THE UNIVERSITY *of* EDINBURGH

Investigating the role of DNA methylation in pluripotency and differentiation using an embryoid body model

THE UNIVERSITY
of EDINBURGH

Philippa Jackson

Thesis presented for the degree of Doctor of Philosophy
The University of Edinburgh
2022

Declaration

I declare that this thesis has been composed by me and that all the work is my own unless otherwise stated

This work has not been submitted for any other degree

Philippa Jackson

September 2022

Acknowledgements

We have a saying on C2: It takes an institute to raise a PhD student. From the number of people I want to thank in these acknowledgements, I'd say it requires a fair few more people than that.

To start with thanks to Richard for being the most patient and supportive PI I could have hoped for. I asked for endless positive energy and inspiration, you provided. Thanks also to my second supervisor Ian and to the rest of my thesis committee for the helpful feedback and encouragement during my yearly reviews.

Thanks also to the other members of the lab. To Heidi, for taking me under your wing and helping me settle in amongst the other Meehens. Donncha, I appreciate all the help with fixing PCRs that wouldn't work, but more importantly for the help with the crosswords. I know you won't miss my squeaky pipette from the other side of the bench, I hope you got your peace and quiet back. Thanks especially to David for joining our group and becoming an indispensable font of PhD survival advice and cloning tricks. Katie, Olga, Mihail – thanks for being the best lab and pub mates, and best of luck in your own projects.

Sisi – we've been through all the highs and lows of a PhD together, and through all of them you were the calm and reassuring presence. Not only were you the experiment troubleshooting Queen and cell culture whizz (you got through those flasks so damn fast!), but you were also always keeping lab meetings topped up with homemade bread, jam, and 'Bulgarian' chocolate (Milka counts as Bulgarian if you bought it there, right?). Not the Rakia though, that stuff should come with a warning label.

Thanks to the technical services staff for keeping us all going. Special thanks to Lizzie down in FACS, I'll miss our cell sorting chats, and to the AIR team for always being on hand no matter how many times I got stuck on a microscope. Those things have minds of their own. Big thanks to Gogo from

the bioinformatics team for all the data analysis help. Thanks for being so patient with my extremely novice queries.

And of course, thanks to everyone of C2. Gillian, Shelagh, Hazel, Abby, Francesca, Giannis – you made the IGMM feel more like a home away from home.

Special thanks to Kamila and Bethan. I never thought something good would have come out of a pandemic, but it brought us together (in a socially distanced way of course), and our zoom wine nights made COVID so much less isolating. I'm looking forward to making up for all those lost lockdown days in the future.

Thanks also to the people who never saw inside my lab but were indispensable pillars of support nonetheless.

To Steven and everyone at GBA for giving me something else to worry about besides my experiments (…how many reps?!). Surely if I could survive that I could survive a thesis.

Thanks to Rachael, Kat and Charlotte for being my proof-readers, and to the rest of the Lincoln crowd for reminding me that life extends beyond the lab walls.

Thanks Giselle and Louisa for all the thesis therapy nights. We got through our Masters together and now we're all doing our PhDs together. I can't wait to see you ladies celebrate submitting too. It does end eventually, I promise.

Thanks to Uncle Roy for always being at the end of the phone with another great (*groan*) joke. You've had snapshots of this project every month for the last five years, and you never doubted for a moment.

Finally, a very special thank you to Mum and Dad (and cats) for supporting me throughout all the many, many years of my education. I couldn't have got this far without you. (I'll go and get a proper job now, promise).

Abstract

Epigenetics is the study of heritable alterations in phenotype caused by changes in cellular properties, but where the genotype is unchanged. At the molecular level these changes include chemical modifications of DNA and histones in chromatin. Specific chromatin states are associated with gene activity or silencing. The proper functioning of these mechanisms is critical for mammalian survival, particularly during embryonic development. One of the best studied epigenetic modifications is DNA methylation, wherein methyl-groups are placed on cytosines in CpG dinucleotide contexts by DNA methyltransferases to form 5mC. The malfunction of this mechanism is associated with failure of embryogenesis and many adult human disease pathologies, including cancer. However, questions remain about how the 5mC patterns are established *de novo*, how the patterns can change between different cell types, and why some cell types can tolerate the absence of 5mC but not others. Genetic removal of the *de novo* (*Dnmt3a* and *Dnmt3b*) or maintenance methyltransferases (*Dnmt1*) in somatic cells can lead to cell death. However, mouse embryonic stem cells (ESCs) can proliferate normally in the absence of all three of these proteins. This suggests that DNA methylation becomes essential at some point after the cells exit pluripotency and begin differentiation. Being able to identify this window of time during development is important for better understanding the dynamics of recruitment of these proteins and deposition of the mark and could therefore provide insight into how their misregulation contributes to disease.

In this thesis I investigate the role of DNA methylation during ESC differentiation using a combination of conditional and reporter cell lines, *in vitro* differentiation models, RNA sequencing, genetic engineering, and high-resolution imaging. The ESC lines include those in which i) the expression of Dnmt1 and key pluripotency genes can be tuned separately and in combination, ii) contain germ layer differentiation reporters, and iii) contain reporters of 5mC distribution. I allow these cells to differentiate to form embryoid bodies (EBs), a validated embryogenesis-like model that enables simulation of the pluripotent-to-differentiated transition *in vitro*. By combining

these approaches, I was able to investigate the impact of loss of Dnmt1 on gene expression pathways including apoptosis, primordial germ cell (PGC) and 2C-like cell formation, germ layer differentiation, and changes in transposable element expression. I was also able to delineate differentiation trajectories by comparing my bulk RNA sequencing data with published single cell RNA sequencing data.

Overall, I observed that inhibition of *Dnmt1* activity consistently led to a significant reduction in EB size though germ layer differentiation was still able to occur. Likewise, EBs were still able to form in the absence of master pluripotency factor Oct4, although to a reduced capacity. Loss of both proteins led to smaller EBs than the wild type. Interestingly *Oct4*$^{-/-}$, *Dnmt1*$^{+/+}$ EBs were the most affected. There were no significant changes in frequency of apoptotic cells, and only LTR-family transposable elements were de-repressed. By comparing data, I was able to identify that loss of *Dnmt1* enriched EBs for PGC-like cell marker genes at late stages of differentiation. I conclude that under normal differentiation initiation conditions the two systems, the pluripotency network and the DNA methylation network, work synergistically to control the gradual switch of cells from the pluripotent to the differentiated state enabling the formation of a limited population of PGC-like cells prior to advanced differentiation of germ layer cell lineages. The purpose of this may be to protect emerging PGC-like cells from Transposon Element activity. It may also be to allow the number of pluripotent cells to reach a threshold level prior to initiating differentiation to prevent the EB size being limited in a mechanism similar to that suggested for primordial dwarfism.

Lay Summary

Every cell of the mammalian body, from brain cells to liver cells, contains an identical copy of the organism's DNA, the double helix structure encoding the genetic instructions for life. How do multiple tissue types arise from this one biological instruction manual if the code in every cell is the same? The answer partly lies in epigenetics; this is the process of decorating the DNA with chemical molecules in a tissue-specific manner that does not alter the underlying sequence but can change the DNA's biological properties. It is akin to actors in a play decorating their scripts; every actor has an identical copy of the script, but the series of post-it notes and which parts are redacted will be different for each role. Which bits are read by which actor, or which genes are expressed by which cell, are indicated by the pattern of markers on top of (epi-) the script (-genetics). By ensuring every actor reads the appropriate lines at the appropriate time, the epigenetic markers ensure the play proceeds as planned. If one of the epigenetic systems fails to place its marks correctly or some are removed by mistake the play can go wrong. In biology this could be the failure of an embryo to develop properly, or the formation of an aberrant tissue like cancer. Therefore, understanding how the mechanisms function has implications for human health. One of the best understood epigenetic systems is DNA methylation, wherein methyl molecules are placed on the cytosines which are next to guanines (CpG) by DNA methyltransferases (Dnmts) to form 5mC. This mark is generally associated with inactive genes, so methylated DNA is comparable to a redacted script.

Epigenetic systems are dynamic during embryonic development. At the early stages the cells have the potential to become any of the cell types (except the placenta) and are called pluripotent. Pluripotent cells cultured in the lab can grow normally without Dnmts and are able to tolerate the absence of 5mC. As the embryo matures and cells exit pluripotency Dnmts become essential and loss of 5mC leads to embryonic death around day 9.5 of development. Thus, we know that while DNA methylation is critical for mammalian development it is not essential from fertilisation. This leads to my thesis question: when does 5mC become essential during development, and why?

We can mimic the early stages of mammalian development using mouse embryonic stem cells in the lab by allowing them to differentiate to form embryoid bodies (EB). These are complex 3D clusters of cells which over time develop all three germ layers, the tissue types found in *in vivo* embryos. EBs share many genetic features with pre-implantation stage embryos making them a useful model for dissecting the dynamics of epigenetic systems, particularly the pluripotent-to-differentiated transition. In this thesis I use the EB model along with ES cells featuring tune-able *de novo* DNA methyltransferase Dnmt1 genes to determine at what stage of EB development DNA methylation becomes essential. I then use molecular biology techniques including RNA sequencing, genome editing and high-resolution imaging to investigate what changes in these cells to deduce why DNA methylation is needed at some stages and in some cell types but not others.

I found that in the absence of Dnmt1 EBs are smaller, have higher expression of potentially harmful transposable elements, and are enriched for genes associated with primordial germ cells, the precursors to sperm and oocytes. These results suggest that the mark is not needed for germ layer differentiation to occur, but it may have a role in the timing of differentiation initiation. This may be required to control the timely emergence of PGCs and to protect them from transposable element activity, which may be detrimental to the next generation if passed on in the gametes. It may also be required to control EB size; by delaying differentiation DNA methylation may ensure a pluripotent cell number threshold is exceeded prior to differentiation initiation. This would maintain a large enough number of rapidly dividing pluripotent cells that the EB can continue to grow in size while still producing slower cycling differentiating cells from the core pluripotent population.

Table of Contents

List of figures

List of tables

Abbreviations

2C	2-cell-stage embryo
5-aza	5-azacytidine
5mC	5-methylcytosine
bp	Base pair
BP	Blood progenitor
CGI	CpG island
CRISPR	Clustered regularly interspaced short palindromic repeats
CSK	Cytoskeleton buffer
ddH_2O	Double-distilled water
DE	Differentially expressed (genes)
DE	Definitive endoderm
DIG	Digoxigenin
DMEM	Dulbecco’s modified Eagle medium
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
Dox	Doxycycline
DSB	Double-strand break
E	Embryonic day
EB	Embryoid body
EBR	EBRTcH3 mESC line
EGFP	Enhanced green fluorescent protein
Epi-LC	Epiblast-like cell

ERV	Endogenous retrovirus
ES	Enrichment score
ESC	Embryonic stem cell
FACS	Fluorescence-activated cell sorting
FC	Fold change
FCS	Foetal calf serum
FDR	False-discovery rate
FISH	Fluorescence *in situ* hybridization
gDNA	Genomic DNA
GFP	Green fluorescent protein
GO	Gene ontology
gRNA	Guide RNA
GSEA	Gene set enrichment analysis
IAP	Intracisternal A particle (retrotransposon)
ICC	Immunocytochemistry
ICE	Imprinting Control Element
ICR	Imprint Control Region
IF	Immunofluorescence
KO	Knockout cell line
LIF	Leukaemia Inhibitory Factor
LINE	Long interspersed nuclear elements (retrotransposon)
lncRNA	Long non-coding RNA
MBD	Methyl-CpG binding domain

MEF	Mouse embryonic fibroblast
MERVL	Murine ERV-L element
MES	Mesoderm
MQ	Milli-Q ultra-pure water
MTZ	Maternal-to-zygotic transition
NC	No template control
NEAA	Non-essential amino acids
NES	Normalised enrichment score
PAM	Protospacer-adjacent motif
PBS	Phosphate buffered saline
PCA	Principle component analysis
PCR	Polymerase chain reaction
pDNA	Plasmid DNA
PFA	Paraformaldehyde
PGC	Primordial germ cell
PI	Propidium iodide
Post-Epib	Post-implantation epiblast
PrE	Primitive endoderm
Pre-Epib	Pre-implantation epiblast
PS	Primitive streak
qPCR	Quantitative polymerase chain reaction
RFP	Red fluorescent protein
RIN	RNA integrity number

rpm	Revolutions per minute
-RT	Reverse transcriptase-free control
RT-PCR	Reverse transcriptase polymerase chain reaction
scRNA-seq	Single cell RNA sequencing
S. Ect	Surface ectoderm
sgRNA	Single-guide RNA
SSEA1	Stage specific embryonic antigen 1
TdT	Terminal deoxynucleotidyl transferase
TE	Transposable element
TKO	Triple knockout cells lacking *Dnmt1*, *Dnmt3a* and *Dnmt3b*
TUNEL	TdT-mediated dUTP nick end labelling
wt	Wild-type
ZGA	Zygotic gene activation

1 Introduction

1.1 A brief history of epigenetics

The term ‘epigenetics’ was coined in 1942 by Conrad Waddington, a geneticist and developmental biologist who wanted to explain how one copy of DNA could contain the information required to develop an entire complex organism (Waddington, 2012; Mattei, Bailly and Meissner, 2022). In his definition he wanted to capture how cell differentiation interacts with genetic mutations during development to create the observed phenotype (Greally, 2018). In other words, he wanted to propose a regulatory mechanism which explained how the mostly unchanging genotype can produce variable phenotypes. Therefore, he proposed the theory of ‘epigenetics’, which he defined as “the branch of biology that studies the causal interactions between genes and their products which bring the phenotype into being” (Waddington, 2014; Deans and Maggert, 2015).

To illustrate his definition, he published a visual metaphor - the now famous ‘Waddington’s Landscape’. A modified version is shown in figure 1.1. In it, he visualises the process of cellular differentiation, in the context of embryonic development, as a hillside containing valleys of varying sizes. These valleys form a series of potential paths, or ‘creodes’, to the bottom of the hill. Importantly, the dimensions of the valleys represent the activities of genes. A ball, representing a pluripotent cell, placed at the top of the hill will roll down and eventually settle into the bottom of one of the valleys. As the ball travels it will arrive at bifurcation points at the tops of valleys from which it is then channelled down into one of the creodes; each bifurcation point represents a cell-fate decision. Waddington envisaged that the valley walls forming the sides of the creodes became steeper towards the bottom of the landscape. This illustrated the difficulty of changing cell fate, as the further the ball proceeds down the landscape the more committed to that differentiation path the cell becomes, which he termed ‘canalisation’. The model implies that the further the ball travels the more differentiated the cell becomes, and the more difficult it is to get the cell back up the landscape in order to change its fate, or de-differentiate. In this way he wanted to connect

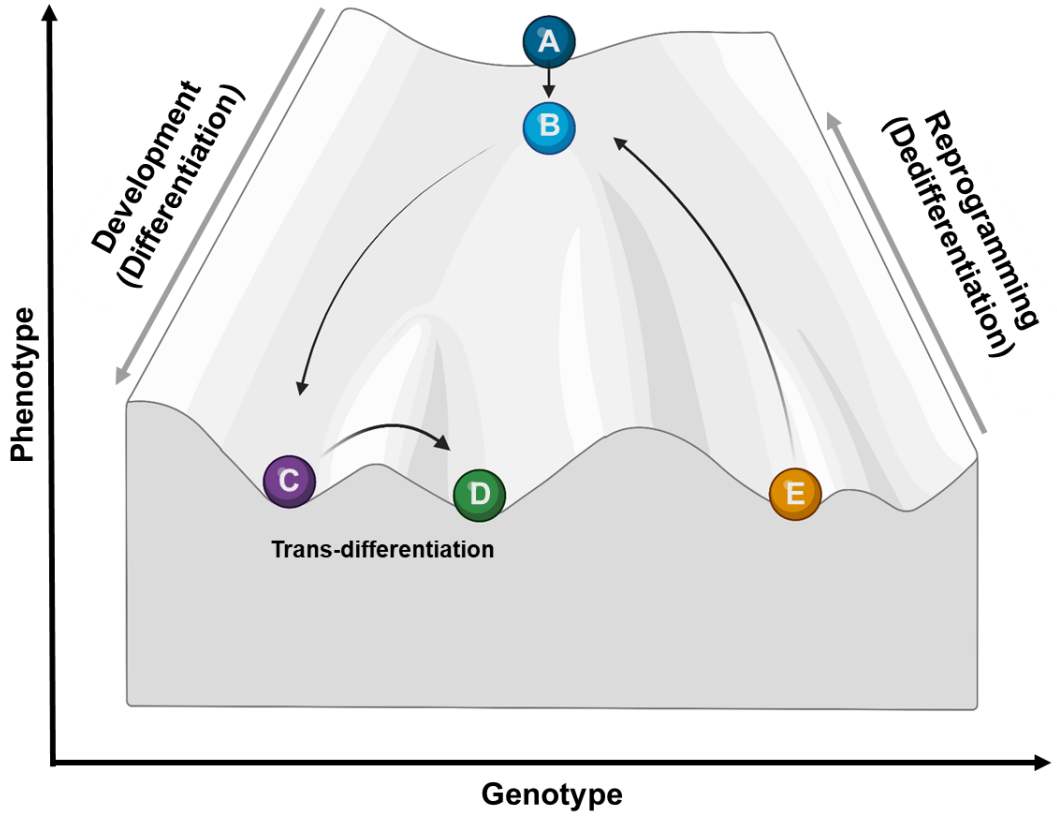


Figure 1.1. An adaptation of Waddington's epigenetic landscape. Waddington's model illustrates the interplay between an organism's genotype and its phenotype. Totipotent cells (A) are canalised during embryonic development into one of multiple differentiated cell types (C-E). Differentiated cells may be artificially reprogrammed back to a pluripotent state (E-B), or directly converted to another somatic cell type by trans-differentiation (C-D). Figure adapted from Waddington (2014) and created with BioRender.com.

the ideas of phenotype plasticity and developmental stability (canalisation) by suggesting the involvement of a gene activity-dependent regulatory pathway. He also proposed the 'canalisation' theory to explain how an organism can consistently produce populations of cells of the same type despite extrinsic genetic and environmental factors, i.e., how certain parts of development, such as the formation of specific organs, were stably reproducible.

However, Waddington's original thesis did not include an element we now know to be key for a mechanism to be considered truly epigenetic.

Namely, the mark must be maintained across cell division, i.e. it must be heritable. This heritable nature was first posited by David Nanney, a contemporary of Waddington, in 1958 (Nanney, 1958) but it was only incorporated into the definition by Holliday in the 90s (Holliday, 1994). Over time as the field of epigenetics became more popular, the use of the original definition was being stretched as it was applied slightly differently among the different areas of biological science. Holliday tried to reconcile these multiple adapted definitions into two new standard ones; one which focused on the molecular aspects of development and which incorporated Nanney's theory of inheritance, and a second definition which focused more on ecology and the impacts of environment on gene expression(Holliday, 1994). Eventually, Wu and Morris streamlined these definitions further into one new standard, re-defining epigenetics as "the study of changes in gene function that are mitotically and/or meiotically heritable and that do not entail change in DNA sequence" (Wu and Morris, 2001).

Although Waddington's theory helped to launch the field of epigenetics by providing a clear, nuanced theoretical framework with which to conceptualise the regulation of development, he did not provide any evidence and it would be many years before this was found. At the time when his theory was published, at the time it was only recently understood that genetic information was encoded within DNA polymers. One potential mechanism was found on the discovery of DNA methylation.

1.2 DNA methylation

A brief history of discovery

DNA methylation was originally discovered in 1925 in bacteria by Johnson and Coghill, who were able to distinguish crystalised cytosine and 5-methylcytosine (5mC) in hydrolysed nucleic acids extracted from *Mycobacterium tuberculosis* by microscopy with polarised light (Johnson and Coghill, 1925). It was later found in mammals by Hotchkiss who performed paper chromatography on an extract of calf thymus and found a slightly larger version of cytosine which he labelled ‘epi-cytosine’. He later hypothesised it to be a methylated form of the nucleotide and suggested it might be 5mC (Hotchkiss, 1948). Wyatt then reported also finding 5mC in other mammal species, as well as in insect and plant DNA (Wyatt, 1950; Wyatt, 1951).

The next important finding was that, in mammals, DNA methylation was not randomly distributed through the genome but was found to most commonly occur on cytosines adjacent to guanines, or CpG dinucleotides (Sinsheimer *et al.*, 1954). However, these CpGs were less common in the genome than expected; for example, only 1% of the nucleic acids in the human genome are 5mCs (Ehrlich *et al.*, 1982). The relative scarcity of these dinucleotides cast doubt on whether 5mC was biologically important (Smith and Markham, 1952; Sinsheimer *et al.*, 1954). Despite this, it was found to occur among such a wide range of species that others argued it had been evolutionarily conserved and therefore must be beneficial. The function of 5mC and its epigenetic nature were still unknown, and its role in the regulation of gene expression wouldn’t be suggested until the 1980s (Holliday and Pugh, 1975; Compere and Palmiter, 1981).

Mapping the distribution of 5mC

During the 1970s, restriction enzymes - particularly *HpaII* (a restriction enzyme which is blocked by DNA methylation) and *MspI* (an enzyme which can cut methylated DNA) - would become critical tools for molecular

biologists investigating DNA methylation (Waalwijk and Flavell, 1978b, 1978a). Using these enzymes scientists would be able to identify all the CpGs present within a genome and then distinguish which ones were methylated and unmethylated, thus enabling them to precisely map the locations of 5mC (Waalwijk and Flavell, 1978a; Singer, Roberts-Ems and Riggs, 1979). This in turn would lead to the determination of the function of DNA methylation.

In a major study, DNA digested with *HpaII* and *MspI* was analysed by Southern blot, a technique for separating and visualising probed DNA by gel electrophoresis (Southern, 1975). This experiment was performed to determine 5mC patterns at tissue-specific genes, improving the resolution of the known 5mC map to the gene level (Southern, 1975; Waalwijk and Flavell, 1978a; Mandel and Chambon, 1979). With this advancement it became possible to identify genes which were always, never, or variably methylated in specific tissues. By investigating the variably methylated genes it was found that a low level of methylation was associated with active gene expression, whereas high methylation levels were present where a gene was inactive. This anti-correlation between gene expression and 5mC level led to the view that DNA methylation was a repressive mark (Kuo, Mandel and Chambon, 1979; McGhee and Ginder, 1979).

With continuing technological advancements, it is now known that CpGs are non-randomly distributed throughout the genome, tending to occur in clusters ~500-1000bp long called CpG Islands (Bird, 1986) (CGIs). Approximately 70% of mammalian CGIs are associated with actively transcribed gene promoters (Saxonov, Berg and Brutlag, 2006) , especially those of housekeeping genes (Gardiner-Garden and Frommer, 1987). These promoter-associated CGIs are highly conserved between mouse and human (Illingworth *et al.*, 2010). CGIs are associated with active genes as they enhance the accessibility of the DNA and promote transcription factor binding (Moore, Le and Fan, 2013).

CpGs found outside CGIs are usually thinly spread and heavily methylated (Bird, 1986). However, CpGs found within CGIs are mostly unmethylated and they are protected from methylation. Deposition of 5mC at these regions is associated with stable silencing (Mohn *et al.*, 2008). This is important for the parent-specific silencing of imprinted genes, as well as regulating gene expression during development and differentiation (Moore, Le and Fan, 2013). Bird proposed CGI protection from methylation was partly mediated by the continuous occupancy of proteins such as transcription factors, which would physically occlude the Dnmts (Bird, 1980). He also suggested that this would be particularly important for housekeeping genes to protect them from spontaneous CpG mutation, such as deamination to form thymine (Coulondre *et al.*, 1978; Bird, 1980), which can occur following addition of a methyl group (Bird, 1986). This tendency to spontaneously mutate may be why 5mC is relatively rare in mammalian genomes, and completely absent in some species.

Although 5mC is mostly found at CpG dinucleotides in mammals, non-CpG DNA methylation is also known to occur, predominantly in plants (Guseinov and Vanyushin, 1975). Non-CpG methylation has been detected in mouse and human ESCs, but so far it has not been found in somatic cells (Ramsahoye *et al.*, 2000; Lister *et al.*, 2009). The role of non-CpG methylation is currently unclear.

The DNA methyltransferases

The first DNA methyltransferase (the family of proteins responsible for methylating nucleotides) was identified and characterised by Bestor *et al*. in 1988, and which they named Dnmt1 (Bestor *et al.*, 1988). Work by Jaenisch *et al*. showed that disrupting the activity of Dnmt1 led to a global loss of DNA methylation, both *in vivo* and *in vitro*, though importantly low levels of 5mC were retained (Li, Timothy H Bestor and Jaenisch, 1992). Knockout of *Dnmt1* in mice led to embryonic lethality between embryonic day (E) 8-10.5 (Li, Timothy H Bestor and Jaenisch, 1992). Knockout embryos featured loss of

two-thirds of their global DNA methylation, as well as numerous apoptotic cells in developing tissues, including the brain. Clearly, the activity of Dnmt1 was critical for development. It was also found that there was a strong up-regulation of IAP-type retroviral elements in these embryos, which suggested a potential role for 5mC in protecting the genome from retroviral activity, especially during embryonic development (Walsh, Chaillet and Bestor, 1998).

That a low level of DNA methylation was retained in *Dnmt1* knockout cells suggested the presence of additional as-yet undiscovered Dnmt proteins (Li, Timothy H Bestor and Jaenisch, 1992; Li, Beard and Jaenisch, 1993). Holliday and Pugh proposed that there could therefore be two enzymes responsible for depositing DNA methylation; one specifically for establishing it *de novo* and a separate one for maintaining it, and between them they could mediate the on/off states of genes (Holliday and Pugh, 1975).

Eventually, the *de novo* methyltransferases Dnmt3a and Dnmt3b were identified; they were highly prevalent in ESCs but lowly expressed in somatic cells (Okano, Xie and Li, 1998). Like Dnmt1, the newly discovered *de novo* methyltransferases proved to be essential in mammals. Mice null for *Dnmt3a* die within weeks of birth, and mice null for *Dnmt3b* die early during embryonic development around E9.5 (Okano, Xie and Li, 1998; Okano, Daphne W Bell, *et al.*, 1999; Rhee *et al.*, 2000). These experiments also showed that both proteins had overlapping functions during embryonic development (Okano, Daphne W. Bell, *et al.*, 1999).

A further essential Dnmt was found. Dnmt3L is a catalytically inactive co-factor to Dnmt1, Dnmt3a and Dnmt3b; it associates with these proteins to stimulate their activity (Hata *et al.*, 2002; Suetake *et al.*, 2004; Jia *et al.*, 2007). It is mainly expressed during early development and its expression is limited to germ cells and the adult thymus As with the other Dnmts, Dnmt3L null ESCs are viable. However, adult mice lacking Dnmt3L can survive and appear phenotypically normal, though both male and female mice are sterile (Bourc'his *et al.*, 2001). It has a prominent role in genomic imprinting; Kaneda

et al. showed that both Dnmt3a and Dnmt3L together are essential for establishing and maintaining imprinted regions (Kaneda *et al.*, 2004), and this was also shown through knockout experiments (Aapola *et al.*, 2000; Bourc'his *et al.*, 2001; Hata *et al.*, 2002).

The mechanism underlying how Dnmts deposit methyl groups onto DNA was outlined by Wu and Santi in 1987 (Wu and Santi, 1987). During the catalytic reaction, Dnmts obtain a donor methyl group from S-adenosylmethionine and covalently add it to a cytosine occurring with a CpG (Smith *et al.*, 2012; Guo *et al.*, 2013). This process is summarised in figure 1.2.

Dnmt1, Dnmt3a and Dnmt3b share common structural features; they each have a large N-terminal regulatory domain and a C-terminal catalytic domain (Yen *et al.*, 1992; Xie *et al.*, 1999). Dnmt3L lacks the catalytic domain. Though they are structurally similar, each Dnmt has a unique function and expression pattern.

Dnmt1 is highly expressed in mammalian tissues, especially the brain (Goto *et al.*, 1994). Unlike most somatic tissues, neural cells don't have a stable methylation pattern so continually require maintenance methyltransferase activity (Moore, Le and Fan, 2013). Dnmt1 preferentially targets and methylates hemi-methylated DNA (Pradhan *et al.*, 1999; Ramsahoye *et al.*, 2000). During DNA replication Dnmt1 localises to the replication fork via Uhrf1 recruitment to copy over the methylation pattern to the newly forming daughter strand (Hermann, Goyal and Jeltsch, 2004; Schermelleh *et al.*, 2007). It is also able to repair methylation when recruited by methyl-binding proteins (Mortusewicz *et al.*, 2005) (MBDs). For this reason, it is known as the maintenance methyltransferase.

Dnmt3a and Dnmt3b can methylate both hemi-methylated and unmethylated DNA with no preference (Okano, Daphne W. Bell, *et al.*, 1999). *Dnmt3a* is expressed widely in most tissues and is required for normal cellular differentiation (Xie *et al.*, 1999). *Dnmt3b* is poorly expressed in

somatic tissues, except for thyroid, testes, and bone marrow, and is mainly required for early development (Xie *et al.*, 1999).

Other Dnmts discovered include Dnmt2, Dnmt1o, and Dnmt3c. Dnmt2 currently is thought to mainly be involved in the methylation of RNAs instead of DNA (Mattei, Bailly and Meissner, 2022). Dnmt1o is an oocyte-specific isoform of Dnmt1 in mice (Howell *et al.*, 2001). Dnmt3c evolved as a duplicate of Dnmt3b and is specifically expressed in male rodent germ cells in which it methylates young retrotransposons (Barau *et al.*, 2016). It therefore has an important role in ensuring mouse fertility (Ambrosi, Manzo and Baubec, 2017).

5mC erasure

In the absence of Dnmts, 5mC is passively lost over multiple cell divisions. However, mammals also have a system of demethylases that actively remove the mark. This mechanism is particularly important for the two waves of epigenetic reprogramming that occur during embryonic development. This process will be described later in the chapter.

DNA demethylation is performed by the ten-eleven translocation methylcytosine dioxygenase proteins 1-3 (TET1-3). These enzymes catalyse the conversion of 5’-methylcytosine (5mC) to 5’-hydroxymethylcytosine (5hmC), and further to 5’-formylcytosine (5fC) and 5’-carboxylcytosine (5caC)(Ito *et al.*, 2010, 2011; He *et al.*, 2011). 5fC and 5caC, but not 5hmC, are recognised and cleaved by thymine DNA glycosylase (TDG), resulting in the formation of an abasic site which the DNA repair machinery then converts to cytosine by base excision repair (BER), completing the cycle (He *et al.*, 2011). This pathway is critical for development as embryos lacking TDG die around E12.5 (Cortázar *et al.*, 2011; Cortellino *et al.*, 2011). The process of DNA demethylation is illustrated in figure 1.2.

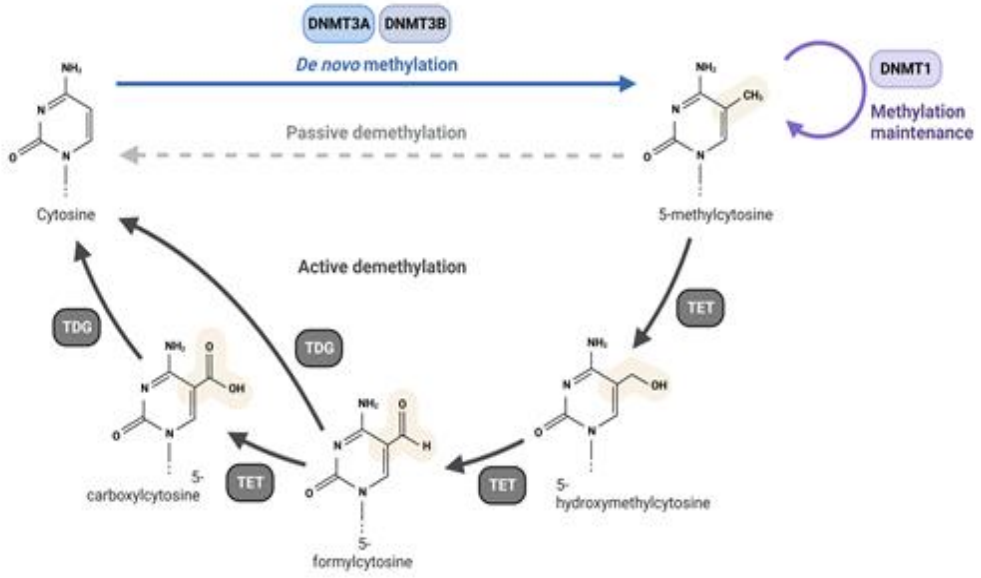


Figure 1.2. The establishment, maintenance, and erasure of DNA methylation. Adapted from “DNA Methylation”, by BioRender.com (2022). Retrieved from https://app.biorender.com/biorender-templates

Recruiting DNA methyltransferases

Maintenance methyltransferase Dnmt1 is recruited to the replication fork specifically by Uhrf1(Bostick *et al.*, 2007; Sharif *et al.*, 2007). At the replication fork, the DNA clamp Proliferating Cell Nuclear Antigen (PCNA) is recognised and bound by Ubiquitin-like with PHD and RING finger domains 1 (Uhrf1, also known as NP95) (SanMiguel and Bartolomei, 2018). Uhrf1 may also be recruited to the replicating region by DNA ligase 1 (LIG1). Uhrf1 binds to hemi-methylated DNA and H3K9me3 at these regions, and then binds the Replication Foci Targeting domain of Dnmt1 (Song *et al.*, 2011; Syeda *et al.*, 2011; Takeshita *et al.*, 2011). Once bound to Uhrf1, the auto-inhibitory conformation of Dnmt1 changes and it is able methylate the emerging daughter strand to match the parental strand (SanMiguel and Bartolomei, 2018). Thus, DNA methylation is mechanistically linked to H3K9 methylation via the activity of Uhrf1 (Rose and Klose, 2014).

Dnmt3a and Dnmt3b have equal affinity for unmethylated and hemi-methylated DNA, suiting their *de novo* roles. Though primarily involved in establishing 5mC *de novo*, they have been shown to participate in maintenance methylation as well, hence why low-level methylation is retained in *Dnmt1* knockout models (Okano, Xie and Li, 1998; Rhee *et al.*, 2000). Dnmt3a and Dnmt3b can bind unmethylated DNA using their PWWP domains, but exactly how they choose their target sites is still unknown (Ge *et al.*, 2004). Mechanisms involving RNA interference have been proposed (Morris *et al.*, 2004), though this seems to be restricted to plants. Dmnt3a is predominantly recruited to intergenic regions by recognising and binding to H3K36me2 (Weinberg *et al.*, 2019). The structure of Dnmt3a, like Dnmt1, is auto-inhibitory; each protein must first bind to unmethylated H3 histone protein. This binding changes the conformation of its catalytic ADD domain which it uses to bind to unmethylated DNA, enabling it to convert cytosine to 5mC (Guo *et al.*, 2014). Dnmt3b is recruited to actively transcribed genes by binding to H3K6me3 with its PWWP domain (Baubec *et al.*, 2015).

The co-factor Dnmt3L localises to chromatin using its plant homeodomain-like domain (PHD) which recognises and binds unmethylated H3K4. The *de novo* and maintenance methyltransferases in turn recognise and bind to Dnmt3L (Dan and Chen, 2016; Jurkowska and Jeltsch, 2016).

Part of Dnmt recruitment is controlled by the indirect negative regulation by other epigenetic marks. For example, most active gene promoters, and therefore CGIs, are enriched for histone protein H3 lysine 4 (H3K4) methylation at the nucleosome (Mikkelsen *et al.*, 2007). This mark protects these regions from Dnmts by physically preventing them from binding, thus inhibiting DNA methylation (Jia *et al.*, 2007; Mikkelsen *et al.*, 2007; Ooi *et al.*, 2007; Weber *et al.*, 2007).

Occlusion by other proteins has also been proposed as a regulatory mechanism. Transcription factors are an example, although there is more evidence to show they primarily protect CGIs from recruiting Dnmts to maintain their unmethylated state. As differentiation proceeds and

transcription factors are downregulated, the CGIs they previously protected are open for Dnmt binding (Lienert *et al.*, 2011).

Dnmts may also be recruited to target sites by methyl-binding proteins (MBDs). There are three main families: the MBD family proteins, Uhrf family proteins, and zinc-finger proteins.

MBD proteins contain a methyl-CpG-binding domain that gives them high affinity for single methylated CpGs (Nan, Meehan and Bird, 1993). MeCP2 was the first MBD to be identified, then MBDs 1-4 were described (Meehan *et al.*, 1989; Lewis *et al.*, 1992; Hendrich and Bird, 1998). Many MBDs, once bound to methylated DNA, recruit transcriptional repressors using their conserved transcriptional repressor domain (TRD) (Nan *et al.*, 1998; Ng *et al.*, 1999). MeCP2, uniquely among MBDs, also has a role in methylation maintenance. It binds Dnmt1 to its TRD and can recruit it to hemi-methylation to perform maintenance methylation (Kimura and Shiota, 2003).

Uhrf (Ubiquitin-like, containing PHD and RING finger domain) proteins bind methylated cytosines via their SET- and RING-associated DNA-binding domain (Hashimoto *et al.*, 2008, 2009). The family includes Uhrf1 and Uhrf2. They bind Dnmt1 and recruit it to hemi-methylated DNA during DNA replication or maintenance (Bostick *et al.*, 2007; Sharif *et al.*, 2007; Achour *et al.*, 2008). Uhrf1 appears to be vital for Dnmt function during replication as deletion of the gene leads to embryonic lethality (Muto *et al.*, 2002).

The last MBD family contains the zinc-finger domain proteins such as Kaiso, Zbtb4 and Zbtb38 (Prokhortchouk *et al.*, 2001; Filion *et al.*, 2006). They, like the MBDs, are generally associated with repression of transcription in a DNA methylation-dependent manner.

1.3 Functions of DNA methylation

DNA methylation is generally associated with the silencing of gene expression. It is key for a number of processes including genomic imprinting, maintenance of X chromosome inactivation in female cells (Riggs, 1975; Gartler and Riggs, 1983; Skvortsova, Stirzaker and Taberlay, 2019), and repression of transposable elements.

DNA methylation controls gene expression directly by physically inhibiting transcription factor binding (Li and Zhang, 2014). The mark can also indirectly control expression by recruiting 5mC-binding proteins (MBDs) such as MBD1, MBD2 (Ng *et al.*, 1999) and MeCP2 (Meehan *et al.*, 1989), which in turn recruit co-repressor complexes which repel transcription factors (Boyes and Bird, 1991). MeCP2 in particular is important for this at CpG-dense regions found at peri-centromeric heterochromatin (PCH) (Meehan *et al.*, 1989). DNA methylation also contributes to gene expression silencing by inducing heterochromatinization, the compaction of DNA into transcriptionally inaccessible, silenced regions (Sager and Kitchin, 1975).

DNA methylation can also mediate gene repression via interactions with other epigenetic mechanisms, including histone modifications. Dnmt1 and Dnmt3a can inhibit gene expression by recruiting histone methyltransferase Suv39h1, a protein which restricts gene expression by methylating histone 3 lysine 9 (H3K9) (Fuks *et al.*, 2003). Both Dnmts can also bind to histone de-acetylases to stimulate the removal of histone acetylation which in turn causes DNA to pack more tightly around the histone, restricting transcription factor access (Fuks *et al.*, 2000; Geiman *et al.*, 2004). In this way DNA methylation can also repress transcription by indirectly modifying chromatin compaction. Additionally, Dnmt3L binds to H3 histone tails and recruits Dnmt3a/b, resulting in DNA methylation at that site (Ooi *et al.*, 2007).

Some studies cast doubt onto the importance of DNA methylation for gene silencing. Two independent studies showed that embryonic carcinoma cells infected with murine MuLV viral DNA quickly silenced the gene following

infection, but DNA methylation was not deposited on the ectopic DNA immediately, it took a few days to become established (Gautsch and Wilson, 1983; Niwa *et al.*, 1983). This showed that DNA methylation was not the primary silencing mechanism. It was then hypothesised that, instead of establishing silencing, 5mC acts to 'lock in' a previously established silenced state. What mechanisms are responsible for this initial establishment, and therefore recruit the *de novo* DNA methylation machinery, is still debated today.

An important tool that was commonly used for studying DNA methylation function was the drug 5-Azacytidine (5aza). 5aza is a nucleoside analog of cytosine which can be incorporated into DNA or RNA in cytosine's place (Pačes, Doskočil and Šorm, 1968). Experiments showed that culturing ESCs or mouse fibroblasts with 5aza led to distinct morphological changes indicative of cellular differentiation (Benedict *et al.*, 1977; Constantinides, Jones and Gevers, 1977). This suggested 5aza incorporation led to re-activation of silenced genes and could therefore act as a specific inhibitor of DNA methylation (Friedman, 1979). It was thought that it was able to do this by impeding the advancement of Dnmts along DNA (Jones and Taylor, 1980). These studies provided further evidence that DNA methylation had a role in regulating development. One important 5aza study using ESCs showed that DNA methylation was not required during early development but that late development was dependent on it; this led to the description of the 'epigenetic paradox' of ESCs (Festuccia, Gonzalez and Navarro, 2017).

X inactivation

Riggs *et al*. proposed that DNA methylation may have a role in X-inactivation and suggested this could possibly also be applied more widely to the rest of the genome (Riggs, 1975). DNA methylation does have a role in maintenance of X inactivation, but it is not required and it is not involved in the initiation of inactivation (Agrelo and Wutz, 2010). However, it is a good example of rare CGI methylation, induction of heterochromatinization as a

result of DNA methylation and the consequent stable silencing which are all observed on the inactive X (Xi) chromosome. Loss of *Dnmt1* does lead to re-activation of the inactive X chromosome during development (Panning and Jaenisch, 1996).

Transposable element silencing

In the 1980s it was suggested that 5mC may have a role in regulating transposable elements, particularly during embryonic development (Stuhlmann, Jähner and Jaenisch, 1981). Harbers *et al*. found evidence for this by transfecting plasmids containing unmethylated viral DNA into fibroblasts. Following the transfection they could detect the viral transcripts suggesting the ectopic genes were active, whereas the endogenous counterparts were not transcribed and were silenced by DNA methylation (Harbers *et al.*, 1981). By testing the effect of injecting the unmethylated virus DNA into embryos at different stages of development it was deduced that pre-implantation embryos rapidly inactivated the ectopic sequence by *de novo* methylation, whereas in post-implantation embryos the ectopic sequence remained unmethylated (Jähner *et al.*, 1982). This suggested that the Dnmts were more active during pre-implantation development and were then downregulated following implantation, emphasising the importance of the mark during early development.

TEs will be described in more detail later in the chapter within the context of embryonic development and the dynamic DNA methylation changes occurring during this time.

Genomic imprinting

One key role of DNA methylation is the establishment and maintenance of genomic imprints. Imprinted genes are those which are expressed from only one allele in a parent-of-origin-specific manner (Surani, Barton and Norris, 1984; Uribe-Lewis *et al.*, 2011; Thamban, Agarwaal and

Khosla, 2020). This involves the differential methylation of CGIs during gametogenesis and early embryonic development (Wutz *et al.*, 1997; Caspary *et al.*, 1998; Zwart *et al.*, 2001; Kantor *et al.*, 2004; Choi *et al.*, 2005; Moore, Le and Fan, 2013).

Imprints are established early in development and maintained across cell division. Approximately 100 imprinted genes have been found in humans, and 150 in mouse, many of which overlap in both species (SanMiguel and Bartolomei, 2018). Each allele acquires a different epigenetic mark pattern to control which one is transcribed. These marks are deposited at regions within the imprinted gene locus known as the Imprint Control Region (ICR). The differential epigenetic modifications on each allele act *in cis* so that one allele remains active while the other is silenced (Thamban, Agarwaal and Khosla, 2020). ICRs remain methylated during embryonic development when the rest of the genome is undergoing 5mC erasure. ICRs are protected from active demethylation by proteins such as Developmental pluripotency-associated 3 (Dppa3) protein, which is highly expressed in oocytes and primordial germ cells (PGCs) (Nakamura *et al.*, 2006)], and zinc finger protein 57 (Zfp57) (Li *et al.*, 2008).

DNA methylation was the first epigenetic mark to be connected with the control of expression of imprinted genes (Reik *et al.*, 1987; Thamban, Agarwaal and Khosla, 2020). This was determined through knockout mouse model studies of *Dnmt1*; loss of the methyltransferase resulted in DNA hypomethylation and loss of expression of the imprinted gene (Li, Beard and Jaenisch, 1993; SanMiguel and Bartolomei, 2018).

Imprinted genes influence multiple developmental pathways including neurodevelopment, stem cell proliferation, and differentiation pathways (Thamban, Agarwaal and Khosla, 2020). Many imprinted genes are important for promoting foetal growth, including *Igf2*, *Igf2r* and *Dlk1* (Cassidy and Charalambous, 2018). Others are important for restricting it, including *Grb10* and Cdkn1c. Deleting or duplicating an imprinted locus, or removing the ICR which regulates the gene's expression, results in disruption to normal

development including placental abnormalities, increased apoptosis, and aberrant differentiation (Fowden *et al.*, 2011).

Human diseases associated with DNA methylation

Disruption of normal Dnmt activity is associated with many human diseases, including cancers, diabetes, and neurological disorders. Many involve the aberrant expression of imprinted genes. Fragile X syndrome is caused by abnormal DNA methylation in the FMR1 gene found on the X chromosome, resulting in a form of mental retardation (Verkerk *et al.*, 1991). Prader-Willi syndrome and Angelman syndrome are both associated with improper methylation of an imprinted allele which causes significant mental impairments (Buiting, 2010).

Some diseases are caused by specific mutations affecting the Dnmt proteins themselves. Immunodeficiency-centromeric instability-facial anomalies (ICF) syndrome is a chromosomal breakage disorder caused by a loss-of-function mutation in the human DNMT3B gene which results in hypomethylation of centromeric satellite DNA (Zoghbi and Beaudet, 2016). Patients with the disease suffer with immunodeficiencies, chromosomal abnormalities, and sometimes cranio-facial defects. Patients are likely to die during childhood due to immunodeficiency complications.

Together these provide clear evidence that changes in DNA methylation can contribute to serious human pathologies, making understanding the regulation of the epigenetic system an important target for new disease therapies.

1.4 DNA methylation during embryonic development

An overview of embryonic development

Mammalian embryonic development is a complex and carefully orchestrated process. It begins with the fusion of an oocyte and sperm to form a totipotent zygote. This is followed by the erasure of DNA methylation from both parental genomes to remove the germ cell differentiation marks, resetting the newly formed genome (Xu *et al.*, 2021). Oocyte activation, maternal-to-zygotic transition (MTZ) and zygotic gene activation (ZGA) are all co-ordinated during this time and are required prior to the first cell-fate decision and the formation of the first differentiation lineages in the developing embryo. At the 8-cell stage the first differentiation step takes place; this involves the emergence of trophectoderm cells from the inner cell mass, which eventually form the placenta. The remaining inner cell mass cells of the blastocyst then divide into pluripotent epiblast cells, which form all the foetal cells, and cells which form the primitive endoderm which eventually give rise extra-embryonic yolk-sac (Chazaud and Yamanaka, 2016). At implantation the primitive streak develops, and this triggers gastrulation (Kojima, Tam and Tam, 2014).

Following onset of gastrulation the developing embryo starts to develop the three primary germ layers from which the remaining adult tissues arise; these layers are the mesoderm, endoderm, and ectoderm (Murry and Keller, 2008). Muscle cells, blood cells, and cardiomyocytes arise from the mesoderm. The organ linings arise from the endoderm. The ectoderm gives rise to neural and skin cells. The precise regulation of germ layer induction is still not fully understood.

The processes which occur during embryonic development are regulated by a series of epigenetic mechanisms, including DNA methylation, histone modifications, and chromatin conformation. They work together to control timely patterns of gene expression which inform cell identity to enable the formation of a new individual animal from the totipotent single-cell zygote (Xu *et al.*, 2021).

Epigenetic reprogramming

Riggs was the first to propose the concept of epigenetic reprogramming when he was investigating X chromosome inactivation in female mammals and discovered a role for DNA methylation (Riggs, 1975). The ability to reprogram epigenetic signatures was then demonstrated in the 1970s when the first nuclear transfer experiments were being performed. In a landmark experiment, Gurdon showed that by removing the nucleus from a somatic cell and placing it in an enucleated oocyte, a process known as somatic cell nuclear transfer (SCNT), the somatic nucleus would become epigenetically reprogrammed to a totipotent-like state(Gurdon and Uehlinger, 1966).

We now know that during embryonic development mammalian cells undergo two waves of epigenetic reprogramming wherein the DNA methylation signature is almost completely erased throughout the genome to then be re-established by Dnmt1(SanMiguel and Bartolomei, 2018). The first wave occurs in the zygote and early pre-implantation embryo. The second wave occurs in the developing germ cells. The dynamics of DNA methylation patterns during embryo development are summarised in figure 1.3. Histone modifications are also reprogrammed but they fall outside the scope of this thesis.

Epigenetic reprogramming is required during development to enable the erasure of somatic imprints from the parental gametes in the newly formed zygote, so that the new pattern of epigenetic marks can be established appropriate to the sex of the embryo (Tada *et al.*, 1997; SanMiguel and Bartolomei, 2018). As the zygote continues to develop into an embryo, the next generation of germ cells form from the pre-implantation epiblast cells (Ginsburg, Snow and McLaren, 1990; Anderson *et al.*, 2000). In order for the germ cells to take on the oocyte- or sperm-specific epigenetic pattern the epiblast-specific patterns must be erased (Monk, Boubelik and Lehnert, 1987; SanMiguel and Bartolomei, 2018). This erasure also protects the next generation from inheriting aberrant epigenetic patterns during germ cell development.

In the hypomethylated blastocysts (following reprogramming in the zygote), methylation is retained at specific regions of the genome: imprinted genes, Intracisternal A particle (IAPs) retrotransposon elements, and a subset of gene promoters associated with gamete generation (Kim *et al.*, 2004; Saitou and Yamaji, 2012). DNA methylation is gradually re-established up to approximately E6.5.

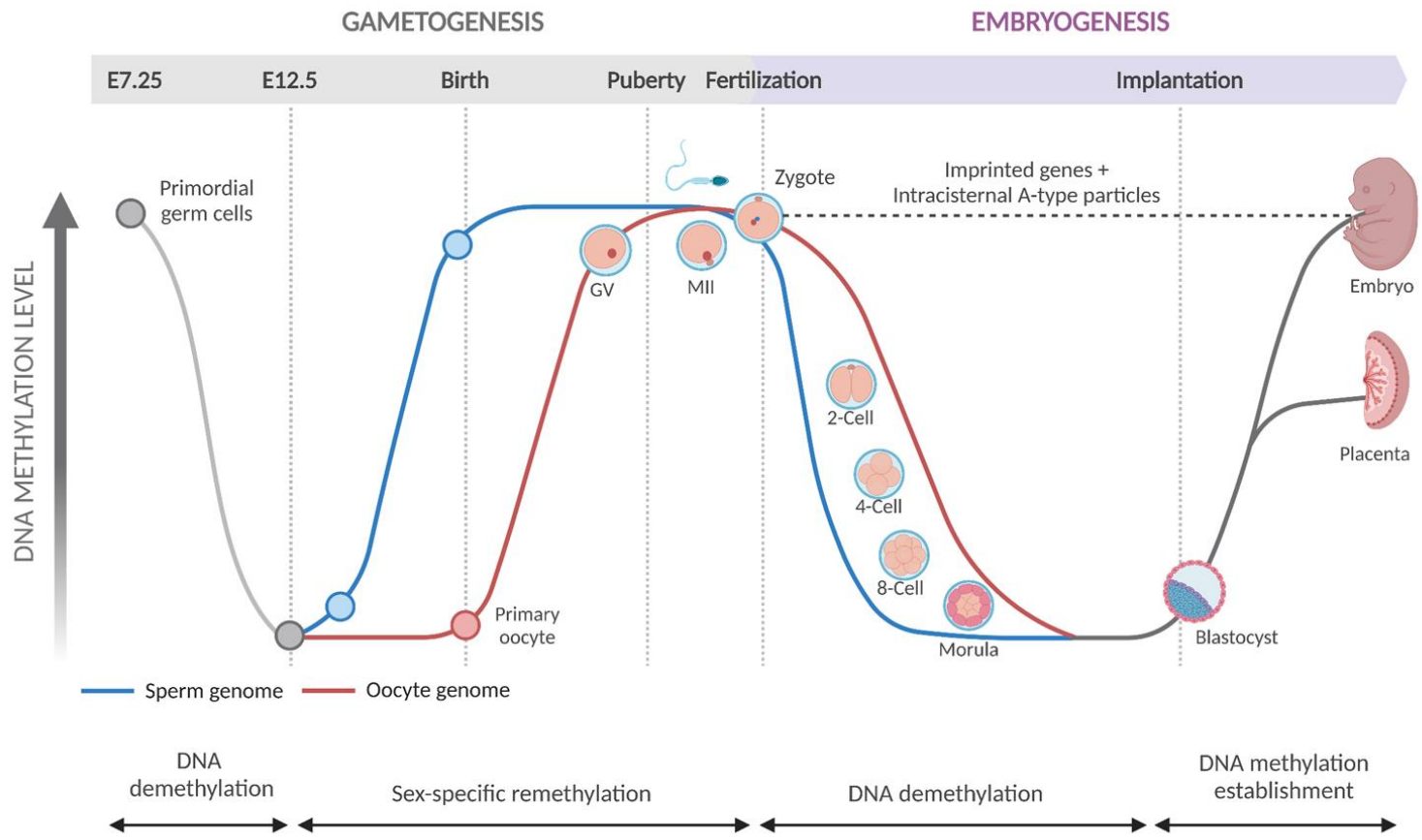


Figure 1.3. Changes in DNA methylation during mammalian embryonic development. Adapted from "DNA Methylation Levels During Mammalian Development", by BioRender.com (2022). Retrieved from https://app.biorender.com/biorender-templates

PGC formation

Proper regulation of DNA methylation, demethylation and re-methylation is key for controlling the timing of emergence of primordial germ cells (PGCs) and appropriate gene expression programs during gametogenesis. In mouse, the PGCs are specified around E7.25 (SanMiguel and Bartolomei, 2018). Following activation of the PGC gene program, the cells proliferate and beginning to migrate towards the developing genital ridge. Once there, the PGCs undergo a wave of demethylation at E8 to erase the epiblast-cell signature and enable establishment of the germ-cell signature, reaching a DNA methylation low-point at E13.5. Around E10.5 the DNA methylation present at imprinted regions which were protected from the previous round of reprogramming in the blastocyst is removed (Saitou and Yamaji, 2012). Other initially resistant regions, such as some repetitive elements and X chromosome CGIs, become hypomethylated in a second wave by E13.5 (Seisenberger *et al.*, 2012). This is through a combination of downregulation of Dnmt3a and Uhrf1, which in turns impairs the targeting of Dnmt1, relocalisation of Dnmt3b to the cytoplasm, and the activity of Tet1 and Tet2 demethylases (Hackett *et al.*, 2013). The timing of this demethylation is carefully regulated to ensure premature differentiation of the PGCs is prevented (Hargan-Calvopina *et al.*, 2016).

Calvopina *et al*. showed that Dnmt1 is responsible for preserving the 5mC signature at ICRs, IAPs and meiotic gene promoters through the first demethylation stage in developing PGCs (Hargan-Calvopina *et al.*, 2016). Using a conditional Cre-LoxP *Dnmt1* mouse cell line they were able to show that loss of Dnmt1 led to a reduction in the number of PGCs formed by the end of the second wave of demethylation. Instead, they found that the cells had undergone early germline differentiation into oocytes and prospermatogonia. They concluded that loss of Dnmt1 leads to precocious germline differentiation between E11.5-13.5. They also hypothesised that the reduction in PGCs formed could have been a consequence of increased expression of TEs as a result of loss of Dnmt1, particularly of IAPs, and that

this was consistent with a previous study in Dnmt1 knockout mouse somatic cells (Walsh, Chaillet and Bestor, 1998).

Dazl is a gene specifically regulated by DNA methylation (Hackett *et al.*, 2012; Chen *et al.*, 2014). It is a key regulator of PGC formation as it controls entry into meiosis and repressed pluripotency gene expression to enable germ cell formation (Hargan-Calvopina *et al.*, 2016). As such it is typically methylated and silent in all other cell types except germ cells, making it a useful indicator of Dnmt activity.

Following completed 5mC erasure by E13.5, re-methylation in the PGCs is initiated. This includes genome-wide DNA methylation, and to establish the sex-specific imprints at ICRs. In male germ cells ICRs are re-methylated rapidly after de-methylation ends and this program is completed by birth (Stewart, Veselovska and Kelsey, 2016). In female germ cells the maternal-specific imprints are not re-established until after birth during oocyte development (Lucifero *et al.*, 2004).

Transposable elements

Transposable elements (TEs) are repetitive sequences of DNA that are able to mobilise and integrate into multiple sites across the host genome (Molaro and Malik, 2016). They make up approximately 45% of the mammalian genome (Schulz, Steinhoff and Florl, 2006). Most of them are found in intergenic regions are silenced by DNA methylation (Walsh, Chaillet and Bestor, 1998). They are divided into two main classes. Class 1 elements, or retrotransposons, use a ‘copy-paste’ mechanism to move their DNA around the genome via an RNA intermediate and reverse transcription into cDNA which is then integrated at a different site (Bourque *et al.*, 2018). This class includes Long Terminal Repeats (LTRs) and non-LTR retrotransposons such as Long and Short Interspersed Nuclear Elements (LINES and SINES). Class 2 elements, or DNA transposons, use a DNA intermediate to mobilize their transposition (Bourque *et al.*, 2018).

The transposable nature of TEs makes them potentially deleterious to cells and their expression is associated with multiple cancers and cellular apoptosis (Bourque *et al.*, 2018). In normal cells they are silenced by DNA methylation to prevent their transposase activity which may trigger aberrant gene activation, silencing or mutation (Xu *et al.*, 2021). For example, some L1 retrotransposon insertions can disrupt tumour suppressor genes and oncogenes which in turn drives cancer formation (Bourque *et al.*, 2018). They can also harm the host in transposition-independent ways such as by interfering with transcription and processing of mRNA; by TE-transcribed protein endonuclease activity causing DNA double strand breaks leading to compromised genome integrity; by triggering an innate immune response when extra-chromosomal TE DNA accumulates which can lead to the formation of an autoimmune disease, such as Aicardi–Goutières syndrome (Bourque *et al.*, 2018). IAPs are the most aggressive class of TEs (Walsh, Chaillet and Bestor, 1998) and are maintained in the silenced state by Dnmt1, even when the rest of the genome is hypomethylated during embryonic development (Gaudet *et al.*, 2004). When Dnmt1 is lost the resulting hypomethylation leads to expression of IAPs (Walsh, Chaillet and Bestor, 1998; Hutnick *et al.*, 2010).

TEs are typically hypermethylated or silenced by repressive histone modifications such as H3K9me2/3 in order to induce a silenced state, and this is protected during periods of global demethylation which occur as part of embryonic development (Xu *et al.*, 2021). The only exception during development is at the 2-cell stage when embryonic genome activation is occurring in mouse embryos and they are highly expressed (Evsikov *et al.*, 2004). Proper regulation of TEs is essential for normal development (Peaston *et al.*, 2004).

Some TE insertions can result in disruption to gene function or harmful chromosomal rearrangements which can result in disease in humans (Gökbuget and Blelloch, 2019). These insertions are particularly harmful in they occur in the germ cells and can lead to sterility. It has been hypothesised that DNA methylation evolved as one of the primary

mechanisms for repressing TE activity in higher eukaryotes (Deniz, Frost and Branco, 2019), and that the system was adapted for other functions over time. This is supported by multiple studies which have shown that IAP type retroviruses increase in expression in the absence of Dnmt1. However, it does not appear to be a global rule as 5mC has been shown to be dispensable for LINE1 silencing in embryonic fibroblasts, which are silenced by Dnmt-independent mechanisms (Dunican *et al.*, 2013; Deniz, Frost and Branco, 2019).

Murine ERV-L elements (MERVLs) make up approximately 10% of both the human and mouse genome (Stocking and Kozak, 2008). They are expressed in 2C-cells where they help drive the expression of transcripts required for zygotic genome activation (ZGA) and totipotency (Kigami *et al.*, 2003; Svoboda *et al.*, 2004). Their expression is strictly controlled in part by *Dux*, another signature gene of the 2C state (Hendrickson *et al.*, 2017; Yang *et al.*, 2020). *Dux* in turn is driven by *Dppa2* and *Dppa4* (Eckersley-Maslin *et al.*, 2019). MERVL elements are also activated by *Zscan* family genes, particularly *Zscan4c*; these genes are known to be key drivers of the 2C/4C-like transcription network (Eckersley-Maslin *et al.*, 2019).

Not all transposable element activities are harmful; some transposable elements play an important role in embryonic development. Long interspersed element 1 (LINE1) elements are the most common retroelements found in mammals (Richardson *et al.*, 2015). They are highly expressed following fertilization and peak at the 2-cell stage. They have a role in maintaining an open chromatin conformation, and they are important for suppressing Dux expression and therefore driving the exit from the 2C-state (Deniz, Frost and Branco, 2019).

Embryonic stem cells

Since they were first isolated from the inner cell mass (ICM) of mouse blastocysts in the 1980s (Evans and Kaufman, 1981), embryonic stem cells (ESCs) have revolutionised molecular biology. *In vivo*, ESCs within the

developing embryo differentiate to form the ~200 tissues of the adult body, except the extra-embryonic cells of the placenta (Murry and Keller, 2008). For this reason, they are described as pluripotent. In vivo populations of these cells occur only transiently (Burdon *et al.*, 1999) but it was found that if they were maintained on a monolayer of feeder cells, mouse embryonic fibroblasts (MEFs), this would prevent them from differentiating and maintain their self-renewal *in vitro.* By defining the chemical cocktail required to maintain them in the pluripotent state scientists unlocked the ability to culture ESCs without MEFs and almost indefinitely. These lab cultured ESCs could contribute to chimera formation when injected into the ICM of mouse blastocysts and were found to be able to incorporate into all of the lineages formed, including germ cells (Smith, 1992). These advances opened up a whole field of opportunities to study the mechanisms underpinning not just epigenetics but embryonic development and cancer.

One advantage of ESCs over in vivo mouse studies is that ESCs can be genetically manipulated to create conditional gene expression or gene knockout cell lines. These cell lines can be used to control the timing or dose of expression of specific genes of interest or remove the genes altogether, making them a versatile tool for the study of molecular mechanisms governing gene expression. They are especially useful for studying genes whose genetic manipulation leads to mortality in adult mice. These cell lines can then be stored in liquid nitrogen, protecting against accidental loss, and they can be more easily shared between research groups for consistent research, and they reduce the requirement for costly mouse experiments. Their ability to differentiate into a wide range of cell types also makes them a promising tool for regenerative medicine.

The key to maintaining ES cells in vitro without the need for feeder cells was discovered in the 1980s. The glycoprotein leukaemia inhibitory factor (LIF) was found to block ESC differentiation when included in the cell culture media (Burdon *et al.*, 1999). LIF suppresses differentiation by binding to the receptor subunits LIFR and gp130 on the ESC membrane, which in turn brings together the receptor-associated kinase JAK. Binding of JAK leads to

activation of the STAT and MAPK signalling pathways. STAT3 directs gene expression programmes required for maintaining pluripotency in ES cells, such as *Oct4* expression, which is absolutely required for the pluripotent state to occur (Hackett and Surani, 2014; Patra, 2020). Thus, LIF maintains pluripotency by activating the JAK-STAT3 pathway which drives pluripotency gene expression which inhibit differentiation. ESCs cultured in calf serum/LIF media are similar to stem cells of the pre-implantation epiblast (E3.5) and as such are considered to be representative of the naïve state of embryonic pluripotency (Boward, Wu and Dalton, 2016).

ESCs have some unique features not shared with somatic cells. They have a relatively fast cell cycle, replicating approximately every 12 hours as opposed to roughly 24 hours for somatic cells (Festuccia, Gonzalez and Navarro, 2017). This is due to them having no G1/S phase transition checkpoint, partly due to low levels of expression of the required G1 cyclins D and E, which results in a shortened G1 phase (Festuccia, Gonzalez and Navarro, 2017). Upon embryoid body (EB) differentiation *in vitro* expression of the G1 cyclins is increased leading to the lengthening of the G1 phase, and a decrease in the rate of cell division (Festuccia, Gonzalez and Navarro, 2017).

Also, ESCs can continue to self-replicate in the absence of DNA methylation (Tsumura *et al.*, 2006; Festuccia, Gonzalez and Navarro, 2017). They have lower global DNA methylation levels compared to somatic cells, but 5mC is retained at satellite repeats, imprinted genes, and transposable elements. Mouse ESCs can tolerate the absence of Dnmt1, Dnmt3a and Dnmt3b singly or together (Triple knockout cells, or TKOs (Tsumura *et al.*, 2006)), and can exist in a hypomethylated state. In TKOs there are no immediate detrimental effects as other epigenetic systems are implemented (Festuccia, Gonzalez and Navarro, 2017). However, the ability to differentiate is compromised in single knockout and TKO ES cells and results in their death, further highlighting the importance of 5mC for development (Li, Timothy H Bestor and Jaenisch, 1992; Li, Beard and Jaenisch, 1993; Lei *et al.*, 1996; Jackson-Grusby *et al.*, 2001).

ESCs are able to maintain their identity across many cell divisions independently of epigenetic systems as, unlike somatic cells, their identity is determined by genetics not epigenetics (Festuccia, Gonzalez and Navarro, 2017). This means their identity is regulated by the expression of transcription factor genes *Oct4*, *Sox2*, *Klf4* and *c-Myc* which make up the core pluripotency network. This network of (OSKM) genes, known as the Yamanaka factors, are sufficient to reprogram somatic cells to a pluripotent-like state when they are ectopically expressed (Takahashi and Yamanaka, 2006; Boward, Wu and Dalton, 2016). Their combined expression is sufficient to overcome and inhibit differentiation. This challenges the idea that gene repression by epigenetic systems, such as DNA methylation, is vital for mammalian cells to survive but endows us with a powerful tool to dissect how those mechanisms come into play once ESCs are stimulated to differentiate and begin to exit pluripotency.

Embryoid bodies

One of the earliest ESC differentiation models used was the generation of embryoid bodies (EBs). Embryoid bodies are non-adherent, three-dimensional aggregates of embryonic stem cells which are typically used as an *in vitro* model of embryonic development due to their ability to produce cells of all three germ layers in a process mirroring gastrulation (Doetschman *et al.*, 1985; Desbaillets *et al.*, 2000; Kurosawa, 2007). EBs as a model of differentiation therefore provide a protocol for testing gene function, particularly as the ESCs exit pluripotency and begin to differentiate. The model provides experimental accessibility which is not possible with *in vivo* models (Brickman and Serup, 2017); ESCs which have been genetically manipulated to contain conditionally expressed genes or in which target genes have been knocked out but which would normally result in a failure to develop an embryo *in vivo* can be differentiated into EBs and the impacts on differentiated analysed.

There are multiple methods of deriving EBs. In all the methods the EBs are generated by culturing ESCs without LIF or feeder cells present, and in a container to which they cannot adhere to a surface. The most commonly used techniques are liquid suspension in a culture dish, culture in a semi-solid media such as methylcellulose, and culture in hanging droplets (Kurosawa, 2007). In the liquid suspension method a single-cell suspension of ESCs is seeded in media in a culture dish with a hydrophobic coating. This is the most efficient method for quickly producing large numbers of EBs, but the EBs produced vary significantly in size (Doetschman *et al.*, 1985).

Alternatively, EBs may also be generated by seeding ESCs in a semi-solid matrix of methylcellulose where single ESCs can form entire EBs without aggregating with other cells (Kurosawa, 2007). This is useful for producing clonal EBs as the matrix prevents them from coalescing. However, as the EBs start from single cells obtaining mature EBs requires more time and extracting the fragile EBs from the methylcellulose media can be difficult.

The final method involves diluting a single-cell suspension of ESCs to a known cell density and hanging drops of a set size from the underside of a culture dish lid (Kurosawa, 2007). The droplets are left in suspension to allow the ESCs to aggregate under gravity at the bottom of the droplet without the risk of the EBs adhering to a surface. The base of the culture dish over which the lid is inverted contains PBS to prevent the droplets from drying out. As the cells aggregate they form the characteristic simple EB spheroid structure. After three to four days in suspension the EBs are moved to culture plates where they are allowed to adhere to the surface and continue to mature until they are harvested. With this method the number of cells per droplet is known and thus the sizes of the EBs harvested are consistent, and no specialist growth matrix is required.

Many studies have reported the derivation of multiple cell types from the three germ layer lineages using EB differentiation, including neural progenitor cells, cardiomyocytes, hepatic cells, and germ cells (Kurosawa, 2007; Lynch, Mazzotta and Hoppler, 2018). This, combined with the relative simplicity of

the hanging droplet culture method, makes EBs an attractive model for investigating the genetic to epigenetic regulation transition which occurs when ESCs begin to differentiate and exit pluripotency.

1.5 Open questions

Despite the decades of research into DNA methylation, many questions remain. This includes improving our understanding of the exact mechanisms of recruitment of the Dnmt and TET proteins. Although DNA methylation is correlated with gene silencing, it is still unclear whether 5mC or the silenced state comes first. How are they targeted to specific substrates, particularly the *de novo* Dnmts? How does their regulation and activity change during differentiation? Why can ESCs tolerate the absence of the Dnmts and 5mC but somatic cells cannot? What function of DNA methylation is essential in somatic cells? Why do embryos lacking Dnmts die? Finally, at what point during embryonic development does the DNA methylation system switch from being dispensable to essential?

1.6 Thesis aims

Epigenetic mechanisms, such as DNA methylation, are thought to have vital roles in controlling tissue-specific gene expression programs, maintaining genomic imprints, and protecting genomic stability by silencing retroelements to prevent dangerous transposition events. Loss of these mechanisms is associated with the failure of embryonic development and somatic cell death, and mutations in epigenetic factors are implicated in multiple human disease pathologies including cancers. However, the pluripotent ESCs are able to maintain their identity and self-renew in vitro independently of epigenetic repressor mechanisms. This suggests there is a point during early embryonic development at which epigenetic mechanisms take over the regulation of cell identity. Understanding when and how this switch occurs will have important implications for our understanding of the fundamental mechanisms governing gene expression, and thus our ability to address aberrant regulation in disease.

In this thesis I will investigate the role of DNA methylation during the transition from the pluripotent to differentiated cell state using genetically modified ESC lines and *in vitro* differentiation models. I aim to:

- Develop a system in which I can control DNA methylation levels during ESC differentiation
- Determine the window during differentiation at which epigenetic regulation by DNA methylation becomes key
- Investigate the effects of loss of DNA methylation on lineage commitment
- Investigate the relationship between the core pluripotency network and epigenetic regulators

2.1 Cell culture

2.2.1 Cell culture

All cell lines used during this project were grown in tissue culture grade flasks (Corning) or plates (ThermoFisher) coated with 0.2% (w/v) porcine gelatin (Sigma; made up with PBS) and incubated at 37°C with 5% CO_2. This excludes mouse embryonic fibroblast cells (MEFs), which were grown without gelatin. Basal ES cell media components are shown in Table I. Mouse ES cell lines used in this project are summarised in Table II. 40-80µL B-mercaptoethanol (50mM, Gibco) and 60µL ESGRO LIF (Millipore, final concentration 1500U/mL) were added to fresh, pre-warmed 40mL aliquots of the ES basal media to make ser/LIF media.

TABLE I: Components of basal ESC media

Reagent	Manufacturer	Volume	% Equivalent
DMEM (High glucose)	Gibco	500mL	
FCS	Hyclone	75mL	15%
Pen/Strep	In-house technical services	5.6mL	1%
L-glutamine	Gibco	5.6mL	1%
NEAA (100X)	Sigma	5.6mL	1%
Sodium pyruvate (100mM)	Sigma	5.6mL	1%

Cells were passaged every two days, or when cell confluency exceeded 85%, using trypsin detachment reagent (Sigma). To passage them, the cells were washed with PBS (Sigma; without added calcium or magnesium) and then incubated in an appropriate volume of trypsin at room temperature for 2-3mins, or until most of the cells had visibly detached. The trypsin was neutralised by adding 1.5x volume ES basal media. The cell suspension was moved to a falcon tube and centrifuged gently at 1000rpm

for 5mins. The cells were then resuspended in fresh, warm ser/LIF and placed in a new flask. Periodically, 500µL media was taken from each flask and submitted for mycoplasma testing (done in-house by technical services), along with a fresh ser/LIF media blank sample.

To harvest cell pellets for extraction, trypsinised cells were spun down, washed with PBS, centrifuged again, then aspirated and moved immediately to dry ice to snap freeze. Pellets were stored at -20°C.

For dox experiments, 5mg/mL dox solution stock was made by dissolving 74mg doxycycline powder (Sigma) in 14.8mL ddH_2O. The solution was filtered using a 0.2µm filter (Millipore), aliquoted, and stored at -20°C. To induce hypomethylation *Dnmt1*$^{tet/tet}$ mESCs were cultured for at least 4 days with 2µg/mL dox, prepared by diluting 16µL dox stock in a 40mL aliquot of ESC ser/LIF media or by adding 2µL straight into a 5mL cell suspension during passaging. Doxycycline powder is dissolved in water; initial experiments included the addition of an equivalent volume of water to untreated cell media but as this proved to be non-toxic to the cells this was discontinued in later experiments.

TABLE II: Details of mouse ESC lines used				
Name	Feature	Source	B-mercaptoethanol	Reference
Dnmt1$^{tet/tet}$	R1 ESCs with dox-regulatable *Dnmt1*$^{tet/tet}$ alleles	Gift from Richard Chaillet	50μM	(Borowczyk *et al.*, 2009)
J1	129/terSv wild-type ESC line	Available in-house	50μM	(Li, Timothy H Bestor and Jaenisch, 1992)
E14	129/Ola wild-type ESC line	Available in-house	50μM	(Hooper *et al.*, 1987)
TKO	ESC cell line lacking *Dnmt1*, *Dnmt3a* and *Dnmt3b* genes	Available in-house	50μM	(Tsumura *et al.*, 2006)
EBRTcH3 (AES0119)	129/Ola wild-type ESC line	Riken cell bank	100μM	(Matoba *et al.*, 2006)
2TS22C (AES0125)	129/Ola ESCs with *Sox2*$^{tet/tet}$	Riken cell bank	100μM	(Masui, Nakatake, Toyooka, Shimosato, Yagi, Takahashi, Okochi, Okuda, Matoba, Alexei A Sharov, *et al.*, 2007)
ZHBTc4 (AES0136)	129/Ola ESCs with *Oct4* $^{tet/tet}$	Riken cell bank	100μM	(Matoba *et al.*, 2006)
EGFP-MBD-nls (AES0134)	E14 ESCs which produce EGFP-tagged MBD1 protein	Riken cell bank	100μM	(Kobayakawa *et al.*, 2007)

2.1.2 Freezing and thawing cells

For long term storage trypsinised cells were washed twice with PBS and resuspended in 500µL CryoStor freezing medium (Sigma) before moving to a cryotube. These were placed in a cell cooler and stored at -80°C for at least 24 hours before long term storage in liquid nitrogen. When required, cells were thawed rapidly in a 37°C water bath, then the cell solution was moved to a falcon tube and 1mL ES basal media was added slowly, drop-wise, to dilute the freezing media without causing osmotic shock. The total volume was made up to 10mL, then the suspension was spun down gently at 1000rpm for 5mins. The cells were resuspended in an appropriate amount of fresh ser/LIF and placed in a new flask.

2.2 Embryoid body culture

2.2.1 EB media

To make EB media, a 40mL aliquot of ES basal media (see Table I) was prepared and supplemented with an additional 2mL FCS (Hyclone) and 80µL B-mercaptoethanol (Gibco). The aliquots were used within 48 hours of preparation and pre-warmed in a 37°C water bath before use.

For EB dox experiments a 40mL aliquot of EB media was prepared and 1.6µL 5mg/mL dox stock added for a final concentration of 0.2µg/mL (low dose), or 16µL for 2µg/mL (high dose). Dox-supplemented media was always prepared fresh.

2.2.2 EB differentiation

On day 0 ESCs were detached from the culture flask by incubating in StemPro™ Accutase™ (Gibco) for up to 5 minutes at 37°C. The Accutase was inactivated by dilution with cell culture media or PBS. The cells were then dissociated by gently pipetting up and down, washed 2x with PBS, counted, and brought to 20 cells/µL density in 10mL EB media. 20µL droplets (approximately 400 cells in total per droplet) were then placed on the

underside of a 15cm culture dish lid using a multi-channel pipette. Approximately 88 droplets could be placed on each lid. 10mL PBS was placed in the base to prevent drying of the droplets during incubation. The lid was then carefully turned over and placed on the base so the droplets became suspended. The plates were kept in an incubator undisturbed at 37°C, 5% CO_2, for 72 hours to allow formation of EB aggregates. After this period the droplets were collected using a p1000, placed onto a suitable gelatinised surface and allowed to develop for up to 14 days. The media was refreshed every two days.

2.2.3 TUNEL assay to quantify apoptosis

TUNEL assays were performed using the APO-DIRECT apoptosis detection kit (Novus Biologicals, Cat. No. NBP2-31159) to measure the frequency of apoptotic cells in EBs with and without *Dnmt1* expression. The experiment was designed by me and carried out by PhD students Katie Pickup and Chetan Srinath. *Dnmt1*$^{tet/tet}$-derived EBs, which were treated with either 0.2µg/mL dox or no dox, were collected at day 7 of differentiation and fixed with 1% paraformaldehyde (PFA) for 60 minutes on ice to cross-link the DNA. They were then treated with the TUNEL kit following the manufacturer's instructions. The kit includes fixed negative and positive control cells. The kit also includes a propidium iodide (PI) counterstain; PI is a cell impermeant dye used to differentiate dead cells and cells which are in the early stages of apoptosis from healthy cells which do not take up PI. Stained EB cell suspensions were stored in 70% ethanol at -20°C until ready for FACS analysis. FACS analysis was performed on a BD LSRFortessa X-20 flow cytometer (BD Biosciences), calibrated using the kit control cells.

2.3 RT-PCR

2.3.1 RNA extraction

RNA was extracted from cell pellets containing no more than $1x10^7$ cells using the RNeasy Plus Mini kit (QIAGEN), following the manufacturer's

instructions. Briefly, the cell pellets were partially thawed on ice and resuspended in RLT Plus buffer. The suspension was homogenised by centrifuging through a QIAShredder column (QIAGEN). The eluate was transferred to a gDNA eliminator column and centrifuged. The eluate was then transferred to an RNeasy spin column and centrifuged again. Following three wash steps the columns were placed in fresh eppendorfs and spun at 13,000rpm for 1 minute to remove residual wash buffer. 20-30µL RNase free water was added to the columns which were then centrifuged to elute the RNA. The concentration of the RNA was quantified by NanoDrop.

2.3.2 RNA quality assessment | Gel electrophoresis

The quality of extracted RNA was checked by gel electrophoresis. Briefly, 500ng per sample, or 2-3µL RiboRuler high range RNA ladder (ThermoFisher), was incubated with an equal volume of 2X RNA gel loading dye (ThermoFisher) at 80°C for 2 minutes, then snap spun and moved straight to ice. The ladders and samples were run on a 0.8% TBE gel (see section 2.3.3 for details on gel preparation), which was then post-stained for 30 minutes with ethidium bromide and washed 2x 15 minutes in ddH_2O. Good quality mouse RNA extracts produce two bands at 4.7kb and 1.9kb, in a roughly 2:1 ratio, which represent the 28s and 18s rRNA fractions. The gel can also indicate the presence of any contaminating DNAs as a large molecular weight smear next to the wells, or degraded RNA as a smear below the lowest ladder bands. Samples with significant DNA contamination were re-extracted. Samples with excessive degradation or without the correct RNA band pattern were discarded.

2.3.3 Synthesis of cDNA

RNA was converted to cDNA by reverse transcription using SuperScript III reverse transcriptase (Invitrogen). Briefly, 1µg each RNA sample was made up to 12µL with RNAse-free water on ice and mixed with 1µL 200ng/µL random primers (Promega) and 1µL 10mM dNTP mix

(Invitrogen). The samples were vortexed and centrifuged briefly, then incubated at 65°C for 5 minutes and returned immediately to an ice bath. 4µL First Strand Buffer and 1µL 0.1M DTT were added, and the samples mixed and centrifuged. 1µL SuperScript III reverse transcriptase was added last, bringing the final volume to 20µL. Each sample had a duplicate which did not include the reverse transcriptase as a control (-RT). Samples and their duplicates were placed in a PCR thermocycler and amplified with the following cycling conditions: 25°C for 5 minutes, 50°C for 30 minutes, 55°C for 30 minutes, 70°C for 15 minutes, 10°C infinite hold. When the samples reached 10°C they were immediately moved to an ice bath. The sample volumes were made up to 100µL with Ultra-pure water (Invitrogen) and stored at -80°C.

2.3.4 Quantitative PCR

To analyse gene expression, cDNAs were amplified by qPCR using a polymerase which incorporates a fluorescent green dye into the newly formed DNA strands during amplification. The fluorescence signal is detected by the thermocycler machine and increases proportionally with the number of copies of the target amplicon during amplification. The sample is assigned a threshold cycle (Ct) value based on the cycle number at which the fluorescence signal passes a signal detection threshold, which is set above the background fluorescence level. The smaller the Ct value, the earlier in the number of cycles the signal passed the threshold, and the greater the initial amplicon concentration. In this way, qPCR can be used to analyse the expression of a gene with greater sensitivity than end-point PCR.

Samples were analysed by qPCR as follows. Per sample, 10µL SYBR select master mix (2X, Applied Biosystems), 7µL PCR-grade water, 1µL 10µM forward primer and 1µL 10µM reverse primer were mixed on ice to a final volume of 19µL. The reagent volumes were scaled to make a reaction master mix depending on the total number of samples needed per primer, including sample cDNAs, no reverse transcriptase controls (-RT), and no

template controls (NC). This reaction master mix was divided into a 96-well plate. To each well 1µL sample cDNA, no transcriptase cDNA control, or water (NC) was added, according to the experimental plate plan. The final reaction volume was 20µL. All samples were run in at least duplicates. The plate was then inserted into a CFX96 Touch Real-Time PCR System (Bio-Rad), and the qPCR run according to the cycling conditions shown in Table III. Melting curve cycles were included at the end to enable the identification of contaminants or primer dimer formation. After the qPCR representative samples from each duplicate were run on a gel to confirm the size of the amplification products and to check for control contamination.

TABLE III: Cycling conditions used for qPCRs			
Stage	Temperature (°C)	Time (mm:ss)	
Pre-incubation 1	50	02:00	
Pre-incubation 2	95	02:00	
Amplification 1	95	00:10	
Amplification 2	60	00:20	X60
Amplification 3	60	00:15	
Melting curve 1	95	00:05	
Melting curve 2	65	01:00	
Melting curve 3	97	Continuous	
Cooling	40	00:30	

Where required, new primers were designed to amplify a ~100bp region of the target gene, and where possible were positioned either side of an exon-exon junction. The melting temperature for the pair was ~60°C and included a GC clamp at the 3'-end to increase primer binding specificity. For each primer pair used the amplification efficiency was determined by running the standard qPCR cycles with neat cDNA, and cDNA diluted 1:10, 1:50, and 1:250. The primer efficiency was calculated using the equation E = (10^(-1/slope)-1)*100. The accepted efficiency range was 95-105%. The primers used are shown in Table IV.

