Gibco), 10 mM Hepes (Fisher BioReagents), 50 µg BSA (Saween & Werner), and 4 µg Heparin (Merck Life Science).

N2 and B27 were internally laboratory made, using the protocol from Weizmann Institute of Science (Science, 2016).

ES cell media

N2B27 media supplemented with 10 µM MEKi and 1 mM GSKi (Sigma, Life Science, CHIR99021) (2i), and LIF (internally laboratory made, expressed from the COS cells and harvested as supernatant).

NSC media

N2B27 media supplemented with EGF and FGFb. EGF (Perprotech (315-09)) used at the final concentration of 10 ng/mL and FGFb (Perprtech (100-18B)) used at the final concentration of 10 ng/mL.

Differentiation media

N2B27 supplemented with 10% FBS.

7.2 Buffer compositions

7.2.1 Genotyping

Alkaline lysis reagent

25 mM NaOH and 0.2 mM EDTA (Oslo University Hospital), to a total volume of 50 mL with ddH_2O.

Neutralization buffer

40 mM Tris-HCL pH 7.5 (Gibco) to a total volume of 50 mL with ddH_2O.

7.2.2 Western blotting

Lysis buffer

50 mM Tris-HCl pH 7.5 (Gibco), 300 mM NaCl (Oslo University Hospital), 0.5% ICEPA® CA-630 (Sigma-Aldrich), 1 mM EDTA pH8 (Oslo University Hospital). 1 mM DTT (Saween & Werner), 1 µg AP (Sigma-Aldrich), 0.5 µg LP (Sigma-Aldrich), and 2 mM PMSF (Sigma-Aldrich) per mL buffer was added to the cold media before use.

20x TBS

0.4 M TRIS pH 7.5 and 3 M NaCl (Oslo University Hospital)

TBS-T

25 mL 20x TBS, 475 mL milli-Q water, and 500 µl Tween® 20 (Sigma)

Blocking milk (5% milk)

30 mL TBS-T, 1.5 g skim milk powder, and 6 µl sodium azide (Sigma-Aldrich)

7.2.3 Immunoprecipitation

Lysis buffer

85 mM KCl (Oslo University Hostpital), 5 mM PIPES pH8 (Sigma), 0.5% ICEPA® CA-630 (Sigma-Aldrich). 1 mM DTT (Saween & Werner), 1 µg AP (Sigma-Aldrich), 0.5 µg LP (Sigma-Aldrich), and 2 mM PMSF (Sigma-Aldrich) per mL lysis buffer was added to the cold media before use.

RSB

10 mM TRIS-HCl pH 7.5, 200 mM NaCl, and 2.5 mM $MgCl_2$ (Oslo University Hospital)

RSB-T

RSB buffer added 0.5% Triton X-100 (Sigma-Aldrich).

Nuclear lysis buffer

RSB buffer added 0.2% NaDoc (Sigma-Aldrich), 0.1% SDS (Oslo University Hospital), 0.05% Na-sarkosyl (Sigma-Aldrich), 0.5% Triton X-100 (Sigma-Aldrich). 1 mM DTT (Saween & Werner), 1 µg AP (Sigma-Aldrich), 0.5 µg LP (Sigma-Aldrich), and 2 mM PMSF (Sigma-Aldrich) per mL lysis buffer was added to the cold media before use.

Decrosslinking buffer

100 mM $NaHCO_3$ and 1% SDS (Oslo University Hospital).

7.3 Primers

Table S1: List of oligonucleotides (Eurofins genomics, Oligonucleotide Synthesis Report) used for NSC, ATRA, and EB differentiation protocols.

Name	Sequence (5`-> 3`)	Name	Sequence (5`-> 3`)
RPO (F)	TCC ATT GTG GGA GCA GA	Brachyury (R)	TCC GAG GCT AGA CCA GTT ATC A
RPO (R)	CAG CAG TTT CTC CCA GAG C	Foxa2 (F)	ACA TGT TCG AGA ACG GCT GCT A
Nanog (F)	TCA AGG CAG CCC TGA TTC T	Foxa2 (R)	CTT CCT TCA GTG CCA GTG GCT T
Nanog (R)	TTC CTG GCA AGG ACC TTG TTC T	Gata4 (F)	GAT GAC TTC TCA GAA GGC AGA G
Oct4_1 (F)	CGG AAG AGA AAG CGA ACT AGC A	Gata4 (R)	ACA GGC ATT GCA CAG GTA GTG T
Oct4_1 (R)	TGT GAG TGA TCT GCT GTA GGG A	Gata6 (F)	CCG AGA ACA GTG ACC TCA AGT A
Sox1 (F)	GCG CAA GAT GGCC CA GGA	Gata6 (R)	AGA GCA CAC CAA GAA TCC TGT C
Sox1 (R)	ATC TCC GAG TTG TGC ATC TTG G	Pax3 (F)	GGA GGC ACA AAG CTG TCT GTA T
Nestin (F)	GCA TTT CCT TGG GAT ACC AGA G	Pax3 (R)	CAC TGT ACA CCA AAG CAC T
Nestin (R)	GTC TCA AGG GTA TTA GGC AAG G	Pax6 (F)	ATT ACG AGA CTG GCT CCA TCA G
Tubb3 (F)	AGA TGC TGG CCA TCC AGA GTA A	Pax6 (R)	TCT TGG CTT ACT CCC TCC GAT T
Tubb3 (R)	CTA CCT TGA CGT TGT TGT GGA T		

Table S2: List of oligonucleotides (Eurofins genomics, Oligonucleotide Synthesis Report) used for DRIP qPCR (Skourti-Stathaki et al., 2019).

Name	Sequence (5`-> 3`)	Name	Sequence (5`-> 3`)
Polycomb-repressed genes		**Active gene**	
Gata4 P (F)	AAG AGC GCT TGC GTC TCT A	β-actin P (F)	GAG GGG AGA GGG GGT AAA
Gata4 P (R)	TTG CTA GCC TCA GAT CTA CGG	β-actin P (R)	GAA GCT GTG CTC GCG G
		β-actin C (F)	CAC CAT TCA CCA TCT TGT C
Inactive gene		β-actin C (R)	TGA TCC ACA TCT GCT GG
Myf5 P (F)	CAC GGT TGT GGT GGG ATA TGC TAA	CyclinB1 P (F)	GCT AGC TTG GAC AGC ACA CA
Myf5 P (R)	GGA GTT TGG GAC TGT CTC TCT G	CyclinB1 P (R)	GTT CCC GTA GAA TGC GTT TC
		CyclinB1 C (F)	AGT TTA GAG CCA GCC AGG ACT
		CyclinB1 C (R)	GAG AAA AGC ACT GAC ATC AGG

Table S3: List of primers used for genotyping.

Primer name	Primer sequence (5`-> 3`)	PCR product size	
		Wild type allele	Conditional knock-out allele
Genotyping of 36.5, 37.4 loxP/loxP, and *Df1* knock-out cells			
182262cre-KLU13	CTA TCA GGA ATA TTG TGT AGG AGA GTC TTG AAG G	171 bp	344 bp
182263cre-KLU13	TTA CAA CAA ACA CTG ATA GCA CAG GAC AGC		
Genotyping of 37.4-creERT2 cells			
Cre-F 10783	CGT ATA GCC GAA ATT GCC AGG	One fragment at 327 bp	
Cre-R 10784	CTG ACC AGA GTC ATC CTT AGC		

Table S4. sgRNA used to generate CRISPR/Cas9 knock-outs.

Name	sgRNA	Sequence with PAM	Sequence without PAM	Usage
DLP4	mY2_e4g1	CCC ATT CAT GGA ACG ATA GGC GG	CCC ATT CAT GGA ACG ATA GG	Used for generation of *Ythdf2* knock-outs
DLP5	mY2_e4g2	GAC CGA AGT TTC TCCA ACA CCG G	GAC CGA AGT TTC TCC AAC AC	
DLP7	mY3_e5g1	TAG GTT GTC AGA TAT GGC ATA GG	TAG GTT GTC AGA TAT GGC AT	Used for generation of *Ythdf3* knock-outs
DLP8	mY3_e5g2	GCT GCG GTG ACG AAA ACT GTA GG	GCT GCG GTG ACG AAA ACT GT	

7.4 Antibodies

Table S4: List of antibodies.