TABLE IV: Table of the primers used for qPCR analysis			
Target	Designed by…	Direction	Sequence (5'-3')
Dazl	Heidi Mjøseng	Forward	TCT TTG CCA GAT ATG GCT CAG T
		Reverse	CTT CTG CAC ATC CAC GTC ATT A
Gata4	Heidi Mjøseng	Forward	GGA AGA CAC CCC AAT CTC GT
		Reverse	ACA TGG CCC CAC AAT TGA CA
Oct4	Heidi Mjøseng	Forward	GCC GTG AAG TTG GAG AAG GT
		Reverse	CTT CTG CTT CAG CAG CTT GG
Otx2	Heidi Mjøseng	Forward	ACT TGC CAG AAT CCA GGG TG
		Reverse	CTT CTT CTT GGC AGG CCT CA
Snai1	Heidi Mjøseng	Forward	GTC TGC ACG ACC TGT GGA AA
		Reverse	GGT CAG CAA AAG CAC GGT TG

2.4 DNA extraction and analysis

2.4.1 DNA extraction using phenol-chloroform

For large-scale DNA extractions (>1x10^7 cells per pellet) the phenol-chloroform method was used. The protocol is as follows: snap frozen cells pellets were partially thawed at room temperature, then resuspended with 1mL lysis buffer. Fresh lysis buffer was made as shown in Table V. 2µL RNase cocktail (Invitrogen) was added, and the samples were placed on a thermoshaker at 37°C for 1 hour. 30µL Proteinase K (ThermoFisher) was added and the samples were incubated overnight at 55°C on a thermoshaker. Before proceeding, the samples and the phenol:chloroform:isoamyl alcohol (P:C:I, ThermoFisher) were allowed to come to room temperature. Steps involving P:C:I and chloroform were always performed inside a fume hood. First, an equal volume of P:C:I was added to the sample, which was then vigorously shaken for 1 minute and centrifuged at 4,000 rpm for 5 minutes. After centrifugation a visible cloudy protein phase becomes visible at the interphase; the supernatant was transferred to a new tube with as little of the protein phase as possible. Another equal volume of the P:C:I was added to the supernatant and the

shake and centrifugation step was repeated. This was repeated a third time. To the final supernatant, an equal volume of chloroform (ThermoFisher) was added. Again, the sample was shaken, centrifuged, and fresh chloroform added to the extraction supernatant a total of three times. The final aqueous phase was moved to a fresh falcon tube and the final steps performed at the lab bench. 2x the sample volume was added of cold (-20°C) absolute ethanol to precipitate the DNA. After mixing, the sample was centrifuged at 4,000rpm for 5 minutes. The ethanol was aspirated, and the sample quickly resuspended in TE buffer (in-house technical services) to prevent over-drying. The DNA was allowed to resuspend for 1 hour at 60°C or overnight at 4°C on a rotator. The DNA extract was then analysed by Nanodrop and stored at 4°C, or -20°C for long-term storage.

TABLE V: Components of cell lysis buffer		
Reagent	Volume (mL)	Final concentration
1M Tris-HCl, pH8	5	100mM
0.5M EDTA	0.5	5mM
20% SDS	0.5	0.2%
5M NaCl	2	200mM
ddH_2O	42	
Total volume	50	

All reagents prepared in-house by technical services.

2.4.2 DNA digestion

For satellite probe/multi-locus DNA blots, 250ng DNA was treated with 1µL RNase cocktail (Invitrogen) for 1 hour at 37°C. The total volume per sample was made up to 25µL with TE buffer (made in-house) and 1X CutSmart buffer (NEB) to ensure glycerol from added reagents made up

<10% of the total volume as this could cause restriction enzyme star activity (non-specific digestion). 1µL *HpyCHIV* (*MaeII*, NEB) was added, and samples were incubated for 2 hours at 37°C. A sample of the digested DNA was checked by gel electrophoresis prior to use.

2.4.3 Gel electrophoresis

0.8-1.5% gels were made by dissolving the appropriate amount of UltraPure Agarose (Invitrogen) in 0.5X TBE (20X TBE, prepared in-house and diluted with MQ water). Once the solution had cooled enough to be warm but safe to touch 5µL/100mL ethidium bromide (VWR) was added (final concentration 0.005%) and mixed well. In a cold room, the warm gel poured into a gel tray sealed at the edges with tape. Any large bubbles were removed from around the combs and the gel was left to set. Once set, the gel was placed in a gel tank filled with 0.5% TBE (0.005% ethidium bromide), and the comb(s) carefully removed to avoid tearing the wells. OrangeG loading dye was made by mixing 3mL glycerol, 100µL 1M Tris (pH8), 6.9mL MQ water and OrangeG powder. GeneRuler DNA ladder mix (ThermoFisher) was loaded into the first well. Then, samples were mixed with OrangeG loading dye and loaded onto the gel. Generally, samples were run on the gel for 1.5 hours at 100V. The results were then viewed under UV light on a gel imaging system.

2.4.4 Southern blotting

For satellite probe/multi-locus DNA blots, 50-100ng DNA was loaded onto a 1-1.5% TBE gel, along with 2-5µL DIG-labelled ladder (DNA molecular weight marker II, Roche), and run at 120V for ~4 hours. The gel was denatured by submerging in denaturing solution (0.5M NaOH, 1.5M NaCl) for 2x 15 minutes, on a rocker at room temperature. The denatured gel was rinsed in ddH_2O and submerged in neutralising buffer (0.5M Tris, pH7.5, 1.5M NaCl) for 2x 15 minutes, on a rocker at room temperature. The gel was

then equilibrated for at least 10 minutes in 20X SSC prior to overnight transfer.

The Southern stack was assembled for overnight transfer as follows: 2x 3MM Whatman blotting paper (Fisher Scientific) wicks were wrapped around a plastic plate, which stood in a tray of 20X SSC. The wicks were soaked with more 20X SSC, and air bubbles were removed using a rolling pin. Another two gel-sized pieces of soaked blotting paper were placed on top, and bubbles removed. The gel was placed face down onto the wicks, and a piece of positively charged Hybond-N+ nylon transfer membrane (VWR) placed on top of the gel. The membrane was handled carefully from the corners with sterilised forceps. One of the corners of the membrane was cut to mark the orientation of the gel. On top of the membrane, a further two pieces of soaked blotting paper were added and rolled out. Cling film or lab parafilm was used to seal the sides of the gel and prevent short-circuiting. A thick pile of paper towels was placed on top of the stack, and on top of that two 500mL bottles of water were balanced as weights to help draw up the solution through the stack. The stack was left to transfer for at least 24 hours before disassembly, and the paper towels were replaced once half the pile was saturated.

After 24 hours the stack was disassembled, the positions of the wells were marked on the back of the membrane with pencil, and the DNA permanently bound to the membrane by UV crosslinking; The wet membrane was laid face-up on a damp piece of blotting paper soaked with 20X SSC, placed in a Stratalinker 1500 UV radiation box and cross-linked at 0.15J for 30s. The membrane was then rinsed in ddH_2O and either hybridised with probe immediately or placed in dry blotting paper, sealed, and stored at 4°C. At this point the gel may also have been post-stained with ethidium bromide to check transfer efficiency.

The DIG-High Prime DNA Labelling and Detection Starter Kit II (Roche) was used for Southern blot detection. From this kit, pre-hybridization buffer was prepared and pre-heated to the hybridization temperature (60°C).

The membrane was incubated in the warmed pre-hybridization buffer in a sealed plastic bag for 30 minutes on an incubated rotator (Enviro-Genie, Scientific Industries) set to 60°C. During this time the pre-made Southern probe targeting major satellite DNA (gift from Sproul lab, HGU, IGC) was thawed and denatured at 65°C for 30 minutes. The pre-hybridization buffer was poured off and the denatured probe buffer added to the membrane bag. The membrane was incubated with the probe overnight on the rotator at 60°C.

Following overnight incubation, the probe solution was collected and stored at -20°C until further use. The membrane was washed in stringency buffers (pre-heated to 50°C) to remove unbound probe. First, the membrane was washed for 2x 5 minutes in low stringency buffer (2X SSC, 0.1% SDS). The membrane was then washed in high stringency buffer (0.5X SSC, 0.1% SDS) for 2x 15 minutes at 65°C.

After the stringency washes, the membrane was rinsed with washing buffer (0.1M Maleic acid, 0.15M NaCl, pH7.5, 0.3% Tween-20). It was incubated for 30 minutes in blocking solution (1X solution prepared by diluting 10X Blocking solution powder from the Roche DIG kit in 1X Maleic acid buffer) on a rocker at room temperature. The membrane was incubated in antibody solution (Anti-digoxigenin-AP conjugate antibody diluted 1:10,000 in blocking solution) in a bag for 30 minutes. Excess antibody was removed by washing 2x 15 minutes in washing buffer on a rocker at room temperature. The membrane was equilibrated for 5 minutes in detection buffer (0.1M Tris-HCl, 0.1M NaCl, pH9.5). 1-2mL CSPD solution (Roche) was added and spread evenly across the membrane between two pieces of plastic, taking care to remove bubbles. Excess liquid was removed, and the membrane incubated at 37°C for 10 minutes prior to imaging. The chemiluminescent signal was detected using an ImageQuant system, or for more sensitive detection X-ray film was used.

2.5 RNA sequencing and analysis

2.5.1 RNA extraction

Whole transcriptome RNA sequencing was performed to compare gene expression profiles of EBs with and without *Dnmt1* at both early and late developmental time points. RNA was extracted from wild-type EBs or those treated with a low or high dose of dox at days 3 and 14 of differentiation. RNA was extracted using the RNEasy Mini Kit (QIAGEN) with QIAshredder columns (QIAGEN), following the manufacturer's instructions. A DNase1 (QIAGEN) step was included to degrade genomic DNA contaminants. The quality of the extracts were checked by gel electrophoresis, as described in section 2.3.3.

2.5.2 RNA quality assessment | Bioanalyzer

To qualify for RNA sequencing extracted RNA must have an RNA Integrity Number (RIN) of at least 9. The RIN scores of the extracted EB RNAs were determined using a 2100 Bioanalyzer (Agilent) with a Nano Chip by the HGU technical services. Three technical replicates per condition were assessed and submitted for sequencing.

2.5.3 RNA sequencing

Library preparation and RNA sequencing were completed by the Edinburgh Clinical Research Facility sequencing service. Triplicate samples were submitted for whole transcriptome sequencing with selection for polyadenylated transcripts. The libraries were prepared using the NEBNEXT Ultra II Directional RNA Library Prep Kit (NEB) with Poly-A mRNA magnetic isolation (NEB). Abundant ribosomal RNAs (rRNA) transcripts were removed during sample preparation. Following amplification, libraries were purified using AMPure XP beads (Beckman Coulter). The libraries were quantified using the Qubit dsDNA HS fluorometry assay (Invitrogen), and fragment quality determined by Bioanalyzer with the DNA HS kit (Agilent). Finally,

sequencing was performed using the NextSeq 500/550 High-Output v2.5 Kit using High Output v2.5 Flow Cells on the NextSeq 550 platform (Illumina).

2.5.4 RNA sequencing bioinformatics

Raw data produced from the RNA-seq experiment were processed and analysed by Dr Philippe Gautier of the HGU Bioinformatics Service. Briefly, the reads were assessed for quality using FASTQC and aligned to the mouse genome (mm10) using STAR. Alignment quality was checked using qualimap, and read counts were obtained using Salmon. All tools were used with the default parameters.

Differential gene expression was analysed in R (v3.6.2) using DeSeq2 (v1.26.0). Genes which had no reads in any sample were removed from the analyses. Log2 fold values were shrinked using the 'apeglm' method.

Figures were prepared in R using ggplot2 (version 3.3.0), and pheatmap (1.0.12). GO term analysis was done with clusterProfiler (version 3.14.3). The volcano plots were generated with EnhancedVolcano (version 1.4.0).

Differentially regulated genes considered for further analysis had an absolute log2 value >2, with an adjusted p-value of <0.05. We used a desktop version of GSEA (version 4.0.3) with the "GSEAPreranked" analysis. The ranked lists were obtained from the Deseq2 analysis, filtering for p-values <0.05 and sorted by Log2fold values. Because of the large size of the resulting lists, the maximum size of the gene set was adjusted to >3100.

2.5.5 Transposable element analysis

The transposable element analysis was kindly performed by Prof. Ian Adams. Briefly, the adaptors were trimmed from the RNA-seq reads using TrimGalore! 0.4.1 (paired end, Illumina, stringency 3). They were then aligned to the mouse genome (mm10) with TopHat 2.1. Very sensitive thresholds were set; inner distance was set to 54 ± 79, with no coverage

search, and max multi-hits set at 1. The TopHat output was intersected with the UCSC genome browser RepeatMasker track co-ordinates using BEDTools 2.25.0. The number of reads within each repeat element type was summed. Reads falling with repeats belonging to "LTR","SINE","LINE" classes were analysed using DESeq2 (v 3.12) to identify differentially expressed retrotransposons within the dataset. These were then plotted using R and GraphPad Prism (v9).

2.6 Protein extraction and analysis

2.6.1 Protein extraction

Pellets of roughly 500k-1x10^6 cells were washed with PBS, aspirated and snap frozen, then stored at -80°C until extraction. Proteins were extracted using freshly made Laemelli buffer, see Table VI.

TABLE VI: Composition of Laemelli buffer for protein extraction	
Laemelli Buffer	Volume (mL)
20% SDS	10
Glycerol	10
Tris	6
ddH_2O	24
	50mL Total

Cell pellets were briefly thawed on ice, then resuspended in 90µL Laemelli buffer and boiled for 5 minutes on a heat block, then returned immediately to ice. The chromatin was manually sheared using a 25mm, 25G needle (BD Microlance) until clumps of chromatin were no longer visible, the consistency was watery and the solution clear, which took approximately 10 minutes. The protein content of the samples was then determined using a Nanodrop on a 1:10 dilution of each sample. The samples were repeatedly measured and diluted until the Nanodrop produced 3 consistent readings. Once obtained, 30µg protein per sample were made up to 10µL with Laemelli buffer. To each sample 5uL 4X NuPage sample buffer (Invitrogen) was added, and the samples boiled for 10 minutes. They were then stored at -

80°C until ready for separation by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

2.6.2 SDS-PAGE

To separate the proteins prior to western blotting, the samples previously diluted and boiled in NuPage sample buffer were loaded onto a NuPAGE 4-12% Bis-Tris precast protein gel (Invitrogen) which was placed in a NuPAGE gel tank filled with 1X NuPAGE MES SDS Running buffer (Invitrogen) at 4°C. A reference protein ladder (Precision Plus Protein Kaleidoscope Pre-Stained Protein Standard, BioRad) was loaded at each end of the gel. The gels were run at 150V for ~1-1.5 hours.

2.6.3 Western blotting

The proteins were transferred from the gel to a membrane using the iBlot2 system (ThermoFisher) following the manufacturer's instructions. The protein transfer was then checked by staining with Ponceau S solution (Sigma) on a rocker for 10 minutes at room temperature.

Blocking buffer was prepared by dissolving 5g BSA powder in 100mL TBS-T (0.2% Tween-20) solution. The membrane was blocked in blocking buffer for 1 hour at room temperature on a rocker. Primary antibodies (see Table VII) were diluted in blocking buffer. The membrane was incubated in primary antibody solution in a falcon tube overnight at 4C on a rotator, then washed 3x 10 minutes with TBS-T on a rocker. Secondary HRP antibodies were prepared in TBS-T. The membranes were incubated with secondary antibodies for 1 hour at room temperature. For chemiluminescent signal detection, a 1:1 mix of SuperSignal West Pico PLUS Chemiluminescent Substrate (ThermoScientific) was prepared, and 2mL spread evenly across the membrane by placing between two pieces of plastic sheet. Excess solution was removed, and the chemiluminescent signal quantified using an ImageQuant LAS800.

Table VII: Antibodies used for western blotting			
Target protein	Cat. No.	Concentration	Host
Primary antibodies			
Oct4	Sc5279	1:200	Mouse
Sox2	Ab92494	1:1000	Rabbit
Dnmt1	Sc20701	1:500	Rabbit
H4	Ab31830	1:10,000	Mouse
Secondary antibodies			
Antibody against	Conjugation	Concentration	
Anti-mouse HRP	Ab6278	1:5000	
Anti-rabbit HRP	Ab6721	1:5000	

2.7 Generating fluorescence reporter cell lines

2.7.1 Bacterial transformation

Plasmids obtained from cloning experiments were first expanded by bacterial transformation. To do this, 2-3µL (concentration depending on the experiment) plasmid was added to a round-bottomed falcon tube (STEMCELL Technologies) kept on ice. Working in a cold room and using pre-chilled pipette-tips, library competent DH5α cells (Life Technologies) were thawed slowly on ice, 50µL added per falcon tube and gently mixed with the pDNA. The mix was incubated on ice for 30 minutes. The cells were then heat-shocked in a 42°C water bath for exactly 45 seconds. The tubes were immediately placed back on ice for 2 minutes. At room temperature, 950µL SOC media (Life Technologies) was added slowly and the mix incubated for 1 hour at 37°C with shaking at 250rpm. During this time selection media was spread on the culture plates if needed, which were then warmed to 37°C. After one hour, 100µL of the transformation mix was spread onto the warmed culture plates under aseptic conditions. Any remaining mix

was stored at 4°C for up to one week. The plates were incubated upside down in a 37°C incubator overnight for a maximum of 16 hours. After this time the plates were checked for colonies, and if successful they were stored at 4°C until the colonies could be picked to L-broth for further expansion overnight.

2.7.2 Gateway cloning into a *piggyBac* vector

2.7.2.1 Obtaining the Lifeact-mScarlet construct

To visualise F-actin in differentiating EBs, the pLifeact-mScarlet_N1 vector (Bindels *et al.*, 2016) was obtained from AddGene (Addgene plasmid #85054, deposited by D. Gadella) and integrated into *Dnmt1*$^{tet/tet}$ cells by *piggyBac* transposition. To generate a stably expressing cell line, the reporter was first integrated into a *piggyBac* transposon vector by Gateway cloning with Clonase II (Invitrogen). The *piggyBac* cloning strategy was used as it enables random integration of target DNA by a cut-and-paste mechanism with high integration efficiency, and without the use of viruses used in retroviral transposon screens are expensive and require more complex safety measures (Woodard and Wilson, 2015). Also, the *piggyBac* system is less prone to overproduction inhibition, wherein the concentration of transposase prevents efficient integration, a phenomenon commonly observed with the *Sleeping Beauty* system.

2.7.2.2 Gateway cloning

The Gateway cloning strategy is a site-specific recombination system adapted from the lambda bacteriophage which uses it to integrate itself into the *E. coli* chromosome (Ptashne, 1992). The recombination is mediated by Gateway Clonase II recombination enzymes which recognise attachment (*att*) recombination sites in DNA. Upon recognition of these sites the DNA segments flanking them are switched. The first recombination step during Gateway cloning is the BP reaction; the linear donor sequence is first amplified by PCR using primers which introduce *att*B sites at either end to

produce an *att*B-PCR product. This is then used in the BP reaction with a donor vector containing *att*P sites, such as pDONR221. BP Clonase II recognizes the *att* sites and switches the contents, producing an entry clone with the PCR product flanked by *att*L sites, and by-product linear DNA flanked by *att*R sites. The second recombination occurs during the LR reaction. Here, the *att*L product in the entry clone is recombined with an *att*R-flanked substrate found in the destination vector, again by BP Clonase II enzyme, to create the final expression clone which will be transfected into ES cells.

Gateway entry, donor and destination vectors contain *ccd*B genes between the *att* sites to enable negative selection of non-recombined clones following recombination and transformation. CcdB protein interferes with *E. coli* DNA gyrase function, inhibiting cell growth (Bernard and Couturier, 1992), so any cells which do not lose the gene by recombination are eliminated from the population following bacterial transformation, which improves cloning efficiency.

The Gateway cloning primers were designed by research assistant Dr David Hay. The primers were designed to make the target sequence compatible with Gateway cloning by introducing overhangs at each end starting with four guanine residues followed by the 25bp *att*B1/B2 site sequence, followed by the first 18-25bp of the target sequence. The primers used are shown in Table VIII.

TABLE VIII: Primers used to generate and screen the Lifeact-mScarlet *piggyBac* clone vectors		
Gateway Reaction	Primer name	Sequence (5'-3')
attB	LifeAct.FOR	GGG GAC AAG TTT GTA CAA AAA AGC AGG CTT AAT GGG CGT GGC CGA C
attB	LifeAct.REV	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTT TTA CTT GTA CAG CTC GTC CAT GCC
Lifeact seq	CMV-F	CGC AAA TGG GCG GTA GGC GTG
Lifeact seq	mCherry-R	TTG GTC ACC TTC AGC TTG G
Expression vector	BGH_R	TAG AAG GCA CAG TCG AGG
Expression vector	PB_CAG	GGC AAA GAA TTC CTC GAC GG

The 25bp *att* sites are highlighted in bold.

The pLifeact-mScarlet _N1 vector was received in bacteria as an agar stab, which were expanded by streak plating on Kanamycin plates (prepared in-house by technical services) overnight at 37°C. The following morning, the plate was moved to 4°C until the early evening when colonies were picked into 5mL aliquots of L-broth with 50µg/mL Kanamycin and incubated overnight for a maximum of 16 hours at 37°C with shaking at 250rpm. The following day, 0.5mL of the broth was used to prepare a glycerol stock by mixing with 0.5mL 50% glycerol (Fisher Scientific; diluted 1:1 with ddH_2O), snap freezing on dry ice and storing at -80°C. The plasmid was extracted from the remaining broth using the DNeasy Blood & Tissue kit (QIAGEN), following the manufacturer's instructions. Extracted plasmid was analysed by NanoDrop, and the Lifeact sequence verified by PCR with mCherry-R and CMV-F 10µM primers (See Table VIII).

2.7.2.3 Gateway cloning, Part 1: BP recombination

For integration into the *piggyBac* Entry vector, the Lifeact-mScarlet sequence was isolated from the plasmid and modified to include *att*B sites by PCR. Per reaction, 12.5µL CloneAmp HiFi PCR Premix (Takara Bio) was mixed with 100ng Lifeact-mScarlet pDNA and 0.5µL each of LifeAct.FOR and LifeAct.REV 10µM primers (Table VIII), in a final volume of 25µL made up with MQ water. The PCR cycle conditions were as follows: 94°C for 30s, 50°C for 30s, 72°C for 20s, for a total of 35 cycles. The *att*B-Lifeact-mScarlet PCR product size (820bp) was checked by gel electrophoresis on a 1.5% TBE gel. The PCR product was purified by column extraction using the DNeasy Blood & Tissue kit (QIAGEN), following the manufacturer's instructions.

The BP reaction to generate the ENTRY clone was performed by combining equimolar concentrations of pDONR221 vector and *att*B-Lifeact-mScarlet PCR product on ice in TE buffer (pH8), then adding 2µL Gateway BP Clonase II enzyme mix (Invitrogen; final reaction volume 10µL). pEXP7-tet vector (Invitrogen) was used as a positive control, and the clonase was excluded from one sample as a negative control. The mixes were incubated at room temperature overnight. The resulting ENTRY vector and control reactions were treated with 1µL Proteinase K (Invitrogen) for 30 minutes at 37°C, then transformed into library competent DH5α cells (Life Technologies; See section 2.3.2 for the detailed transformation procedure) and incubated on agar plates with Kanamycin overnight at 37°C. The *att*B-Lifeact-mScarlet-transformed colonies were expanded in L-broth overnight with Kanamycin selection, and the ENTRY vector extracted and purified using the DNeasy Blood & Tissue kit (QIAGEN), following the manufacturer's instructions.

2.7.2.4 Gateway cloning, Part 2: LR recombination

The Lifeact-mScarlet sequence was transferred from the purified ENTRY vector into a pPB-PGK destination vector to create the final expression vector by LR recombination with Gateway LR Clonase II

(Invitrogen). 150ng LifeAct-mScarlet-ENTRY vector was mixed with 150ng destination vector and 2µL Gateway LR Clonase II (final reaction volume to 10µL with TE buffer, pH8) and incubated overnight at room temperature. A positive control reaction was run in parallel using pENTR-gus entry clone (provided with the clonase II kit) with the destination vector. A negative control reaction was included by excluding clonase II. The following day, the reactions were treated with 1µL Proteinase K (2µg/mL, Invitrogen) for 10 minutes at 37°C. The LR reaction products were transformed into library competent DH5α cells (Life Technologies) and grown overnight on agar plates with ampicillin (prepared in-house by technical services). A transformation control plate was included by transforming DH5α cells with pUC19. The following day, four colonies were picked from the expression vector plate and expanded in L-broth with ampicillin (50µg/mL) overnight at 37°C with shaking at 250rpm. The expression vectors were extracted and purified from the overgrowths using the DNeasy Blood & Tissue kit (QIAGEN), following the manufacturer's instructions. A sample from each extraction was submitted for sequencing at the in-house technical services sequencing facility with BGH_R and PB_CAG 10µM primers (See Table VIII) to confirm the expression clone construct. Sequencing confirmed the integration of the Lifeact-mScarlet sequence in the *piggyBac* vector in all four expression clones; one clone was used for the *Dnmt1*$^{tet/tet}$ transfection.

2.7.2.5 Transfecting PB-Lifeact-mScarlet into *Dnmt1*$^{tet/tet}$ ESCs

Following sequencing validation of the Lifeact-mScarlet *piggyBac* vector, the plasmid was transfected into *Dnmt1*$^{tet/tet}$ mESCs by reverse transfection to generate a stable cell line. Briefly, 600ng Lifeact-mScarlet *piggyBac* vector was mixed with 600ng *piggyBac* transposase expression vector in a total volume of 100µL with Optimem (Gibco). To this, 4.5µL Attractene reverse transfection reagent (QIAGEN) was added. The mix was incubated at room temperature for at least 15 minutes prior to adding to the cell suspension, up to a maximum of 30 minutes. Dissociated cells were seeded at a density of $3x10^5$ per 2mL serum/LIF media per well in a

gelatinized 6-well plate (Corning), and to this the reverse transfection mix was added. The media was refreshed after 24 hours to remove the transfection mix. At this time, approximately 10% of the transfected ESCs had visible RFP signal. 48 hours post-transfection, the cells were screened for mScarlet expression by fluorescence associated cell sorting (FACS).

2.7.2.6 FACS analysis of reporter expression

To analyse fluorescence reporter signal in Lifeact-mScarlet transfected ESCs, cells were detached, dissociated, and resuspended in 500µL PBS (5% FCS) in FACS tubes, then analysed on a FACS JAZZ cell sorter (BD Biosciences) by a member of the FACS facility team (Lizzie Freyer/Robbie Pineda). Gating was applied to identify the target cells for sorting. First, single cells were identified by plotting forward scatter height against forward scatter area, then live cells were distinguished from debris by plotting forwards scatter vs side scatter. Finally, target RFP+ cells were identified by sorting for signal between 561-582nm. Non-transfected cells were used to control for background autofluorescence. Cells with no / low / medium / high RFP were collected (24 cells each) and single cells were seeded in gelatinized 96-well plates in serum/LIF. The wells were checked for the presence of colonies after 6 days; 41/96 wells had colonies (43% survival rate). Some cells which had been RFP+ at the time of sorting were no longer expressing Lifeact-mScarlet. To ensure the picked RFP+ clones were stably expressing Lifeact-mScarlet, RFP+ colonies and RFP- control colonies were picked to 6-well plates and subjected to 200µg/mL hygromycin selection (the *piggyBac* destination vector contained a hygromycin resistance gene) until 48 hours after the RFP- and wild-type *Dnmt1*$^{tet/tet}$ ESCs had died (5 days total). After 3 more passages without selection, the remaining *Dnmt1*$^{tet/tet}$-Lifeact RFP+ clones were collected, and stocks stored in liquid nitrogen.

2.8 Generating knockout cell lines with CRISPR-Cas9 genome editing

2.8.1 Designing CRISPR guides

CRISPR guides targeting regions either side of *Dnmt1* exon 1 and 38 were designed by research assistant Dr David Hay using web tool Benchling. In total four ~20bp guide sequences were identified, one for the C-terminus region (guide 3), and 3 for the N-terminal region (guides 1, 4 and 5). The positions of the guides are shown in figures 2.1 A-C. Guides were checked for potential off-target mismatches using web tool CHOPCHOP (v3) (Montague *et al.*, 2014; Labun *et al.*, 2019). For each guide sequence identified, the forward sequence and its reverse complement sequence were modified to add overhangs to enable integration of the self-annealed sgRNA oligos into a CRISPR vector (Table IX). The finalised custom guide DNA oligos were ordered from SIGMA.

TABLE IX: Sequences of sgRNAs used for CRISPR/Cas9 deletion of *Dnmt1*				
Target	Name	Primer direction	Sequence (5'-3')	CRISPR vector
Exon 1	Guide 1	Forward	CA CCG GAC GAG CCC ACT ATA GCC AGG	PX458
Exon 1	Guide 1	Reverse	C CCT GGC TAT AGT GGG CTC GTC CAA A	PX458
Exon 1	Guide 4	Forward	CA CCG CGG ACG AGC CCA CTA TAG CC	PX458
Exon 1	Guide 4	Reverse	C GGC TAT AGT GGG CTC GTC CGC AAA	PX458
Exon 1	Guide 5	Forward	CA CCG AGC CAC CTC TTC AGG CAA GG	PX458
Exon 1	Guide 5	Reverse	C CCT TGC CTG AAG AGG TGG CTC AAA	PX458
Exon 38	Guide 3	Forward	CA CCG TCA GCA GGC ACC GAT CTC AA	PX459
Exon 38	Guide 3	Reverse	C TTG AGA TCG GTG CCT GCT GAC AAA	PX459

Overhangs for cloning into backbone vector are shown in bold.

The oligos were prepared for cloning into CRISPR backbone vectors following the protocol from Zhang lab (available from: zhang-lab-general-cloning-protocol-target-sequencing_1.pdf (addgene.org)). Guide 3 was cloned into PX459 (pSpCas9(BB)-2A-Puro v2.0, AddGene plasmid #62988, deposited by Feng Zhang) (Ran *et al.*, 2013), which contains a puromycin resistance gene for transformation screening. The other three C-terminus guides were cloned into PX458 (pSpCas9(BB)-2A-GFP, AddGene #48138, deposited by Feng Zhang) (Ran *et al.*, 2013). PX458 contains GFP for screening selection. The backbone CRISPR vectors were linearised by digesting 1µg of plasmid with 1µL FastDigest *BbsI* (Fermentas), 1µL FastAP

(Fermentas), and 2µL 10X FastDigest Buffer (Fermentas) in a total volume of 20µL with MQ water. The mix was incubated at 37°C for 30 minutes, then the digested plasmid extracted and purified using the QIAquick Gel Extraction Kit (QIAGEN), following the manufacturer's instructions. The 4 oligo pairs were annealed by incubating 100µM each forward and reverse primer in CutSmart Buffer (NEB), total volume 10µL, for 30 minutes at 37°C on a thermoshaker set to 700rpm. Then, they were placed in a heat block set to 100°C and the block allowed to cool to room temperature on the lab bench. 1:200 dilutions of each annealed oligo pair were made with MQ water. 1µL each diluted oligo was ligated into 50ng linearised backbone CRISPR vector (see Table IX) using the Quick Ligation Kit (NEB) by incubating the mix at room temperature for 4 hours.

Each PX458/PX459 guide ligation product was transformed into library competent DH5α cells (Life Technologies; see section 2.6.1 for transformation protocol), which were then spread on ampicillin plates. Colonies were picked and expanded overnight, and plasmids extracted using the NucleoSpin Mini kit (Machery-Nagel), following the manufacturer's instructions. Glycerol stocks were prepared from excess overgrowth. The gRNA section of the ligated vector was validated by sequencing with the hU6-F primer (5' - GAG GGC CTA TTT CCC ATG ATT – 3').

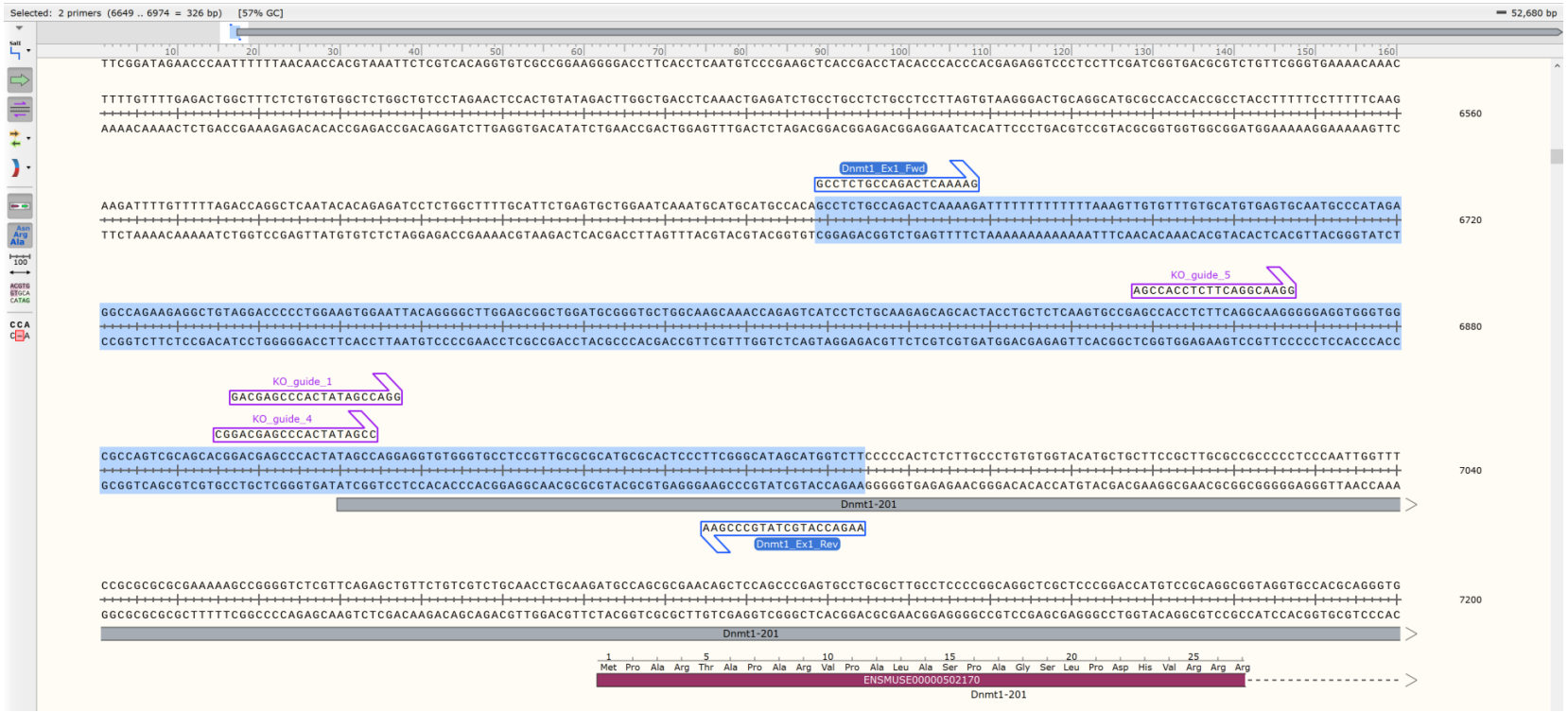


Figure 2.1 A. Positions of sgRNA oligos and PCR primers used during *Dnmt1* CRISPR KO and clone screening. Part 'a' shows the exon 1 region, and the blue highlights the area targeted in the exon 1 micro-deletion screen PCR. A 326bp band with the highlighted primers would indicate that the CRISPR KO was unsuccessful.

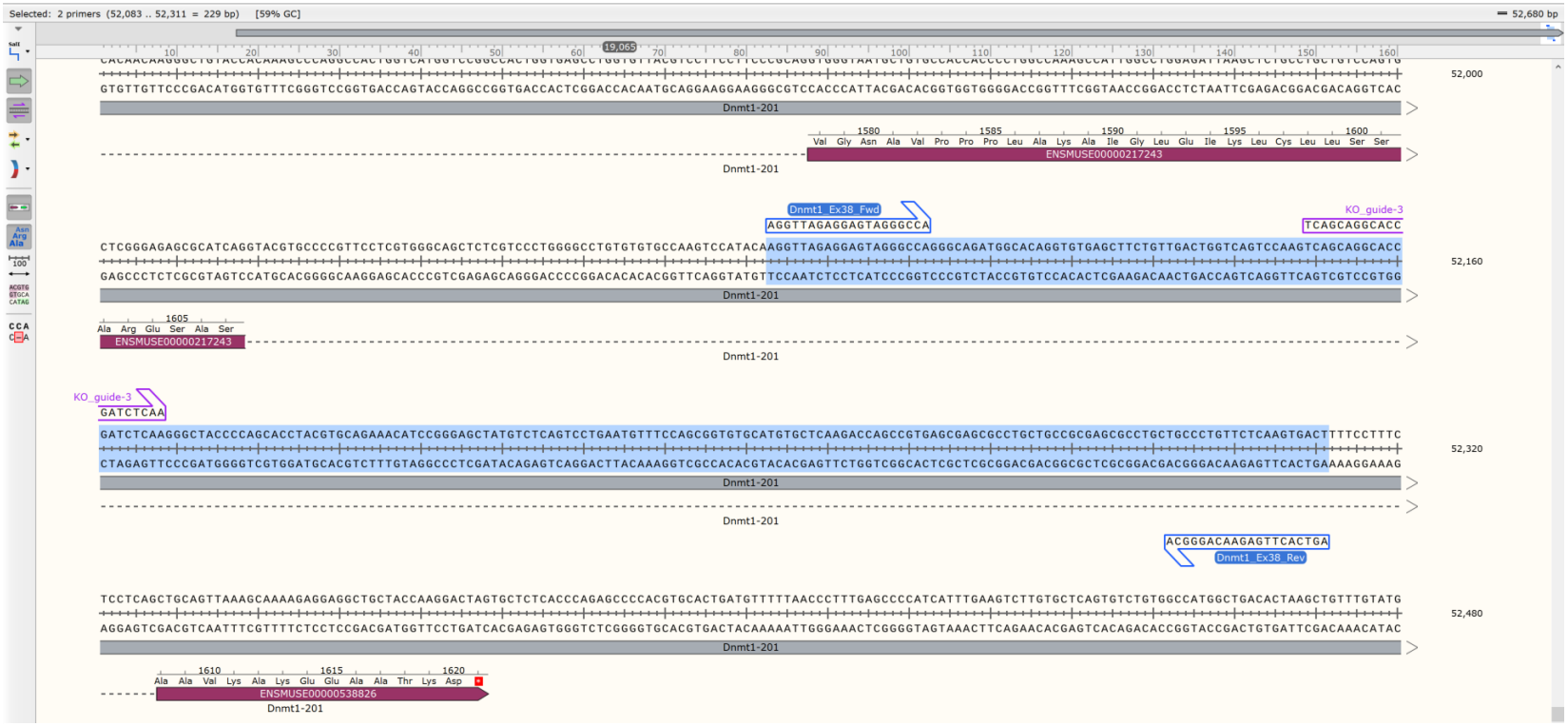


Figure 2.1 B. Positions of sgRNA oligos and PCR primers used during *Dnmt1* CRISPR KO and clone screening. Part ‘b’ shows the exon 38 region, and the blue highlights the area targeted in the exon 38 micro-deletion screen PCR. A 229bp band with the highlighted primers would indicate that the CRISPR KO was unsuccessful.

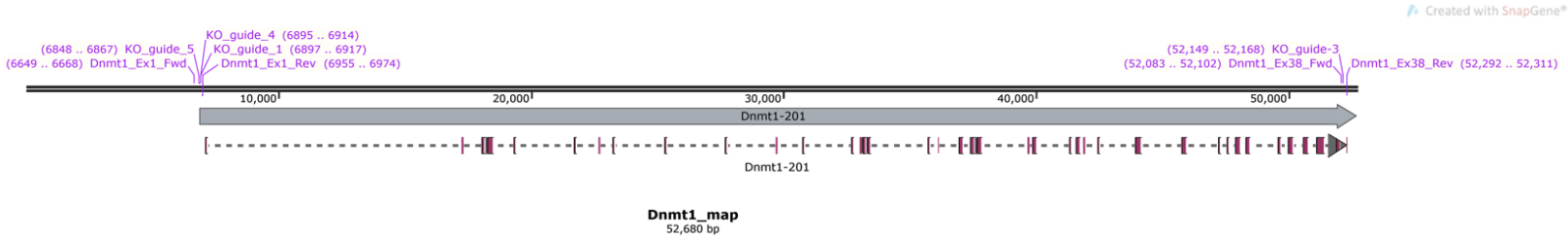


Figure 2.1 C. Positions of sgRNA oligos and PCR primers used during *Dnmt1* CRISPR KO and clone screening. Part 'c' gives an overview of the *Dnmt1* gene and positions of KO guides and primers used. For the macro-deletion screen, Dnmt1_Ex1_Fwd and Dnmt1_Ex38_Rev primers were used and a 432bp band indicated the successful deletion of *Dnmt1*.

2.8.2 Testing the sgRNAs and CRISPR knock out PCR screening strategy

To test the efficiencies of the guides and determine the optimum pair for *Dnmt1* KO, 1.2µg total plasmid DNA of exon 1- and exon 38-targeting guide vectors (600ng each; guide 3 + guide 1/4/5) were reverse transfected into wild-type E14 mESCs with Attractene (QIAGEN; see Section 2.6.2.5 for reverse transfection method). The next day, the cells were subjected to 2µg/mL puromycin selection for 24 hours. Following selection and ES cell expansion, the DNA was extracted from the mixed population of transfected E14 cells using the DNeasy Blood and Tissue kit (QIAGEN). The mixed population DNA was analysed by PCR using primers designed to detect a macro-deletion of the whole gene, or micro-deletions testing for the presence of either end exon of *Dnmt1* (Primer sequences shown in Table X; primer positions and micro-deletion target regions shown in Fig 2.1 A-C).

TABLE X: Details of the PCR primers used to screen for *Dnmt1* KO

Primer name	Deletion type	Sequence (5'-3')
Dnmt1_Ex1_Fwd	Macro & micro	GCC TCT GCC AGA CTC AAA AG
Dnmt1_Ex1_Rev	Micro	AAG ACC ATG CTA TGC CCG AA
Dnmt1_Ex38_Fwd	Micro	AGG TTA GAG GAG TAG GGC CA
Dnmt1_Ex38_Rev	Macro & micro	AGT CAC TTG AGA ACA GGG CA

Representative test PCRs are shown in figures 2.2 and 2.3. The transfected cell mixed population DNA was from E14s transfected with guides 1 and 3 (See figure 2.1 C). For this PCR, 100ng DNA was mixed with 12.5µL Platinum Green Hot Start PCR Master Mix (2X; Invitrogen) and 0.5µL each primer (10µM), in a final volume of 25µL with MQ water. The PCR cycling conditions are shown in Table XI. The PCR products were analysed by gel electrophoresis on a 1.5% TBE gel with ethidium bromide.

TABLE XI: The PCR cycling conditions used to screen for *Dnmt1* KO		
Number of cycles	Temperature (°C)	Time (mm:ss)
	94	02:00
	94	00:30
X35	55	00:30
	72	00:30
	72	05:00
	10	Infinity

The PCR test showed that the micro-deletion primers can produce specific bands of the expected size with wild-type DNA, which we would expect to be absent when *Dnmt1* is knocked out and the primer site within the exon sequence is removed. The macro-deletion PCR was unable to produce a band in the wild-type DNA due to the size of the locus being too large for the polymerase to amplify across, as expected. In the transfected population however, a band was produced suggesting that in some cells the *Dnmt1* had been successfully removed and thus the primer sites, which are external to each end of the gene, had been brought close enough together to be able to produce an amplicon of the predicted size. Following this initial test with E14 DNAs, these PCR conditions were used to screen for *Dnmt1* KO in further cell lines. As the first combination of CRISPR guides was able to produce a clear band in the macro deletion PCR, and as time was limited, the remaining guide RNAs were not tested.

PCR products were submitted for sequencing to confirm the removal of Dnmt1 in the clones. A representative sequence alignment view is shown in Figure 2.4.

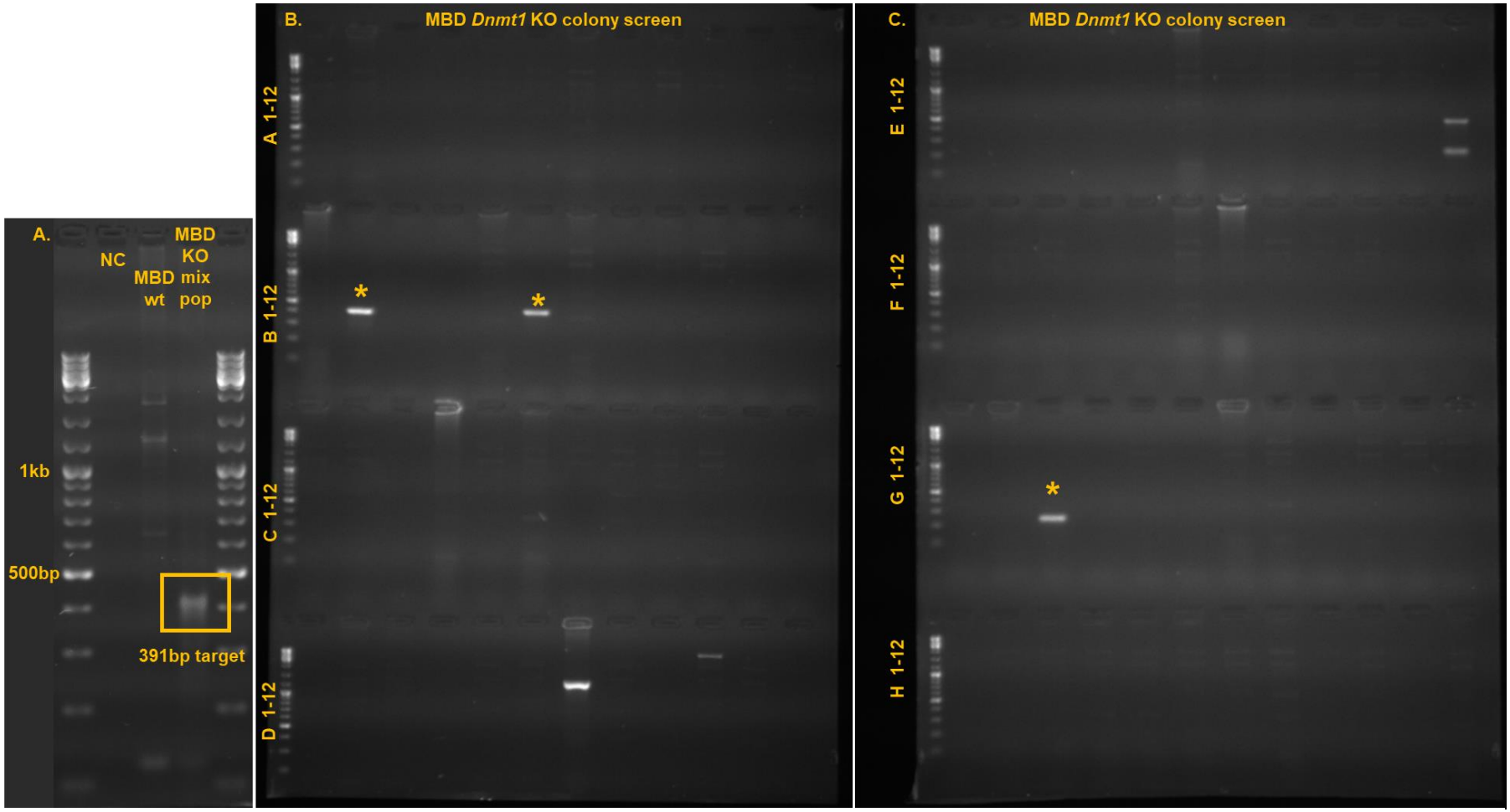


Figure 2.2. Legend on next page…

Figure 2.2 (previous page). A macro-deletion PCR screen of EGFP-MBD-nls cells transfected with *Dnmt1* KO guides identified 3 potential clones. Gel A shows the PCR controls; NC = No template control; MBD wt = DNA from untransfected EGFP-MBD-nls cells; MBD KO mix pop = DNA from the mixed transfection population which was used as a positive control. Gel B and C show the PCR products from single picked transfection colonies. Clones which were identified as potential homozygous KO clones are highlighted by asterisks.

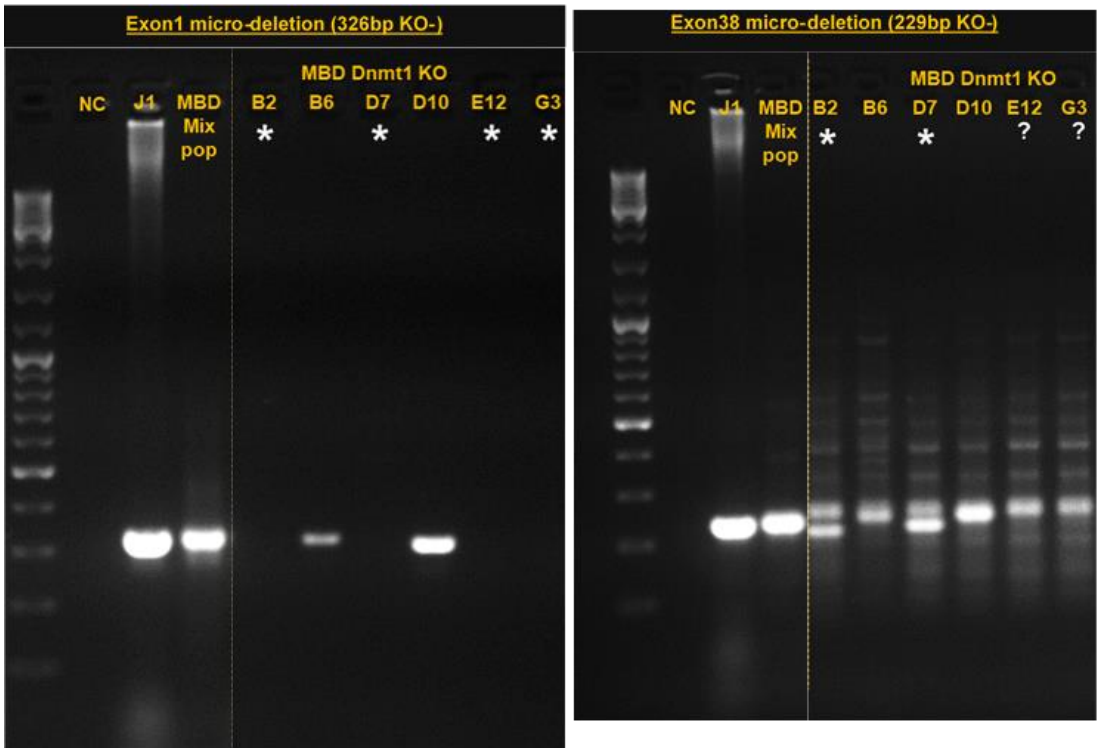


Figure 2.3. Micro deletion PCRs targeting exons 1 and 38 of *Dnmt1* identified two clones for further screening. The micro deletion PCRs were designed to detect *Dnmt1* KO clones by loss of the amplicon as one primer site per pair is within an exon. NC = No template control; J1 = wild type ESC control which should produce an amplicon; MBD mix pop = positive control for the PCR conditions. Potential KOs are indicated by asterisks.

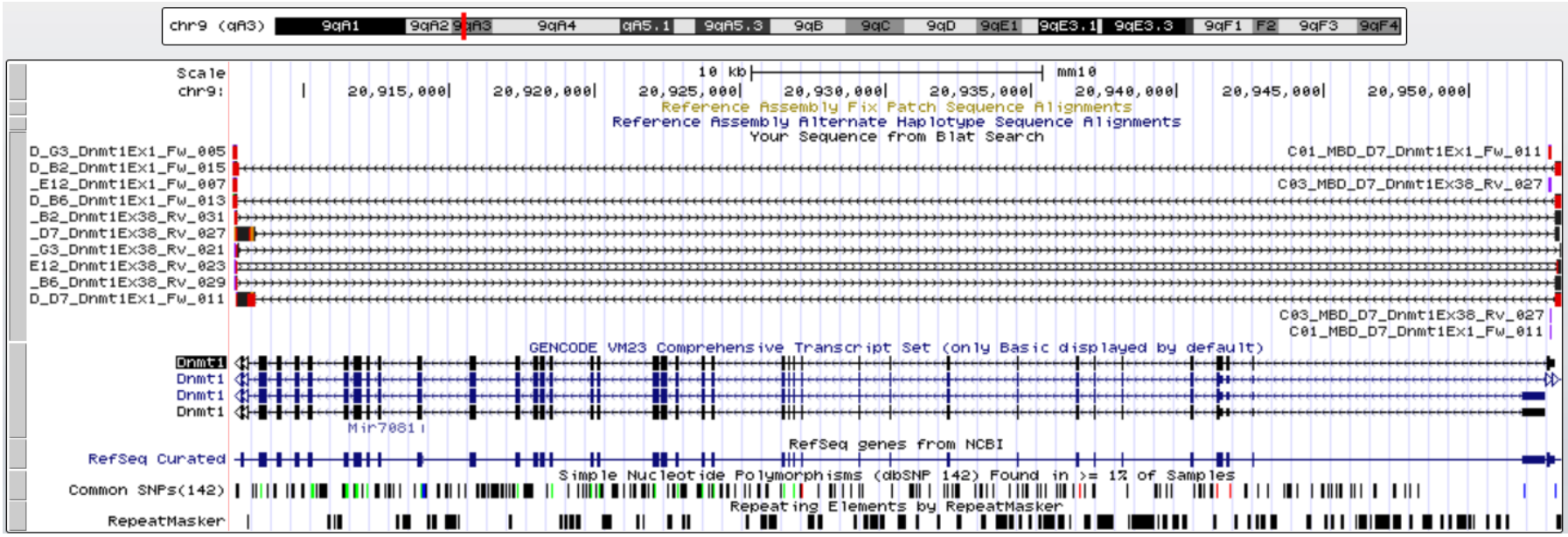


Figure 2.4. Sequencing the PCR products from the EGFP-MBD-nls *Dnmt1* KO macro-deletion screen and analysing by BLAST alignment confirmed the removal of *Dnmt1*. The products were sequenced using the PCR forward and reverse primers, and the resulting sequence data were aligned to the mouse genome (grcm38/mm10) using the UCSC genome browser.

2.9 Flow cytometry

2.9.1 Ssea1 assay for pluripotency

To validate the doxycycline inhibition of *Oct4* in *Oct4*$^{tet/tet}$ ESCs, I performed an Ssea1 assay. This assay quantifies the presence of Ssea1 cell surface proteins which are only expressed on mouse pluripotent stem cells (Furusawa *et al.*, 2006). The mark is lost as cells differentiate.

Ssea1 solution was prepared by diluting 100µL Ssea1 reagent (eBioscience) in 1.9mL PBS (5% BSA). To perform the assay, samples of live cells were washed 1x with PBS. They were collected by enzymatic dissociation with Accutase (Gibco) and counted. 500k cells per sample were placed into a new tube, spun down, and resuspended in 100µL cold Ssea1 solution. The samples were incubated at room temperature for 10 minutes prior to flow cytometry analysis. Control cells were resuspended in cold PBS/BSA without Ssea1 reagent. DAPI was added as a counterstain, and the samples were made up to ~300µL with PBS. Ssea1 signal was analysed on an LSR Fortessa flow cytometer. Ssea1 signal was measured at 488-520/550nm, DAPI (dead cells) at 408-450/550nm, and background at 561-586/15nm.

2.9.2 Reporter for 2C-like cells

A plasmid construct was obtained encoding an EGFP reporter of 2C-like cells (MERVL:EGFP) (Ishiuchi *et al.*, 2015) from AddGene (plasmid #69071). Dr David Hay cloned the MERVL:EGFP construct into a pMVP/PB/mCherry-DEST PiggyBac cloning vector obtained from AddGene (#121871). This removed the *neo* selection marker from the MERVL:EGFP reporter. He then transfected the final construct into wild-type J1 and *Dnmt1*$^{tet/tet}$ ESCs. The resulting cell lines constitutively express SV40:mCherry, and the 2C reporter is expressed in MERVL+ ES cells, marking those transitioning through a 2C-like state. The 2C::EGFP signal was then analysed in *Dnmt1*$^{tet/tet}$ MERVL:EGFP cells cultured with and without dox by Chetan Srinath by microscopy and flow cytometry analysis.

2.10 Imaging and imaging analysis

2.10.1 Live imaging

Live EBs were imaged using a microscope equipped with an incubator chamber to enable control of the sample temperature and humidity during imaging. The chamber was set to 37°C, 5% CO_2.

The LifeAct reporter EBs were imaged on a spinning disk confocal microscope at 20X. Beating EBs were imaged by capturing an image every 50ms x100, and these were converted to a video format (.AVI) in FIJI (Schindelin *et al.*, 2012).

2.10.2 Immunocytochemistry

2.10.2.1 Fixing cells

For imaging, EBs were taken from suspension culture and seeded into either glass-bottomed 24-well plates (maximum of 4 EBs per well) or 10-well glass bottom CELLview culture slides (Greiner; one EB per well). The plates were previously coated with 0.2% gelatin at least 24 hours prior to seeding and were pre-warmed at 37°C in an incubator. To fix the cells, they were washed once with PBS, very gently, and then incubated in 4% PFA for 10 minutes at room temperature. The wells were then washed three times with PBS and stored in PBS at 4°C until imaging. Cells were stored for a maximum of 2 weeks.

To improve the resolution of imaging of nuclear proteins, such as Oct4 and Gata4, EBs were treated with cytoskeleton (CSK) buffer (Dunican *et al.*, 2013) prior to fixation. This buffer enhances the accessibility of the chromatin for improved antigen binding. To make the buffer, 7mL 5M NaOH was added to 15mL ddH_2O in a conical flask on a heated stirrer plate with a magnetic flea. 15.11g PIPES powder was added slowly, then the pH was adjusted to 6.8 using 5M NaOH solution. The final volume was brought to 50mL with ddH_2O. Once cooled, the solution was filtered through a 0.2µm pore filter and stored at room temperature.

An aliquot of buffer was chilled to 4°C then kept on ice prior to EB staining. After a PBS wash, the EBs were placed on ice and covered with CSK buffer for exactly 1 minute. The EBs were then washed 1x with PBS and fixed with PFA as per the protocol.

2.10.2.2 Staining cells

Fixed cells were washed once with PBS, then incubated in 10% donkey serum (Sigma) solution for 1 hour at room temperature to block. Primary antibodies (see Table XII) were diluted in 1% donkey serum, 0.1% TritonX. EBs were then incubated with the primary antibody overnight at 4°C. They were washed 3x with PBS, 10 minutes per wash at room temperature. Secondary antibody was diluted 1:200 in 1% donkey serum solution and kept protected from light. EBs were incubated in the secondary antibody for 1 hour at room temperature wrapped in tin foil. The EBs were washed 3x for 10 minutes at room temperature. DAPI counterstain was prepared by diluting 50mg/mL stock solution 1:1000 in PBS. EBs were incubated with DAPI for 10 minutes at room temperature, then washed 2x with PBS, and once with ddH_2O. Finally, the EBs were sealed using VectaShield (Vector Labs), and then were stored in the dark at 4°C.

For cells cultured in the CELLview plates, one antibody scheme was used per plate. The black plastic culture chamber and glue seal were separated from the slide prior to staining. Once stained, the samples were sealed with VectaShield and then a rectangular glass coverslip was placed on top, and the edges sealed with clear nail varnish. For cells cultured in 24-well plates, one drop of VectaShield was added to the centre of a round glass coverslip, which was carefully placed into each well.

TABLE XII: Antibodies used for immunocytochemistry			
Target protein	Cat. No.	Concentration	Host
Primary antibodies			
Oct4	Sc5279	1:500	Mouse
Ncam1	Ab220360	1:1000	Rabbit
Gata4	Sc25310	1:500	Mouse
Secondary antibodies			
Antibody against	Conjugation	Concentration	Host
Anti-mouse	AF488	1:200	Donkey
Anti-rabbit	AF647	1:200	Donkey

2.10.2.3 Fixed cell imaging

Fixed, stained EBs were imaged at 20X using epifluorescence microscopes. To image entire EBs the microscope was set to scan the sample in overlapping tiles, which were then stitched to make the single final automatically at the end of acquisition.

2.10.3 Image analysis

Thresholding in FIJI to measure EB area

The thresholding feature in FIJI image processing software (v1.53f51) (Schindelin *et al.*, 2012) was used to measure EB size (Fig. 2.5). The analysis pipeline used depended on the fluorescence reporter or stain being used to highlight the EB cells.

EBs expressing the Lifeact-mScarlet actin reporter were imaged at days 7, 10 and 14 on a Dragonfly confocal microscope (Brand) at 20X magnification. Individual EBs were captured as tiled Imaris dataset images (.ims). The Fusion Stitcher file was then opened in FIJI using the Bio-Formats importer to stitch and auto-scale the tiled images. Once opened, the brightness and contrast of the image was adjusted for clarity, a scale bar

added, and the resulting image saved as a JPEG and TIFF. The area of the EB was measured by deleting the scale bar and then using the Threshold tool to highlight EB cells and exclude the background. The red highlighted area was then measured to obtain the area, as well as mean, minimum and maximum gray values, perimeter length in pixels, and the minimum and maximum threshold values. Area was measured either in pixels^2 or μm^2, depending on if the pixels/μm for the microscope was known. FIJI automatically measures the area in pixels^2 unless the image scale has been calibrated using ‘Set scale…’, then area is reported in μm^2.

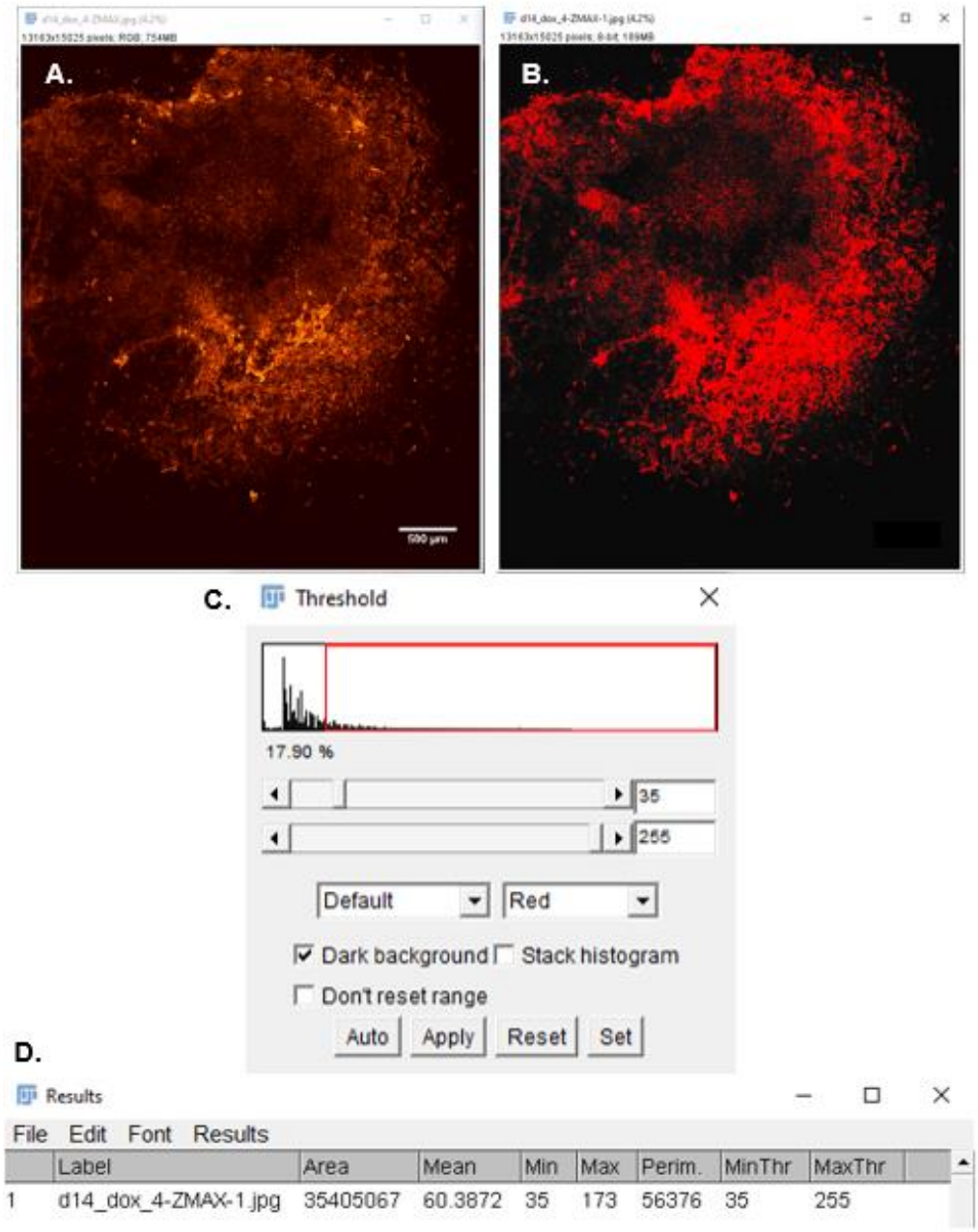


Figure 2.5. EB area was measured using the Threshold tool in FIJI. To highlight the EB cells (A), the bottom slider (sets maximum threshold value) in the Threshold tool was moved to the far right and the top slider (sets minimum threshold value) was adjusted to the point where most EB cells had been highlighted without introducing background noise (B). The red area was then measured to obtain the EB area (D).

Chapter 3 | Investigating the role of DNA methylation during differentiation using an embryoid body model

3.1 Introduction

The mammalian body is composed of a diverse range of cell types which function together to keep the organism alive. All these cell types contain identical copies of the organism's genetic code encoded in DNA. The DNA sequence itself contains sections called genes, stretches of DNA which encode instructions for specific products such as proteins and long non-coding RNAs (lncRNAs). Which combination of these genes are 'read' by the cell's transcriptional machinery determines which cell type is produced and maintained, enabling the formation of a large variety of cells from one source code. Multiple factors determine which genes are 'read' including cell signalling pathways, gene regulatory networks, chromatin conformation, and the presence or absence of epigenetic marks on DNA and chromatin proteins such as histones. Epigenetic (meaning over and above (epi-) the genome) marks are molecular tags which are placed on the DNA strands and histones that alter the way that chromatin can be interacted with by transcription factor 'readers' without changing the underlying sequence (Allis and Jenuwein, 2016).

One such tag is the methyl group which when deposited on the fifth carbon of cytosines found in CpG di-nucleotides forms 5-methylcytosine (5mC), a process known as DNA methylation which is performed by DNA methyltransferases (Stricker and Götz, 2018; Ambrosi, Manzo and Baubec, 2017; Schübeler, 2015). In mammals the proteins responsible for regulating 5mC patterns are the *de novo* methyltransferases Dnmt3a and Dnmt3b, and the maintenance methyltransferase Dnmt1 (Ambrosi, Manzo and Baubec, 2017). They operate during embryonic development with co-factors such as Dnmt3L, Uhrf1 and Hells (Hata *et al.*, 2002; Nishiyama *et al.*, 2013). Genome-wide profiling experiments have shown that 5mC is prevalent throughout mammalian genomes, except at CpG-dense regions known as CpG islands (CGIs) which are regulatory landmarks mostly associated with gene promoters (Bird, 1986). By varying the DNA methylation pattern, as well as DNA conformation and other epigenetic mark patterns, genes can be silenced or expressed in different combinations and at different times

enabling the formation of multiple cell types at different stages of an organism's life from one genetic code. These altered genomic DNA methylation profiles correlate with different cellular states and tissue types (Meissner *et al.*, 2008), thus DNA methylation profiling has been used extensively to chart changes in cell state during development.

DNA methylation is generally associated with transcriptionally silent genomic regions. DNA methylation is key for genomic imprinting, regulating germline-specific gene expression (e.g. X-chromosome inactivation), maintaining genomic stability by silencing repetitive element expression, and during cellular differentiation (Ambrosi, Manzo and Baubec, 2017). There are two leading theories as to how DNA methylation helps induce this silenced state. The first suggests this silencing is achieved by preventing transcription factors from binding to DNA at their target binding sites within gene promoter regions through steric hindrance, by blocking the binding site or changing the shape of the DNA helix (Isbel, Grand and Schübeler, 2022). The second theory suggests the 5mC mark is recognised and bound by methylation-binding proteins (MBDs) which then prevent transcription factor binding directly or through competition for binding motifs (Isbel, Grand and Schübeler, 2022). Either way, it remains unclear if the presence of 5mC is a cause or consequence of transcriptional repression is still debated (Shukla *et al.*, 2020).

The importance of 5mC in mammals is highlighted by its prevalence in many human diseases and cancers (Weinberg *et al.*, 2019; Greenberg and Bourc'his, 2019). In mouse and human somatic cells, disruption of the DNA methylation pattern results in mis-expression of genes and loss of genome stability which may predispose individuals to diseases such as cancer (Hackett *et al*., 2012; Schübeler, 2015) or developmental disorders (Zoghbi and Beaudet, 2016). Proper 5mC patterning is particularly important during embryonic development as hypomethylation through loss of *Dnmt3b* or *Dnmt1* expression result in embryonic lethality (Ambrosi, Manzo and Baubec, 2017). In mouse models, the global loss of DNA methylation caused by loss of *Dnmt1* or double knockout of *Dnmt3a* and *Dnmt3b* is known to be lethal by

embryonic day 9.5 (E9.5) ((Ambrosi, Manzo and Baubec, 2017; Li, Bestor and Jaenisch, 1992; Okano, Bell and Haber, 1999). However, in pluripotent mouse embryonic stem cells (ESCs) loss of DNA methylation has no effect on cell phenotype or growth; ES cells which lack the three main DNMTs (triple knockout (TKO) cells lacking *Dnmt1*, *Dnmt3a* and *Dnmt3b*) can grow and divide indefinitely in cell culture (Tsumura, 2006). These observations present a paradox: how are some cell types able to tolerate hypomethylation but not others? It suggests a point is reached during early mouse embryonic development where DNA methylation becomes essential for viability. At what point during development does this occur? In this thesis I aim to address some of these questions using an ESC to embryoid body (EB) differentiation model.

EBs are spheroids of ES cells grown in the absence of Leukaemia Inhibitory Factor (LIF), the component of ES culture media which prevents their escape from pluripotency. When grown in suspension culture without LIF, ESCs aggregate to form an EB containing a mixed population of cells differentiating towards the three germ layers (endoderm, ectoderm, and mesoderm). Due to the similarities between EBs and pre-implantation blastocysts, EBs are commonly used to model *in vivo* embryonic development *in vitro* (Desbaillets *et al*., 2000; Jackson *et al.*, 2010).

To generate EBs I started with ESCs which have conditional *Dnmt1*$^{tet/tet}$ alleles, a gift from Prof. R. Chaillet. In these cells *Dnmt1* expression can be reversibly inhibited by culturing with the tetracycline-class drug doxycycline (dox; Figure 3.1 and 3.2) (Borowczyk *et al*., 2009). Both Dnmt*1* alleles are controlled by *tet* regulatory cassettes inserted into exon 1. A tTA activator gene is expressed from the endogenous *Dnmt1* promoter, and the resulting protein binds to the *tet* operator (tetO) sequence which is located at the 5’ end of the *Dnmt1* initiation codon, promoting the transcription of the methyltransferase gene *in trans*. Doxycycline binds the tTA protein, thus inclusion of dox in the culture medium prevents *trans* initiation of *Dnmt1* transcription (Borowczyk *et al.*, 2009). This leads to passive loss of 5mC and global hypomethylation. DNA methylation becomes

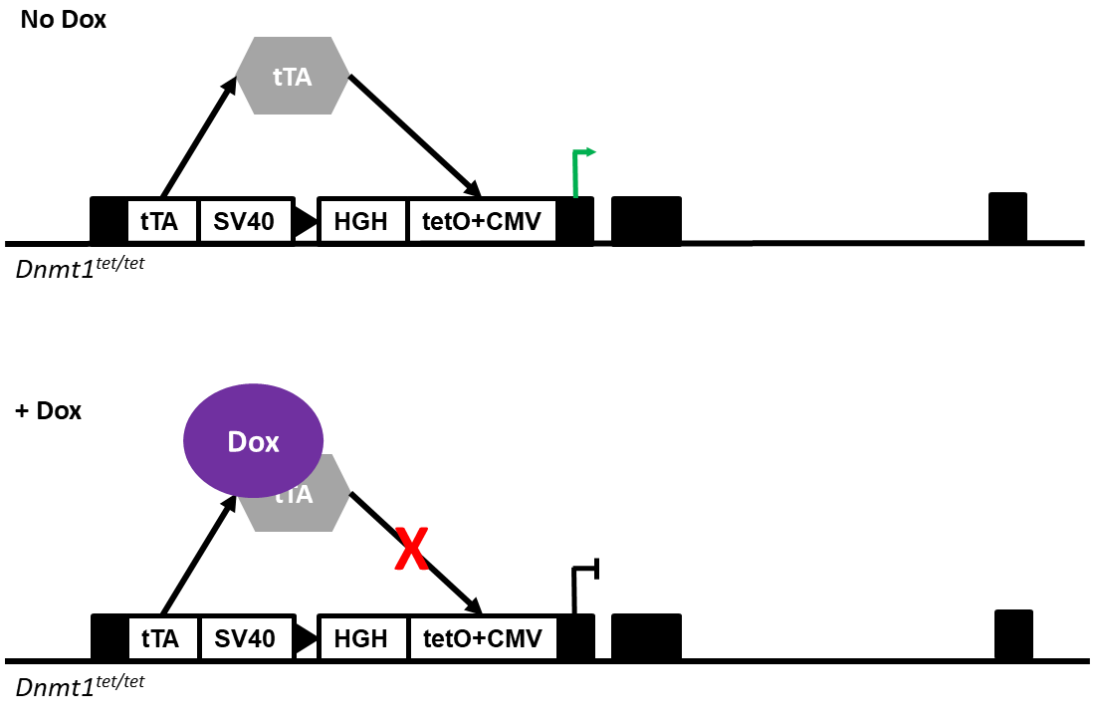


Figure 3.1. *Dnmt1* expression in *Dnmt1*$^{tet/tet}$ ES cells is inhibited in the presence of doxycycline (Dox). The cell line was a gift from R. Chaillet.

undetectable within 4 days of dox culture, but then returns to untreated levels within 7 days of cessation of dox treatment (Dunican *et al.*, 2020). This inhibition is reversible, as is illustrated in figure 3.2. This makes it a powerful system for studying the role of this epigenetic mark during differentiation as the levels of DNA methylation are easily tuneable. This line thus enables investigations into the effects of absence of *Dnmt1* and consequent loss of global DNA methylation during *in vitro* differentiation.

The aim of this chapter is to optimise the differentiation protocol and ensure EBs are compatible with dox culture, and to characterise EBs cultured with and without *Dnmt1*.

3.2 Results

3.2.1 Loss of DNA methylation impacts on embryoid body formation and gene expression

Loss of Dnmt1 is tolerated by ESCs but not by embryoid bodies

Previously in the lab it has been shown than 2µg/mL dox is sufficient to reduce Dnmt1, and consequently global 5mC, to nearly undetectable levels in ESCs within 4 days (S. Dimova 2021, unpublished thesis). The ESCs can be cultured long-term under these conditions with no adverse effects on pluripotency or cell growth, and the effect is reversible (Dunican *et al.*, 2020) (Fig. 3.2). The initial experiments with no dox versus high dox EBs showed a clear distinction between the EBs with normal DNA methylation and those cultured under 'hypomethylation conditions', i.e., those treated with doxycycline. The untreated *Dnmt1*$^{tet/tet}$ ESCs reliably formed smooth, large EB spheroids after 3 days incubation in suspension. However, *Dnmt1*$^{tet/tet}$ ESCs cultured with 2µg/mL dox for four days and then suspended in droplets of EB media with dox frequently failed to aggregate, and most EBs that did form were smaller, granular, and prone to disintegrating during transfer to adherent culture (Fig 3.3). By day 4 the surviving dox-treated EBs were typically half the size of the untreated EBs. This difference in size and EB quality was striking, and immediately suggested that although ESCs can tolerate the absence of Dnmt1 and relative paucity of DNA methylation, following initiation of differentiation the mark becomes much more significant for EB formation.

To try to get to the root of this phenomenon, I went on to optimise the EB differentiation protocol to find a dox treatment condition in which I could obtain enough EB material from hypomethylation conditions to perform further downstream analyses.

Optimising the differentiation protocol

A sub-population of *Dnmt1*$^{tet/tet}$ ESCs were cultured with dox for at least 4 days to induce hypomethylation, and these were used to generate the

dox-treated EBs. To produce EBs, untreated and hypomethylated ESCs were diluted to 200cells/μL in EB media (see Materials and Methods for details) and suspended from the lids of 15cm tissue culture dishes in 20μL droplets, each containing approximately 400 ESCs. The plates were incubated for 3 days at 37°C, 5% CO_2, and the cells allowed to aggregate together under gravity. After 3 days in suspension the resulting EBs were transferred to a gelatinised plate for further culture for up to a maximum of 14 days. 14 days was set as the culture time limit as EBs are reported to have reached the end stages of differentiation by this point (Jackson *et al.*, 2010). This was also suggested by the significant reduction in speed of colony growth and stabilisation of colony size between days 10-14 suggesting an increase in cell cycle time indicative of exit from pluripotency in most cells. An overview of the experimental setup is shown in figure 3.4.

In line with the *Dnmt1*$^{tet/tet}$ ESCs, I initially cultured dox-treated *Dnmt1*$^{tet/tet}$ EBs with 2μg/mL dox (Referred to later as the 'high dox' treatment). As previously mentioned, the high dox treated EBs frequently failed to form EBs, and those that did form were small and fragile. Furthermore, once plated the high dox treated EBs were easily washed off and damaged when changing media, so it quickly became clear that obtaining sufficient high-quality material from later time points would be difficult. To address this, I experimented with other doses of dox. A previous student showed that in *Dnmt1*$^{tet/tet}$ ESCs a dose range from 0.0002-2μg/mL was sufficient to deplete Dnmt1 to similarly undetectable levels and induce hypomethylation within 4 days of culture (I. MacGregor 2019, unpublished thesis; Fig. 3.5).

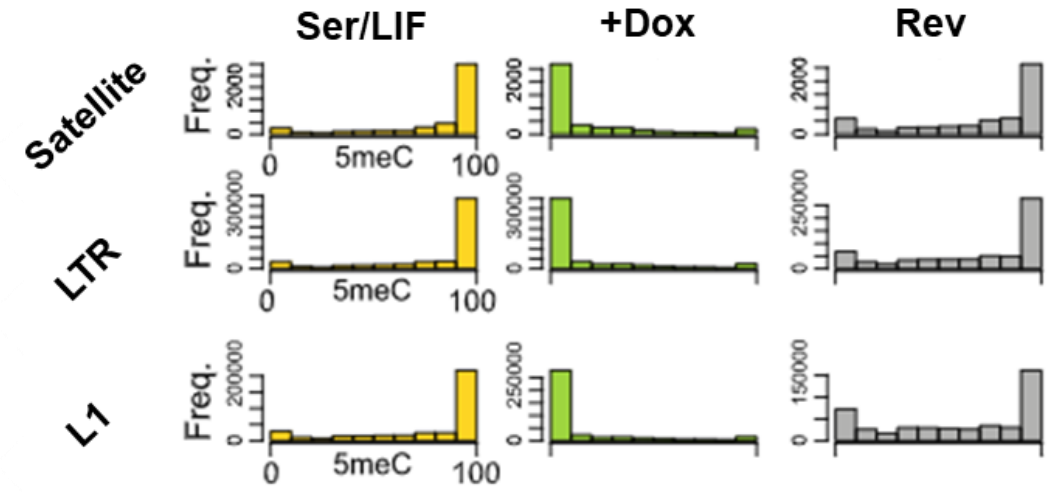


Figure 3.2. Inhibition of *Dnmt1* expression in *Dnmt1*$^{tet/tet}$ ESCs with dox is reversible. This plot is adapted from Dunican *et al.,* 2020. The histograms show the proportion of methylated CpGs (5meC) found in repetitive DNA (Satellite, LTR and L1 regions) by whole genome bisulfite sequencing. In untreated ESCs (Ser/LIF) most of the cytosines found within CpGs are methylated, whereas the addition of dox for 7 days (+Dox) leads to absence of the methyl mark at most cytosines. However, 7 days after cessation of dox treatment most cytosines revert to methylation levels similar to those found in untreated ESCs (Rev).

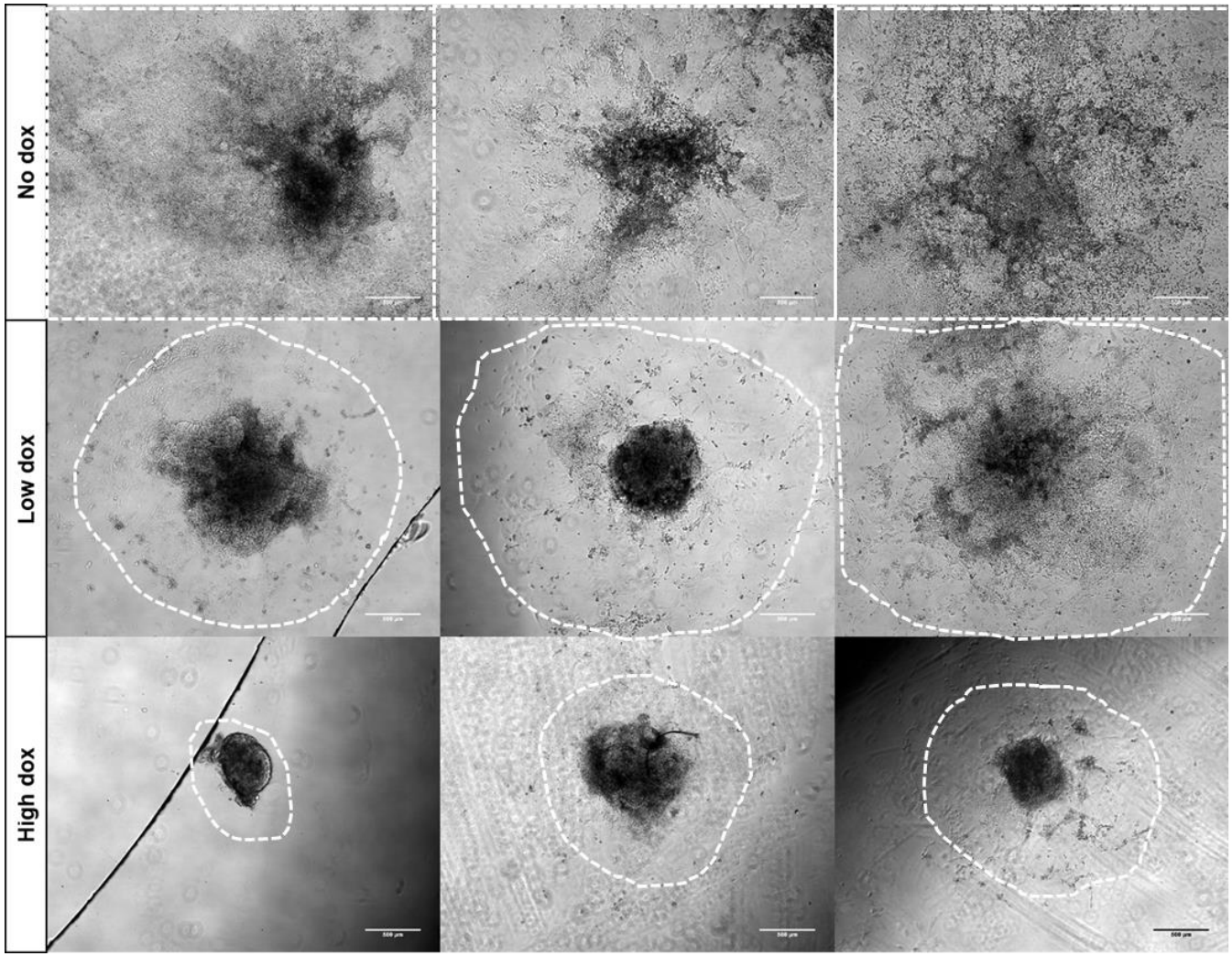


Figure 3.3. Removal of *Dnmt1* leads to reduced EB size and complexity. EBs were derived from *Dnmt1*$^{tet/tet}$ ESCs (a conditional cell line wherein addition of dox inhibits *Dnmt1* expression) and imaged at day 14 of culture. These images were captured on an inverted epifluorescent microscope at 5X magnification. Dox-treated EBs were smaller overall and less spread than untreated EBs. This phenotype was more extreme with the higher dose. The white dashed lines indicate the approximate outer boundaries of the EBs. The 'no dox' EBs exceed the size of the frame. Low dox = 0.2µg/mL; High dox = 2µg/mL. Scale bars represent 500µm.

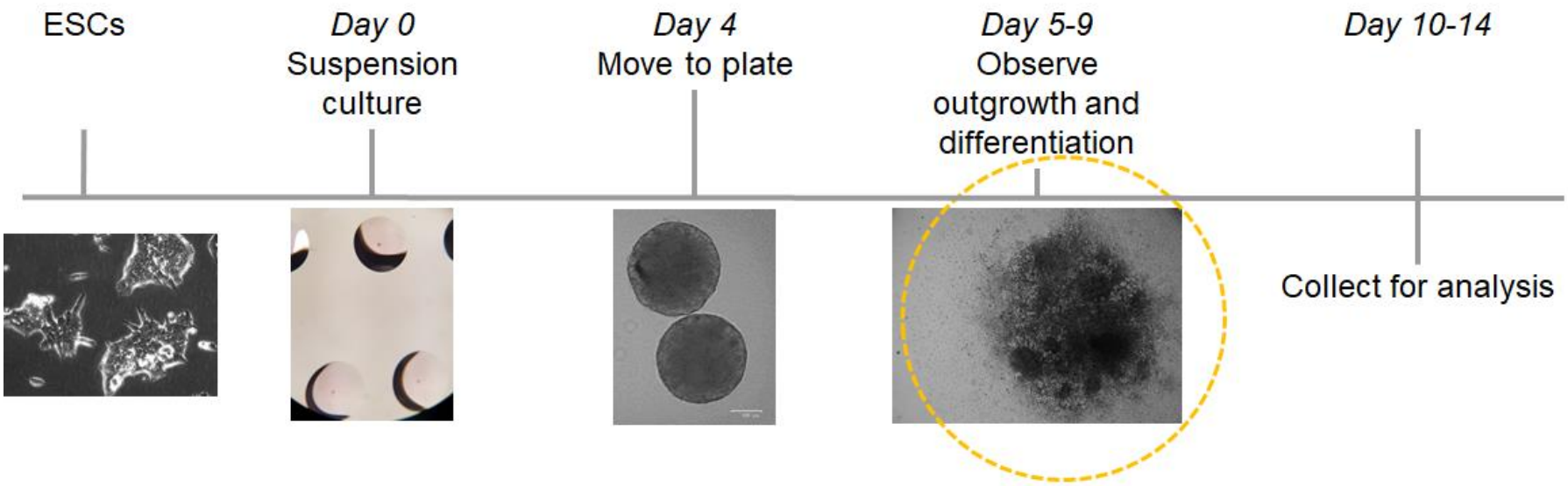


Figure 3.4. Overview of embryoid body generation protocol. Embryoid bodies are formed by suspending 20µL droplets of EB media on the underside of a 15cm cell culture dish lid (Day 0). Each droplet contains a single cell suspension of approximately 400 *Dnmt1*$^{tet/tet}$ ESCs. Over 72 hours the cells aggregate under gravity to form spheres of differentiating cells. After 96 hours these spheres are moved to a gelatinised surface where they adhere and spread out to cover the surface over time. From day 10 beating cardiomyocyte cells may be observed. The EBs are cultured for a maximum of 14 days before they are collected for analysis. The yellow dashed line highlights the approximate full size of the EB.

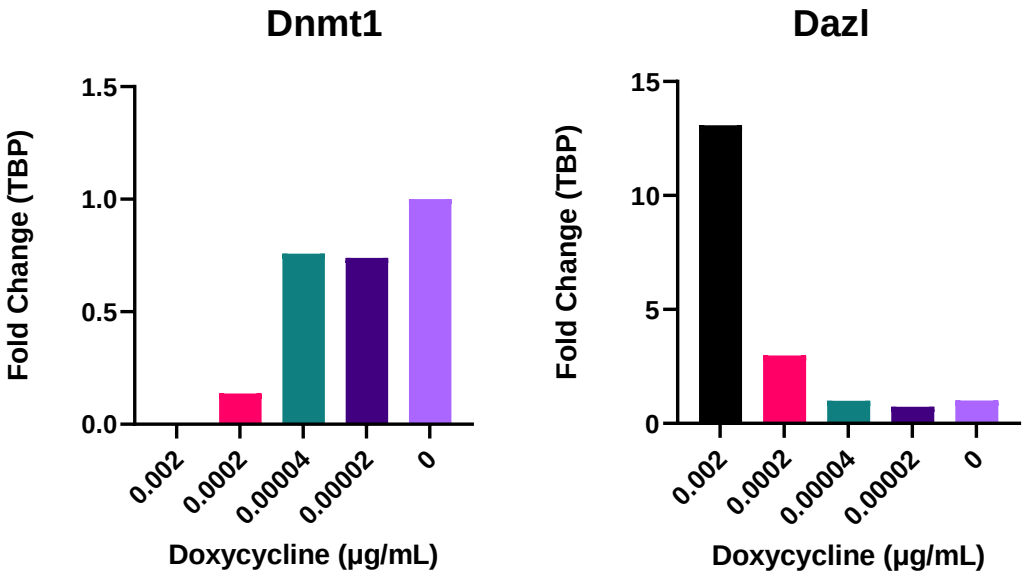


Figure 3.5. At least 0.002µg/mL doxycycline is required to reduce *Dnmt1* expression to undetectable levels. *Dazl* is a germ-cell gene which is specifically silenced by DNA methylation; increased expression of *Dazl* indicates loss of expression of *Dnmt1*. This data was generated by Dr Issie MacGregor. Expression fold change normalised to TBP. Further experimental details are available in her thesis.

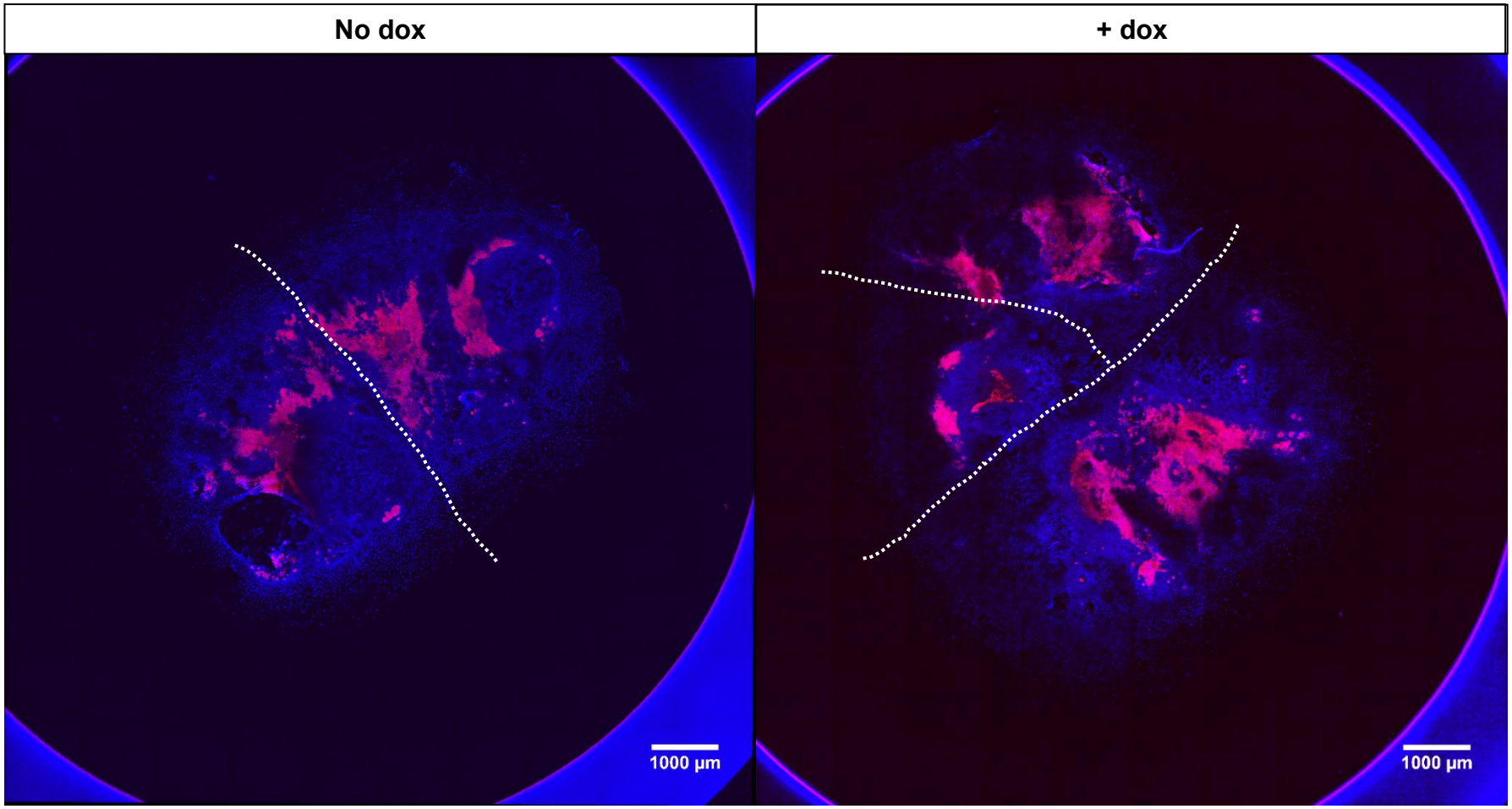


Figure 3.6. Untreated (no dox) and hypomethylated (+ dox) EBs maintain populations of Oct4+ pluripotent cells within the centre region. The left image contains 2 untreated overlapping *Dnmt1*$^{tet/tet}$ EBs at day 14 of differentiation. The right has ~3-4 dox-treated, hypomethylated EBs. All cells are stained with DAPI (blue), whereas *Oct4*-expressing cells are stained red. The dotted lines roughly mark EB boundaries. The scale bar represents 1000µm.

As EBs form very large 3D structures with dense patches of cells that might be hard to penetrate, and as the media would be refreshed every 48 hours, I decided the test 'low dox' condition should be no more than ten-fold lower than the high dox condition, i.e. 0.2µg/mL. In this way I hoped to find a dox dose which would maintain the hypomethylated state without compromising EB viability to the point where experiments became impractical. For this reason, when the EBs were seeded on adherent plates the dox dosage was either maintained at 2µg/mL (High dose) or reduced to 0.2µg/mL (Low dose). The dosage was only reduced upon transfer to adherent culture to ensure hypomethylation was maintained during 3-4 days suspension culture as the media could not be refreshed during this time.

Upon transfer to adherent culture, untreated EBs typically formed large (>1mm^2), flattened structures with dense mounds of cells at the centre which spread to form a monolayer at the outer edges. By day 14 single EBs were large enough to be seen with the naked eye. In contrast, dox treated EBs were much smaller; high dose treated EBs remained mostly spherical with very few cells spreading from the central mass, which blackened and were loosely adherent (Fig 3.3). The low dox treated EBs grew slightly more than the high dose EBs but were still considerably smaller than untreated EBs. They could spread to form adherent colonies and were able to survive to day 14 at greater frequency without losing significant numbers to washes when changing media. When stained, it was seen that EBs maintained populations of Oct4+ pluripotent cells near the centre of the mass; this is highlighted in figure 3.6 and will be explored in more detail in Chapter 5.

Testing the differentiation protocol showed that EBs could be derived from the *Dnmt1*$^{tet/tet}$ ES cell line, but that the dox dose had to be moderated to enable sufficient numbers of treated EBs to reach day 14 of development. Having optimised the dox protocol, I moved on to confirm the hypomethylation status in treated EBs.

3.2.2 EBs cultured under hypomethylation conditions were consistently smaller than wild-type EBs

Induced hypomethylation was maintained throughout differentiation

To confirm that the dox treatment was inhibiting *Dnmt1* expression and maintaining hypomethylation throughout the experiment, DNA was collected at multiple time points during 14 days of culture in EB media with or without a low dose of dox. The DNA was digested with restriction enzyme *MaeII* which recognises and cuts at A'CG'T, but whose activity is blocked by the presence of 5mC. The digested DNA was analysed by Southern blot using a probe targeting major satellite DNA (the probe was a gift from the Bickmore lab, IGC). Satellite DNA is found in peri-centromeric heterochromatin and is rich in CpG repeats. This DNA is normally highly methylated to maintain chromatin compaction and prevent transcription of the repeats, which is key for chromosome stability (Ambrosi *et al*., 2017; Chen *et al*., 2004). The digest patterns were compared to digested wild-type J1 ESC DNA, and to digested DNA from constitutively hypomethylated *Dnmt* TKO cells (Fig. 3.7). The Southern showed that *MaeII* was able to digest the dox-treated EB DNA and produce a blot pattern of low molecular weight bands similar to that of the TKO DNA. In contrast, DNA from the untreated EBs was digested to a similar level as wild-type J1 ESC DNA which is normally methylated and therefore mostly undigested at days 3, 5 and 14; this can be seen as high molecular-weight bands and smears at the top of the blot. Interestingly there appears to be programmed partial hypomethylation at the satellite repeats in day 7 and day 10 wild-type EBs, which is restored after this time point. This may be indicative of a period of programmed demethylation in a subpopulation of cells within the EBs; *in vivo*, this occurs during 2C-like cell formation and primordial germ cell (PGC) reprogramming. I will investigate this intriguing observation further in the next chapter.

In addition to the Southern blot, I performed western blot analysis to track the depletion of Dnmt1 protein in *Dnmt1*$^{tet/tet}$ ESCs and during EB differentiation in the presence of doxycycline (Fig. 3.8). Protein was collected

from untreated and low dox dose treated ESCs and EBs at days 4, 7 and 10 of differentiation. 30µg protein per sample were loaded onto a gel, separated by electrophoresis, and probed with α-Dnmt1or α-H4 primary antibodies. Antibody signal was detected by chemiluminescence imaging. H4 protein was used as the loading control as it is similarly abundant across the tissue types tested. Protein from hypomethylated TKOs was used as a negative control. The blot showed that the 184kDa Dnmt1 protein was undetectable in TKOs and dox-treated samples, including the late-stage differentiating EBs. However, in the untreated samples the methyltransferase was clearly detectable.

Altogether these results suggest the low dox dose and drug refresh rate used were sufficient to maintain very low levels of DNA methylation in treated EBs throughout the 14 days of differentiation.

Oct4 is not repressed following differentiation in the absence of Dnmt1

Having established that dox treated EBs are hypomethylated throughout culture I went on to compare gene expression between dox treated and untreated EBs to see if differentiation and loss of pluripotency could be detected by quantitative PCR (qPCR). I extracted RNA from EBs collected at day 14 of differentiation which had been cultured with no (normally methylated), low (hypomethylated) or the high dose (hypomethylated) of dox, and from *Dnmt1*$^{tet/tet}$ ESCs as a pluripotent line control. For the qPCR I used primers targeting genes associated with pluripotency (*Oct4* or *Pou5f1*), and specific lineage marker genes (*Gata4* for endoderm, *Otx2* for ectoderm, *Snai1* for mesoderm). As a secondary measure of dox efficacy in inhibiting *Dnmt1* expression in *Dnmt1*$^{tet/tet}$ ESC-derived EBs, primers for detecting *Dazl* transcripts were included in the qPCR. In mouse *Dazl* is only expressed in germ cells and is specifically repressed by DNA methylation in other cell types (Maatouk *et al*., 2006). Studies have shown that inducing hypomethylation in mouse ESCs leads to aberrant expression of *Dazl (*Maatouk *et al*., 2006; Hackett *et al*., 2012; Shukla *et al*., 2020*)*. The qPCR results are summarised in figure 3.9.

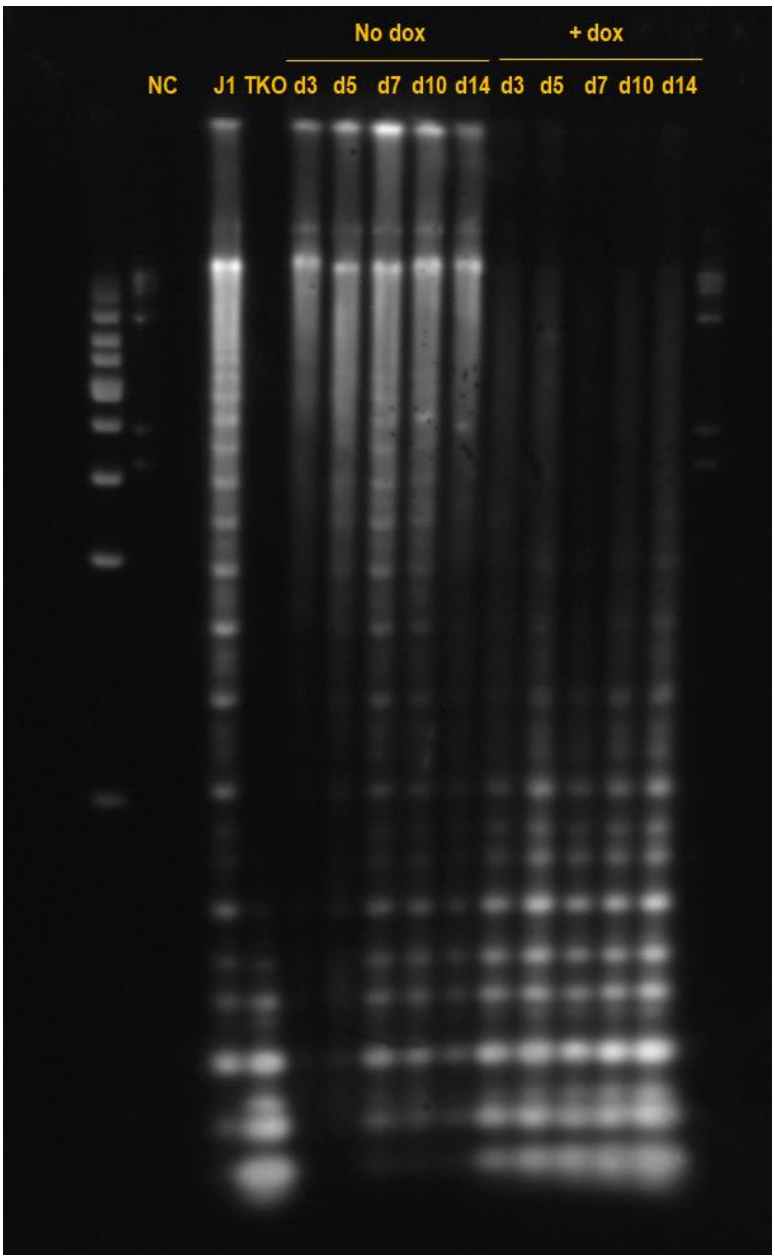


Figure 3.7. Dox treatment maintains hypomethylation throughout EB differentiation. To check if hypomethylation induced by low dose dox treatment (0.2µg/mL) was maintained throughout 14 days of EB differentiation, DNA from EBs generated from untreated *Dnmt1*$^{tet/tet}$ cells (normally methylated) and dox-treated *Dnmt1*$^{tet/tet}$ (hypomethylated) cells were collected and digested with a methylation-sensitive restriction enzyme, *MaeII*. The digested DNA was analysed by Southern blot with a probe targeting CpG-rich major satellite DNA. The dox-treated EBs show similar levels of digestion compared to constitutively hypomethylated TKO cell DNA throughout the time course, whereas untreated EBs have normal levels of methylation similar to wild-type J1 ESCs. This suggests a low dose of dox is sufficient to maintain hypomethylation during EB differentiation.

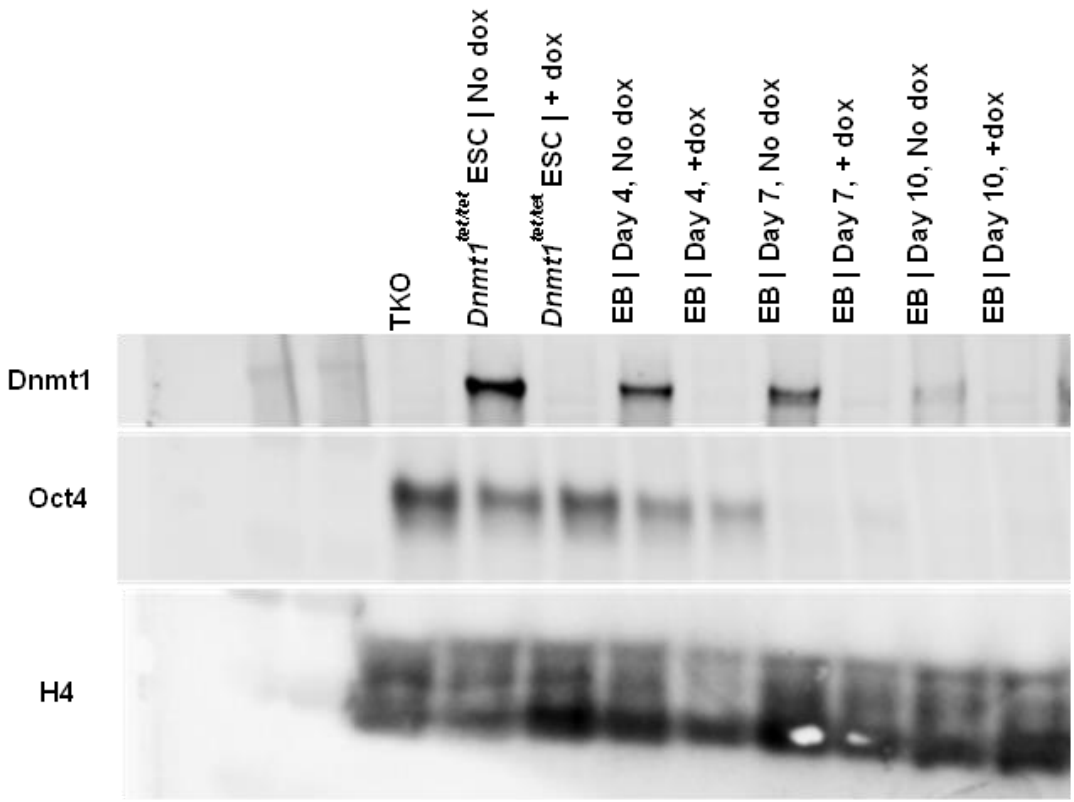


Figure 3.8. Dnmt1 is lost in *Dnmt1*$^{tet/tet}$ ESCs and EBs cultured in the presence of doxycycline (dox). Oct4 abundance decreases with differentiation but is slightly higher in +dox EBs. 30µg protein per sample was loaded. H4 protein was used as the loading control. Protein extracted from TKO ES cells, which lack the *de novo* and maintenance DNA methyltransferases, was used as a negative control.

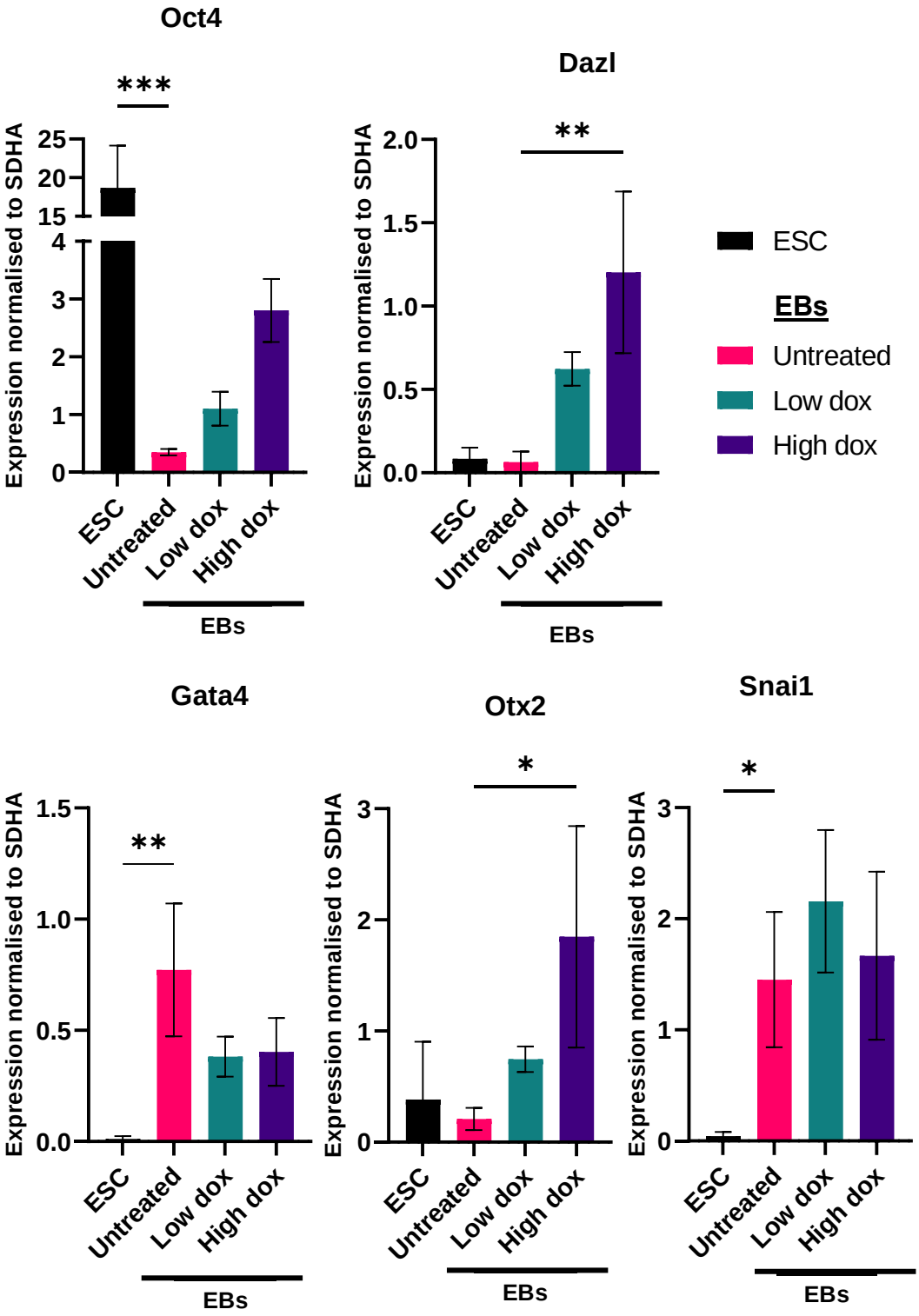


Figure 3.9. Dose-dependent changes in pluripotency and differentiation gene expression caused by loss of *Dnmt1* in EBs, compared to wt ESCs and untreated EBs. Relative expression calculated by ΔΔCt method for triplicate samples normalised to Sdha. Mean expression compared to untreated EBs analysed by one-way ANOVA. Only statistically significant changes shown (*=p≤0.05, **=p≤0.001, ***=p≤0.0001).

In the untreated ESCs *Dazl* was almost undetectable, as expected. The lineage marker genes were also very lowly expressed, as they are only activated in differentiating cells. Pluripotency gene *Oct4* is normally highly expressed in ESCs, which was observed. In comparison, *Oct4* is lowly expressed in EBs and the lineage marker genes are more highly expressed, particularly *Gata4* and *Snai1*. In the absence of dox *Dazl* is also almost undetectable in EBs.

In the dox treated EBs the normal pattern of expression changed, regardless of dox dose. Firstly, *Dazl* expression was high indicating the loss of *Dnmt1* led to hypomethylation and de-repression of the 5mC-sensitive gene. *Oct4* was much more highly expressed than in untreated EBs, though not to the same levels as in ESCs. *Gata4* expression was slightly lower in dox-treated EBs, *Otx2* was higher, and *Snai1* expression was similar between treated and untreated EBs. Comparing the dox-treated EBs only, expression of *Oct4*, *Dazl* and *Otx2* is highest in the high dox condition, suggesting a dose-dependent effect. These observations were subsequently confirmed by RNA sequencing studies (see Chapter 4; genome browser views of these genes are shown in the Appendix), but also suggested *Snai1* expression is reduced in the high dose EBs compared to low dose and normal EBs.

Altogether this experiment provides more evidence that dox treatment, at either the high or low dose, prevented re-expression of *Dnmt1* and maintained hypomethylation throughout 14 days of differentiation as shown by the expression of *Dazl*. It also suggests that in the absence of Dnmt1 key pluripotency gene *Oct4* is not silenced upon initiation of differentiation but is co-expressed with the lineage marker genes, which are increasing in expression.

It is interesting to compare the gene expression changes between the dox conditions. EBs of both dox dosages were derived from hypomethylated ESCs and the dox treatment was acting to maintain a pre-existing hypomethylated state, not induce it. However, the effect on gene expression is more subtle in the low dox EBs. This could suggest that in the high dox

EBs the gene expression is being affected by further downstream consequences of DNA hypomethylation, and perhaps indicates indirect consequences of *Dnmt1* loss are enhancing the changes in gene expression. This may explain the increased severity of the phenotype seen in high dox EBs, wherein an increase in downstream off-target effects of loss of Dnmt1 may trigger apoptosis in more cells resulting in smaller and more fragile EBs. It should also be noted that even in hypomethylated conditions some residual 5mC is present due to compensatory activity by the *de novo* Dnmts, particularly at repetitive regions containing TEs which have been found to retain the mark the longest when DNA methylation is removed, such as during cellular reprogramming (Deniz *et al*., 2019). It could be that in the high dox condition the global levels of 5mC become so low that the mark is lost even from the repetitive regions leading to expression of TEs, which in turn would cause loss of genome stability. This instability may be the cause of the extreme apoptotic-like phenotype observed in high dox EBs.

3.2.3 Does loss of Dnmt1 trigger p53-driven apoptosis in EBs?

A pattern which was frequently observed across experiments was that loss of *Dnmt1* expression led to smaller EBs, loss of adherence, and blackening of the central mass suggestive of cell mortality (Fig. 3.3). The differences in size were considerable, but the cause was unclear. I initially hypothesised that the inhibition of *Dnmt1* expression with dox caused treated EBs to be smaller by disrupting the normal signalling patterns and stimulating the intrinsic apoptosis pathway in many cells, leading to a reduction in total cell number and therefore EB size and growth rate.

I tested this hypothesis with the help of PhD students Katie Pickup and Chetan Srinath, who together performed the experiment I planned. We tested for changes in apoptotic cell frequency by performing a Terminal deoxynucleotidyl transferase dUTP Nick End Labelling (TUNEL) assay, then analysed the assay results by flow cytometry. The TUNEL assay detects apoptotic cells by labelling the 3’-OH ends of DNA fragments with fluorescent FITC-dUTP (Gavrieli, Sherman and Ben-Sasson, 1992). These fragments

are found at DNA double-strand breaks (DSBs) which occur more frequently in cells undergoing programmed cell death as endonucleases are activated by the apoptotic caspases (D'Arcy, 2019). The FITC signal is then detected and quantified by FACS. The assay used included a propidium iodide (PI) stain step to distinguish healthy cells from fragmented cells and debris; PI is a cell-impermeant dye which can only stain cells which have compromised cell membranes, i.e. those which have been damaged mechanically, are necrotic, or are at the late stage of apoptosis.

Untreated *Dnmt1*$^{tet/tet}$ EBs and EBs treated with low dox were collected and fixed at day 7 of differentiation. Either 16 or 64 EBs were pooled and treated with the TUNEL kit (see Materials and methods); two different sized pools were used to test for assay efficiency with different quantities of starting material. The kit used provided apoptosis-positive and -negative control cells to enable the calibration of the flow cytometer. Flow cytometry was then used determine what proportion of the intact cells (PI^-) were undergoing apoptosis ($FITC^+$) at the time of fixation. The results are shown in figure 3.10 A and B.

Figure 3.10 part A shows the data from the control cells and highlights the regions of interest within the histograms. Both conditions have clearly segregated populations of healthy diploid cells (Peak highlighted to the right of the PI histogram) and haploid or fragmented cells. The bottom panel highlights the FITC+ cells which are inferred as apoptotic, demarcated in the right-hand box.

In figure 3.10 part B the EB TUNEL data are shown. Initial observations suggested that there are generally more FITC+ (apoptotic) cells in dox-treated EBs as shown by the second peak to the right of the FITC plot. However, there were inconsistencies between the 16- and 64-EB pools for each condition, particularly the proportion of cells within each population which passed the initial gating steps. In particular, EB samples are difficult to gate based on cell size due to the heterogeneous nature of the cell types and morphologies contained within maturing EBs, and between treatment conditions. Generous gating was performed to capture as wide a variety of the cells as possible for the samples to be representative and comparable.

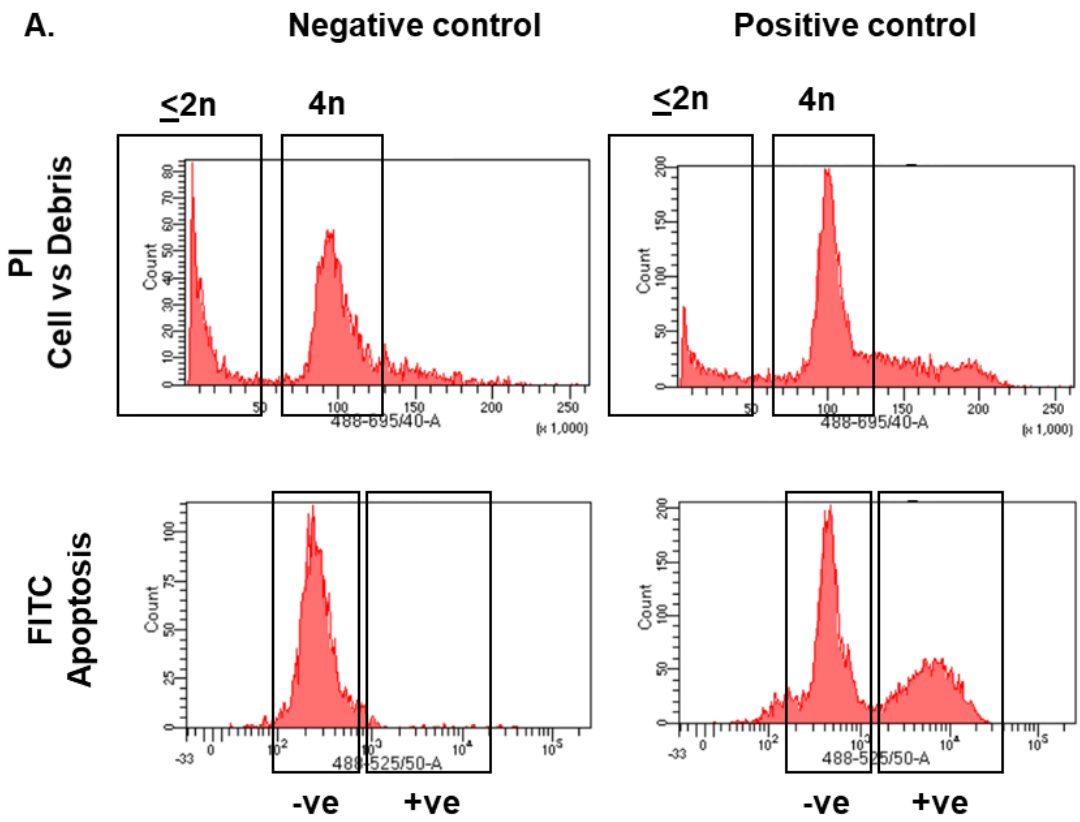


Figure 3.10, A. A TUNEL assay comparing apoptotic cell frequency in hypomethylated and wild-type EBs was inconclusive. Part A illustrates the controls used to calibrate the flow cytometer. Both control cell populations have distinct 2n (pre-replicative or fragmented cells) and 4n (healthy diploid cells) profiles as detected by propidium iodide (PI) staining, highlighted by the boxes. The positive control cells also show the expected pattern formed by the presence of FITC+ cells, which are absent in the negative control. 10,000 events were detected in total per sample.

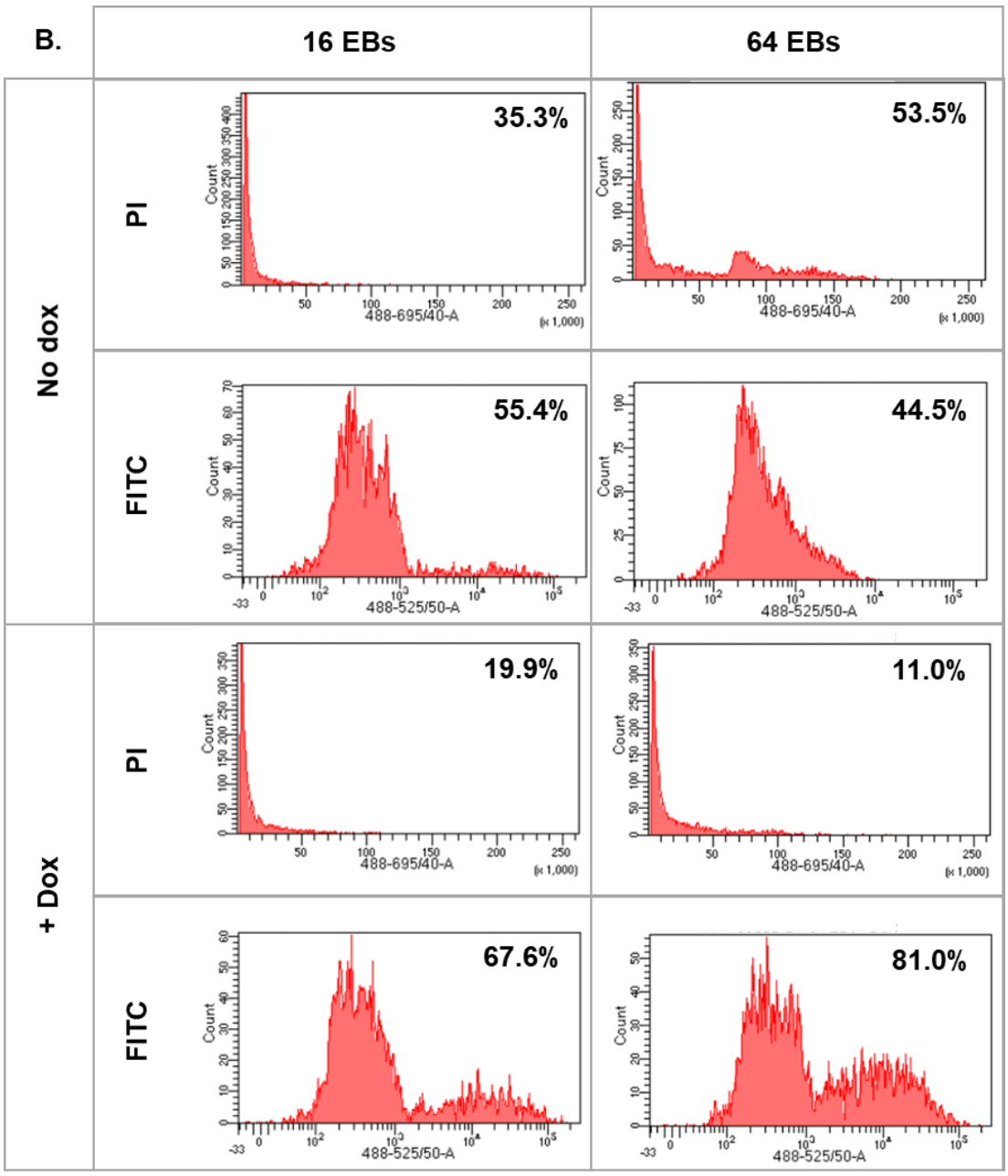


Figure 3.10, B. A TUNEL assay comparing apoptosis frequency in hypomethylated and wild-type EBs was inconclusive. Part B shows the results of the EB TUNEL assay. The percentages indicate the relative proportion of the cell population after gating for single cell size and granularity (Forward vs side-scatter calibration). The assay was performed using either 16 or 64 pooled EBs. 10,000 events were detected in total per sample. Both populations, regardless of dox treatment or EB number, show high frequencies of fragmented cells as indicated by the clustering of the histogram to the left of the PI plot. The +dox EBs generally showed a greater number of FITC+ cells, but consistency between samples and overall cell quality was poor.

From observing EBs with and without dox in culture, it was found that dox-treated EBs were generally less adherent and easily became detached during wash steps. It was thus expected that more damaged PI+ cells would be detected in these samples. However, as shown in figure 3.10 the 64 ‘no dox’ EB sample had the highest number of PI+ cells (peaks at the centre of the plot). It was expected that a small proportion of cells within an otherwise healthy untreated (no dox) EB population would become damaged during processing and/or naturally be undergoing apoptosis, and this would be more easily detected due to the larger number of cells in these samples as untreated EBs were relatively larger and more cell-dense compared to dox-treated EBs. That so few PI+ cells were observed in the equivalent plot for the dox-treated EBs was unexpected. It suggested either the number of cells per EB in this condition was too low, despite controlling for the number of EBs collected, possibly due to the lower density of cells In these EBs as observed previously (see EB size difference plots). Or, there were concerns more dox-treated EB cells were being lost during EB collection and the TUNEL staining process due to the low-adhesion phenotype noted under dox conditions. Technical issues meant further experimental optimisation and replicates were not performed, so this observation could not be checked and therefore these preliminary results were considered inconclusive.

The conclusion is also made unclear by the observation that the PI histograms consistently cluster to the left which suggested very few cells were PI+. This implies that most cells had intact membranes which the PI couldn’t penetrate, thus that most of the cells analysed were healthy. This contradicts the FITC+ cell numbers which suggest many are apoptotic. One caveat of this assay is that many cells are lost or became damaged during the TUNEL process, potentially leading to aberrant staining and making interpretation of the results difficult. This is particularly true of fragile cells, including those which are apoptotic or necrotic. This may be part of the reason why so many of the cells analysed were PI^-. Although considered the gold-standard method for detecting apoptosis, with slightly greater sensitivity

than the alternative Annexin V assay (Kylarová *et al.*, 2002), this test is limited by the fragile nature of the population being analysed.

As the TUNEL assay was inconclusive, other indicators of apoptosis were looked for. One stage of apoptosis involves the condensation of chromatin in preparedness for degradation which results in the formation of distinctive pyknotic nuclei, a process called pyknosis (Mooren and Krüger, 2015; Kerr *et al*., 1972; Ortiz, 2000)', . These nuclei can be seen using a light microscope; however, no enrichment was observed in hypomethylated EBs.

Another sign of high levels of apoptosis is the formation of nucleosomal ladders in raw DNA (Matassov *et al.*, 2004). Apoptotic cells destroy their nuclear contents by condensing the chromatin and increasing the activity of endonucleases to shear the DNA (Matassov *et al.*, 2004). The endonucleases cut the linker DNA between nucleosomes resulting in accumulation of fragments of DNA. When the undigested DNA is spread on a gel these fragments create distinctive nucleosomal ladder patterns.

Figure 3.11 shows a DNA digest which was performed as part of the preparation for the Southern blot shown earlier in the chapter. The purpose of the digest blot was to check the hypomethylation established in the dox-treated ESCs was maintained as they differentiated to form EBs and were continually cultured with dox. By comparing to normally methylated mouse ESC DNA (from J1 mouse strain ESCs) and hypomethylated mouse ESC DNA (ESCs with the three main Dnmts removed), it was observed that when digested with methylation-sensitive enzyme *HpaII* the hypomethylated EBs had digest patterns similar to the hypomethylated TKO ESCs throughout the time course. The Southern blot shows this more clearly. When using a methylation-insensitive restriction enzyme, *MspI*, dox-treated and untreated EBs and control J1s and TKOs were digested to similar levels. The digested DNA figure was shown here as it includes undigested DNA from each EB time point as a control to check the quality of the raw DNA extracts prior to digestion. What this also happens to show is that nucleosomal ladders are not observable in any of the undigested EB DNAs at any of the time points, in

either of the treatment conditions with either of the restriction enzymes. This suggests both the normally methylated and hypomethylated EBs had low levels of apoptotic cells. This opposes the hypothesis that dox-treated EBs are smaller due to an increase in apoptosis. However, it should be noted that DNA fragmentation is one of the final steps of apoptosis, so this simple assay may overlook populations of cells which are just initiating the pathway or are removed during media changes between sample collection times.

Overall, I observed that dox-treated EBs did not show an enrichment for apoptotic markers such as pyknotic nuclei or nucleosomal laddering, and the results of a TUNEL assay were inconclusive. Thus, I concluded that apoptosis was unlikely to be an important contributor to the size disparity observed between wild-type and hypomethylated EBs. Instead, it could be that cells in hypomethylated EBs have longer cell cycles due to the absence of Dnmt1, perhaps due to its absence at the replication fork causing it to progress or be recruited more slowly. Slower growth would lead to a smaller number of total cells in a hypomethylated EB compared to equivalent wild-type EBs, which would appear as a difference in size in a process similar to that proposed for primordial dwarfism (Klingseisen and Jackson, 2011).

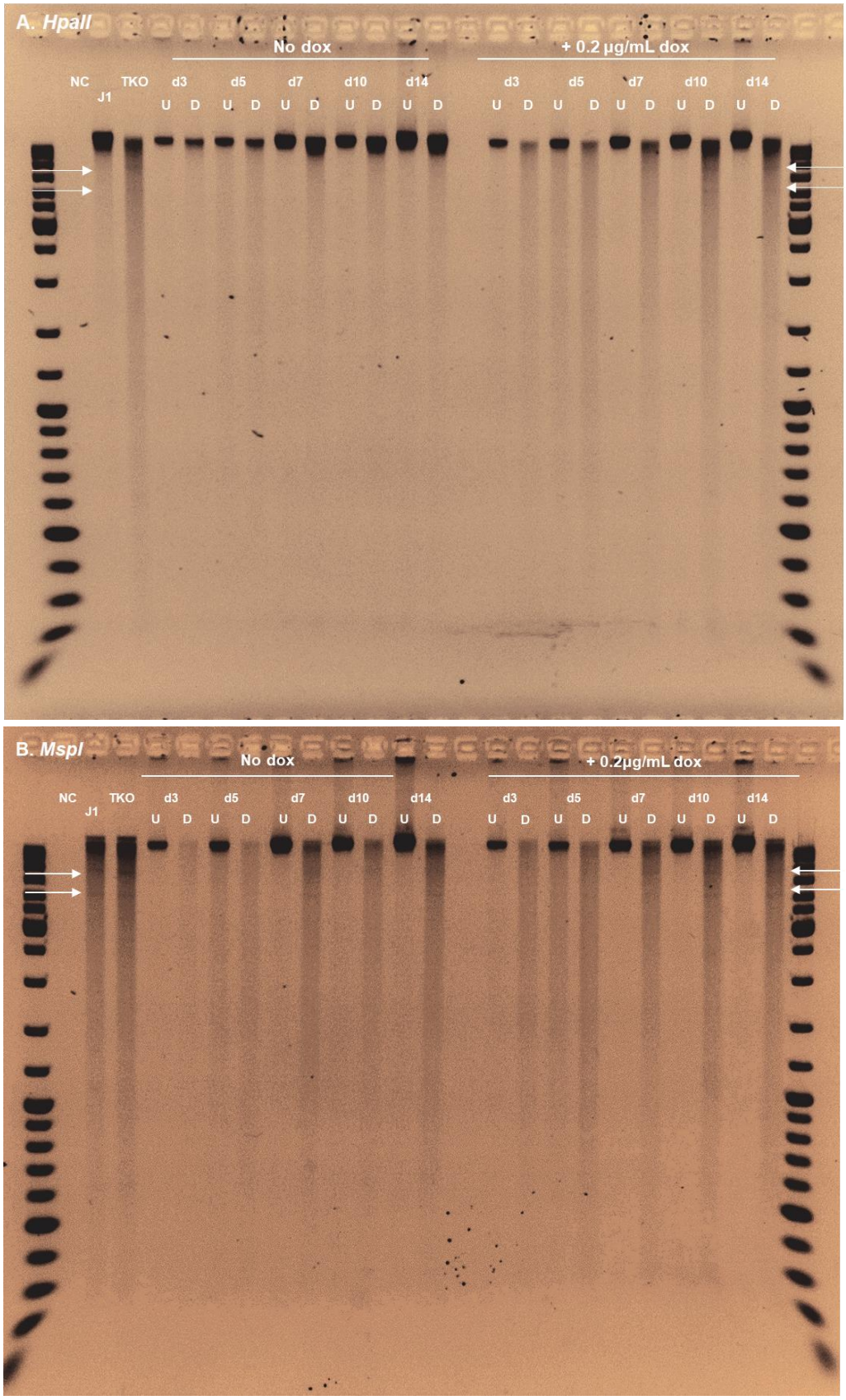


Figure 3.11. Legend on next page…

Continued from previous page…

Figure 3.11. *HpaII*-digested DNA showed hypomethylation was maintained in *Dnmt1*$^{tet/tet}$ EBs cultured with dox for up to 14 days, and shows the absence of nucleosomal laddering which would indicate an abundance of apoptotic cells. DNA from EBs with (No dox) and without (+ dox) *Dnmt1* expression were collected at intervals (e.g. d3 = day 3), digested with methylation sensitive (*HpaII*) and methylation insensitive (*MspI*) restriction enzymes and visualised by gel electrophoresis.

Gel A shows that, from dox-treated EBs, the DNA samples which were treated with *HpaII* became highly digested, developing smears down the gel. The untreated EBs retained *Dnmt1* expression and therefore their DNAs were normally methylated, so were resistant to digestion. Compared to the +dox EB DNA from the equivalent time-point, the DNAs from untreated EBs which were digested with *HpaII* (D) more closely resembled their uncut control samples (U) as the DNAs were unable to migrate through the gel. Overall, this shows that the dox treatment was able inhibit *Dnmt1* expression to induce and maintain hypomethylation in *Dnmt1*$^{tet/tet}$ EBs. The white arrows highlight LINE1 repeat DNA bands which appear in digested DNA samples.

Gel B was run as a control. As *MspI* is able to digest methylated DNA, the dox-treated and -untreated digested samples should look similar. Digested DNA (D) is able to migrate and thus forms a smear down the gel, with the smallest fragments travelling the furthest. Gel B therefore shows the DNAs of all the samples were permissive to digestion by restriction enzymes.

In both gels A and B, undigested samples (U) were included to show the quality of the DNAs used, and that none of the samples were significantly degraded prior to digestion; this degradation would be visible as smearing in the lanes. As shown here, good quality undigested DNA (U) is too large to be able to migrate through the gel and thus forms thick bands at the top. These samples would also indicate if a large proportion of the cells used for DNA extraction were undergoing apoptosis, as late-stage apoptotic cells break down the chromatin which then forms distinctive nucleosomal bands when separated on a gel. These bands were not observed in any of the undigested bands on either gel, suggesting the proportions of apoptotic cells in both untreated and +dox EB populations was low, and that dox treatment did not increase this proportion.

NC = No template control; J1s = ESCs with wild-type DNA methylation levels; TKO = Constitutively hypomethylated ESCs; U = Undigested; D = Digested.

3.2.4 Hypomethylated EBs still develop beating cardiomyocytes

It is known that ESCs can proliferate normally in the absence of 5mC but that the mark becomes essential at some point during early development as *Dnmt1* KO is embryonic lethal in mice by E9.5 (Li *et al.*, 1992). At what stage of development does *Dnmt1* expression become essential, and what pathways are being affected and leading to developmental failure when its activity is disrupted? Is DNA methylation required to regulate the expression of differentiation pathways? To investigate this, I wanted to be able to visualise the formation of specific lineages in live EBs. To this end, I integrated into *Dnmt1*$^{tet/tet}$ ESCs a Lifeact-mScarlet reporter by *piggyBac* cloning to obtain a modified cell line with both a tuneable *Dnmt1* cassette and a live fluorescent reporter for actin localisation, a product of the mesoderm differentiation pathway.

Fusions of actin-binding domains to mTurquoise2 have been used to visualise actin, notably the N-terminal 17 amino acids of ABP120 in mammalian cells, as its small size and absence from mammals make it an attractive actin marker (Goedhart *et al.*, 2012). The Lifeact fusion protein was designed by Riedl *et al.*, 2008 who determined that it is the first 17 amino acids of the ABP140 protein that are sufficient to localise the reporter to F-actin, the protein which makes up the cellular cytoskeleton. Lemieux *et al.* (2014) investigated the Lifeact protein further and concluded that it is better than rival reporters because it is smaller, its binding is not inhibited by endogenous proteins, and it is not cytotoxic. The super-bright red fluorescent protein mScarlet was developed by Bindels *et al.* (2016) who showed that it is more than 3x brighter than mCherry. They tagged the fluorophore to Lifeact to produce the Lifeact-mScarlet construct used here. The resulting construct is thus able to specifically highlight F-actin without harming the cells.

As an actin-binding reporter this construct would enable the visualisation of cardiomyocyte cell formation, thus acting as a reporter of mesoderm differentiation. The reporter was obtained from AddGene (pLifeact-mScarlet_N1 vector; deposited by Dr D. Gadella) and I modified it

by Gateway cloning to be compatible with the *piggyBac* vector system so that the reporter could be stably integrated into the *Dnmt1*$^{tet/tet}$ genome. The Gateway cloning *att*B guide sequences were designed by research assistant Dr David Hay. Red, mScarlet^{+} transfected cells were identified and sorted by FACS. Colonies grown from sorted single cells were subjected to hygromycin selection to select for stably integrated clones (Fig. 3.12). One clone of medium mScarlet intensity was picked and used for all subsequent experiments.

To determine the effect of global loss of 5mC on cardiomyocyte formation, the previous dox experiment was repeated with the new *Dnmt1*$^{tet/tet}$ Lifeact-mScarlet cells. EBs were cultured with the low dose or without dox and observed over 14 days of development (Fig. 3.13 A-C). EB size was measured by imaging the RFP signal at days 7, 10 and 14, and determining the area of the EBs in FIJI using colour thresholding (see Materials and methods; Fig. 3.14).

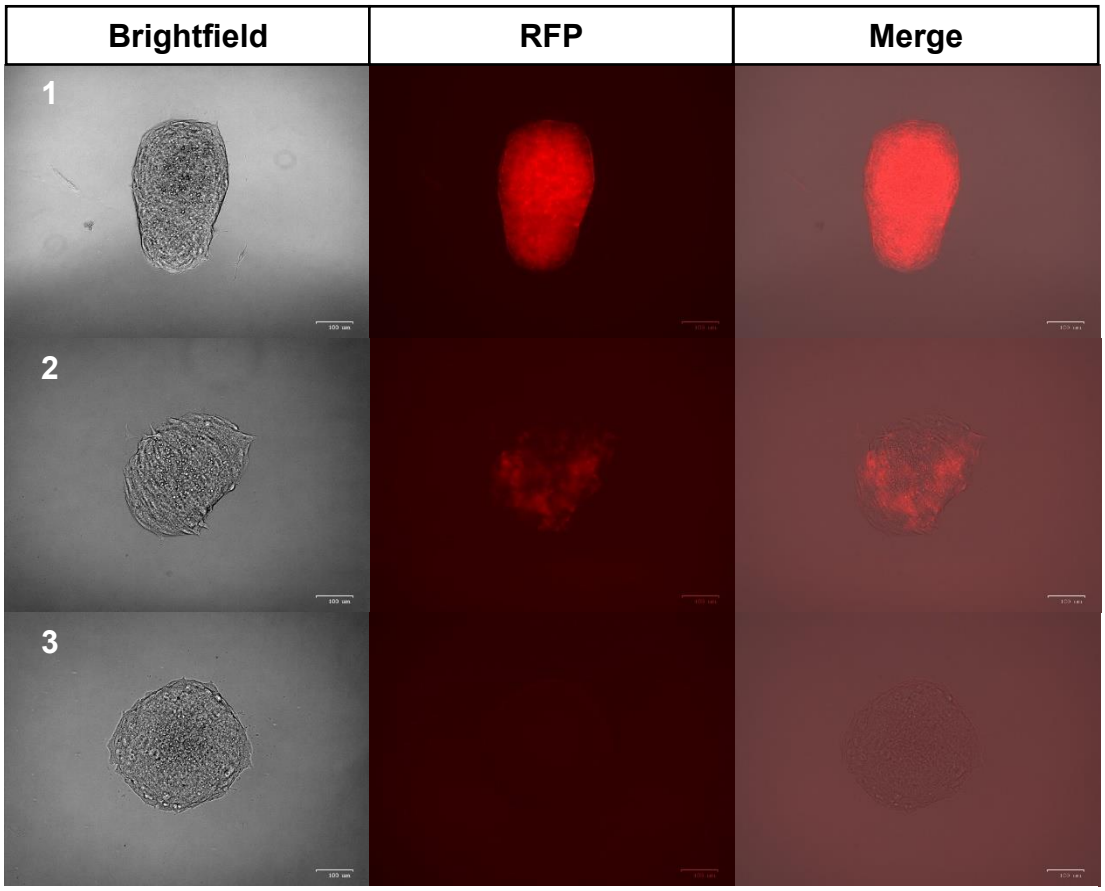


Figure 3.12. *Dnmt1*$^{tet/tet}$ ESCs transfected with Lifeact-mScarlet 6 days after FACS sorting for RFP. Some clones stopped expressing RFP signal after sorting (e.g., row 3), so all picked clones were subjected to hygromycin selection to ensure the Lifeact construct was stably integrated. Clonal colonies were imaged on a ZOE fluorescent cell imager (Bio-Rad). Scale bars represent 100μm.

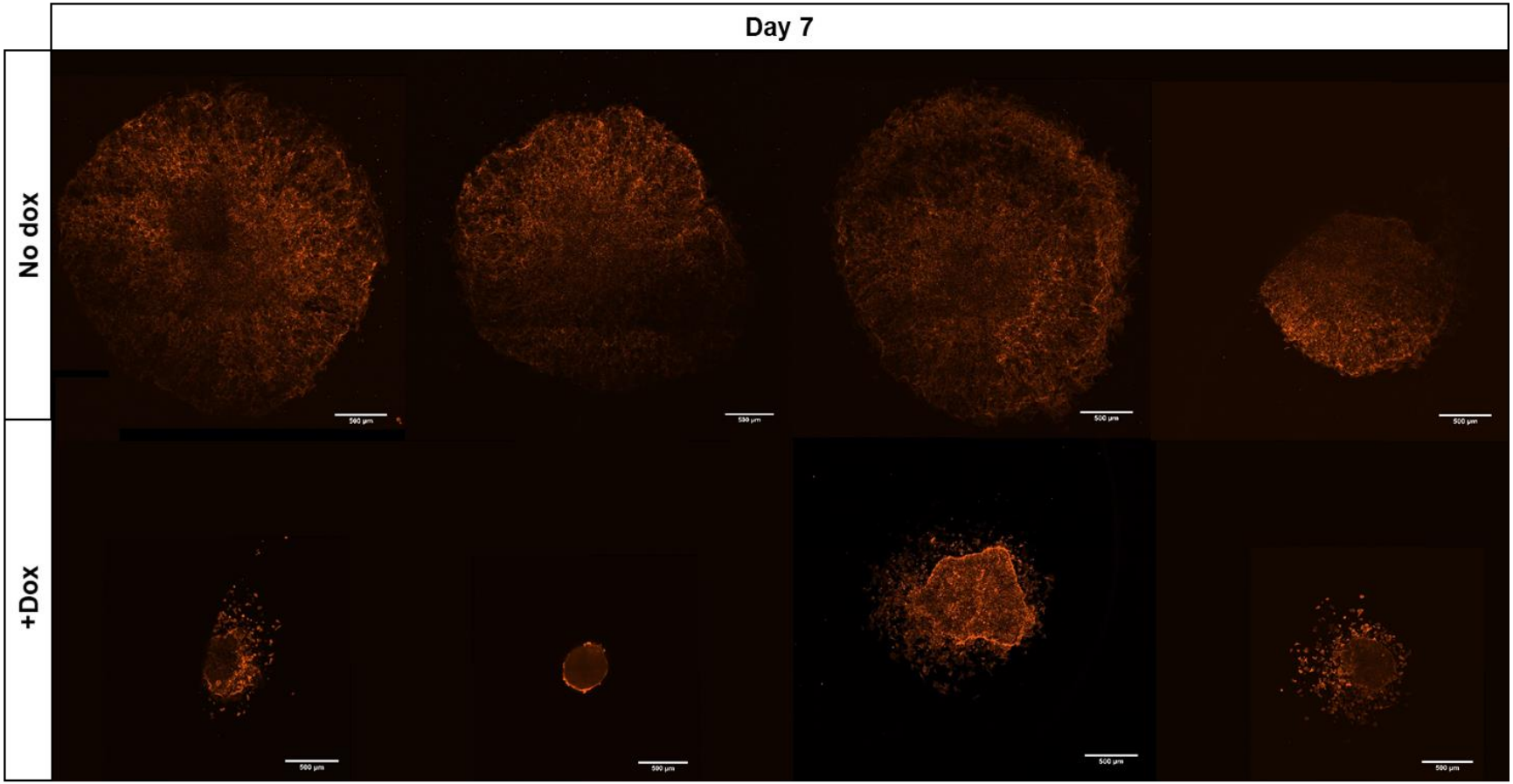


Figure 3.13, A. Lifeact-*Dnmt1*$^{tet/tet}$ EBs at day 7 of differentiation show clear differences in size. Lifeact-*Dnmt1*$^{tet/tet}$ mESCs were treated with water or 0.2µg/mL dox for 4 days then differentiated to form EBs and imaged at days 7, 10 and 14 on a confocal microscope. By day 7 dox-treated EBs are smaller and less complex than untreated EBs. Tiled images were captured at 20X and stitched using Fusion software (Andor). Scale bars represent 500µm.

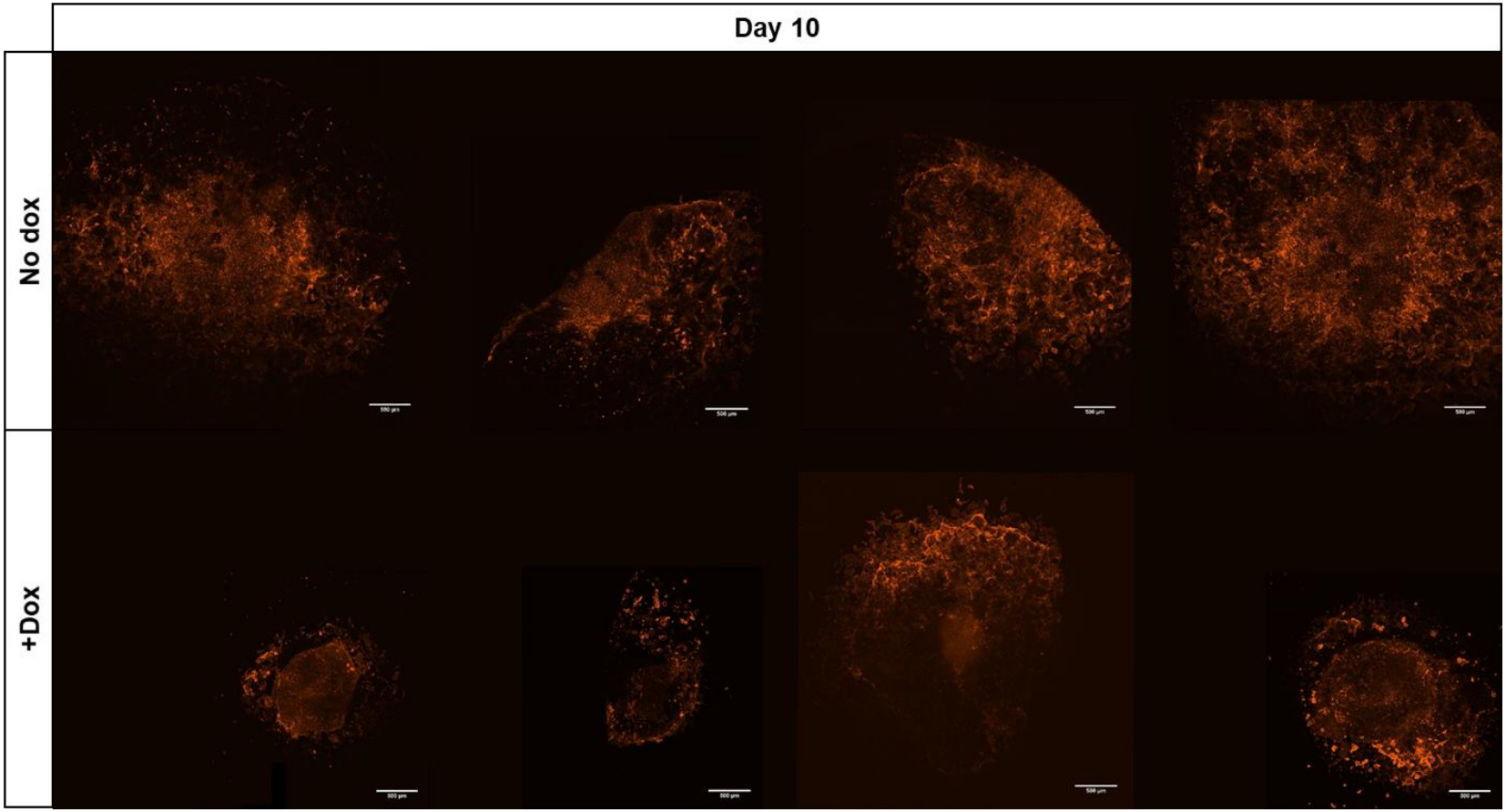


Figure 3.13, Bb. Lifeact-*Dnmt1*$^{tet/tet}$ EBs at day 10 of differentiation show clear differences in size. Lifeact-*Dnmt1*$^{tet/tet}$ mESCs were treated with water or 0.2µg/mL dox for 4 days then differentiated to form EBs and imaged at days 7, 10 and 14 on a confocal microscope. By day 7 dox-treated EBs are smaller and less complex than untreated EBs. Tiled images were captured at 20X and stitched using Fusion software (Andor). Scale bars represent 500µm.

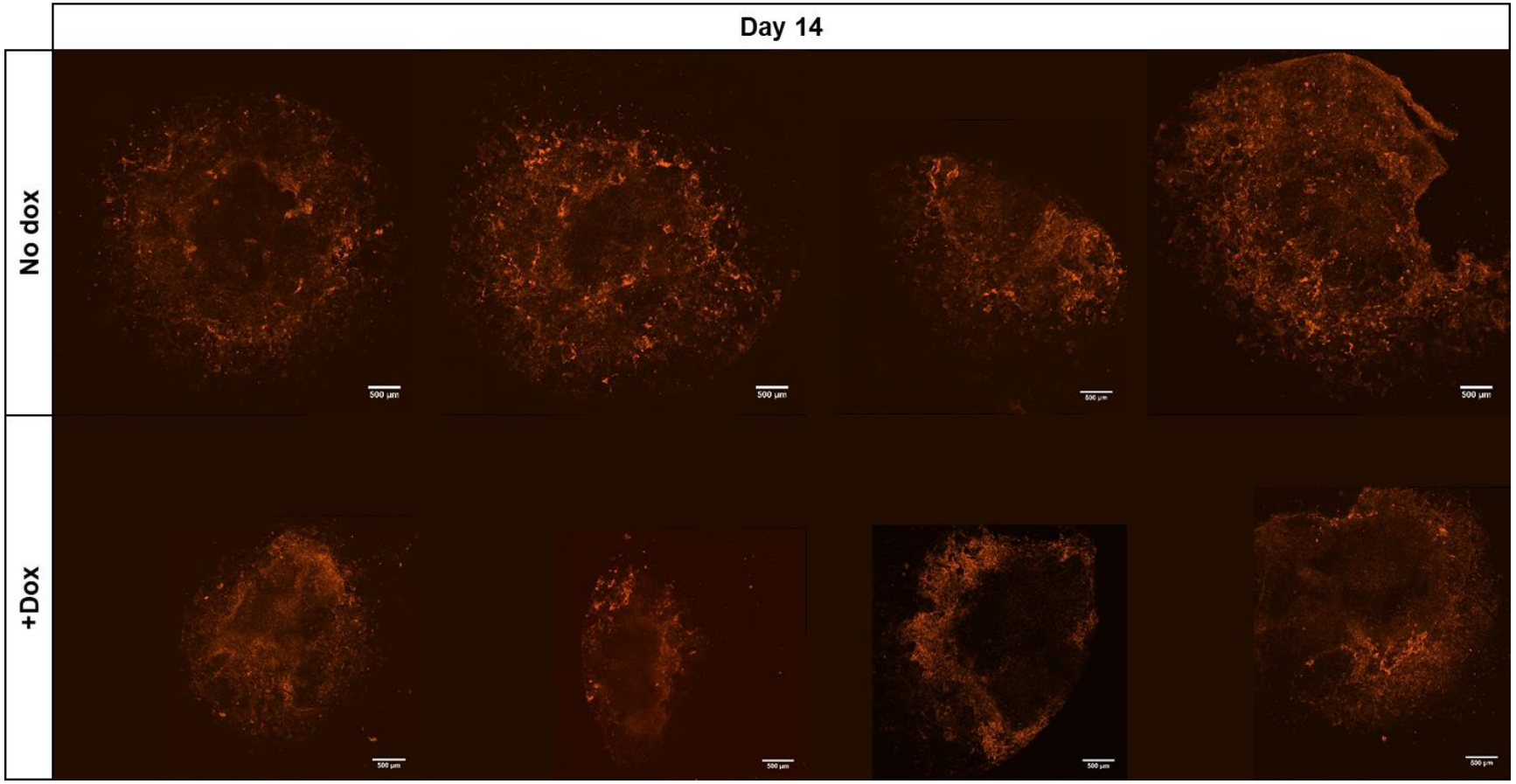


Figure 3.13, C. Lifeact-*Dnmt1*$^{tet/tet}$ EBs at day 14 of differentiation show clear differences in size. Lifeact-*Dnmt1*$^{tet/tet}$ mESCs were treated with water or 0.2µg/mL dox for 4 days then differentiated to form EBs and imaged at days 7, 10 and 14 on a confocal microscope. By day 7 dox-treated EBs are smaller and less complex than untreated EBs. Tiled images were captured at 20X and stitched using Fusion software (Andor). Scale bars represent 500µm.

As was observed in previous *Dnmt1*$^{tet/tet}$ experiments, dox-treated Lifeact-mScarlet EBs were consistently smaller than their untreated counterparts from day 7 onwards. In untreated EBs, the Lifeact signal highlighted networks of cells within the cardiomyocyte lineage, which could be seen as beating clusters of cells from ~day 10 of differentiation (Fig. 3.15). Interestingly, dox-treated EBs could still form beating clusters of cells around day 10 of development, suggesting loss of methylation does not prevent mesoderm differentiation (see supplementary videos). The clusters of cells were smaller and less connected to other clusters, reflecting the differences in size and complexity observed between EBs grown with and without *Dnmt1* expression. However, this is not consistent with the previous qPCR data which showed that the mesoderm marker gene *Snai1* decreased slightly with dox treatment (Fig. 3.9), particularly in the high dox condition.

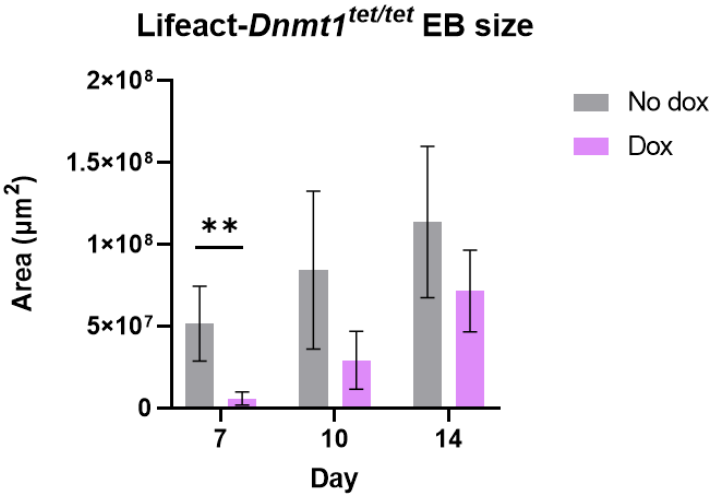


Figure 3.14. Lifeact-*Dnmt1*$^{tet/tet}$ EBs grown under hypomethylation conditions were typically much smaller than untreated EBs. EB size was determined by thresholding in Fiji to detect EB cells and determining the total EB area in µm^2. Mean EB areas per time point (n=4) were compared by two-tailed t test. Significantly different samples are highlighted (** = p<0.01).

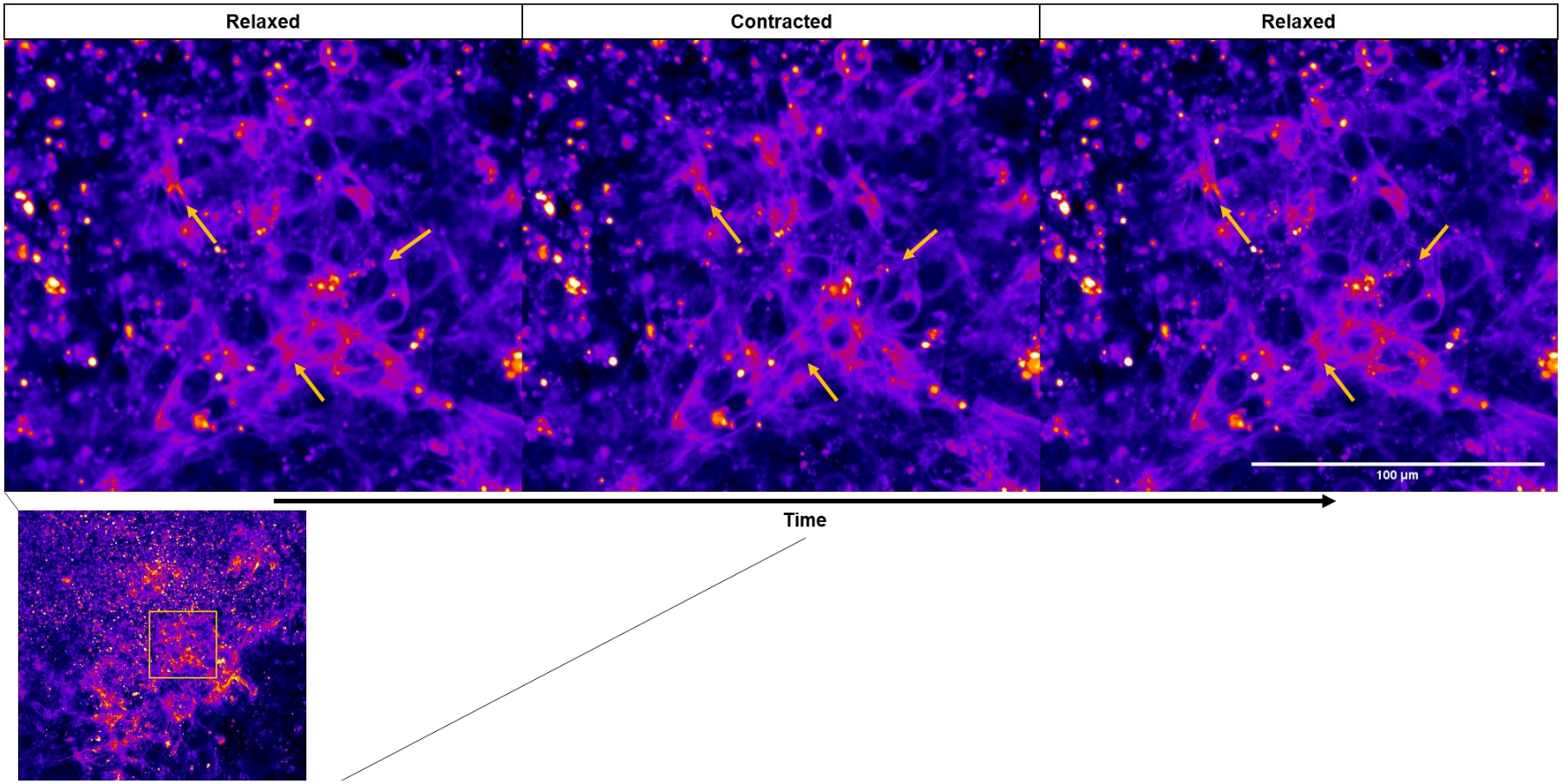


Figure 3.15. Stills from a time-lapse capture of a single beating cluster of progenitor cardiomyocyte cells of a *Dnmt1*$^{+/+}$ EB. The EB was imaged at day 10 of differentiation on a confocal microscope. Here the image has been false coloured in FIJI to highlight the elongated cells branching between beating clusters. The arrows highlight some of the contracting cells. Inset shows the area of the whole EB from which this cluster was isolated.

3.2.5 Tracking changes in DNA methylation distribution during differentiation

Thus far, the effect of loss of *Dnmt1* on EB differentiation has been investigated by Southern blot and RT-qPCR using conditional loss culture with *Dnmt1*$^{tet/tet}$ ESCs and derived EBs cultured with or without dox. As an orthogonal approach I also obtained an ES cell line from the Riken cell bank in which there is a fluorescent reporter of 5mC, named EGFP-MBD-nls (AES0134). Under normal conditions 5mC-rich regions are visible as EGFP-positive foci in these cells (Kobayakawa *et al*., 2007) (Fig. 3.16). I used CRISPR/Cas9 genome editing to knock out *Dnmt1* in these cells. Figure 3.17 shows a schematic of the *Dnmt1* gene and highlights the positions of the guide RNA (gRNA) sequences and PCR primers. The gRNAs were designed with help from research assistant Dr David Hay. Once the knockout was validated by PCR and sequencing (see Materials and Methods), I generated EBs from these cell lines and imaged them at days 4, 7, 10 and 14 of differentiation (Fig. 3.18). Almost no foci were visible in the *Dnmt1* KO MBD-EGFP EB cells (Fig. 3.19). Foci of EGFP-MBD are still visible in the *Dnmt1* knockout ES cells as the *de novo* DNMTs are highly expressed in these cells relative to differentiated cells, so the reporter can still form foci where these enzymes are active; this is consistent with what was reported by Kobayakawa *et al*.

Next, I wanted to use these cells to see if I could observe the pattern I saw previously in the *MaeII* restriction digest Southern blot; in untreated *Dnmt1*$^{tet/tet}$ EBs there was a wave of partial demethylation between days 5-10 of differentiation. I theorised that this may be indicative of the formation of a population of PGC-like or 2C-like cells, which are hypomethylated compared to differentiated cells. PGCs *in vivo* undergo waves of demethylation to enable the activation of germ-cell genes, such as *Dazl* (Maatouk *et al*., 2006; Irie *et al*., 2014). In EB models PGC-like cells have been reported to emerge at the second differentiation event (at approximately differentiation day 6; Kim *et al*., 2020) wherein cells of the EB pre-implantation epiblast-like cell population diverge to become either PGC-like cells (which form at gastrulation and become

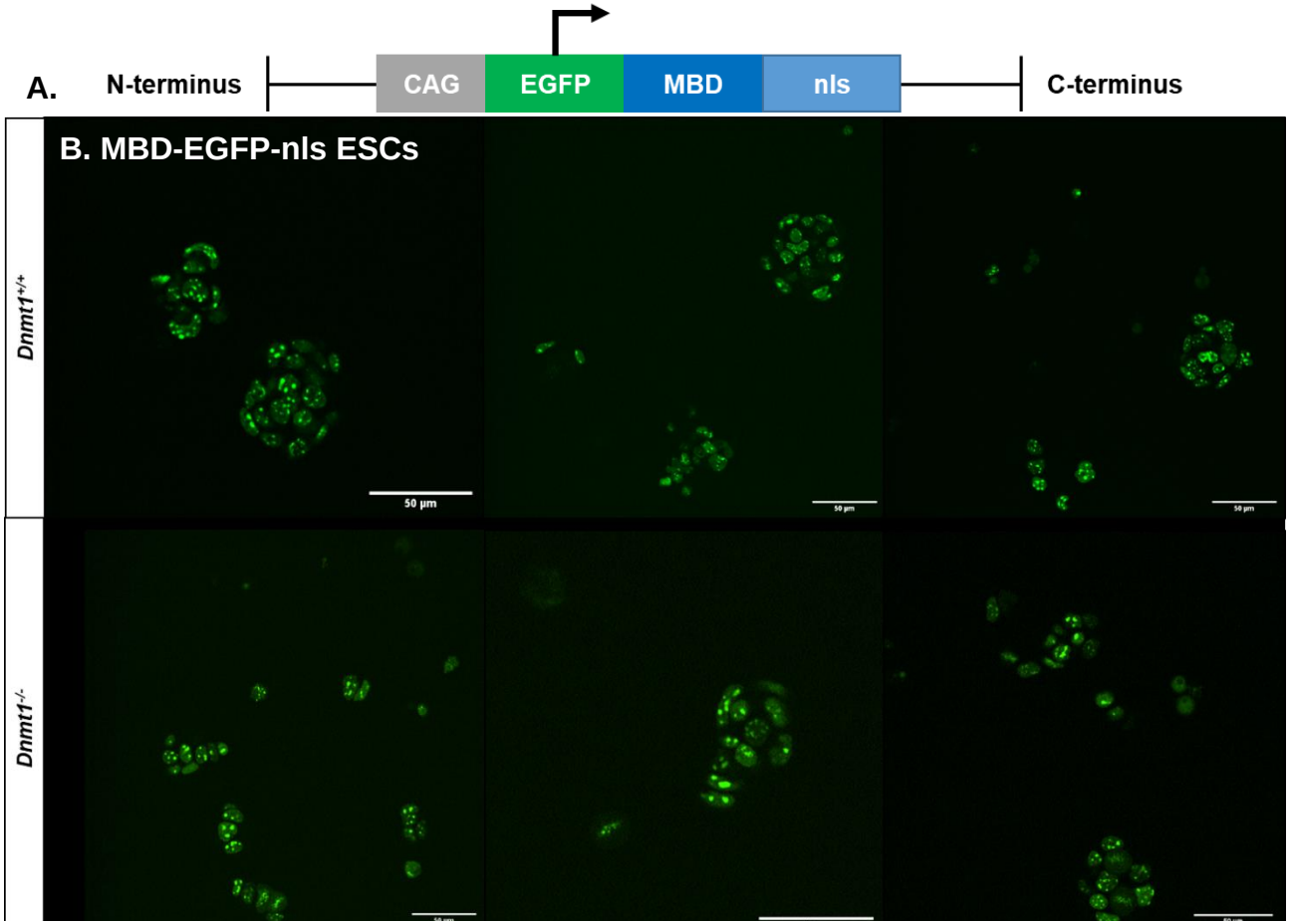


Figure 3.16. Knockout of *Dnmt1* does not affect ESC morphology. A) Schematic representation of the fluorescence-tagged methyl-binding domain (MBD) protein used to visualise DNA methylation distribution. B) DNA methylation foci are visible in EGFP-MBD-nls ESCs in which *Dnmt1* was deleted due to the activity of the *de novo* methyltransferases Dnmt3a and Dnmt3b. Images captured on a confocal microscope at 20X by Dr S. Dimova. Scale bar = 50µm.