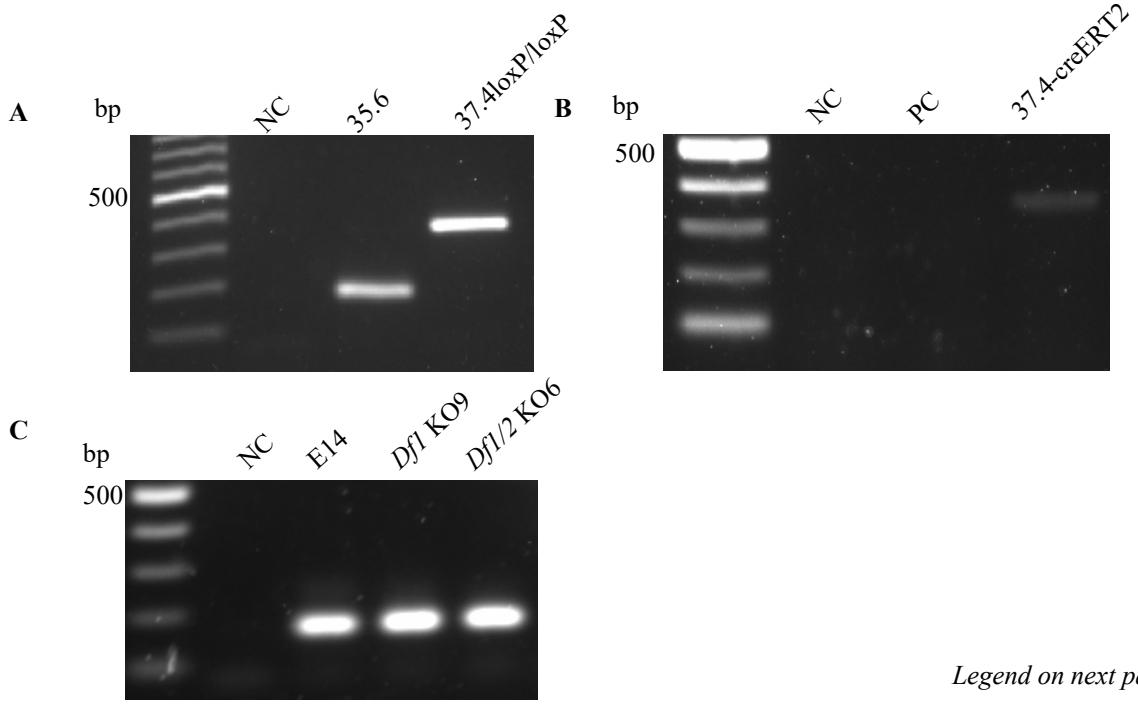
Antibody name	Type	Application	Unit	Working concentration	Obtained from
Primary antibody					
Anti-phospho-Histone H2A.X (Ser139) Clone JBW301	Mouse monoclonal IgG	IF	1.0 mg/mL	1 : 500	Sigma-Aldrich (05-636-I)
Anti-53BP1	Rabbit	IF	1.0 mg/mL	1 : 2,000	Abcam (ab36823)
Anti-DNA-RNA hybrid S9.6	Goat IgG monoclonal	IF	100 µg	1 : 500	Kerafast (Kf-Ab01137-24.1)
GAPDH Antibody	Rabbit polyclonal IgG	WB	200 µg/mL	1 : 5,000	Santa Cruz Biotechnology (FL-335)
Recombinant Anti-YTHDF1	Rabbit monoclonal	WB	0.396 mg/mL	1 : 2,000	Abcam (ab220162)
Recombinant Anti-YTHDF2	Rabbit monoclonal	WB	0.52 mg/mL	1 : 2,000	Abcam (ab220163)
Recombinant Anti-YTHDF3	Rabbit monoclonal	WB	0.482 mg/mL	1 : 2,000	Abcam (ab220161)
Anti-GFP	Rabbit polyclonal	WB		1 : 50,000	Abcam (ab290)
Anti-Vinculin	Mouse monoclonal	WB	200 µg	1 : 5,000	Merck life sciences: V9131
Ty1	Mouse monoclonal	WB		1 : 1,000	Internally lab-made*
MMSET (8C3)	Mouse monoclonal	WB		1 : 3	Internally lab-made

(Table continues on next page)