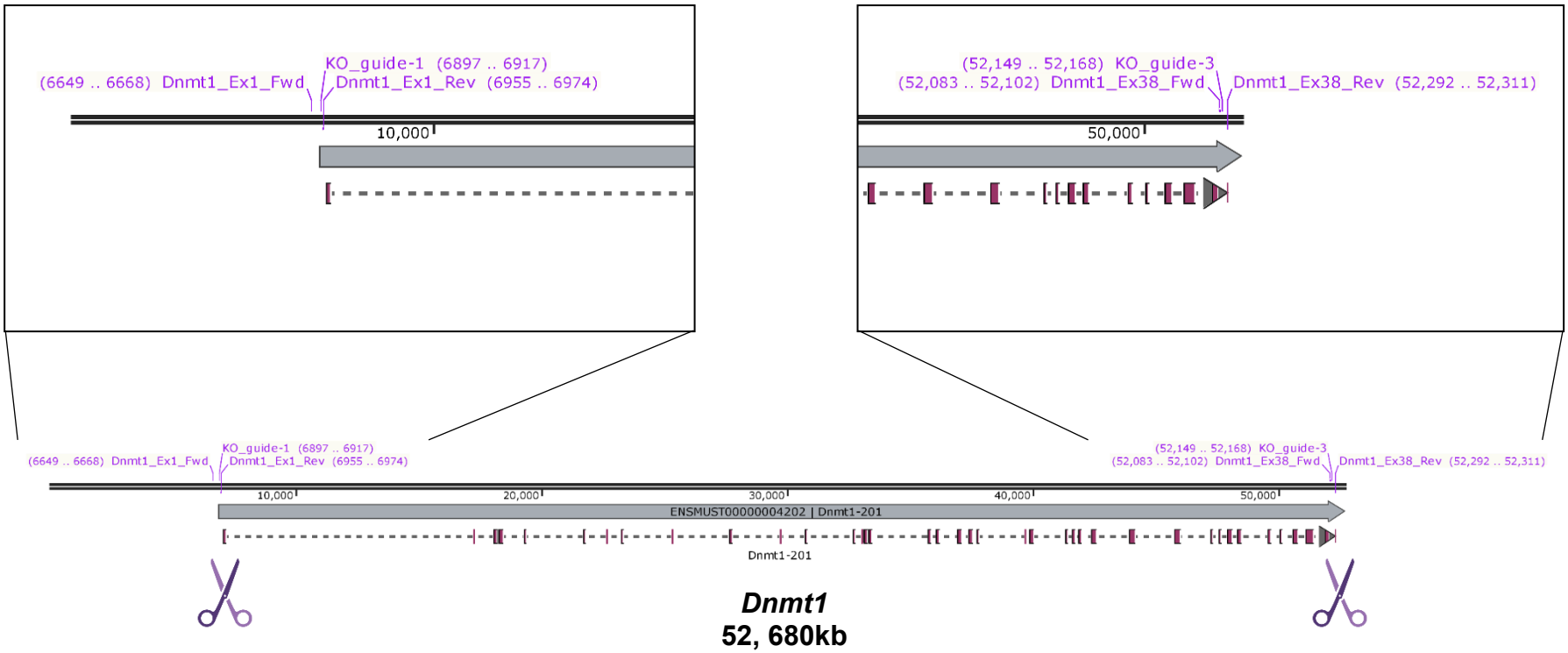


Figure 3.17. Schematic of the *Dnmt1* gene highlighting the locations of the CRISPR KO guide cut sites, and primer positions for the PCR macro- and micro-deletion screens. In total a 52kb region was removed. The schematic was drawn using SnapGene software.

gametes *in vivo*), or post-implantation epiblast-like cells which go on to form the terminally differentiated cell types of the mesoderm, endoderm and ectoderm lineages (Rossant *et al*., 2009; Kojima *et al*., 2014; Fu *et al*., 2020). As such, I hypothesised that PGC-like cells in EBs might be visible in this cell line as a sub-population of cells with relatively few foci or diffuse EGFP signal. I enlisted the help of Dr Silviya Dimova to perform foci analysis using image analysis pipelines she designed on the images I acquired.

The EGFP-MBD EBs were analysed for foci frequency. In the *Dnmt1* KO version of the line a negligible number of cells per EB had detectable foci as expected; any foci observed were probably due to the compensatory activity of the *de novo* DNMTs. The KO line was used to provide a visual reader of the successful CRISPR knockout strategy for the *Dnmt1* gene; this strategy was used in other cell lines. The foci analysis was performed primarily to try to determine if the distribution of 5mC, and therefore MBD binding, changed over time in developing EBs, and in comparison to the pluripotent ESCs. For this I used an image processing algorithm designed by Dr Silviya Dimova; the algorithm involved thresholding and segmentation to isolate single cells and their nuclei, then foci counting. The counts were then visualised as histograms. It was observed that most EB cells counted had 0 foci in. The histograms showed very similar patterns across conditions, with most cells having 0-5 foci.

As part of the foci analysis, one of the questions I hoped to address was if I could identify populations of cells within an EB which had a small number of large foci as opposed to a large number of small foci. This could be indicative of a population of hypomethylated cells where 5mC was limited to demethylation-resistant regions, i.e. cells which may be 2C-like or PGC-like as these are the cells which are hypomethylated or undergo programmed hypomethylation during development *in vivo*. This was to follow up on the observed loss of DNA methylation in untreated EBs as detected by the Southern blot earlier in the project. Unfortunately, the algorithm used was not able to categorise cells by foci size or identify cells with relatively few foci from the images provided.

The process was limited by the quality of the EB images. Due to the sizes of the largest EBs being imaged and the image processing power available, it was not possible to perform z-stacking at 40X magnification to capture the entire 3D structure of the EB and with high resolution of individual cells. The focus was placed on the single EB cells at the outer edges of the EBs at 20X magnification to try to capture as many cells in the analysis as possible at a magnification which would enable visualisation of single cells. Unfortunately, due to the compaction of cells within the densest central regions of the EBs single cells in those regions were impossible to distinguish. This meant that for every EB imaged some portion of it was un-analysable, which inevitably leads to loss of information and an incomplete picture of the patterns of foci. Also, as the EBs were imaged live and then returned to the incubator to be re-imaged at another time point it was not possible to stain the nuclei, either with DAPI which is for fixed cells or Hoechst which is known to be partially cytotoxic. Instead, the nuclei were isolated by artificially creating ‘nuclei’ using ‘Gaussian blur’ tool during processing to approximate the borders of the nuclei from the position of foci within the cells. This meant that the algorithm was often unable to tell apart cells which had clear foci but were clustered together, often identifying them as a single large cell with 100+ foci. These aberrant counts were manually removed prior to data analysis, and for consistency the number of foci was capped at 50.

This experiment was also limited by the absence of a control; I had no reference for what 5mC, and therefore MBD, distribution looks like in 2C- or PGC-like cells. I therefore looked instead for those cells which have patterns dissimilar to MBD-EGFP ESCs and those which don’t have multiple foci indicative of high DNA methylation activity. Cells with low levels of DNA methylation, I presumed, would have very few foci, possibly a small number of bright foci where 5mC was maintained at demethylation-resistant regions and therefore MBD would be concentrated. I tried to approach these problem using an alternative method, 2C::EGFP reporter cells, which I describe in the next chapter.

Finally, in addition to the foci analysis experiment, the images of the EGFP-MBD-nls EBs were used to again quantify differences in size between EBs with and without *Dnmt1*. This analysis was previously done with the Lifeact-mScarlet-expressing *Dnmt1*$^{tet/tet}$ EBs; in this orthogonal analysis the EBs are derived from ESCs of an independent mouse background (E14 vs R1) in which the endogenous *Dnmt1* locus has been removed by CRISPR, and the resulting EBs were imaged in greater numbers and from an earlier time point (Day 4). With the help of Dr Murphy at the IGC Advanced Imaging Resource, the size quantification by thresholding method was refined to semi-automate the analysis. Figure 3.18 gives an overview of the differences in size between the *Dnmt1* KO and wild-type EGFP-MBD EBs. The panel in figure 3.19 zooms in to view single cells, highlighting the difference in foci frequency. Finally, figure 3.20 shows the results of the size quantitation. Consistent with the Lifeact EB data, *Dnmt1* KO EGFP-MBD-nls EBs were significantly smaller than *Dnmt1*$^{+/+}$ EBs (Fig. 3.20).

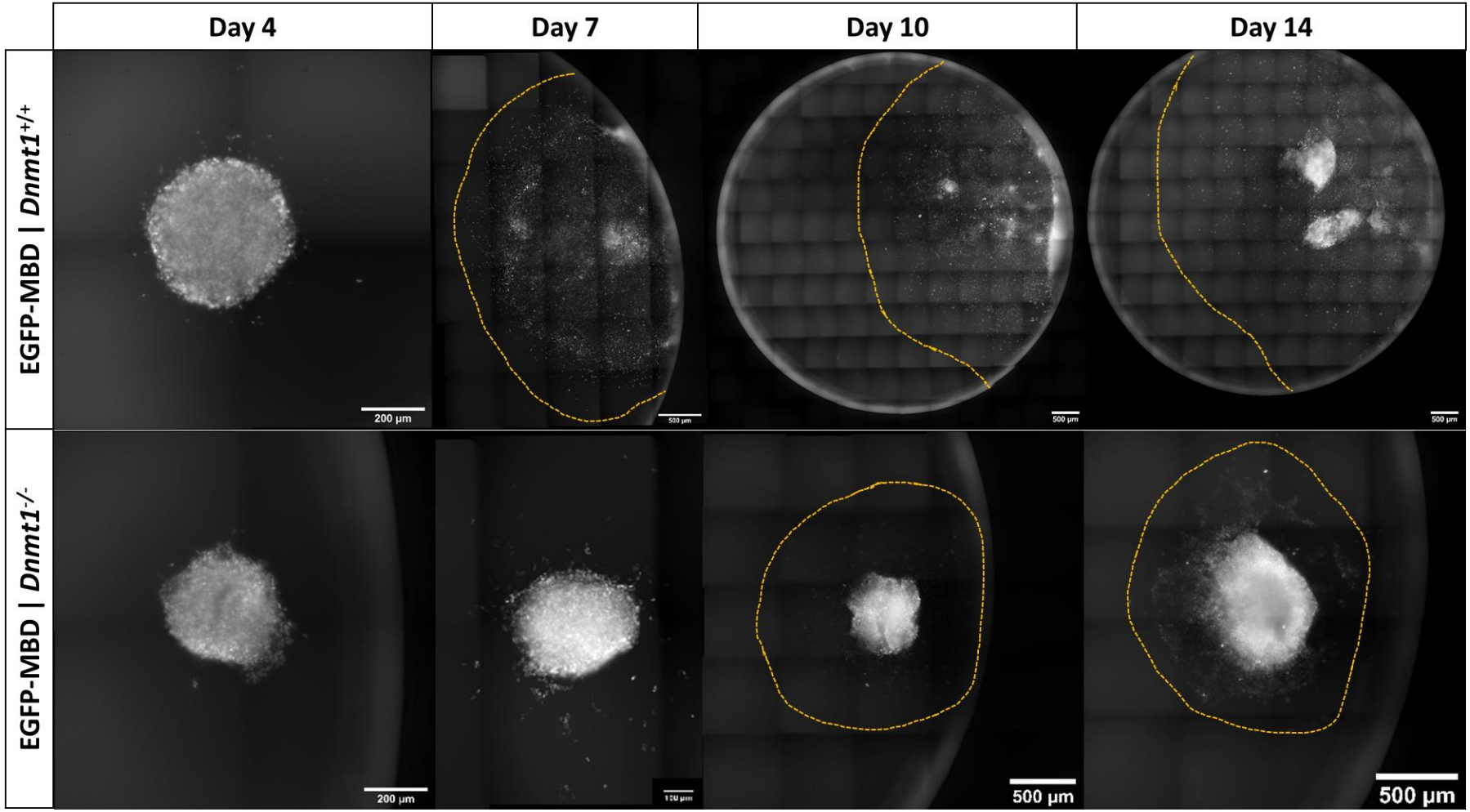


Figure 3.18. Legend on next page…

Figure 3.18 on previous page. CRISPR knockout of *Dnmt1* in a cell line with a fluorescent live reporter of 5mC localisation resulted in loss of foci and reduced EB growth. ES cells expressing an EGFP-tagged methyl-CpG binding domain (EGFP-MBD) protein were modified to knockout *Dnmt1* and differentiated to form EBs. EB development over time was captured by imaging on an inverted epifluorescence microscope at 20X and tiling images to capture the whole EB. As with conditional *Dnmt1* expression cell lines, EBs formed from *Dnmt1*$^{-/-}$ cells showed restricted growth patterns. Removal of *Dnmt1* can be seen as a loss of EGFP-MBD foci.

Figure 3.19 *on next page.* An EGFP-5mC reporter ES cell line showed that *Dnmt1* KO results in smaller EBs. ESCs expressing EGFP-MBD were modified to KO *Dnmt1* using CRISPR and differentiated to form EBs. The loss of MBD foci in the KO line EBs showed the deletion of *Dnmt1* and consequently most 5mC. Similar to conditional *Dnmt1* KO line EBs, MBD *Dnmt1* KO EBs were smaller than their wild-type counterparts but could still form beating cardiomyocyte clusters indicative of mesoderm differentiation.

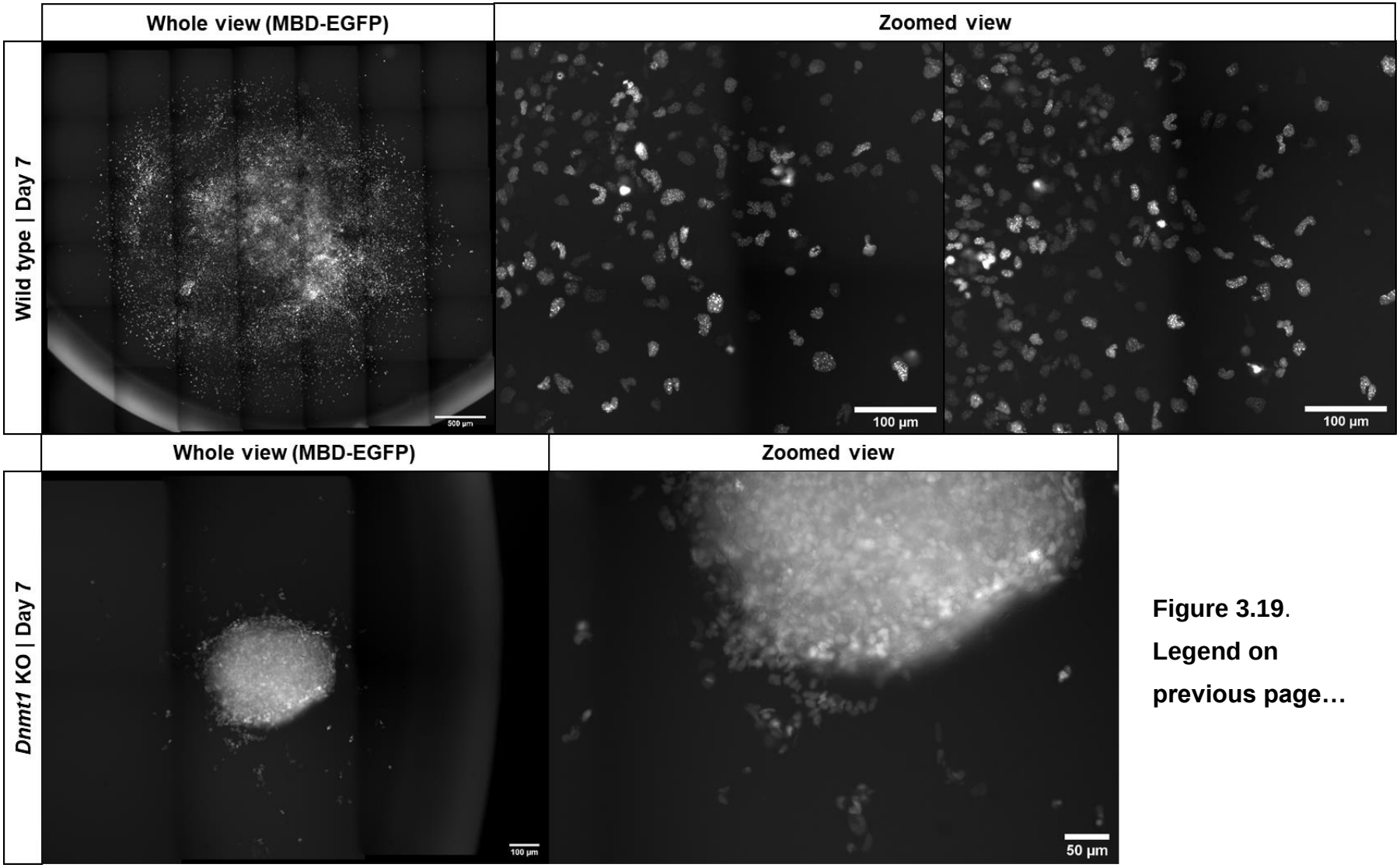


Figure 3.19. **Legend on previous page…**

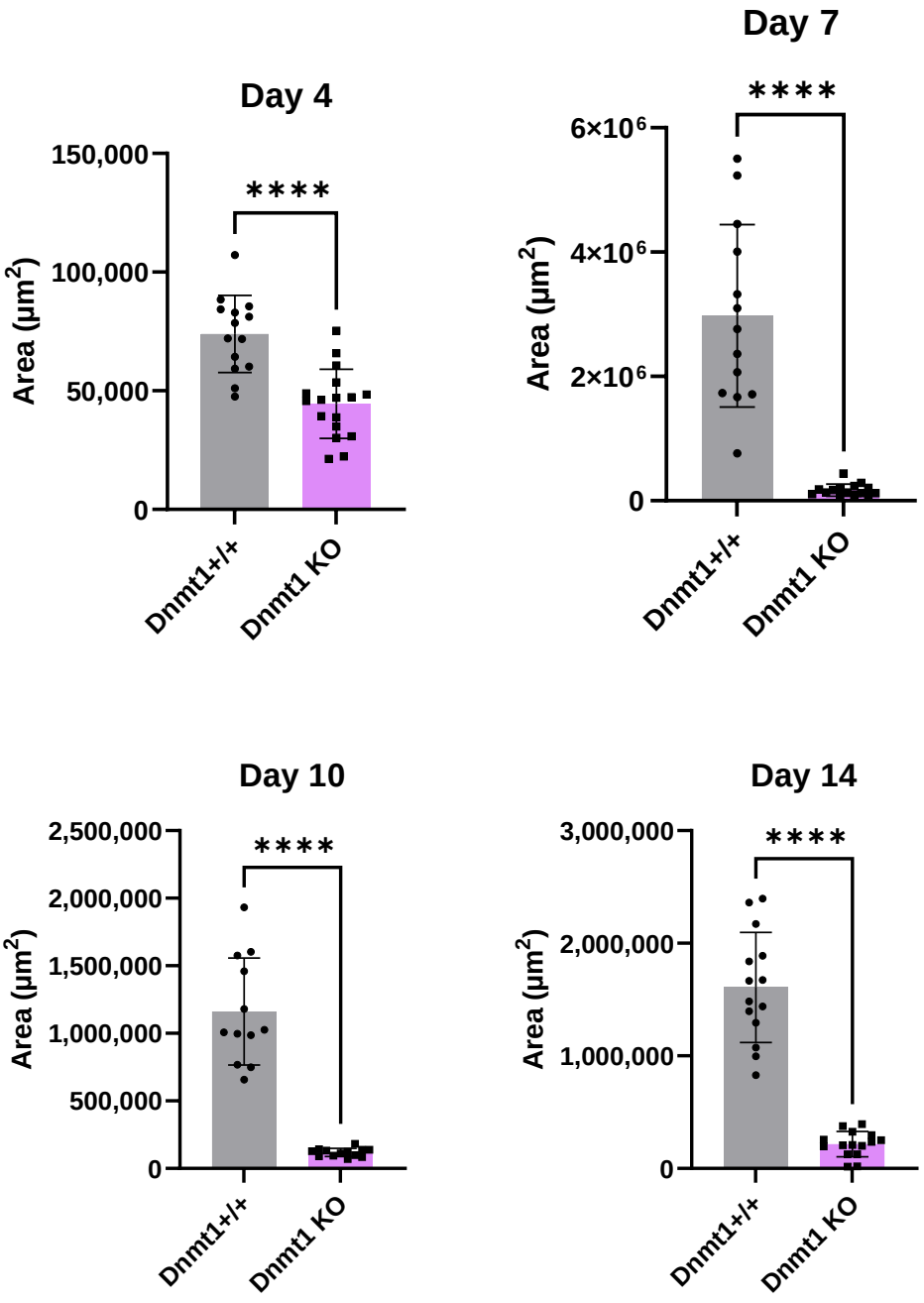


Figure 3.20. CRISPR knockout of *Dnmt1* impairs EB growth. Similarly to when *Dnmt1* is removed conditionally in *Dnmt1*$^{tet/tet}$ ESC derived EBs, CRISPR knockout of *Dnmt1* in E14 EGFP-MBD-nls-derived EBs resulted in significantly smaller EBs relative to wild-type EBs, and this difference was observed from an earlier time point. Mean EB sizes compared by unpaired, two-tailed t test (**** = p<0.0001; 12-17 EBs measured per sample).

3.3 Discussion

The mechanisms of how DNA methylation is deposited and propagated across cell division are well understood but fully delineating its role during differentiation, which is arguably when it becomes essential for tissue specification, is ongoing. In particular, it is unknown at what exact time point or period during development the mark becomes essential for survival. In this chapter I aimed to find and optimise an ESC differentiation system that would allow me to model early embryonic development *in vitro* and investigate the role of DNA methylation during cellular differentiation.

I started by converting *Dnmt1*$^{tet/tet}$ ESCs with or without maintenance methyltransferase *Dnmt1* into EBs. In this cell line the *Dnmt1* alleles are dox-regulatable; the drug binds to the trans-activator protein and prevents it from driving the gene promoter, resulting in reduced *Dnmt1* mRNA levels (Fig. 3.1). This in turn prevents maintenance of 5mC patterns across cell division leading to global genomic hypomethylation, to at least 5% of normal levels (S. Dimova, unpublished thesis). I tested two different concentrations of dox for maintaining *Dnmt1* inhibition during EB culture; at day 3 of differentiation, it was frequently observed that the high dox dose cells failed to aggregate and form EBs, and those which did typically failed to adhere or grow when transferred from suspension to adherent culture. I concluded that the high dose of dox was not practical to use. Low dox treated EBs fared better and usually formed aggregates of a similar size to untreated cells, but following adhesion they grew more slowly and were a fraction of the size of their untreated counterparts at all time points measured. Untreated EBs with normal methylation rapidly expanded in size and cell type complexity, as seen in brightfield images (Fig. 3.3). Large, beating patches of cardiomyocyte cells were typically seen from day 8-10 onwards. These cells were also observed in some dox treated EBs showing advanced mesoderm differentiation could still occur in the absence of *Dnmt1*. However, these cell clusters were smaller and did not feature the elongated cell branches which formed between beating clusters in the untreated EBs (Fig. 3.15). It was unclear at this point if the differences in size were due to an increase in cell

cycle time due to loss of Dnmt1, or if the loss induced apoptosis which reduced the total numbers of cells in each EB. I tried to address this using a TUNEL assay, which will be discussed later.

Wild-type ES cells were previously used to test for non-specific or cytotoxic side-effects of dox treatment, and none were found. The results of experiments performed later in this thesis which removed *Dnmt1* by alternative means, such as by genetic knock-out, support the view that dox does not have non-specific or cytotoxic effects in EBs either; similar phenotypes were found between dox-treated and *Dnmt1* KO EBs suggesting the effects observed were from the relative absence of Dnmt1, not the presence of dox. It was therefore deemed unlikely that the dox treatment could be having other effects on the ES cells and EBs which could have affected the results presented here.

I confirmed by Southern blot that the hypomethylation induced in the *Dnmt1*$^{tet/tet}$ ES cells was sustained in the low dox treated EBs throughout the 14 days differentiation period. An unexpected observation came from this result; in the untreated EB DNA there appeared to be a temporary decrease in DNA methylation which occurred between days 5-10. This was unexpected as generally differentiation is correlated with an increase in global DNA methylation deposition and expression of Dnmts (Gökbuget and Blelloch, 2019), meaning I expected to see progressive patterns of high molecular weight bands indicating the lack of digestion of satellite DNA by *MaeII* as EB differentiation progressed. I speculate that the pattern observed could indicate a brief window of reprogramming in a subset of cells wherein DNA methylation is temporarily removed. This occurs *in vivo* during mouse PGC formation at gastrulation (~E6.5); following stimulation by BMP signalling PGCs emerge from the embryonic epiblast population and undergo epigenetic reprogramming during their migration to the developing genital ridge, including global erasure of DNA methylation (Irie *et al.*, 2014). This first wave of demethylation removes the somatic patterns of DNA methylation established during epiblast differentiation in the precursor cell and enable re-

activation of pluripotency genes required for PGC formation (Messerschmidt *et al*., 2014).

The Southern digest pattern could indicate the presence of PGC-like cells which, in the absence of extracellular BMP4 signalling, don't commit to the germ cell lineage but return to a somatic/epiblast-like state or consequently arrest at this stage and then undergo apoptosis; either route could cause the 5mC digest pattern to return to pre-demethylation levels by day 10. It has been shown that PGCs can be derived *in vitro* in the absence of BMP signalling. Magnusdottir *et al*. (2014) showed that forced overexpression of *Prdm1* (*Blimp*), *Prdm14* and *Tcfap2c* (*AP2γ*) in Epi-LCs can induce PGC-like cell formation *in vitro*, independently of the normal cytokine signalling. This could explain how PGC-like cells could form during EB culture which lacks the extra-cellular cytokine signalling found *in vivo*. If the PGC identity could be confirmed this would provide evidence that EBs generate an environment permissive of PGC formation and therefore can be used to study their properties *in vitro*. In particular, using these conditional *Dnmt1* EBs it would then be interesting to investigate how inducing the loss of *Dnmt1* impacts their formation. If PGCs naturally undergo hypomethylation as part of their development process *in vivo*, would inducing hypomethylation enrich the EB population for PGC-like cells? The phase of hypomethylation was observed between days 5-10; does the timing of loss of 5mC affect PGC emergence? The conditional *Dnmt1* expression line could be used to identify a window during EB differentiation in which DNA methylation has a role in PGC formation. Maatouk and Resnick (2003) investigated the impact of demethylating agent 5-azacytidine (5-aza) on PGC development and found that application of the drug increased the rate with which the post-migratory PGC marker Gcna1 was expressed by pre-migratory PGCs in culture, thus loss of 5mC was found to enhance germ cell differentiation in their system. However, 5-aza is not a specific inhibitor of *Dnmt1* and is toxic to some cell types, therefore it is possible other cell signalling systems contributed to their observations. Here, the *Dnmt1*$^{tet/tet}$ cells would provide a more refined system to study the effects of hypomethylation on PGC formation. Also, their

experiments targeted the later stages of PGC-development, i.e. when pre-migratory PGCs differentiate to become post-migratory PGCs similar to those that occupy the gonads. This contrasts with our system which encompasses multiple differentiation steps starting from the exit of pluripotency.

Further work is needed to confirm if PGC-like cell formation is contributing to the observed demethylation phase in our data. In chapter 4 I investigate changes in gene expression patterns of EBs with and without *Dnmt1* by RNA-seq; this includes checking for differential expression of PGC-associated genes as a consequence of hypomethylation.

Up to this point I had seen that culturing ESCs under EB conditions produced rapidly expanding populations of differentiating cells, with clear differentiation towards the mesoderm lineage. I also saw that loss of *Dnmt1* negatively impacted this phenotype, although mesoderm differentiation could still occur. So next, to confirm the differentiation on a molecular level and to check for loss of 5mC, I investigated changes in gene expression by qPCR. Under normal EB conditions lineage-specific marker genes were up-regulated and pluripotency genes down-regulated compared to pluripotent ES cells, showing that germ layer formation was occurring as expected with the culture conditions applied.

Next, I found that there was co-expression of pluripotency and differentiation marker genes within the dox-treated EBs. During normal development, exit from pluripotency leads to a decrease in expression of pluripotency genes. Here, it appeared that the absence of *Dnmt1* in the developing EBs prevented the repression of *Oct4*, though it is unclear whether this is a direct or indirect mechanism. Is the timely repression of *Oct4* by DNA methylation at the onset of differentiation key for EB survival, and therefore also perhaps the process of early embryonic development? If so, this could suggest that although DNA methylation is dispensable for ES cell viability it becomes essential at the exit of pluripotency when pluripotency networks need to be repressed as differentiation pathways become activated. This co-expression could result in conflicting signalling pathways which ultimately lead to developmental arrest followed by induction of apoptosis.

This is consistent with the observed phenotype noted earlier. Low levels of some of the differentiation genes were detected in the ESCs; this could be due to the serum/LIF culture used when growing these cells. This culture produces heterogeneous populations of ES cells, some of which are in a 'primed' state of pluripotency with slightly higher methylation and expression of some differentiation genes ready for epiblast formation (Weinberger *et al*., 2016).

Dnmt1 expression was not quantified directly in this assay as the locus is modified in these cells, meaning new primers would have to be designed and validated for the qPCR specifically. Instead *Dazl*, a gene which is specifically repressed by DNA methylation and is silenced in ESCs, was used as detecting these transcripts confirms the loss of Dnmt1 activity and loss of 5mC at this locus, as well as being an indirect indicator of loss of *Dnmt1* transcription.

This experiment was done with EBs collected at day 14 of differentiation. It would be good to follow up with an earlier time point to see if the dynamics of changes in gene expression can be quantified for these markers.

Following up on the observed differences in size between EBs with and without *Dnmt1*, I collaborated with PhD students Katie Pickup and Chetan Srinath to perform a TUNEL assay to determine if p53-apoptosis could explain the discrepancy. This assay measures the abundance of DNA double-strand breaks in cells which occur naturally as part of the apoptosis mechanism wherein the dying cell compacts its DNA and starts to break it down (Gavrieli *et al*., 1992). The TUNEL system works similarly to the Annexin V assay, which measures the abundance of phosphatidylserine cell surface receptors which are produced during the early stages of apoptosis (Koopman *et al*., 1994). Both strategies are considered sensitive methods for the detection of apoptosis, although the TUNEL assay is generally used for fixed cells (Kylarová *et al*., 2002). Analysis of EBs collected at day 7 of differentiation initially suggested apoptotic cells occurred more frequently in the dox-treated conditions, but the pervasiveness of cell debris and variability

between samples led us to conclude that the test was inconclusive. Combining this analysis with the observations that nucleosomal ladders, another indicator of apoptosis, were not present in raw EB DNA in the presence or absence of dox throughout culture (Fig 3.11), nor were there obvious differences in the abundance of pyknotic nuclei (not shown), I conclude that apoptosis is unlikely to be a significant contributing factor to the difference in size observed between wild-type and hypomethylated EBs.

Jackson-Grusby *et al.* (2001) showed that in somatic MEF cells Cre-Lox deletion of *Dnmt1* resulted in p53-dependent apoptosis, as determined using a TUNEL assay. They found that these cells had a lower proliferative capacity compared to wild-type cells, dividing every 72 hours instead of every 24, and they also observed an increase in expression of cell growth CDK inhibitors. They suggest that loss of *Dnmt1* may trigger apoptosis in somatic cells by one of several mechanisms: 1) by de-repressing genes, such as *c-myc*, which directly activate p53; 2) by activating genes which indirectly activate p53 via other apoptosis sensor proteins; 3) by destabilising the DNA replication complex and stalling the cell cycle. Although ESC proliferation is not impacted by loss of *Dnmt1*, the exit from pluripotency triggered by removal of LIF in EB culture would quickly leave hypomethylated cells vulnerable to the above mechanisms.

Importantly, they showed using a p53 double-mutant line that removal of p53 activity reversed the apoptotic effects observed upon hypomethylation in their Cre-Lox *Dnmt1* MEFs. However, in their mutant only two exons of *Dnmt1* were removed to prevent protein localisation or catalytic activity. It would be interesting to see if the same phenomenon is observed in our conditional *Dnmt1* EBs in which *Dnmt1* expression is completely suppressed; would transfection with an siRNA targeting p53 at onset of differentiation (EB day 0) reduce the differences in size observed between dox treated and untreated EBs? If so, it would suggest a direct link between hypomethylation and apoptosis in differentiating cells primed towards multiple lineages. If not, it would suggest a greater role is attributable to other apoptotic factors, such as loss of repression of repeat elements leading to increased transposition

and consequent chromosomal instability. Another factor could be induced senescence due to a stalled cell cycle from loss of stability of the replisome at replication forks (Greenberg *et al*., 2019).

In conclusion, the Jackson-Grusby study found a similar effect of loss of *Dnmt1* in a mouse embryo derived somatic cell line compared to my early differentiation model, indicating slower cycling cells.

In this project a fluorescent marker of actin (Lifeact-mScarlet) was used to highlight precursor cells committed to the mesodermal lineage in live differentiating EB cells, and to test if loss of *Dnmt1* affected this differentiation pathway. What was observed was that in wild-type EBs RFP^{+} cells formed expansive networks of cells, and by day 10 of differentiation many of these were capable of organising into contracting patches of cardiomyocytes (Fig 3.13 A-C, 3.15). Interestingly, EBs which had been maintained under hypomethylated conditions without *Dnmt1* were still able to form beating clusters of cells from around the same time point. Just as these EBs were smaller than the untreated EBs, so the beating clusters of cells were smaller and less connected. These observations suggest DNA methylation is not required for mesoderm differentiation to occur. It remains to be seen if this is also true for ectodermal and endodermal differentiation, which will be investigated later by staining for markers of these germ layers in fixed EBs.

The exact proportion of EBs which developed beating cardiomyocytes is unknown; the contraction periods may have been temporary, and without 24 hour monitoring of EB growth it is difficult to determine if, when, and how many beating clusters occurred per EB per condition. However, broadly, more beating clusters were observed where *Dnmt1* was present.

To date, very little research has gone into understanding the role of DNMTs in mouse embryonic cardiac development. However, two papers from Wendler lab sought to address this; in their most recent paper Fang *et al.* (2021) extracted primary cardiomyocytes from CD-1 mouse embryos at E13.5 and knocked down *Dnmt1* expression using siRNA. They chose this time point as during this period of heart development the cells are proliferating and differentiating rapidly, and therefore undergoing epigenetic

reprogramming. Their method enabled them to remove between 75-92% of Dnmt1 within 72 hours of transfection. Although they saw an increase in the incidence of apoptotic cells, they observed no change in the percentage of cardiomyocytes in culture; this is similar to the findings here which showed beating cells formed under both normal and hypomethylation conditions. However, they did see changes in the size of individual cardiomyocyte (α-actinin+) cells, which became larger after treatment. Using microelectrode array technology, they also were able to measure the effects of loss of *Dnmt1* on cardiomyocyte contractility showing that treatment effectively slowed the beating rate, though this gradually returned to normal with time as the siRNA was lost. Overall, they concluded that *Dnmt1* has a significant role in cardiomyocyte formation and function. However, in a previous study they carried out similar experiments with the focus on the *de novo* Dnmts, *Dnm3a* and *Dnmt3b*. That data suggested that it is actually *Dnmt3a* whose disruption has the greatest effect on embryonic cardiomyocytes, causing loss of sarcomere formation (Fang *et al*., 2016), even though *Dnmt1* is normally the most highly expressed of the three at this stage. It should also be noted that in their earlier paper siRNA knockdown of *Dnmt1* did not reduce cardiomyocyte beating frequency; only *Dnmt3a* knockdown did.

These results are consistent with those found by Gilsbach *et al.* (2014) who sequenced the methylomes of cardiomyocytes from embryonic development until postnatal maturation stage. They also concluded that it is the activity of the *de novo* DNA methyltransferases that are more important for cardiomyocyte formation. In addition, they suggested that the other epigenetic repressor system of Polycomb complexes and their marks, particularly H3K27me3, may also be involved in regulating this differentiation pathway. Overall, these results suggest that although DNA methylation is important for differentiation, the sustained activity of the *de novo* methyltransferases may have been sufficient to enable mesoderm differentiation to occur in my EBs. Although the genome was globally hypomethylated due to suppression of *Dnmt1*, it may be that *Dnmt3a* was

capable of establishing sufficient 5mC to enable mesoderm programs to occur, perhaps by repressing upstream inhibitors of this pathway.

Next, I leveraged a mouse ES cell line containing a live reporter of DNA methylation distribution, EGFP-MBD-nls cells. The construct used to visualise 5mC distribution was generated by Kobayakawa *et al.* (2007). They took the domain that recognises and binds methylated CpGs (MBD), and the nuclear localisation signal domain (nls) from MBD1 protein and fused them to a CAGGS-driven EGFP reporter. Their aim was to devise a strategy which would enable them to view changes in epigenetic mark distribution with nuclear structures at the same time. Here, I tried to use the cell line to compare MBD foci frequency between ESCs and EBs with and without Dnmt1. The experiment showed that the frequencies were comparable between ESCs and wild-type EBs, but very few foci were visible in the *Dnmt1* KO EBs. I used this as a measure of the effectiveness of the CRISPR knockout strategy to show that *Dnmt1* had been successfully removed and thus maintenance methylation was lost and the cells had very low levels of DNA methylation resulting in the loss of MBD foci. Some foci were retained in rare cells but this is probably due to compensatory activity of the *de novo* Dnmts. I also tried to use the cell line to see if I could find any signs of 2C-like or PGC-like cells, which I expected would have very few foci due to the low methylation found in these cell types. However, I was not able to detect such sub-populations by manual checking or by the foci analysis pipeline. The experiment was limited by the 3D and compact nature of the EBs making cell segmentation difficult to achieve. The main conclusions from the EGFP-MBD reporter EB differentiation experiments were that the MBD foci were lost in the *Dnmt1* CRISPR knockout cell-derived EBs suggesting the gene targeting was successful. Secondly, the line was used to obtained EB size quantification from a cell line derived from a different mouse line to those previously used, and the same pattern was seen; EGFP-MBD EBs in which *Dnmt1* had been removed and were therefore hypomethylated were significantly smaller than the EGFP-MBD EBs which had the wild-type *Dnmt1* alleles. The pattern was therefore re-confirmed in a differentiation experiment

in which *Dnmt1* repression was not drug-dependent but genetic, supporting my previous experiments with dox-controlled ESCs.

In future work I could try to overcome some of the caveats of these experiments by dissociating EBs at each time point, using trypsin for example, to break up the EBs into a single cell suspension and then fixing these cells to a slide in order to be able to image more of the cells and at a higher magnification. Once the cells were in a monolayer it would be more straightforward to image and analyse them. I could also experiment with directed differentiation techniques in order to obtain a reference for what 2C- and PGC-like MBD-EGFP reporter cells look like to then reanalyse the EB images.

Embryoid bodies are used as an *in vitro* model for early embryonic development because they form 3D structures composed of mixed cell types which reflect those found *in vivo*. However, the earliest time point analysed is usually day 3 or 4 of differentiation. By this time many changes in cell signalling and gene expression have occurred, differentiation from the pluripotent state has advanced significantly, and a population of multiple cell types has arisen. As an alternative protocol, I tried to differentiate ESCs to epiblast-like cells (Epi-LC). The transition from a naïve pluripotent cell of the inner cell mass to an epiblast cell (considered a primed state of pluripotency), identified by expression of *Bmp4*, *Fgf4*, *Nanog* and *Sox2* (Boroviak *et al*., 2015), is one of the first differentiation steps to occur in the developing pre-implantation embryo (Rossant *et al*., 2009). ESCs can be converted to Epi-LCs *in vitro* by culturing them on fibronectin in DMEM/Neurobasal media supplemented with FGF2, Activin A, N2 and B27, producing a homogeneous population of Epi-LCs from four days of culture. They provide a simpler model of differentiation in which all the cells have committed to the same lineage decision and progress along the differentiation pathway is limited to one ‘step’ from naïve to formative pluripotency (Chen *et al*., 2018). This makes analyses of changes in gene expression and other epigenetic or chromatin architecture as a result of genetic manipulation more straightforward. Unfortunately, it is commonly found with this and other

similar Epi-LSC derivation protocols that the cells will undergo crisis with most of the population entering apoptosis at day 3, and it is hard to sustain a healthy number of cells through this phase (sometimes multiple crisis phases). Many papers report data from day 3 converted cells only, for this reason (Hayashi *et al*., 2011; Buecker *et al*., 2014). After multiple attempts I was unable to consistently derive healthy populations of EpiLSCs as most cells died at day 3 of culture. Therefore, I decided to proceed with EB differentiation only.

In this chapter I have shown that the EB culture model can be used to investigate the impact of loss of maintenance methylation, and therefore induced hypomethylation, during differentiation *in vitro*. These results are summarised in figure 3.21. I found that pluripotency genes were no longer silenced in differentiating EBs lacking DNA methylation, but this did not impact differentiation as lineage marker genes were also co-expressed. I also observed temporary loss of 5mC in wild-type EBs, suggesting a wave of demethylation had occurred. As these were bulk cell population analyses it remains to be determined if the pluripotency and differentiation genes were being expressed in separate cells or if they truly were being co-expressed in single cells, or if the cells which underwent demethylation increased expression of *Oct4* as is seen in PGC-like cells. I aim to address this using ICC staining in a later chapter.

Altogether these experiments show that although loss of *Dnmt1* negatively impacts EB growth it does not inhibit differentiation pathways. This raises further questions: Is DNA methylation required for a specific window at the start of differentiation, or must it be maintained constantly once pluripotency has been exited? Having tested the EB differentiation model with our conditional *Dnmt1* expression ES cells, I proceeded to try to address some of these questions by investigating the observed effects on a more molecular level by quantifying changes in gene expression by RNA-seq.

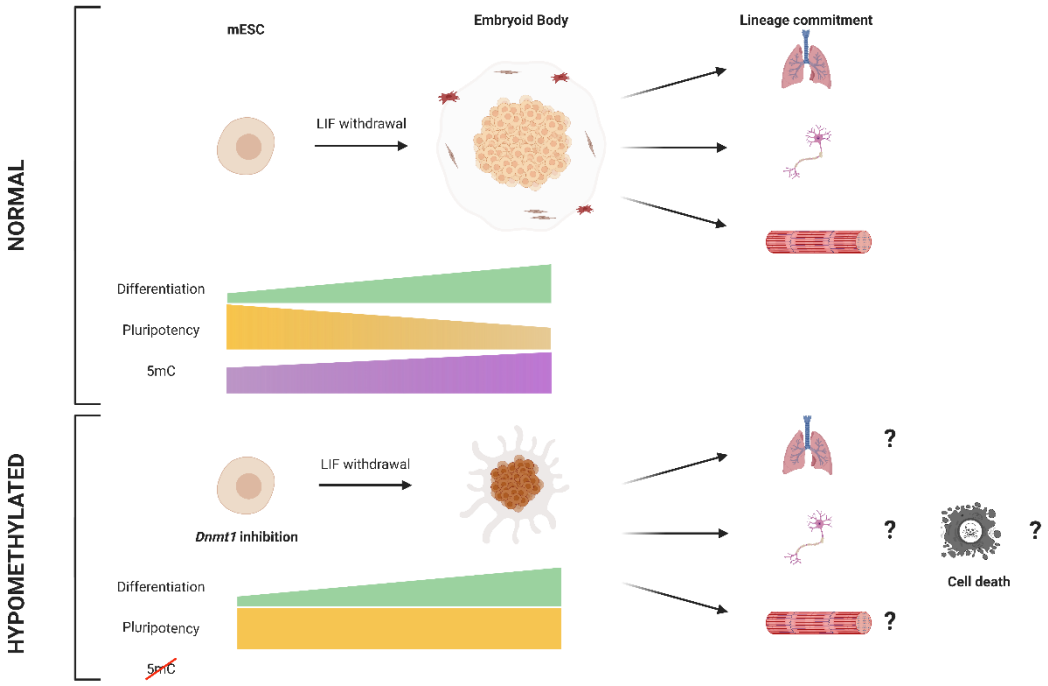


Figure 3.21. Summary of Chapter 3 results. Drug-induced inhibition of *Dnmt1* expression leads to a sustained loss of DNA methylation, aberrant sustained expression of pluripotency genes, and a small EB phenotype. The impact on specific germ layer lineage formation is unclear. Figure made using BioRender.com.

Chapter 4 | Assessing the impact of loss of *Dnmt1* on gene expression in early and late differentiating embryoid bodies by RNA-seq

4.1 Introduction

In Chapter 3 I tested an EB differentiation protocol in tandem with a conditional *Dnmt1* ES cell line as a tool to investigate the role of DNA methylation during cellular differentiation in an *in vitro* system. I found that inhibition of *Dnmt1* and consequent DNA hypomethylation did not prevent cellular differentiation, as evidenced by the emergence of beating cardiomyocytes. However, the hypomethylated EB phenotype was characterised by a significant reduction in size relative to the wild-type and loss of cell adhesion resulting in loose, fragile EB structures.

To confirm that the dox-treated EBs were indeed hypomethylated, a Southern blot was performed using EB DNA digested by methylation-sensitive enzymes. Dox-treated EB DNA was consistently digested throughout the 14 day experiment, showing 5mC was no longer being maintained in the absence of DNMT1. This was supported by qPCR analysis which showed that the methylation-sensitive gene *Dazl* was highly expressed only in dox-treated EBs.

Having confirmed the hypomethylated state of dox-treated EBs, it was hypothesised that the disparity in size could have been a consequence of increased apoptosis due to loss of 5mC-associated gene repression; a TUNEL assay was performed to address this which did not find evidence of a significant change in abundance of FITC-tagged double-strand breaks indicative of apoptotic DNA degradation.

Although beating mesodermal cardiomyocytes could be seen in dox-treated EBs, it was not clear from the phenotype alone if the other germ layer differentiation pathways were affected. To investigate changes induced by hypomethylation at the transcription level in mature EBs at day 14 of culture, qPCR was performed on select genes representative of pluripotency and differentiation networks. This analysis revealed that genes associated with the three germ layers were expressed at levels similar to the wild-type, suggesting DNA methylation is not required for germ layer lineage formation. However, it was also found that the master pluripotency factor *Oct4 (Pou5f1)*

was being aberrantly expressed in the hypomethylated EBs, suggesting a role for 5mC in the specific repression of this gene and possibly the pluripotency network in general.

The Southern blot also revealed a potential surprising regulatory role of 5mC. The analysis showed a brief period of reduced DNA methylation in wild-type EBs; this was reminiscent of a period of epigenetic reprogramming which occurs during *in vivo* pre-implantation embryonic development wherein developing primordial germ cells (PGCs) undergo a wave of global demethylation (Saitou and Yamaji, 2012; Leitch, Tang and Surani, 2013; Tan and Tee, 2019). If I could confirm that PGC-like cells were indeed occurring in wild-type EBs, would I then observe an enrichment of this cell type in EBs cultured under hypomethylation conditions? This led me to my second hypothesis: that DNA methylation although not required for germ layer differentiation, may instead be required for PGC differentiation to regulate the window of their emergence during EB development.

Overall, the findings from chapter 3 led to the hypothesis that DNA methylation, although not required to regulate germ layer differentiation networks directly, may instead be involved in regulating the pluripotency and PGC-like differentiation pathways. More specifically, it may be required to silence pluripotency genes at the onset of differentiation to enable the cell to exit pluripotency and fully commit to a somatic lineage. Also, it may be required to control the timing of PGC-like cell emergence during EB development. It may even be possible that the two pathways are interlinked; with pluripotency genes being required during the window of PGC-like cell commitment but silenced at the exit of pluripotency and somatic-cell fate commitment.

However, the gene expression analyses performed so far were limited to a small number of representative genes from each pathway, and this did not include PGC-associated genes. Also, although TUNEL assays are considered the gold standard method for apoptosis detection (Kylarová *et al.*, 2002), they can be difficult to perform accurately due to the fragile nature of

dead and dying cells. For this reason, I performed whole transcriptome RNA sequencing to obtain a detailed view of global changes in gene expression caused by differentiating EBs under hypomethylation conditions in order to start to address the above proposed hypotheses. I used the data to compare the top differentially expressed genes and to look for changes in enrichment of the previously discussed lineages and apoptotic signalling pathways.

Thus, in this chapter I aim to address the hypotheses that *Dnmt1*-dependent 5mC is required to regulate the pluripotency and PGC-like differentiation networks during EB differentiation, to clarify changes in apoptosis and lineage commitment signalling, and to investigate the reduced size phenotype using RNA sequencing.

4.2 Results

4.2.1 *Dnmt1* inhibition leads to global changes in gene expression during EB differentiation

In order to determine the impact of hypomethylation on global gene expression profiles during EB differentiation, triplicate samples of *Dnmt1*$^{tet/tet}$ EBs and EBs treated with a low (0.2µg/mL; Day 14 only) or high (2µg/mL) dose of dox were collected at days 3 and 14 of differentiation and submitted to the Edinburgh Clinical Research Facility for Illumina whole transcriptome sequencing. The raw data were processed and converted to normalised count data by Dr Philippe Gautier. Details of the data processing and analysis pipelines are given in the materials and methods section. The day 14 sample data are summarised in a Principal Component Analysis (PCA) plot in figure 4.1. Performing PCA enables the clustering of data based on co-variance which has been reduced to at least two principal components. Figure 4.1 shows the consistency of the triplicate data and clustering behaviour; all EB samples cluster away from the serum/LIF *Dnmt1*$^{tet/tet}$ ESCs along PCA 1, which accounts for 73% of the total transcriptome variance. The dox-treated EBs cluster more closely together with clear segregation from the untreated EBs along PCA 2, which explains 20% of the total variance. The day 3 data are discussed later in this chapter; the day 14 EB samples were collected and analysed first to confirm the efficacy of the dox treatment throughout the experiment.

Using the normalised count data, it was confirmed that at both time points *Dnmt1* expression was reduced to negligible expression levels in dox-treated EBs relative to wild-type EBs (Fig. 4.2a and b; Day 14 - No vs Low, one-way ANOVA, $padj<0.0001$). Despite the low dose of dox being 10-fold less concentrated than the high dose, the difference in *Dnmt1* transcript counts between the low and high dox-treated EBs were not significant (ANOVA, $p>0.99$). Expression of the other *Dnmts* also changed (Fig. 4.3a and b); at day 14 *Dnmt3b* increased significantly upon hypomethylation (No vs Low, one-way ANOVA, $padj=0.029$), along with co-factor *Dnmt3L* (No vs Low, one-way ANOVA, $padj=0.011$), perhaps as a compensatory mechanism

as a consequence of loss of *Dnmt1*. *Dnmt3a* did not change in expression significantly between EB conditions (No vs Low, one-way ANOVA, *padj*=0.97), fitting with the pattern that it is lowly expressed in differentiated cell types (Chédin, 2011). At day 3 only *Dnmt1* was expressed significantly differently between the wild-type and dox-treated EBs (unpaired two-tailed t test, *p*=0.001). This suggests that any compensation by the other DNMTs for inhibition of the maintenance methyltransferase does not begin until later than day 3 of differentiation.

The reduction in *Dnmt1* expression is also shown in figure 4.3, where the transcriptome data from EBs collected at day 14 of differentiation have been aligned to the mouse genome (mm10) using the IGV genome browser (Robinson *et al.*, 2011) (v2.9.4). As a secondary indicator of DNA methylation status, the expression level of *Dazl* was checked (Fig. 4); where the EBs have been treated with dox, the germ-cell specific, 5mC-regulated gene is aberrantly expressed.

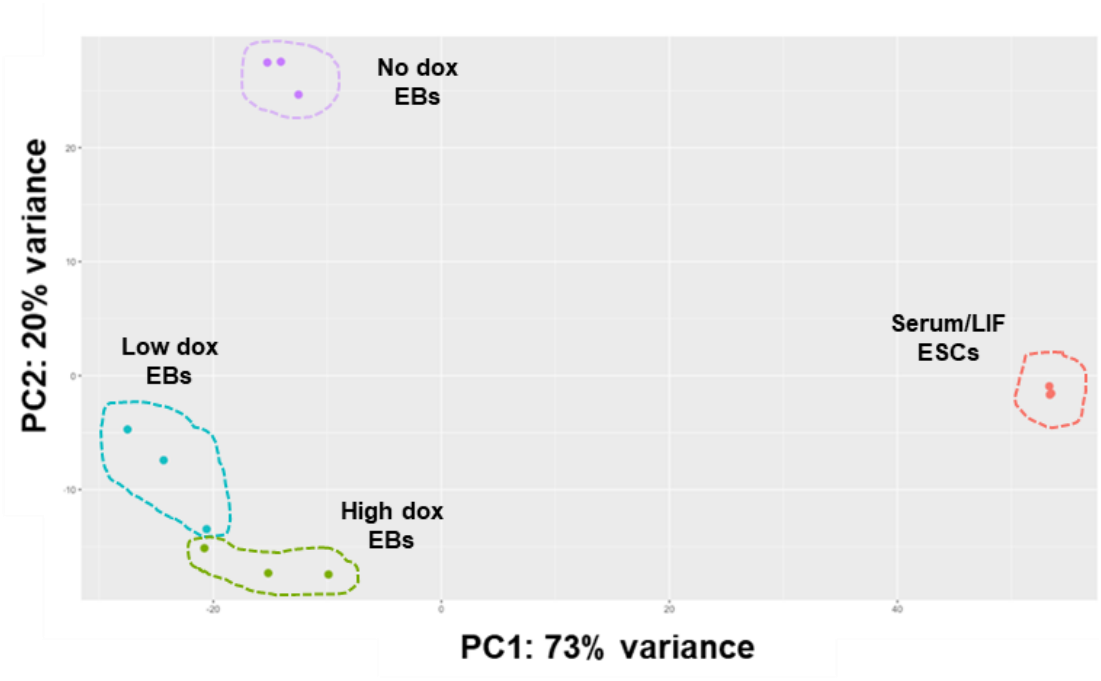


Figure 4.1. Principal component analysis (PCA) showing gene expression profiles of day 14 EBs cluster separately from ESCs, and hypomethylated EBs are distinct from wild type EBs (no dox). The profiles of the low and high dox treated EBs are highly similar, clustering closely in both PC1 and PC2.

Following sequencing of the Day 3 EBs, the earlier PCA analysis was repeated and more ESC samples incorporated. Figure 4.5 shows the day 3 and day 14 EBs, with and without dox, in a combined PCA with RNA-seq data from ground state (2i) and naïve (serum) ESCs with wild-type (J1s) or conditional *Dnmt1* alleles (*Dnmt1*$^{tet/tet}$ cells). In the top plot the samples are colour-coded by sample group name, and in the bottom plot by dox treatment. The top plot shows that the day 3 EBs have a similar PCA pattern to the day 14 EBs with clear separation between wild-type and dox-treated EBs along PC1 (70% of the total variance), although the day 3 and day 14 EBs are also clearly separated along PC2 (9% of the total variance). Interestingly, the dox-treated day 3 EBs cluster more closely with untreated ESCs, particularly the parental *Dnmt1*$^{tet/tet}$ cells, and serum/LIF cultured J1s. All the dox treated samples show consistent separation from the equivalent untreated sample along PC2, showing that hypomethylation induces consistent change in gene expression across cell types and degree of differentiation or potency (Fig. 4.5, bottom plot).

Similarly to the PCA, hierarchical clustering analysis further illustrated the similarity between low and high dox-treated EB transcriptomes at day 14. In figure 4.6 this is shown as a darker shade of blue indicating higher similarity. The plot shows that within each condition the triplicate samples show high similarity to each other. Between conditions the dox-treated samples are positioned more closely together, as shown by the clustering branches, showing they are more similar than when compared to the untreated EB samples.

Thus far I have confirmed the overall quality and consistency of the grouped EB RNA-seq samples by quantifying *Dnmt* transcript counts, 5mC status by *Dazl* expression, and obtaining confirmatory PCA and clustering analyses. I have also shown that dox-treated EBs are very different to the wild-types at the overall gene expression level. I next go on to investigate more specific changes and how select signalling pathways were affected.

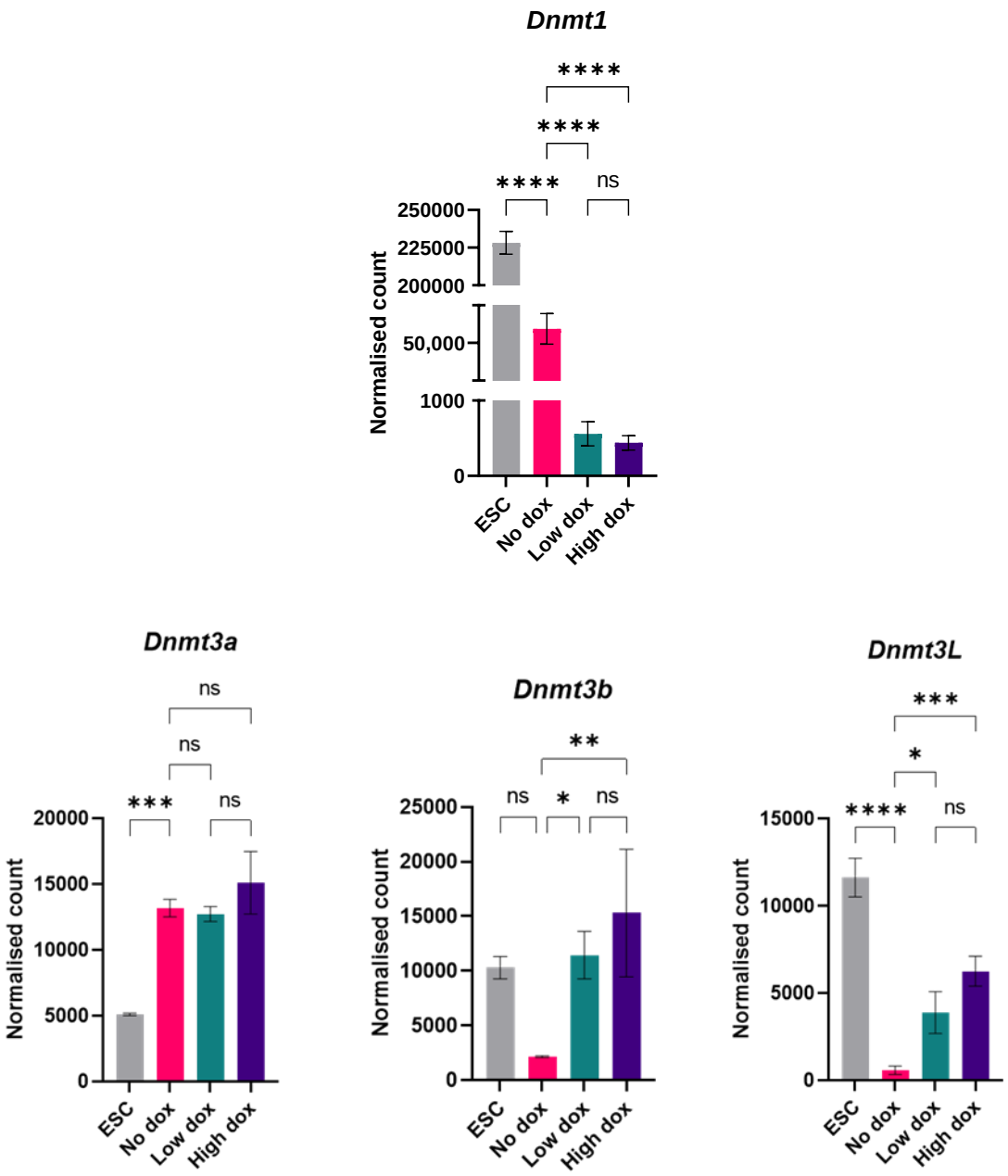


Figure 4.2, A.. Low levels of *Dnmt1* expression were maintained in dox-treated *Dnmt1*$^{tet/tet}$ EBs at day 14 of differentiation. The difference in *Dnmt* expression between the dox doses used was not significant. Expression of *de novo* methyltransferase *Dnmt3b* and catalytically inactive co-factor *Dnmt3L* were upregulated under hypomethylation conditions, but *Dnmt3a* did not change significantly in EBs. Levels of *Dnmt* expression varied significantly between pluripotent ESCs and differentiated EBs. Standard deviations of three technical replicates are shown. Normalised counts compared to the No dox EB counts by one-way ANOVA (ns = not significant; * = *padj*<0.05; ** = *padj* <0.005; *** = *padj* <0.0005; **** = *padj* < 0.0001).

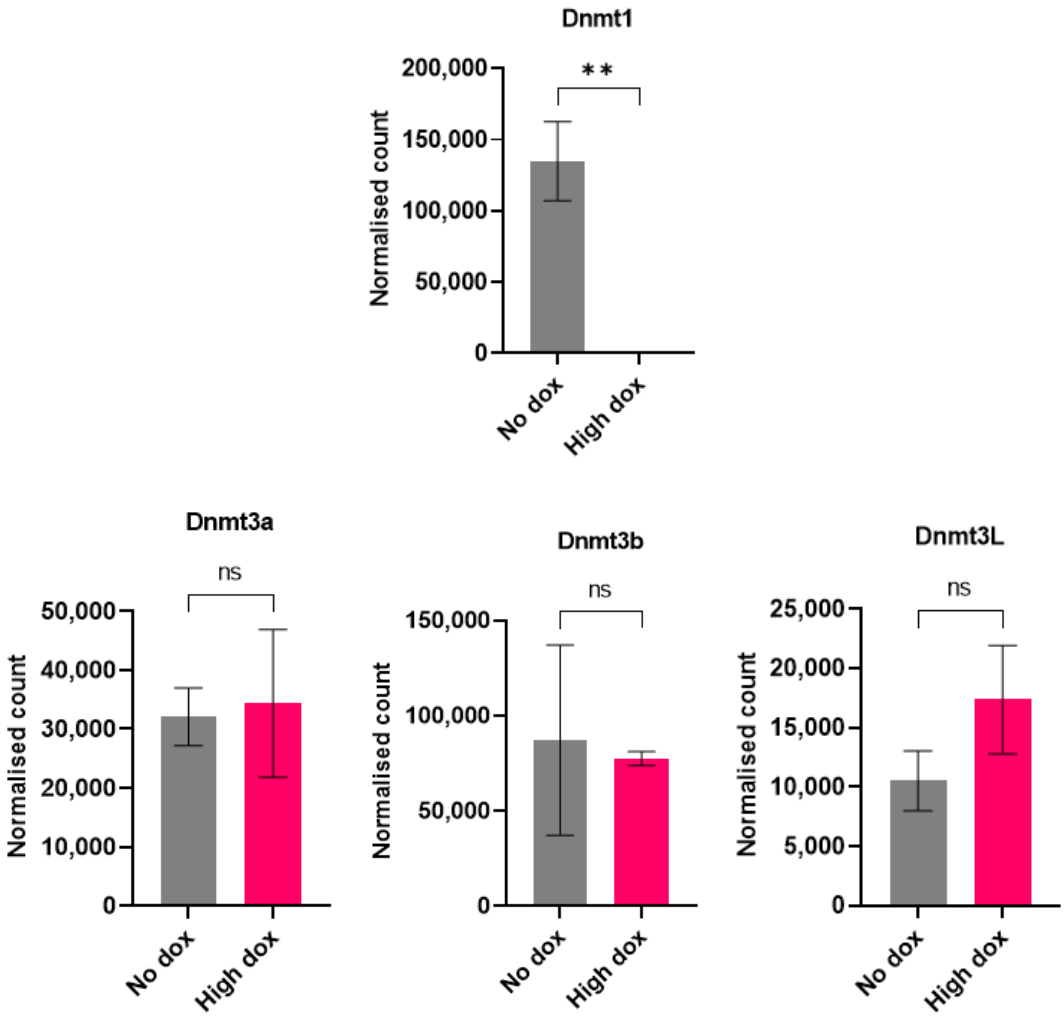


Figure 4.2, B. Dox treatment effectively maintained *Dnmt1* inhibition in *Dnmt1*$^{tet/tet}$ ESCs during 3 days of EB suspension culture. ESCs were suspended in EB media with or without dox (High dox = 2μg/mL doxycycline) for 72 hours to form the EB aggregates before being moved to adherent culture with fresh culture media. Expression of other *Dnmt*s were not significantly different between treated and untreated EBs. Normalised transcript counts per gene were compared by two-tailed unpaired t-test (ns = not significant; ** = p<0.01).

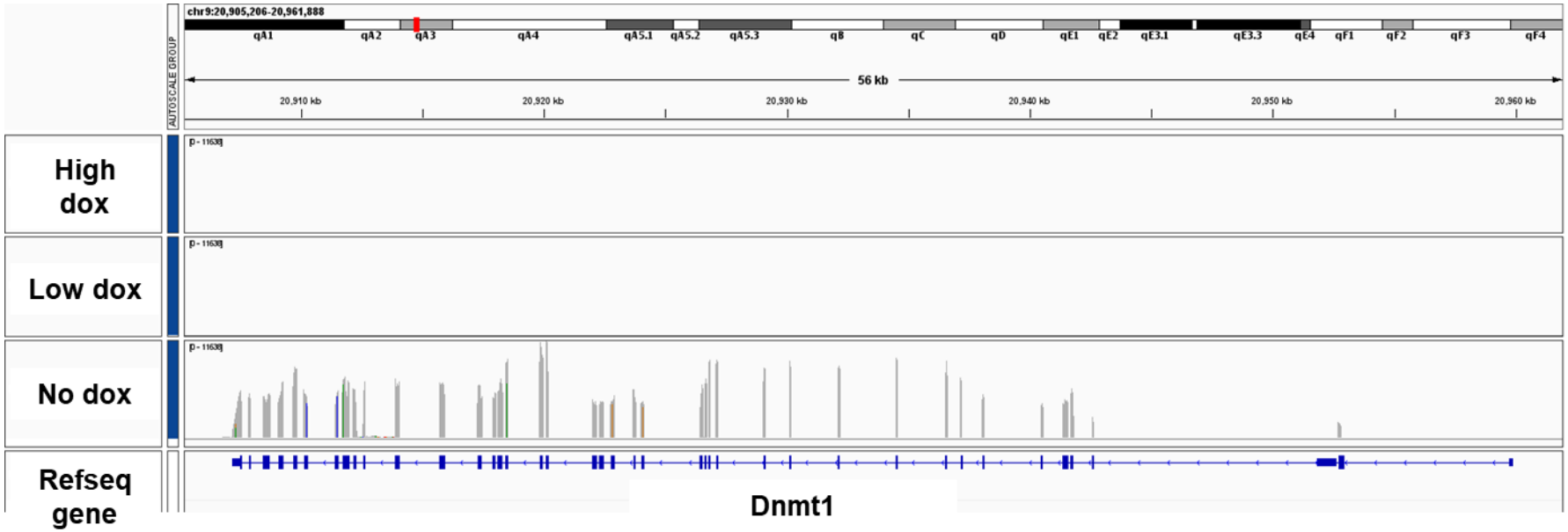


Figure 4.3. Genome browser view of the *Dnmt1* locus in RNA-seq data showing expression was inhibited in dox treated *Dnmt1*$^{tet/tet}$ EBs at day 14 of differentiation. RNA-seq data from triplicate samples were merged and the data shown here aligned to the mouse genome (mm10) using IGV genome browser (v2.9.4).

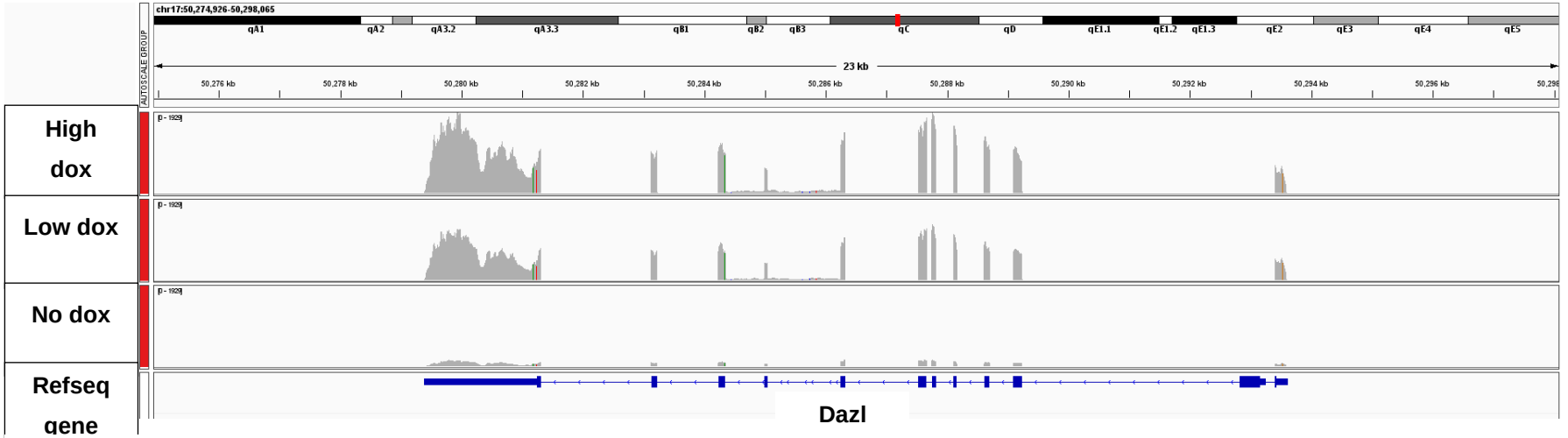


Figure 4.4. Germ cell-specific gene *Dazl* was detectable in day 14 hypomethylated EBs as repressive 5mC was lost due to *Dnmt1* inhibition. RNA-seq data from triplicate samples were merged and the data shown here aligned to the mouse genome (mm10) using IGV genome browser (v2.9.4).

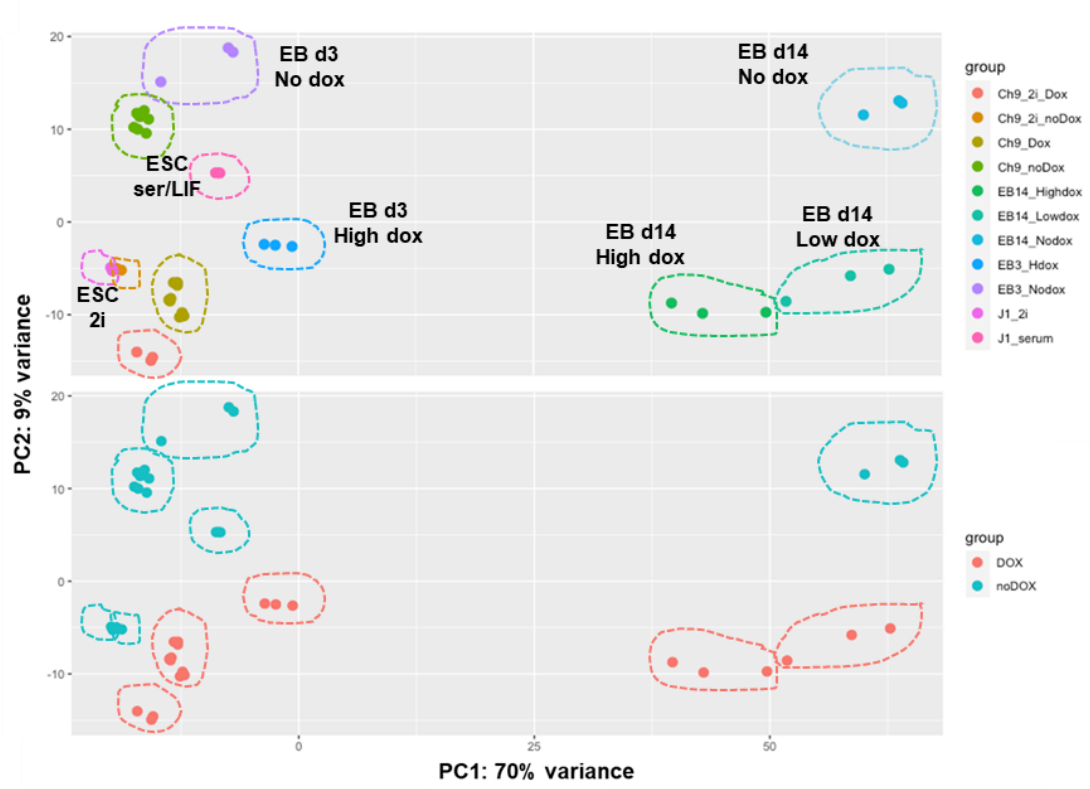


Figure 4.5. PCA comparing transcription profiles of ground state and naïve ESCs, and early and late EBs with and without *Dnmt1* inhibitor doxycycline (dox). Day 14 EBs cluster together in PC1 with clear segregation between the wild type and hypomethylated EBs. Day 3 EBs cluster more closely with ESCs. The top plot shows samples colour-coded by cell and culture type. The bottom plot shows samples colour-coded by treatment condition. Dox-treated and untreated samples are segregated along PC2.

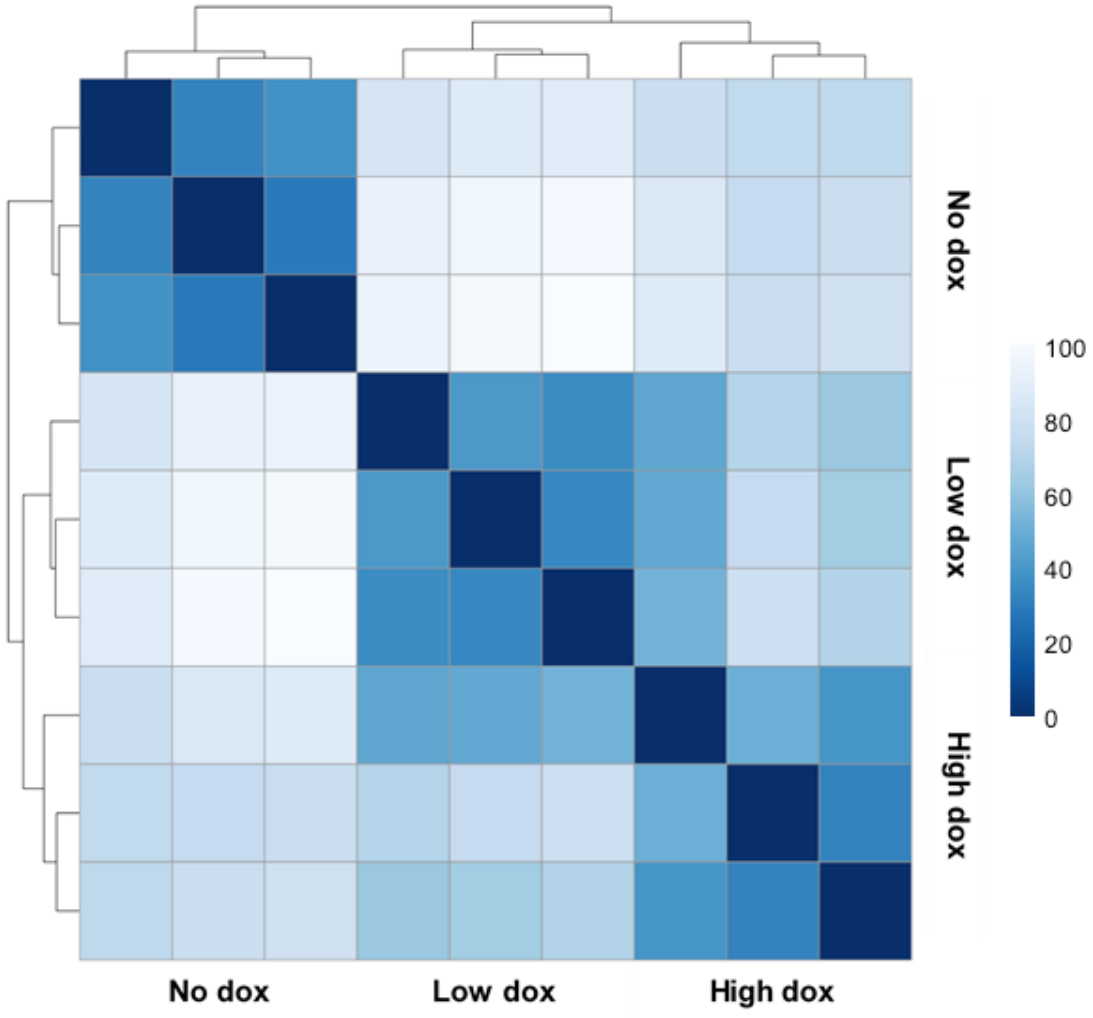


Figure 4.6. Clustered distance map showing that dox-treated EB samples were more similar to each other and distinct from untreated EBs. RNA-seq samples grouped by similarity of gene expression profile in R. Deeper blue to black indicates fewer dissimilarities, whereas pale blues to white indicates greater dissimilarity.

4.2.2 The EB RNA-seq data recapitulate the earlier single gene qPCR analyses

Having confirmed the efficacy of the dox treatment in maintaining inhibition of *Dnmt1* in EBs for up to 14 days at the transcriptomic level, I then re-examined the expression of the marker genes which were quantified by qPCR in Chapter 3. In figure 4.7a I show that the RNA-seq data from day 14 EBs broadly recapitulate patterns observed by qPCR. *Dazl*, a germ cell-specific gene which is repressed specifically by 5mC, is significantly upregulated in dox-treated *Dnmt1*$^{tet/tet}$-derived EBs (No vs Low, one-way ANOVA, *padj*=0.020). The master pluripotency transcription factor *Oct4* (*Pou5f1*), which is normally silenced in differentiated cells and highly expressed in ES cells, is upregulated in dox-treated EBs (No vs Low, one-way ANOVA, *padj*=0.012). The embryonic germ layer-specific genes also showed similar expression patterns; endoderm marker *Gata4* was expressed less in hypomethylated EBs relative to the wild-type (No vs Low, one-way ANOVA, *padj*=0.0002); ectoderm marker *Otx2* was expressed significantly more (No vs Low, one-way ANOVA, *padj*=0.0413); finally, mesoderm marker *Snai1* was similarly expressed in untreated and low-dox EBs as observed in the qPCR data, although its expression was lower in high dox EBs (No vs Low, one-way ANOVA, *padj*=0.555; No vs High, one-way ANOVA, *padj*=0.0043).

In the day 3 EBs *Dazl*, which was nearly undetectable in untreated EBs, was highly expressed in dox-treated EBs (Two-tailed, unpaired t test, *p*=0.0085). The changes in expression of the differentiation genes were much more subtle, with only mesodermal marker *Snai1* (Two-tailed, unpaired t test, *p*=0.047) and endodermal marker *Gata4* (Two-tailed, unpaired t test, *p*=0.0389) increasing in expression, whereas *Oct4* and *Otx2* (Ectoderm marker) were not expressed significantly differently from the wild-type.

These data support the previous observation that suppression of *Dnmt1* activity leads to aberrant gene expression, including potential co-expression of pluripotency and differentiation markers, and show the change is reproducible. They also suggest that this effect does not occur immediately

after withdrawal of pluripotency regulator LIF, but takes at least 3 days in differentiation culture before gene expression changes of some lineage markers become detectable at the transcript level. It remains to be determined if these gene expression changes occur within the same cells or represent separate populations of pluripotent and differentiated cells within single EBs. I attempt to address this in Chapter 5.

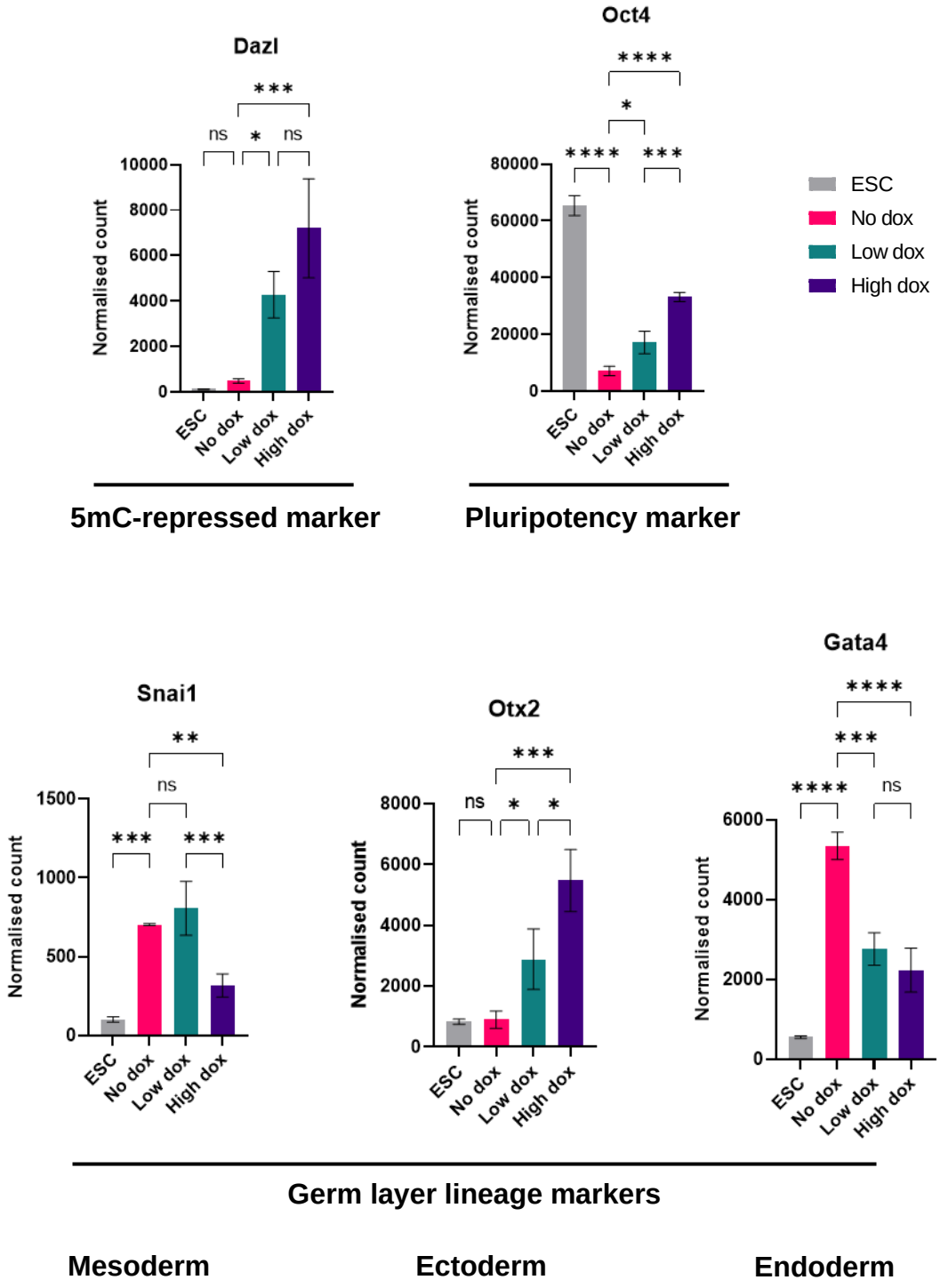


Figure 4.7 A. The day 14 EB +/- dox RNA-seq data recapitulates the expression patterns determined by qPCR. The expression of *Dazl* demonstrates the functional inhibition of *Dnmt1* under dox culture conditions (ns = not significant; * = *padj*<0.05; ** = *padj* <0.005; *** = *padj* <0.0005; **** = *padj* < 0.0001).