Antibody name	Type	Application	Unit	Working concentration	Obtained from
Secondary antibody					
Donkey anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 555	Donkey polyclonal	IF Used for γH2AX	2 mg/mL	1 : 1,000	Invitrogen (A-31570)
Donkey anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 555	Donkey polyclonal	IF Used for 53BP1	2 mg/mL	1 : 1,000	Invitrogen (A-31572)
Donkey anti-Goat IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	Donkey polyclonal	IF Used for S9.6	2 mg/mL	1 : 400	Invitrogen (A-11055)
HRP Goat Anti-Rabbit IgG Antibody (Peroxidase)	Goat	WB	1 mg	1 : 5,000	Vector laboratories (PI-1000-1)
HRP Horse Anti-Mouse IgG Antibody (Peroxidase)	Horse	WB	1 mg	1 : 5,000	Vector laboratories (PI-2000-1)

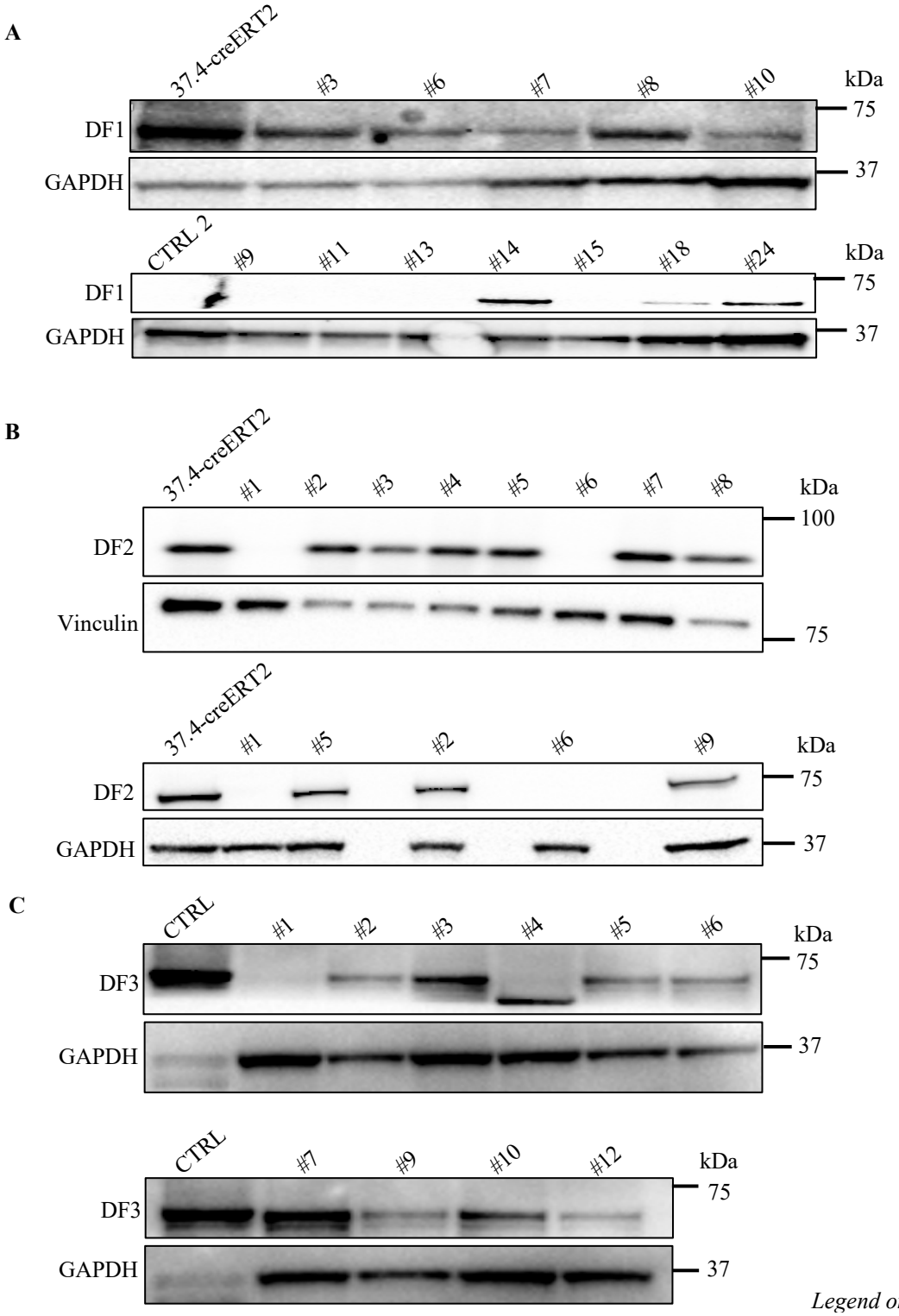
* a kind gift from the group of Prof. Ragnhild Eskeland, IMB, UiO

7.5 Genotyping and western blot analyses



Legend on next page

Figure S1. Genotyping of 35.6, 37.4, 37.4-creERT2, and *Df1* KO cells. Genotyping performed by PCR and separation by agarose gel (1.25%). H_2O loaded as negative control (NC) and E14 used a positive control.
(A) Genotyping of 35.6 (wild type) and 37.4 loxP/loxP (conditional knock-out). Agarose gel picture indicating 35.6 as wild type at ~171 bp and the conditional knock-out at ~344 bp.
(B) Genotyping of 37.4-creERT2**.** Agarose gel picture showing successfull insert of creERT2 in the 37.4 cell line using lentivirus, as a band at ~ 327 bp.
(C) Genotyping of *Df1* KO and *Df1/2* KO**.** Genotyping of *Df1* KOs shows a slightly higher band than the positive control as expected upon *Df1* KO. Genotyping of *Df1/3* KO (and the other *Df1* and *Df1/2* KOs) lack due to technical issues in the laboratory.



Legend on next page

Figure S2. Western blot analysis of treated or transfected cells after single cell sorting (related to figure 14).
All clones were probed with the indicated antibodies, and GAPDH or vinculin were used as loading controls.
(A) 12 clones tested for possible *Df1* KO, where #9, #11, #13, and #15 are *Df1* KOs, as no bands are visible when probed with DF1 antibody.
(B) 8 clones tested for possible *Df2* KO, where #1 and #6 are *Df1/2* KOs. The bottom panel show these two KOs loaded, including two clones that later were used as control cells. A last clone, #9, was also tested for *Df1/2* KO.
(C) 10 clones tested for possible *Df3* KO, and #1 was shown to be a clean KO for *Df*3.

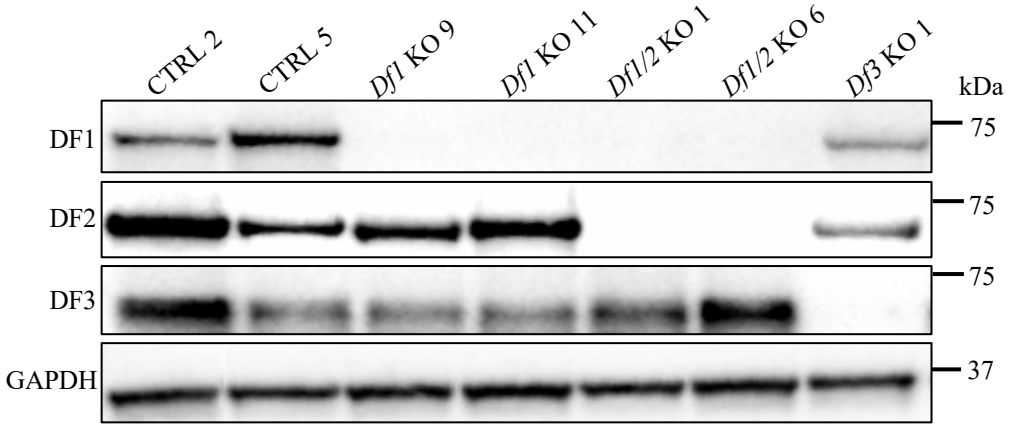


Figure S3: Western blot of *Df* knock-out cells (related to figure 14).
Western blot showing the different *Df* knock-outs, all knock-out cells were probed with DF1-3 antibodies. Control 2 and control 5 are clone #2 and #5 that turned out not to be KOs for *Df2* (Fig. S2). GAPDH was used as loading control.