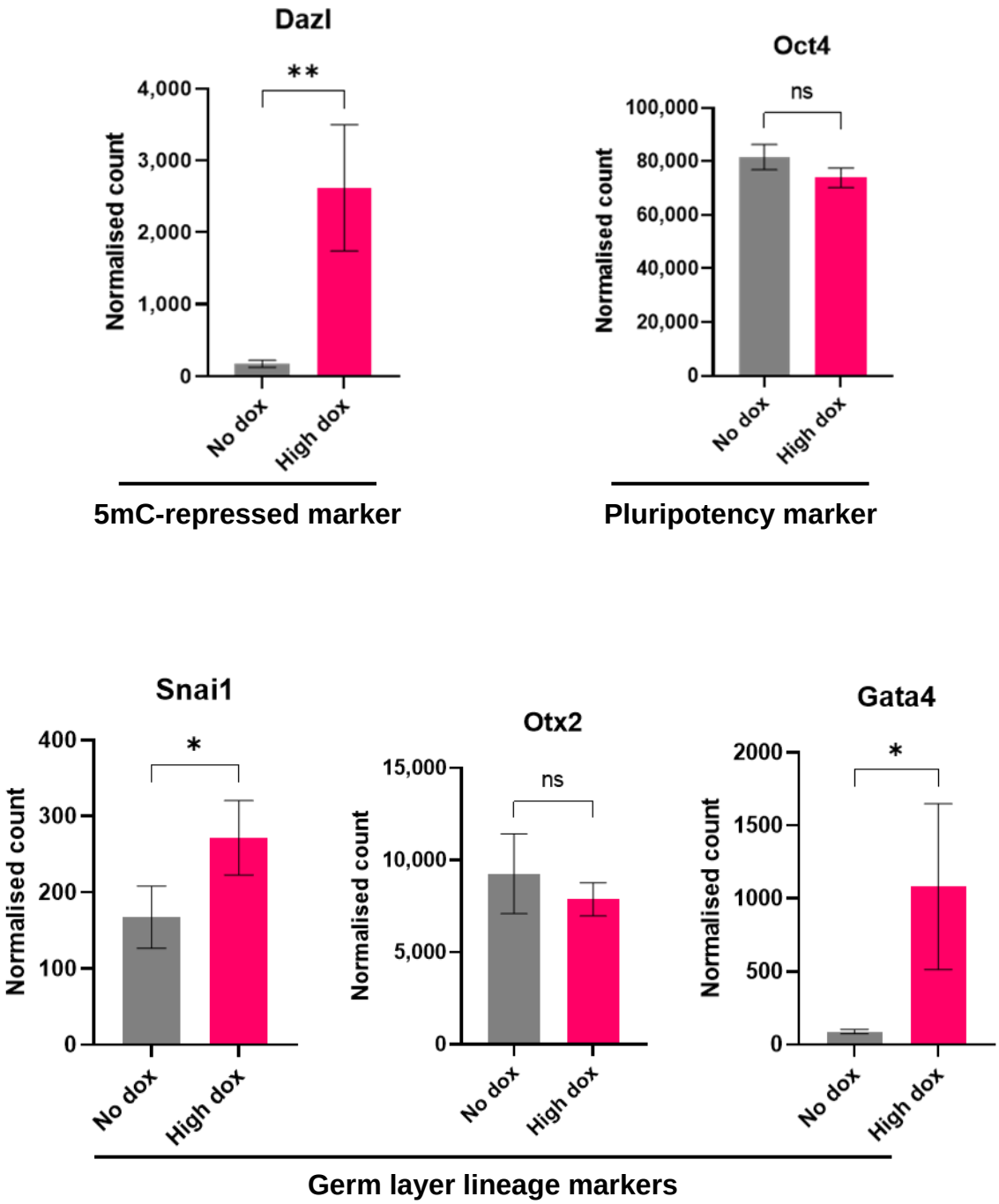


Figure 4.7 B. The day 3 EB +/- dox RNA-seq data recapitulates the expression patterns determined by qPCR. The expression of *Dazl* demonstrates the functional inhibition of *Dnmt1* under dox culture conditions. Expression of pluripotency marker *Oct4* remained similar between conditions up to day 3. Some changes in differentiation gene expression have begun. Normalised counts were compared by unpaired two-tailed t-test (ns = not significant; * = p <0.01; ** = p <0.001).

4.2.3 Differential gene expression analysis shows hypomethylation impacts multiple important signalling pathways

To assess the overall impact of inhibition of *Dnmt1* on EB gene expression, Dr Gautier performed differential gene expression analysis using DESeq2 (Love, Huber and Anders, 2014) (v1.26.0). In figure 4.8, the changes between untreated EBs and those treated with a low dose of dox at day 14 are shown as a volcano plot. 180 genes decreased in expression >-2-fold (*padj* <0.05), whereas the majority of differentially expressed (DE) genes increased in expression >2-fold (n=1036). When untreated and high dox-treated EBs were compared, more genes decreased in expression >-2-fold (n=817), but again the majority increased in expression (n=1079, Fig. 4.9). In both dox treated conditions most differentially expressed (DE) genes were upregulated which fits with the theory that DNA methylation is generally a repressive epigenetic mark, therefore removal of the mark releases previously silenced genes. Although a similar number of genes increase in expression >2-fold in both dox conditions, many more genes were downregulated in the high dox condition. This could be due to indirect effects of aberrant gene expression as a result of hypomethylation, or it could be due to an increase in expression of transposable elements (TEs) which are usually highly DNA methylated and cause further aberrant expression directly or indirectly through their transposable nature (Wang, Huang and Shi, 2020). I will revisit TE expression in more detail later in this chapter. In both conditions the most upregulated genes included imprinted genes such as *Meg3* and *Rian*. As expected, the most downregulated gene was *Dnmt1*.

As the most significant changes were observed at the later stages of differentiation, I will first address DE genes and affected pathways in day 14 EBs and describe changes at day 3 in the subsequent section.

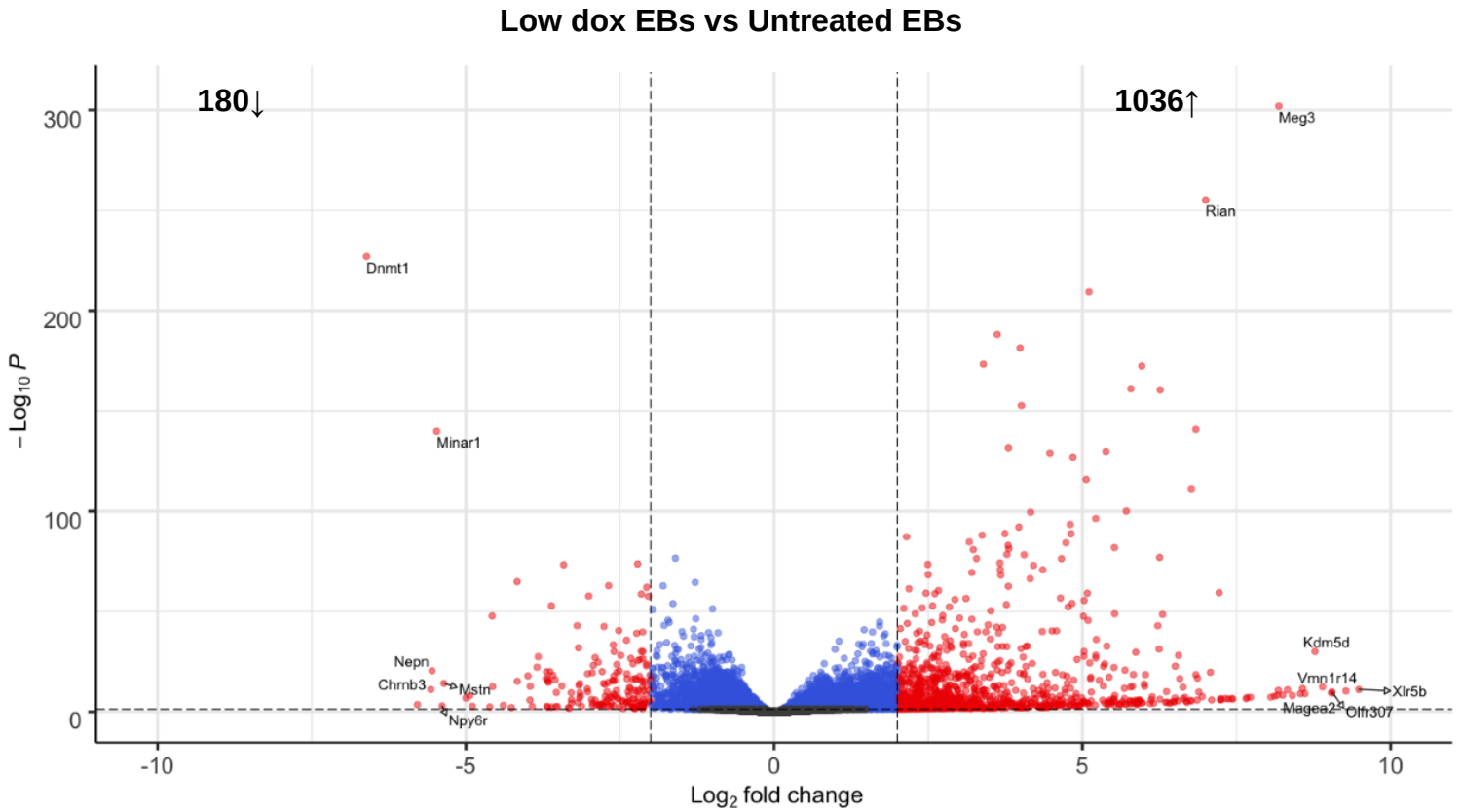


Figure 4.8. Inhibition of *Dnmt1* with a low dose of dox during EB differentiation leads to up-regulation of many genes at day 14 of culture with relatively few being downregulated >2-fold. The top 6 up- or down-expressed genes are labelled. 10,112 genes were differentially expressed (*padj*<0.05; red and blue dots); of these, 1036 increased and 180 decreased >2-fold in expression (red dots). The blue dots are genes whose expression changed significantly but <+2/-2 fold-change. Black dots are genes whose expression did not change significantly between conditions. In total 34,331 genes are shown.

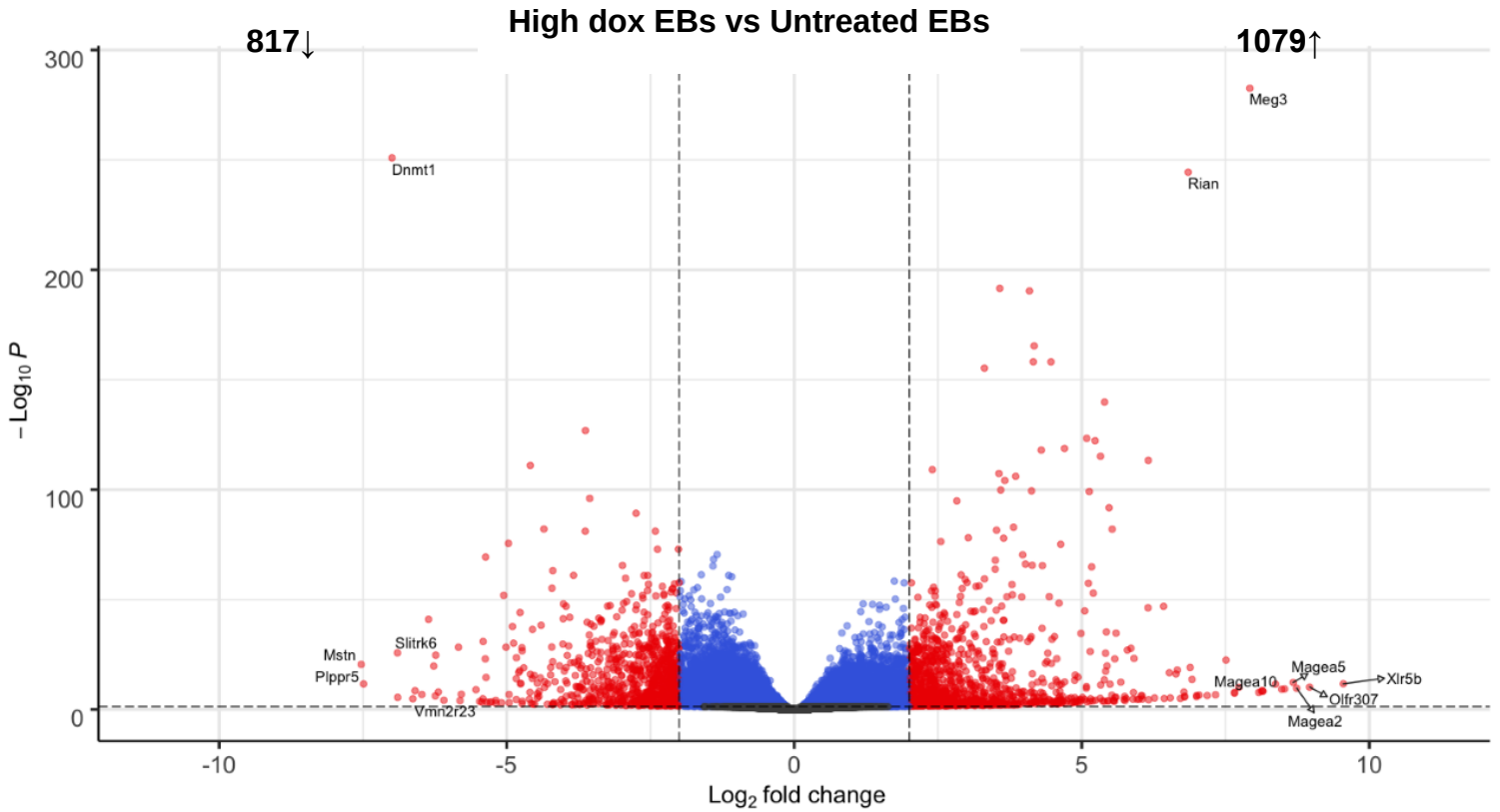


Figure 4.9. Inhibition of *Dnmt1* with a high dose of dox during EB differentiation leads to up-regulation of a similar number of genes compared to the low dose EBs at day 14 of culture, but with far more being downregulated >2-fold. The top 6 up- or down-expressed genes are labelled. In total 13,399 genes were differentially expressed (*padj*<0.05, red and blue dots); of these, 1079 increased in expression and 817 decreased >2-fold (red dots). The blue dots are genes whose expression changed significantly but <+2/-2 fold-change. Black dots are genes whose expression did not change significantly between conditions. In total 34,331 genes are shown.

To gain a broad overview of changes in biological processes induced by differentiation under hypomethylation conditions at day 14, lists of differentially expressed (DE) genes ranked by *padj* value were analysed using GO term enrichment analysis (Fig. 4.10). When comparing high vs no dox day 14 EBs, >2-fold upregulated genes were frequently associated with reproductive GO terms such as ‘reproductive process’, ‘gamete generation’, and ‘meiotic cell cycle’. The most downregulated terms were linked to differentiation, including ‘cell differentiation’, ‘organ development’, and ‘morphogenesis’. Many of the same GO terms were found in the low dox condition analysis, though the gene counts were lower (not shown). Together these suggested that during differentiation dox-treated EBs had a skewed gene expression profile, with an increase in expression of meiosis-related genes.

GO term enrichment is also shown as a GO term network map which I generated using Metascape, a web-based tool built for gene list analysis that incorporates over 40 continuously updated, independent databases (Zhou *et al.*, 2019) (Fig. 4.11). For the network analysis plots, the Metascape algorithm identified representative enriched GO terms in lists of the 1000 most-upregulated genes (*padj*<0.05) per condition at day 14 of differentiation and displayed them as clustered network maps. Each term was shown as a colour-coded node where the size of the node indicates the number of genes represented. The nodes were connected by similarity, and the thickness of the connection is proportional to the similarity score. For the top map in figure 4.11, the figure legend shows one term from each colour-coded cluster which is representative of the contents of that cluster. Related terms are clustered by colour. In the bottom panel, the same map is shown as above but the node colour now indicates the experimental condition, with red representing genes present in the high dox EB list, blue the low dox EB list, and green the untreated EB list. Nodes with multiple colours show genes which overlap between conditions. Together the maps show which terms are enriched in which condition; the main terms enriched in the high and low dox conditions but not the no dox condition are highlighted by grey brackets. These nodes

are represented by terms such as ‘meiotic cell cycle’, ‘cellular processes involved in reproduction’, and others such as ‘DNA alkylation’, ‘DNA metabolic process’, and ‘regulation of cell adhesion’. These are consistent with those found in the previous GO term analysis in figure 4.10 and support the view that meiotic differentiation pathways are upregulated in hypomethylated EBs. It is also interesting that cell adhesion-related terms appeared in two independent enrichment analyses, as high dox-treated EBs in particular had a less adherent phenotype.

Also using Metascape, differences and similarities in the top 1000 DE genes between dox-treated and untreated conditions were visualised as Circos plots (Zhou *et al.*, 2019) (Fig. 4.12, Top). In this type of plot the outer coloured rings indicate the experimental condition. The inner ring shows the multiple conditions (dark orange). Unique genes are shown by the light orange rings. Genes expressed in multiple conditions are connected by purple lines. Between EB conditions, the untreated sample (ServsNo) had the most unique genes, whereas the LowvsNo and HighvsNo samples had more than half the same genes in common. This analysis was extended to include more samples in a second plot, including somatic mouse cells (MEFs), *Dnmt1*$^{tet/tet}$ ESCs, and day 3 wild-type EBs (Fig. 4.12, Bottom). This showed that the dox-treated day 14 EBs had the most genes in common. Untreated EBs were more similar to dox-treated EBs than somatic MEFs, pluripotent ESCs or wild-type day 3 EBs. The day 3 EBs shared a little over half their genes with other conditions, mostly with ESCs. MEFs had the most unique genes. Of the few they shared with other conditions, most were in common with day 14 EBs. These observations reflect the stages of differentiation captured by RNA-seq; the day 14 EBs, regardless of treatment, were more similar to somatic MEFs than to pluripotent ESCs, which were most similar to day 3 EBs. This fits with the view that differentiation proceeds slowly, with early EBs retaining many pluripotent cell-like features until at least day 3, and not obtaining terminally differentiated cell-like expression features until much later in EB culture.

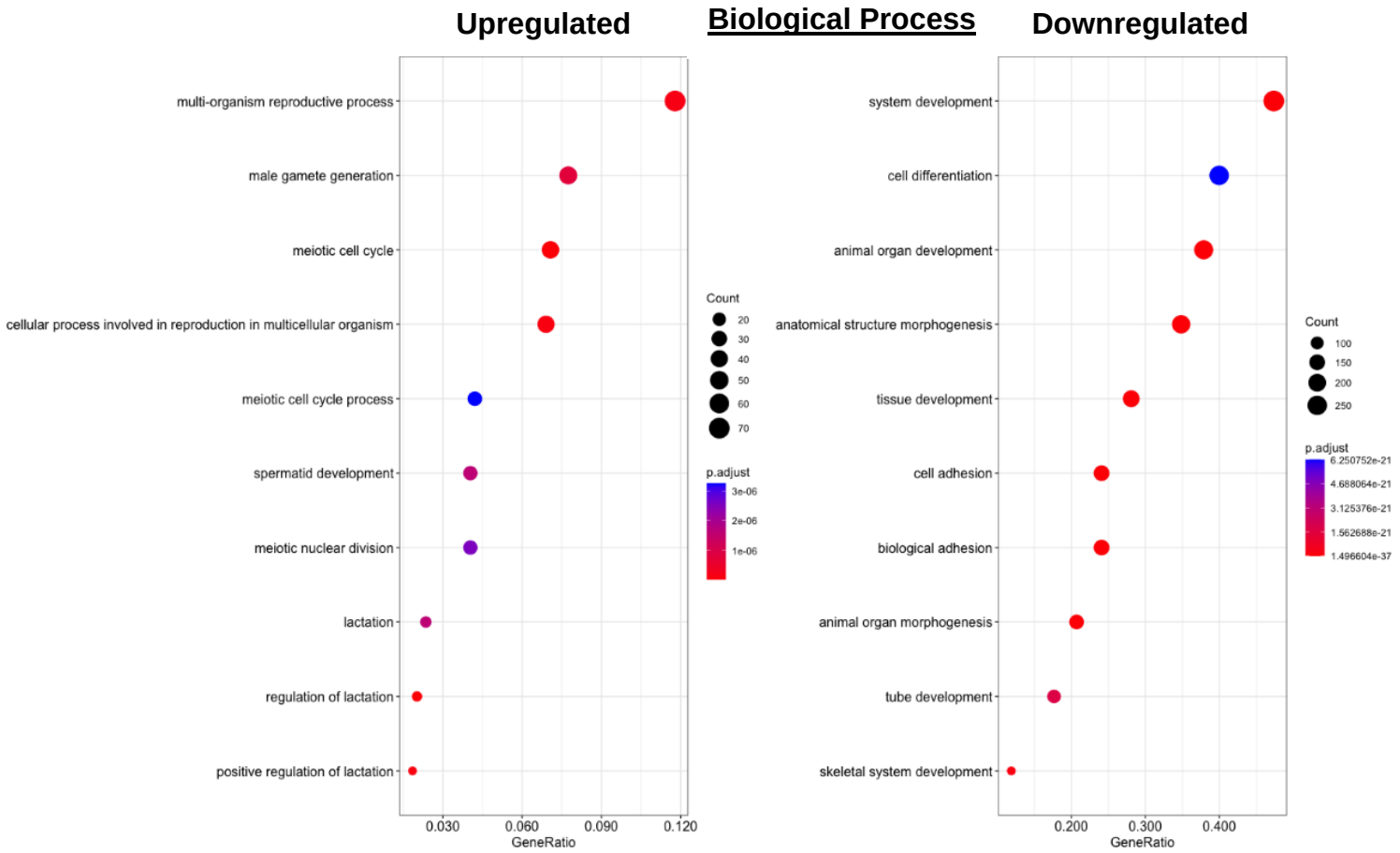


Figure 4.10. EBs treated with a high dose of dox were enriched for GO terms associated with meiotic processes and germ cell development. Biological Process GO term enrichment was compared between day 14 Untreated and High dox EBs. The size of the circle indicates the number of genes associated with the GO term. The colour of the circle indicates the *padj* value.

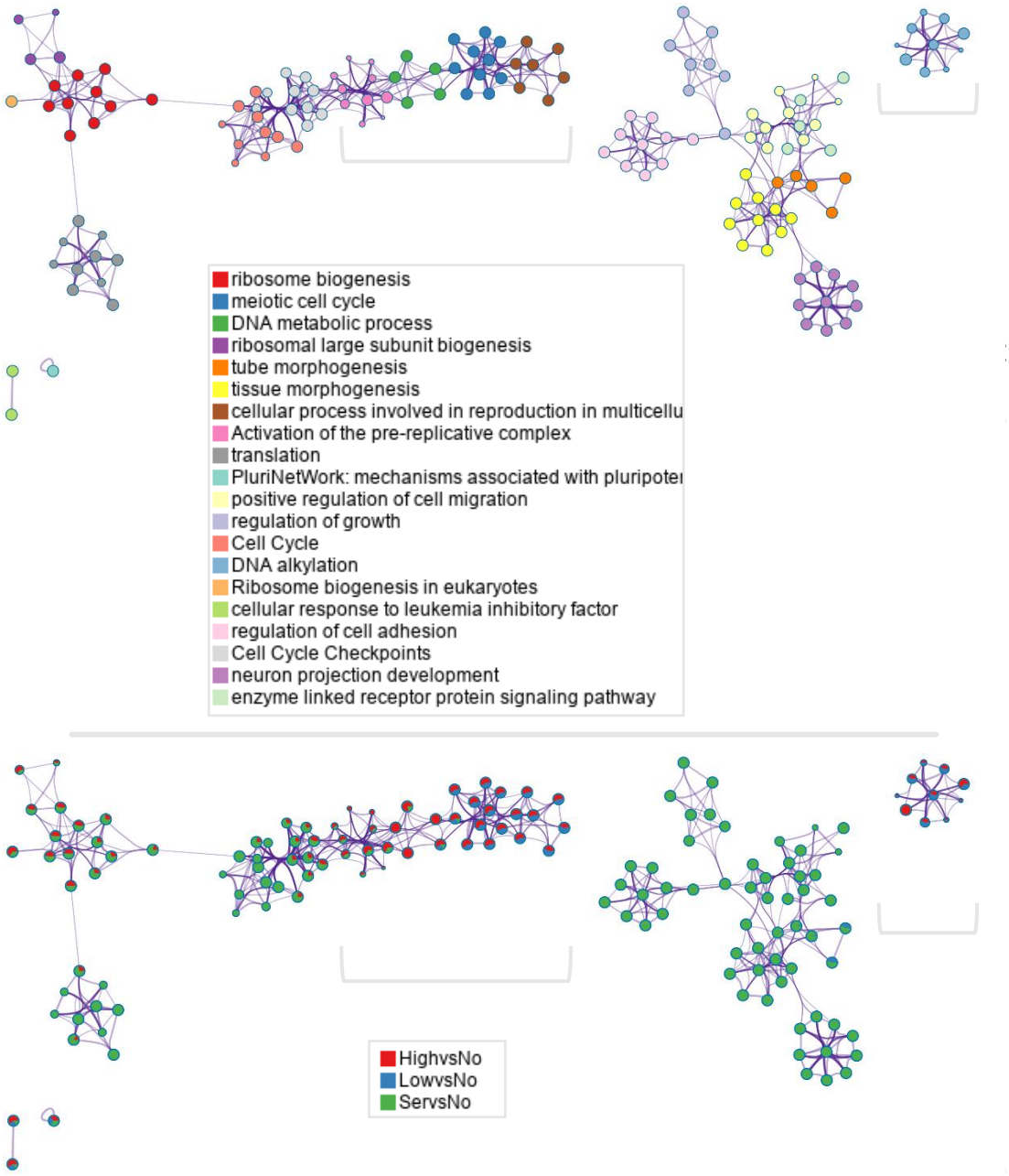


Figure 4.11. Dox-treated EB gene profiles show an enrichment for meiotic cell cycle-associated genes and processes involved in multicellular reproduction (grey brackets). Plots were generated using *Metascape.org*. Lists of the top 1000 upregulated genes (compared to untreated EBs; *Padj*<0.05) in Low and High dox EBs and serum/LIF ESC cells were analysed. The genes were then organised by their associated Gene Ontology (GO) terms and sorted into a network of nodes (coloured circles). The size of each node is proportional to the number of genes in the input list associated with that GO term, and the colour indicates cluster identity. Nodes are connected by lines; the thickness of the line indicates the similarity score ($\geq$0.3) of the connected nodes. A term representative of each cluster is shown in the colour key (Top).

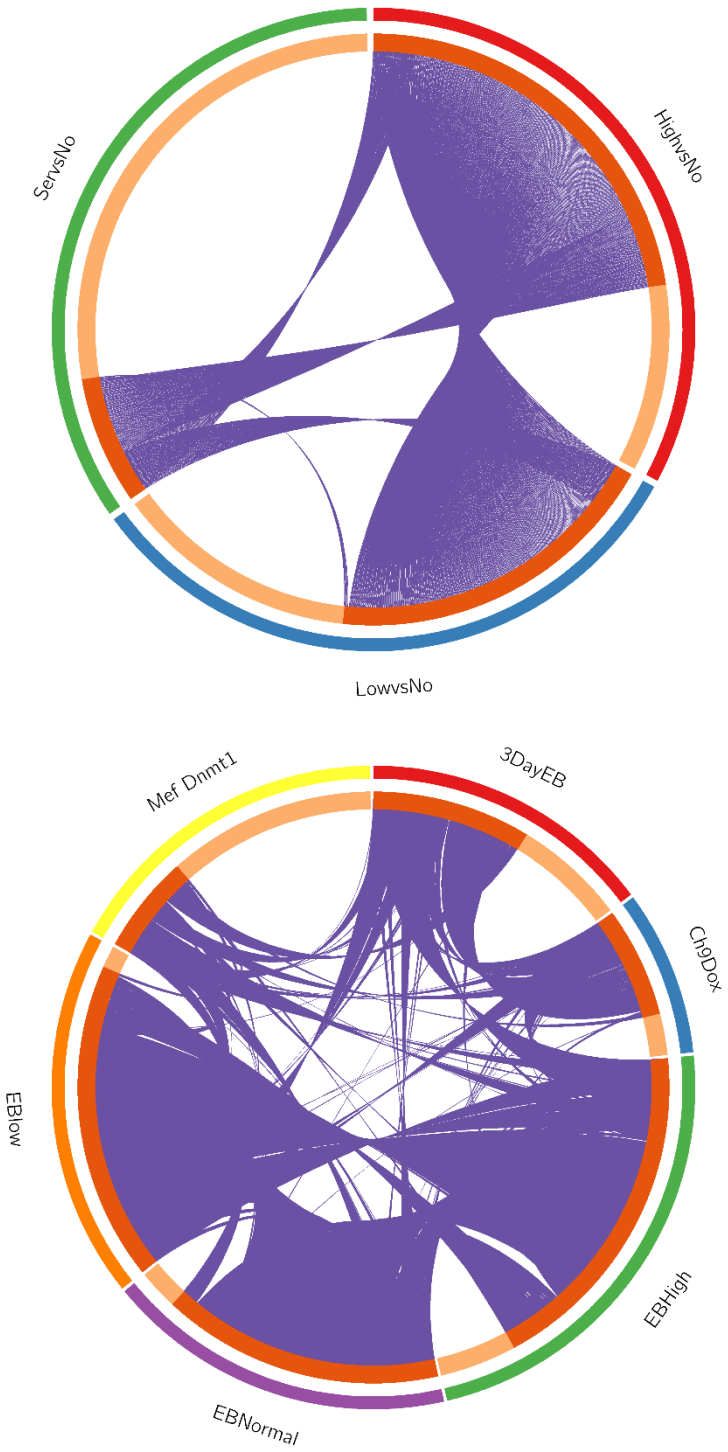


Figure 4.12. (Top) Dox-treated EBs had more overlapping genes compared to untreated EBs. (Bottom) Treated and untreated day 14 EBs shared more overlapping genes compared to other groups; day 3 EBs shared genes mostly with *Dnmt1*$^{tet/tet}$ ESCs; somatic MEF cells had the most unique genes. Genes which are common between treatments are connected by a purple line. The length of the light and dark orange segments in the inner ring indicate the proportion of unique and shared genes per treatment respectively.

4.2.4 Changes in gene expression at an early stage of EB

The first RNA-seq experiment performed in this project captured the gene expression profiles of day 14 EBs which contain advanced differentiation cell types and therefore represent late-stage differentiation. To gain insight into changes occurring during early differentiation, a second RNA-seq experiment was performed with high dox or untreated EBs collected at day 3 of EB culture. As EBs spend the first three days of culture forming undisturbed in hanging droplets of media, only the high dose of dox was used to ensure the drug was not depleted while the media could not be refreshed. Transcriptome analysis showed that at day 3 most differentially expressed genes were upregulated; 734 genes increased >2-fold (*padj* <0.05), whereas only 24 decreased >-2-fold (*padj* <0.05; Fig. 4.13). GO analysis of the upregulated genes revealed an enrichment for reproductive- and development-related terms (Fig. 4.14), consistent with the appearance of multiple imprinted and X-linked genes in the volcano plot, such as *Xlr* (X-linked, lymphocyte regulated) family genes and *Usp26*, which is also located on the X chromosome. This shows that at this point during differentiation most of the differentially expressed genes are those which had previously been silenced by DNA methylation specifically, and suggests that effects on differentiation seen in the day 14 EBs occur after this time point.

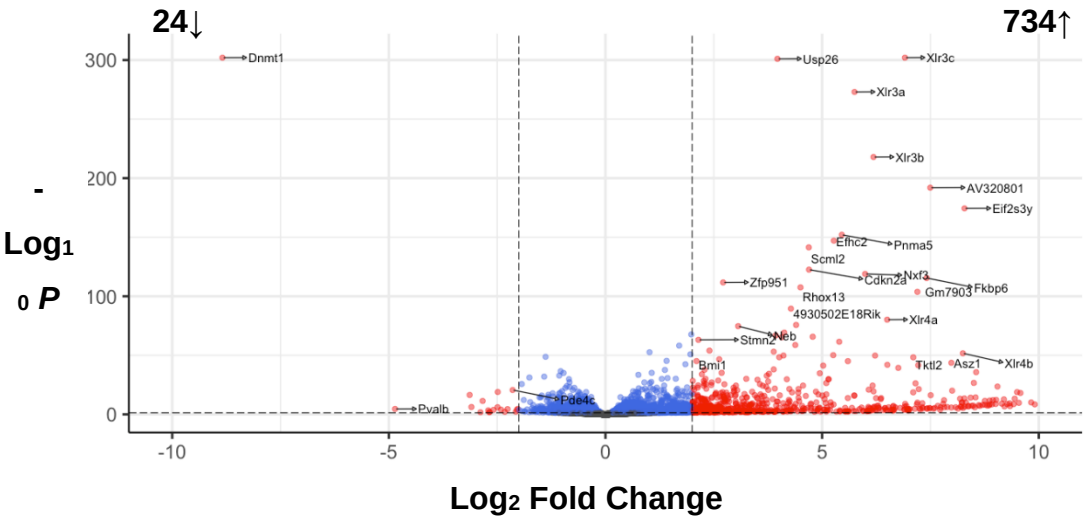
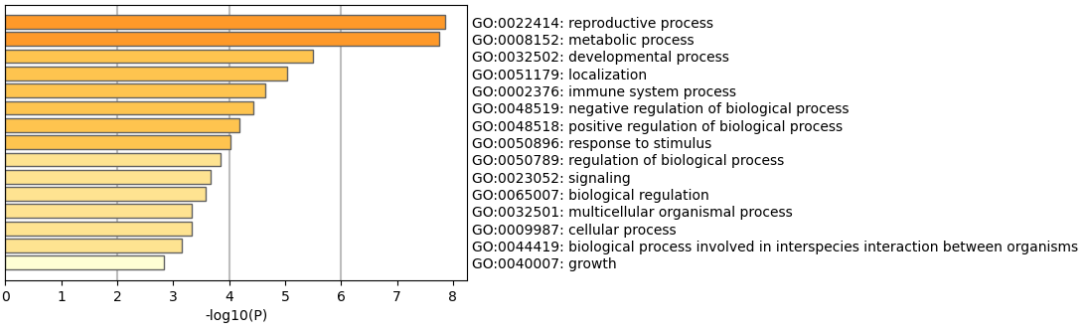


Figure 4.13. Most differentially expressed genes in hypomethylated EBs at day 3 of differentiation were upregulated and associated with repression by DNA methylation. Total events = 18,910. Red = *padj*<0.05, Log_2 Fold change ≤-2/≥2; Blue = *padj*<0.05, Log_2 Fold change ≥-2/≤2; Black = *padj*>0.05, Log_2 Fold change ≥-2/≤2.

Figure 4.14. Hypomethylated EBs at day 3 were enriched for meiotic and developmental GO terms. A list of genes which increased ≥2-fold (*padj* <0.05) in High dox vs untreated day 3 EBs was analysed for GO-term enrichment using Metascape.org. The top 20 enriched terms are



4.2.5 ESC-associated pluripotency genes were more predominantly expressed than EB-associated differentiation genes at day 3

The lab has previously performed RNA-seq using both *Dnmt1*$^{tet/tet}$ ESCs and an independent wild-type cell line, J1s, both in serum/LIF culture and 2i culture and with and without dox (Dunican *et al.*, 2020; Shukla *et al.*, 2020). In comparison to serum/LIF culture, 2i culture produces homogenous populations of ES cells which are considered to be in the 'ground state' of pluripotency and more akin to cells of the *in vivo* blastocyst (Ying *et al.*, 2008; Schlesinger and Meshorer, 2019). This homogeneity can be seen in figure 4.5 wherein the 2i replicates cluster much more tightly together than serum/LIF ESCs. I leveraged these datasets to see if hypomethylation induced a more ES-like state in developing EBs and, if so, was that ES-like state more similar to ground-state 2i ESCs or naïve serum/LIF ESCs?

The datasets were used in a bulk RNA-seq analysis along with the day 3 and 14 EB datasets, with and without dox, performed by Dr Philippe Gautier. The PCA plots in figure 4.5 give an overview of how all the datasets are clustered. The middle PCA plot shows the day 14 EBs clustering to the right along PCA 1 (70%), away from the pluripotent ES cells. Interestingly, day 3 EBs cluster mostly to the left. This may be indicative of the continued activity of the pluripotency network at this stage of differentiation which would increase the similarity in gene expression profile to undifferentiated ES cells which highly express pluripotency transcription factors. PCA 2 (9%) highlights the difference between dox-treated and untreated samples, which is emphasised in the bottom plot where samples are colour-coded by treatment type. Interestingly, all *Dnmt1*$^{tet/tet}$ samples show a bimodal distribution; the untreated samples cluster together in the top half of the plot, along with J1s which have wild-type *Dnmt1* expression unaffected by dox, and dox-treated samples clustering at the bottom. This shows that induced hypomethylation with dox produced a consistent change in gene expression versus the equivalent wild-type condition across cell line backgrounds and cell types. Overall, this plot shows that these independent datasets could be used together to compare how induced hypomethylation impacts gene expression

profiles in differentiating cells and if it brings about a more ES-like profile, and if so, if the profile is more akin to ground-state or naïve pluripotency. Using the above datasets, we re-examined expression of lineage-specific markers, 2C-markers, imprinted genes and apoptosis markers to see if hypomethylation in EBs changed expression levels to ESC levels. Lists of genes associated with specific germ layer lineages and apoptosis regulation were compiled by searching for the appropriate GO term in the Gene Ontology Resource website (Geneontology.org (Ashburner *et al.*, 2000; Carbon *et al.*, 2021)). The 2C list was adapted from the list published by Eckersley-Maslin *et al*.(Eckersley-Maslin *et al.*, 2016). The gene lists used for the above analyses are shown in the appendices, and the results are displayed as heat maps in figures 4.15 and 4.16.

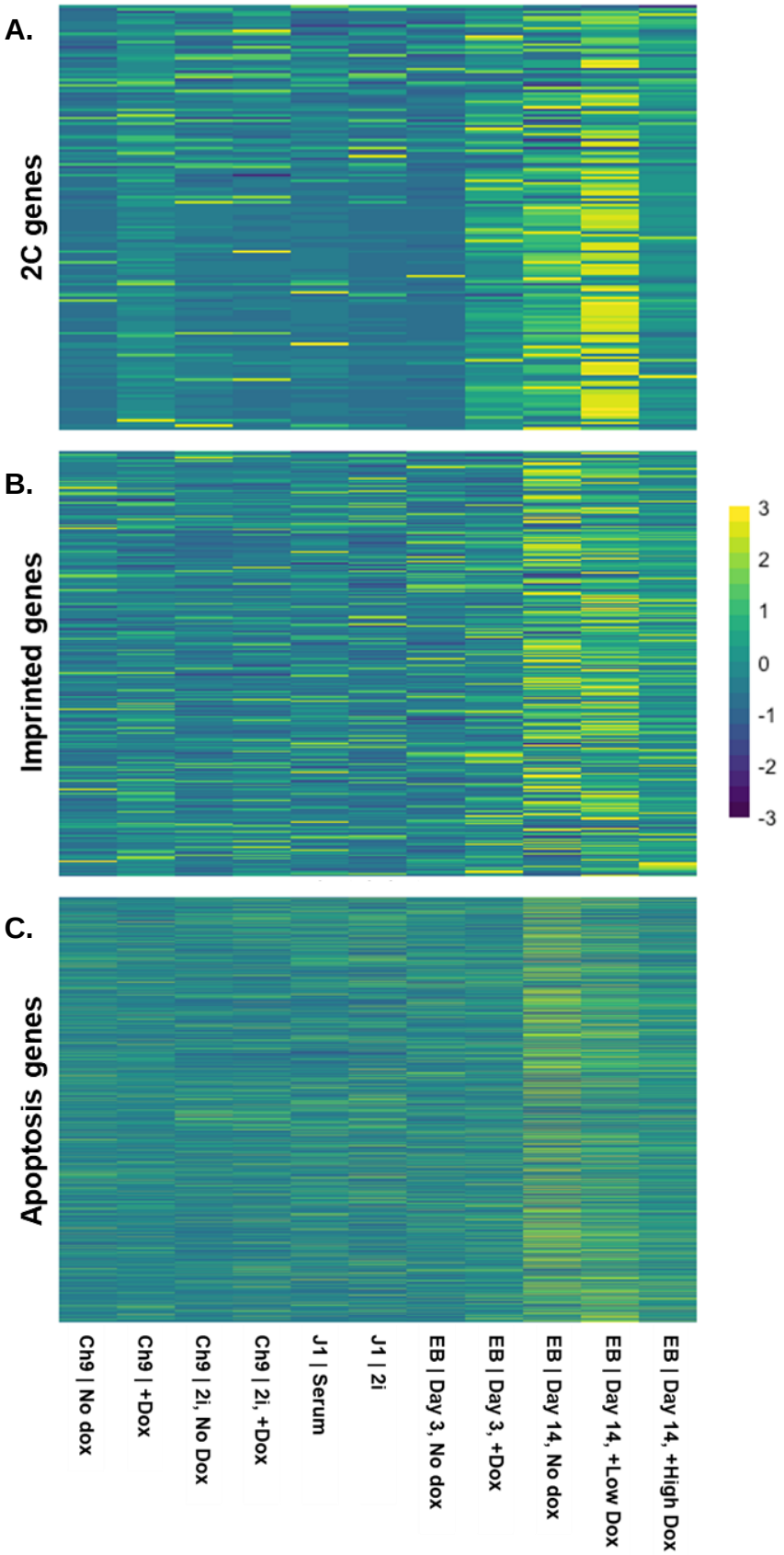


Figure 4.15 | Legend on next page…

Figure 4.15 | Figure on previous page…

Figure 4.15, A. 2C-associated genes, which are lowly expressed in ESCs and early EBs, were relatively highly expressed in dox-treated early EBs and day 14 EBs. Gene expression was highest in low-dox treated day 14 EBs.

Figure 4.15, B. Imprinted genes are highly expressed in day 14 EBs, though slightly reduced under hypomethylation conditions. Imprinted genes are silenced in all ESCs and highly expressed in late-differentiation wild-type EBs. They are also upregulated in dox-treated day 14 EBs, though less than in wild-types. Expression remains at a similar level to ESCs in early day 3 EBs, with and without *Dnmt1*.

Figure 4.15, C. Hypomethylated EBs did not show an enrichment for signature apoptosis genes relative to the wild-type at either of the time-points analysed. Wild type EBs at day 14 expressed relatively high levels of apoptosis signature genes relative to ground state and naïve ESCs, day 3 EBs, and late-stage hypomethylated EBs.

The scale bar shows each gene's z-score, which is calculated as the number of standard deviations from the group mean.

The heat map in figure 4.15 (B) shows relative changes in expression of mouse imprinted genes between naïve (serum/LIF culture) and ground-state (2i culture) ESCs, and in early and late stage EBs, with and without dox. Proper maintenance of imprinted gene epigenetic state is important as these genes have key roles during foetal development, particularly during neurodevelopment and formation of the placenta (reviewed by Thamban *et al.* (Thamban, Agarwaal and Khosla, 2020)). *In vivo*, developing primordial germ cells (PGCs) undergo two waves of global demethylation between ~E8.5-13.5 orchestrated by Tet demethylases and passive demethylation. The second wave starts at ~E10.5, removing 5mC from previously demethylation-resistant imprint control regions (ICRs) so that the epigenetic

imprint is reset for the next generation of totipotent cells and to enable establishment of sex-specific imprints in developing gametes (Hackett *et al.*, 2013; Leitch, Tang and Surani, 2013; SanMiguel and Bartolomei, 2018). In the developing spermatogonia and oocytes imprinted genes are expressed from one allele in a parent-of-origin manner; the other allele is silenced by deposition of DNA methylation at the appropriate ICR during gametogenesis by Dnmt3a and Dnmt3L (reviewed in Barlow *et al*. (Barlow and Bartolomei, 2014)). The mark is then maintained across cell division by Dnmt1, and in somatic cells by histone modifications (Thamban, Agarwaal and Khosla, 2020).

In figure 4.15 B the heat map shows that imprinted genes are most highly expressed in wild-type day 14 EBs and low dox treated EBs. However, they are expressed at relatively low levels in high dox EBs at both day 3 and 14, and in wild-type EBs at day 3, with expression levels similar to ESCs. As imprinted genes are maintained in the silenced state by DNA methylation, one might expect to see an increase in expression in dox-treated *Dnmt1*$^{tet/tet}$ EBs, but that is not observed. However, loss of methylation at these loci would result in de-repression of the silenced alleles resulting in ncRNA being produced. These non-coding transcripts overlap with imprinted gene promoters (which normally regulate the expression of multiple imprinted genes as they tend to occur in clusters in the genome) and block RNA Pol III activity at the protein-coding allele, leading to a decrease in expression of the protein-coding transcripts (Thamban, Agarwaal and Khosla, 2020).

These findings are similar to those of Hargan-Calvopina *et al.* (Hargan-Calvopina *et al.*, 2016). To explain the 2-stage demethylation process in developing PGCs, they proposed that DNA methylation functions to prevent premature meiotic activation by maintaining the appropriate ICR methylation states through the first wave of demethylation, preserving the ICR methylation state until the PGCs arrive in the gonadal ridge and can be instructed of which sex of gamete to develop into, and therefore which imprints to silence. By generating a conditional *Dnmt1* KO mouse strain which deletes *Dnmt1* in PGC cells at the time of PGC specification, they

observed that loss of DNA methylation led to premature initiation of the meiotic program in both male and female cells, and consequently hypogonadism in adult male mice. These findings are in agreement with the EB model, in that I observed an increase in meiosis-related terms in the day 14 dox-treated EBs and altered lineage trajectories occurring from day 3.

They also observed that embryos derived from the *Dnmt1* KO PGCs had significantly fewer PGCs in the developing gonads. Initially they thought the drop in PGC number may have been caused by aberrant TE expression but found expression of IAP retrotransposons and a type of Long Interspersed Nuclear Element (LINE) retrotransposon were not affected and remained repressed, so concluded Dnmt1 is not required for their repression during PGC development.

Heat map C shows relative expression of apoptosis genes. Although higher in wild-type EBs, there is no enrichment for these genes under hypomethylation conditions at either the early or late differentiation stage EBs. The z-scores in the high dox EBs are similar to those of ESCs. This suggests apoptosis may not be the cause of the reduced size of mature hypomethylated EBs which is consistent with the previous data obtained from the TUNEL assay in Chapter 3.

In figure 4.16 there are heat maps for genes which are expressed in each of the mouse germ layers (the lists were obtained from GeneOntology.org and heat maps generated by Dr Gautier; this is described in detail in the materials and methods section). They show that in early EBs the levels of expression of lineage-specific genes is more similar to those found in pluripotent ES cells than day 14 EBs. Expression is higher in late-stage EBs; mesoderm and endoderm genes are expressed highest in the wild-types, but ectoderm genes are expressed to similar levels across conditions. The ectoderm heat map suggests these genes are enriched in J1 ES cells when cultured in 2i; this is a known feature of J1 cells cultured in 2i which is not yet fully understood (Shukla *et al.*, 2020). The list is relatively short due to the overlapping nature of ectoderm and mesoderm gene

expression profiles; for this analysis, genes which occurred in more than one of the germ layer gene lists were removed to obtain lists of genes specific to the associated germ layer.

Overall, all six heat maps described above suggest that the gene expression profiles of day 3 EBs, regardless of treatment, more closely resembled the profiles of pluripotent ES cells than late-stage differentiating EBs. This suggests that differentiation pathway activation and pluripotency pathway silencing occur later than day 3 during differentiation. In future work, more time-points could be analysed to try to pinpoint this stage. Also, these heat maps do not suggest a bias towards a particular lineage when cells are differentiated to EBs under hypomethylation conditions but a general subtle decline in expression of differentiation genes.

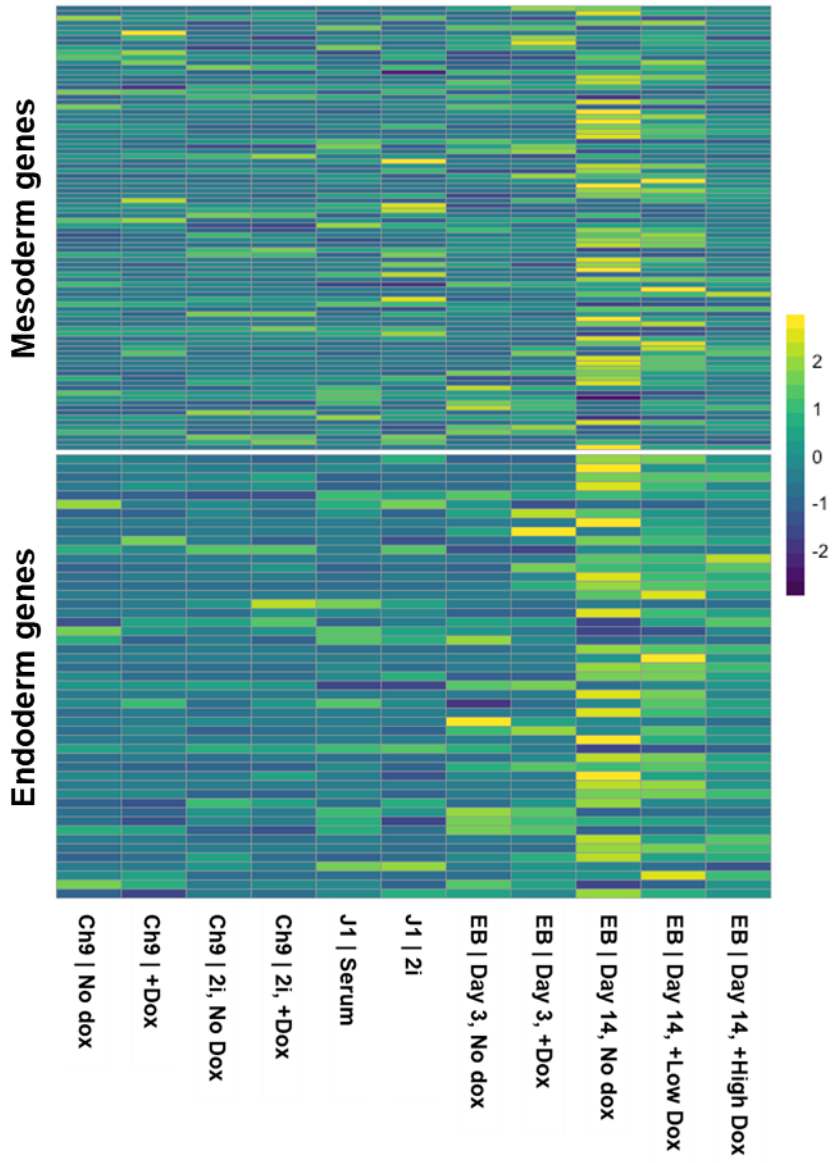
Figure 4.16 on next page…

Figure 4.16, Top left. Signature mesoderm genes highly expressed in wild-type EBs are lowly expressed in hypomethylated day 14 EBs. Expression is similar between early EBs and ESCs.

Figure 4.16, Top right. Signature ectoderm gene expression is similar between wild-type and hypomethylated EBs at both early and late–stage differentiation.

Figure 4.16, Bottom left. Signature endoderm gene expression is reduced in late-stage EBs when *Dnmt1* is inhibited. Endoderm genes were enriched in wild type EBs at day 14, but were expressed less in dox-treated EBs and at day 3 of differentiation in both wild type and treated EBs. Endoderm genes are lowly expressed in ESCs regardless of parental line, culture media, or dox treatment.

The scale bars show each gene's z-score, which represents the number of standard deviations from the group mean.



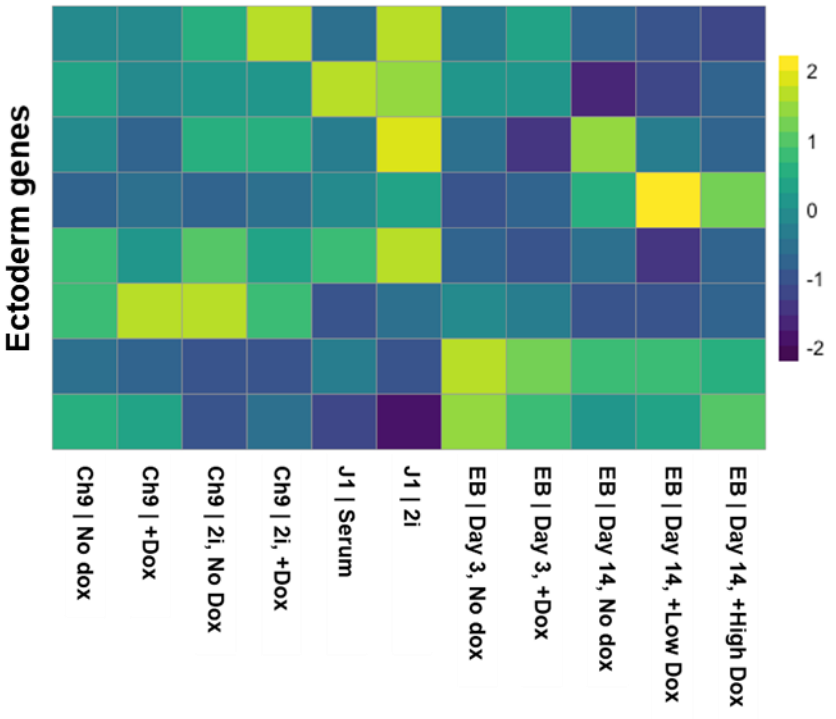


Figure 4.16 | Legend on previous page…

4.2.6 The relative absence of 5mC during differentiation leads to increased expression of transposable elements

DNA methylation is particularly concentrated at regions of the mammalian genome which harbour transposable element (TE) DNAs which are able to translocate and re-integrate into other areas of the genome with potentially harmful consequences (Wang, Huang and Shi, 2020; Wells and Feschotte, 2020). It is thought that epigenetic silencing mechanisms, particularly DNA methylation, could have evolved as a response to this type of viral invasion, helping to prevent potentially harmful transposition events (Deniz, Frost and Branco, 2019). This is particularly important in germ cells where genetic changes, including any TE transposition events, are passed onto the next generation (Richardson and Faulkner, 2018; Deniz, Frost and Branco, 2019). For this reason we sought to determine if there were any changes to TE expression in EBs formed from hypomethylated ES cells in the RNA-seq data. This analysis was done with the help of Professor Ian Adams, who created and ran the bioinformatics pipeline (as described in the Materials and Methods section, Chapter 2).

The day 14 EB RNA-seq analysis showed that generally TEs increased in expression between untreated and dox-treated EBs, but the change was very similar in both dox conditions (Fig. 4.17, A and B). When analysing changes in total transcript read counts it was found that TEs accounted for 2% of the total read counts in untreated EBs, and this increased to 4-6% in dox-treated EBs. Long terminal repeat (LTR)-class TEs, mainly Intracisternal A particles (IAPs), changed expression the most, almost doubling between untreated and treated conditions (Fig. 4.17, C & D). This is unsurprising as IAPs are normally silenced by 5mC (Deniz, Frost and Branco, 2019). An increase in expression of TEs could partially explain the difference in phenotype observed between the dox treatments. Slightly more transposition events in the high dox condition could have increased the likelihood that some occurred in essential gene regions which would cause aberrant downstream signalling. This could potentially result in stimulation of

the apoptosis pathway which would lead to a more extreme phenotype like the one observed.

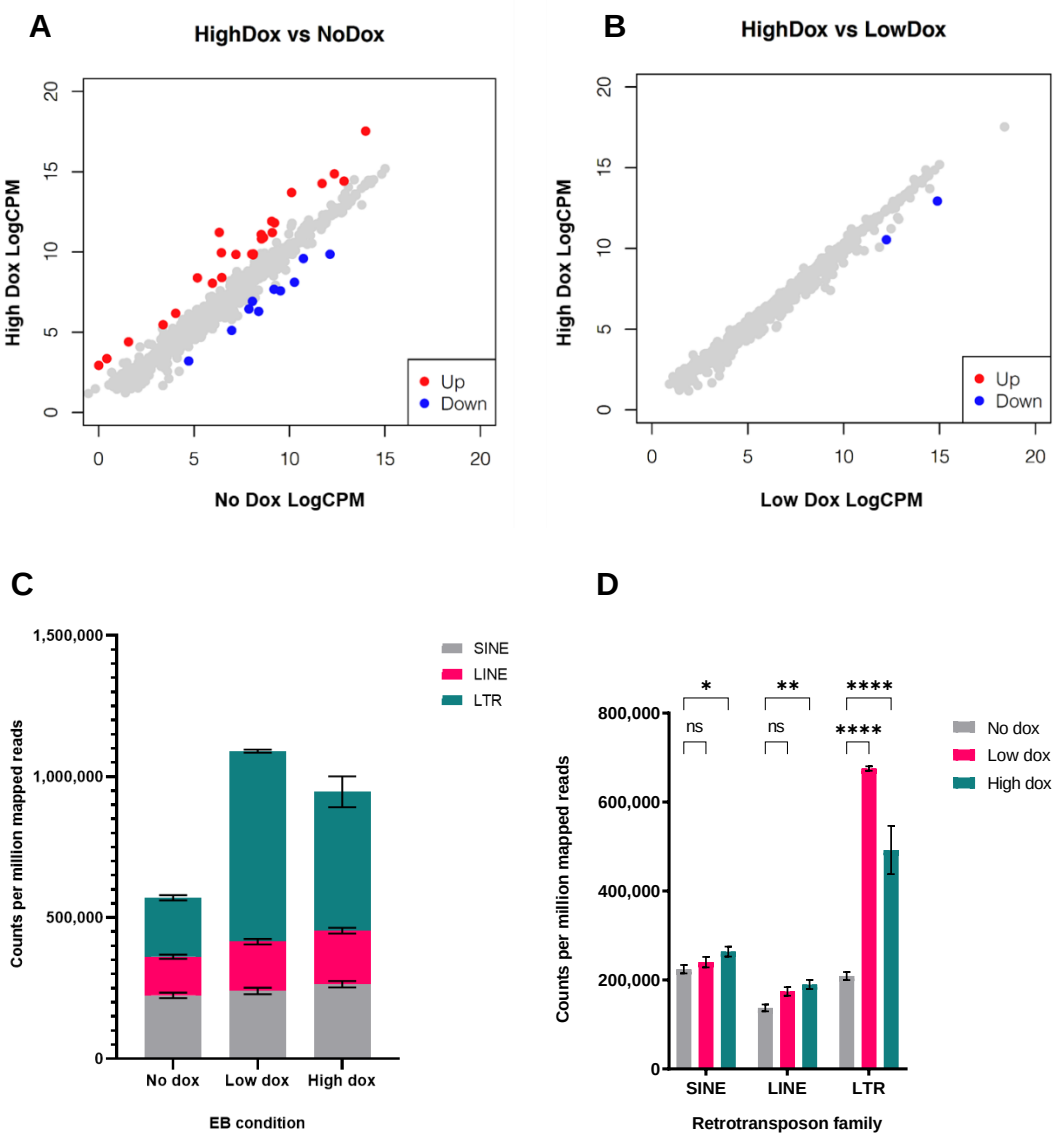


Figure 4.17. Inhibition of *Dnmt1* expression correlated with an increase in expression of transposable elements (TEs) by day 14 of EB differentiation. A). RTs were strongly upregulated in High dox vs No dox EBs, but B) there were few differences between High and Low dox treatments. A & B) Red and blue dots represent RTs whose expression increased or decreased at least 2-fold respectively (*padj*<0.05) C & D). Expression of LTR-class RTs changed most significantly in the absence of Dnmt1.

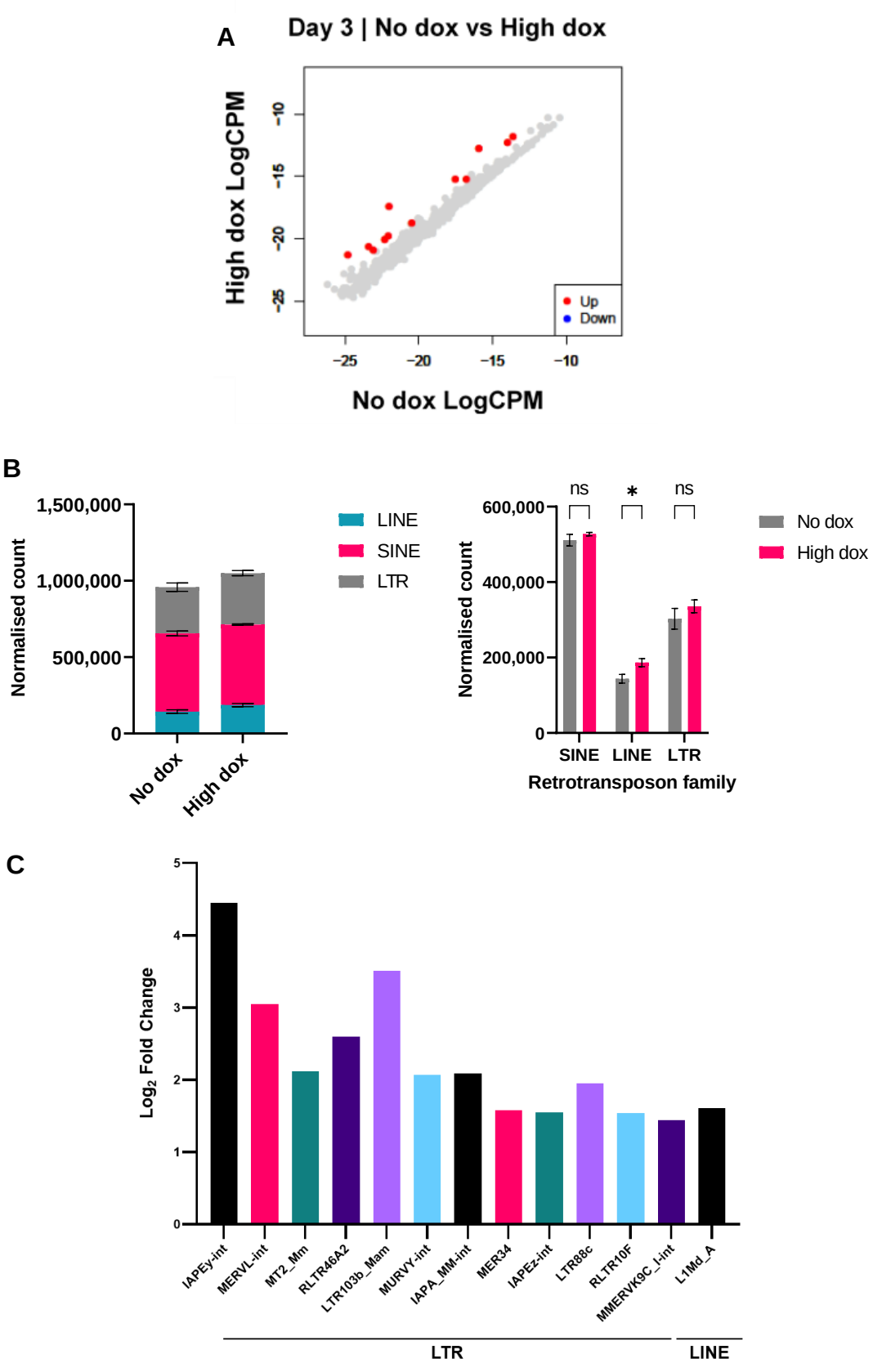


Figure 4.18. Legend on next page…

Figure 4.18 on previous page…

Figure 4.18. Significant changes in retrotransposon expression at day 3 of EB differentiation under hypomethylation conditions were limited to 11 LTRs and one LINE. A) No TEs were significantly downregulated during this time period, and a small number were upregulated (red dots). B) TE expression was similar between untreated and high dox-treated EBs, but LINEs were expressed slightly more in treated EBs. Total normalised counts were compared by unpaired two-tailed t test (ns = not significant, *= $p<0.01$). C) Most of the significantly upregulated TEs were representative of the LINE family. TEs are ordered from lowest to highest *padj* value.

In contrast, the day 3 hypomethylated EBs show relatively few changes in TE expression compared to the wild-type (Figure 4.18). Only approximately 12 TEs changed expression significantly, and all of those were upregulated (Figure 4.18, A). When comparing the average TE counts normalised to the total number of mapped reads most retrotransposon families were similarly expressed between the two conditions, though LINES were slightly upregulated under hypomethylation conditions (unpaired, two-tailed t test, $p=0.0092$; Figure 4.18, B). However, the most differentially expressed individual TEs were members of the LTR family of retrotransposons (Figure 4.18, C). This is consistent with other studies which have shown that LTRs are upregulated in mouse embryos which lack *Dnmt1* and are therefore strongly associated with silencing by DNA methylation(Walsh, Chaillet and Bestor, 1998), thus we might expect them to be the first to respond to an induced absence of 5mC.

Overall, these data suggest that loss of DNA methylation through inhibition of *Dnmt1* leads to strong up-regulation of TEs, particularly those of the LTR family, but this effect takes more than three days to become evident. The most significant changes affect the LTR and LINE family

retrotransposons which are normally highly methylated in somatic cells as they have the most potent transposition behaviour and are implicated in contributing to many cancer types (Deniz, Frost and Branco, 2019).

4.2.7 An EB single-cell RNA sequencing study showed similar findings

In 2020 a paper was published by Kim *et al.* (Kim *et al.*, 2020) in which a genetic barcoding system was used to record cell fate decisions occurring during embryoid body differentiation. By collecting scRNA-seq data from barcoded single cells they constructed a pseudo-time tree mapping the differentiation process. Using this they determined the timings of lineage commitment decisions made by single developing EB cells, such as when a pre-implantation epiblast-like cell differentiates either along the PGC-like pathway or the post-implantation epiblast-like pathway. They also used the tree to identify lists of highly expressed 'signature' genes specifically associated with each lineage. They also investigated the effects of demethylating agent 5-aza cytidine (Kelly, De Carvalho and Jones, 2010) (5-aza) on differentiation, and identified genes associated with early and late differentiation more generally. Here, we compared the expression of genes in their pre-defined lists to our bulk EB RNA-seq datasets to see if the two studies had comparable results.

As part of their study they briefly investigated the effects of hypomethylation on EB differentiation trajectories by treating EBs with demethylating agent 5-aza for 14 days of differentiation, collecting cells for single-cell sequencing every 2 days. When we compared their gene expression data with our day 14 EB bulk RNA-seq (untreated, low dox and high dox), we found that the genes which were most enriched in their day 14 5-aza-treated EBs were also highly expressed in our high dox EBs, and lowly expressed in our untreated EBs (Fig. 4.19). This shows that two independent studies which induced hypomethylation in developing EBs were able to produce similar results, with similar genes being differentially expressed in both. Our system specifically inhibits *Dnmt1* whereas 5-aza is known for its pleiotropic effects on gene expression, partly as a consequence of cellular toxicity caused by the drug (Kelly, De Carvalho and Jones, 2010). That we had similar findings with a more broad-range drug system for reducing 5mC indicates that the common gene expression changes are indicative of consequences of loss of DNA methylation.

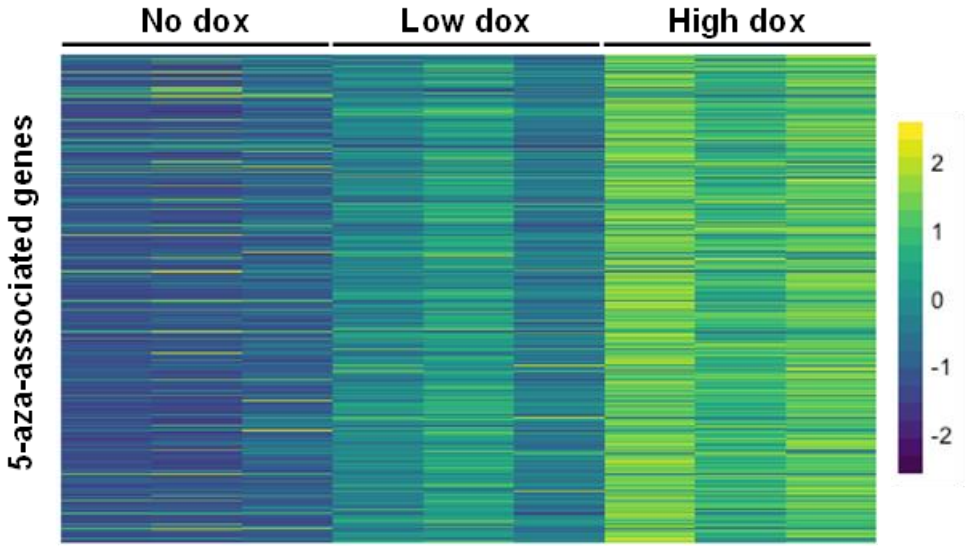


Figure 4.19. The genes most upregulated at day 14 of 5-azacytidine treatment in the scRNA seq. paper (*Kim et al*., 2020) are also highly expressed in our dox-treated EBs, but not in the untreated EBs. This shows that a similar study using an alternative approach to DNA demethylation found similar changes in gene expression as a result of loss of 5mC, supporting our observations.

The scRNA-seq study also made available lists of genes which were highly differentially expressed in their datasets during the early differentiation process (day 0-4), and at late stages (days 8-14). The 'early' list contained genes associated with the first lineage decisions made by ESCs escaping pluripotency, including genetic markers of pluripotency, PGC-like and epiblast-like cells. The 'late' stage list contained genes associated with advanced differentiation states such as primitive streak and germ layer markers. We compared our no, low and high dox transcriptomes to these 'early' and 'late' gene lists, and the results are shown in figure 4.20. What we observed is that the 'early' genes were most highly expressed in the high dox EBs, and lowly expressed in untreated EBs. In contrast, the 'late' differentiation genes were most highly expressed in untreated EBs, and decreased in expression in dox-treated EBs, being particularly lowly expressed in the high dox condition. These heat maps suggest that dox-

treated EBs sustain the expression of genes associated with early stages of differentiation; does this suggest a sub-population of early-stage cells is being maintained throughout differentiation, or is the bulk of the dox-treated EB cell population arresting at this stage? The low expression of late differentiation genes in dox treated EBs suggests the latter, as a mix of sub-populations would lead to enrichment of both datasets in the bulk transcriptomes.

The scRNA-seq paper also provided lists of genes which were used to define each lineage identified in the pseudo-time course. The expression of these genes in our bulk day 14 RNA-seq is shown in Fig. 4.21. In untreated EBs the most highly expressed genes (light green - yellow) roughly fall into the primitive streak (PS), surface ectoderm (S. Ect), mesoderm (Mes) and Definitive endoderm (DE) groups. This correlates well with the previous suggestion that untreated EBs contain cells expressing late-stage differentiation markers belonging to terminally differentiated cell types and their immediate precursors. In the high dox EBs the most highly expressed genes tended to fall in the ESC, pre- and -post implantation epiblast (Pre-Epib, Post-Epib) and primordial germ cell (PGC) groups. These groups can generally be classed as early differentiation lineages. Interestingly, the blood progenitor (BP) group of genes were lowly expressed in both untreated and high dox EBs but were relatively highly expressed in low dox EBs. Overall, these data agree with the hypothesis that dox-treated EBs are enriched for early differentiation genes that are lowly expressed in untreated EBs, which instead show enrichment for markers of late differentiation.

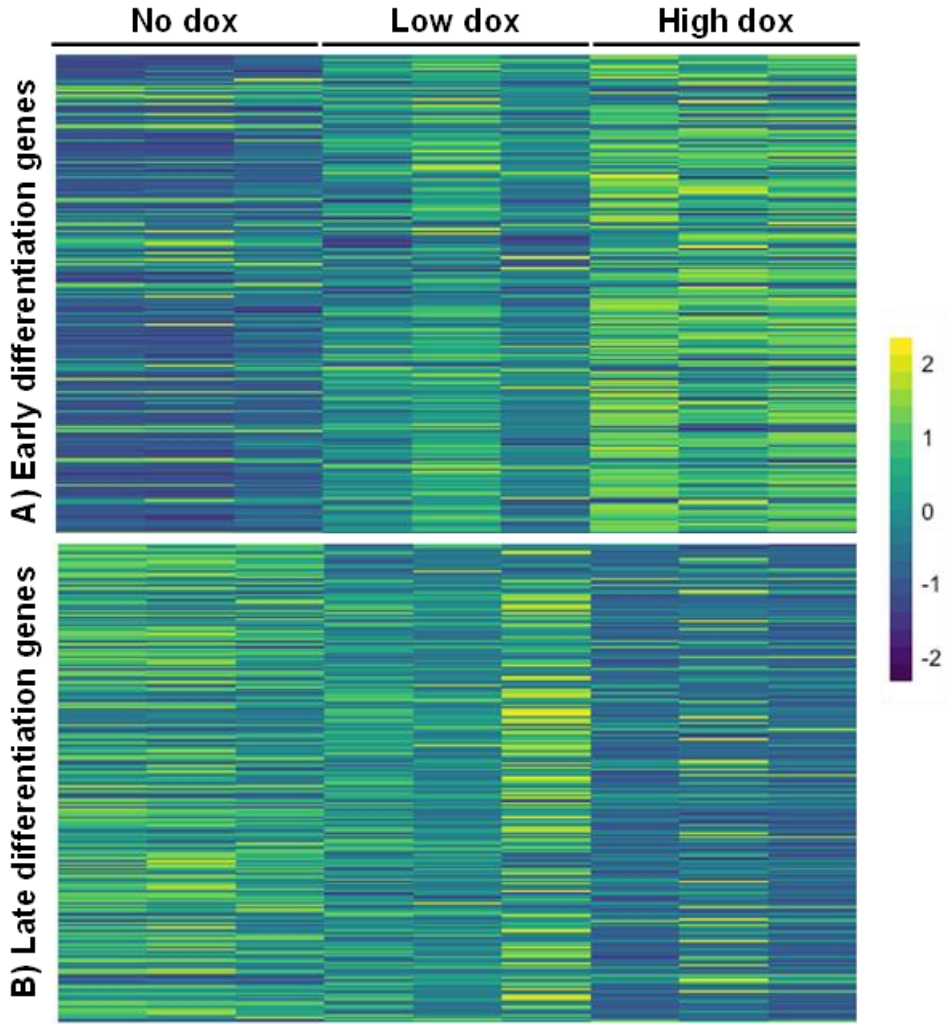


Figure 4.20. Gene expression profiles from EBs differentiated with dox for 14 days showed enrichment for genes associated with early differentiation (A), whereas untreated EBs were enriched for late differentiation genes (B). Bulk RNA-seq data from untreated and hypomethylated EBs collected at day 14 of differentiation were compared to gene lists from a published EB RNA-seq study which determined gene expression profiles for each day of differentiation up to 14 days (Kim *et al*., 2020). These profiles provided lists of genes characteristic of 'early' (days 0-4, from ESC to epiblast stage markers) and 'late' (day 8, from primitive streak to germ layer markers) differentiation.

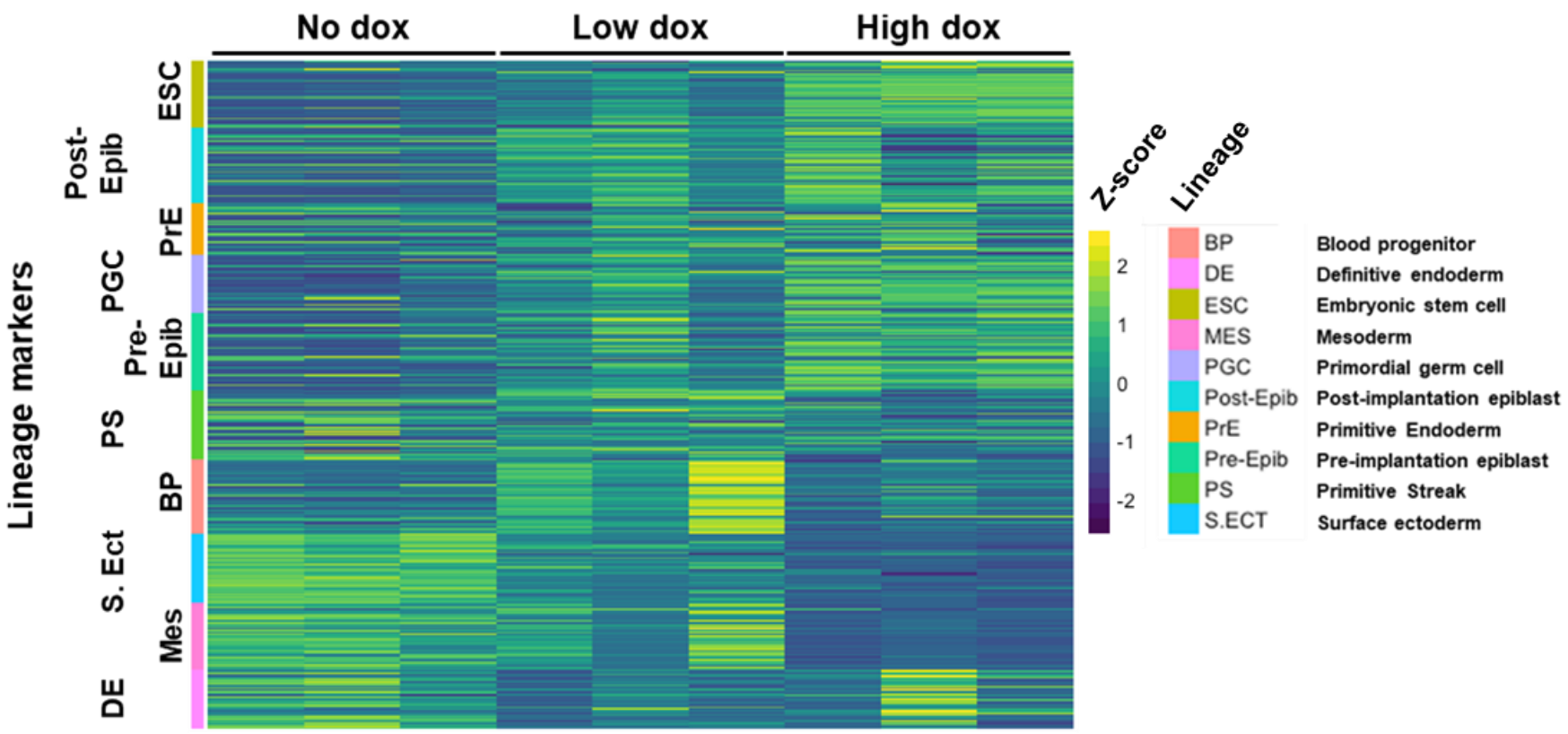


Figure 4.21. EBs differentiated under hypomethylation conditions expressed pluripotency genes associated with ESCs and early epiblast states, and lowly expressed advanced differentiation markers. The opposite was true for untreated EBs. Lists of genes associated with each lineage were obtained from the Kim *et al.* (2020) EB RNA-seq study[25] which determined tissue-specific gene expression profiles over time.

We used the lineage lists for further Gene Set Enrichment Analysis (GSEA) to obtain an overview of any changes in expression of genes associated with the germ layer lineages in our hypomethylated EB differentiation experiments. GSEA is a bioinformatics tool designed to determine if genes associated with a particular phenotype are over- or under-represented within a ranked list of genes obtained from large sequencing experiments (Subramanian *et al.*, 2005). It produces a graphical overview of the distribution of the genes in a pre-defined gene s*et al*ong the ranked background gene list. The background list used was the total list of DE genes (*padj*$\leq$0.05) produced from the RNA-seq experiment ranked from highest to lowest expression, and the pre-defined list of genes used were the lineage lists published in the scRNA-seq paper (Kim *et al*., 2020).

The GSEA algorithm determines enrichment by performing a rolling sum analysis; the position of each gene in the pre-defined list, the fold-change in expression of that gene, and presence of genes not in the pre-defined list contribute to a rolling sum statistic which is illustrated as a green 'Enrichment Profile' line in the final GSEA plot (see figure 4.22). The positions of the pre-defined list genes within the ranked list are shown as black 'hits' bars beneath the Enrichment Profile, and the relative change in expression is shown by a colour scale bar beneath the 'hits'. Red indicates higher expression and positive correlation with the ranked list, and blue indicates decreased expression and correlation. Finally, the ranked genes and their relative expression are shown as grey 'Ranking metric scores' at the bottom of the plot. The final ES is taken as the highest point the rolling sum reached from zero. The GSEA algorithm also determines the null enrichment score, which is used to normalise the ES to give a normalised ES (NES). Finally, the algorithm carries out multiple-testing correction and produces a False Discovery Rate (FDR) *q*-value of significance for the NES. Overall the plot gives a clear overview of positive or negative enrichment of the pre-defined list within the ranked gene list, with a strong peak to the right or left side indicates a strong correlation and high ES, while a random distribution suggests no or weak correlation with a low ES.

In this study, GSEA was used to check for changes in enrichment of germ layer genes in dox treated and untreated EBs to see if hypomethylation disrupted normal differentiation. The pre-defined gene lists were obtained from wild-type EB analyses published by Kim *et al.*(Kim *et al.*, 2020), and enrichment compared between Low or High dox vs No dox conditions at day 14. The results are shown in figure 4.22. In total 13,361 genes were in the ranked background list. There were 44 genes in the pre-defined 'Primordial germ cell' list, and 32 were found in the ranked High dox vs No dox list. The resulting NES was 2.176 (q<0.0001), indicating a strong positive correlation between the 'Primordial germ cell' gene list and the DE genes upregulated in the high-dox EBs. 26 of the genes were identified in the Low dox ranked list, and although a positive correlation was observed the change was more subtle (q=0.051). In contrast, the GSEA analyses for the Blood progenitor/Mesoderm and Surface ectoderm/transition gene lists showed significantly negative correlations for both Low and High dox datasets. In high dox EBs, 88/113 Blood progenitor/Mesoderm genes were detected (NES = -2.95, q<0.0001), and in the Low dox 59/113 (NES = -2.11, q<0.0001). Of the Surface ectoderm/Transition gene list, 45/56 were detected in the High dox dataset (NES = -2.51, q<0.0001), and 37/56 in the Low dox dataset (NES = -2.06, q<0.0001). The list for PrE/Pre-implantation epiblast genes was also analysed but showed a bi-modal distribution with an insignificant NES in both conditions (not shown). Altogether these data suggest that maintaining a hypomethylated state during EB differentiation leads to an increase in expression of PGC-associated genes, whereas genes associated with mesoderm and ectoderm genes are downregulated, suggesting loss of DNA methylation does impact lineage commitment pathways. Kim *et al.* did not include a list of endoderm-exclusive genes in their study, so we cannot investigate the impact on this particular germ layer in this analysis.

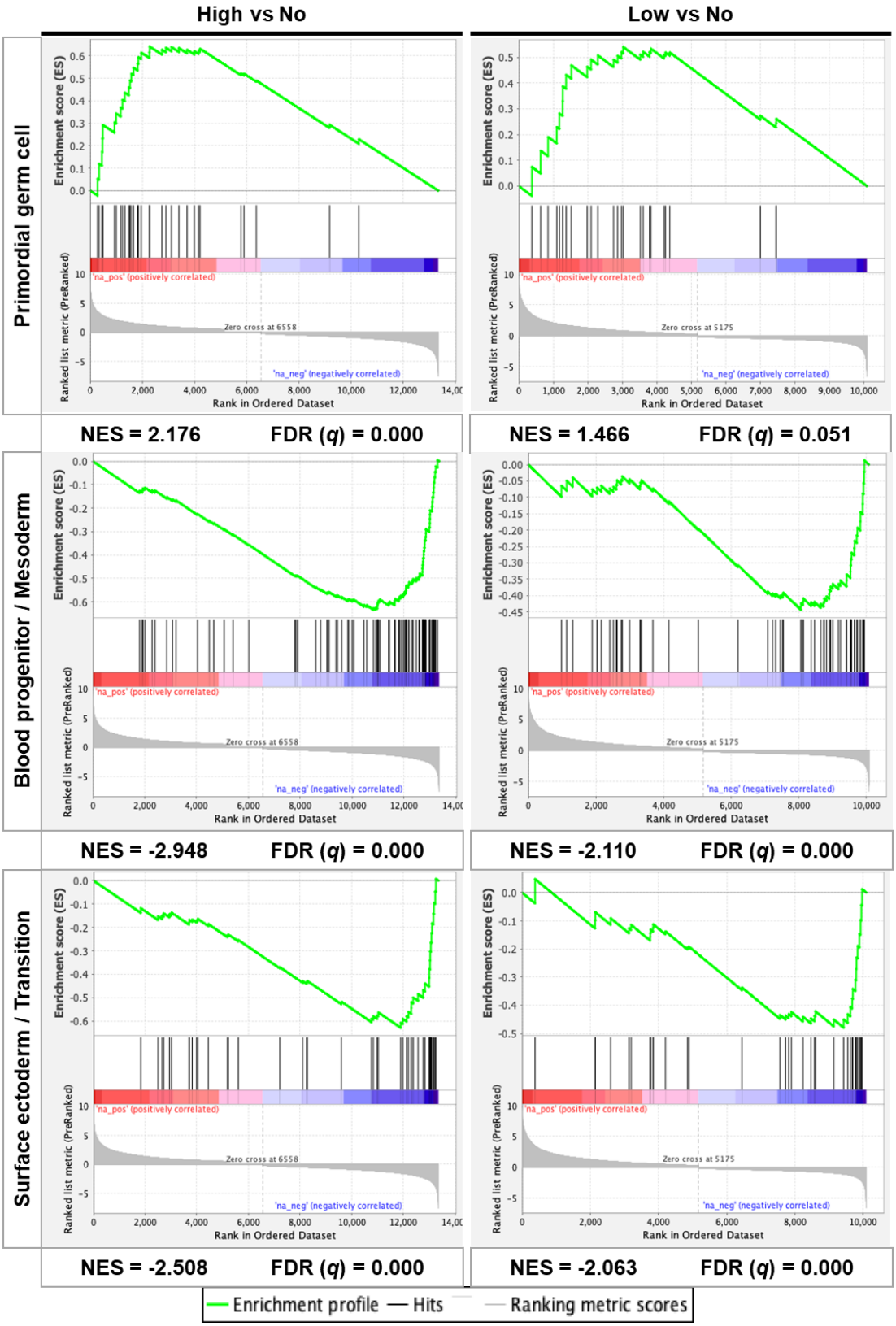


Figure 4.22. GSEA showed dox-treated EBs were positively enriched for PGC-associated genes, but negatively enriched for mesoderm and ectoderm genes. NES = Normalised enrichment score.

4.2.8 Hypomethylation enriched EBs for cells expressing 2C-like markers from as early as day 3 of differentiation

The first division of the zygote produces a 2-cell (2C) embryo which is totipotent and has the ability to form all the tissues of the adult body (Genet and Torres-Padilla, 2020). In comparison, pluripotent cells of the blastocyst can form all tissues of the body except the extra-embryonic placenta. In ESC culture 2C-like cells are very rare, typically making up $\leq$1% of the population (Lu and Zhang, 2015; Genet and Torres-Padilla, 2020). It is thought that all ES cells in culture transiently pass in and out of this state, potentially to repair telomeres (Lu and Zhang, 2015; Genet and Torres-Padilla, 2020). An important feature of mouse 2C cells is that, unlike ES cells, they have high expression of MERVL elements and *Zscan* family genes (Lu and Zhang, 2015; Genet and Torres-Padilla, 2020). In a recent paper, Reik *et al.* (Eckersley-Maslin *et al.*, 2016) isolated MERVL+/Zscan4+ ES cells, which they labelled 2C-like cells, and performed sequencing experiments to obtain a 2C-like cell gene expression profile. In this study we investigated if these genes were present in our EB datasets, and if their expression changed as a consequence of hypomethylation. It was observed that genes associated with the totipotent 2C state, including many members of the *Zscan4* family, were significantly upregulated in hypomethylated EBs compared to wild-type EBs from as early as day 3 of differentiation, as shown in figure 4.23. Here, all genes increased in expression by at least 2-fold, with some increasing as much as 10-fold.

Following this observation a 2C-reporter construct was obtained in which a MERVL element was tagged with GFP, and which constitutively expressed SV40-driven mCherry (Ishiuchi *et al.*, 2015). Working with PhD student Emma Yang, this construct was integrated into *Dnmt1*$^{tet/tet}$ cells by piggyBac cloning to obtain stable reporter line clones expressing low, medium or high levels of mCherry as determined by FACS. A medium-expressing clone was picked for further experiments. Cells were then treated with dox by PhD student Chetan Srinath under the supervision of myself and Dr David Hay; the cells were FACS sorted at day 0, 4, and 7 to see if

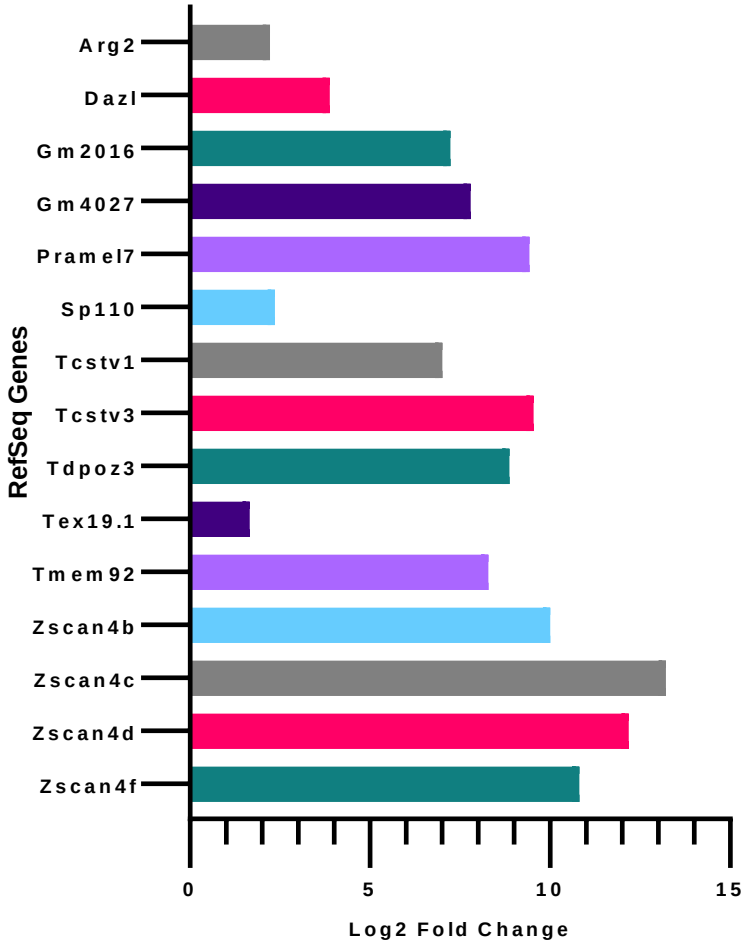


Figure 4.23. Day 3 EBs differentiated under hypomethylation conditions upregulated expression of genes associated with the 2C-like state. Genes shown had a *padj*<0.05.

inducing hypomethylation increased the proportion of MERVL-GFP cells within the population (Fig. 4.24). This work showed that in the untreated cells the reporter activity was present in ~1% of the population throughout the experiment, which is consistent with what has been reported elsewhere for 2C-like cell abundance in ESC populations. Following addition of dox the proportion increased to 16%. However, it was noted that mCherry signal also increased in these cells; it was suggested that under –dox conditions the reporter promoter, which is viral in origin, may have become methylated leading to a discrepancy in mCherry signal between the two treatments. Future work could involve replacing the viral promoter with a non-viral

alternative and seeing if the results change. In future work we would like to perform qPCR with these cells and analyse expression of 2C marker genes, as well as 5mC-sensitive gene *Dazl*, to further validate the 2C-like state of these cells at the gene expression level. I would also like to use this system to investigate if the 2C-like state is inherited by daughter cells when MERVL-GFP+ cells divide and to quantify how long the GFP+ signal is expressed as an indicator of the lifetime of the 2C-like state.

Furthermore, Kim *et al*. (Kim *et al.*, 2020) did not look for 2C gene enrichment in their EB scRNA-seq 5-aza data. In future work we would like to investigate that ourselves and see if it correlates with our data. This will also provide an independent 2C-signature list to complement the one published by Eckersley-Maslin *et al*. (Eckersley-Maslin *et al.*, 2016), hopefully showing the same patterns of gene expression.

4.2.9 Hypomethylation may trigger an alternative differentiation trajectory, leading to smaller EBs

It was observed that at day 3 of EB differentiation dox-treated EBs had increased expression of transcription factors, particularly of the *Gata* family, which are associated with somatic differentiation (Fig. 4.25). This suggested that hypomethylation could be causing an early departure from the wild-type differentiation trajectory. This may partly explain the observed EB size phenotype, as somatic cells have a significantly elongated cell cycle compared to pluripotent cells; they divide once every 24 hours or more as opposed to once every 8-10 hours for pluripotent cells (White and Dalton, 2005). An accumulation of these differentiating, slow-cycling cells early on in EB development would lead to an overall reduction in total cell number compared to wild-type EBs of the same age, and this would result in EBs which are smaller than the wild-type.

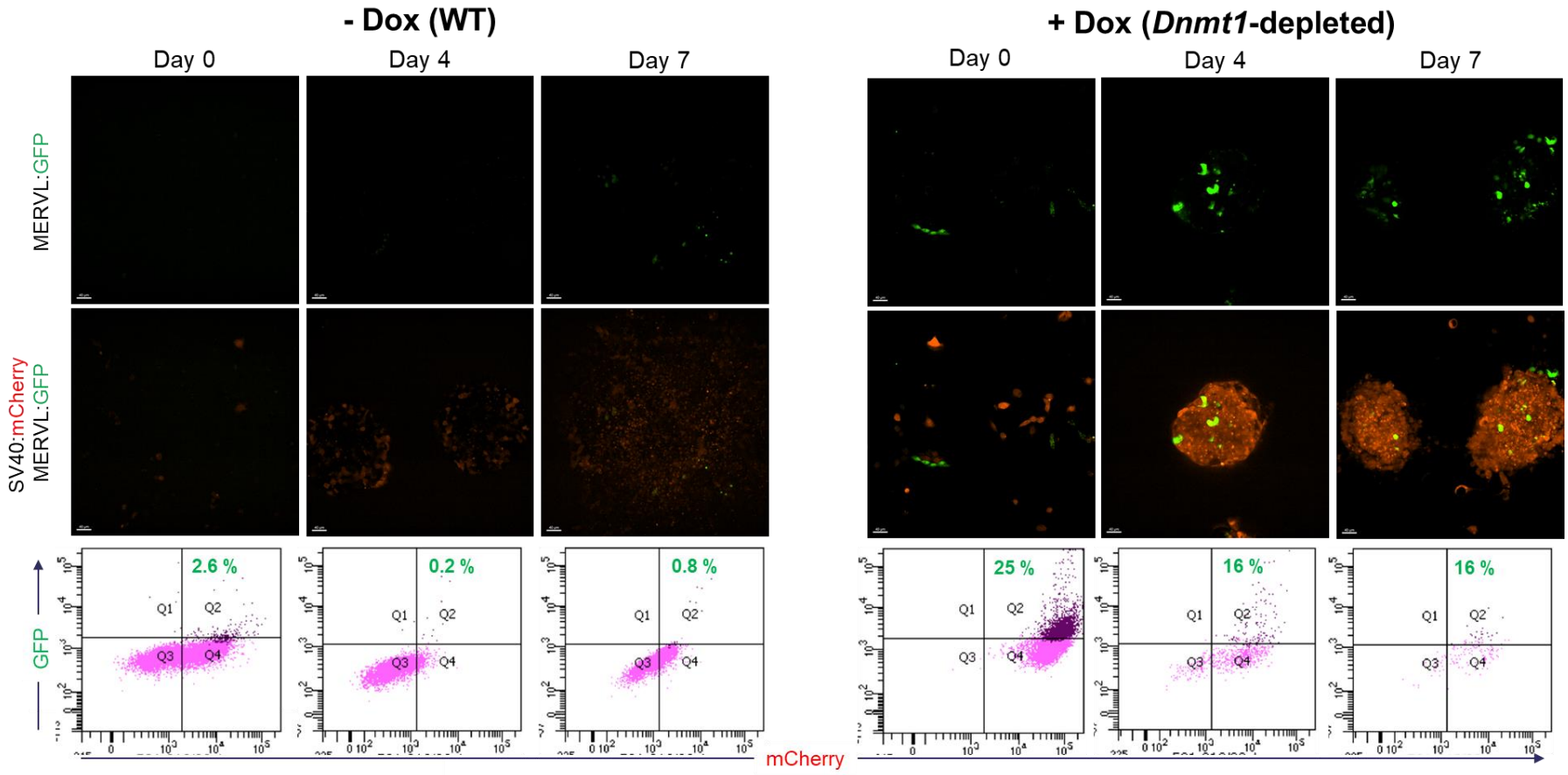


Figure 4.24. Cells expressing the 2C-state marker MERVL-GFP occurred more frequently in *Dnmt1*-depleted embryoid bodies. The reporter signal was observed in live cells by microscopy and quantified in fixed cells by FACS. The cell line was prepared by Emma Yang with supervision from myself and Dr David Hay. The FACS work was performed by Chetan Srinath under the supervision of Dr David Hay. Scale bars represent 40μm.

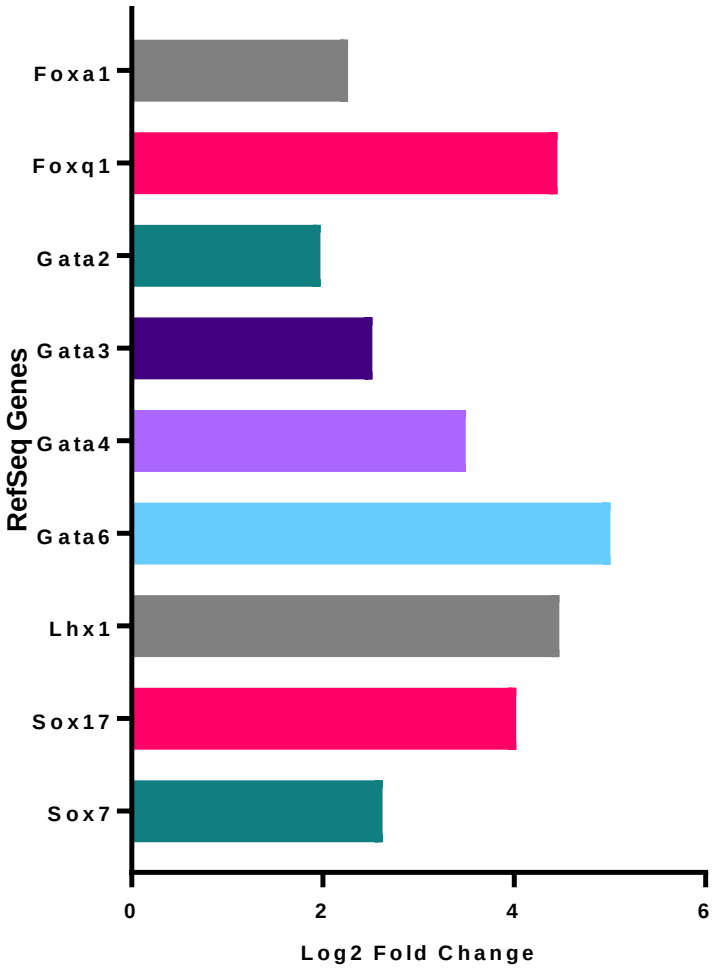


Figure 4.25. Day 3 EBs differentiated under hypomethylation conditions upregulated transcription factor genes which suggest an early departure from the wild type differentiation trajectory. All *padj* <0.05.

4.2.10 An updated model for the role of DNA methylation during EB differentiation

Under normal embryoid body differentiation conditions, EBs increase expression of differentiation genes, increase deposition of 5mC, and decrease expression of pluripotency transcription factors relative to the original ESCs. This combination of changes brings about the emergence of cells committed to one of the three germ layer differentiation pathways (mesoderm, endoderm, and ectoderm). Similarly to ESC cell culture, the numbers of 2C-like and PGC-like cells in wild-type EB cell populations remains low (<1%). EBs grow rapidly during the first 14 days of culture, doubling in size every few days, and the number of visible dead and dying cells is relatively low. In contrast, when hypomethylated ESCs are released from LIF culture and allowed to form EBs we see significant phenotypic and genetic changes. EB growth rates approximately halve leading to significant reductions in size compared to wild-type EBs of the equivalent age. Hypomethylated EBs are also much less adherent, with many floating single cells coming off the main body suggesting an increase in apoptosis. I have shown that both differentiation and pluripotency genes are expressed when normally they show inverse expression patterns. By the end of Chapter 3, questions remained as to how 2C, PGC and lineage specific genes were being affected.

In this chapter further molecular experiments were performed to try to identify the genetic cause of the observed phenotype. Results suggested an increase in frequency of cells expressing 2C-like and PGC-like markers, indicating absence of 5mC during differentiation was enabling increased expression of these gene groups, enriching the EB populations for these previously rare cell types. Apoptotic pathway gene profiles were also investigated. No enrichment for these were found, which is in agreement with the TUNEL assay performed in chapter 3 in that no significant difference in apoptotic double-strand breaks between dox treated and untreated EBs was seen. It was also observed that no change in frequency of pyknotic nuclei or nucleosomal laddering were present in cells or DNA digests throughout the

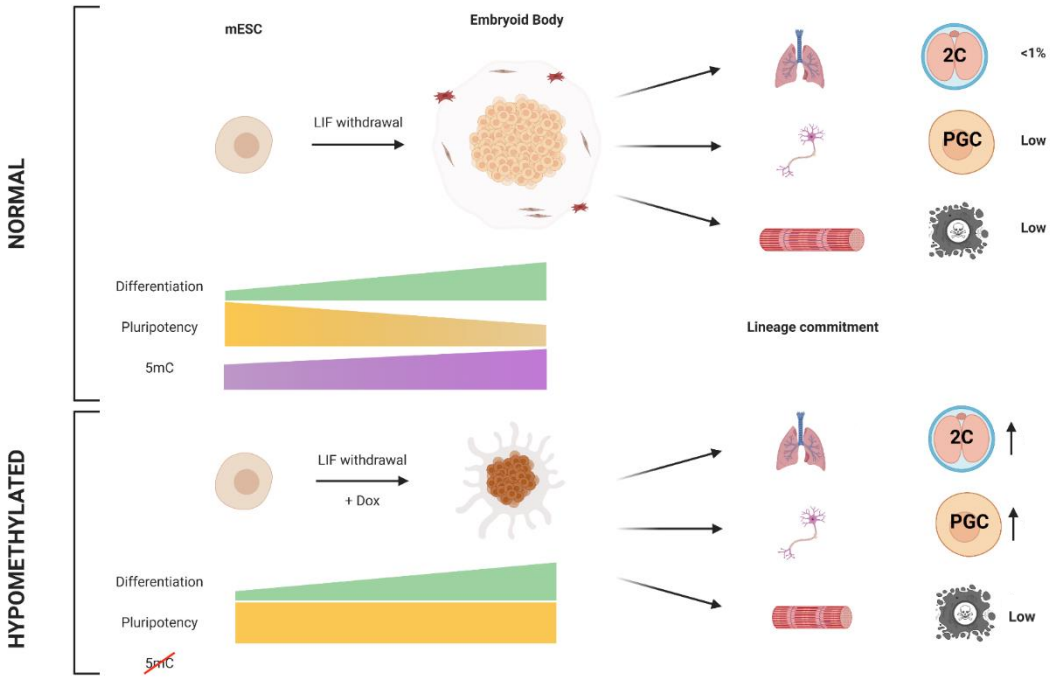


Figure 4.26. A model for the role of *Dnmt1* during EB differentiation. The data so far suggests that inhibition of *Dnmt1* expression during EB differentiation prevents the silencing of pluripotency factors and leads to an increase in PGC- and 2C-like gene expression. Germ layer differentiation is not inhibited, and the incidence of apoptosis does not change significantly.

study (Fig 3.11). Overall, these results suggest that while hypomethylation leads to an increase in germ cell and totipotent state gene expression programmes, it does not lead to increased cell lethality via apoptosis (Fig. 4.26). The phenotype observed may instead be a consequence of down-regulation of cell-cell adhesion genes, whose expression was shown to be upregulated in GO analyses.

4.3 Discussion

In this chapter I aimed to investigate the impact of absence of *Dnmt1* during early and late EB differentiation at the gene expression level using RNA-seq to build on the model established at the end of Chapter 3 (Figure 3.21). I found that the global gene expression profiles of wild-type and hypomethylated EBs were very different, particularly at day 14 (Figures 4.5 and 4.14). However, these changes did not significantly impact germ layer differentiation pathways; beating cardiomyocytes formed under hypomethylation conditions, and signature genes all three germ layers were still expressed albeit at lower levels to the wild-type (Figure 4.16). The most significant changes affected genes associated with meiotic pathways and TE repression (Figures 4.10, 4.11 and 4.17). I also investigated apoptosis gene expression in an attempt to address the size disparity between wild-type and hypomethylated EBs; I found no enrichment for these genes (Figure 4.15, heat map C), suggesting cell death is not the primary cause of the observed phenotype. Overall, this led to the hypothesis that Dnmt1-dependent 5mC is required for regulating PGC- and 2C-like gene expression programmes during EB development as loss of the mark leads to upregulation of these gene signatures. This is summarised in figure 4.26.

In 2020 a large study investigating gene expression in EBs was published by Kim *et al.*(Kim *et al.*, 2020). They tracked differentiation pathways using scRNA-seq and constructed a pseudo-time tree highlighting lineage commitment decisions over time. They then identified the genes associated with each branch point. Using their lists I was able to look for enrichment of the various lineages, including PGC signatures, in my EB bulk RNA-seq datasets. They also explored the impact of 5-azacytidine on EB differentiation, and I was able to compare their findings from single cells with my bulk EB conditional *Dnmt1* KO data.

Considering the evidence so far from the RNA-seq experiment and looking at the most highly expressed genes across the EB datasets, it appears that although there are clear differences in EB size and gene

expression profiles the broad patterns of differentiation themselves are not significantly different. This leads to the conclusion that although marker gene expression is reduced relative to the wild-type they are not lost completely, suggesting DNA methylation is not required for germ layer differentiation to occur.

So, what causes the differences in size between hypomethylated and normally methylated EBs? Initially it was hypothesised that increased levels of apoptosis were occurring as a product of co-expression of pluripotent and differentiation signalling pathways. However, comparing the gene expression profiles did not indicate a significant change in expression of apoptosis or necrosis related genes. There was no evidence of a relative prevalence of pyknotic nuclei in hypomethylated EB cells, nor was there any indication of nucleosomal DNA laddering in DNA digests run on a gel (see chapter 5). In Chapter 3 I tried to quantify the proportion of apoptosis-related DNA breaks by TUNEL assay; although the assay itself required further optimisation, initial experiments did not indicate a significant difference in FITC signal between treated and untreated EBs which is in agreement with the RNA-seq analysis results presented in this chapter. Overall, I conclude that apoptosis is unlikely to be the cause of the differences in size of wild-type and hypomethylated EBs.

An alternative hypothesis is that loss of DNA methylation leads to a change in the cell cycle as Dnmt1 is strongly associated with recruitment of replication machinery to replication forks (Schermelleh *et al.*, 2007). Perhaps the absence of Dnmt1 leads to replication fork stalling or slowing. This could be addressed in future work by introducing a catalytically inactive Dnmt1 mutant rescue and investigating if the reduction in EB size still occurs.

It may also be that the cell cycle itself is affected. During normal development the exit from pluripotency is associated with a lengthened cell cycle, particularly the G1 phase. In EBs this could lead to a reduction in growth rate over time, as a greater number of cells entering a somatic state with lengthened cell cycle and fewer cell divisions would lead to a reduction in the population of rapidly dividing pluripotent cells and thus overall

reduction in EB size in a process similar to that observed in primordial dwarfism (Klingseisen and Jackson, 2011).

Another hypothesis to explain the size discrepancy observed is that in the absence of Dnmt1 differentiation pathways are activated earlier; this would lead to a greater proportion of cells within each EB having an increased cell cycle length as they exit pluripotency prematurely which would appear to reduce the overall growth rate of the entire EB. From the experiments so far it is not clear how cell cycle regulators are affected. Interestingly, at day 3 multiple important growth factors were differentially expressed in hypomethylated EBs such as *Gata2/3/4/6*, *Sox7/17*, and *Foxa* and *Hnf1* genes. This supports the hypothesis suggested that inhibition of *Dnmt1* expression leads to early initiation of differentiation pathways, which as suggested earlier may result in a greater number of EB cells increasing their cell cycle length as they exit pluripotency which would consequently lead to smaller EBs. It is also worth noting that key growth factor *Igf2* is downregulated in day 14 EBs. In future work, the contribution of *Igf2* expression in EB expansion could be tested by adding *Igf2* to the culture media in addition to dox and seeing if this leads to an increase in hypomethylated EB size.

At the end of this chapter I propose a change to the model presented at the end of Chapter 3 (Fig. 3.21). The earlier model summarised the observations from the first hypomethylation differentiation experiments in Chapter 3; wild-type EBs grown in the absence of dox form large, spread colonies of cells of all the three germ layers, including visibly beating cardiomyocyte cell clusters. In contrast, dox-treated ESCs formed small, compact EBs which appeared blackened and formed smaller, less complex clusters of beating cardiomyocyte cells. At this stage it was unclear if hypomethylation affected formation of any of the three lineages, or if it increased incidence of cell death.

Following the Southern blot result in Chapter 3, it was suggested that wild-type EBs undergo a wave of demethylation during early differentiation

which it was hypothesised may have indicated the transient emergence of hypomethylated 2C-like or PGC-like cells, raising the question of whether DNA methylation has a role in controlling the emergence of these cell types. I found both 2C and PGC genes were enriched in the hypomethylated EBs from as early as day 3 of differentiation. Fu *et al*. (Fu *et al.*, 2019, 2020) have shown that *Myc* and *Dnmt1* impede the pluripotent-to-2C transition; this complements our findings, as conditional inhibition of *Dnmt1* led to an increase in the prevalence of MERVL-GFP+ cells in EB culture.

However, in this chapter I also investigated the impact of hypomethylation on TE expression. It was observed that between 3-14 days of EBs differentiation LINE and IAP-family TEs become significantly upregulated compared to the wild-type. It was hypothesised that an increase in retrotransposition activity could have contributed to the size difference between treated and untreated EBs, however, as has been discussed there was no observable increase in apoptosis which would be expected as a consequence of TE activity. Interestingly, this increase in TE activity, particularly the LINE1 elements, may contradict my previous 2C-like cell enrichment hypothesis; LINE1 expression has recently been implicated in the regulation of ESC proliferation by blocking *Dux* expression, a master transcription factor-encoding gene which maintains the totipotent 2C-like state (Percharde *et al.*, 2018; Mangiavacchi *et al.*, 2021). This contradicts our observation that more 2C-like cells are present in hypomethylated EBs as de-repression of LINE1 TEs should have a repressive effect on this signalling pathway.

The aim of this chapter was to address some of these questions using RNA sequencing and following up with more targeted molecular experiments. By the end of the chapter it has been determined that germ layer differentiation still occurs in the absence of DNA methylation. Also, in line with the tentative results of the TUNEL assay in chapter 3, no enrichment of cell death pathways was found in the hypomethylated EB gene expression profiles. As an alternative hypothesis for the size disparity between hypomethylated and wild-type EBs, it was posited that dox-treated EB cells

have an altered cell cycle which leads to an overall reduction in cell numbers and therefore EB size. In line with this, evidence was found for an earlier differentiation trajectory with changes in transcription factor expression starting as early as day 3. This fits with the observation that the size disparity becomes significant from 4-5 days of differentiation (Figure 3.20). Early down-regulation of key growth factors, such as *Igf2*, may also be a contributing factor. Finally, I have shown evidence that differentiating in the absence of *Dnmt1* leads to an increase in 2C- and PGC-like cell signature genes, with the proportion of MERVL+ cells increasing and hypomethylated EBs being enriched for GO terms associated with meiosis.

In summary, DNA methylation may serve to suppress expression of 2C- and PGC-like cells and to prevent premature expression of genes associated with late-stage differentiation. Perhaps 5mC is required to allow PGC formation by strictly partitioning PGC signalling from differentiation and pluripotency signalling pathways, but why can't these occur at the same time? Perhaps a threshold number of cells needs to be reached first before the EB as a whole can start along the next lineage trajectory as depicted in the Kim paper. Another question is raised, why are 2C cells so rare in both pluripotent and differentiating cell populations?

In the next chapter I will try to clarify some of the observations made earlier in the thesis. Chiefly, I will try to determine if the aberrant expression of pluripotency genes occurs within the same cells which are expressing differentiation genes using fluorescence in situ hybridization (FISH) imaging. I also knock out *Dnmt1* in cell lines with conditional pluripotency genes to try to determine if the same gene changes occur as were presented in this chapter. Finally, I use an alternative 5mC reporter cell line to investigate changes in MBD distribution as a consequence of *Dnmt1* KO.

Chapter 5 | Investigating the interactions between DNA methylation and pluripotency networks during EB development

4.4 Introduction

In Chapter 3 I tested an *in vitro* model of early embryonic development, embryoid body formation, using mouse ESCs with drug controllable *Dnmt1* expression so I could investigate the impact of loss of DNA methylation on ESC differentiation. I observed that absence of Dnmt1 led to a significant reduction in EB size and caused changes to gene expression patterns, particularly of pluripotency factor *Oct4*. In Chapter 4 I explored the changes in gene expression in greater detail using RNA sequencing and by comparing data with a published single-cell RNA sequencing study. From this I gained a global overview of the effects of hypomethylation on EB transcription and development during early and late stages of differentiation. I was able to determine that the size differences between wild-type and hypomethylated EBs were unlikely to be a result of apoptosis as no enrichment of genes associated with the intrinsic apoptosis pathway were found.

Throughout, I observed that cells from all three germ layers were still able to form in hypomethylated EBs, suggesting DNA methylation is not an essential requirement for somatic differentiation to occur. Instead, it was found that transcription factors strongly associated with growth and development, such as *Igf2*, were being aberrantly expressed from day 3. That these genes were expressed from this early time point suggested that hypomethylated EBs had an altered differentiation trajectory, embarking towards a somatic-like state earlier than the wild-type EBs. It was also found that at both early (day 3) and late (day 14) differentiation points hypomethylated EBs had much higher levels of expression of meiosis-related genes, including those associated with 2C-like cells and PGC-like cells; particularly differentially expressed genes included *Zscan*-family genes and imprinted genes such as *Rian* and *Meg3*. These changes were occurring at the same time as key pluripotency factors, including *Oct4* and *Sox2*, were being aberrantly expressed. It is thus far unclear what effect simultaneous expression of pluripotency and differentiation gene networks has on EB development, as it was not associated with an increase in rates of apoptosis.

Along with *Sox2* and *Nanog*, *Oct4* (also known as *Pou5f1* or *Oct3/4*) makes up the core pluripotency network required to prevent differentiation and maintain self-renewal in ESCs (Pauklin, Pedersen and Vallier, 2011; Chen *et al.*, 2017). *Oct4* is considered to be the master factor as it acts upstream of the other pluripotency network genes and loss of its expression leads to spontaneous differentiation. Overexpression of *Oct4* is also able to compensate for the loss of *Sox2* expression (Chen *et al.*, 2017). It is highly expressed in ESCs, unfertilized oocytes, PGCs, epiblast stem cells and some cancer types, such as embryonic carcinoma cells, but it is rarely expressed in somatic cell types (Patra, 2020). *Oct4* and *Sox2* complex together and control other downstream pluripotency factors, such as *Nanog*. As such they are strongly associated with the pluripotent cell state. They control this state by regulating the expression of many genes, including important growth factors such as *Fgf4* and *Tdgf1*.

As a result of the work so far, key open questions have been raised: in the absence of *Dnmt1*, are the pluripotency genes and differentiation genes being expressed in the same cells or did they represent sub-populations within the EBs? Does loss of 5mC lead to sustained pluripotency network expression? Does loss of a key pluripotency gene affect EB development? Finally, what happens to EBs when these master genes are not expressed? I attempt to address these questions in this chapter using new conditional cell lines in which I can control the expression of *Oct4* or *Sox2* using doxycycline. I further manipulate these lines with CRISPR genome editing to remove Dnmt1 and analyse the impact of single or double loss of the pluripotency gene and DNA methylation using fluorescence imaging techniques. With these experiments I aimed to further resolve how these pathways interact during EB differentiation, and the results then form the end of this thesis.

5.2 Results

5.2.1 Pluripotency marker Oct4 and differentiation marker Gata4 occupy different territories within EBs

In earlier chapters I showed by qPCR and RNA sequencing that key pluripotency factors such as *Oct4* (*Pou5f1*) and *Sox2* are expressed in late-stage hypomethylated EBs (day 14 of differentiation) whereas they are lowly expressed in the wild-types. It was unclear if this aberrant expression was connected to the observed phenotype; the hypomethylated EBs grew significantly more slowly from day 4 (Fig. 3.14 and 3.20), and by day 14 were approximately half the size or less of the untreated EBs. However, these smaller EBs still expressed genes of the three germ layer lineages, showing somatic differentiation was not affected by the aberrant presence of the pluripotency factors or absence of DNA methylation. As 5mC is generally associated with silent genes, I hypothesised that the mark may be required to silence certain gene expression pathways upon initiation of differentiation to prevent conflicting systems being co-expressed. However, the maintenance of expression of the core pluripotency network genes beyond the stage they were normally silenced did not correlate with an increase in apoptotic cells. Next, I sought to determine if removal of both Dnmt1 (and subsequently DNA methylation) and a key pluripotency factor such as Oct4 or Sox2 would then be sufficient to induce apoptosis in this model as both networks are required for proper embryonic development (Patra, 2020).

Having explored the changes in patterns of gene expression at the molecular level in *Dnmt1*$^{tet/tet}$ EBs with and without *Dnmt1* expression, I next sought to investigate how these genes were expressed in single cells within the whole EB context by immunocytochemistry (ICC). I aimed to obtain images of these EBs which had been stained with markers for each of the three germ layers to obtain an overview of how the cells of these tissues were distributed throughout a wild-type EB and if this changed in EBs grown under hypomethylation conditions. I would then go on to co-stain EBs for pluripotency markers along with a differentiation marker to determine if there was co-localisation of the proteins within single cells or if they represented

distinct sub-populations. The results of the earlier molecular analyses suggested that the pluripotency markers may be being aberrantly expressed in cells of hypomethylated EBs which were expressing differentiation markers due to the inability of the cells to silence the pluripotency network in the absence of Dnmt1. However, I earlier hypothesised that if these two contradictory gene networks overlapped they would trigger intrinsic apoptotic pathways, but when I performed tests to detect apoptotic markers I did not observe enrichment for apoptotic cells in the hypomethylation conditions. It may be that the cells can tolerate co-expression, or that the networks were not being co-expressed in the same cells. My previous molecular experiments were performed on bulk EB tissue extracts so could not provide information on the content of single cells. Thus, I decided to use ICC to obtain more single-cell resolution data on gene expression patterns while the cells are still within the EB context.

To begin with, I cultured *Dnmt1*$^{tet/tet}$ EBs for 14 days with or without dox on glass culture plates, treated them with cytoskeleton (CSK) buffer (Dunican *et al.*, 2013) (see materials and methods), fixed them with PFA and stained them with antibodies targeting a range of lineages. CSK buffer was used to remove loosely bound nuclear proteins from the chromatin to try to improve the specificity of antibody binding and the resolution of the final images. EBs were stained with antibodies targeting core pluripotency proteins (Oct4), proteins associated with specific differentiation lineages (Gata4, endoderm/mesoderm marker; Acta/2, mesoderm marker; Cdh2 and Ncam1, ectoderm markers), and proteins associated with general differentiation (Fos and Fosb). Following antibody testing, only the Oct4 and Gata4 antibodies were found to reliably stain when compared to secondary-only control stains (not shown). The results, previewed in chapter 3, are shown in full here in figures 5.1 A and B.

The staining showed that Oct4 could be detected in both the wild-type and hypomethylated EBs (Figure 5.1 A). In both conditions single Oct4+ cells can be seen, and overall these cells tend to occur in patches within the central regions of the EBs. Though I expected there to be

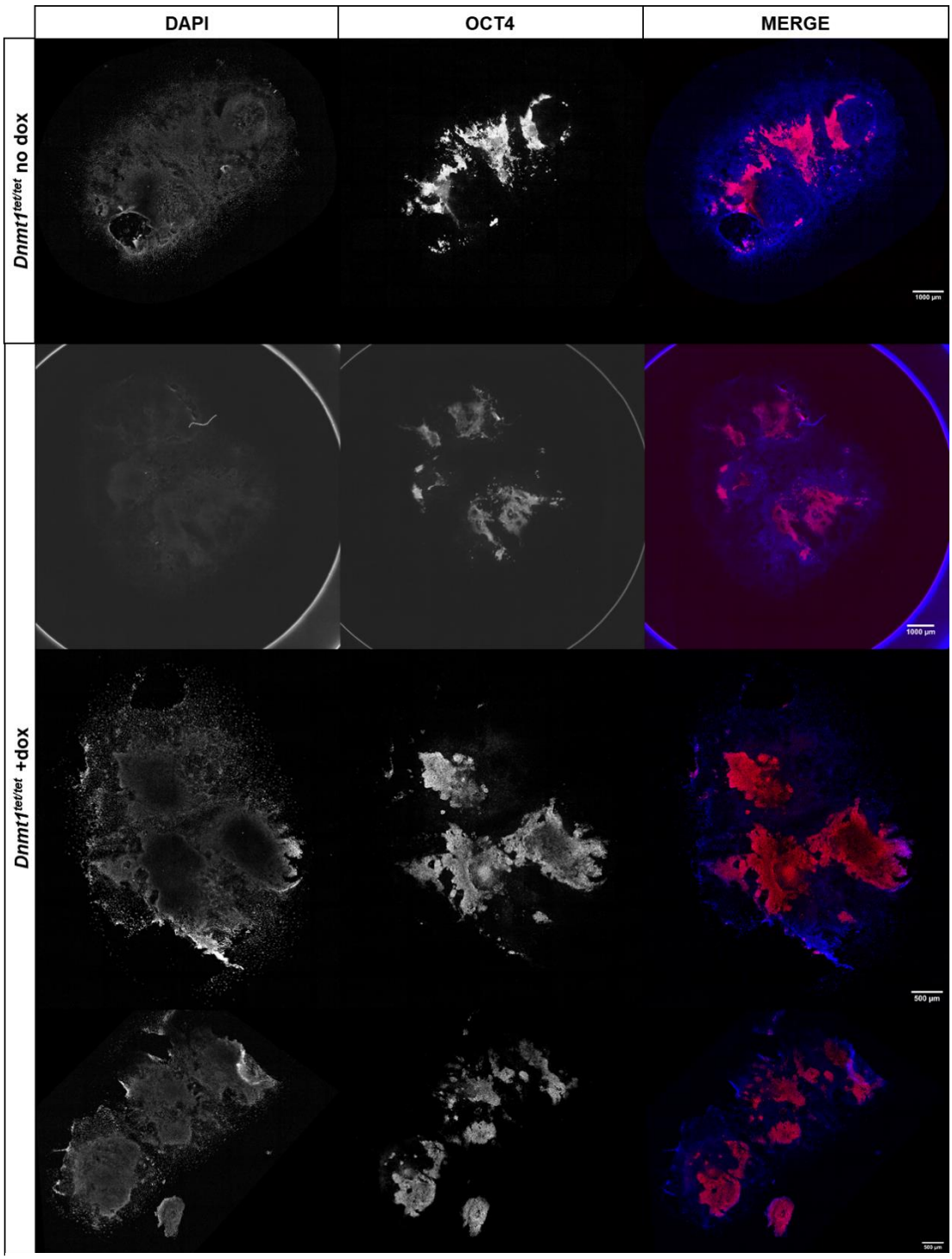


Figure 5.1 A. *Dnmt1*$^{tet/tet}$ EBs with and without Dnmt1 maintain populations of Oct4+ pluripotent cells at late differentiation. These populations consistently occur in the central areas of the EBs. EBs were fixed at day 14 of culture. Scale bar units are shown within each panel.

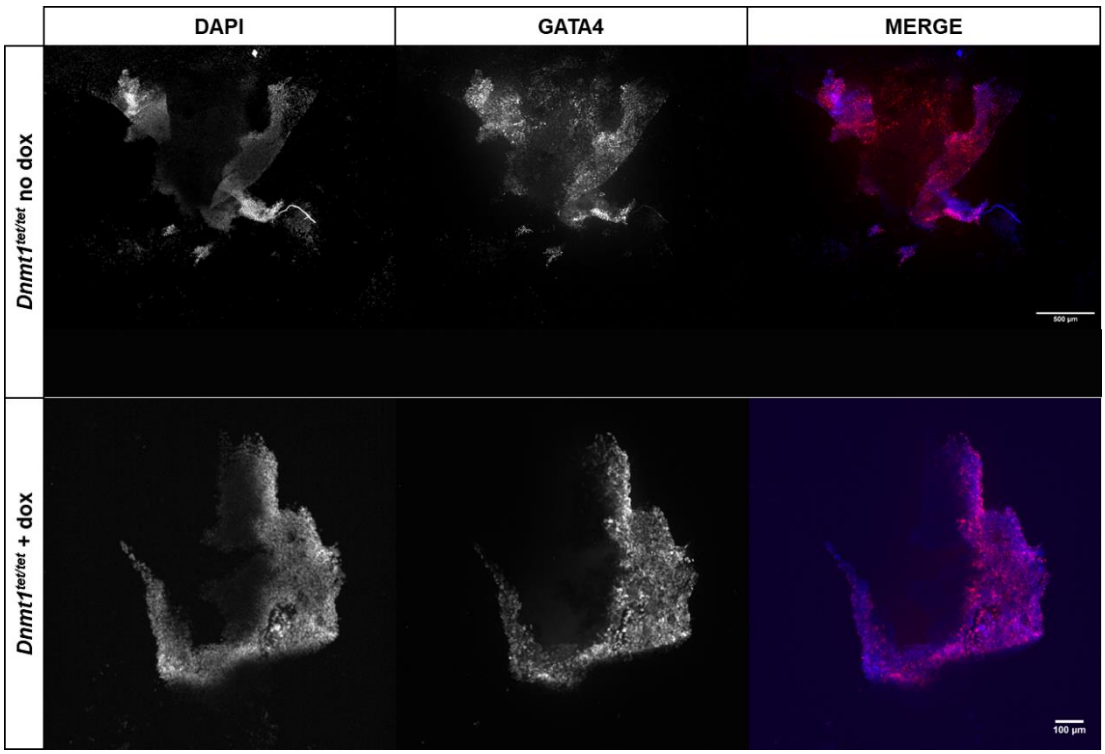


Figure 5.1 B, part 1. *Dnmt1*$^{tet/tet}$ EBs with and without express *Gata4* to similar levels. Gata4+ cells occur throughout the EBs and are not limited to the central areas like Oct4. EBs were fixed at day 14 of culture. Scale bar units shown in the panels.

significantly more Oct4+ cells in the hypomethylated (+dox) EBs at day 14, populations can be seen in the wild-type as EBs and embryos maintain populations of pluripotent cells to provide a pool from which differentiating cells can emerge. These ICC differences for *Oct4* expression are not as distinct as the observations from the qPCR and RNA-seq analyses performed in earlier chapters, however, in each assay *Oct4* was lowly expressed in the wild-type EBs at day 14 with more expression in hypomethylated EBs. This observation has been validated independently by Olga Stepanova in her PhD project. I had expected that even if a few cells within each EB expressed *Oct4* this signal would be a relatively small proportion of the overall pool of transcripts. However, the proportion of Oct4+ cells seen here in the ICC experiment do not support that view, as it appears that Oct4+ cells are present in both EB conditions. I believe the results of the RNA sequencing more accurately represent the *Oct4* expression in the EBs.

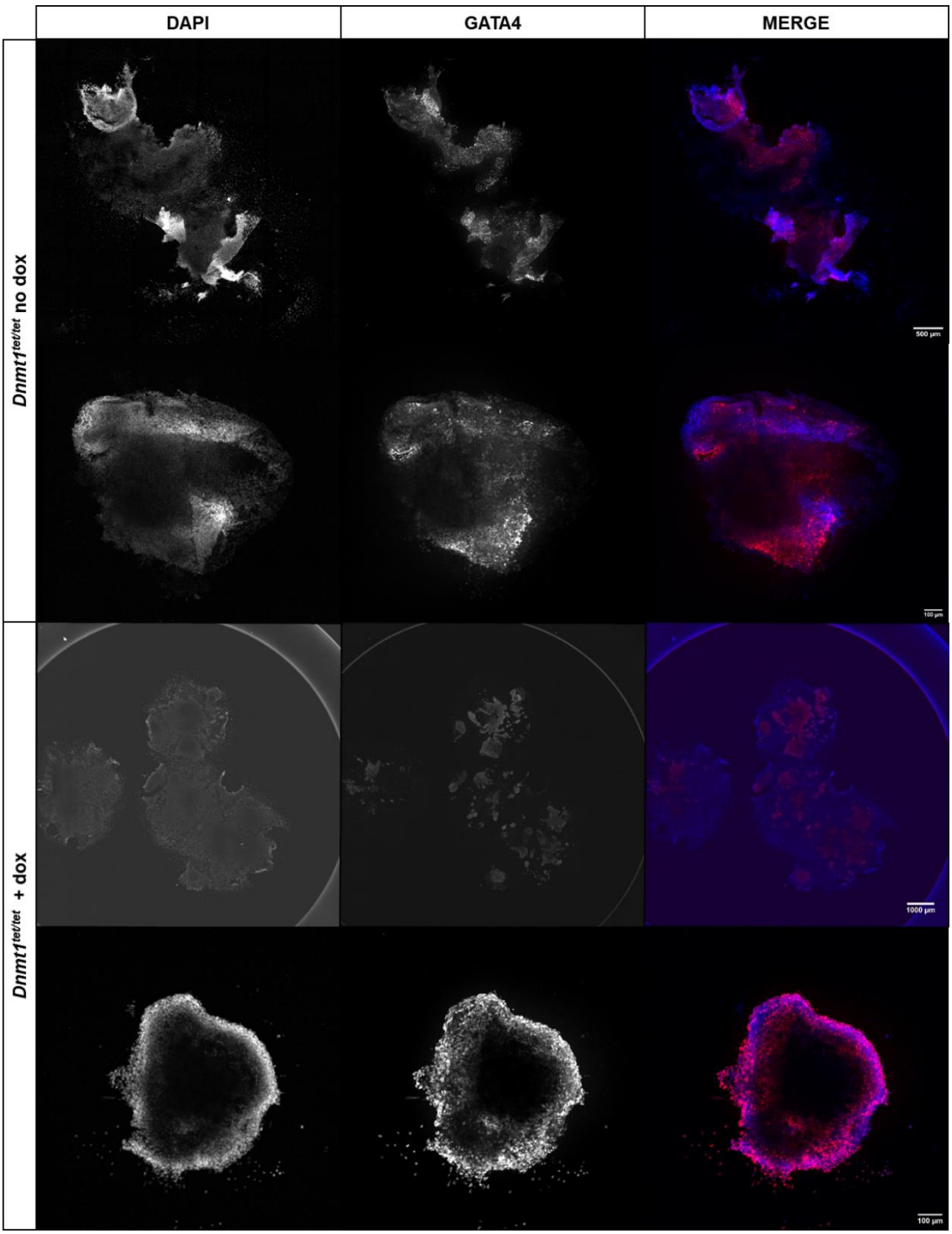


Figure 5.1 B, part 2. *Dnmt1*$^{tet/tet}$ EBs with and without dox express *Gata4* to similar levels. Gata4+ cells occur throughout the EBs and are not limited to the central areas like Oct4. EBs were fixed at day 14 of culture. Scale bar units shown in the panels.

The Gata4+ cells were also similarly abundant in both conditions (Figure 5.1 B), which supports the earlier observation that cardiomyocyte differentiation still occurs in dox-treated EBs. Broadly, the Gata4+ cells tended to occur towards the edges of the EBs in comparison to the Oct4+ cells which were concentrated in central regions. This supports the hypothesis that cells exiting pluripotency migrate towards the outer EB boundary, perhaps to leave room for a population of pluripotent cells to be maintained at the centre of the mass to continually produce more pluripotent cells as the population is lost to differentiation. This is reflective of the re-organisation of differentiating cells from the pluripotent inner cell mass of developing embryos to form the specialised tissues, such as the specification and separation of primitive endoderm and epiblast cells (Chazaud and Yamanaka, 2016; Fu, Zhang and Zhang, 2020).

It should be noted that in this experiment the wells contained multiple EBs, so it's difficult to distinguish them and thus quantify their individual sizes and proportion of each stain. This is important as what appears to be one large EB may be a cluster of smaller EBs as they do have a tendency to group together while still floating following transfer from the suspension culture. From a general overview of the images, it appears the proportions of cells per EB stained with each antibody are similar between conditions, suggesting inhibiting *Dnmt1* expression does not affect *Gata4* expression or the size of the pluripotent Oct4+ cell population in late-stage EBs. This is contrary to previous suggestions in this thesis, such as the data from the qPCR analysis in Chapter 3.

Unfortunately for the ICC analysis, one of the features of Dnmt1-deficient EBs is a loss of cell adhesion resulting in much looser structures. This becomes apparent during staining when many EBs become wholly or partially detached from the plate, and is exacerbated by growing them on glass which they adhere to poorly compared to plastic culture plates, despite both being coated with gelatin. Upon imaging it can be seen where some EBs are missing sections, or some simply detach at the edges and roll up

(see Figure 5.1 B top panel for an example). This makes obtaining consistent sample sizes and overall quantitation difficult.

In future work I would like to co-stain with these two antibodies to see if I can observe Oct4+/Gata4+ cells, which I did not have time to do during this project. I would also like to cover earlier time points to see if the individual patterns change. In addition, to try to combat uncertainty over the specificity of *Oct4* antibodies, I could try co-staining with another pluripotency marker such as *Nanog*, *Sox2* or *Klf4*. Strong overlap between these genes would suggest those cells were genuinely pluripotent and still being formed in wild-type day 14 EBs, or it would confirm the somatic expression of *Oct4*.

5.2.2 Obtaining and validating TET-OFF pluripotency cell lines

To investigate the interplay between DNA methylation and pluripotency networks I obtained from the Riken Institute in Japan conditional ES cell lines in which the master pluripotency genes are under the control of doxycycline-regulatable TET-OFF promoter systems. One line was originally engineered to contain a tetracycline-controlled *Oct4* transgene (*Oct4* $^{tet/tet}$), with the wild-type alleles being disrupted by KO vector sequences (Niwa, Miyazaki and Smith, 2000). The second line had the *Sox2* alleles under the *tet* operon control (Masui, Nakatake, Toyooka, Shimosato, Yagi, Takahashi, Okochi, Okuda, Matoba, Alexei A. Sharov, *et al.*, 2007) (*Sox2* $^{tet/tet}$). As a control cell line I also obtained ES cells which were derived from the same mouse line (129/Ola mice) but with the wild-type *Oct4* and *Sox2* alleles, EBRTcH3 cells (Masui *et al.*, 2005) (EBR). The new cell lines and their key features are summarised in figure 5.2.

To confirm the functionality of the TET-OFF systems in these cell lines I cultured the cells with 1µg/mL dox for up to four days, as per the published protocols (Niwa, Miyazaki and Smith, 2000; Masui *et al.*, 2007), and noted changes in cell morphology. Representative images are shown in figures 5.3 and 5.4. Masui *et al.* reported that *Sox2*$^{tet/tet}$ ESCs became completely differentiated by 72 hours of culture with dox, with most cells acquiring a trophectoderm-like cell morphology (Masui *et al.*, 2007). Here, differences in

the *Sox2*$^{tet/tet}$ ESCs' phenotype could clearly be seen when comparing 0 hours to 72 hours in dox as most of the cells became smaller and more rounded, which is consistent with the reported phenotype (Figure 5.4). Phenotypic changes were also seen in the *Oct4*$^{tet/tet}$ cells when they were cultured with dox. Niwa *et al.* reported the cells flattened and enlarged their nuclei over 48 hours in dox culture(Niwa, Miyazaki and Smith, 2000). However, in this study the *Oct4*$^{tet/tet}$ cells took on a more flattened, epithelial-like structure by 72 hours suggesting they were differentiating down a different lineage path (Fig. 5.3). Though these phenotypes differ, the morphology change observed here still indicates the cells exited the pluripotent state, suggesting loss of *Oct4* activity and confirming the efficacy of the dox treatment. The morphology of the EBRTcH3 control cells remained consistent in the presence of dox, as expected as they do not contain TET-OFF promoters.

To confirm the loss of the respective proteins, I performed western blot analyses with proteins extracted from ES cells cultured with dox for 4 days. In the *Sox2*$^{tet/tet}$ cells the Sox2 protein levels fell by roughly 2-fold to become nearly undetectable after 24 hours of culture which is consistent with what was previously reported (Masui *et al.*, 2007) (Figure 5.5). The *Oct4*$^{tet/tet}$ cells also appeared to lose Oct4 within 24 hours of initiating dox treatment (Fig. 5.6); Niwa *et al*. did not publish a western blot result but they showed by northern blot that *Oct4* RNA became undetectable by 48 hours (Niwa, Miyazaki and Smith, 2000). Both blots presented here confirm the TET-OFF systems were functioning in the two modified cell lines.

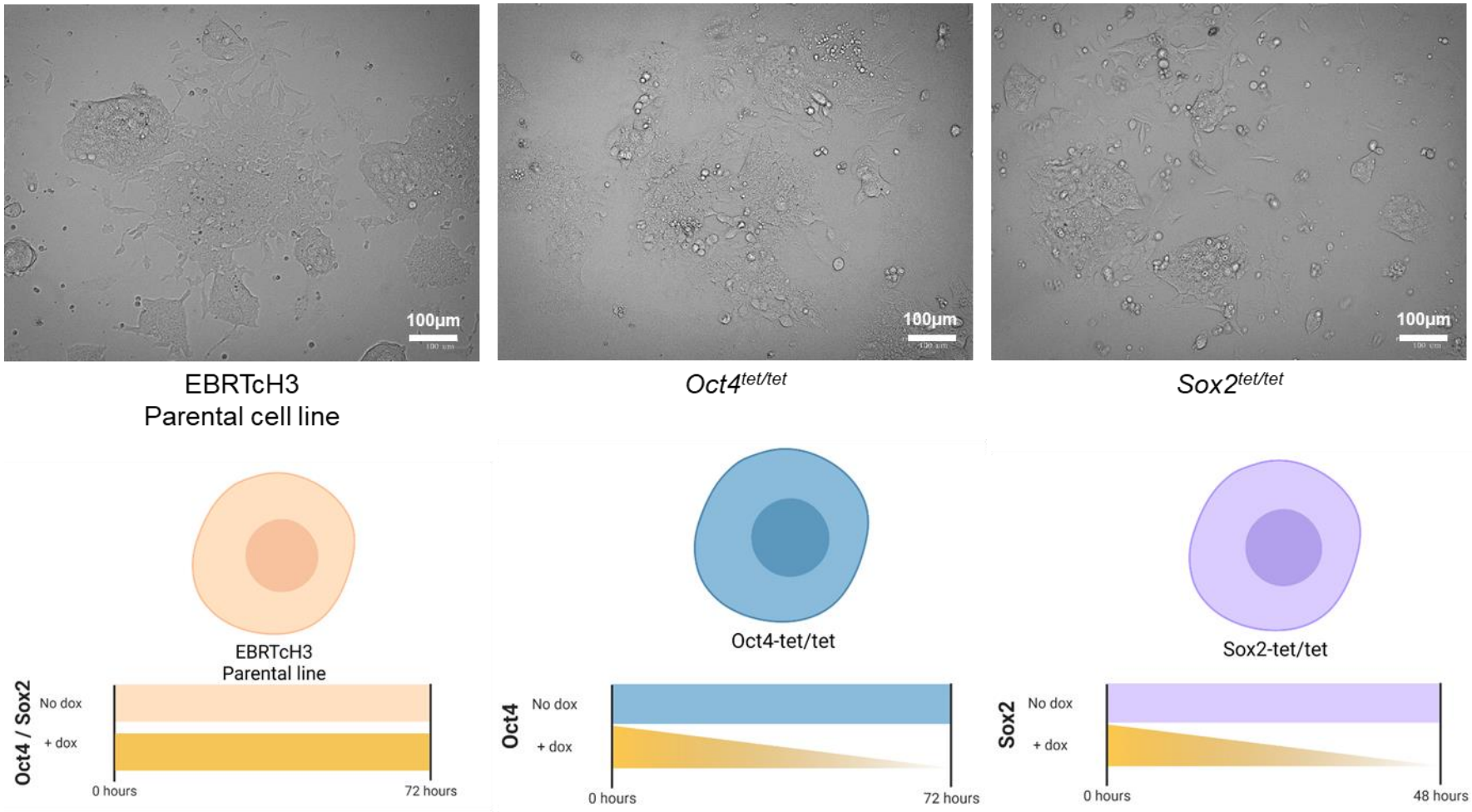


Figure 5.2. Cell lines were obtained in which expression of one of the master pluripotency genes, *Oct4* or *Sox2*, could be inhibited with dox. Figure legend on next page…

Figure 5.2. Cell lines were obtained in which expression of one of the master pluripotency genes, *Oct4* or *Sox2*, could be inhibited with dox. The top row shows bright-field images of the ES cells growing in media without dox. The schematic beneath illustrates which genes are under the control of the TET-OFF system in each line. The parental cell line, named EBRTcH3 cells, has wild-type *Oct4* and *Sox2* alleles and was used as a control. The *Oct4*$^{tet/tet}$ cells are reported to lose *Oct4* expression within 3 days of treatment with dox. The *Sox2* $^{tet/tet}$ cells lose expression of Sox2 are reported to lose within 48 hours of culture with dox. The cell lines were obtained from the Riken Cell Bank, see main text for references. Scale bars represent 100μm.

As the Oct4 protein appeared to become depleted much more quickly than expected in the *Oct4*$^{tet/tet}$ ESCs when treated with dox, I performed an orthogonal assay to quantify any changes in pluripotency in treated cells. Ssea1 is a cell-surface marker specifically expressed by pluripotent cells (Donovan *et al.*, 1986). I performed an Ssea1 assay to detect the protein using antibodies labelled with a fluorescent marker which is then detected by flow cytometry. I collected dox-treated EBRTcH3 and *Oct4*$^{tet/tet}$ cells every 24 hours up to 96 hours, stained them as per the manufacturer's instructions, and quantified the intensity of Ssea1-FITC signal by flow cytometry. The results are shown in figure 5.7. The proportion of cells which are Ssea1+ in the control condition fluctuates over the 96 hour period but never drops below 79%. It is expected that during serum/LIF culture a small proportion of cells will exit a full pluripotency state, which would lead to a drop in observed Ssea1 signal. In contrast, 43% of *Oct4*$^{tet/tet}$ cells were Ssea1+ prior to the addition of dox, and this fell rapidly to <5% after 96 hours with dox. This again suggests the dox treatment is effectively inhibiting the expression of Oct4 in these cells. However, the starting proportion of Ssea1+ cells was lower than in the parental cells; this could be due to an inability of the transgene to recapitulate endogenous Oct4 expression or to an inadequate volume of antibody being used for the number of cells being stained as there were no visible indications of mass differentiation occurring in these cells

under no dox conditions. I conclude therefore that, although the assay may require further optimisation, the trend observed suggested dox treatment effectively inhibited *Oct4* expression in these cells and consequently leads to loss of pluripotency and therefore Ssea1 expression.

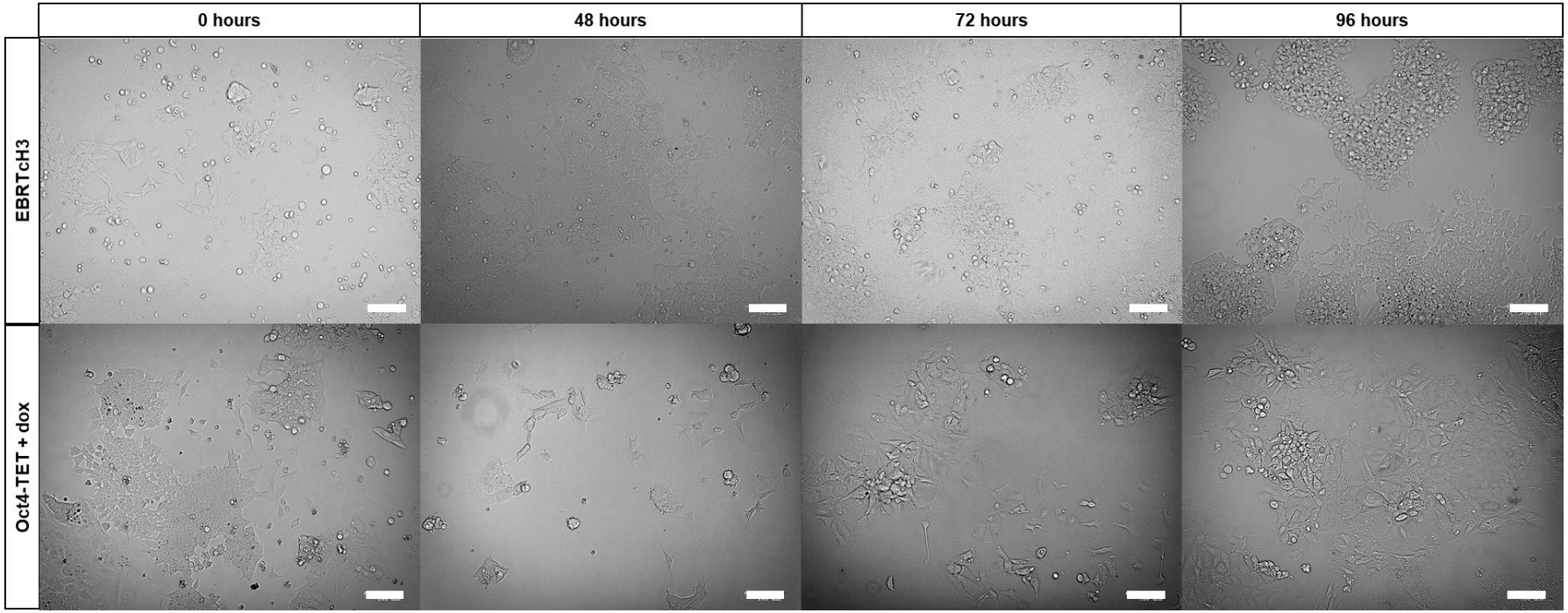


Figure 5.3. *Oct4*$^{tet/tet}$ ES cells change morphology when cultured with dox. Bright-field microscopy images show the parental EBRTcH3 cell line with wild-type *Oct4* alleles do not change morphology in the presence of dox, whereas the ES cells with *Oct4* alleles under control of a TET-OFF system become progressively less ES-like in the presence of dox. Scale bars represent 100μm.

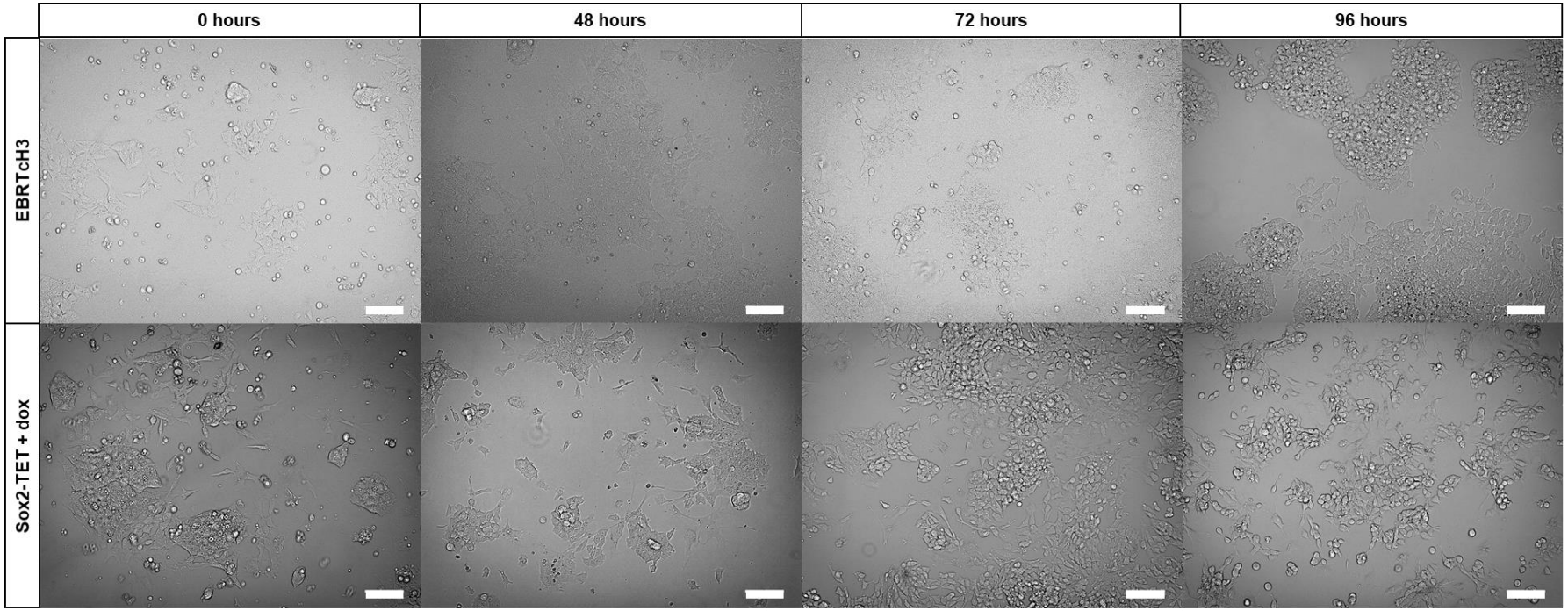


Figure 5.4. *Sox2*$^{tet/tet}$ ES cells change morphology when cultured with dox. Bright-field microscopy images show the parental EBRTcH3 cell line with wild-type *Sox2* alleles do not change morphology in the presence of dox, whereas the ES cells with *Sox2* alleles under control of a TET-OFF system become progressively less ES-like in the presence of dox. Scale bars represent 100μm.

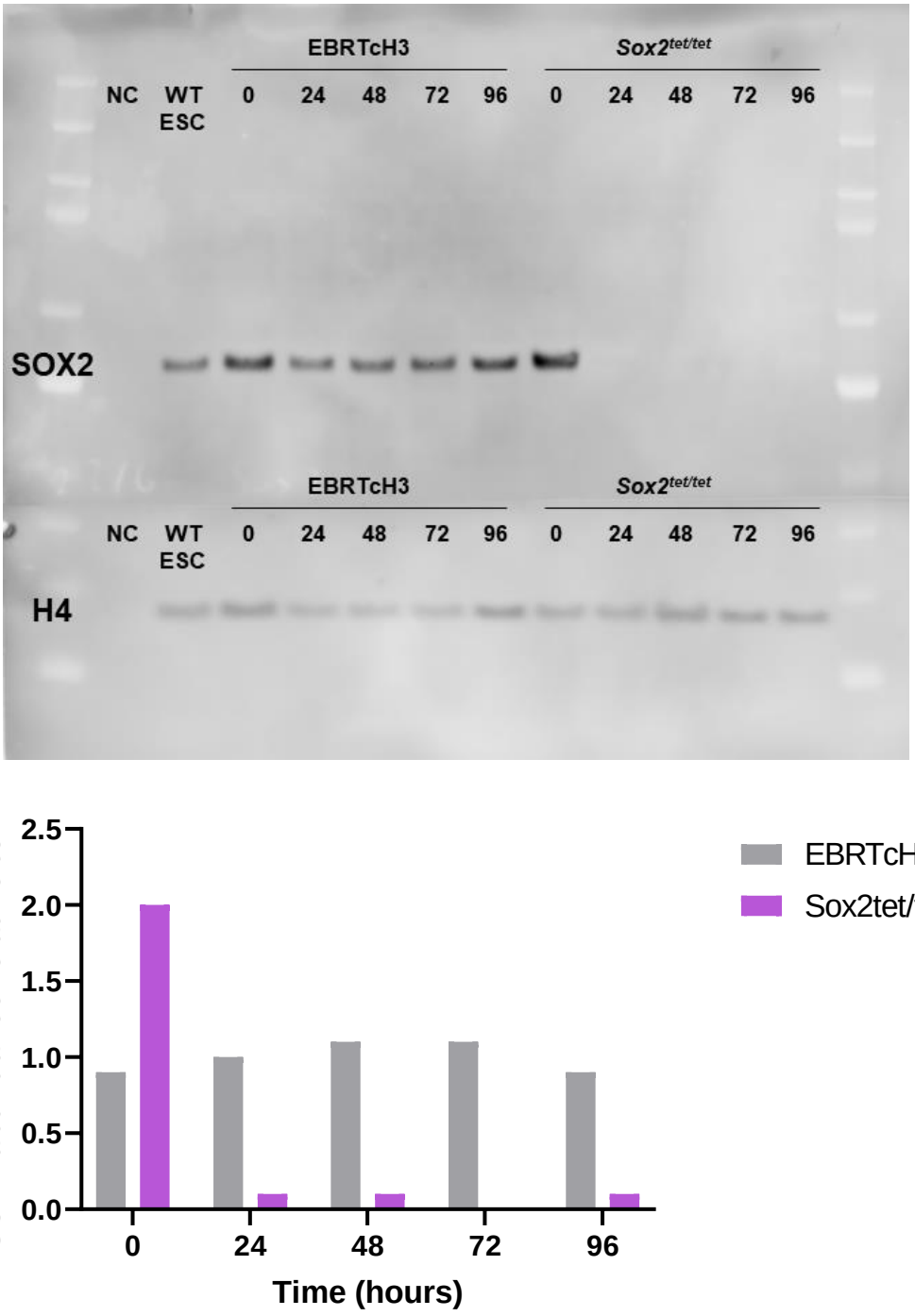


Figure 5.5. Western blot analysis confirmed the inhibition of *Sox2* expression by dox in *Sox2*$^{tet/tet}$ ESCs. Sox2 levels remained stable in the EBRTcH3 cells which have wild-type *Sox2* alleles, but the protein became almost undetectable after 24 hours in dox culture in the *Sox2*$^{tet/tet}$ ESCs. Protein extract from an independent ES line (WT ESC) was included as a control to compare endogenous Sox2 levels. The changes in protein levels are quantified in the bar plot and are shown normalised to H4. H4 was used as a loading control. NC = No template control.

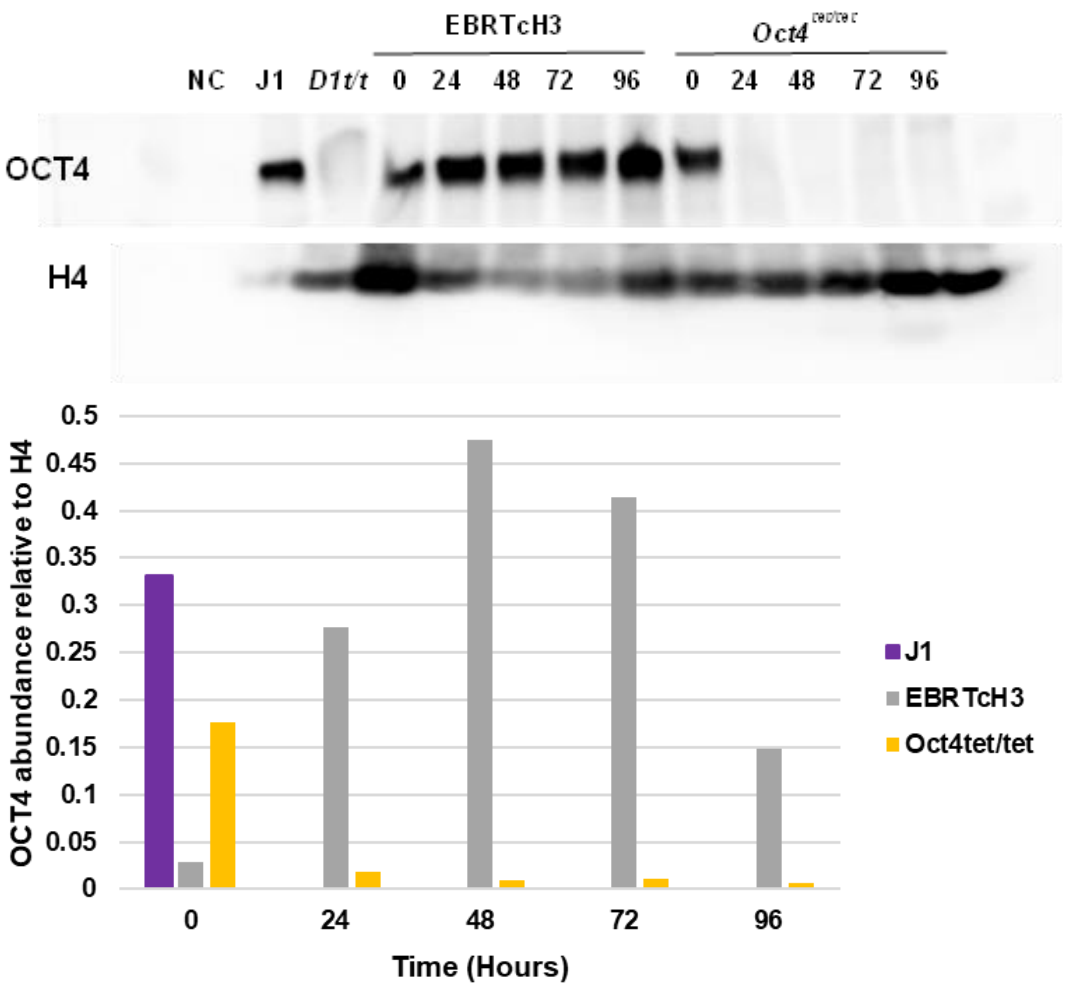


Figure 5.6. Oct4 was significantly reduced in *Oct4*$^{tet/tet}$ ESCs after 24 hours in dox culture. The ESCs were cultured in 1µg/mL dox for up to 96 hours. EBRTcH3 is the parental cell line with wild-type *Oct4* alleles. J1 ESCs were used as a wild-type protein control from an alternative mouse background. *D1t/t* = *Dnmt1*$^{tet/tet}$ ESC protein control. H4 was used as the protein loading control. Oct4 is shown in the bar plot normalised to H4 for each sample.

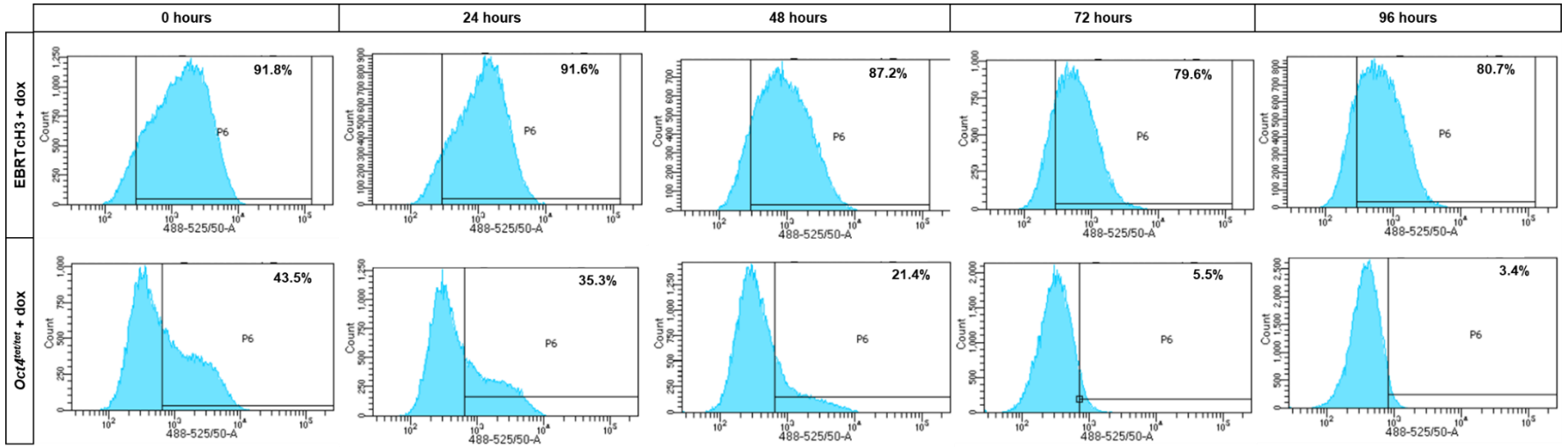


Figure 5.7. Ssea1 staining showed *Oct4*$^{tet/tet}$ ESCs became less pluripotent when cultured with dox. EBRTcH3 ES cells have wild-type *Oct4* alleles and were used as controls. Cell culture medium was supplemented with 1μg/mL doxycycline for the duration of the time course, refreshed daily.

5.2.3 Generating new cell lines lacking *Dnmt1* using CRISPR genome editing

So far in this chapter I have validated the TET-OFF system in the newly acquired cell lines by assessing cell morphology changes, by western blot analysis for the TET-regulated proteins, and by Ssea1 pluripotency assay in the inducible *Oct4*$^{tet/tet}$ cell line. My next aim was to study the consequences of simultaneous absence of DNA methylation and a master pluripotency transcription factor on EB development. To achieve this, I attempted to generate homozygous *Dnmt1* knockout (KO) cell lines from the control EBRTcH3 cells, and from the *Oct4* and *Sox2* TET-OFF cell lines using CRISPR Cas9 genome editing. I used the same strategy as that used for *Dnmt1* knockout in the EGFP-MBD-nls cell line which was described in Chapter 3. A schematic of the deletion strategy is shown in figure 3.17. This process took 8 weeks from initial transfection with the CRISPR guides to final colony selections, and a further 8 weeks to perform the validation experiments.

As before, I performed PCR analyses to identify successful KO clones. In summary, I designed two PCR screens to confirm the deletion; the first screening assay was designed to detect the macro-deletion of the whole gene by placing PCR primer sites external to each end of the gene such that retention of the 52kb *Dnmt1* sequence would inhibit the newly synthesised sequences from overlapping and therefore cause the PCR to fail to produce a band when analysed by gel electrophoresis. Single clones assayed with these PCR primers were considered for the second round of the screen only if a ~500bp PCR band was produced and visible on a gel as formation of this band was strongly suggestive of removal of one of the *Dnmt1* alleles.

In the second screen PCR primer pairs were produced to target a ~500bp region at each end of the gene such that one primer was outside the gene and the other was within the exon sequence. Two primer pairs were designed, one for each end of the gene. The pairs were designed in this way to determine if the clone being tested was homozygous or heterozygous for the gene deletion. The results for the EBRTcH3 and *Oct4*$^{tet/tet}$ ES line

CRISPR knockout experiments are shown in figures 5.8 and 5.9. Figure 5.8 includes the control PCR reaction results, wherein the primers were tested with DNA from a cell line which had not been transfected with the CRISPR guides (*Dnmt1*$^{tet/tet}$) and DNA from the transfection mixed populations (labelled 'EBR mix' and 'OCT mix'). In figure 5.9 the macro-deletion PCRs for the chosen clones were repeated and run in the same gel with the micro-deletion PCR samples. One clone each for the EBRTcH3 and *Oct4*$^{tet/tet}$ knockout lines were identified which produced the target band in the macro-deletion screen and no bands in either of the two micro-deletion screens, suggesting both alleles of *Dnmt1* had been removed in these clones. The results for the EGFP-MBD-nls cell line knockout experiment are shown and described in Chapter 3. Unfortunately, I was unable to derive a knockout line for the *Sox2*$^{tet/tet}$ cells after 3 attempts, so I continued with the EBRTcH3 and *Oct4*$^{tet/tet}$ lines only.