7.6 Sorting of mES cells by FACS

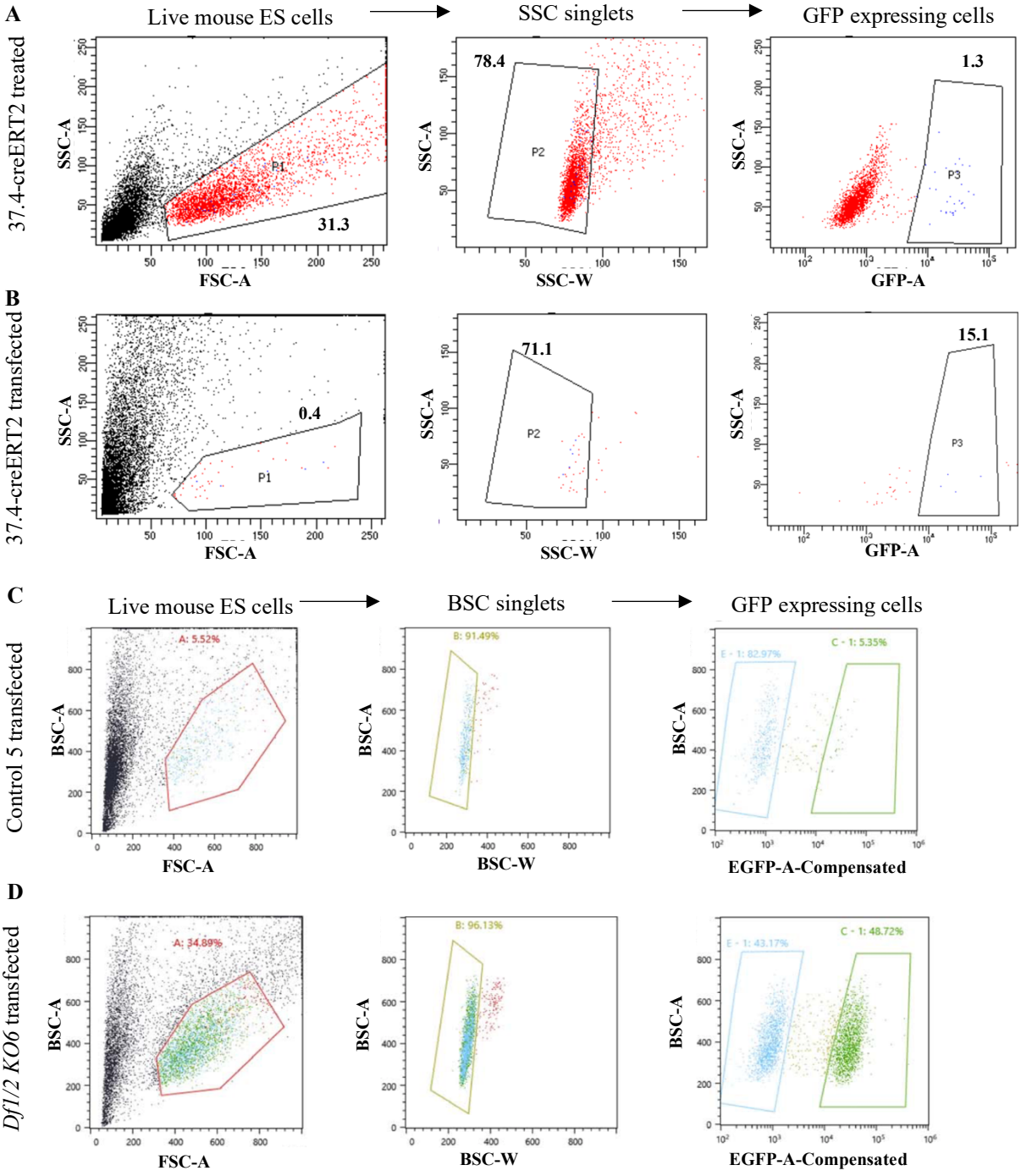


Figure S4. Treated and transfected cells analyzed by flow cytometry for GFP expression
(A) Flow cytometry of treated 37.4-creERT2 cells. Sorting of treated 37.4-creERT2 cells for generation of *Df1* KO cells. Sorted into 96-well plate.
(B) Flow cytometry of transfected 37.4-creERT2 cells. Sorting of transfected 37.4-creERT2 cells for generation of *Df2* KO, later found out to be *Df1/2* KO. Sorted into 96-well plate.
(C) Flow cytometry of transfected control 5 cells. Sorting of transfected control 5 cells for generation of *Df3* KO cells. Sorted into 96-well plate.
(D) Flow cytometry of transfected *Df1/2 KO 6* cells. Sorting of transfected *Df1/2 KO6* cells for generation of triple *Df1/2/3* KO cells. Sorted into 96-well plate.