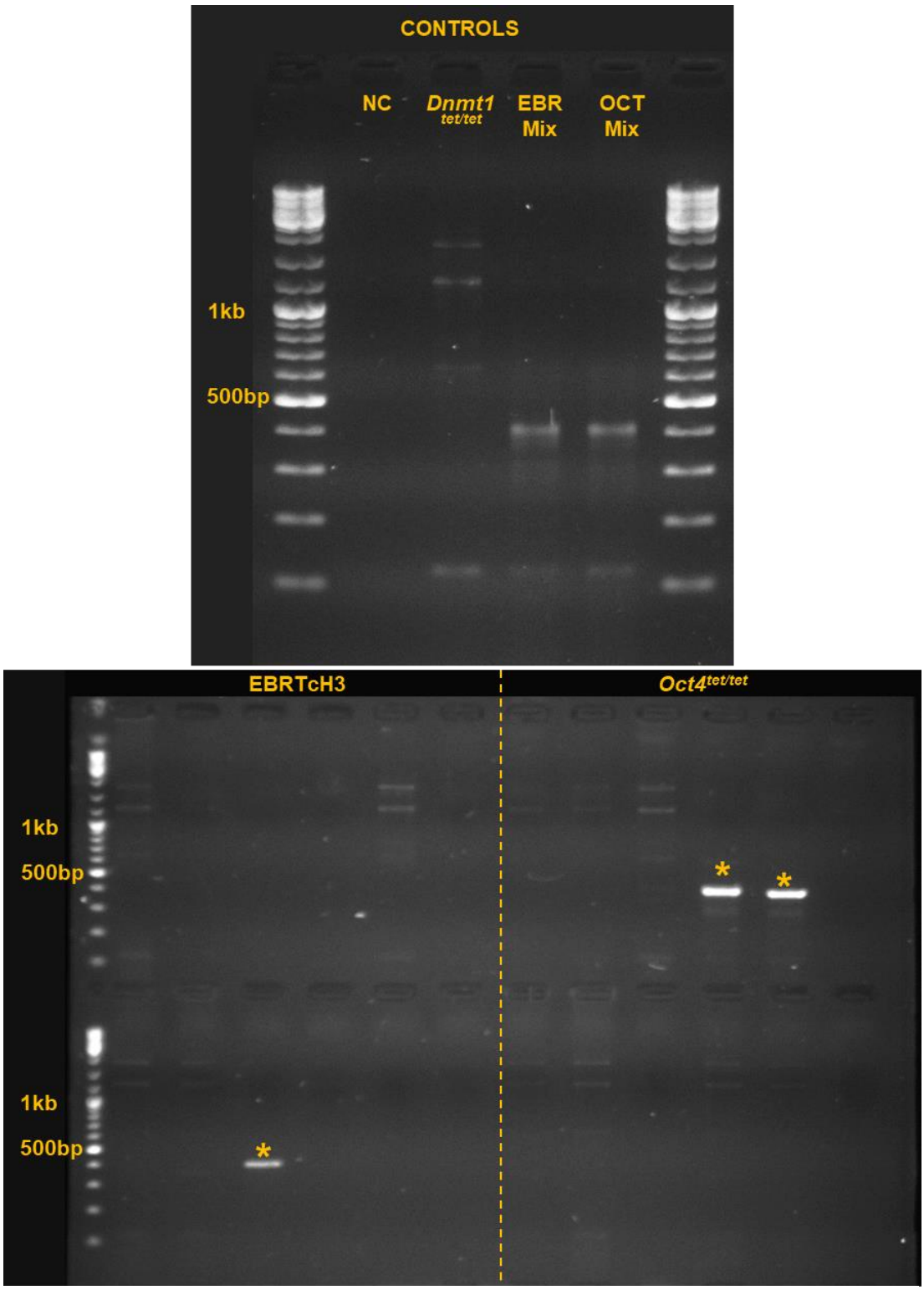


Figure 5.8. *Dnmt1*$^{-/-}$ clones were identified by PCR using primers positioned outside each end of the gene which create a ~500bp amplicon only if the 52kb *Dnmt1* gene is removed. Wild-type *Dnmt1*$^{tet/tet}$ DNA and DNA extracts from the mixed transfected cell populations were used as PCR controls (top gel). Representative single clone samples are shown in the bottom gel. The asterisks highlight clones which were selected for further screening.

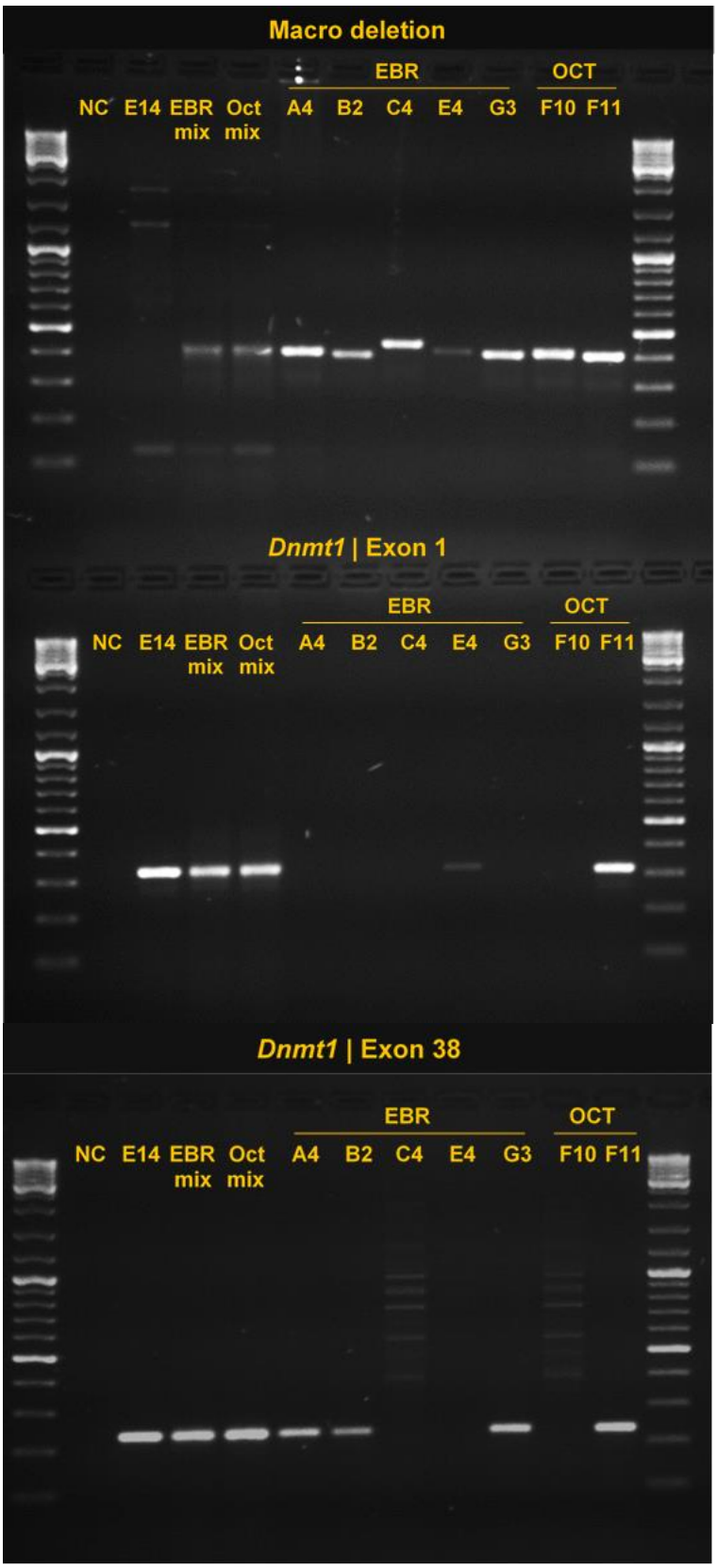


Figure 5.9. Legend on next page…

Figure 5.9 (Previous page). *Dnmt1*$^{-/-}$ clone homozygosity was checked by PCR targeting regions internal to the outer exons of *Dnmt1*. For the micro-deletion PCRs deletion of the gene removed the internal primer sites leading to negative PCRs. Clones considered to be homozygous knockouts had one band in the macro-deletion PCR (top gel) and two negative micro-deletion PCRs (middle and bottom gels). As a result, EBRTcH3 clone C4 and *Oct4*$^{tet/tet}$ clone F10 were identified. Wild type E14 ESC DNA and DNA from the transfection mixed populations were used as negative controls. NC = no template control.

Having identified potential *Dnmt1*$^{-/-}$ clones for the EBRTcH3 and *Oct4*$^{tet/tet}$ cell lines by PCR I then moved on to validating the knockout by orthogonal methods. First, I performed a western blot to confirm the Dnmt1 protein was not being produced. Protein was extracted from wild-type ES cells derived from an independent mouse line as a control (J1). Protein was also used from a cell line in which *Dnmt1*, *Dnmt3a* and *Dnmt3b* have all been removed to use as an additional control (TKO cells) (Tsumura *et al.*, 2006). For the final control I used an ES cell line in which *Dnmt1* is overexpressed due to the presence of a strong CAG promoter (HEK CAG-DNMT1). Finally, I included protein from the unmodified EBRTcH3 and *Oct4*$^{tet/tet}$ cell lines. The results are shown in figure 5.10. As in the no template control and the TKO control lanes, no protein was detected in the EBRTcH3 *Dnmt1* KO clone sample, supporting the PCR results. Surprisingly a faint band was repeatedly observed in the *Oct4*$^{tet/tet}$ *Dnmt1* knockout clone sample lane suggesting the knockout was incomplete.

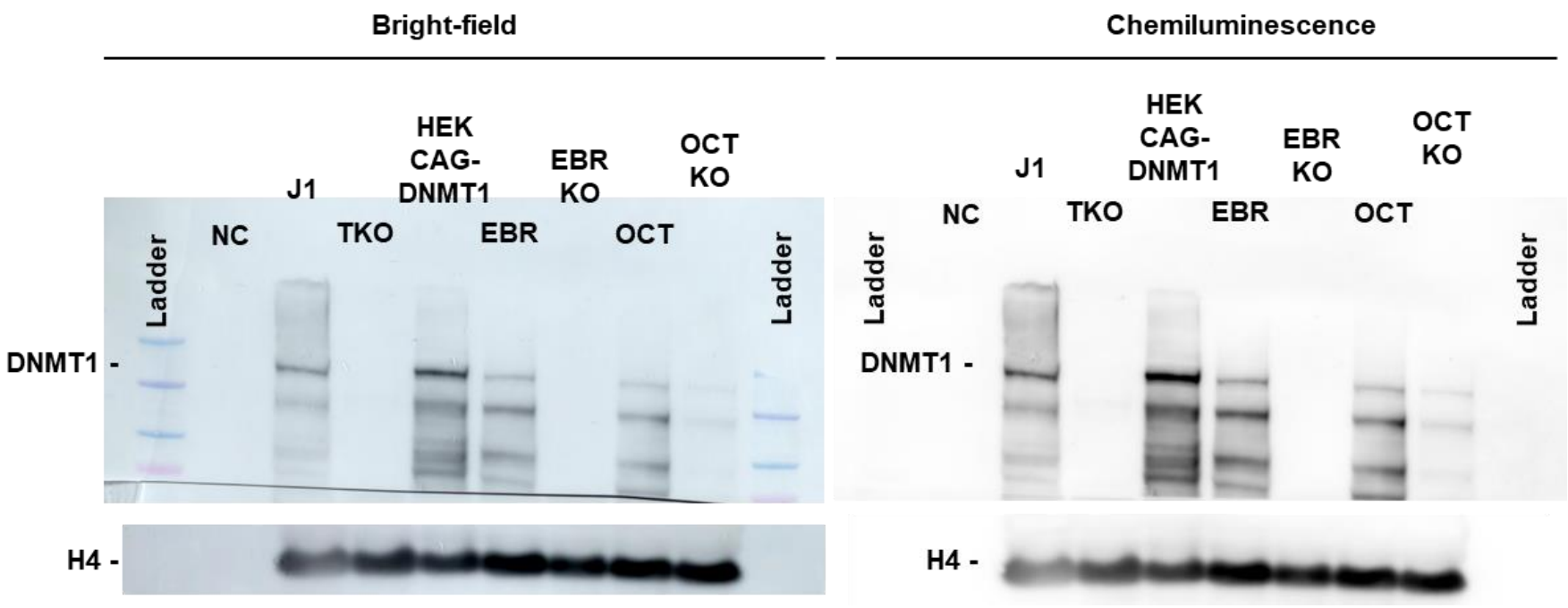


Figure 5.10. Western blot analysis of the *Dnmt1* KO cell lines showed Dnmt1 protein was undetectable in the EBRTcH3 knockdown clone (EBR KO), but some was detectable in the *Oct4*$^{tet/tet}$ knockdown clone (OCT KO) suggesting a hypomorph line was formed. H4 was used as a loading control. J1 = wild type ES cell line protein extract. TKO = ES line lacking *Dnmt3a*, *Dmnt3b* and *Dnmt1*. HEK CAG-DNMT1 = DNMT1 overexpression ES cell line. EBR = Unmodified EBRTcH3 cell line. OCT = Unmodified *Oct4*$^{tet/tet}$ cell line.

To investigate the *Oct4*$^{tet/tet}$ *Dnmt1* knockout clone further, I performed a Southern blot using clone (F10) DNA which was digested with methylation-sensitive restriction enzyme *MaeII* (*HpyChIV*) and probed for repetitive major satellite DNA as described in Chapter 3. The results are shown in figure 5.11. The *Oct4*$^{tet/tet}$ *Dnmt1* KO clone DNA was digested to an intermediate level between the patterns observed for wild type ESC DNA (normally methylated and mostly undigested) and control TKO DNA (constitutively hypomethylated and mostly digested), suggesting some *Dnmt1* activity was retained in this clone. The no template control (NC) showed there was no contamination in the reagents used. The TKO DNA is hard to see in this plot; this is most likely because the hypomethylated DNA is easily digested by *MaeII* and forms very small fragments which can pass through the membrane and into the adjacent blotting paper when the transfer step is run overnight.

The Southern blot validated the western blot analysis (Figure 5.10) that suggested around 10% of normal Dnmt1 protein was present, and I conclude that I have produced a cell line in which *Dnmt1* was partially deleted in the Oct4$^{tet/tet}$ background cell line (labelled as -/+ in figure 5.11). As the nature of the partial knockdown could not be fully defined due to time constraints, and as the activity of Dnmt1 was consequently significantly reduced as shown by the Southern blot, I hereafter define the new cell line as expressing a hypomorphic version of Dnmt1, and therefore refer to this line as the Dnmt1 hypomorph cell line. It is possible that only the outer exons were removed and so enough of the gene was retained to produce a partially functioning protein which could establish low-level *de novo* methylation; this protein was sufficient to be bound by the anti-Dnmt1 western blot antibody, and to establish sufficient DNA methylation that the restriction enzyme was not able to fully digest the DNA. However, the levels of protein and DNA methylation are still significantly reduced when compared to the unmodified parental cell line samples. Unfortunately, there was insufficient time remaining in the PhD project timeline to attempt a repeat of the knockout experiment and screening, so I continued with this clone and noted its hypomorph status. In future work I could add an additional primer pair to the

PCR screen validation steps to try to detect the catalytic region of the protein, or the middle exon to further support the macro-deletion screen conclusion.

Following the partial knock-out of *Dnmt1* in the *Oct4*$^{tet/tet}$ cell line and full knock out in the control EBRTcH3 line, differences in ES cell morphology became apparent during serum/LIF culture (Figure 5.12). Single *Oct4*$^{tet/tet}$ *Dnmt1* hypomorph cells became larger and more flattened in appearance, with multiple small protrusions along each edge, a phenotype suggestive of differentiation despite the presence of LIF. This was unexpected as ESCs are known to be able to tolerate the loss of 5mC without undergoing spontaneous differentiation and can continue to self-renew, unlike somatic cells which die. I therefore didn't expect to see a significant change in morphology in the *Oct4*$^{tet/tet}$ *Dnmt1* KO clone cells. The EBRTcH3 *Dnmt1* KO clone cells, however, retained an ES-like phenotype with tapered, round cells, though they too appeared to be larger than in the unmodified Dnmt1$^{+/+}$ cell line. These observations suggest that the parental *Oct4*$^{tet/tet}$ line does not recapitulate full wild-type *Oct4* expression, resulting in differentiated cells that are more sensitive to *Dnmt1* loss than compared to wild-type EBRTcH3 ESCs.

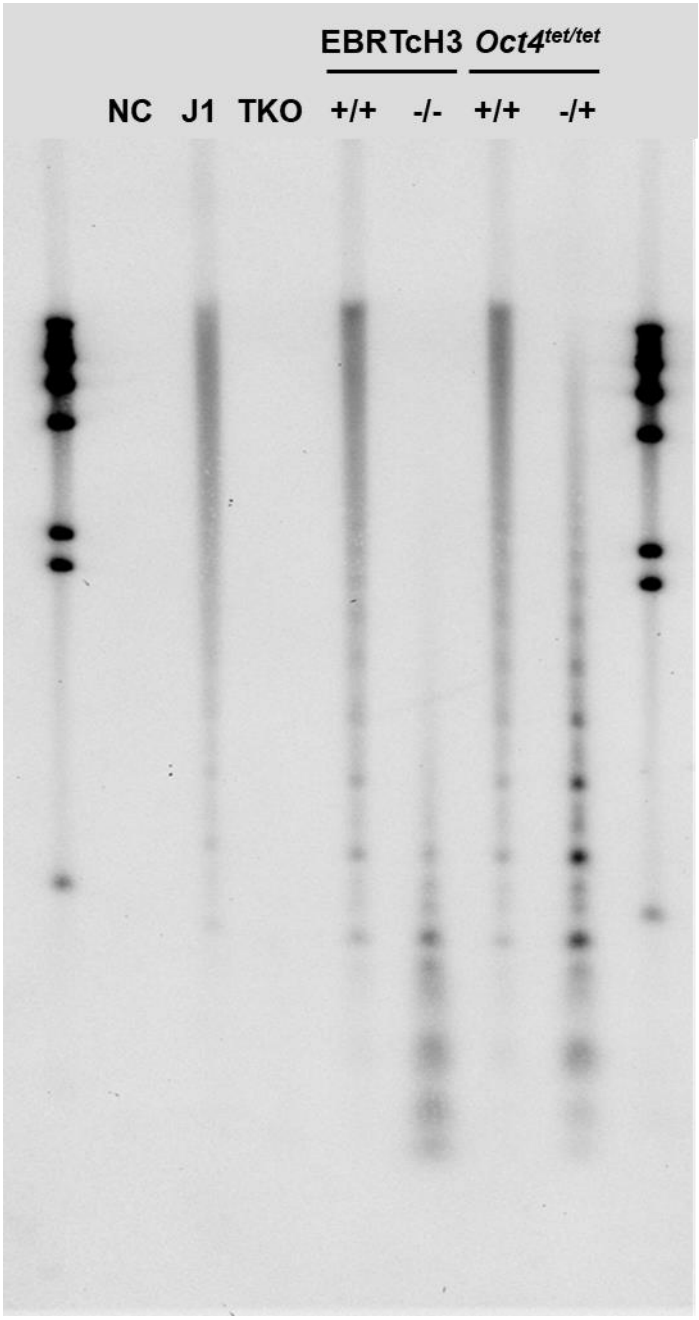


Figure 5.11. Figure legend on next page…

Figure 5.11. Figure on previous page…

Figure 5.11. Southern blot showing the CRISPR deletion of *Dnmt1* led to total hypomethylation in EBRTcH3 cells and partial hypomethylation in *Oct4*$^{tet/tet}$ ESCs. DNA extracts were digested with methylation-sensitive restriction enzyme *MaeII* (*HpyCH4IV*) and analysed by Southern blot with DIG-probe targeting major satellite repeats. The parental EBRTcH3 ES cells and *Oct4*$^{tet/tet}$ ES cells with wild-type *Dnmt1* (+/+) had satellite repeat patterns similar to the wild-type J1 ES cell line, with mostly high molecular weight bands at the top of the gel. In comparison, the CRISPR-modified EBRTcH3 cell line in which *Dnmt1* was targeted and knocked out (-/-) had mostly low-molecular weight bands towards the bottom of the gel. The CRISPR-modified *Oct4*$^{tet/tet}$ ESC DNA had an intermediate band pattern with some high but mostly low molecular weight bands. This suggested the knockout was incomplete (-/+), but sufficient to significantly reduce global DNA methylation in these cells so the DNA was permissive to digestion by *MaeII*.

NC = No template control. J1 = ES cell line with wild-type *Dnmt1* alleles and DNA methylation. TKO = constitutively hypomethylated ESC control cell line. +/+ = both *Dnmt1* alleles present. -/- = *Dnmt1* homozygous knockout. -/+ = Incomplete *Dnmt1* knockout resulting in hypomorphic gene expression.

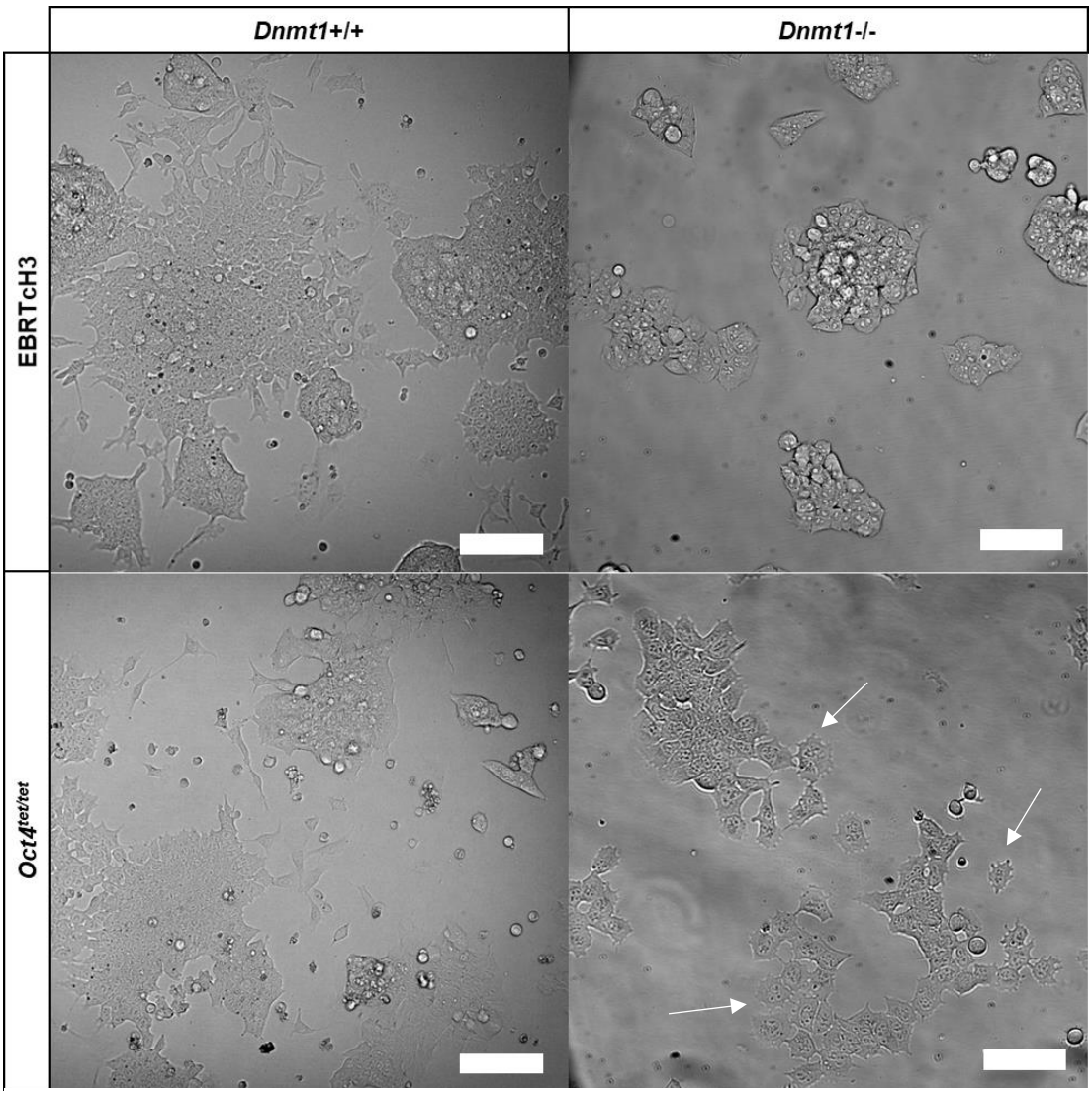


Figure 5.12. CRISPR knockout of *Dnmt1* led to ES cell differentiation.

Arrows indicate differentiating cells. Scale bars represent 100μm.

5.2.4 Investigating the interplay between DNA methylation and pluripotency networks in differentiating EBs

Having established an *Oct4*$^{tet/tet}$ *Dnmt1* hypomorph cell line and EBRTcH3 *Dnmt1*$^{-/-}$ cell line, I went on to investigate the impact of combining loss of *Dnmt1* with repression of *Oct4* expression on EB differentiation. I performed immunocytochemistry (ICC) staining on wild-type EBRTcH3, EBRTcH3 *Dnmt1*$^{-/-}$, *Oct4*$^{tet/tet}$, and *Oct4*$^{tet/tet}$ *Dnmt1* F10 hypomorph EBs cultured in the presence and absence of dox and collected at day 14 of differentiation. As in previous experiments, the appropriate ES cell lines were pre-treated with doxycycline for at least 4 days prior to EB differentiation. I used anti-Oct4 antibodies to mark pluripotent cells within the EBs, anti-Ncam1 antibodies to mark cells differentiating towards the ectoderm germ layer, and counterstained with DAPI to highlight the cell nuclei. I imaged the stained EBs using an inverted epifluorescence microscope with Nikon JOBS tiling array acquisition enabled to image the EBs at 20X magnification. The JOBS algorithm was designed and written by Dr Matthew Pearson of the IGC AIR imaging team to enable me to quickly identify EBs within the wells and select the area of the well for the microscope to image at high magnification, reducing the total time required to acquire the images. The 20X tiles were then stitched automatically by the NIS elements imaging software (Nikon). Once acquired, I processed the stitched images to produce the final montages and performed size quantification using FIJI (Schindelin *et al.*, 2012). The final montages are shown in Figures 5.13 A-D.

Figure 5.13 A shows representative images of the control EB condition wherein both lines are expressing *Dnmt1* and *Oct4*. As with wild-type EBs of other lines tested, these EBs generally formed large, far-spread bodies containing both pluripotent and differentiating populations of cells.

Figure 5.13 B shows examples of EBs generated from the *Dnmt1* KO cell lines in which *Oct4* was being expressed (no dox). Oct4+ and Ncam1+ cells were still detected in this condition suggesting loss of DNA methylation as a consequence of absence of Dnmt1 did not inhibit ectodermal differentiation or the formation of pluripotent cells. This is consistent with

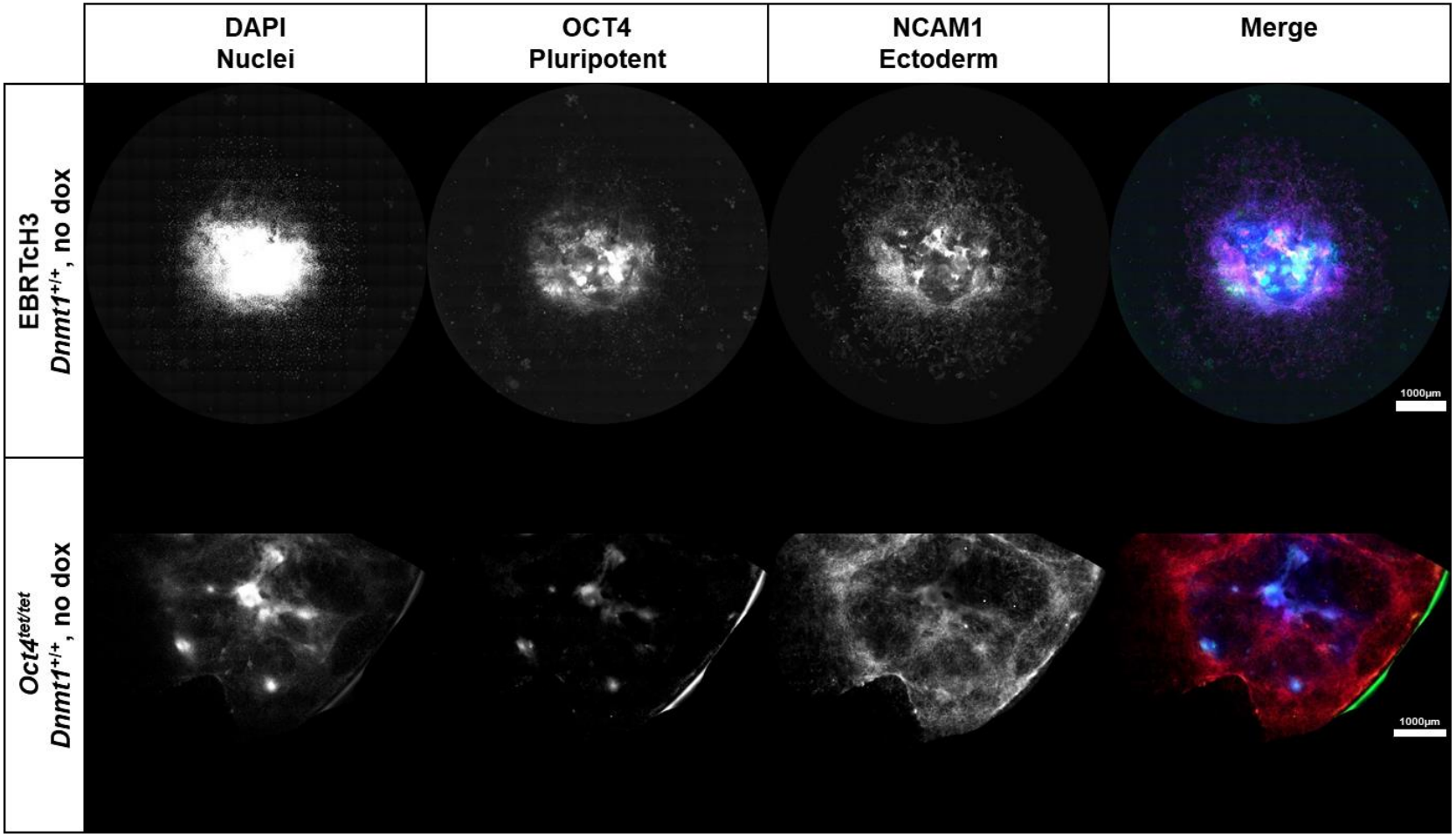


Figure 5.13 A. With wild-type *Dnmt1* alleles and normal *Oct4* expression, *Oct4*$^{tet/tet}$ EBs are similarly sized and complex compared to the control EBRTcH3 cell line EBs.

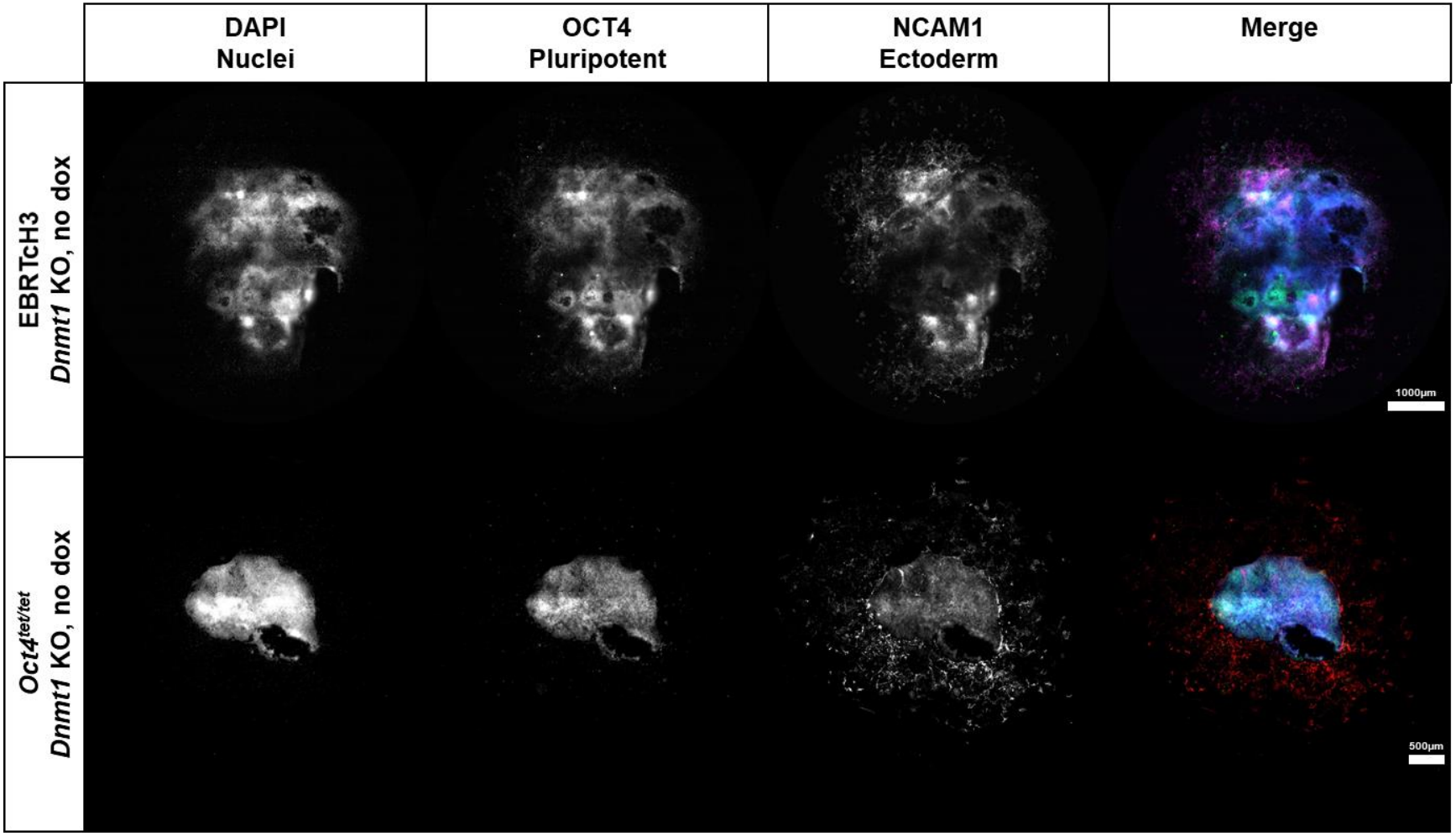


Figure 5.13 B. CRISPR knock-out of *Dnmt1* leads to a reduction in EB size. The change in growth is consistent with the pattern observed in EBs containing TET-cassette regulated *Dnmt1* alleles used in earlier chapters. EBs shown were fixed at day 14 of culture.

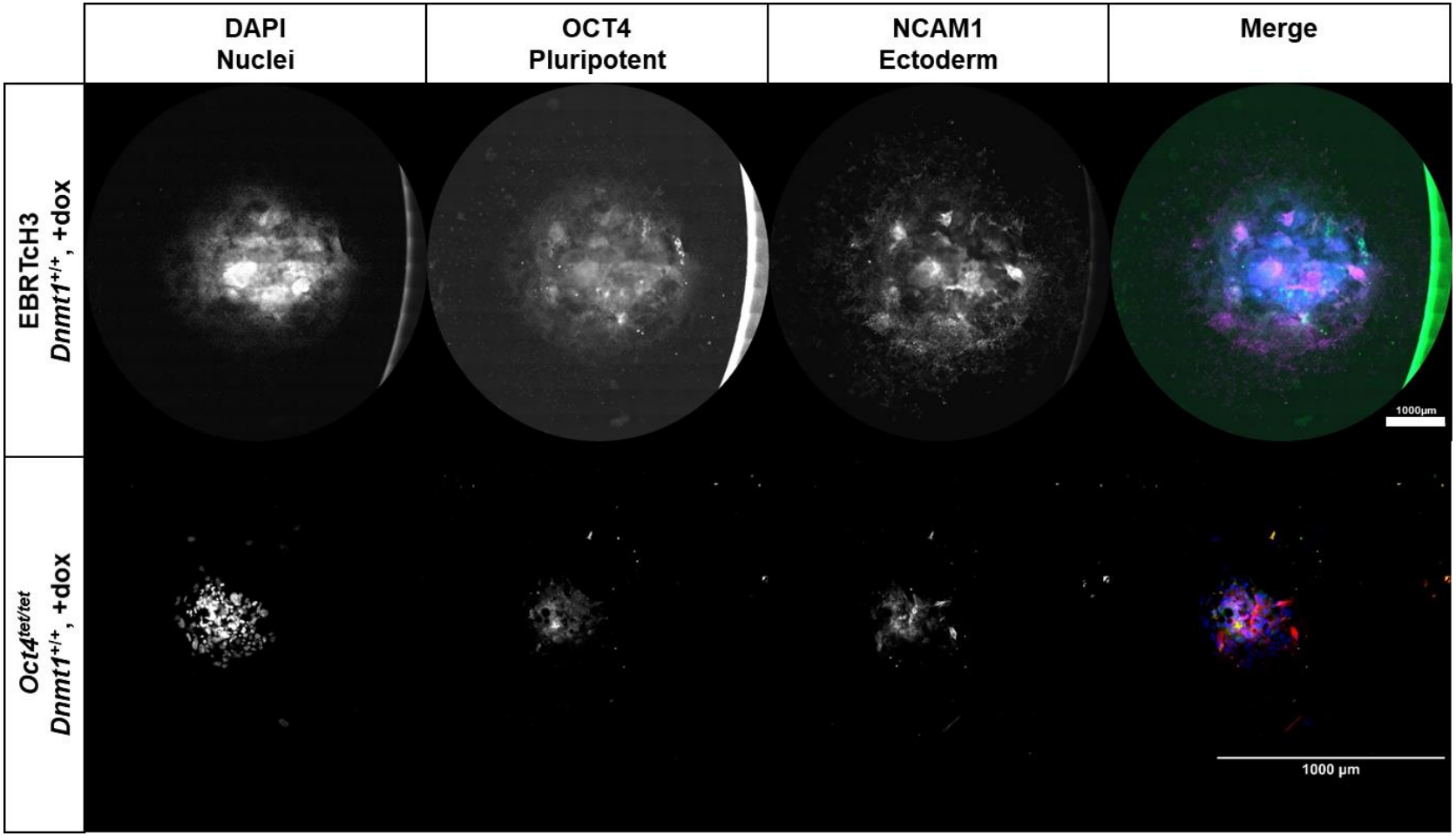


Figure 5.13 C. Inhibition of *Oct4* expression severely impacted EB growth. Ncam1+ cells were still detectable in the relative absence of *Oct4* signalling. EBs shown were fixed at day 14 of culture.

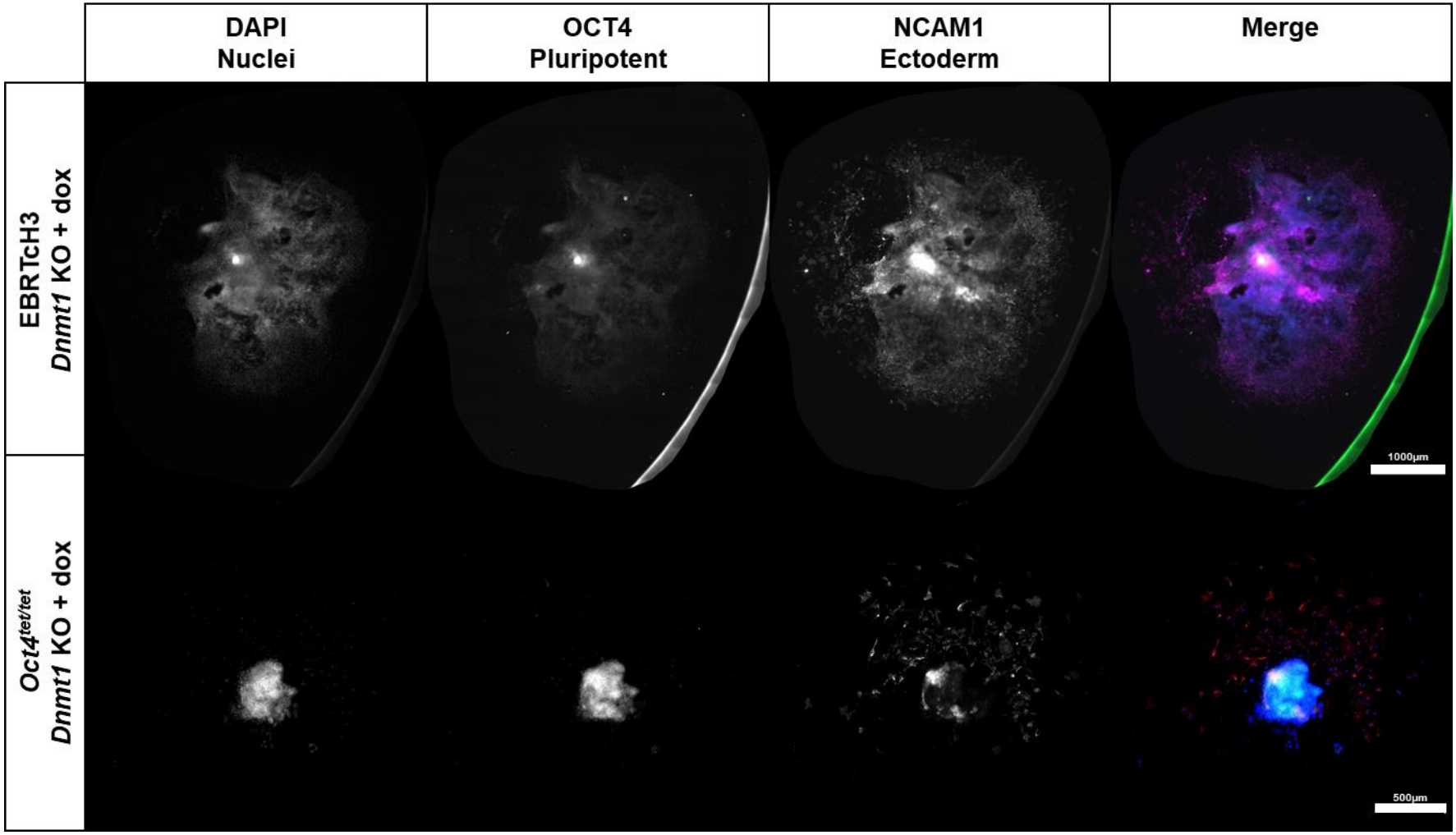


Figure 5.13 D. Loss of both *Dnmt1* and Oct4 expression in *Oct4*$^{tet/tet}$-derived EBs leads to a severe reduction in EB size compared to control EBRTcH3-derived EBs. Ncam1+ cells of the ectoderm lineage were detectable in both lines. EBs shown were imaged at day 14.

previous experiments in which I have shown all the germ layers are represented in the EBs lacking Dnmt1 when the genes are controlled by the TET-OFF system. The primary change observed was a reduction in size, supporting previous results.

In the third experimental condition the parental cell lines with wild type *Dnmt1* alleles were treated with dox, which in the *Oct4*$^{tet/tet}$-derived EBs resulted in differentiation without Oct4 (Figure 5.13 C). Unsurprisingly, the EBRTcH3 control cell line showed the same phenotype and size as EBs in the untreated condition. However, the *Oct4*$^{tet/tet}$-derived EBs consistently showed an extreme phenotype wherein they were significantly smaller. The anti-Oct4 signal was very weak. A green secondary antibody was used to mark the Oct4+ cells, but cells and culture plates tend to auto-fluoresce strongly in the green part of the colour spectrum, so most Oct4 'signal' was background as the contrast was poor. Ncam1+ cells were still detectable in defined single cells, so ectoderm differentiation was not inhibited.

In the final condition the *Dnmt1* KO lines were treated with dox to generate the double-repressed condition lacking both *de novo* methylation and, in the *Oct4*$^{tet/tet}$ *Dnmt1* hypomorph EBs, the master pluripotency factor (Figure 5.13 D). The EBRTcH3 *Dnmt1*$^{-/-}$ line-derived EBs were much larger than expected as they looked to be a similar size to the unmodified parental cell line. This may be a cell line-specific phenomenon; *Dnmt1*$^{tet/tet}$ ESCs could not form epiblast-like stem cells for example. It is not likely to be due to a mistake with drug dosage or quality as this line is a stable knock-out cell line for which I have shown evidence the CRISPR deletion was complete. Possibly the EBRTcH3 cells are more tolerant of loss of *Dnmt1* than the other lines tested, or they are not competent at forming EBs well generally. In comparison, the *Oct4*$^{tet/tet}$ *Dnmt1* hypomorph line when maintained and differentiated in the presence of dox shows a severe size phenotype similar to the *Dnmt1*$^{+/+}$ Oct4- condition. Again, differentiated Ncam1+ cells were detectable. Only the central mass of the EB can be seen in the *Oct4* channel due to the accumulation of cells which auto-fluoresce in this part of the spectrum.

Considering all of the ICC data together, it appears that neither Oct4 nor Dnmt1-dependent 5mC are required for ectoderm differentiation to occur. However, there are clear differences in the abilities of the EBs to grow in size. To quantify this further, I re-used the thresholding method described in earlier chapters to measure the size of the EBs from their respective merged channels images. The data are summarised in figure 5.14. While I did not expect to see a difference in size caused by the presence of dox in the EBRTcH3 EB culture media I did expect the EBs lacking *Dnmt1* to be roughly half the size of their wild-type counterparts as a result of previous analyses. This was not observed; although the *Dnmt1* knockout EBs were smaller the difference was not significant or as extreme as previously seen in other experiments. However, the number of EBs which were available to image and quantify by the end of the staining process were limited, so it may be that the sample size was too small and selective EBs were retained after processing. Certainly, in the 'KO, no dox' condition the range of sizes is large which makes assessment of the impact on size in this condition for this line difficult. Also, the EBRTcH3-derived EBs are all significantly smaller than the equivalent EBs derived from the *Oct4*$^{tet/tet}$ line cells. This suggests that EB derivation for this line needs to be further optimised for further quantitation studies.

As for the *Oct4*$^{tet/tet}$ line derived EBs, I observed again the pattern found with the previous cell lines wherein EBs lacking Dnmt1 were significantly smaller than those EBs which retained the gene. This difference was assessed by one-way ANOVA with multiple comparisons versus the *Dnmt1*$^{+/+}$ no dox condition (*padj*<0.05). Interestingly, the condition with the greatest negative impact on EB size was the +dox only, suggesting the loss of Oct4 is more significant for EB growth than DNA methylation. However, as this line was determined to be a hypomorph with incomplete deletion of *Dnmt1* it is unclear how much the residual *Dnmt1* is impacting this phenotype.

In future work I could re-attempt the CRISPR KO strategy on the *Dnmt1* hypomorph line to complete the deletion and see how this impacts the

EB development compared to the wild-type and hypomorph lines. Alternatively, I could use the new *Dnmt1* inhibitor GSK-3484862 (Azevedo Portilho *et al.*, 2021) that was recently published to induce hypomethylation and see how the EB sizes are impacted.

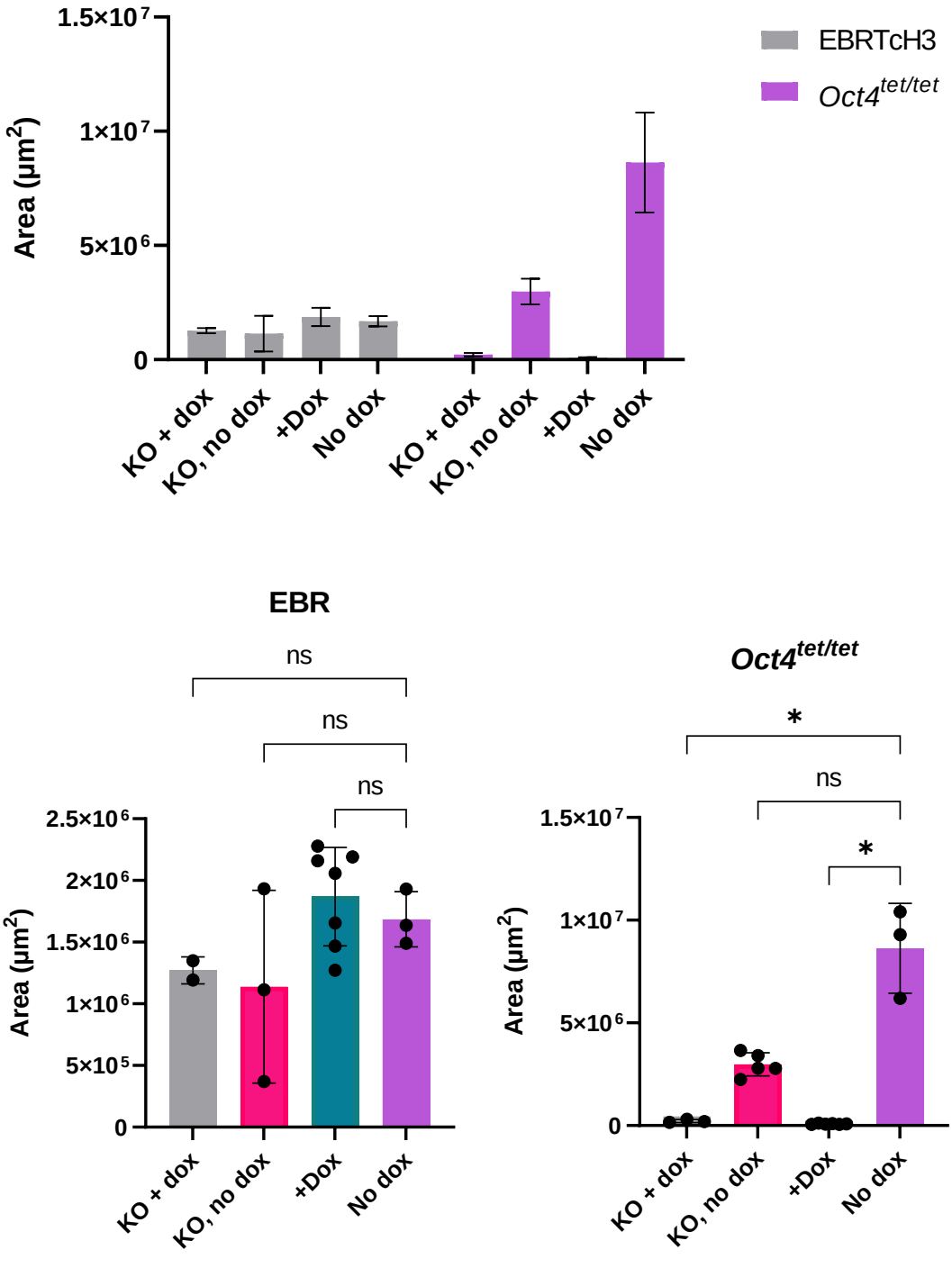


Figure 5.14. Loss of Oct4 negatively impacts EB growth more significantly than loss of Dnmt1. EBs were fixed at day 14 and sizes quantified by thresholding. Mean EB sizes were compared to the No dox condition by one-way ANOVA (EBR = EBRTcH3 parental control cell line-derived EBs; ns = not significant; * = *padj*<0.05)

5.3 Discussion

By the end of Chapter 4 I had data that suggested DNA methylation may have a role in regulating the timing of differentiation initiation, as it was observed that in EBs lacking *Dnmt1* there was high aberrant expression of master pluripotency transcription factor *Oct4*. However, it was unclear if *Oct4* was being aberrantly expressed in the same cells that were expressing differentiation genes, which are not normally expressed in pluripotent ES cells where *Oct4* is normally highly expressed, or if the Oct4+ cells represented a separate sub-population within the late stage hypomethylated EBs. In this chapter I aimed to investigate this further. To do so I obtained new cell lines in which I could control the expression of key pluripotency genes using the TET-OFF system. I then modified these to generate parallel cell lines in which I had removed *Dnmt1* using CRISPR. I planned to use these modified and unmodified cell lines to visualise by immunocytochemistry if pluripotency genes and differentiation markers overlapped in the same cells, and if that pattern changed in *Dnmt1*$^{-/-}$ EBs.

Unfortunately, due to technical difficulties I was unable to produce a full *Dnmt1* KO clone in either the *Oct4*$^{tet/tet}$ or the *Sox2*$^{tet/tet}$ ES cell lines as they appeared resistant to CRISPR editing after multiple attempts. The same CRISPR strategy successfully produced homozygous knockout clones in three other cell lines (wild-type E14 ESCs, wild-type EBRTcH3 cells, and in the EGFP-MBD reporter cell line). A partial knockout was achieved in the *Oct4*$^{tet/tet}$ line. However, differentiating cells were frequently observed in the F10 hypomorph ESC population even when dox was absent suggesting the cells were not recapitulating the wild-type *Oct4* expression which inhibits this spontaneous differentiation in other wild-type ESC lines. Further to this, as the clone is hypomorphic for *Dnmt1* expression further study is required to more precisely establish what levels of DNA methylation are present in these cells.

Nonetheless, when comparing the *Oct4*$^{tet/tet}$-derived EBs cultured in the absence of dox with and without *Dnmt1* expression I found that in the

Dnmt1 KO EBs the patterns of the Oct4 and DAPI stains more closely overlapped, whereas in the wild-type there appear to be many more DAPI+ cells than Oct4+ cells (Figure 5.13 B). This might suggest that the relative absence of *Dnmt1* led to an increase in frequency of OCT4+ cells. This supports previous data; both qPCR and RNA sequencing data suggested from bulk cell extracts that *Oct4* increased in expression following chemical inhibition of *Dnmt1* expression. Here, ICC enables the visualisation of the Oct4 signal distribution across individual cells within the context of a whole EB, though the precise signal cannot be quantified. I would need to perform single cell RNA sequencing or qPCRs for single EBs to address this. In the *Dnmt1* hypomorph images the cell borders can be seen in the Oct4+ cells, distinguishing the antibody signal from auto-fluorescence or blurred cell masses. The signal can be isolated by eye to discrete cells within a concentrated area. It is also possible to see single cells in the other channels which suggests the resolution is good enough.

One of the other main questions I wanted to address in this chapter was if the aberrant expression of pluripotency genes in late-stage embryos observed in the earlier results chapters represented a persistent sub-population or aberrant expression in differentiating cells which could be seen by co-staining in an ICC experiment. What I observed when I stained day 14 *Oct4*$^{tet/tet}$ EBs without dox, and which were hypomorphic for *Dnmt1*, was that there seemed to be no co-expression in single cells. However, this experiment was limited by the hypomorphic nature of the knockout; from the ICC it is unknown how similar the *Dnmt1* levels are between these EBs and those used for qPCR and RNA sequencing analysis, therefore it is unclear what effect the hypomorph is having. Instead, it is possible that a sub-population of Oct4+ cells may have been allowed to escape silencing in the relative absence of *Dnmt1*. The relative proportion of Ncam1+ cells did not appear to change significantly between conditions suggesting the *Ncam1* expression, and therefore ectoderm differentiation, was not significantly affected. Ncam1+ cells mostly appeared to be separated from the Oct4+ cells.

The key change observed between the conditions affected the EBs' size. When dox was added to the cell lines to inhibit *Oct4* expression (no effect was observed in the EBRTcH3 control cell line), there was a significant negative impact on EB size, similar to the loss of *Dnmt1* experiments but more severe (Figure 5.14). The EBs lacking both *Oct4* and *Dnmt1* expression (*Oct4*$^{tet/tet}$-derived EBs without *Dnmt1* cultured with dox) produced the second smallest EBs; surprisingly *Dnmt1*$^{+/+}$ +dox EBs were smaller, suggesting *Oct4* has the greater influence on EB growth (Figure 5.13 D, figure 5.14). This supports the notion that pluripotent populations of cells are required for rapid size expansion up to a limit.

Intriguingly, the *Oct4*$^{tet/tet}$ EBs with wild-type *Dnmt1* that were treated with dox were the smallest EBs at day 14 overall (Figure 5.13 C). It may support the hypothesis that the DNA methylation and pluripotency networks act in opposition to counterbalance each other. When Dnmt1 is present but Oct4 is not the EBs are very small, barely forming the core cell aggregate. This suggests that either overexpression of *Dnmt1* through lack of counteractive *Oct4* expression, or simply overexpression of *Dnmt1* at the same time as loss of expression of *Oct4* has an extremely negative effect on EB development. Possibly due to interference with timing of differentiation initiation and inability to produce rapidly dividing pluripotent cells. However, in the *Dnmt1* knockout dox-treated condition I still saw Oct4+ single cells forming; the TET-OFF system inhibits gene expression but not completely. It could be that residual low-level Oct4 was detected by the staining and this low signal was enhanced during image processing. When compared to the other stains the signal remains relatively low.

To generate the hypomethylated TET-OFF cell lines, I used CRISPR genome engineering to delete *Dnmt1*. This proved to be more difficult than anticipated, especially when under a time constraint. I was able to successfully perform the deletion in three of the five cell lines attempted; the EBRTcH3 parental control ES cell line, the EGFP-MBD-nls DNA methylation reporter ES cell line, and wild-type E14 ESCs which I used to test the CRISPR guides during the 6 months I was waiting to obtain the new cell lines

from Riken. CRISPR editing in the *Sox2*$^{tet/tet}$ ES cells never worked, and the deletion in the *Oct4*$^{tet/tet}$ ES line was eventually found to be incomplete. Though the hypomorphic clone identified passed the PCR screens (Figures 5.8 and 5.9) Dnmt1 protein could still be faintly detected by western blot. Also, a Southern blot with DNA digested with a methylation-sensitive restriction enzyme showed an incomplete digest pattern suggesting DNA methylation was still being deposited *de novo* at low levels. This deletion was attempted multiple times in this line and in the *Sox2*$^{tet/tet}$ ES line, which proved to be completely resistant to CRISPR activity, despite the knockout being successful in the parental cell line. This suggests that unfortunately in this case, CRISPR genome editing was not as efficient with the guide RNAs used in the test ESC line.

So, although this series of experiments produced some interesting early observations, it requires much more analysis and experimental refinement that unfortunately was not possible during my period of study. This was partly due to the impact of the COVID-19 pandemic, which disrupted access to the labs, upset key laboratory supply chains and generally disrupted research momentum. Additionally, it took a considerable amount of time to obtain the *Oct4*$^{tet/tet}$ and *Sox2*$^{tet/tet}$ ESCs from the Riken repository in Japan. Overall, it took 6 months for the cells to be shipped and this considerably impacted the time I had to perform the gene editing experiments and subsequent validation prior to using the modified cell lines. Accordingly, this aspect of the work needs more investigation. The availability of a new reliable *Dnmt1* inhibitor GSK-3484862 (Azevedo Portilho *et al.*, 2021) may expediate experimental analysis; it might negate the requirement for a *Dnmt1* knock-out cell line to be able to study the impact of global hypomethylation on cell viability and ESC differentiation programmes.

5.3.1 Future work

In future work I would like to test the GSK-3484862 (Azevedo Portilho *et al.*, 2021) *Dnmt1* inhibitor on the *Sox2*$^{tet/tet}$ cells which were resistant to CRISPR knockout, and in the *Oct4*$^{tet/tet}$ cells to see if I get the same results as were observed with the hypomorph line I generated.

To compensate for some of the limitations of the ICC experiment, I could analyse gene co-expression and relative expression between conditions by FACS to obtain quantitative data. FACS analysis would be able to speed up this process and provide additional quantitative data, such as total number of marked cells, potentially with PI counterstain for cell viability, co-staining, and information wouldn't be lost from the cells making up the central mass as this bulk could be broken up during processing giving greater single-cell coverage of the whole population. It wouldn't be able to describe the distribution of the markers within the EBs; in the previous chapter the stains suggested that pluripotent Oct4+ cell populations tended to occur more to the centre of the EBs, whereas differentiating cells (e.g. those marked with Ncam1 etc.) were found at the edge. This supports the idea that EBs maintain a population of rapidly dividing cells to the centre and then the cells migrate to the edges as they specialise. Also, FACS is not suited to small sample sizes, and the smallest EBs which are also the most interesting from an experiment point of view as these are the ones lacking either *Dnmt1*, *Oct4* or both, look to be fewer than 10,000 cells per EB. This is the minimum, number of cells required for reliable FACS analysis. However, FACS would be a form of bulk data analysis, i.e. multiple EBs would be combined in one sample in order to obtain enough cells for detecting the minimum of 10,000 events required for statistical rigour. This would mean I would be unable to quantify the sizes of individual EBs, which is one of the main changes observed as a result of loss of *Dnmt1* and/or *Oct4* expression. Therefore, performing both ICC and FACS would be beneficial.

Immunocytochemistry was able to reveal broad patterns of gene expression but not with as much resolution as was expected. The staining experiment was limited by the microscopes available; the wild-type EBs set

the upper limit for EB size and spread, and therefore the scope of the imaging experiments. Imaging an entire wild-type EB at day 14, when they are at their largest, with at least 20X resolution (the minimum magnification required to be able to distinguish single nuclei) demanded a microscope with image tiling capabilities. Image tiling is the process of taking sequential, partially overlapping images in a grid pattern in order to be able to capture an imaging area that is larger than the microscope's field of view. The individual image tiles are then be stitched together either *in situ* by the microscope image acquisition software or in post-processing. The largest EBs could generate up to 100 tiles, and the total size of the stitched image would be multiple GB of data. This file size is also limiting, as the processing software which had been used throughout the project, and therefore needed to continue to be used for consistency, was FIJI (Schindelin *et al.*, 2012). FIJI has a maximum memory capacity available, whatever computer specification is used, and it was less than the maximum file sizes I needed to analyse. Ideally, I wanted to use 40X magnification, but the files produced per EB were simply too large for the image acquisition and processing software to handle, required more computer memory (RAM and storage) than the microscope computers had available, was prohibitively expensive, and single images took too long to acquire (up to 30mins per EB at 20X). In future work, I would like to try staining EBs with methylene blue and scoring the sizes of the colonies, as this has recently been shown to be a reliable alternative method.

In previous chapters I observed clearly defined regions of marker staining (e.g. Figures 3.6 and 5.1 A and B), you can see pluripotent cell populations cluster in the central regions while Gata4+ cells are usually found at the EB edges), however in the *Oct4*$^{tet/tet}$ +/- dox +/- *Dnmt1* and equivalent EBRTcH3 control stains I did not see this clear demarcation for the Oct4 stain.

When I compared the sizes of EBs with and without dox, with and without *Dnmt1*, I observed that the dox-treated *Oct4*$^{tet/tet}$ EBs were always smaller whether *Dnmt1* was present or not. The loss of *Dnmt1* seemed to

cause an intermediate size disadvantage, emphasised when Oct4 was also absent. Oddly, in the control parental cell line EBs (EBRTcH3 EBs) the size differences were not as expected. The loss of *Dnmt1* did not lead to the reduction in size observed in all previously tested cell lines. This could be a feature of this sub-clone ESC line; however, it is curious that the *Oct4*$^{tet/tet}$ line derived from it did not behave the same way phenotypically in the presence or absence of Dnmt1. As the *Oct4*$^{tet/tet}$ line did show differences similar to previous experiments I can presume the lack of changes in the EBRTcH3 line was not down to technical error but is more probably due in some way to the genotype, causing them to be unsuited to EB differentiation.

In future work I would reduce the final time point to day 10 as differentiating cells of the three germ lineages are present by this stage, as evidence by beating cardiomyocytes which emerge around this time, and the EBs are slightly smaller so easier to image and process, plus takes less time to complete the experiment and uses few cell culture reagents as a result. I don't think much more information is gained by waiting an extra four days.

I would also like to test my hypothesis that the timing of *Oct4* repression is important for proper EB differentiation. In this study *Oct4*$^{tet/tet}$ ESCs were pre-treated with dox for at least 4 days prior to EB differentiation to minimise *Oct4* expression while still in the pluripotent state. In future work I would like to test if starting dox treatment at day 0 of EB differentiation affects the EB phenotype. I showed by western blot that the Oct4 protein is reduced to undetectable levels within 48 hours of treatment with 1µg/mL dox (Figure 5.6), and it is published in the literature that it takes at least 72 hours when using 10ng/mL or 1µg/mL dox (Niwa, Miyazaki and Smith, 2000). In Chapter 3 I observed that the differences in size between hypomethylated and wild type EBs started to become significant from around day 4 of culture. In Chapter 5 I observed that loss of *Oct4* alone had the most significant negative impact on EB growth rate. Therefore, I predict that starting dox treatment in *Oct4*$^{tet/tet}$ cells at day 0 of EB differentiation will lead to at least a partial rescue of the size disparity as this conditional loss of *Oct4* will be more similar to the wild type change in *Oct4* expression.

6.1 Thesis summary

Since Waddington first proposed his theory illustrating how a fixed genotype could control a plastic phenotype through a heritable epi-genotype (Waddington, 1957), we have since discovered DNA methylation to be one of the main members of this epigenetic system (Mattei, Bailly and Meissner, 2022). We now know that DNA methylation is responsible for maintaining gene repression, not only to maintain tissue-specific gene expression programs but to control imprinted genes and repress transposable elements (Greenberg and Bourc'his, 2019). We know that the system is evolutionarily conserved in multiple species across kingdoms, being found in both plants and animals (Feng *et al.*, 2010; Feng, Jacobsen and Reik, 2010). The system is also critically important for development, as experiments which knocked out the DNA methyltransferases showed this was lethal to embryos and somatic cells (Li *et al.*, 1999). Finally, misregulation of the system is associated with many human disease pathologies, including cancers (Murrell, 2006; Thomson and Meehan, 2017). All of this suggests that DNA methylation is an essential epigenetic system for mammalian cells to function. However, pluripotent ESCs can self-renew in vitro almost indefinitely in the absence of the mark, and embryos undergo multiple programmed waves of demethylation during pre-implantation development (Tsumura *et al.*, 2006; Greenberg and Bourc'his, 2019). This suggests that 5mC is dispensable under certain conditions. This somewhat paradoxical observation has raised important questions in the field of epigenetics which are still to be addressed.

The aim of this thesis was to investigate the role of DNA methylation as ESCs exit pluripotency and begin to differentiate, as little is currently known about how Dnmts affect transcription during this time. To do this I used an embryoid body (EB) model of differentiation as a proxy for early embryonic development. I also employed a range of modified cell lines which enabled the manipulation of various signalling pathways, including pluripotency and lineage markers, to dissect the function of DNA methylation. A summary of my findings are shown in figure 6.1.

The main findings of this thesis were:

- Hypomethylation induced by inhibiting *Dnmt1* expression leads to loss of EB growth which is independent of apoptosis
- Hypomethylation does not inhibit germ layer differentiation
- Hypomethylation leads to sustained aberrant expression of pluripotency genes
- Inhibition of *Oct4* expression in the absence of Dnmt1 does not rescue the EB size phenotype
- Induced hypomethylation may enrich EBs for 2C-like and PGC-like cells

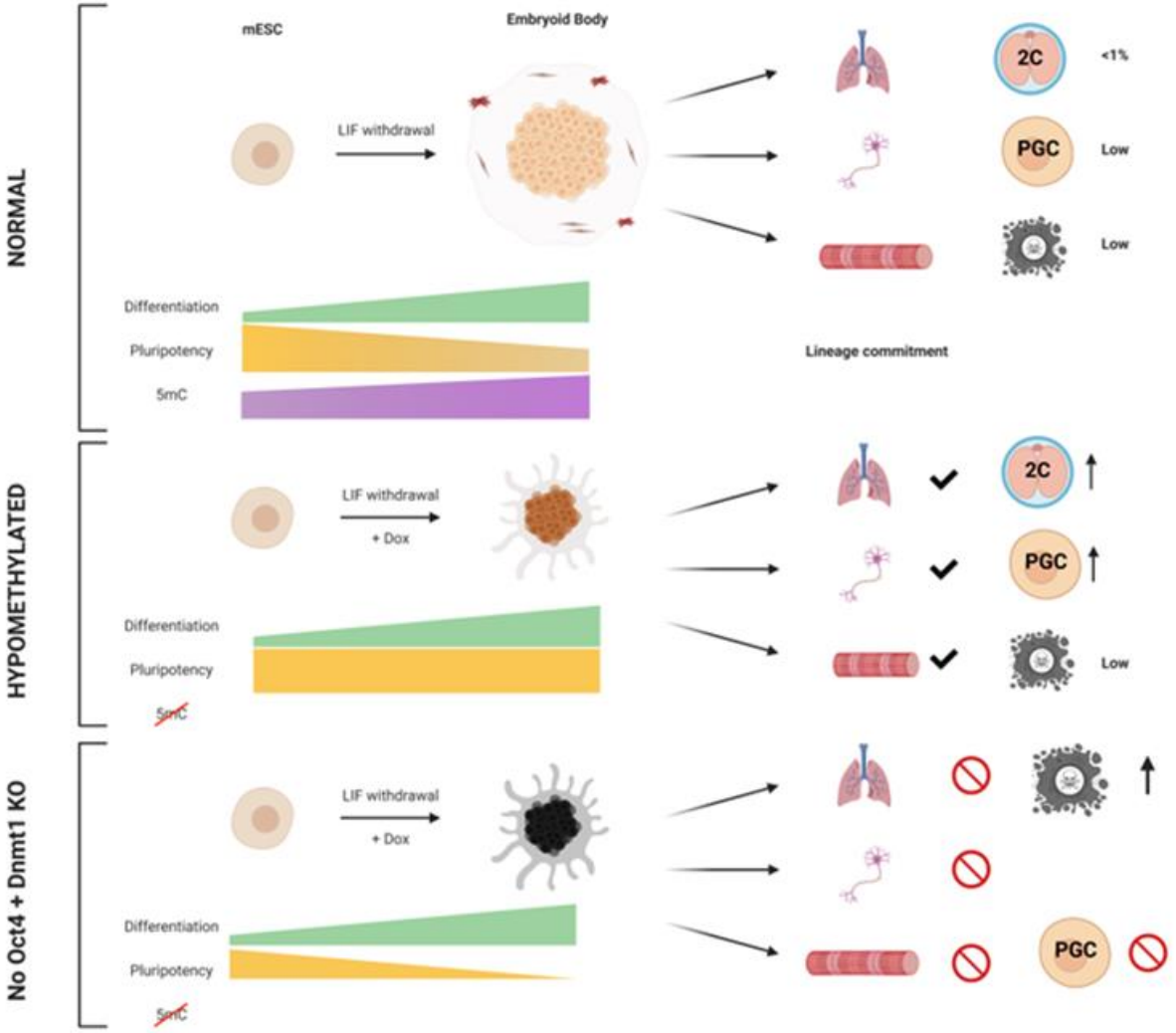


Figure 6.1. Summary of thesis results. Embryoid bodies do not require DNA methylation to develop cells of the germ layers but differentiating in a hypomethylated state results in impaired growth, and simultaneous loss of the pluripotency network leads to developmental failure. Made using BioRender.com.

6.2 DNA hypomethylation does not inhibit germ layer differentiation during embryoid body development

As DNA methylation is not required to maintain the pluripotent state, one might instead assume that the mark must therefore play a role in differentiation as cells exit pluripotency. For this reason, I investigated whether loss of *Dnmt1* changed the differentiation outcomes in developing EBs, perhaps biasing cells towards a particular lineage or blocking the emergence of one germ layer. However, in my experiments hypomethylation did not appear to skew the differentiation towards or away from any of the germ layers; expression of signature genes was still detected at the transcript level by RNA sequencing, at the protein level as indicated by the LifeAct mesoderm reporter, and in the ICC staining experiments with antibodies targeting ectoderm and endodermal proteins. Advanced mesoderm differentiation was confirmed by the visual detection of beating cardiomyocyte clusters. This was consistent with findings of Sakaue *et al*. (Sakaue *et al.*, 2010) who were able to generate cells of all three germ layers from TKO-derived EBs using retinoic acid- (RA) directed differentiation. They also observed an increased population of cells at the G2/M phase and a strong γH2A.X signal (an indicator of DNA double-strand breaks) in the RA-treated cells. They generated EBs from wild-type and TKO ESCs using the non-adherent method in 96-well plates and observed a small number of beating cardiomyocytes were formed in the TKO-derived EBs at day 5. However, their TKO EBs had a very apoptotic phenotype relative to their wild-type control. Notably, nuclear transfer experiments showed that preimplantation development occurred normally until the blastocyst stage using either TKO or wild-type ESCs as nuclear donors. They thus concluded *Dnmts* are dispensable for initiation of differentiation and specification of lineages. These experiments are not cited to the extent to which they should be as they challenge many of the general attributes for the role of DNA methylation in development noted by many reviewers. The work I have presented here may lead to those important experiments being revisited.

It is known that DNA methylation does not establish gene silencing, but rather the Dnmts are recruited to previously silenced regions and establish 5mC to further ‘lock-in’ and stabilise the repressed state. They may be recruited by other epigenetic marks, such as histone modifications, but how those first systems are recruited and, similarly, how the *de novo* methyltransferases are recruited to their nascent target sites is still not understood. The presence of these alternative epigenetic mechanisms and the unknown factors, which initially set which genes are active or silent in a specific lineage, may explain why lineage-specific gene expression patterns were still established upon differentiation in hypomethylated EBs.

Although germ layer lineage differentiation was not affected, I did observe in my RNA sequencing experiments the aberrant expression of pluripotency factors *Oct4* and *Sox2* in late stage hypomethylated EBs. These factors are normally downregulated upon activation of differentiation, and as they promote ESC self-renewal they are generally considered antagonistic to differentiation. As they were being co-expressed in the hypomethylated EBs along with differentiation genes I investigated if this combination of potentially conflicting signals was causing the severe size phenotype observed. However, simultaneous knockout of both *Dnmt1* and *Oct4* did not result in an increase in size in the hypomethylated EBs. In fact, surprisingly, they were significantly smaller than the EBs in which only *Dnmt1* was inhibited. This suggests some synergy, not antagonism, between the two systems and that at least one of these signalling networks must be initially active for EB development to proceed, though that development will be impaired without both networks present.

In the wider field of epigenetics research, very few studies have investigated the effect of *Dnmt* deficiency on EB development. This may in part be due to the association of loss of Dnmts with lethality or limited proliferation seen in somatic and embryo models. The first such study was reported in 1996 (Panning and Jaenisch, 1996); Panning *et al*. used a mutant *Dnmt1* ESC line to explore the impacts of hypomethylation on X inactivation within an EB differentiation model. They found aberrant expression of the

imprinted genes and a severe apoptotic phenotype as determined by TUNEL assay.

In a more recent 2012 study, *Dnmt1*$^{-/-}$ and TKO ESCs (lacking *Dnmt1/3a/3b* (Tsumura *et al.*, 2006)) were allowed to undergo undirected differentiation to form EBs by the hanging drop method (Schmidt *et al.*, 2012). Consistent with my data, they reported that their *Dnmt* mutant EBs were smaller, especially the TKOs. They also analysed the expression and promoter methylation of *Oct4* and *Nanog*. Unlike in my data, they showed that in all three of their lines the pluripotency network genes were downregulated in response to the differentiation initiation. They found both *Oct4* and *Nanog* were downregulated by day 4 in wild-type and *Dnmt1*$^{-/-}$ EBs (though expression was still higher in mutant vs wild-type), but interestingly this was significantly delayed in the TKO EBs. Though the initial changes they observed varied from those found here, their TKO EBs showed a similar delay in pluripotency downregulation.

As in this thesis, they asked if *Oct4* was homogenously reduced in cells of mutant EBs or if it was retained in sub-populations. They used cell dissociation and FACS to explore this. Using an ICC approach, I observed that large sub-populations of Oct4+ cells were present within central regions in both wild-type and hypomethylated EBs, though I could not quantify expression in single cells with this method. This was also observed by PhD student Olga Stepanova in an independent hypomethylated EB ICC experiment. In their study, Schmidt *et al* reported more Oct4+ cells in their *Dnmt1* mutant EBs compared to the 2% of cells which were Oct4+ in the wild-type EBs after 4 days. They concluded the *Oct4* was expressed at uniform low levels across conditions. In future work I would like to emulate their FACS experiment using my conditional *Dnmt1* EBs

They also analysed global changes in gene expression by performing micro-array analysis. Though this form of analysis is less powerful than RNA sequencing they found similar changes in expression patterns between hypomethylated *Dnmt* mutant and wild-type EBs. They found concordance in

gene expression between their day 4 EBs and wild-type ESCs; this is similar to what was observed here, the day 3 EBs had high concordance of gene expression with the untreated EBs. These similarities were then lost as the EBs underwent differentiation and their expression profiles diverged.

They reported that differentiation gene expression programs were disrupted more in the TKO mutants, suggesting what we have suspected earlier, that the Dnmt3 proteins enabled the *Dnmt1*$^{-/-}$ derived EBs to differentiate in a pattern that was more akin to the wild-type ESCs. They also noticed upregulation of genes associated with determining neural cell fates in TKO EBs, which they said was consistent with a report showing involvement of DNMT3B in human ESC neural differentiation via PRC recruitment (Martins-Taylor *et al.*, 2012). They reported an increase in terms associated with cell adhesion, suggesting cell-to-cell contact within the EBs contributes to silencing of *Oct4*, *Nanog*, and *Tet1*, and this was released upon dissociation. This is consistent with the RNA sequencing data presented here, which showed an enrichment of GO terms associated with cell adhesion in the hypomethylated EBs and may explain the relatively non-adherent phenotype of the dox-treated *Dnmt1*$^{tet/tet}$ EBs, although in my data *Oct4*, *Nanog* and *Tet1* all increased in expression with hypomethylation.

The top upregulated genes were linked with meiotic processes, such as gamete formation, and imprinted genes, though they did not explore impact on PGC formation. They too concluded that any genes which decreased in expression was due to indirect effects as DNA methylation is mostly a repressive mark. They also analysed changes between ‘early’ and ‘late’ EBs but with RNA from days 4 and 16 of differentiation; by PCA analysis they showed that longer culture time led to greater separation between the undifferentiated ESCs and the differentiating EBs, with Dnmt1$^{-/-}$ EBs holding an intermediate position between the ESCs and TKOs. Ultimately, they also concluded that DNA methylation was dispensable for differentiation to occur but that it was critical for enforcing canalisation of cell fate once established upon differentiation.

As mentioned earlier, developing embryos undergo waves of demethylation and re-methylation which together is known as epigenetic reprogramming. I attempted to emulate this by converting wild-type and hypomethylated *Dnmt1*$^{tet/tet}$ EBs at day 10 of culture to ESC culture media containing LIF to see if the EBs could revert to an ESC-like state. I analysed changes in gene expression by qPCR but was unable to obtain consistent data between sample repeats, so I have not shown the experiment in the thesis. I had hypothesised that the absence of Dnmt1 would allow the EBs to more easily reprogram to a pluripotent-like state when LIF was introduced due to the aberrant Oct4 expression observed, and to test if loss of Dnmt1 meant they would have greater differentiation plasticity. Schmidt *et al*. (Schmidt *et al.*, 2012) performed a similar experiment, reverting *Dnmt1*$^{-/-}$ and TKO dissociated EBs from day 12 of culture by culturing them in ESC media for three days. They found the mutant EBs, but not the wild-types, were able to fully reprogram to an ESC-like state by analysing changes in gene expression by qPCR. They concluded that this underscored the importance of DNA methylation for restricting developmental potential by locking in the transcriptional states adopted during differentiation. It also suggests a use for *Dnmt1* inhibition in improving the efficiency of cellular reprogramming, which may have implications for regenerative medicine.

Overall, my data, supported by data from Schmidt *et al*. (Schmidt *et al.*, 2012), Kim *et al*. (Kim *et al.*, 2020), and Sakue *et al*. (Sakaue *et al.*, 2010) suggest that DNA methylation is not required for the establishment of lineage commitment upon EB differentiation. Instead, these studies agree that the mark is more likely to act as a secondary locking mechanism to maintain a previously established transcriptional state. My data built upon these other studies by providing a model in which *Dnmt1* could be conditionally expressed or silenced, and provides further nuance to these findings through the use of additional mutation and reporter cell lines.

6.3 DNA methylation may be required for regulating the timing of expression of the PGC program to protect them from transposon activity

During this project I performed a Southern blot to check the relative levels of DNA methylation present in the EBs over 14 days of culture, with and without *Dnmt1* expression. DNA from the wild-type EBs which I expected to produce a pattern of undigested, high molecular-weight major satellite DNA bands was found to become relatively demethylated towards the middle of the culture period, accumulating some low molecular-weight bands indicative of high digestion. The DNA was digested with an enzyme whose activity is blocked by the presence of 5mC, so the transient presence of the low-molecular weight band pattern suggested the EBs had undergone their own period of demethylation. The digest pattern then returned to the original high-molecular weight band pattern. This dynamic change was unexpected and was reminiscent of a process which occurs naturally in embryos.

During embryonic development the genome undergoes multiple waves of carefully timed demethylation. The first occurs after the zygote is formed to erase the methylation patterns established in the parental gametes, resulting in a totipotent 2C embryo which then maintains this hypomethylated state until approximately the 8-cell stage, shortly after which the first lineage commitment decisions occur, and DNA methylation is re-established. The second wave occurs as the primordial germ cells are formed to erase the epiblast cell signature which must occur for the sex-specific germ cell methylation pattern to then be established.

Hargan-Calvopina *et al*.(Hargan-Calvopina *et al.*, 2016) proposed that DNA methylation has a role in preventing premature PGC emergence, perhaps to protect them from transposable element activity. Harmful transposable element integration within a developing germ cell could negatively impact fertility or harm the offspring. It is thought that this is why DNA methylation is typically retained at these transposable element regions when it is being erased elsewhere during reprogramming.

As I mentioned previously, a small number of studies have looked into the impact of loss of Dnmts specifically on gene expression and lineage specification in EBs. Fewer still have investigated the impacts on PGC formation, which we know undergo carefully orchestrated de- and re-methylation. Following the Southern blot result I tried to investigate this further. The RNA sequencing data showed hypomethylated EBs were enriched for meiotic markers. One recent study which aimed to map the gene expression patterns occurring at each of the cell fate decisions occurring in developing embryoid bodies found similar up-regulation of meiotic factors. Kim *et al*. showed that by inducing hypomethylation using 5-azacytidine treatment they could enrich the EBs for PGC-like cells, concluding that lack of DNA methylation promotes PGC specification. By using a system which specifically inhibits *Dnmt1* I was able to discern similar expression changes. Thus, I have demonstrated an effective system for further investigation into this relationship between DNA methylation and PGC specification in future studies.

6.4 EB size was not limited because of increased apoptosis but may have been a result of loss of growth factor activity in a mechanism analogous to dwarfism

It was consistently observed throughout this thesis that embryoid bodies generated without *Dnmt1*, either through genetic knockout or conditional expression, were much smaller than their wild-type counterparts. This phenotype was also found by PhD student Katie Pickup when she used a new *Dnmt1* inhibitor during EB culture using wild-type cells, which she will discuss in her thesis. Further to this, she also observed a similar size effect when she generated gastruloids under hypomethylation conditions. Gastruloids are a newly emerging differentiation model which more closely recapitulate the gastrulation-stage embryo (Chan *et al.*, 2021; van den Brink and van Oudenaarden, 2021), which will complement the EB studies which are more representative of pre-implantation development. As well as forming

multiple tissue types similarly to EBs, gastruloids also undergo symmetry breaking and polarisation. Katie will make use of the work in EBs presented here to inform her gastruloid experiments and further our knowledge of the role of DNA methylation during differentiation in this more advanced model. I initially speculated that this size phenotype was a consequence of loss of *Dnmt1* leading to activation of p53 apoptosis, which had been reported in previous studies (Jackson-Grusby *et al.*, 2001). However, multiple attempts to quantify apoptotic markers suggested this was unlikely to be the cause.

RNA sequencing then revealed that key growth factor genes, including *Igf2*, were downregulated in these EBs. Insulin-like growth factor two (*Igf2*, or *IgfII*), an imprinted gene, encodes a key protein that regulates growth, particularly during foetal development (Murrell, 2006; Sélénou *et al.*, 2022). This gene was found to be downregulated in hypomethylated EBs at both day 3 and 14. Studies have shown that misregulation of *Igf2* as a result of loss of methylation at the imprint control region leads to impaired proliferation. It has also been implicated in multiple human diseases with phenotypes showing impaired pre- and post-natal growth and foetal growth restriction (Murrell, 2006; Sélénou *et al.*, 2022). In mouse, targeted disruption of the gene was found to produce pups exhibiting proportional dwarfism (DeChiara, Efstratiadis and Robertsen, 1990; Bale and Conover, 2005). It seems promising that the loss of expression of this gene in the hypomethylated EBs, probably as a result of loss of DNA methylation at the imprinted gene promoter, could have contributed to the phenotype observed.

In future work I could investigate this further by supplementing the culture media of hypomethylated EBs with Igf2 to see if that rescues the phenotype, or if the sizes of the EBs are impacted at all. This has recently been tried in a study investigating the impact of *Nox4*, a ROS-generating enzyme, knockout on EB differentiation (Kim *et al.*, 2022). The *Nox4*$^{-/-}$ EBs displayed impaired growth. Igf2 was one of the most downregulated genes in these EBs. They found addition of Igf2 to the EB culture media led to a significant increase in the size of the EBs, though not quite to wild-type size.

6.5 Future work

This work has presented numerous new avenues that warrant further investigation, but I will discuss a few that I would particularly like to explore. Firstly, I would further investigate the impact of loss of *Igf2* expression on EB development. As mentioned earlier, this has previously been linked with knockout of the *Nox4* gene, a gene implicated in self-renewal and differentiation in ESCs through ROS generation. Knockout of the gene led to a reduction in EB size, which was then partially restored by exogenous expression of Igf2. I would be interested in seeing if the same size restoration effect is observed in the conditional *Dnmt1*$^{tet/tet}$ EBs used here in which *Igf2* was also shown to be lost upon hypomethylation.

EBs are often used as models for differentiation to derive cells of the three germ layers, but little study has gone into the specific derivation of PGC-like cells. To further consolidate the results of the RNA sequencing experiment which showed strong up-regulation of meiosis-related genes, it would be worth confirming these results with further ICC and FACS analyses to determine the degree of enrichment. This could also be applied to 2C-like cells which are very transient within ESC cultures and difficult to isolate. Dnmt1-deficient EBs could provide a model whereby isolation of these rare cells could be made easier, with implications for improving our understanding of totipotency.

Throughout my experiments hypomethylated EBs were derived from ESCs which had been converted to a hypomethylated state prior to differentiation. It is unclear what effect inducing hypomethylation at onset of differentiation would have on the gene expression pattern changes observed, or indeed on the size phenotype. Work by Kim *et al*. (Kim *et al.*, 2020) showed that there is a precise window during EB differentiation at which PGC-associated genes were up-regulated briefly prior to post-implantation differentiation. This period occurred approximately between days 4-6. As the *Dnmt1*$^{tet/tet}$ ESCs require roughly 4 days to lose DNA methylation upon initiation of dox treatment, I would hypothesis that starting dox treatment in

the EBs would lead to a maximum level of hypomethylation at approximately the time it is predicted the PGC-like genes are up-regulated. It would be interesting to see if the PGC-like cell population was still enriched in these EBs but perhaps without incurring the loss of EB growth observed when they start from a hypomethylated population.

Finally, DNA methylation is also known to interact with other epigenetic systems, such as Polycomb proteins which are also associated with gene repression. These interactions are key for proper embryonic development to occur (Atlasi and Stunnenberg, 2017). It would be interesting to attempt to unpick the crosstalk between these mechanisms using the models and cell lines I have generated here.

6.6 Final comments

DNA methylation has repeatedly been shown to play a vital role in mammalian development, and yet there are still many aspects of the mechanism and regulation of DNA methylation which are not fully understood. In this thesis I have highlighted the versatility of the embryoid model for improving our understanding of the roles of epigenetic marks during differentiation. This work also revealed that there is an unexpected requirement for simultaneous expression of both pluripotency and differentiation networks for successful EB formation, which warrants further investigation. As more advanced differentiation models emerge this work will form an important base from which to expand our understanding of the epigenetic systems during differentiation, which in turn could have important implications for human therapeutics and regenerative medicine.

References

Aapola, U. *et al.* (2000) 'Isolation and Initial Characterization of a Novel Zinc Finger Gene, DNMT3L, on 21q22.3, Related to the Cytosine-5-Methyltransferase 3 Gene Family', *Genomics*. Academic Press, 65(3), pp. 293–298. doi: 10.1006/GENO.2000.6168.

Achour, M. *et al.* (2008) 'The interaction of the SRA domain of ICBP90 with a novel domain of DNMT1 is involved in the regulation of VEGF gene expression', *Oncogene*. Oncogene, 27(15), pp. 2187–2197. doi: 10.1038/SJ.ONC.1210855.

Agrelo, R. and Wutz, A. (2010) 'X inactivation and disease', *Seminars in Cell and Developmental Biology*. Academic Press, 21(2), pp. 194–200. doi: 10.1016/j.semcdb.2009.09.017.

Allis, D. C. and Jenuwein, T. (2016) The molecular hallmarks of epigenetic control*. Nature Reviews Genetics* 17 (8).

Ambrosi, C., Manzo, M. and Baubec, T. (2017) 'Dynamics and Context-Dependent Roles of DNA Methylation', *Journal of Molecular Biology*. Academic Press, pp. 1459–1475. doi: 10.1016/j.jmb.2017.02.008.

Anderson, R. *et al.* (2000) 'The onset of germ cell migration in the mouse embryo', *Mechanisms of Development*, 91(1–2), pp. 61–68. doi: 10.1016/S0925-4773(99)00271-3.

Ashburner, M. *et al.* (2000) 'Gene ontology: Tool for the unification of biology', *Nature Genetics*. NIH Public Access, pp. 25–29. doi: 10.1038/75556.

Atlasi, Y. and Stunnenberg, H. G. (2017) 'The interplay of epigenetic marks during stem cell differentiation and development', *Nature Reviews Genetics*. Nature Publishing Group, 18(11), pp. 643–658. doi: 10.1038/nrg.2017.57.

Azevedo Portilho, N. *et al.* (2021) 'The DNMT1 inhibitor GSK-3484862

mediates global demethylation in murine embryonic stem cells', *Epigenetics and Chromatin*. BioMed Central Ltd, 14(1), doi: 10.1186/S13072-021-00429-0.

Bale, L. K. and Conover, C. A. (2005) 'Disruption of insulin-like growth factor-II imprinting during embryonic development rescues the dwarf phenotype of mice null for pregnancy-associated plasma protein-A', *The Journal of endocrinology*. J Endocrinol, 186(2), pp. 325–331. doi: 10.1677/JOE.1.06259.

Barau, J. *et al.* (2016) 'The DNA methyltransferase DNMT3C protects male germ cells from transposon activity.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 354(6314), pp. 909–912. doi: 10.1126/science.aah5143.

Barlow, D. P. and Bartolomei, M. S. (2014) 'Genomic imprinting in mammals', *Cold Spring Harbor Perspectives in Biology*. Cold Spring Harbor Laboratory Press, 6(2), p. a018382. doi: 10.1101/cshperspect.a018382.

Baubec, T. *et al.* (2015) 'Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation', *Nature 2015 520:7546*. Nature Publishing Group, 520(7546), pp. 243–247. doi: 10.1038/nature14176.

Benedict, W. F. *et al.* (1977) 'Induction of Morphological Transformation in Mouse C3H/10T/2 Clone 8 Cells and Chromosomal Damage in Hamster A(T,)C1-3 Cells by Cancer Chemotherapeutic Agents1', *CANCER RESEARCH*, 37, pp. 2202–2208.

Bernard, P. and Couturier, M. (1992) 'Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes', *Journal of Molecular Biology*. Academic Press, 226(3), pp. 735–745. doi: 10.1016/0022-2836(92)90629-X.

Bestor, T. *et al.* (1988) 'Cloning and sequencing of a cDNA encoding DNA

methyltransferase of mouse cells: The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases', *Journal of Molecular Biology*. Academic Press, 203(4), pp. 971–983. doi: 10.1016/0022-2836(88)90122-2.

Bindels, D. S. *et al.* (2016) 'mScarlet: a bright monomeric red fluorescent protein for cellular imaging', *Nature Publishing Group*, 14(1), pp. 53–56. doi: 10.1038/nmeth.4074.

Bird, A. P. (1980) 'DNA methylation and the frequency of CpG in animal DNA', *Nucleic Acids Research*. Oxford Academic, 8(7), pp. 1499–1504. doi: 10.1093/NAR/8.7.1499.

Bird, A. P. (1986) 'CpG-Rich islands and the function of DNA methylation', *Nature*. Nature Publishing Group, 321(6067), pp. 209–213. doi: 10.1038/321209a0.

Boroviak, T. *et al.* (2015) Lineage-specific profiling delineates the emergence and progression of naïve pluripotency in mammalian embryogenesis. *Developmental Cell* 35, 366-382.

Borowczyk, E. *et al.* (2009) 'Identification of a region of the DNMT1 methyltransferase that regulates the maintenance of genomic imprints.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 106(49), pp. 20806–11. doi: 10.1073/pnas.0905668106.

Bostick, M. *et al.* (2007) 'UHRF1 plays a role in maintaining DNA methylation in mammalian cells', *Science*. American Association for the Advancement of Science, 317(5845), pp. 1760–1764.

Bourc'his, D. *et al.* (2001) 'Dnmt3L and the establishment of maternal genomic imprints.', *Science (New York, N.Y.)*, 294(5551), pp. 2536–2539. doi: 10.1126/SCIENCE.1065848.

Bourque, G. *et al.* (2018) 'Ten things you should know about transposable

elements', *Genome Biology*. BioMed Central Ltd., 19(1), pp. 1–12. doi: 10.1186/s13059-018-1577-z.

Boward, B., Wu, T. and Dalton, S. (2016) 'Concise Review: Control of Cell Fate Through Cell Cycle and Pluripotency Networks', *Stem Cells*, pp. 1427–1436. doi: 10.1002/stem.2345.

Boyes, J. and Bird, A. (1991) 'DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein', *Cell*. Cell Press, 64(6), pp. 1123–1134. doi: 10.1016/0092-8674(91)90267-3.

Brickman, J. M. and Serup, P. (2017) 'Properties of embryoid bodies', *Wiley Interdisciplinary Reviews: Developmental Biology*. John Wiley and Sons Inc., 6(2), p. e259. doi: 10.1002/wdev.259.

van den Brink, S. C. and van Oudenaarden, A. (2021) '3D gastruloids: a novel frontier in stem cell-based in vitro modeling of mammalian gastrulation', *Trends in Cell Biology*. The Authors, 31(9), pp. 747–759. doi: 10.1016/j.tcb.2021.06.007.

Buecker, C. *et al.* (2014) Reorganizing of enhancer patterns in transition from naïve to primed pluripotency. *Cell Stem Cell* 14, 838-853.

Buiting, K. (2010) 'Prader-Willi syndrome and Angelman syndrome', *American journal of medical genetics. Part C, Seminars in medical genetics*. Am J Med Genet C Semin Med Genet, 154C(3), pp. 365–376. doi: 10.1002/AJMG.C.30273.

Burdon, T. *et al.* (1999) 'Signaling mechanisms regulating self-renewal and differentiation of pluripotent embryonic stem cells', *Cells Tissues Organs*, 165(3–4), pp. 131–143. doi: 10.1159/000016693.

Carbon, S. *et al.* (2021) 'The Gene Ontology resource: Enriching a GOld mine', *Nucleic Acids Research*, 49(D1), pp. D325–D334. doi: 10.1093/nar/gkaa1113.

Caspary, T. *et al.* (1998) ‘Multiple mechanisms regulate imprinting of the mouse distal chromosome 7 gene cluster’, *Molecular and cellular biology*. Mol Cell Biol, 18(6), pp. 3466–3474. doi: 10.1128/MCB.18.6.3466.

Cassidy, F. C. and Charalambous, M. (2018) ‘Genomic imprinting, growth and maternal-fetal interactions’, *The Journal of experimental biology*. J Exp Biol, 221(Pt Suppl 1). doi: 10.1242/JEB.164517.

Chan, W. K. *et al.* (2021) ‘Using organoids to study human brain development and evolution’, *Developmental Neurobiology*. John Wiley & Sons, Ltd, 81(5), pp. 608–622. doi: 10.1002/DNEU.22819.

Chazaud, C. and Yamanaka, Y. (2016) ‘Lineage specification in the mouse preimplantation embryo’, *Development (Cambridge)*, 143(7), pp. 1063–1074. doi: 10.1242/dev.128314.

Chédin, F. (2011) ‘The DNMT3 family of mammalian de novo DNA methyltransferases’, in *Progress in Molecular Biology and Translational Science*. Academic Press, pp. 255–285. doi: 10.1016/B978-0-12-387685-0.00007-X.

Chen, A. F. *et al.* (2018) GRHL2-dependent enhancer switching maintains a pluripotent stem cell transcriptional subnetwork after exit from naïve pluripotency. *Cell Stem Cell* 23, 226-238.e4

Chen, C. Y. *et al.* (2017) ‘Mechanisms of pluripotency maintenance in mouse embryonic stem cells’, *Cellular and Molecular Life Sciences*. Springer International Publishing, 74(10), pp. 1805–1817. doi: 10.1007/s00018-016-2438-0.

Chen, H. H. *et al.* (2014) ‘DAZL limits pluripotency, differentiation, and apoptosis in developing primordial germ cells’, *Stem Cell Reports*. Cell Press, 3, pp. 892–904. doi: 10.1016/j.stemcr.2014.09.003.

Chen, T., Tsujimoto, N. and Li, E. (2004) The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite

repeats at pericentromeric heterochromatin. *Molecular and Cellular Biology* 24, 9048-9058.

Choi, J. D. *et al.* (2005) 'A novel variant of Inpp5f is imprinted in brain, and its expression is correlated with differential methylation of an internal CpG island.', *Molecular and Cellular Biology*. American Society for Microbiology, 25(13), pp. 5514–5522. doi: 10.1128/MCB.25.13.5514-5522.2005.

Compere, S. J. and Palmiter, R. D. (1981) 'DNA methylation controls the inducibility of the mouse metallothionein-I gene in lymphoid cells', *Cell*. Cell Press, 25(1), pp. 233–240. doi: 10.1016/0092-8674(81)90248-8.

Constantinides, P. G., Jones, P. A. and Gevers, W. (1977) 'Functional striated muscle cells from non-myoblast precursors following 5-azacytidine treatment', *Nature 1977 267:5609*. Nature Publishing Group, 267(5609), pp. 364–366. doi: 10.1038/267364a0.

Cortázar, D. *et al.* (2011) 'Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability', *Nature 2010 470:7334*. Nature Publishing Group, 470(7334), pp. 419–423. doi: 10.1038/nature09672.

Cortellino, S. *et al.* (2011) 'Thymine DNA Glycosylase Is Essential for Active DNA Demethylation by Linked Deamination-Base Excision Repair', *Cell*. Cell Press, 146(1), pp. 67–79. doi: 10.1016/J.CELL.2011.06.020.

Coulondre, C. *et al.* (1978) 'Molecular basis of base substitution hotspots in Escherichia coli', *Nature 1978 274:5673*. Nature Publishing Group, 274(5673), pp. 775–780. doi: 10.1038/274775a0.

D'Arcy, M. S. (2019) 'Cell death: a review of the major forms of apoptosis, necrosis and autophagy', *Cell Biology International*. John Wiley & Sons, Ltd, 43(6), pp. 582–592. doi: 10.1002/CBIN.11137.

Dan, J. and Chen, T. (2016) 'Genetic studies on mammalian DNA methyltransferases', *Advances in Experimental Medicine and Biology*. Springer New York LLC, 945, pp. 123–150. doi: 10.1007/978-3-319-43624-

1_6/TABLES/1.

Deans, C. and Maggert, K. A. (2015) ‘What do you mean, “Epigenetic”?’, *Genetics*. Genetics Society of America, 199(4), pp. 887–96. doi: 10.1534/genetics.114.173492.

DeChiara, T. M., Efstratiadis, A. and Robertsen, E. J. (1990) ‘A growth-deficiency phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted by targeting’, *Nature*. Nature, 345(6270), pp. 78–80. doi: 10.1038/345078A0.

Deniz, Ö., Frost, J. M. and Branco, M. R. (2019) ‘Regulation of transposable elements by DNA modifications’, *Nature Reviews Genetics*. Nature Publishing Group, pp. 417–431. doi: 10.1038/s41576-019-0106-6.

Desbaillets, I. *et al.* (2000) ‘Embryoid Bodies: An In Vitro Model of Mouse Embryogenesis’, *Experimental Physiology*. John Wiley & Sons, Ltd, 85(6), pp. 645–651. doi: 10.1111/J.1469-445X.2000.02104.X.

Doetschman, T. C. *et al.* (1985) *The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium*, *Embryol. exp. Morph*. Available at: https://dev.biologists.org/content/develop/87/1/27.full.pdf.

Donovan, P. J. *et al.* (1986) ‘Migratory and postmigratory mouse primordial germ cells behave differently in culture’, *Cell*. Elsevier, 44(6), pp. 831–838. doi: 10.1016/0092-8674(86)90005-X.

Dunican, D. S. *et al.* (2013) ‘Lsh regulates LTR retrotransposon repression independently of Dnmt3b function’, *Genome Biology*. BioMed Central Ltd., 14(12), pp. 1–23. doi: 10.1186/gb-2013-14-12-r146.

Dunican, D. S. *et al.* (2020) ‘Bivalent promoter hypermethylation in cancer is linked to the H327me3/H3K4me3 ratio in embryonic stem cells’, *BMC biology*. NLM (Medline), 18(1), p. 25. doi: 10.1186/s12915-020-0752-3.

Eckersley-Maslin, M. *et al.* (2019) ‘Dppa2 and Dppa4 directly regulate the Dux-driven zygotic transcriptional program’, *Genes and Development*. Cold Spring Harbor Laboratory Press, 33(3–4), pp. 194–208. doi: 10.1101/gad.321174.118.

Eckersley-Maslin, M. A. *et al.* (2016) ‘MERVL/Zscan4 Network Activation Results in Transient Genome-wide DNA Demethylation of mESCs’, *Cell Reports*. Cell Press, 17(1), pp. 179–192. doi: 10.1016/j.celrep.2016.08.087.

Ehrlich, M. *et al.* (1982) ‘Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells’, *Nucleic Acids Research*, 10(8), pp. 2709–2721. doi: 10.1093/nar/10.8.2709.

Evans, M. J. and Kaufman, M. H. (1981) ‘Establishment in culture of pluripotential cells from mouse embryos’, *Nature*. Nature Publishing Group, 292(5819), pp. 154–156. doi: 10.1038/292154a0.

Evsikov, A. V. *et al.* (2004) ‘Systems biology of the 2-cell mouse embryo’, *Cytogenetic and genome research*. Cytogenet Genome Res, 105(2–4), pp. 240–250. doi: 10.1159/000078195.

Fang, X. *et al.* (2021) Knockdown of DNA methyltransferase 1 reduces DNA methylation and alters expression patterns of cardiac genes in embryonic cardiomyocytes. FEBS Open Bio 11, 2364–2382.

Fang, X. *et al.* (2016) Knockdown of DNA methyltransferase 3a alters gene expression and inhibits function of embryonic cardiomyocytes. FASEB Journal 30, 3238–3255.

Feng, S. *et al.* (2010) ‘Conservation and divergence of methylation patterning in plants and animals’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 107(19), pp. 8689–8694.

Feng, S., Jacobsen, S. E. and Reik, W. (2010) ‘Epigenetic reprogramming in plant and animal development’, *Science*, 330(6004), pp. 622–627. doi:

10.1126/science.1190614.

Festuccia, N., Gonzalez, I. and Navarro, P. (2017) ‘The Epigenetic Paradox of Pluripotent ES Cells’, *Journal of Molecular Biology*. Academic Press, 429(10), pp. 1476–1503. doi: 10.1016/j.jmb.2016.12.009.

Filion, G. J. P. *et al.* (2006) ‘A family of human zinc finger proteins that bind methylated DNA and repress transcription.’, *Molecular and Cellular Biology*. American Society for Microbiology, 26(1), pp. 169–181. doi: 10.1128/MCB.26.1.169-181.2006.

Fowden, A. L. *et al.* (2011) ‘Imprinted genes and the epigenetic regulation of placental phenotype’, *Progress in biophysics and molecular biology*. Prog Biophys Mol Biol, 106(1), pp. 281–288. doi: 10.1016/J.PBIOMOLBIO.2010.11.005.

Friedman, S. (1979) ‘The effect of 5-azacytidine on E.coli DNA methylase’, *Biochemical and Biophysical Research Communications*. Academic Press, 89(4), pp. 1328–1333. doi: 10.1016/0006-291X(79)92154-5.

Fu, X. *et al.* (2019) ‘Myc and Dnmt1 impede the pluripotent to totipotent state transition in embryonic stem cells’, *Nature Cell Biology*. Nature Publishing Group, 21(7), pp. 835–844. doi: 10.1038/s41556-019-0343-0.

Fu, X. *et al.* (2020) ‘A transcriptional roadmap for 2C-like-to-pluripotent state transition’, *Science Advances*, 6(22), pp. 5181–5210. doi: 10.1126/sciadv.aay5181.

Fu, X., Zhang, C. and Zhang, Y. (2020) ‘Epigenetic regulation of mouse preimplantation embryo development’, *Current Opinion in Genetics & Development*. Elsevier Current Trends, 64, pp. 13–20. doi: 10.1016/J.GDE.2020.05.015.

Fuks, F. *et al.* (2000) ‘DNA methyltransferase Dnmt1 associates with histone deacetylase activity’, *Nature genetics*. Nat Genet, 24(1), pp. 88–91. doi: 10.1038/71750.

Fuks, F. *et al.* (2003) ‘The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase.’, *Nucleic Acids Research*, 31(9), pp. 2305–2312. doi: 10.1093/NAR/GKG332.

Furusawa, T. *et al.* (2006) ‘Gene expression profiling of mouse embryonic stem cell subpopulations.’, *Biology of Reproduction*, 75(4), pp. 555–561. doi: 10.1095/BIOLREPROD.105.049502.

Gardiner-Garden, M. and Frommer, M. (1987) ‘CpG Islands in vertebrate genomes’, *Journal of Molecular Biology*. Academic Press, 196(2), pp. 261–282. doi: 10.1016/0022-2836(87)90689-9.

Gartler, S. M. and Riggs, A. D. (1983) ‘Mammalian X-chromsome Inactivation’, *Annual Review of Genetics*, 17, pp. 155–190.

Gaudet, F. *et al.* (2004) ‘Dnmt1 expression in pre- and postimplantation embryogenesis and the maintenance of IAP silencing’, *Molecular and cellular biology*. Mol Cell Biol, 24(4), pp. 1640–1648. doi: 10.1128/MCB.24.4.1640-1648.2004.

Gautsch, J. and Wilson, M. (1983) ‘Delayed de novo methylation in teratocarcinoma suggests additional tissue-specific mechanisms for controlling gene expression’, *Nature 1983 301:5895*. Nature Publishing Group, 301(5895), pp. 32–37. doi: 10.1038/301032a0.

Gavrieli, Y., Sherman, Y. and Ben-Sasson, S. A. (1992) ‘Identification of programmed cell death in situ via specific labeling of nuclear DNA fragmentation’, *Journal of Cell Biology*, 119(3), pp. 493–501. doi: 10.1083/jcb.119.3.493.

Ge, Y. Z. *et al.* (2004) ‘Chromatin Targeting of de Novo DNA Methyltransferases by the PWWP Domain’, *Journal of Biological Chemistry*. Elsevier, 279(24), pp. 25447–25454. doi: 10.1074/JBC.M312296200.

Geiman, T. M. *et al.* (2004) ‘DNMT3B interacts with hSNF2H chromatin remodeling enzyme, HDACs 1 and 2, and components of the histone

methylation system', *Biochemical and Biophysical Research Communications*. Biochem Biophys Res Commun, 318(2), pp. 544–555. doi: 10.1016/j.bbrc.2004.04.058.

Genet, M. and Torres-Padilla, M. E. (2020) 'The molecular and cellular features of 2-cell-like cells: a reference guide', *Development (Cambridge, England)*. NLM (Medline). doi: 10.1242/dev.189688.

Gilsbach, R. *et al.* (2014) Dynamic DNA methylation orchestrates cardiomyocyte development, maturation and disease. Nature Communications 5, 1–13.

Ginsburg, M., Snow, M. H. L. and McLaren, A. (1990) 'Primordial germ cells in the mouse embryo during gastrulation', *Development (Cambridge, England)*. Development, 110(2), pp. 521–528. doi: 10.1242/DEV.110.2.521.

Goedhart, J. *et al.* (2012) 'Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%', *Nature Communications*. Nature Publishing Group, 3(1), pp. 1–9. doi: 10.1038/ncomms1738.

Gökbuget, D. and Blelloch, R. (2019) 'Epigenetic control of transcriptional regulation in pluripotency and early differentiation', *Development (Cambridge)*. Company of Biologists Ltd. doi: 10.1242/dev.164772.

Goto, K. *et al.* (1994) 'Expression of DNA methyltransferase gene in mature and immature neurons as well as proliferating cells in mice', *Differentiation*. Elsevier, 56(1–2), pp. 39–44. doi: 10.1046/J.1432-0436.1994.56120039.X.

Greally, J. M. (2018) 'A user's guide to the ambiguous word "epigenetics"', *Nature Reviews Molecular Cell Biology*, pp. 207–208. doi: 10.1038/nrm.2017.135.

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

Guo, J. U. *et al.* (2013) ‘Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain.’, *Nature Neuroscience*, 17(2), pp. 215–222. doi: 10.1038/NN.3607.

Guo, X. *et al.* (2014) ‘Structural insight into autoinhibition and histone H3-induced activation of DNMT3A’, *Nature 2014 517:7536*. Nature Publishing Group, 517(7536), pp. 640–644. doi: 10.1038/nature13899.

Gurdon, J. B. and Uehlinger, V. (1966) ‘“Fertile” Intestine Nuclei’, *Nature 1966 210:5042*. Nature Publishing Group, 210(5042), pp. 1240–1241. doi: 10.1038/2101240a0.

Guseinov, V. A. and Vanyushin, B. F. (1975) ‘Content and localisation of 5-methylcytosine in DNA of healthy and wilt-infected cotton plants’, *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis*. Elsevier, 395(3), pp. 229–238. doi: 10.1016/0005-2787(75)90193-8.

Hackett, J. A. *et al.* (2012) ‘Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline.’, *Development (Cambridge, England)*. Oxford University Press for The Company of Biologists Limited, 139(19), pp. 3623–32. doi: 10.1242/dev.081661.

Hackett, J. A. *et al.* (2013) ‘Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine’, *Science*. American Association for the Advancement of Science, 339(6118), pp. 448–452. doi: 10.1126/science.1229277.

Hackett, J. A. and Surani, M. A. (2014) ‘Regulatory Principles of Pluripotency: From the Ground State Up’, *Cell Stem Cell*. Cell Press, 15(4), pp. 416–430. doi: 10.1016/J.STEM.2014.09.015.

Harbers, K. *et al.* (1981) ‘DNA methylation and gene expression: Endogenous retroviral genome becomes infectious after molecular cloning’, *Proceedings of the National Academy of Sciences of the United States of*

America. Proceedings of the National Academy of Sciences, 78(12 II), pp. 7609–7613. doi: 10.1073/pnas.78.12.7609.

Hargan-Calvopina, J. *et al.* (2016) 'Stage-Specific Demethylation in Primordial Germ Cells Safeguards against Precocious Differentiation', *Developmental Cell*. Cell Press, 39(1), pp. 75–86. doi: 10.1016/j.devcel.2016.07.019.

Hashimoto, H. *et al.* (2008) 'The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix.', *Nature*, 455(7214), pp. 826–829. doi: 10.1038/NATURE07280.

Hashimoto, H. *et al.* (2009) 'UHRF1, a modular multi-domain protein, regulates replication-coupled crosstalk between DNA methylation and histone modifications', *Epigenetics : official journal of the DNA Methylation Society*. NIH Public Access, 4(1), p. 8. doi: 10.4161/EPI.4.1.7370.

Hata, K. *et al.* (2002) 'Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice', *Development*. The Company of Biologists, pp. 1983–1993. doi: 10.1242/dev.129.8.1983.

Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. (2011) Reconstitution of the Mouse Germ Cell Specification Pathway in Culture by Pluripotent Stem Cells. Cell 146, 519–532.

He, Y. F. *et al.* (2011) 'Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA', *Science*. American Association for the Advancement of Science, 333(6047), pp. 1303–1307.

Hendrich, B. and Bird, A. (1998) 'Identification and Characterization of a Family of Mammalian Methyl-CpG Binding Proteins', *Molecular and Cellular Biology*. American Society for Microbiology, 18(11), pp. 6538–6547.

Hendrickson, P. G. *et al.* (2017) 'Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL

retrotransposons', *Nature genetics*. Nat Genet, 49(6), pp. 925–934. doi: 10.1038/NG.3844.

Hermann, A., Goyal, R. and Jeltsch, A. (2004) 'The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites', *Journal of Biological Chemistry*. Elsevier, 279(46), pp. 48350–48359. doi: 10.1074/jbc.M403427200.

Holliday, R. (1994) 'Epigenetics: An overview', *Developmental Genetics*. John Wiley & Sons, Ltd, 15(6), pp. 453–457. doi: 10.1002/DVG.1020150602.

Holliday, R. and Pugh, J. E. (1975) 'DNA Modification Mechanisms and Gene Activity During Development', *Science*. American Association for the Advancement of Science, 187(4173), pp. 226–232. doi: 10.1126/SCIENCE.187.4173.226.

Hooper, M. *et al.* (1987) 'HPRT-deficient (Lesch–Nyhan) mouse embryos derived from germline colonization by cultured cells', *Nature*. Nature Publishing Group, 326, pp. 292–295. doi: 10.1038/326292a0.

Hotchkiss, R. D. (1948) 'THE QUANTITATIVE SEPARATION OF PURINES, PYRIMIDINES, AND NUCLEOSIDES BY PAPER CHROMATOGRAPHY', *Journal of Biological Chemistry*. Elsevier, 175(1), pp. 315–332. doi: 10.1016/S0021-9258(18)57261-6.

Howell, C. Y. *et al.* (2001) 'Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene', *Cell*. Cell Press, 104(6), pp. 829–838. doi: 10.1016/S0092-8674(01)00280-X.

Hutnick, L. K. *et al.* (2010) 'Repression of retrotransposal elements in mouse embryonic stem cells is primarily mediated by a DNA methylation-independent mechanism.', *The Journal of Biological Chemistry*, 285(27), pp. 21082–21091. doi: 10.1074/JBC.M110.125674.

Illingworth, R. S. *et al.* (2010) 'Orphan CpG Islands Identify Numerous Conserved Promoters in the Mammalian Genome', *PLOS Genetics*. Public

Library of Science, 6(9), p. e1001134. doi: 10.1371/JOURNAL.PGEN.1001134.

Irie, N., Tang, W. W. C. & Azim Surani, M. (2014) Germ cell specification and pluripotency in mammals: a perspective from early embryogenesis. Reproductive Medicine and Biology 13, 203–215.

Isbel, L., Grand, R. S. and Schübeler, D. (2022) 'Generating specificity in genome regulation through transcription factor sensitivity to chromatin.', *Nature reviews. Genetics*. doi: 10.1038/s41576-022-00512-6.

Ishiuchi, T. *et al.* (2015) 'Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly', *Nature Structural & Molecular Biology 2015 22:9*. Nature Publishing Group, 22(9), pp. 662–671. doi: 10.1038/nsmb.3066.

Ito, S. *et al.* (2010) 'Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification', *Nature 2010 466:7310*. Nature Publishing Group, 466(7310), pp. 1129–1133. doi: 10.1038/nature09303.

Ito, S. *et al.* (2011) 'Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine', *Science*. American Association for the Advancement of Science, 333(6047), pp. 1300–1303.

Jackson-Grusby, L. *et al.* (2001) 'Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation', *Nature Genetics 2001 27:1*. Nature Publishing Group, 27(1), pp. 31–39. doi: 10.1038/83730.

Jackson, M. *et al.* (2010) 'The culture of mouse embryonic stem cells and formation of embryoid bodies.', *Methods in molecular biology (Clifton, N.J.)*. Humana Press, 633, pp. 1–18. doi: 10.1007/978-1-59745-019-5_1.

Jähner, D. *et al.* (1982) 'De novo methylation and expression of retroviral genomes during mouse embryogenesis', *Nature 1982 298:5875*. Nature Publishing Group, 298(5875), pp. 623–628. doi: 10.1038/298623a0.

Jia, D. *et al.* (2007) ‘Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation’, *Nature 2007 449:7159*. Nature Publishing Group, 449(7159), pp. 248–251. doi: 10.1038/nature06146.

Johnson, T. B. and Coghill, R. D. (1925) ‘Researches on pyrimidines. C111. The discovery of 5-methyl-cytosine in tuberculinic acid, the nucleic acid of the tubercle bacillus’, *Journal of the American Chemical Society*. American Chemical Society, 47(11), pp. 2838–2844. doi: 10.1021/ja01688a030.

Jones, P. A. and Taylor, S. M. (1980) ‘Cellular differentiation, cytidine analogs and DNA methylation’, *Cell*. Cell Press, 20(1), pp. 85–93. doi: 10.1016/0092-8674(80)90237-8.

Jurkowska, R. Z. and Jeltsch, A. (2016) ‘Mechanisms and biological roles of DNA methyltransferases and DNA methylation: From past achievements to future challenges’, *Advances in Experimental Medicine and Biology*. Springer New York LLC, 945, pp. 1–17.

Kaneda, M. *et al.* (2004) ‘Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting’, *Nature*, 429(6994), pp. 900–903. doi: 10.1038/NATURE02633.

Kantor, B. *et al.* (2004) ‘Establishing the epigenetic status of the Prader–Willi/Angelman imprinting center in the gametes and embryo’, *Human Molecular Genetics*. Oxford University Press, 13(22), pp. 2767–2779. doi: 10.1093/hmg/ddh290.

Kelly, T. K., De Carvalho, D. D. and Jones, P. A. (2010) ‘Epigenetic modifications as therapeutic targets’, *Nature Biotechnology*. Nature Publishing Group, pp. 1069–1078. doi: 10.1038/nbt.1678.

Kerr, J. F. R., Wyllie, A. H. & Currie, A. R. (1972) Apoptosis: A basic biological phenomenon with wide-ranging implications in tissue kinetics. British Journal of Cancer 26, 239–257.

Kigami, D. *et al.* (2003) ‘MuERV-L is one of the earliest transcribed genes in

mouse one-cell embryos', *Biology of reproduction*. Biol Reprod, 68(2), pp. 651–654. doi: 10.1095/BIOLREPROD.102.007906.

Kim, I. S. *et al.* (2020) 'Parallel Single-Cell RNA-Seq and Genetic Recording Reveals Lineage Decisions in Developing Embryoid Bodies', *Cell Reports*. Elsevier B.V., 33(1), p. 108222. doi: 10.1016/j.celrep.2020.108222.

Kim, Jusong *et al.* (2022) 'Nox4-IGF2 Axis Promotes Differentiation of Embryoid Body Cells Into Derivatives of the Three Embryonic Germ Layers', *Stem Cell Reviews and Reports*. Stem Cell Rev Rep, 18(3), pp. 1181–1192. doi: 10.1007/s12015-021-10303-x.

Kim, S. H. *et al.* (2004) 'Differential DNA methylation reprogramming of various repetitive sequences in mouse preimplantation embryos', *Biochemical and biophysical research communications*. Biochem Biophys Res Commun, 324(1), pp. 58–63. doi: 10.1016/J.BBRC.2004.09.023.

Kimura, H. and Shiota, K. (2003) 'Methyl-CpG-binding protein, MeCP2, is a target molecule for maintenance DNA methyltransferase, Dnmt1', *The Journal of biological chemistry*. J Biol Chem, 278(7), pp. 4806–4812. doi: 10.1074/JBC.M209923200.

Klingseisen, A. and Jackson, A. P. (2011) 'Mechanisms and pathways of growth failure in primordial dwarfism', *Genes and Development*, 25(19), pp. 2011–2024. doi: 10.1101/gad.169037.

Kobayakawa, S. *et al.* (2007) 'Dynamic changes in the epigenomic state and nuclear organization of differentiating mouse embryonic stem cells', *Genes to Cells*, 12(4), pp. 447–460. doi: 10.1111/j.1365-2443.2007.01063.x.

Kojima, Y., Tam, O. H. and Tam, P. P. L. (2014) 'Timing of developmental events in the early mouse embryo', *Seminars in Cell & Developmental Biology*. Academic Press, 34, pp. 65–75. doi: 10.1016/J.SEMCDB.2014.06.010.

Koopman, G. *et al.* (1994) Annexin V for flow cytometric detection of

phosphatidylserine expression on B cells undergoing apoptosis. Blood 84, 1415–1420.

Kuo, M. T., Mandel, J. L. and Chambon, P. (1979) ‘DNA methylation: correlation with DNase I sensitivity of chicken ovalbumin and conalbumin chromatin’, *Nucleic Acids Research*. Oxford Academic, 7(8), pp. 2105–2113. doi: 10.1093/NAR/7.8.2105.

Kurosawa, H. (2007) ‘Methods for inducing embryoid body formation: in vitro differentiation system of embryonic stem cells’, *Journal of Bioscience and Bioengineering*. Elsevier, 103(5), pp. 389–398. doi: 10.1263/jbb.103.389.

Kylarová, D. *et al.* (2002) ‘Comparison of the TUNEL, lamin B and annexin V methods for the detection of apoptosis by flow cytometry’, *Acta Histochemica*. Urban & Fischer, 104(4), pp. 367–370. doi: 10.1078/0065-1281-00674.

Labun, K. *et al.* (2019) ‘CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing’, *Nucleic Acids Research*. Oxford Academic, 47(W1), pp. W171–W174. doi: 10.1093/NAR/GKZ365.

Lei, H. *et al.* (1996) ‘De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells’, *Development*. The Company of Biologists, 122(10), pp. 3195–3205. doi: 10.1242/DEV.122.10.3195.

Leitch, H. G., Tang, W. W. C. and Surani, M. A. (2013) ‘Primordial Germ-Cell Development and Epigenetic Reprogramming in Mammals’, in *Current Topics in Developmental Biology*. Academic Press Inc., pp. 149–187. doi: 10.1016/B978-0-12-416027-9.00005-X.

Lemieux, M. G. *et al.* (2014) Visualization of the actin cytoskeleton: Different F-actin-binding probes tell different stories. Cytoskeleton 71, 157–169.

Lewis, J. D. *et al.* (1992) ‘Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA.’, *Cell*, 69(6), pp. 905–914. doi: 10.1016/0092-8674(92)90610-O.

Li, E., Beard, C. and Jaenisch, R. (1993) ‘Role for DNA methylation in genomic imprinting’, *Nature*. Nature Publishing Group, 366(6453), pp. 362–365. doi: 10.1038/366362a0.

Li, E., Bestor, Timothy H. and Jaenisch, R. (1992) ‘Targeted mutation of the DNA methyltransferase gene results in embryonic lethality’, *Cell*. Cell Press, 69(6), pp. 915–926. doi: 10.1016/0092-8674(92)90611-F.

Li, E., Bestor, Timothy H and Jaenisch, R. (1992) *Targeted Mutation of the DNA Methyltransferase Gene Results in Embryonic Lethality*, *Cell*. Available at: https://ac.els-cdn.com/009286749290611F/1-s2.0-009286749290611F-main.pdf?_tid=e8858e98-101f-4c11-abe7-054b4020a5b2&acdnat=1532598697_11372af618c7d00a76ed6fe250f89492 (Accessed: 26 July 2018).

Li, E. and Zhang, Y. (2014) ‘DNA methylation in mammals’, *Cold Spring Harbor Perspectives in Biology*. Cold Spring Harbor Laboratory Press, 6(5), p. a019133. doi: 10.1101/cshperspect.a019133.

Li, X. *et al.* (2008) ‘A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints.’, *Developmental Cell*, 15(4), pp. 547–557. doi: 10.1016/J.DEVCEL.2008.08.014.

Lienert, F. *et al.* (2011) ‘Identification of genetic elements that autonomously determine DNA methylation states’, *Nature Genetics*. Nature Publishing Group, 43(11), pp. 1091–1097. doi: 10.1038/ng.946.

Lister, R. *et al.* (2009) ‘Human DNA methylomes at base resolution show widespread epigenomic differences’, *Nature*. Nature Publishing Group, 462(7271), pp. 315–322. doi: 10.1038/nature08514.

Love, M. I., Huber, W. and Anders, S. (2014) ‘Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2’, *Genome Biology*. BioMed Central Ltd., 15(12), pp. 1–21. doi: 10.1186/s13059-014-0550-8.

Lu, F. and Zhang, Y. (2015) ‘Cell totipotency: Molecular features, induction,

and maintenance', *National Science Review*, pp. 217–225. doi: 10.1093/nsr/nwv009.

Lucifero, D. *et al.* (2004) 'Gene-specific timing and epigenetic memory in oocyte imprinting', *Human molecular genetics*. Hum Mol Genet, 13(8), pp. 839–849. doi: 10.1093/HMG/DDH104.

Lynch, A. T., Mazzotta, S. and Hoppler, S. (2018) 'Cardiomyocyte differentiation from mouse embryonic stem cells', *Methods in Molecular Biology*. Humana Press Inc., 1816, pp. 55–66. doi: 10.1007/978-1-4939-8597-5_4.

Maatouk, D. M. *et al.* (2006) DNA methylation is a primary mechanism for silencing postmigratory primordial germ cell genes in both germ cell and somatic cell lineages. Development 133, 3411–3418.

Maatouk, D. M. & Resnick, J. L. (2003) Continuing primordial germ cell differentiation in the mouse embryo is a cell-intrinsic program sensitive to DNA methylation. Developmental Biology 258, 201–208.

Magnúsdóttir, E. *et al.* (2013) A tripartite transcription factor network regulates primordial germ cell specification in mice. Nature Cell Biology 15, 905– 915.

Mandel, J. L. and Chambon, P. (1979) 'DNA methylation: organ specific variations in the methylation pattern within and around ovalbumin and other chicken genes', *Nucleic Acids Research*. Oxford Academic, 7(8), pp. 2081–2103. doi: 10.1093/NAR/7.8.2081.

Mangiavacchi, A. *et al.* (2021) 'New insights into the functional role of retrotransposon dynamics in mammalian somatic cells', *Cellular and Molecular Life Sciences*. Springer, pp. 5245–5256. doi: 10.1007/s00018-021-03851-5.

Martins-Taylor, K. *et al.* (2012) 'Role of DNMT3B in the regulation of early neural and neural crest specifiers', *Epigenetics*. Taylor and Francis Inc., 7(1),

pp. 71–82. doi: 10.4161/epi.7.1.18750.

Masui, S. *et al.* (2005) ‘An efficient system to establish multiple embryonic stem cell lines carrying an inducible expression unit’, *Nucleic Acids Research*, 33(4), pp. 1–8. doi: 10.1093/nar/gni043.

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, Alexei A, *et al.* (2007) ‘Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells’, *NATURE CELL BIOLOGY*, 9. doi: 10.1038/ncb1589.

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, Alexei A., *et al.* (2007) ‘Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells’, *Nature Cell Biology*, 9(6), pp. 625–635. doi: 10.1038/ncb1589.

Matassov, D. *et al.* (2004) ‘Measurement of apoptosis by DNA fragmentation’, *Methods in molecular biology (Clifton, N.J.)*. Humana Press, 282, pp. 1–17. doi: 10.1385/1-59259-812-9:001/FIGURES/3.

Matoba, R. *et al.* (2006) ‘Dissecting Oct3/4-Regulated Gene Networks in Embryonic Stem Cells by Expression Profiling’, *PLOS ONE*. Public Library of Science, 1(1), p. e26. doi: 10.1371/JOURNAL.PONE.0000026.

Mattei, A. L., Bailly, N. and Meissner, A. (2022) ‘DNA methylation: a historical perspective’, *Trends in Genetics*. Elsevier, pp. 676–707. doi: 10.1016/j.tig.2022.03.010.

McGhee, J. D. and Ginder, G. D. (1979) ‘Specific DNA methylation sites in the vicinity of the chicken β-globin genes’, *Nature 1979 280:5721*. Nature Publishing Group, 280(5721), pp. 419–420. doi: 10.1038/280419a0.

Meehan, R. R. *et al.* (1989) ‘Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs.’, *Cell*, 58(3), pp. 499–507. doi: 10.1016/0092-8674(89)90430-3.

Meissner, A. *et al.* (2008) ‘Genome-scale DNA methylation maps of pluripotent and differentiated cells’, *Nature*. Nature Publishing Group, 454(7205), pp. 766–770. doi: 10.1038/nature07107.

Messerschmidt, D. M., Knowles, B. B. & Solter, D. (2014) DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. Genes and Development vol. 28 812–828.

Mikkelsen, T. S. *et al.* (2007) ‘Genome-wide maps of chromatin state in pluripotent and lineage-committed cells’, *Nature*. Nature Publishing Group, 448(7153), pp. 553–560. doi: 10.1038/nature06008.

Mohn, F. *et al.* (2008) ‘Lineage-Specific Polycomb Targets and De Novo DNA Methylation Define Restriction and Potential of Neuronal Progenitors’, *Molecular Cell*. Cell Press, 30(6), pp. 755–766. doi: 10.1016/J.MOLCEL.2008.05.007.

Molaro, A. and Malik, H. S. (2016) ‘Hide and seek: how chromatin-based pathways silence retroelements in the mammalian germline’, *Current Opinion in Genetics & Development*. Elsevier Current Trends, 37, pp. 51–58. doi: 10.1016/J.GDE.2015.12.001.

Monk, M., Boubelik, M. and Lehnert, S. (1987) ‘Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development’, *Development*. The Company of Biologists, 99(3), pp. 371–382. doi: 10.1242/DEV.99.3.371.

Montague, T. G. *et al.* (2014) ‘CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing’, *Nucleic Acids Research*. Oxford Academic, 42(W1), pp. W401–W407. doi: 10.1093/NAR/GKU410.

Moore, L. D., Le, T. and Fan, G. (2013) ‘DNA methylation and its basic function’, *Neuropsychopharmacology*, 38(1), pp. 23–38. doi: 10.1038/npp.2012.112.

Mooren, F. C. and Krüger, K. (2015) ‘Exercise, Autophagy, and Apoptosis’, in

Progress in Molecular Biology and Translational Science. Elsevier B.V., pp. 407–422. doi: 10.1016/bs.pmbts.2015.07.023.

Morange, M. (2009) 'What history tells us XVII. Conrad Waddington and the nature of life', *Journal of biosciences*. J Biosci, 34(2), pp. 195–198. doi: 10.1007/S12038-009-0022-6.

Morris, K. V. *et al.* (2004) 'Small interfering RNA-induced transcriptional gene silencing in human cells', *Science*. American Association for the Advancement of Science, 305(5688), pp. 1289–1292.

Mortusewicz, O. *et al.* (2005) 'Recruitment of DNA methyltransferase I to DNA repair sites', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 102(25), p. 8905. doi: 10.1073/PNAS.0501034102.

Murrell, A. (2006) 'Genomic Imprinting and Cancer: From Primordial Germ Cells to Somatic Cells', *The Scientific World Journal*. Hindawi Limited, 6, p. 1888. doi: 10.1100/TSW.2006.318.

Murry, C. E. and Keller, G. (2008) 'Differentiation of Embryonic Stem Cells to Clinically Relevant Populations: Lessons from Embryonic Development', *Cell*. Cell Press, 132(4), pp. 661–680. doi: 10.1016/j.cell.2008.02.008.

Muto, M. *et al.* (2002) 'Targeted Disruption of Np95 Gene Renders Murine Embryonic Stem Cells Hypersensitive to DNA Damaging Agents and DNA Replication Blocks', *Journal of Biological Chemistry*. Elsevier, 277(37), pp. 34549–34555. doi: 10.1074/JBC.M205189200.

Nakamura, T. *et al.* (2006) 'PGC7/Stella protects against DNA demethylation in early embryogenesis', *Nature Cell Biology 2006 9:1*. Nature Publishing Group, 9(1), pp. 64–71. doi: 10.1038/ncb1519.

Nan, X. *et al.* (1998) 'Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex', *Nature*. Nature, 393(6683), pp. 386–389. doi: 10.1038/30764.

Nan, X., Meehan, R. R. and Bird, A. (1993) ‘Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2.’, *Nucleic Acids Research*, 21(21), pp. 4886–4892. doi: 10.1093/NAR/21.21.4886.

Nanney, D. L. (1958) ‘Epigenetic Control Systems’, *Proceedings of the National Academy of Sciences*. Proceedings of the National Academy of Sciences, 44(7), pp. 712–717.

Ng, H. H. *et al.* (1999) ‘MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex’, *Nature genetics*. Nat Genet, 23(1), pp. 58–61. doi: 10.1038/12659.

Nishiyama, A. *et al.* (2013) ‘Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication’, *Nature*. Nature Publishing Group, 502(7470), pp. 249–253. doi: 10.1038/nature12488.

Niwa, H., Miyazaki, J. I. and Smith, A. G. (2000) ‘Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells’, *Nature Genetics*, 24(4), pp. 372–376. doi: 10.1038/74199.

Niwa, O. *et al.* (1983) ‘Independent mechanisms involved in suppression of the moloney leukemia virus genome during differentiation of murine teratocarcinoma cells’, *Cell*. Cell Press, 32(4), pp. 1105–1113. doi: 10.1016/0092-8674(83)90294-5.

Okano, M., Bell, Daphne W, *et al.* (1999) ‘DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development.’, *Cell*. Elsevier, 99(3), pp. 247–57. doi: 10.1016/S0092-8674(00)81656-6.

Okano, M., Bell, Daphne W., *et al.* (1999) ‘DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development’, *Cell*. Cell Press, 99(3), pp. 247–257. doi: 10.1016/S0092-8674(00)81656-6.

Okano, M., Xie, S. and Li, E. (1998) ‘Cloning and characterization of a family

of novel mammalian DNA (cytosine-5) methyltransferases', *Nature Genetics 1998 19:3*. Nature Publishing Group, 19(3), pp. 219–220. doi: 10.1038/890.

Ooi, S. K. T. *et al.* (2007) 'DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA', *Nature 2007 448:7154*. Nature Publishing Group, 448(7154), pp. 714–717. doi: 10.1038/nature05987.

Ortiz, A. (2000) Role and regulation of apoptotic cell death in the kidney. Y2K update. Frontiers in Bioscience 5, 735–749.

Pačes, V., Doskočil, J. and Šorm, F. (1968) 'Incorporation of 5-Azacytidine into nucleic acids of Escherichia coli', *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis*. Elsevier, 161(2), pp. 352–360. doi: 10.1016/0005-2787(68)90113-5.

Panning, B. and Jaenisch, R. (1996) 'DNA hypomethylation can activate Xist expression and silence X-linked genes', *Genes & development*. Genes Dev, 10(16), pp. 1991–2002. doi: 10.1101/GAD.10.16.1991.

Patra, S. K. (2020) 'Roles of OCT4 in pathways of embryonic development and cancer progression', *Mechanisms of Ageing and Development*. Elsevier, 189, p. 111286. doi: 10.1016/j.mad.2020.111286.

Pauklin, S., Pedersen, R. A. and Vallier, L. (2011) 'Mouse pluripotent stem cells at a glance', *Journal of Cell Science*, 124(22), pp. 3727–3732. doi: 10.1242/jcs.074120.

Peaston, A. E. *et al.* (2004) 'Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos', *Developmental cell*. Dev Cell, 7(4), pp. 597–606. doi: 10.1016/J.DEVCEL.2004.09.004.

Percharde, M. *et al.* (2018) 'A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity', *Cell*. Cell Press, 174(2), pp. 391-405.e19. doi: 10.1016/J.CELL.2018.05.043/ATTACHMENT/623A0267-775D-4D5F-BCF6-F52061E03C13/MMC5.XLSX.

Pradhan, S. *et al.* (1999) 'Recombinant Human DNA (Cytosine-5) Methyltransferase: I. EXPRESSION, PURIFICATION, AND COMPARISON OF DE NOVO AND MAINTENANCE METHYLATION', *Journal of Biological Chemistry*. Elsevier, 274(46), pp. 33002–33010. doi: 10.1074/JBC.274.46.33002.

Prokhortchouk, A. *et al.* (2001) 'The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor', *Genes & Development*. Cold Spring Harbor Laboratory Press, 15(13), p. 1613. doi: 10.1101/GAD.198501.

Ptashne, M. (1992) *A genetic switch : phage λ and higher organisms.* Cell Press.

Ramsahoye, B. H. *et al.* (2000) 'Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a', *Proceedings of the National Academy of Sciences of the United States of America*. The National Academy of Sciences, 97(10), pp. 5237–5242.

Ran, F. A. *et al.* (2013) 'Genome engineering using the CRISPR-Cas9 system', *Nature Protocols 2013 8:11*. Nature Publishing Group, 8(11), pp. 2281–2308. doi: 10.1038/nprot.2013.143.

Reik, W. *et al.* (1987) 'Genomic imprinting determines methylation of parental alleles in transgenic mice.', *Nature*, 328(6127), pp. 248–251. doi: 10.1038/328248A0.

Rhee, I. *et al.* (2000) 'CpG methylation is maintained in human cancer cells lacking DNMT1', *Nature 2000 404:6781*. Nature Publishing Group, 404(6781), pp. 1003–1007. doi: 10.1038/35010000.

Richardson, S. R. *et al.* (2015) 'The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes', *Microbiology Spectrum*. American Society for Microbiology, 3(2).

Richardson, S. R. and Faulkner, G. J. (2018) 'Heritable L1 Retrotransposition

Events During Development: Understanding Their Origins', *BioEssays*. John Wiley and Sons Inc. doi: 10.1002/bies.201700189.

Riedl, J. *et al.* (2008) Lifeact: a versatile marker to visualize F-actin. doi:10.1038/NMETH.1220.

Riggs, A. D. (1975) 'X inactivation, differentiation, and DNA methylation', *Cytogenetic and Genome Research*. Karger Publishers, 14(1), pp. 9–25. doi: 10.1159/000130315.

Robinson, J. T. *et al.* (2011) 'Integrative genomics viewer', *Nature Biotechnology*. Nature Publishing Group, pp. 24–26. doi: 10.1038/nbt.1754.

Rose, N. R. and Klose, R. J. (2014) 'Understanding the relationship between DNA methylation and histone lysine methylation', *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*. Elsevier, 1839(12), pp. 1362–1372. doi: 10.1016/J.BBAGRM.2014.02.007.

Rossant, J. & Tam, P. P. L. (2009) Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. Development vol. 136 701–713.

Sager, R. and Kitchin, R. (1975) 'Selective Silencing of Eukaryotic DNA', *Science*. American Association for the Advancement of Science, 189(4201), pp. 426–433. doi: 10.1126/science.189.4201.426.

Saitou, M. and Yamaji, M. (2012) 'Primordial germ cells in mice', *Cold Spring Harbor Perspectives in Biology*. Cold Spring Harbor Laboratory Press, 4(11), p. a008375. doi: 10.1101/cshperspect.a008375.

Sakaue, M. *et al.* (2010) 'DNA Methylation Is Dispensable for the Growth and Survival of the Extraembryonic Lineages', *Current Biology*. Cell Press, 20(16), pp. 1452–1457. doi: 10.1016/j.cub.2010.06.050.

SanMiguel, J. M. and Bartolomei, M. S. (2018) 'DNA methylation dynamics of genomic imprinting in mouse development', *Biology of Reproduction*, pp.

252–262. doi: 10.1093/biolre/ioy036.

Saxonov, S., Berg, P. and Brutlag, D. L. (2006) ‘A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 103(5), pp. 1412–1417. doi: 10.1073/PNAS.0510310103/SUPPL_FILE/10310TABLE6.XLS.

Schermelleh, L. *et al.* (2007) ‘Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation’, *Nucleic Acids Research*. Oxford Academic, 35(13), pp. 4301–4312. doi: 10.1093/nar/gkm432.

Schindelin, J. *et al.* (2012) ‘Fiji: an open-source platform for biological-image analysis’, *Nature Methods 2012 9:7*. Nature Publishing Group, 9(7), pp. 676–682. doi: 10.1038/nmeth.2019.

Schlesinger, S. and Meshorer, E. (2019) ‘Open Chromatin, Epigenetic Plasticity, and Nuclear Organization in Pluripotency’, *Developmental Cell*. Cell Press, pp. 135–150. doi: 10.1016/j.devcel.2019.01.003.

Schmidt, C. S. *et al.* (2012) ‘Global DNA Hypomethylation Prevents Consolidation of Differentiation Programs and Allows Reversion to the Embryonic Stem Cell State’, *PLoS ONE*, 7(12), p. 52629. doi: 10.1371/journal.pone.0052629.

Schübeler, D. (2015) Function and information content of DNA methylation. Nature 517, 321–326.

Schulz, W. A., Steinhoff, C. and Florl, A. R. (2006) ‘Methylation of endogenous human retroelements in health and disease’, *Current topics in microbiology and immunology*. Curr Top Microbiol Immunol, 310, pp. 211–250. doi: 10.1007/3-540-31181-5_11.

Seisenberger, S. *et al.* (2012) ‘The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells.’, *Molecular Cell*,

48(6), pp. 849–862. doi: 10.1016/J.MOLCEL.2012.11.001.

Sélénou, C. *et al.* (2022) 'IGF2: Development, Genetic and Epigenetic Abnormalities', *Cells*. Cells, 11(12). doi: 10.3390/CELLS11121886.

Sharif, J. *et al.* (2007) 'The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA', *Nature 2007 450:7171*. Nature Publishing Group, 450(7171), pp. 908–912. doi: 10.1038/nature06397.

Shukla, R. *et al.* (2020) 'Activation of transcription factor circuity in 2i-induced ground state pluripotency is independent of repressive global epigenetic landscapes', *Nucleic Acids Research*, (1). doi: 10.1093/nar/gkaa529.

Singer, J., Roberts-Ems, J. and Riggs, A. D. (1979) 'Methylation of Mouse Liver DNA Studied by Means of the Restriction Enzymes Msp I and Hpa II', *Science*. American Association for the Advancement of Science, 203(4384), pp. 1019–1021. doi: 10.1126/SCIENCE.424726.

Sinsheimer, R. L. *et al.* (1954) 'THE ACTION OF PANCREATIC DESOXYRIBONUCLEASE: I. ISOLATION OF MONO- AND DINUCLEOTIDES', *Journal of Biological Chemistry*. Elsevier, 208(1), pp. 445–459. doi: 10.1016/S0021-9258(18)65663-7.

Skvortsova, K., Stirzaker, C. and Taberlay, P. (2019) 'The DNA methylation landscape in cancer', *Essays in Biochemistry*. Portland Press, pp. 797–811. doi: 10.1042/EBC20190037.

Smith, A. G. (1992) 'Mouse embryo stem cells: their identification, propagation and manipulation', *Seminars in cell biology*. Semin Cell Biol, 3(6), pp. 385–399. doi: 10.1016/1043-4682(92)90010-S.

Smith, J. D. and Markham, R. (1952) 'Polynucleotides from Deoxyribonucleic Acids', *Nature 1952 170:4316*. Nature Publishing Group, 170(4316), pp. 120–121. doi: 10.1038/170120a0.

Smith, Z. D. *et al.* (2012) ‘A unique regulatory phase of DNA methylation in the early mammalian embryo’, *Nature 2012 484:7394*. Nature Publishing Group, 484(7394), pp. 339–344. doi: 10.1038/nature10960.

Song, J. *et al.* (2011) ‘Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation’, *Science*. American Association for the Advancement of Science, 331(6020), pp. 1036–1040. doi: 10.1126/SCIENCE.1195380/SUPPL_FILE/SONG-SOM.PDF.

Southern, E. M. (1975) ‘Detection of specific sequences among DNA fragments separated by gel electrophoresis’, *Journal of Molecular Biology*. Academic Press, 98(3), pp. 503–517. doi: 10.1016/S0022-2836(75)80083-0.

Stewart, K. R., Veselovska, L. and Kelsey, G. (2016) ‘Establishment and functions of DNA methylation in the germline.’, *Epigenomics*. Future Medicine Ltd., 8(10), pp. 1399–1413. doi: 10.2217/EPI-2016-0056.

Stocking, C. and Kozak, C. A. (2008) ‘Murine endogenous retroviruses’, *Cellular and molecular life sciences : CMLS*. Cell Mol Life Sci, 65(21), pp. 3383–3398. doi: 10.1007/S00018-008-8497-0.

Stricker, S. H. & Götz, M. (2018) DNA-Methylation: Master or Slave of Neural Fate Decisions? Frontiers in Neuroscience 12, 5.

Stuhlmann, H., Jähner, D. and Jaenisch, R. (1981) ‘Infectivity and methylation of retroviral genomes is correlated with expression in the animal’, *Cell*. Cell Press, 26(2), pp. 221–232. doi: 10.1016/0092-8674(81)90305-6.

Subramanian, A. *et al.* (2005) ‘Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles’, *Proceedings of the National Academy of Sciences of the United States of America*, 102(43), pp. 15545–15550. doi: 10.1073/pnas.0506580102.

Suetake, I. *et al.* (2004) ‘DNMT3L Stimulates the DNA Methylation Activity of Dnmt3a and Dnmt3b through a Direct Interaction’, *Journal of Biological Chemistry*. Elsevier, 279(26), pp. 27816–27823. doi:

10.1074/JBC.M400181200.

Surani, M. A. H., Barton, S. C. and Norris, M. L. (1984) ‘Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis’, *Nature*. Nature, 308(5959), pp. 548–550. doi: 10.1038/308548A0.

Svoboda, P. *et al.* (2004) ‘RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos’, *Developmental Biology*. Academic Press Inc., 269(1), pp. 276–285. doi: 10.1016/j.ydbio.2004.01.028.

Syeda, F. *et al.* (2011) ‘The Replication Focus Targeting Sequence (RFTS) Domain Is a DNA-competitive Inhibitor of Dnmt1’, *Journal of Biological Chemistry*. Elsevier, 286(17), pp. 15344–15351. doi: 10.1074/JBC.M110.209882.

Tada, M. *et al.* (1997) ‘Embryonic germ cells induce epigenetic reprogramming of somatic nucleus in hybrid cells’, *The EMBO Journal*, 16(21), pp. 6510–6520.

Takahashi, K. and Yamanaka, S. (2006) ‘Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors’, *Cell*. Cell Press, 126(4), pp. 663–676. doi: 10.1016/j.cell.2006.07.024.

Takeshita, K. *et al.* (2011) ‘Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1)’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 108(22), pp. 9055–9059. doi: 10.1073/PNAS.1019629108/SUPPL_FILE/PNAS.1019629108_SI.PDF.

Tan, H. and Tee, W. W. (2019) ‘Committing the primordial germ cell: An updated molecular perspective’, *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*. John Wiley & Sons, Ltd, 11(1), p. e1436. doi: 10.1002/wsbm.1436.

Thamban, T., Agarwaal, V. and Khosla, S. (2020) ‘Role of genomic imprinting

in mammalian development.', *Journal of biosciences*, 45(20). doi: 10.1007/s12038-019-9984-1.

Thomson, J. P. and Meehan, R. R. (2017) 'DNA Methylation Changes in Cancer', in *Cancer Drug Discovery and Development*. Humana Press Inc., pp. 75–96. doi: 10.1007/978-3-319-59786-7_4.

Tsumura, A. *et al.* (2006) 'Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b', *Genes to Cells*. Wiley/Blackwell (10.1111), 11(7), pp. 805–814. doi: 10.1111/j.1365-2443.2006.00984.x.

Uribe-Lewis, S. *et al.* (2011) 'Molecular mechanisms of genomic imprinting and clinical implications for cancer', *Expert Reviews in Molecular Medicine*. Cambridge University Press, 13, p. e2. doi: 10.1017/S1462399410001717.

Verkerk, A. J. M. H. *et al.* (1991) 'Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome', *Cell*. Cell, 65(5), pp. 905–914. doi: 10.1016/0092-8674(91)90397-H.

Waalwijk, C. and Flavell, R. A. (1978a) 'DNA methylation at a CCGG sequence in the large intron of the rabbit β-globin gene: tissue-specific variations', *Nucleic Acids Research*. Oxford Academic, 5(12), pp. 4631–4642. doi: 10.1093/NAR/5.12.4631.

Waalwijk, C. and Flavell, R. A. (1978b) 'MspI, an isoschizomer of hpaII which cleaves both unmethylated and methylated hpaii sites', *Nucleic Acids Research*. Oxford Academic, 5(9), pp. 3231–3236. doi: 10.1093/nar/5.9.3231.

Waddington, C. H. (1957) *The strategy of the genes: A discussion of some aspects of theoretical biology*. London: George Allen & Unwin Ltd.

Waddington, C. H. (2012) 'The Epigenotype', *International Journal of Epidemiology*. Oxford Academic, 41(1), pp. 10–13. doi: 10.1093/IJE/DYR184.

Waddington, C. H. (2014) *The strategy of the genes: A discussion of some aspects of theoretical biology*. London: Routledge. doi: 10.4324/9781315765471.

Walsh, C. P., Chaillet, J. R. and Bestor, T. H. (1998) 'Transcription of IAP endogenous retroviruses is constrained by cytosine methylation', *Nature Genetics 1998 20:2*. Nature Publishing Group, 20(2), pp. 116–117. doi: 10.1038/2413.

Wang, J., Huang, J. and Shi, G. (2020) 'Retrotransposons in pluripotent stem cells', *Cell Regeneration*. Springer. doi: 10.1186/s13619-020-00046-4.

Weber, M. *et al.* (2007) 'Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome', *Nature Genetics 2007 39:4*. Nature Publishing Group, 39(4), pp. 457–466. doi: 10.1038/ng1990.

Weinberg, D. N. *et al.* (2019) 'The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape', *Nature*. Nature Publishing Group, 573(7773), pp. 281–286. doi: 10.1038/s41586-019-1534-3.

Weinberger, L., Ayyash, M., Novershtern, N. & Hanna, J. H. (2016) Dynamic stem cell states: Naive to primed pluripotency in rodents and humans. Nature Reviews Molecular Cell Biology vol. 17 155–169.

Wells, J. N. and Feschotte, C. (2020) 'A Field Guide to Eukaryotic Transposable Elements', *Annual Review of Genetics*, pp. 539–561. doi: 10.1146/annurev-genet-040620-022145.

White, J. and Dalton, S. (2005) 'Cell cycle control of embryonic stem cells', *Stem Cell Reviews*, 1(2), pp. 131–138. doi: 10.1385/SCR:1:2:131.

Woodard, L. E. and Wilson, M. H. (2015) 'piggyBac-ing models and new therapeutic strategies', *Trends in Biotechnology*. Elsevier Current Trends, 33(9), pp. 525–533. doi: 10.1016/J.TIBTECH.2015.06.009.

Wu, C. T. and Morris, J. R. (2001) ‘Genes, Genetics, and Epigenetics: A Correspondence’, *Science*. American Association for the Advancement of Science, 293(5532), pp. 1103–1105. doi: 10.1126/SCIENCE.293.5532.1103.

Wu, J. C. and Santi, D. V. (1987) ‘Kinetic and catalytic mechanism of HhaI methyltransferase.’, *Journal of Biological Chemistry*. Elsevier, 262(10), pp. 4778–4786. doi: 10.1016/S0021-9258(18)61263-3.

Wutz, A. *et al.* (1997) ‘Imprinted expression of the Igf2r gene depends on an intronic CpG island’, *Nature*. Nature, 389(6652), pp. 745–749. doi: 10.1038/39631.

Wyatt, G. R. (1950) ‘Occurrence of 5-Methyl-Cytosine in Nucleic Acids’, *Nature 1950 166:4214*. Nature Publishing Group, 166(4214), pp. 237–238. doi: 10.1038/166237b0.

WYATT, G. R. (1951) ‘Recognition and estimation of 5-methylcytosine in nucleic acids’, *Biochemical Journal*. Portland Press, 48(5), pp. 581–584. doi: 10.1042/BJ0480581.

Xie, S. *et al.* (1999) ‘Cloning, expression and chromosome locations of the human DNMT3 gene family’, *Gene*. Elsevier, 236(1), pp. 87–95. doi: 10.1016/S0378-1119(99)00252-8.

Xu, R. *et al.* (2021) ‘Insights into epigenetic patterns in mammalian early embryos’, *Protein and Cell*. Higher Education Press Limited Company, pp. 7–28. doi: 10.1007/s13238-020-00757-z.

Yang, F. *et al.* (2020) ‘DUX-miR-344-ZMYM2-Mediated Activation of MERVL LTRs Induces a Totipotent 2C-like State’, *Cell Stem Cell*. Cell Press, 26(2), pp. 234-250.e7. doi: 10.1016/j.stem.2020.01.004.

Yen, R. whay C. *et al.* (1992) ‘Isolation and characterization of the cDNA encoding human DNA methyltransferase.’, *Nucleic Acids Research*, 20(9), pp. 2287–2291. doi: 10.1093/NAR/20.9.2287.

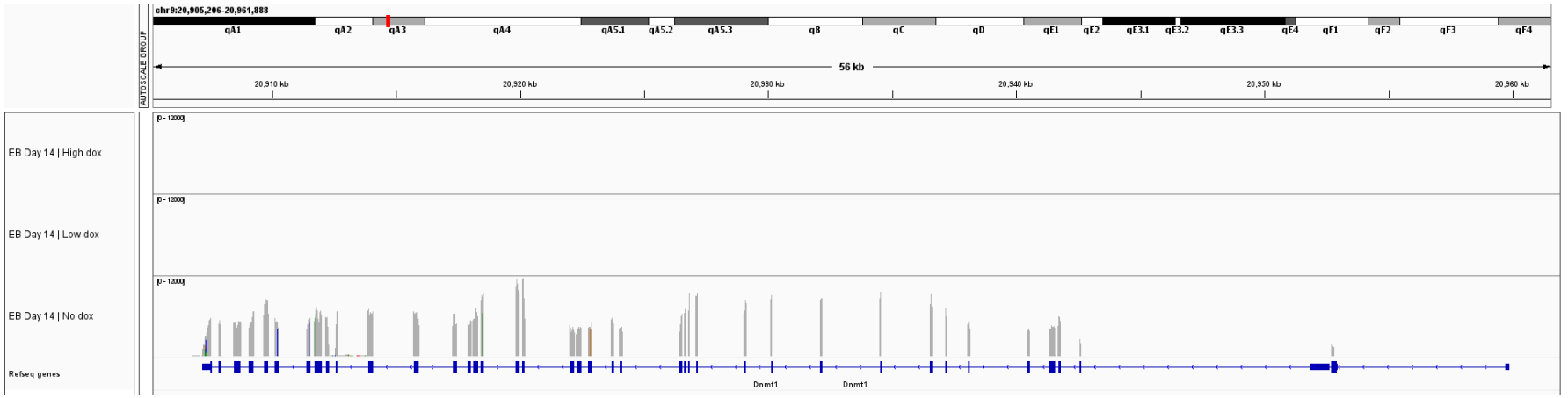
Ying, Q. L. *et al.* (2008) ‘The ground state of embryonic stem cell self-renewal’, *Nature*, 453(7194), pp. 519–523. doi: 10.1038/nature06968.

Zhou, Y. *et al.* (2019) ‘Metascape provides a biologist-oriented resource for the analysis of systems-level datasets’, *Nature Communications*, 10(1), pp. 1–10. doi: 10.1038/s41467-019-09234-6.

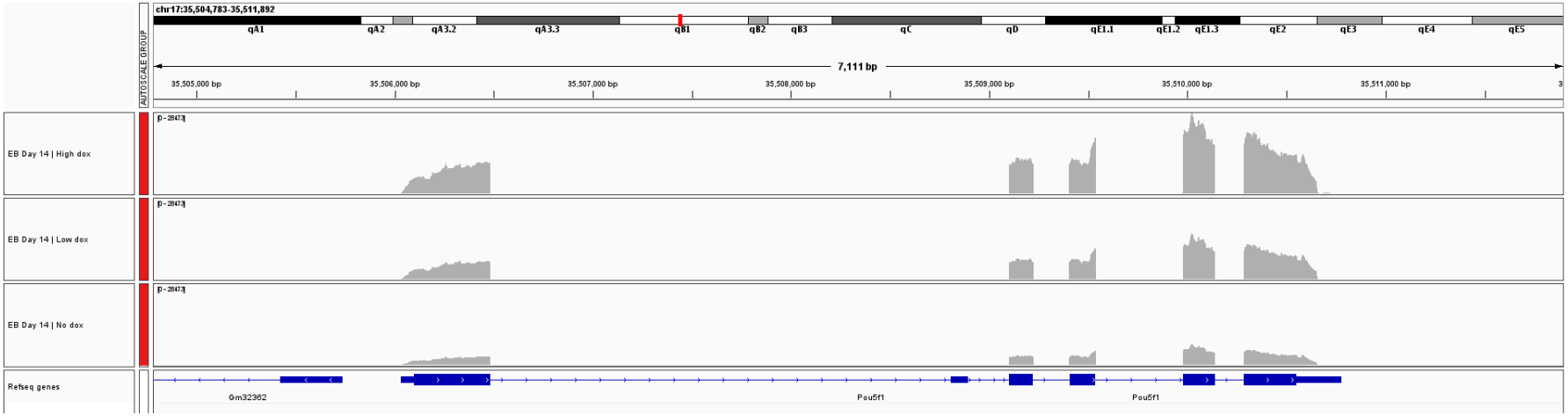
Zoghbi, H. Y. and Beaudet, A. L. (2016) ‘Epigenetics and human disease’, *Cold Spring Harbor Perspectives in Biology*. Cold Spring Harbor Laboratory Press, 8(2), pp. 1–28. doi: 10.1101/cshperspect.a019497.

Zwart, R. *et al.* (2001) ‘Bidirectional action of the Igf2r imprint control element on upstream and downstream imprinted genes.’, *Genes & Development*, 15(18), pp. 2361–2366. doi: 10.1101/GAD.206201.

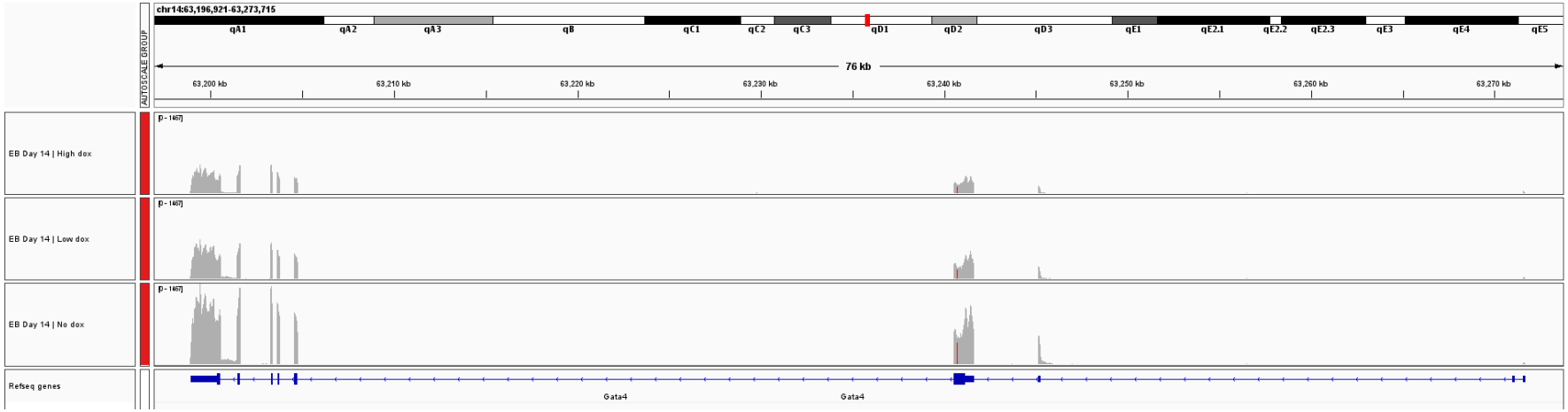
Appendices



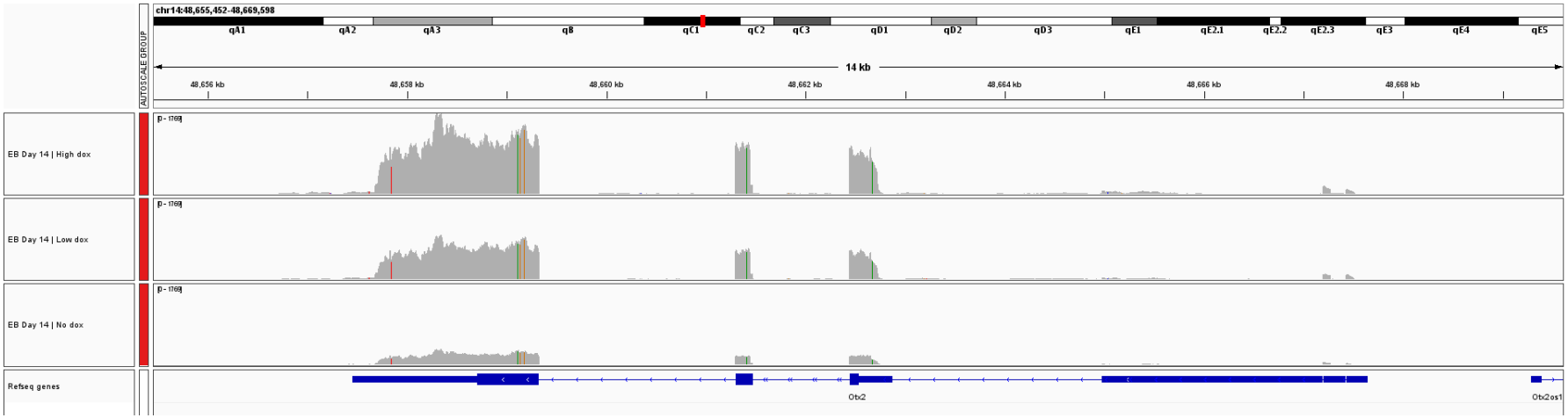
Appendix figure 1. Genome browser view of *Dnmt1* transcripts from *Dnmt1*$^{tet/tet}$ EBs cultured with and without dox at day 14 of differentiation. Scale is 0-12,000.



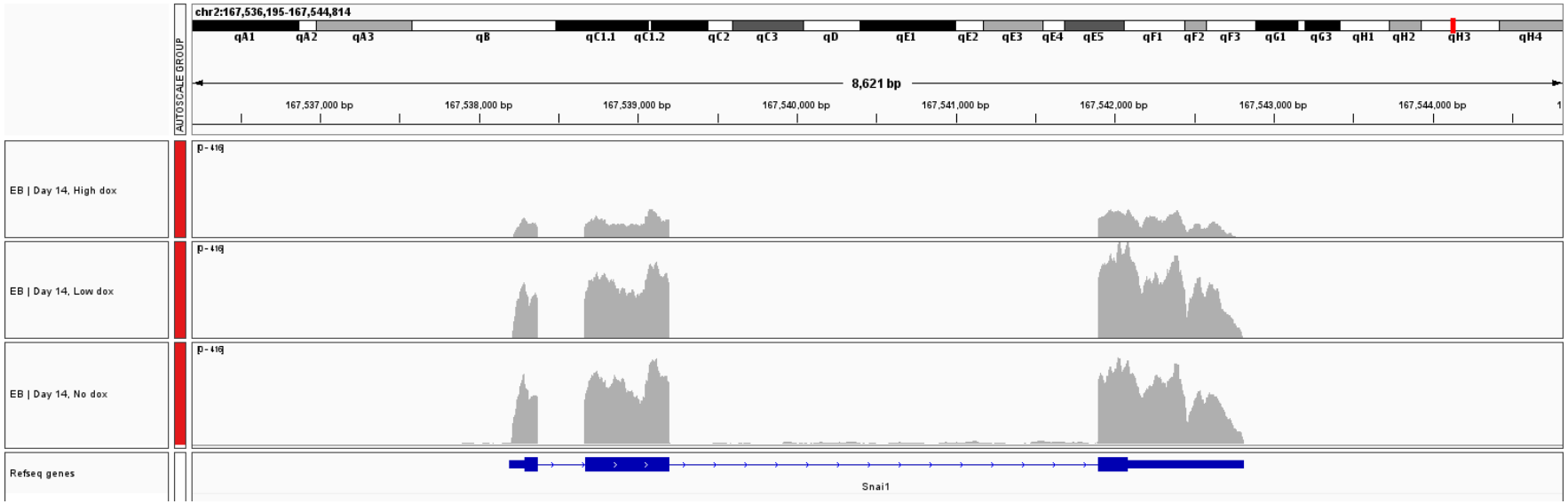
Appendix figure 2. Genome browser view of *Oct4* (*Pou5f1*) transcripts from *Dnmt1*$^{tet/tet}$ EBs cultured with and without dox at day 14 of differentiation. *Oct4* is aberrantly expressed in late-stage hypomethylated (dox treated) EBs when it is normally silenced upon initiation of differentiation in wild-type (No dox) EBs. Scale is 0-28,473.



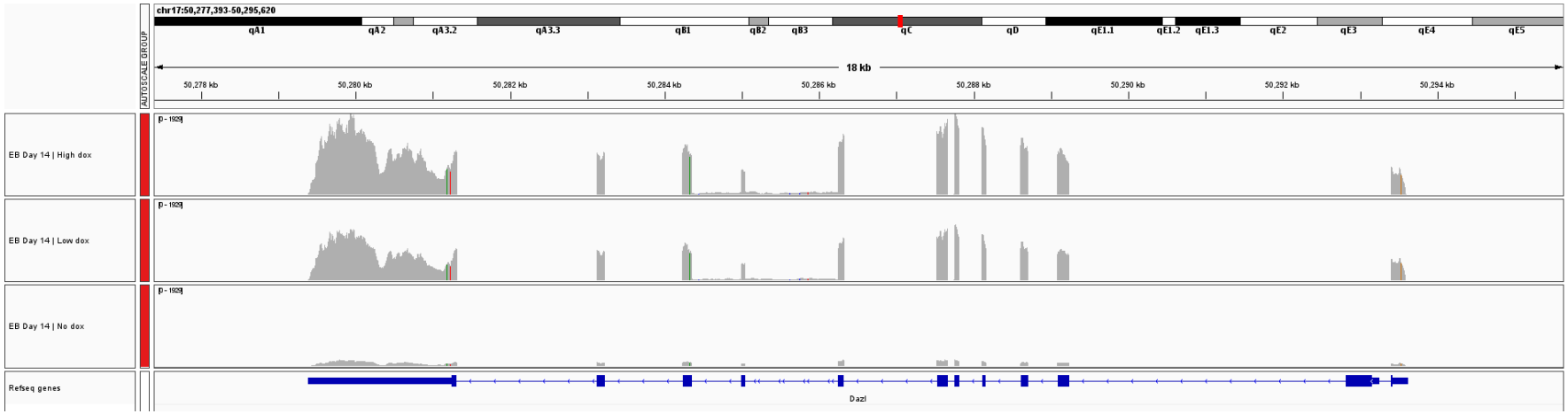
Appendix figure 3. Genome browser view of *Gata4* transcripts from *Dnmt1*$^{tet/tet}$ EBs cultured with and without dox at day 14 of differentiation. *Gata4* is associated endoderm differentiation. Scale is 0-1,467.



Appendix figure 4. Genome browser view of *Otx2* transcripts from *Dnmt1*$^{tet/tet}$ EBs cultured with and without dox at day 14 of differentiation. *Otx2* is associated ectoderm differentiation. Scale is 0-1,765.



Appendix figure 5. Genome browser view of *Snai1* transcripts from *Dnmt1*$^{tet/tet}$ EBs cultured with and without dox at day 14 of differentiation. *Snai1* is associated with mesoderm differentiation. Scale is 0-416.



Appendix figure 6. Genome browser view of *Dazl* transcripts from *Dnmt1*$^{tet/tet}$ EBs cultured with and without dox at day 14 of differentiation. *Dazl* is expressed exclusively in PGCs, and is normally silenced by DNA methylation. Scale is 0-1,925.