7.7 DRIP western blot and qPCR analyses

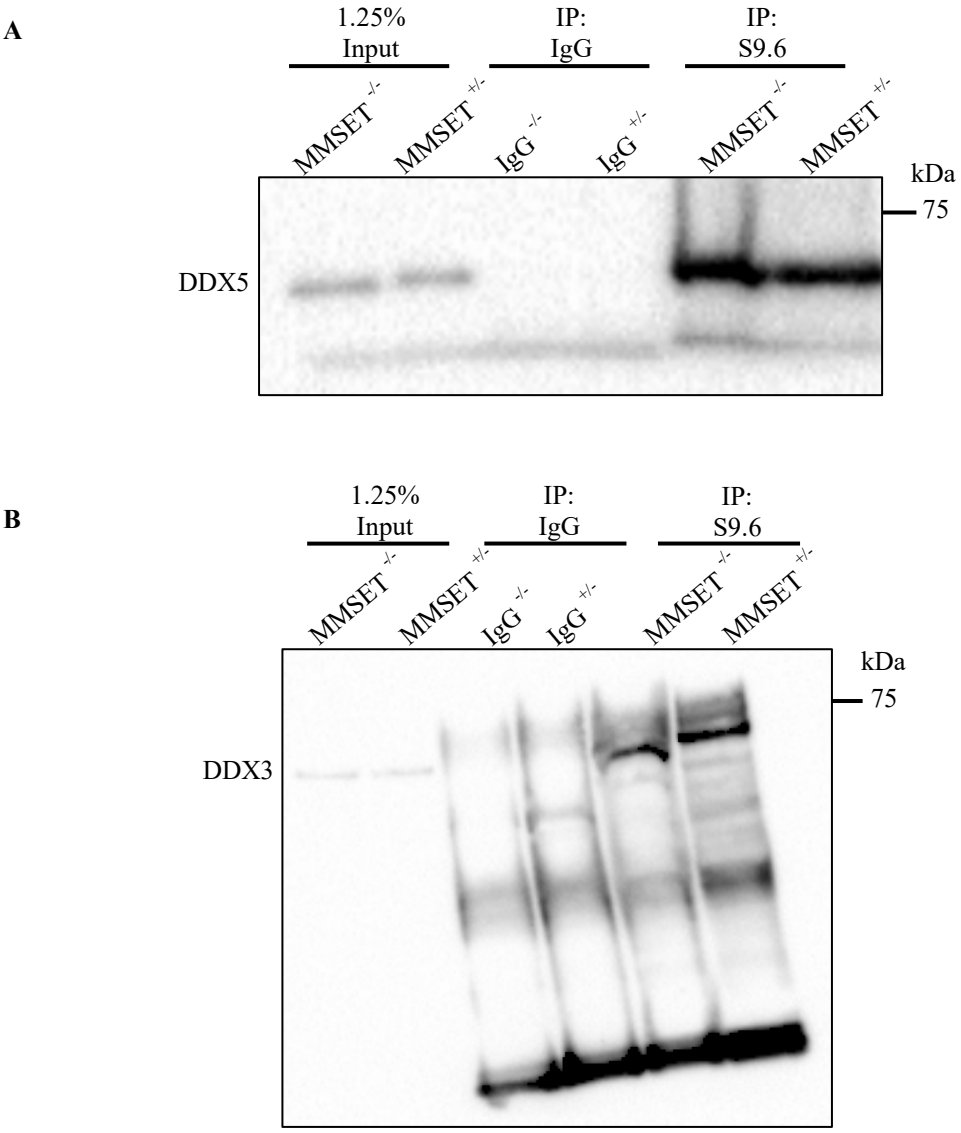


Figure S5. Western blot analysis of R-loop IP probed with the indicated antibodies (related to figure 21). The same western blot was probed for DHX9 (A) and MMSET (B).
(A) DRIP western blot analysis, probed with DDX5 antibody. Loading control for IgG, validating that protein extracts where loaded for IP IgG.
(B) DRIP western blot analysis, probed with DDX3 antibody. Loading control for IgG, validating that protein extracts where loaded for IP IgG.

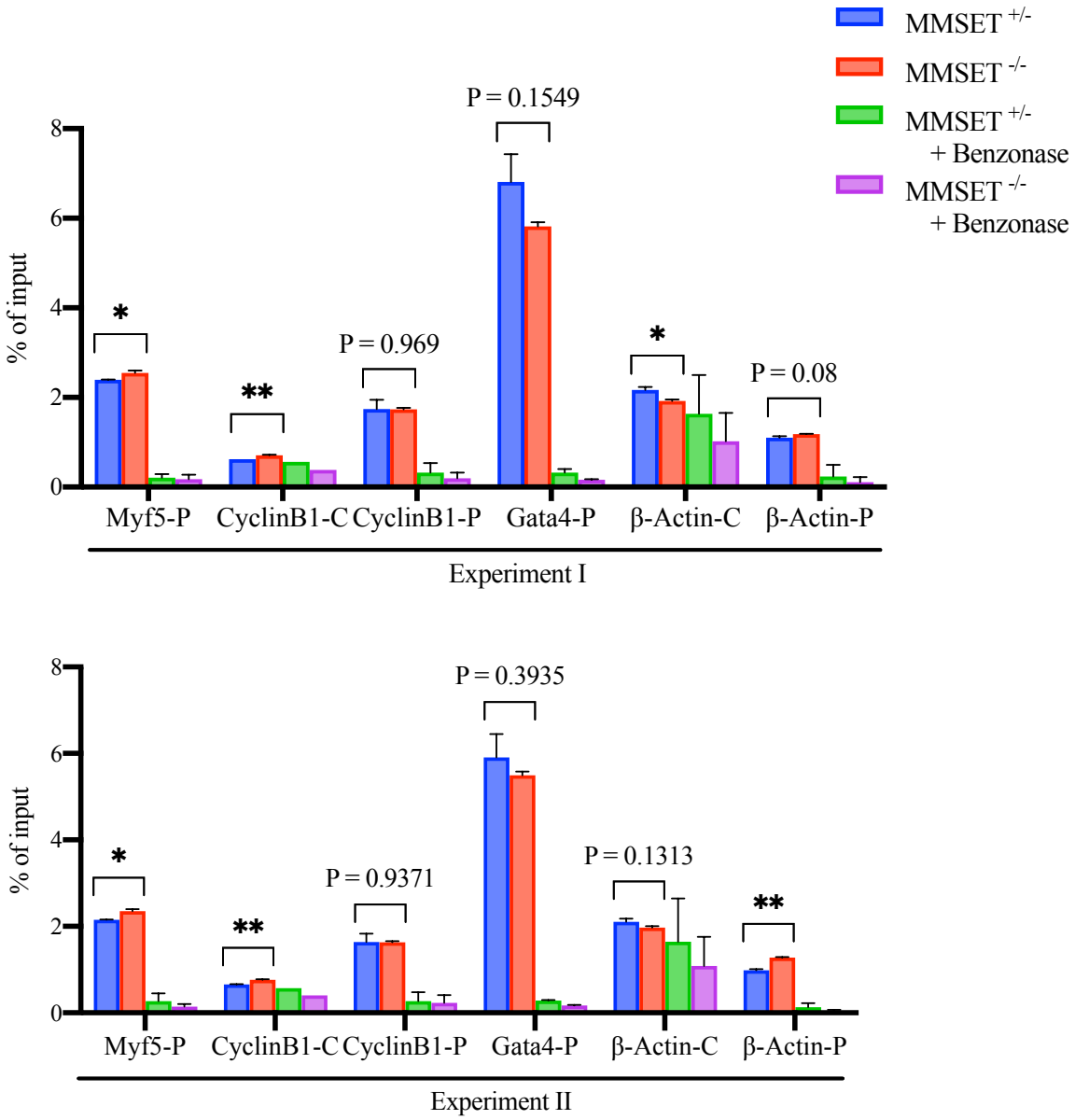


Figure S6. DRIP qPCR (related to figure 22).
Performed on heterozygous MMSET and MMSET null cells. DRIP analysis performed over the PcG active genes, *Cyclin B1-C/P* and *β-actin-C/P,* the polycomb repressed gene, *Gata4-P,* and the inactive gene *Myf5-P*. Where indicated, the cells were treated with benzonase. Error bars indicate ±SD; n=2, and P values were calculated by two-tailed unpaired t-test. *p <0.05, **p <0.01.

Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway