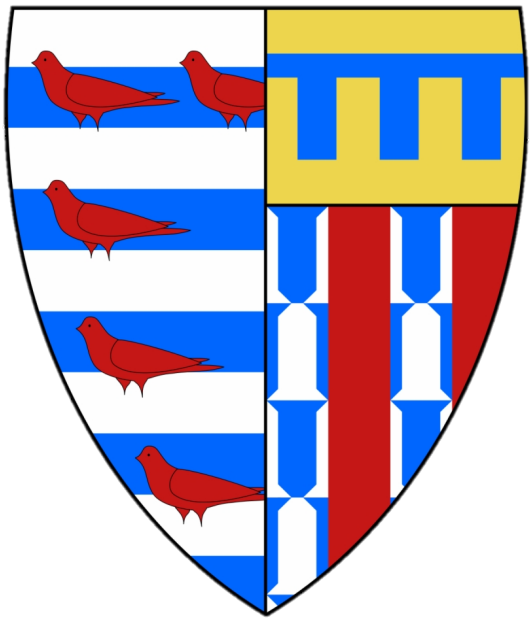
# Role of Histone Lysine Demethylase, KDM1B, in Trophoblast Stem Cell Self-Renewal and Differentiation



Georgia Lea

The Babraham Institute

Pembroke College

University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy

September 2021

# Declaration

This dissertation is the results of my own work and included nothing which is the outcome of work done in collaboration except where specifically indicated in the text. Neither this dissertation, nor the work presented herein, has been submitted for any other qualification at the University of Cambridge or any other institution. The research presented in this dissertation was carried out between October 2017 and September 2021 under the supervision of Prof. Wolf Reik at the Babraham Institute, Cambridge, with Prof Myriam Hemberger of the University of Calgary. This dissertation does not exceed the limit of 60,000 words imposed by the Cambridge University Faculty of Biology Degree Committee.

Georgia Lea

September 2021

# Acknowledgements

This PhD would not have been possible without many very important people.

Firstly, I would like to thank my original supervisor, Myriam Hemberger, without you this project would not have existed and the completion of this PhD and thesis would not have been possible. You have been an instrumental supervisor and mentor throughout the last four years and I cannot thank you enough for taking me on and introducing me to this fascinating area of biology. Secondly, I have to thank Vicente Perez-Garcia. Vicen, you were a rock for me through a lot of hard changes and I cannot express how grateful I am for you always being there when I needed to vent or had a question, or for an extra pair of hands on the trip down to stores. Courtney, you were so kind and generous with your time. Thank you for being a wonderful desk-buddy. Natasha, Claire and Laura, I am so lucky to have met you all and I'm so grateful for our friendship.

Wolf, I am so grateful to you for adopting me at the end of my first year. Moving labs was a big change and I thank you for being there when I needed and for your patience and guidance. Aled, thank you for being my mentor and for providing insights and help over the last three years. To everyone in the Reik Lab (you know it's too risky to try to name everybody), thank you for being so welcoming and helpful.

Jules, Dil, Izzy, Maggs and Caz, you have made my time at BI worth it. I know we're going to be friends for a long time to come and I cannot wait to see what the next few years bring with us all moving on. Caz you are and will always be my work wife and it's going to be so weird not seeing you every day.

Now, on to family. Karl, Sam, Frezgi, Noah, Regan, Steph, Dean and Mon. You are my chosen family and I would not have made it this far in life, let alone through the last four years, without you all. I'm so lucky to have met you and kept you. I love you so much.

Mumma and Dad, you have been so patient and supportive. Thank you so much for being mine. Thank you for not prying and for waiting for me to come to you with news and issues. And thank you for always being there. Without question. Even when I fuck up. Hats and Bibs, it has been an honour to be your big sister. I love you both so much. Hats, thank you for being such a loving, calm and fun influence in my life and Bibs thank you for being so compassionate and nurturing. You're all wonderful and I'm so lucky to be a Lea.

Finally, thanks to me, without whom this, literally, would not have been possible.

Vaccines are real. Trans rights are human rights. Love is Love. Black lives matter.

# Assistance to be acknowledged

## Initial training in techniques, laboratory practice and subsequent mentoring:

Myriam Hemberger (original supervisor), Vicente Perez-Garcia (lab member), Courtney Hanna (lab member), Wolf Reik (supervisor), Aled Parry (lab member)

## Data obtained from a technical service provider:

Next-generation sequencing of mRNA and ChIP libraries was carried out at The Babraham Institute Next Generation Sequencing Facility by Kristina Tabbada, Nicole Forrester and Paula Kokko-Gonzales.

Next-generation sequencing data was processed and aligned by Felix Kreuger and Simon Andrews in the Bioinformatics group at The Babraham Institute. Initial inspection of H3K36me3 ChIP data was performed by Simon Andrews.

Flow cytometry analysis was performed by Atilla Bebes and Christopher Hall of The Babraham Insititute Flow Cytometry Facility.

# Summary

The first cell fate decision in development occurs at the blastocyst stage with the emergence of the trophectoderm (TE) and the inner cell mass (ICM). The TE is the precursor population of all major placental cell types. Reflecting this developmental plasticity, trophoblast stem cells (TSCs) can be derived from the TE of mouse blastocysts. TSCs have proven an invaluable research tool to study processes of early placentation in vitro. Despite the placenta's central role in reproduction, our understanding of the regulatory networks that orchestrate TSC self-renewal and differentiation remains incomplete.

In this project, I characterised an epigenetic modifier, KDM1B, for its role in TSC self-renewal and differentiation. I identified this factor as a putative novel regulator of trophoblast stem cell fate and *in vitro* differentiation from transcriptomics data as its expression is markedly induced at the onset of differentiation. Furthermore, *Kdm1b* had been implicated in mouse development and placentation, via directing DNA methylation of maternal imprints in the oocyte. KDM1B is a histone lysine demethylase whose activity is directed to H3K4me1 and H3K4me2, particularly within the gene body of actively transcribed genes. By generating CRISPR-Cas9-mediated knockout TSCs ablated for *Kdm1b*, I show that *Esrrb* is consistently down-regulated but *Gcm1* is up-regulated in *Kdm1b*$^{-/-}$ TSC clones as measured by RT-qPCR, indicative of precocious differentiation into the syncytiotrophoblast lineage. By performing a large cohort of integrated genome-wide analyses, notably RNA-seq and chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) for the relevant histone modifications H3K4me1, H3K4me2, H3K4me3 and H3K36me3, I show that KDM1B regulates intragenic H3K4me1-marked enhancers, while not impacting H3K4me3. Remarkably, KDM1B null trophoblast cells also show an increased instability of chromosome 13, the same chromosome on which *Kdm1b* itself is located. This chromosome hosts several gene families that arose from gene duplication events, with vital roles in trophoblast development. The instability of chromosome 13 that gave rise to these gene families is apparently exacerbated by loss of KDM1B, or by CRISPR Cas9-induced cutting of the chromosome.

In addition to interrogating the role of KDM1B during differentiation, I utilised the transcriptomic and ChIP-seq data to identify several trophoblast-specific transcripts via location of distal H3K4me3 peaks. Using previously published ChIP-seq data in combination with my own, I link enhancer activity in stem cells to gene expression throughout differentiation. Finally, I identify a striking and novel redistribution of H3K36me3 in 5 day differentiated trophoblast cells, to the promoter regions of expressed genes.

In conclusion, this work presents an in-depth analysis of the transcriptional and epigenomic rearrangements that occur both in WT trophoblast and as a consequence of *Kdm1b* deletion. These

data lend important insights into the functions of this epigenetic modifier in the fine-tuning of the transcriptional networks that direct TSC self-renewal and differentiation.

# Abbreviations

| | |
|---|---|
| CGI | CpG island |
| ChIP | Chromatin immunopreciptation |
| ChIP-seq | Chromatin immunopreciptation followed by next generation sequencing |
| (1)D | (1) Day of differentiation |
| DE | differentially expressed |
| DNA | deoxyribonucleic acid |
| EPI | Epiblast |
| ESC | Embryonic stem cell |
| FISH | Fluorescence in situ hybridisation |
| Gly-T | Trophoblast glycogen cell |
| GO | Gene ontology |
| H3K36me3 | Histone H3 lysine 36 trimethylation |
| H3K4me1 | Histone H3 lysine 4 monomethylation |
| H3K4me2 | Histone H3 lysine 4 dimethylation |
| H3K4me3 | Histone H3 lysine 4 trimethylation |
| ICM | Inner cell mass |
| *Kdm1b* | Lysine demethylase 1b, gene name |
| KDM1B | Lysine demethylase 1b, protein |
| *Lsd1* | Lysine demethylase 1a, gene name |
| LSD1 | Lysine demethylase 1a, protein |
| mRNA | Messenger ribonucleic acid |
| NGS | Next generation sequencing |
| PCA | Principal component analysis |
| PCR | polymerase chain reaction |
| PE | Primitive endoderm |
| RNA-seq | Next generation sequencing of mRNA |
| RT-qPCR | Quantitative reverse transcription polymerase chain reaction |
| SynT | Syncytiotrophpblast |
| TE | Trophectoderm |
| TF | Transcription factor |
| TGC | Trophoblast Giant Cell |
| TSC | Trophoblast Stem Cell |
| WB | Western blot |

# Contents

# 1 Chapter One

## Introduction

## 1.1 Mammalian preimplantation development

### 1.1.1 Epigenome orchestrates cell fate decisions

Epigenetics has been defined as the study of the heritable mechanisms adopted by eukaryotic cells to regulate gene expression and thereby define distinct cellular lineages. The epigenome mediates the control of gene expression that enables a single nucleus to give rise to the more than 200 different somatic cell types of an adult organism. A carefully balanced combination of external signals, transcription factors and epigenetic modifications enables cell fate 'decisions' to take place. The earliest of such cell fate decisions occurs during preimplantation development of the mammalian embryo, and has been particularly well studied in the mouse leading to the commitment to either the embryonic or extraembryonic lineages (Ilgren, 1983; Dyce *et al.*, 1987). These distinct lineages first emerge at the blastocyst stage (Figure 1.1a): the trophectoderm (TE), primitive endoderm (PE) and epiblast (EPI). The TE and PE give rise to the extraembryonic tissues of the placenta and yolk sac, respectively, whilst the cells of the EPI are pluripotent and give rise to all cells of the embryo proper and the foetal vasculature of the placenta.

These cell types express transcription factors (TFs) that activate distinct programmes of gene expression. In both human and mouse blastocysts, the transcription factors *Pou5f1* (*Oct4*), *Nanog* and *Sox2* are markers of EPI-derived cells; conversely, TE-derived cells are enriched in *Cdx2* and *Gata3* and cells of the PE are enriched for *Sox17* and *Gata6* (Blakeley et al. 2015; Schrode et al. 2013; Figure 1.1b).

Transcription factors are proteins that bind to specific sequences of DNA called motifs, to which they recruit the cells' transcription machinery. TF binding motifs are often located in promoter/enhancer regions, upstream of the transcription start site (TSS), of a protein-coding gene. They therefore regulate cell type-specific gene expression; the pattern of TFs expressed by a cell predicts and controls its lineage and differentiative plasticity, thereby governing any cell fate decisions that occur during differentiation. Transcription factor expression, however, does not account for the heritable nature of cell fate decisions. This information is stored in the epigenome.

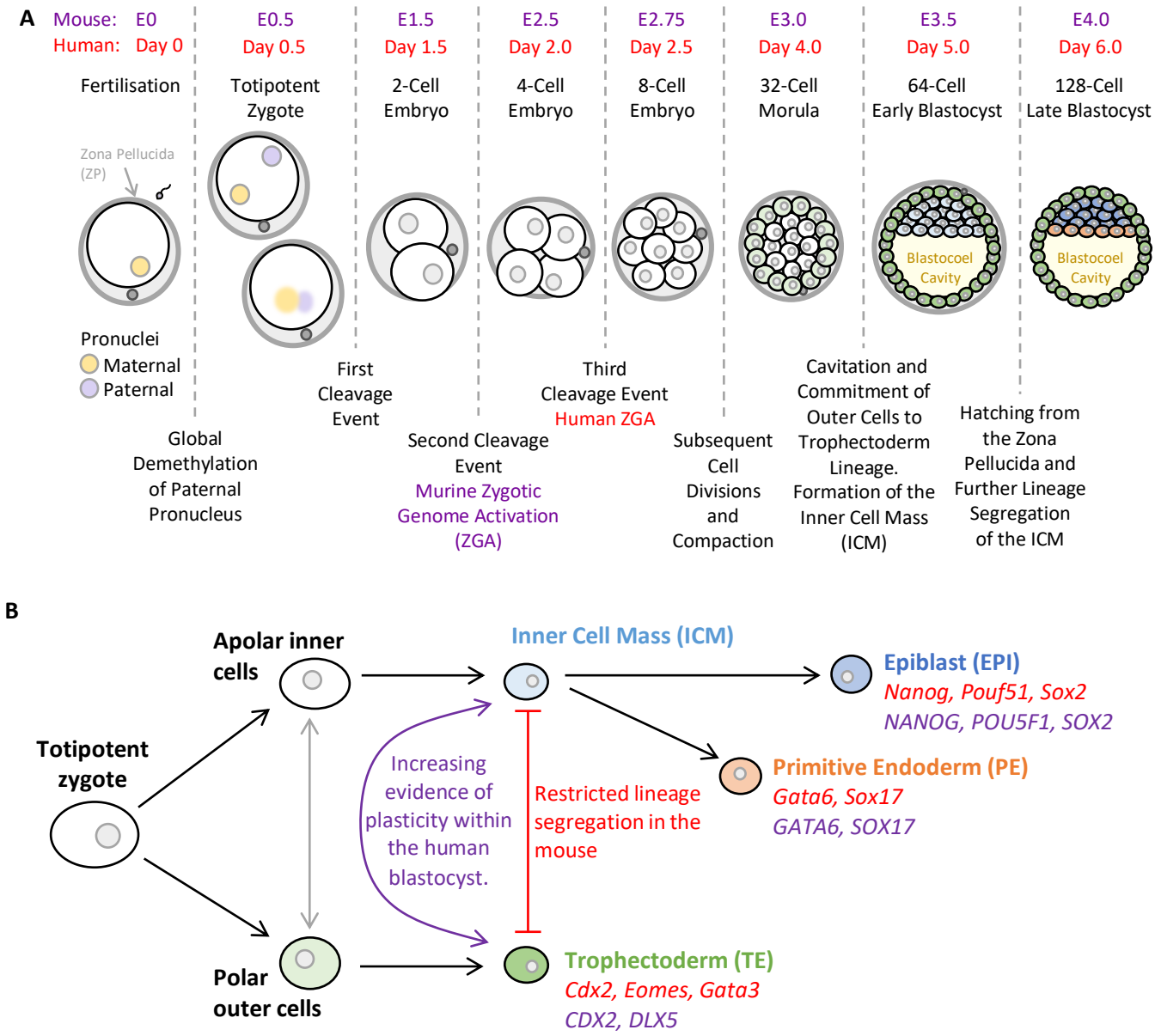



**Figure 1.1 Preimplantation development of human and mouse embryos**

**A**. Schematic detailing certain aspects of preimplantation development of human (purple) and mouse (red) embryos. **B**. More detailed description of cell types as they arise during development from zygote to blastocyst. Black arrows indicate naturally-occurring route of cellular differentiation *in vivo* with grey arrows indicating potential for cell-fate plasticity at this stage of differentiation. Purple arrow indicates potential plasticity as evidenced by *in vitro* transdifferentiation experiments from human stem cell populations comparative to the distinct segregation of these lineages in the mouse as denoted by the red barred arrow.

### 1.1.2 Epigenetic modifications

The epigenome is the interpretative layer of information that is superimposed onto the genetic information residing in the DNA, regulating chromatin structure and thereby, in conjunction with TFs, gene expression. Within the nucleus, DNA is packaged into chromatin (Figure 1.2). This consists of many layers of packaging, initiating with the genomic DNA being wrapped around histones to produce protein:DNA complexes called nucleosomes. Nucleosomes are the smallest unit of chromatin, each comprising around 147 bp of DNA wrapped approximately 1.65 times around the histone octamer. Each nucleosome is separated from the next by a 33bp stretch of DNA bound by the linker histone H1 (Luger *et al.*, 1997).

The histone octamer consists of a heterotetramer of two molecules each of histones H3 and H4 and a pair of heterodimers of H2A:H2B. Histones are rich in the positively charged amino acids, lysine (K) and arginine (R), allowing the formation of strong interactions with the negatively charged phosphate moieties of DNA. The core histones contain globular carboxyl (C)-terminal domains that form most of the protein:protein and protein:DNA interactions of the assembled nucleosome, as well as soluble amino (N)-terminal tails. The histone tails can be modified by many covalent post-translational modifications, such as methylation, acetylation, phosphorylation ADP-ribosylation and ubiquitination (Turner, 2002). Such modifications alter the higher-order chromatin structure by changing the strength of protein:protein interactions between neighbouring nucleosomes, condensing or relaxing the higher-order chromatin architecture. The more tightly condensed the chromatin, the less accessible the underlying DNA sequence and vice versa (Figure 1.2). DNA accessibility is one of the key obstacles to efficient transcription. Tightly packaged, inaccessible chromatin is referred to as heterochromatin while the more accessible, active chromatin, as euchromatin. Transcription requires open, euchromatin along with binding of trans-activating transcription factors, the assembly of the pre-initiating complex upstream of the TSS and often, activating signals from cis-regulatory elements, such as enhancers (Wray *et al.*, 2003).

The combinations of these histone modifications are referred to as the histone code. This code is used by epigeneticists to identify and define regulatory elements of the genome. For example, trimethylation of lysine 4 of histone H3 (H3K4me3) is enriched at active promoters (Noma, Allis and Grewal, 2001; Bernstein *et al.*, 2002; Noma and Grewal, 2002); conversely, trimethylation of lysine 27 of histone H3 (H3K27me3) confers a repressive chromatin structure and is therefore associated with transcriptional repression. Somewhat paradoxically, actively transcribed gene bodies are marked with repressive H3K36me3 and H3K9me2, to prevent spurious transcription (Carrozza *et al.*, 2005; Vakoc *et al.*, 2005). Other activating histone modifications, H3K27ac and H3K4me1 are found in concert at classical active enhancers (Heintzman *et al.*, 2009; Creyghton *et al.*, 2010). These modifications can be

read, written and erased by histone modifiers. Certain histone modifications are mutually exclusive, such as H3K27me3 and H3K27ac (Tie *et al.*, 2009). Further, so-called 'activating' histone modifications such as acetylation are not usually found in the same domains as repressive histone marks such as trimethylation of lysine 9 of histone H3 (H3K9me3) or H3K27me3. These modifications also confer opposing effects on chromatin organisation often by marking unique genomic features. This is with the exception of bivalently marked domains hosting both H3K4me3 and H3K27me3 that are characteristic in particular of pluripotent cells (Pan *et al.*, 2007).

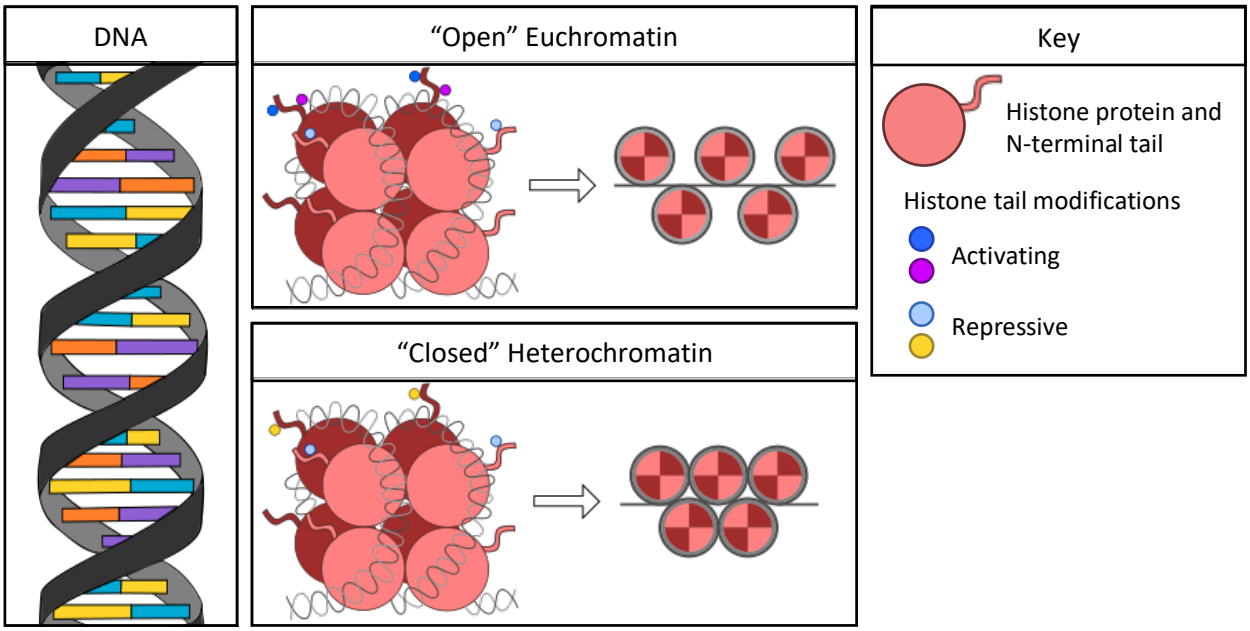



**Figure 1.2 Chromatin organization**

Schematic representation of the DNA helix wrapped around the histone octamer to create a nucleosome. Different modifications of the histone tails result in "open" euchromatin or "closed" heterochromatin.

DNA replication constitutes a challenge to the propagation of nucleosomes, and with that the propagation of their epigenetic information. In general, nucleosomes are dismantled during DNA replication, and reformed behind the replication fork. The process of histone protein recycling after replication fork and the segregation of the H3:H4 tetramer and H2A:H2B dimers to the leading/lagging strands is highly complex and tightly regulated (Zhang, Feng and Li, 2020).

Another type of epigenetic modification is the covalent modification of DNA itself. Methylation of cytosines in CpG dinucleotides at position 5 (5mC) of the pyrimidine ring, is catalysed by the DNA methyltransferases (DNMTs). *De novo* DNA methylation is catalysed by DNMT3A, DNMT3B or, in mice, DNMT3C with the cofactor, DNMT3L, and is maintained following replication by DNMT1 (Yoder *et al.*,

1997; Okano *et al.*, 1999; Bourc'his *et al.*, 2001; Barau *et al.*, 2016). Of the *de novo* methylation machinery, the oocyte dominantly expresses DNMT3A, whereas, in mice, DNMT3C is solely expressed in developing spermatogonia (Barau *et al.*, 2016). Active DNA demethylation occurs by sequential oxidation of the 5-methyl group by the TET DNA dioxygenase enzymes, TET1 and TET3, as summarised by Seisenberger *et al.*, (2012). This results in the iterative formation of 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine (5hmC, 5fC and 5caC, respectively). The modified cytosine is a target for excision and the creation of an abasic site by thymine DNA glycosylase (TDG). The base excision repair (BER) machinery is then recruited to this site and an unmodified cytosine is inserted (Reviewed by Wu and Zhang 2014). DNA methylation can also be lost passively when not actively transferred to the daughter strand by DNMT1 in replicating cells.

DNA methylation of CpG dinucleotides is a vital method of transcriptional repression, genomic imprinting and X-inactivation. The process of stable X-inactivation in female somatic tissues, is dependent on maintenance of DNA methylation (Sado *et al.*, 2000), in addition to the continued expression of *Xist* mRNA (Lee *et al.*, 1996; Penny *et al.*, 1996; Herzing *et al.*, 1997; Marahrens *et al.*, 1997). Certain developmental genes, called "imprinted genes", are heritably methylated in either maternal or paternal gametes, to ensure mono-allelic expression (Reik and Walter, 2001). CpG-rich promoters, referred to as CpG island (CGI) promoters, are refractory to DNA methylation, and can become aberrantly methylated in cancers (Sproul and Meehan, 2013). Promoter methylation is a key mode of transcriptional repression, and is often associated with developmental genes. An example of such, are the *HOX* cluster genes, that are not strictly CGI promoters but share a strong affinity for the CXXC binding motif (Illingworth et al. 2008; Illingworth et al. 2010), which preferentially binds stretches of unmethylated CpG, such as CGIs (Kui *et al.*, 2000; F., Ittai and P., 2004; Jin *et al.*, 2014).

### 1.1.3 Epigenetic remodelling during preimplantation development and lineage commitment

Throughout preimplantation development, dramatic changes to the epigenetic landscape occur. Upon fertilization of the oocyte by a spermatozoon, two pronuclei are formed. The paternal pronucleus is hypermethylated compared to the maternal pronucleus and undergoes a wave of global demethylation prior to the formation (Gu *et al.*, 2011) of the totipotent zygote (Figure 1.1). Subsequent cleavage divisions occur at intervals of 12-24 hours, producing the 2-, 4- and 8-cell stage embryos. Various developmental milestones occur in these early stages such as zygotic genome activation (ZGA). In mouse embryos, ZGA occurs at the two-cell stage (Flach *et al.*, 1982), whereas this occurs later in human pre-implantation embryos, around the 4-cell stage (Braude, Bolton and Moore, 1988). Cleavage divisions continue until the embryo reaches the 16-cell stage when it forms the so-

called morula. In the mouse compaction takes place at the 8-cell stage when the embryo becomes polarised as an obligate forerunner to lineage establishment.

Within the compacted morula, two cell populations arise: polar outer cells surrounding apolar inner cells. Further cell divisions and cavitation give rise to the blastocyst. It is at this stage that the divergent populations, the TE and inner cell mass (ICM), arise. Further cell divisions and transition from the early to late blastocyst sees the commitment of these cells to their respective lineages, with the further segregation of the inner cell mass to the PE and EPI. The late blastocyst stage is reached at E4.0-4.5 in mouse and at 6 days post-fertilisation in humans (Niakan *et al.*, 2012). These populations are spatially segregated in the blastocyst with the TE forming the outer layer surrounding an inner fluid-filled cavity called the blastocoel. Within the blastocyst, at one pole, are the cells of the EPI; these are adjacent to a layer of cells which are committed to the PE cell fate (Figure 1.1). The ICM and EPI are the populations from which embryonic stem cells (ESCs) are derived (Evans and Kaufman, 1981); murine trophoblast stem cells (TSCs) can be derived from the TE of the blastocyst or also from the extraembryonic ectoderm at later stages of trophoblast development (Tanaka *et al.*, 1998; Uy, Downs and Gardner, 2002) . As such, stem cell potential is retained much longer in the extra-embryonic compartment compared to the embryonic lineage. The significance of the early-to-late blastocyst transition is that at this stage, cells become committed to their respective fate, i.e. EPI cells cannot contribute to the TE any longer, and vice versa. These first, incisive cell fate decisions are accompanied and underpinned by global epigenetic differences between the lineages.

In mice, the DNA of blastocyst-stage ICM cells is more highly methylated than TE cells (Santos *et al.*, 2002). Additionally, distribution of H3K27me3 across the genome appears to be one of the first global differences between cells of the TE and ICM lineages (Dahl *et al.*, 2010). It is these differences that create a barrier to spontaneous switching from one lineage to another and maintenance of cell fate commitment, in cells (including stem cells) that are descendent from the late blastocyst.

Further, cells of the epiblast contain so-called bivalent domains, where both repressive (H3K27me3) and activating (H3K4me3) histone marks are found (Bernstein *et al.*, 2006). In the pluripotent cells of the EPI, these domains are primed for transcription, often being bound by RNA polymerase II (RNAPII). In the trophoblast lineage, rather than being primed for expression, these regions are silenced and bound by H3K9me3 methyltransferase, SUV39H1. Therefore, rather than bivalent domains, the TE contains trivalent domains harbouring H3K27me3, H3K9me3 and H3K4me3 marks (Alder *et al.*, 2010).

The cell lineage-specific epigenomes are generally believed to “lock in” the early cell fate decisions made by differential transcription factor expression. As mentioned above, EPI and ESCs are characterized, for example, by OCT4, NANOG, SOX2 and other pluripotency TFs that establish a self-

reinforcing gene network; conversely TE and TSCs are characterized by expression of CDX2, EOMES, GATA3, ELF5 and other TFs in the mouse. There are also some differences between species. The human orthologues of murine TE-specific transcription factors, *Elf5* and *Eomes*, *ELF5* and *EOMES* are absent in human trophectoderm cells of late-stage blastocysts analysed by single-cell RNA-seq (Blakeley *et al.*, 2015), although, ELF5 is detected in first trimester CTB and EVT cells (Hemberger *et al.*, 2010; Soncin *et al.*, 2018). Another TE-specific factor in mice, *Id2*, is also differentially expressed in human blastocysts, with *ID2* expression being most enriched in human PE cells.

In early human development, there is increasing evidence that the naïve cells of the preimplantation epiblast are more plastic than their murine-derived counterparts, with several recent studies (Cinkornpumin *et al.*, 2020; Dong *et al.*, 2020; Guo *et al.*, 2021; Io *et al.*, 2021) demonstrating an increased potential of naïve hPSCs to be reprogrammed to trophoblast(-like) stem cells. By the time these epiblast cells have undergone capacitation and become primed for differentiation, they appear to lose this plasticity. The biological need for a delay in firm lineage commitment during human preimplantation development might be explained by shuttling of damaged, for example polyploid epiblast cells, to the trophectoderm, where genome instability is more tolerated (Kalousek and Dill, 1983; Kalousek, 1994).

## 1.2 The placenta

Species of both primates and rodents form an invasive, discoid placenta. Thus, trophoblast cells penetrate the uterine epithelium and the underlying endometrial stroma to target maternal arteries where they displace the endothelial cell lining to create trophoblast-lined canals carrying maternal blood into the placenta. This process is termed spiral artery remodelling. This type of placentation is called haemochorial (Figure 1.3, Krishnan et al. 2013). Other types of placentation are less invasive, for example the epitheliochorial placenta produced by horses, which leave all layers of the maternal uterus intact with the chorion only coming into contact with the endometrial epithelium (Samuel, Allen and Steven, 1974).

Placentas perform the roles of the respiratory, renal, hepatic, cardiovascular, digestive and immune systems throughout much of pregnancy. Figure 1.3 illustrates the structures of both human and murine term placentas. Briefly, the murine placenta is formed of three layers from the foetal side, the labyrinth, the junctional zone (consisting of spongiotrophoblast and a layer of secondary trophoblast giant cells (TGCs)) and finally the maternally derived decidua basalis. In human placentas the approximately equivalent structures are called foetal placenta, the basal plate and the placental bed. These structures contain similar but distinct trophoblast populations in mice and humans. The murine

labyrinth and human foetal placenta is the site of nutrient exchange between the maternal and foetal blood (Reviewed Georgiades et al. 2002).

Placentas are short-lived organs that must function up until the end of their lifespan, i.e. upon successful delivery of live, healthy offspring. Placentation defects resulting in sub-optimal placental development and subsequent function are highly correlated with foetal growth defects including intrauterine growth restriction (IUGR) leading to low birth weight (Lang *et al.*, 2003; Zohdi *et al.*, 2012). Further, it has recently been illustrated by Perez-Garcia *et al*. (2018) that placental defects can be detected in 68 % of pregnancies arising from genetic mouse mutants which are embryonic lethal. These placental dysmorphologies are closely associated with embryonic abnormalities and are likely causative of many embryonic failures.

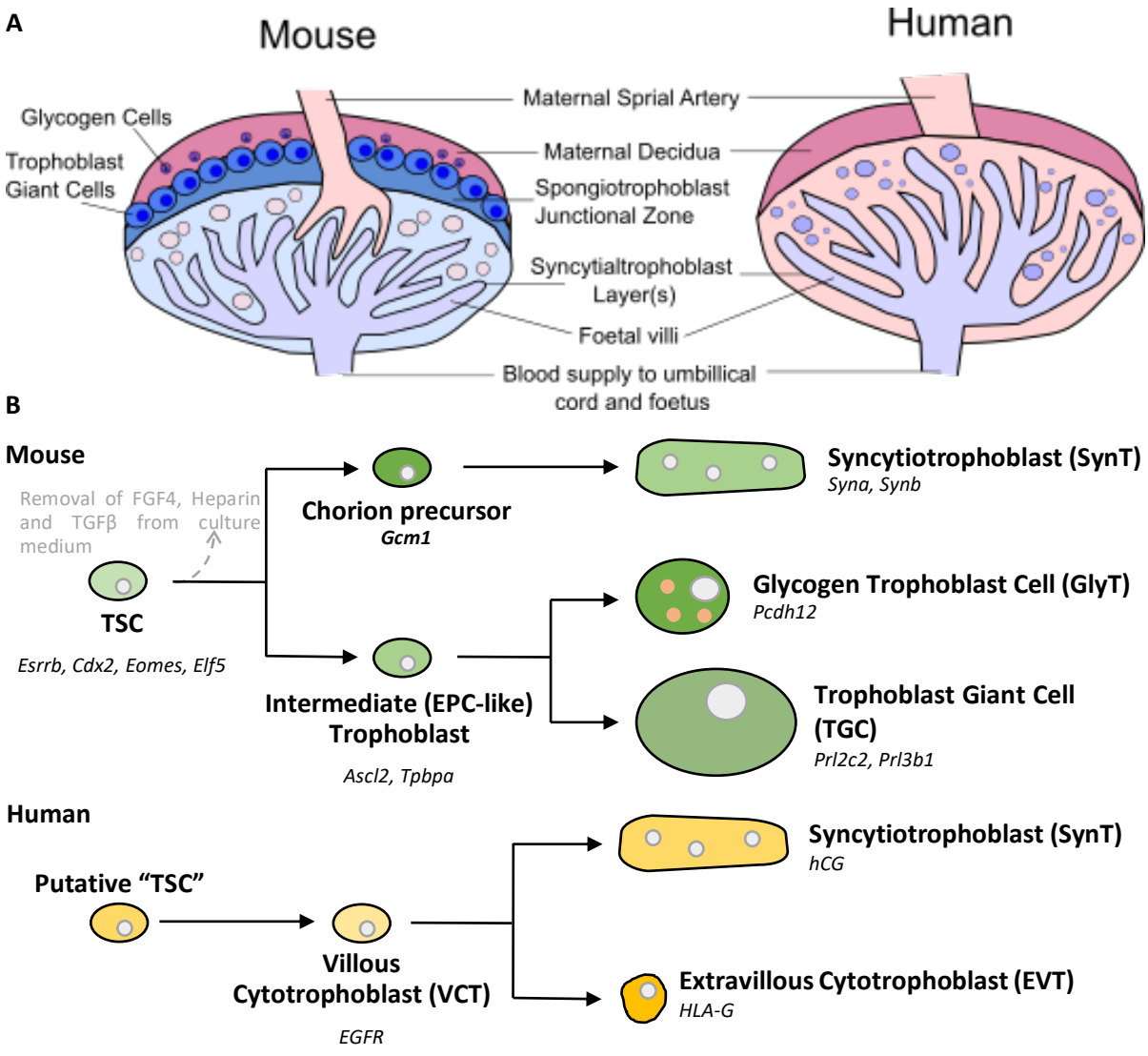


**Figure 1.3 Comparison of human and mouse placentas and trophoblast subtypes**

**A**. Gross structures of the human and mouse placenta **B**. Trophoblast stem cells and the main differentiation trajectories they undertake during *in vitro* culture.

A healthy placenta can have lifelong impact on the offspring it supported during prenatal development (Barker, 2004; Barker *et al.*, 2012). Many studies have shown that birth weight and intrauterine growth restriction (IUGR) are key risk factors for cardiovascular diseases at middle age; birth weight is directly linked with placental development and function (Hack, Klein and Taylor, 1995). The relative invasive capacity of the placenta also directly causes a spectrum of pregnancy complications, with under-invasive placentation often attributed to pre-eclampsia, which if left untreated, can lead to fatal eclampsia for the mother. Conversely, an overly invasive placental phenotype can cause cancer such as choriocarcinoma. In other circumstances, the trophoblast can invade too deeply into the uterus, into the uterine muscle or even beyond the uterus to surrounding tissue, called placenta accrete, placenta increta or placenta percreta, respectively. Compared to the extensive knowledge base around embryonic development, little is known about placental development, despite the fundamental requirement of the organ for successful reproduction.

### 1.2.1 Placental structure and cell types

In the mouse placenta the labyrinth is the site of foetal-maternal exchange of hydrophilic molecules. This nutrient exchange occurs across several cell types from maternal to foetal circulation, sinusoidal TGCs (S-TGCs), two layers of syncytiotrophoblast (SynT) cells and foetal endothelium. The syncytiotrophoblast is a single multinucleated cell layer formed by fusion, or syncytialisation, of precursor cytotrophoblast cells. The mouse syncytium consists of two tightly apposed layers, termed SynT-I and SynT-II, which face the maternal and foetal blood supplies, respectively. During early gestation, it is the S-TGCs that are in direct contact with the maternal blood spaces of the labyrinth, and only after E14.5 do perforations in the S-TGC layer appear allowing direct contact of SynT-I with the maternal blood (Coan, Ferguson-Smith and Burton, 2005).

Conversely, in the human placenta nutrient exchange occurs across a single syncytiotriophoblast layer, which is in direct contact with the maternal blood supply in the intervillous space. The SynT layer is formed by syncytialisation of the underlying cytotrophoblast population, it's precursor population (Richart, 1961; Kliman *et al.*, 1986). The human SynT does not express any HLA class I proteins, and hence appears 'invisible' to the maternal immune system (Juch *et al.*, 2012).

There are four subtypes of TGC in the mature murine placenta, which are of differing origin and inhabit distinct niches but are all polyploid (Simmons, Fortier and Cross, 2007). Parietal TGCs (P-TGCs) are formed, in part, from the mural trophectoderm (Rossant, Gardner and Alexandre, 1978), line the implantation site at E8.5 (Scott *et al.*, 2000), boarder the decidua and are a highly invasive subtype of TGC, with the highest level of polyploidy of all TGCs (Simmons, Fortier and Cross, 2007; Hannibal *et al.*, 2014). S-TGCs arise from the chorion, uniquely expressing *Ctsq* (Simmons, Fortier and Cross, 2007; Simmons *et al.*, 2008). The spiral artery-associated TGCs (SpA-TGC) and canal TGCs (C-TGCs) arise from *Tpbpa*$^{+}$ and both *Tpbpa*$^{+}$ and *Tpbpa*$^{-}$ spongiotrophoblast cells, respectively (Simmons, Fortier and Cross, 2007). SpA-TGCs, C-TGCs and S-TGCs are all in direct contact with the maternal blood, lining the surface of the maternal spiral arteries, junctional zone's canals and labyrinthine synusoids, respectively. The various TGC cell types can be defined by their distinct expression of the prolactins, *Prl2c2*, *Prl3d1* and *Prl3b1* and cathepsin, *Ctsq*, as well as their function and position in the murine placenta (Hemberger *et al.*, 2003).

In the human placenta the equivalent structure, the basal plate, is formed of cytotrophoblast cell columns (CCC), which extend from the chorionic villi of the foetal placenta. The CCC is formed from proliferative cytrotrophoblast, the cells undergoing epithelial to mesenchymal transition (EMT; Vićovac and Aplin, 1996) becoming progressively more invasive as they transition from the base to the tip of the CCC, differentiating to extravillous cytotrophoblast (EVT) cells. The EVTs are highly invasive, polyploid and express the non-canonical HLA protein, HLA-G (Damsky, Fitzgerald and Fisher, 1992; Wakuda and Yoshida, 1992; McMaster *et al.*, 1995; Zybina *et al.*, 2004), to prevent immune cells rejecting the paternal antigens expressed in the zygote-derived placental cells. The cytotrophoblast cells at the base of the column likely represent the elusive human trophoblast stem cell population (Hemberger *et al.*, 2010; Chang and Parast, 2017; Lee *et al.*, 2018).

As great apes and Old-world monkeys, and certain rodents exhibit a relatively invasive, haemochorial type of placentation, mice have been widely exploited in this respect and are an excellent model organism to study pre-implantation development and placentation. In addition, vital to our understanding of early mammalian development has been the advent of *in vitro* culture of embryonic stem cell (ESC) together with the derivation of trophoblast stem cell (TSC) lines. This has enabled further discoveries and the elucidation of key molecular pathways at play during early mammalian embryonic development.

## 1.3 Trophoblast stem cells and reprogramming

Though many of the regulatory networks at play in the murine trophoblast have been elucidated, our knowledge is by no means complete. Differentiation of TSCs into the derivative cell types of the

placenta can be recapitulated and therefore studied *in vitro*. Murine TSCs are cultured in media supplemented with FGF4 and TGFβ, the latter supplied by murine embryonic fibroblast (MEF)-conditioned media (Tanaka *et al.*, 1998; Chiu, Maruyama and Hsu, 2010). Removal of these factors from the culture medium induces differentiation of TSCs along two principle paths, either towards SynT, or towards spongiotrophoblast and terminally differentiated TGCs (Hughes *et al.*, 2004; Latos and Hemberger, 2016; Figure 1.3). At around day 2 differentiation (2D), SynT cells begin to appear due to an increase in *Gcm1* expression, which demarcates both the SynT precursor population, as well as later SynT-II cells. Cells committing to the SynT-I and SynT-II layers can be differentiated based on expression of *Syna* and *Synb*, respectively, which are both expressed during *in vitro* differentiation (Simmons *et al.*, 2008). Simultaneously, other cells differentiating *in vitro* commit to the spongiotrophoblast lineages and transcription of markers, *Ascl2* and *Tpbpa*, can be detected and based on the expression of *Prl3d1*, the TGC population appear around 5D (Tanaka *et al.*, 1998; Hughes *et al.*, 2004).

During unsupervised differentiation in vitro, i.e withdrawal of FGF4 and TGFβ, it has been estimated that only around of 5 % of cells *in vitro* enter the chorionic lineages of SynT or S-TGC cells (Maltepe *et al.*, 2005). The major route of differentiation is that of spongiotrophoblast and more particularly, the TGC. Directed differentiation of TSCs towards the principal lineages, TGCs or SynT cells, can also be performed, to increase the propensity of TSCs towards chorionic or spongiotrophoblast lineages. HDAC inhibition of the global TSC population *in vitro* maintained into differentiation, skews the cells towards SynT cells (Maltepe *et al.*, 2005).

A large gap in our knowledge of human placentation stems from the fact that an equivalent population of human TSCs has only recently been isolated from human blastocysts and first trimester placentas. In 2018, Okae et al. derived a population of cells from both human blastocysts and the cytotrophoblast compartment of first trimester human placentas, which they argue may represent the elusive human trophoblast stem cell. These cells have been cultured *in vitro* for over 60 passages over the course of five months, without seeming to lose proliferative capacity. Differentiation experiments were performed on these cells and Okae et al. demonstrated the ability of these cells to give rise to EVT-like and SynT-like cells, based on morphology and gene expression, as determined by RNA-seq (Okae *et al.*, 2018). These cells have been widely disseminated among researchers and a consensus over their faithful recapitulation of a human TSC remains to be confirmed. Later in 2018 two papers were published outlining the derivation of trophoblast organoids from first trimester placental samples (Haider *et al.*, 2018; Turco *et al.*, 2018). Both systems demonstrated a 3D culture system of a proliferative human trophoblast able to differentiate towards SynT and EVT cells upon culture in defined conditions.

As previously mentioned, several recent papers have demonstrated an increased plasticity of naïve human pluripotent stem cells, which represent the preimplantation epiblast, that are able to transdifferentiate towards trophoblast-like cells (Cinkornpumin *et al.*, 2020; Dong *et al.*, 2020; Guo *et al.*, 2021; Io *et al.*, 2021). Thus, the placental field has expanded rapidly during the last three years with many thrilling new avenues of studying various aspects of placentation and development opening up.

TSCs are a research tool of immense biomedical relevance. They represent a stem cell population with the potential to differentiate into all trophoblast cell types found in the mature placenta. Importantly, most placental-based pregnancy complications have their roots in the very early stages of trophoblast differentiation, yet this developmental time is a major challenge for research as tissue reflecting these earliest stages is almost impossible to obtain. Hence, establishing such systems in humans, and gaining a profound understanding of early regulatory processes of trophoblast expansion and differentiation in both humans and model systems such as the mouse, are of key importance.

## 1.4 Endoreduplication and TGCs

One of the key features of trophoblast giant cells is their acquisition of many hundreds (up to 1000N) of copies of the genome known as polyploidy. Ploidy refers to the overall number of chromosomal units, for instance somatic cells are usually 2n and gametes, n, where n refers to the complete set of chromosomes 1-22 or 1-19 plus X/Y, in humans or mice, respectively. As previously described, these cells are a terminally differentiated, often highly invasive cell type of the rodent placenta, which do not proliferate.

In order to become polyploid, eukaryotic cells utilise atypical cell cycles whereby they skip mitosis (endoreduplication) or do not complete mitosis (endomitosis), between rounds of DNA replication. In *Drosophila*, for example, it is only the neuronal and imaginal disk tissues which do not become polyploid during maturation. In mammals, however, acquisition of polyploidy is rarer, with extreme polyploidy beyond 8N being seen solely in the trophoblast and hence a characteristic of this component of the placenta. Trophoblast giant cells utilise endoreduplication, cycling between growth and synthesis phases without entering mitosis, to acquire many copies of their genome. Whether there is selective amplification of specific chromosomes or portions of chromosomes has been a long-standing cell biological and gene regulatory question among placental biologists trying to understand the remarkable nuclear content of TGCs. According to studies from the Baker lab, this process does not occur uniformly across all chromosomes, with several regions becoming over- or under-amplified in these cells (Hannibal *et al.*, 2014; Hannibal and Baker, 2016). The under-represented regions have

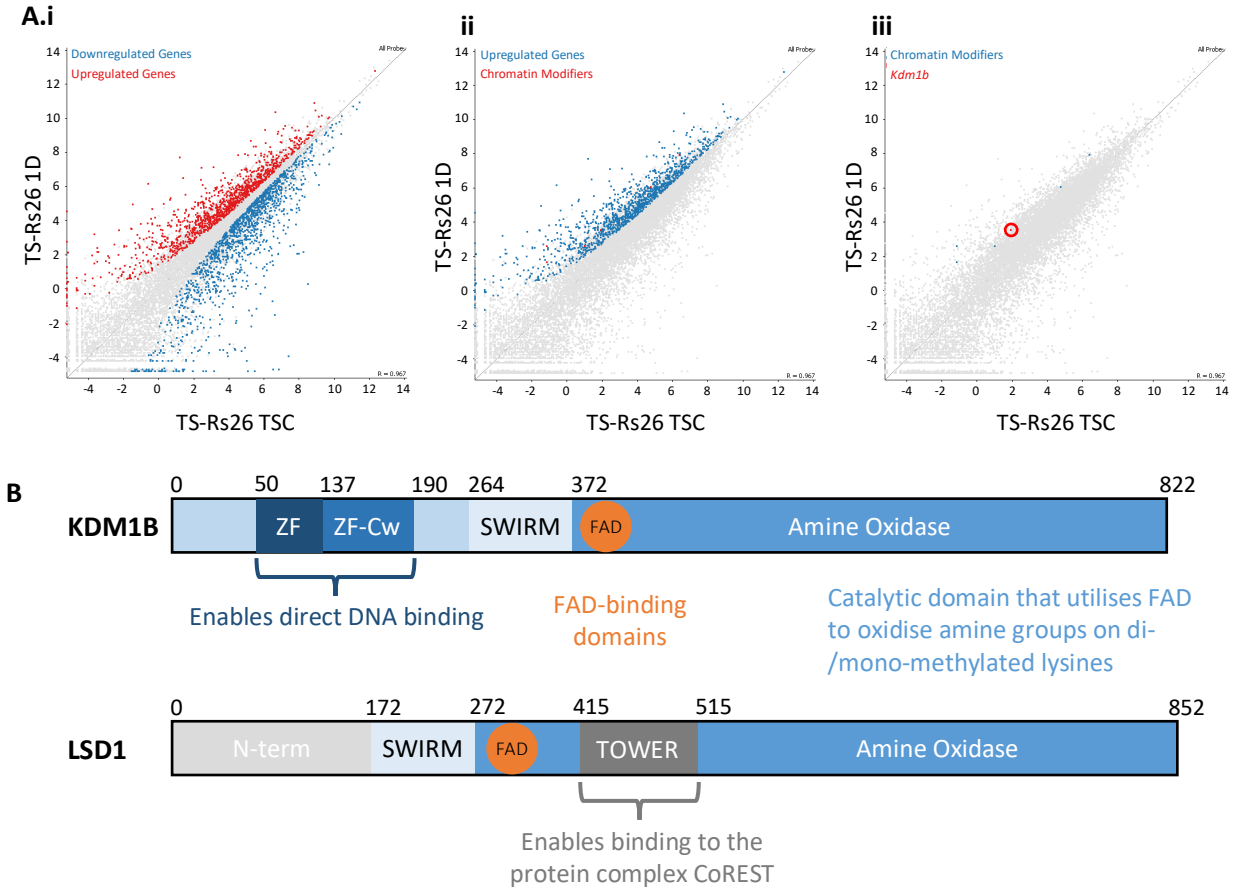
been demonstrated to be late replicating. This means that with each round of DNA replication, these regions are increasingly under-represented in the expanding TGC compartment of the rodent placenta.

In contrast, the overrepresented regions seem to have more of a functional role to play. Four of the five regions expanded in TGCs *in vivo* are found on chromosome 13, with the fifth on chromosome 6. Interestingly, these regions contain gene families that arose via gene duplications and are highly expressed in the trophoblast. On chromosome 13, there are two regions containing the prolactin-like (Prl) hormones, one containing the cathepsins (serine, cysteine or aspartic acid proteases) and the last, the serpins (serine protease inhibitors). The chromosome 6 region contains NK/CLEC complex genes, which are involved in natural killer cell interactions; the equivalent human allorecognition system, KIR/HLA-C, has been shown to play a role in preeclampsia and IUGR. What is important to note is the fact that parietal trophoblast giant cells appear to require further expansion of evolutionarily expanded gene regions. This implies a need to acquire even further copies of already highly expanded gene families, for their function.

## 1.5 Histone lysine demethylase 1B, Kdm1b

The epigenome is instrumental in the establishment and maintenance of cellular identity and direction of differentiation. This work is centred on the hypothesis that changes to the expression of an epigenetic modifier during trophoblast differentiation might indicate that the modifier plays a vital role in stem cell fate and/or differentiation.

Unpublished RNA-seq data generated by the Hemberger group, maps the transcriptional changes occurring over the first six days of TSC differentiation *in vitro*. These data have the potential to highlight novel regulators of both murine trophoblast stem cell self-renewal and lineage specification during differentiation. One such factor that has been identified from this dataset is the histone lysine demethylase, *Kdm1b*. *Kdm1b* encodes a lysine demethylase whose activity is specific to histone H3 lysine 4 di- and mono-methylation (H3K4me2/1). Its expression is up-regulated during the early stages of trophoblast differentiation *in vitro*, (Figure 1.4). The factor was selected for further analysis based on three criteria: we wanted to find a factor whose expression is upregulated early during TSC differentiation, with the hypothesis that preventing expression of this factor, or inhibition of its activity, could “enhance” the stem state and block differentiation. Secondly, the upregulated factor would be an epigenetic modifier, such as a histone acetyltransferase or, in our case a histone lysine demethylase. And finally, the candidate would ideally be known for having a role in placental development (Figure 1.4a).




**C** Schematic representation of the likely role KDM1B plays in chromatin landscape maintenance and gene expression

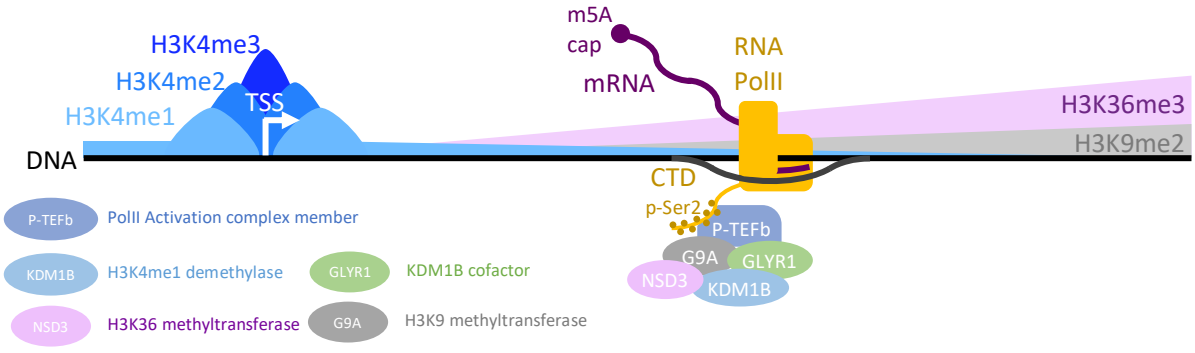



**Figure 1.4 Identification of *Kdm1b* as a target for studying TSCs and differentiation *in vitro***

**A.** Demonstration of isolation of *Kdm1b* from RNA-seq data previously generated in the lab. Scatter plots highlighting **i)** differentially expressed genes at 1 day (1D) differentiation; **ii)** significantly upregulated genes that were also annotated as being a chromatin modifier; **iii)** upregulated chromatin modifier(s) with a placental phenotype, circled is *Kdm1b*. **B.** Comparing the protein structures of KDM1B and family member, LSD1. **C.** KDM1B in complex with its cofactor, GLYR1 as well as methyltransferases, NSD3 and G9A and elongating RNA PolII, within the gene body of an actively transcribed gene.

The other factors identified were three SWI/SNF related proteins, *Smarcd2*, *Smarcd3*, *Smarca2* (Wilson *et al.*, 2014; Jiang *et al.*, 2020; Schim van der Loeff *et al.*, 2021), histone deacetylase, *Sirt7* (Tang *et al.*, 2021) and the transcriptional regultor, *Ctr9* (Bahrampour and Thor, 2016). Interestingly, the only histone methyltransferase in the opposite category, i.e expression rapidly repressed at the onset of differentiation, was *Suv39h2*, an H3K9-methyltransferase that deposits H3K9me3 (Rice *et al.*, 2003; Alder *et al.*, 2010).

Ciccone et al. demonstrated in their 2009 paper that *Kdm1b* is required for the establishment of DNA methylation imprints during oogenesis. They disrupted the *Kdm1b* gene in mice using the Cre/*loxP* recombinase system (Sauer and Henderson, 1988) excising exons 14-16 to induce a frameshift mutation, generating a *Kdm1b* functionally null allele (*Kdm1b*$^{-}$). Mice heterozygous for the *Kdm1b* null allele are developmentally normal and fertile and heterozygous intercrosses produce healthy offspring of wild type (WT), heterozygous (*Kdm1b*$^{+/-}$) and homozygous null (*Kdm1b*$^{-/-}$) genotypes at the predicted ratio, 1:2:1. The male *Kdm1b*$^{-/-}$ offspring are fertile, however the females are not; they display a maternal-effect embryonic lethality. Thus, heterozygous embryos developed from *Kdm1b* null oocytes do not survive beyond embryonic day (E) 10.5 and have various abnormalities including growth impairment, neural tube defects and pericardial oedema, in conjunction with severe placental defects.

These defects likely arise from the absence of maternal methylation imprints at several imprinted differentially methylated regions (DMRs), due to loss of KDM1B. This is due to methylation of H3K4 being protective against DNA methylation, repelling DNA methyltransferase machinery (Ooi *et al.*, 2007). The DMRs effected by loss of KDM1B are those controlling *Mest*, *Grb10*, *Zac1* and *Impact* expression. Consequently, these genes are either expressed bi-allelically (*Mest, Zac1*, *Impact*) or are not expressed at all (*Grb10*). Notably, maternal DMR methylation of other imprinted genes, such as *Igf2r*, *H19*, *Rasgrf1* or *Lit1*, occurs normally in these embryos, pointing to a highly site-specific function of KDM1B.

KDM1B is a histone H3 di/mono-methylated lysine 4 (H3K4me1/2) demethylase of the FAD-dependent amine oxidase family of lysine demethylases. It is closely related to the better-studied histone demethylase, LSD1 (also known as KDM1A), which was the first histone lysine demethylase to be identified (Shi *et al.* 2004). LSD1 catalyses the demethylation of both H3K4me1 and H3K4me2 but can also affect H3K9me1/2, thus acting on both activating and repressive chromatin marks. LSD1 is primarily associated with gene promoters and enhancers through which it regulates gene expression (Kerenyi *et al.*, 2013). Since it removes active marks at these sites, LSD1 is often associated with gene silencing. It sets boundaries between heterochromatin and euchromatin (Chosed and Dent, 2007). A prominent role of LSD1 has also been reported in enhancer decommissioning in ESCs (Whyte *et al.*, 2012). In the trophoblast compartment, LSD1 regulates differentiation onset and the invasive capacity

of TSCs (Zhu *et al.*, 2014) and was found to directly interact with the key trophoblast-specific TF, ESRRB(Latos *et al.*, 2015). LSD1-deficiency results in a reduction in TSC capacity. The TOWER domain of LSD1 located within its amine oxidase (AO) domain recruits a cofactor, CoREST, which is required for its catalytic activity (Shi et al. 2005, Figure 1.4b).

Comparatively less is known about LSD1's family member, KDM1B. Human and mouse KDM1B (89.3 % sequence identity and 94 % sequence similarity) contain zinc finger, SWIRM and AO domains (Figure 1.4b). The AO domain contains an FAD-binding motif, without which no catalytic activity is detected (Fang et al. 2010). KDM1B does not have an obligate requirement for a catalytic-cofactor for activity *in vitro*, however, *in vivo* the cofactor, glyoxylate reductase 1 (GLYR1), interacts with KDM1B to confer function (Fang et al. 2013). GLYR1 binding aids the formation of interactions between KDM1B and the histone H3 tail, specifically with leucine 20 (L20). In marked contrast to LSD1, KDM1B is not associated with gene promoters but is enriched within gene bodies that are highly marked with H3K36me3, a modification indicative of active transcription (Fang et al. 2010; Pokholok et al. 2005). Fang et al. showed that the level of H3K4me2 decreases towards the 3' end of actively transcribed genes, which become concurrently more enriched for H3K36me3. Depletion of KDM1B results in an increase of H3K4me2 as well as a decrease of H3K9me2 at KDM1B-binding sites and a consequent dysregulation of target gene transcription. Furthermore, KDM1B forms active protein complexes with euchromatic histone methyltransferases G9a and NSD3 as well as with cellular factors involved in transcriptional elongation (Fang et al. 2010, Figure 1.4c). The fact that transcription of *Kdm1b* is observed early on in TSC differentiation *in vitro,* leads us to hypothesize that *Kdm1b* may play a role in trophoblast stem cell self-renewal or differentiation.

## 1.6 CRISPR/Cas9 genome editing

The CRISPR/Cas9 genome editing tool allows base-resolution alterations to genomic DNA. This naturally-evolved bacterial defence (Rodolphe *et al.*, 2007; Wiedenheft, Sternberg and Doudna, 2012) has been adapted for use in eukaryotic systems (Qi *et al.*, 2013). The CRISPR/Cas9 genome editing technique takes advantage of the cells' inherent DNA double-strand break (DSB) repair pathways to induce genetic mutations. In eukaryotic cells, two pathways of DNA repair can be employed in response to a DSB. Most often used is the non-homologous end-joining (NHEJ) pathway, which is active in both proliferative and non-dividing cells (reviewed, Lieber, 2010). This is an error-prone pathway utilised during V(D)J recombination and class switch recombination; processes required for lymphocyte development (Ma *et al.*, 2002; Dudley *et al.*, 2005; Yan *et al.*, 2007). NHEJ often results in insertions, deletions, point mutations and even chromosomal translocations (Jäger *et al.*, 2000; Welzel

*et al.*, 2001; Lieber, 2010). During S or G2 phases, when a sister chromatid is available, another pathway of error-free DNA DSB repair can be employed: homologous recombination (HR; San Filippo, Sung and Klein, 2008). In the absence of a template for repair, CRISPR Cas9-induced DSBs are majorly repaired by the error-prone NHEJ pathway, introducing mutations with some predictability (Molla and Yang, 2020).

The CRISPR/Cas9 genome editing tool consists of an engineered guide RNA (gRNA) specific to a region of interest, which anneals to the genome and recruits the Cas9 protein. The *Streptococcus pyogenes* Cas9 (*Sp*Cas9) protein most widely used by researchers contains two catalytic endonuclease domains, HNH and RuvC, which cleave the target and non-target strands, respectively. The target strand is complementary to the sgRNA sequence, to which it anneals. Guide RNAs are designed to target Cas9 to a protospacer adjacent motif (PAM), upon recognition of which, a DSB is then introduced into the genome (Reviewed, Jiang and Doudna, 2017).

Compared to other methods of genetic editing, such as Zinc-Finger Nucleases (ZFNs; Klug, 2010) and Transcriptional Activator-Like Effector Nucleases (TALENs; Sun and Zhao, 2013), CRISPR is highly adaptable. The simplicity with which the target sequence of CRISPR/Cas9 can be changed, makes it a highly attractive system for further engineering and adaptation for both genetic and epigenetic editing. Point mutations in either the HNH or RuvC domains have been used to create Cas9 proteins with nickase capabilities (Jinek *et al.*, 2012). Such nickase Cas9 enzymes induce a single-strand break (SSB). Further, catalytically dead Cas9 (dCas9), containing inactivating mutations in both endonuclease domains, can be fused to epigenetic modifiers to either activate or repress gene expression, known as CRISPR activation (CRISPRa) or interference (CRISPRi), respectively (Larson *et al.*, 2013; Perez-Pinera *et al.*, 2013; Qi *et al.*, 2013; Konermann *et al.*, 2015). An important advancement these activating tools provide, is the opportunity to epigenetically edit endogenous loci, potentially providing more biologically relevant information than traditional overexpression systems.

## 1.7 Hypotheses

Our knowledge of the regulation of self-renewal and differentiation of the trophoblast remains incomplete. By identifying those genes that are up regulated at the onset of differentiation, our hypothesis is that it may be possible to identify novel factors, competent to drive differentiation. The removal or chemical inhibition of such factors could prevent the stem cell population responding to such differentiation cues, thus 'confining' them to the stem state, will be assessed herein. An alternative result of the loss of such factors, could lead to skewed differentiation of trophoblast stem cells into one trophoblast cell lineage over another. In my project I will investigate one such factor,

KDM1B, by studying the effects of protein knock out on TSC potential and differentiation. The aim of my PhD project has been to elucidate the function of the histone demethylase KDM1B in TSC self-renewal and differentiation.

This work aims to

- Characterise the role of histone lysine demethylase, *Kdm1b*, on key regulators of trophoblast stem cells by CRISPR Cas9-induced gene ablation.
- Perform in-depth transcriptomic and epigenomic analysis of the effect of loss of KDM1B on TSC differentiation.
- Integrate transcriptomic and epigenomic data of WT TSC differentiation to identify novel regulatory elements of trophoblast gene expression.

# 2 Chapter Two

# Materials and Methods

## 2.1 Buffers and Media

### 2.1.1 Mammalian cell culture

#### TS Base media, 500 mL

| Reagent | Volume | Final Concentration | Manufacturer, Product Code |
|---|---|---|---|
| FBS | 100 mL | 20 % (v/v) | Gibco, A4766801 |
| Sodium pyruvate, 100 mM | 5 mL | 1 mM | Gibco, 11360039 |
| Penicllin-streptomycin, 5000 U/mL | 5 mL | 50 U | Gibco, 15070063 |
| β-mercaptoethanol, 50 mM | 500 µL | 50 µM | Gibco, 31350010 |
| RPMI 1640 | to 500 mL | | Gibco, 21875034 |

#### Complete TS media, 50 mL

| Reagent | Volume | Final Concentration | Manufacturer, Product Code |
|---|---|---|---|
| bFGF, 50 µg | 25 µL | 25 ng/mL | Sigma, SRP4038-50UG |
| Heparin 1 mg/mL | 50 µL | 1 µg/mL | |
| MEF-conditioned media | 35 mL | | |
| TS Base media | 15 mL | | |

### 2.1.2 Buffers

#### PBST, 500 mL

| Reagent | Amount (unit) | Final Concentration | Manufacturer, Product Code |
|---|---|---|---|
| Tween 20 | 1 mL | 0.1 % | Promega, H5151 |
| PBS | to 500 mL | | |

#### PBST–5 % milk, 50 mL

| Reagent | Amount (unit) | Final Concentration | Manufacturer, Product Code |
|---|---|---|---|
| Skimmed milk | 2.5 g | 5 % (w/v) | Marvel |
| PBST | to 50 mL | | |

#### PBST–1 % BSA, 500 mL

| Reagent | Amount (unit) | Final Concentration | Manufacturer, Product Code |
|---|---|---|---|
| BSA | 5 g | 1 % (w/v) | Sera Care, 1900-0011 |
| PBS | to 500 mL | | |

### PBS–Triton X-100, 500 mL

| Reagent | Amount (unit) | Final Concentration | Manufacturer, Product Code |
|---|---|---|---|
| Triton X-100 | 1.5 mL | 0.3 % (v/v) | ThermoFisher, 28313 |
| PBS | to 500 mL | | |

### Western blot running buffer 10X, 1 L

| Reagent | Amount (unit) | Final Concentration | Manufacturer, Product Code |
|---|---|---|---|
| Tris Base | 30.3 (g) | 3.03 % (w/w) | |
| Glycine | 144 (g) | 14.4 % (w/w) | Thermo Scientific, J16407.A1 |
| 20 % SDS | 50 (mL) | 1 % | Bio-Rad, 1610418 |
| Ultrapure $dH_2O$ | to 1L | | |

Dilute 10X buffer 1:10 in $H_2O$ for working buffer

### Western blot transfer buffer 10X, 1 L

| Reagent | Amount (unit) | Final Concentration | Manufacturer, Product Code |
|---|---|---|---|
| Tris Base | 30.3 (g) | 3.03 % (w/w) | |
| Glycine | 144 (g) | 14.4 % (w/w) | Thermo Scientific, J16407.A1 |
| Ultrapure $dH_2O$ | to 1L | | |

Dilute 10X buffer 1:10 in $H_2O$, 20 % (v/v) methanol for working buffer

## 2.1.3 ChIP Buffers

### MNase dilution buffer, 1 mL

| Reagent | Volume (µL) | Final Concentration | Manufacturer, Product Code |
|---|---|---|---|
| 1 M Tris, pH7.5 | 10 | 10 mM | Sigma, T2319 |
| 1 M NaCl | 50 | 10 mM | Sigma, 71386 |
| 0.5 M EDTA | 2 | 1 mM | |
| Glycerol | 500 | 50 % | Sigma, G5516 |
| Ultrapure $dH_2O$ | 438 | | |

### Complete immunoprecipitation (IP) buffer, 5 mL

| Reagent | Volume (µL) | Final Concentration | Manufacturer, Product Code |
|---|---|---|---|
| 1 M Tris, pH8.0 | 100 | 20 mM | Sigma, T2694 |
| 5 M NaCl | 150 | 150 mM | Sigma, 71386 |
| 0.5 M EDTA | 20 | 2 mM | |
| Triton-X100 | 5 | 0.1 % | Sigma, G5516 |
| Protease Inhibitor Cocktail | 50 | 1X | Sigma, P8340 |
| PMSF | 50 | 1 mM | Sigma, P7626 |
| Ultrapure $dH_2O$ | 4.625 mL | | |

### Low salt wash buffer, 10 mL

| Reagent | Volume (µL) | Final Concentration | Manufacturer, Product Code |
|---|---|---|---|
| 1 M Tris, pH8.0 | 200 | 20 mM | Sigma, T2694 |
| 5 M NaCl | 300 | 150 mM | Sigma, 71386 |
| 0.5 M EDTA | 40 | 2 mM | |
| Triton-X100 | 100 | 1 % | Sigma, G5516 |
| 20 % SDS | 10 | 0.2 % | |
| Ultrapure $dH_2O$ | 9.350 mL | | |

### High salt wash buffer, 10 mL

| Reagent | Volume (µL) | Final Concentration | Manufacturer, Product Code |
|---|---|---|---|
| 1 M Tris, pH8.0 | 200 | 20 mM | Sigma, T2694 |
| 5 M NaCl | 1000 | 500 mM | Sigma, 71386 |
| 0.5 M EDTA | 40 | 2 mM | |
| Triton-X100 | 100 | 1 % | Sigma, G5516 |
| 20 % SDS | 10 | 0.2 % | |
| Ultrapure $dH_2O$ | 8.650 mL | | |

### ChIP elution buffer, 2 mL

| Reagent | Volume (µL) | Final Concentration | Manufacturer, Product Code |
|---|---|---|---|
| 1 M $NaHCO_3$ | 200 | 100 mM | |
| 20 % SDS | 1000 | 1 % | |
| Ultrapure $dH_2O$ | 1700 | | |

## 2.2 Oligos and antibodies

### 2.2.1 Table 2.2.1 DNA Oligos

| Experiment | | Oligo ID | Sequence 5'-3' |
|---|---|---|---|
| Diagnostic PCR to identify knockouts | Pair 1 | Intron4_F | GCTGTAGGAATCACCATGGAC |
| | | Exon5_R | GGTAACTCCACAAGCCAACTC |
| | Pair 2 | Intron4_F | GCTGTAGGAATCACCATGGAC |
| | | Intron10_R | AGGAGCAGTGAACCATCTATCC |
| | Pair 3 | Intron4a_F | CTGGATGCGGGAAACAGTAAAC |
| | | Intron11a_R | CAGCACTTGAAAGGCAGAGG |
| | Pair 4 | Intron4b_F | GATAAGGGTGCAGACACTTGG |
| | | Intron11b_R | GTTCAACCAACCCATCCCAC |
| gRNA oligos for CRISPR | gRNA 1a | kdm1b_Intron4_1aF | CACCGTGCACGCTCCTCCTGATAG |
| | | kdm1b_Intron4_1aR | AAACCTATCAGGAGGAGCGTGCAC |
| | gRNA 1b | kdm1b_Intron4_1bF | CACCGACCAGCTCAAGTTGTAGTG |
| | | kdm1b_Intron4_1bR | AAACCACTACAACTTGAGCTGGTC |
| | gRNA 2a | kdm1b_Intron5_2aF | CACCGGGCCTCGACTTGCTAGACC |
| | | kdm1b_Intron5_2aR | AAACGGTCTAGCAAGTCGAGGCCC |
| | gRNA 2b | kdm1b_Intron5_2bF | CACCGTGGGCCATATGGCGTTTCTA |
| | | kdm1b_Intron5_2bR | AAACTAGAAACGCCATATGGCCCAC |
| | gRNA 3a | kdm1b_Intron10_2aF | CACCGTACCATCACGTACCGTACCG |
| | | kdm1b_Intron10_2aR | AAACCGGTACGGTACGTGATGGTAC |
| | gRNA 3b | kdm1b_Intron10_2bF | CACCGTCACGTACCGTACCGTGGGA |
| | | kdm1b_Intron10_2bR | AAACTCCCACGGTACGGTACGTGAC |
| RT-qPCR | *Kdm1b* | Kdm1b_Forward | GCCATCAACAGCTTGGGTGC |
| | | Kdm1b_Reverse | CTGGCACTGGGAGGAACATG |
| | *Esrrb* | Esrrb_Forward | AGTACAAGCGACGGCTGG |
| | | Esrrb_Reverse | CCTAGTAGATTCGAGACGATCTTAGTCA |
| | *Cdx2* | Cdx2_Forward | AGTGAGCTGGCTGCCACACT |
| | | Cdx2_Reverse | GCTGCTGCTGCTTCTTCTTGA |
| | *Eomes* | Eomes_Forward | TCGCTGTGACGGCCTACCAA |
| | | Eomes_Reverse | AGGGGAATCCGTGGGAGATGGA |
| | *Elf5* | Elf5_Forward | ATTCGCTCGCAAGGTTACTCC |
| | | Elf5_Reverse | GGATGCCACAGTTCTCTTCAGG |
| | *Gcm1* | Gcm1_Forward | GAAGAGCAGTTTCAGCCTCC |
| | | Gcm1_Reverse | GATCTAAGCCCATGCATGCC |
| | *Syna* | Syna_Forward | CCTCACCTCCCAGGCCCCTC |
| | | Syna_Reverse | GGCAGGGAGTTTGCCCACGA |
| | *Synb* | Synb_Forward | TCCGGAAAGGGACCTGCCCA |
| | | Synb_Reverse | CAGCAGTAGTGCGGGTGCC |
| | *Ascl2* | Ascl2_Forward | AGCCCGATGGAGCAGGAG |
| | | Ascl2_Reverse | CCGAGCAGAGGTCAGTCAGC |
| | *Tpbpa* | Tpbpa_Forward | ACTGGAGTGCCCAGCACAGC |
| | | Tpbpa_Reverse | GCAGTTCAGCATCCAACTGCG |
| | *Plf* | Plf_Forward | TTCCCATGTGTGCAATGAGG |
| | | Plf_Reverse | AGTCATTGTCTAGGCAGCTG |
| | *Pl2* | Pl2_Forward | GCACTCGGGGAACAGCAGCC |
| | | Pl2_Reverse | ACTGCCAGGAACAGGAGTGCC |
| | *Lsd1* | Lsd1_Forward | GCCACCTCTTCCTGAGTGGAA |
| | | Lsd1_Reverse | TTGGGTCCCAGAACACACGG |
| | *Sdha* | Sdha_Forward | TGGTGAGAACAAGAAGGCATCA |
| | | Sdha_Reverse | CGCCTACAACCACAGCATCA |

### 2.2.2 Table 2.2.2 Antibodies

| Experiment | Species | Antibody (dilution/ amount) | Manufacturer | Product code |
|---|---|---|---|---|
| WB/IF | Rabbit | KDM1B (1:1000) | Abcam | ab193080 |
| WB | Rabbit | LSD1 (1:1000) | Abcam | ab17721 |
| WB | Rat | TUBULIN (1:2000) | Abcam | ab6160 |
| ChIP-seq | Rabbit | H3K4me3 (250 ng) | Diagenode | C15410003 |
| WB/ChIP-seq | Rabbit | H3K4me2 (62.5 ng) | Abcam | ab7766 |
| WB/ChIP-seq | Rabbit | H3K4me1 (125ng) | Abcam | Ab8895 |
| ChIP-seq | Rabbit | H3K36me3 (250 ng) | Diagenode | C15410192 |
| Histone WB | Rabbit | H3K4me3 (1:1000) | Abcam | ab8580 |
| Histone WB | Rabbit | H3K36me3 (1:1000) | Abcam | ab194677 |
| Histone WB | Rabbit | H3K9me3 (1:1000) | Abcam | ab8898 |
| Histone WB | Mouse | H3K9me2 (1:1000) | Abcam | Ab1220 |
| Histone WB | Mouse | Total H3 (1:1000) | NEB | 14269S |
| Histone WB | Rabbit | Total H3 (1:1000) | Abcam | Ab1791 |
| WB | Mouse | RNAP2 (1:1000) | Biolegend | 664912 |
| WB | Rabbit | phosphoSer2-RNAP2 (1:1000) | Abcam | Ab193468 |
| WB | Rabbit | phosphoSer5-RNAP2 (1:1000) | Abcam | Ab193467 |
| Histone WB | Goat | Rabbit IRDye 800CW (1:5000) | LICOR | 926-32211 |
| Histone WB | Goat | Mouse IRDye 680RD (1:5000) | LICOR | 926-68070 |
| WB | Goat | Rat-HRP (1:5000) | BIO-Rad | 5204-2504 |
| WB | Goat | Rabbit-HRP (1:5000) | BIORAD | 170-6515 |
| WB | Goat | Mouse-HRP (1:5000) | BIORAD | 170-6516 |

WB, western blot; IF, immunofluorescence; ChIP, Chromatin immunoprecipitation

## 2.3 Mammalian Cell Culture

TSC lines, TS-Rs26 and TS-EGFP, were cultured in Complete TS media consisting of 70 % pre-conditioned TS Base media on murine embryonic fibroblasts, at 37 °C in a 5 % $CO_2$ atmosphere. Human embryonic fibroblast cell lines, IMR90 (Sigma) and WS1 (ATCC-1502), were cultured in IMR90 and WS1 media (Buffers and Media), respectively. All mammalian cell lines were passaged upon reaching 80 % confluency, to maintain cell viability (splitting between 1:20 and 1:50 for Ts-RS26 cells and 1:4 with IMR90, WS1 and HEK293T cells). The cells were passaged by washing with phosphate buffered saline (PBS), before being lifted by incubation in 0.05 % (fibroblasts) or 0.25 % (TSCs) Trypsin (Gibco).

## 2.4 Generating clonal *Kdm1b*$^{-/-}$ TS-Rs26 cell lines

### 2.4.1 gRNA design and cloning

The online CRISPR.mit.edu design software was used to generate gRNA sequences targeting the murine *Kdm1b* gene. Three locations were chosen: intron 4 by gRNAs 1a/b, intron 5 by gRNAs 2a/b and intron 10 by gRNAs 3a/b (sequences listed in Table 2.1). Oligonucleotides of the selected sequences were generated by Sigma-Aldrich and resuspended to a 100 µg.mL$^{-1}$ solution in $H_2O$. These gRNAs were cloned into the pCas9.2A.EGFP plasmid (Addgene #48138) and their sequences verified.

### 2.4.2 Transfection of TS-Rs26 cells

The gRNA-containing vectors were transfected into ESCs, using Lipofectamine 2000 reagent, as per the manufacturers' instructions in combinations such that both upstream gRNA (targeting intron 4 of *Kdm1b*) was co-transfected with each downstream gRNA (targeting either intron 5, Combination1/2, or intron 10, Combination3/4; Table 2.4.3). The *Kdm1b* gRNA 1a (Table 2.2.1) did not appear to produce any positive band in the bulk ES cell PCR reaction, so only gRNA 1b was used for generation of knock-out clones in TS-Rs26 cells. Twenty-four hours post-transfection, the six TS-Rs26 populations were trypsinised, passed through a 40 µm filter and stained with 3 µM 4,6-diamidino-2-phenylindole (DAPI) for flow cytometry. DAPI$^-$/EGFP$^+$ cells were single cell-sorted into 96 well plates, collecting one plate (i.e. 96 cells) per transfection condition.

### 2.4.3 Table 2.4.3 Combination of pCas9.2A.EGFP.*Kdm1b*gRNA vectors transfected into TS-Rs26 cells for knock out of *Kdm1b*

| Transfection | 2.6 µg total DNA per transfection | | Deletion | Knockouts Recovered? |
|---|---|---|---|---|
| Combination 1 | p.gRNA 1b | p.gRNA 2a | Exon 5 | No |
| Combination 2 | p.gRNA 1b | p.gRNA 2b | Exon 5 | No |
| Combination 3 | p.gRNA 1b | p.gRNA 3a | Exons 5-10 | Yes |
| Combination 4 | p.gRNA 1b | p.gRNA 3b | Exons 5-10 | Yes |
| Control 1 | pCas9.2A.EGFP | | WT | n/a |
| Control 2 | pCas9.2A.EGFP | | WT | n/a |

## 2.5 Diagnostic polymerase chain reaction (PCR)

Genomic DNA (gDNA) extraction was performed using QuickExtract reagent (QE09050; Epicentre, Illumina) as per the manufacturer's instructions. PCRs were set up containing 1 µL extracted gDNA, 200 µM each dNTP, 1 µM of each forward and reverse primer (Table 2.2.1), 2 µL 10X CoralLoad Buffer and 0.5 U HotStar Taq Plus DNA polymerase (203605, QIAGEN) in water, to a volume of 20 µL.

## 2.6 RNA extraction and Turbo DNase® treatment

RNA extraction was performed using TRI reagent (T9424, Sigma) as per the manufacturer's instructions and stored at -80 °C in DEPC-$H_2O$ for further use. TRI reagent, 800 µL, was added to one 80 % confluent well of a 6well plate, on ice, the cells thoroughly lysed by pipetting up and down and scraping with the pipette tip. The TRI reagent was then transferred to a clean Eppendorf, then 160 µL of chloroform was added and thoroughly mixed by inversion. The mixture was incubated at room temperature with frequent thorough mixing by inversion. The mixture was then centrifuged at 13,000 rpm at 4 °C for 15 minutes. The aqueous phase was then combined with 400 µL isopropanol containing glycogen. The samples mixed thoroughly by inversion and the RNA pelleted by centrifugation at 13,000 rpm, 4°C for 15 minutes. The pellet was then washed with ice cold 80 % ethanol in DEPC-$H_2O$ and re-centrifuged at 13,000 rpm for 5 minutes at 4°C. The pellet was air-dried at room temperature until matte, then resuspended in TE or DEPC-$H_2O$ using a volume dependent on the size of the pellet.

Following RNA extraction, 2 µg RNA was treated with TURBO DNase (AM1907M, Thermo-Fisher Scientific) to remove any genomic DNA prior to RNA-seq library preparation. Following resuspension of the RNA in 1X TURBO DNase buffer and with 1 U TRUBO DNase enzyme, the mixture was incubated at 37°C for 20-30 minutes. Following digestion, 2 µL (1/$10^{th}$ volume) SLURRY was added and the mixture incubated at room temperature for 5 minutes with frequent mixture by flicking the tube. The mixture was then centrifuged for 2 minutes at 13,000 rpm at room temperature and the DNase-treated RNA collected in a clean, labelled, DNA lobind® Eppendorf tube; then stored -80 °C.

## 2.7 Reverse transcription coupled with quantitative PCR (RT-qPCR)

Reverse transcriptase (RT) reactions were performed on each RNA extraction using equivalent amounts of RNA; relative RNA concentrations were quantified by visualisation on 1 % agarose (WebScientific) gel or measurement on Nanodrop. Initially a mixture of RNA and 0.3 ng.$\mu L^{-1}$ mix of Random Hexamer and Oligo $dT_{(18)}$ primers (SO142 and SO132 respectively; ThermoFisher, MA, USA) were heated to 70 °C for 5 minutes to denature the RNA and break up any secondary structures. The RT reaction was set up using RevertAid H Minus First Strand reverse transcriptase (EP0451; ThermoFisher) as per the manufacturer's instructions. The cDNA was then diluted 1:30 in DEPC-$H_2O$.

Real time quantitative PCR (qPCR) was performed in triplicate on each cDNA, per gene analysed. Each 12 µL reaction contained 5 µL cDNA, 6 µL SYBR Green Jump Start Ready Mix (S4438; Sigma-Aldrich), 0.12 µL Reference Dye and 0.24µl (1 µM each) forward and reverse primer (Table 2.2.1) in DEPC-$H_2O$. The qPCR reactions were performed on a Bio-Rad CFX-384 real time PCR thermocycler.

The generated Cq (quantification cycle) values were converted to relative expression values by calculating delta Cq (ΔCq) values, obtained by subtracting the Cq of the housekeeping gene *Sdha* per sample, followed by power calculation using the exponent –ΔCq to the base of 2 ($2^{-\Delta Cq}$).

## 2.8 Protein extraction

Cells for protein extraction were plated in 6 cm or 10 cm dishes, washed in PBS and the plate snap frozen on dry ice. The plates were then stored at -80 °C or placed on ice for protein extraction. Using around 100 µL or 250 µL RIPA buffer for 6 or 10 cm dishes, respectively. The cells were thoroughly scraped lysate collected in labelled eppendorfs, on ice. Samples were then rotated at 25 rpm for 15 minutes followed by centrifugation at 13,000 rpm for 15 minutes, all at 4 °C. The lysate was removed to a new labelled Eppendorf and stored at -80 °C.

## 2.9 Western blot analysis

Cells were lysed in an appropriate volume of RIPA Buffer with proteinase and phosphatase inhibitors and stored at -80 °C (2.8). Prior to loading an 8 % SDS-PAGE gels, the protein concentration of each sample was quantified by BCA Assay (23227; Thermo) and samples were diluted in 6X Laemmli buffer before denaturing the mixture at 95 °C for 5 minutes.

The proteins were separated along the gel by electrophoresis at 180 V for up to 90 minutes. The protein was transferred to an activated membrane in 1X transfer buffer by passing a constant current of 250 mA, from gel to membrane, for 90 minutes. After staining the membranes with Ponceau S to check equal protein loading and successful transfer, the membranes were washed in 0. 1 % TBS-Tween (TBS-T), then blocked in 0. 1 % TBS-T-5 % skimmed milk (TBS-T-Milk) for up to 2 hours at room temperature. The blots were then incubated overnight in primary antibody diluted as appropriate (Table 2.2.2) in TBS-T-Milk, at 4 °C. The blots were washed three times in TBS-T before incubation in secondary antibody, diluted in TBS-T-Milk, for 1 hour at room temperature. Following another wash step, the blots were briefly incubated in ECL™ reagent (Sigma, GERPN3245) before exposing X-ray film, which was developed using MI-5 X-ray film processor.

## 2.10 Histone western blot analysis

Histone extractions were performed on cell pellets suspended in 6X Laemmli buffer; the samples were then heated to 95 °C then sonication for 15 minutes, which was repeated. Histone samples were then

diluted in 18 µL $H_2O$ and run on a 14 % SDA-PAGE gel. The gel was run for 60 minutes at 170 V, the protein then transferred to an activated membrane, as above (2.9). The proteins were visualised by Ponceau S staining to gauge relative concentration for equal loading.

The above steps were repeated adjusting for apparent concentrations to achieve more equal loading. After transferring proteins to an activated membrane, it was immediately blocked in TBS-milk. The membrane was then incubated in TBS-milk with primary antibodies diluted as required (Table 2.2.2), rocking at 25 rpm, overnight, at 4 °C. Blots were incubated with an antibody specific to one of H3K4me1/2/3, H3K9me2/3 or H3K36me3 as well as a total H3 antibody, ensuring both antibodies were raised in different species. After washing in TBS, the blots were incubated at room temperature in both anti-mouse and anti-rabbit secondary antibodies (Table 2.2.2) diluted as appropriate in TBS, rotating at 25 rpm for 1 hr. Blots were visualised using LICOR Odyssey infrared blot scanner.

## 2.11 Immunofluorescence

TS-Rs26 cells were plated in 6 well plates onto cover slips (sterilized by baking at 200 °C) and incubated in TS-Base or complete TS media. All further steps were carried out at room temperature. Cells were fixed in 4 % paraformaldehyde (PFA) for 10 minutes, then washed with 2 x 2 mL PBS before being permeabilised by incubation in PBS-0.3 % Triton X-100 (PBS/Triton) for 30 minutes. The cells were then blocked in PBS-0.1 % Tween-1 % BSA (PBT/BSA) for 30 minutes before another wash step, followed by incubation in 100 µL Kdm1b antibody (Table 2.2.2) diluted as appropriate in PBT/BSA for 90 minutes. Following another wash step, the cells were then incubated in the dark in 100 µL PBT-BSA containing the appropriate fluorescence-labelled secondary antibody for 60 minutes. After another wash, DAPI was added to the cells diluted 1:1000 in PBS, for 5 minutes, and the cover slips washed again. The cover slips were then mounted onto slides with 50 % glycerol (G/0650/17; Fisher Scientific) in PBS.

## 2.12 RNA-seq Library preparation

Libraries were prepared from 350 ng DNase-treated RNA (2.6) using the Ilumina TruSeq® Stranded mRNA Library Preparation kit (20020595), adapted for use with PCR tubes rather than using a PCR plate. Following library synthesis, cDNA was purified using AMPure XP beads. Libraries were quantitated and their quality assessed by Bioanalyser 2100 system (Agilent) and KAPA Library Quantification Complete Kit (KAPA Biosystems, KK4824). Pooled libraries were then sequenced on an Illumina HiSeq2500 sequencer using a 50bp single-end protocol. Raw fastq data were mapped to the

Mus musculus GRCm38 genome assembly using Hisat 2 by the Bioinformatics facility at Babraham Institute.

## 2.13 Ultra-low input native chromatin immunoprecipitation (ChIP) followed by next generation sequencing (ULI-nChIP-seq)

The ULI-nChIP-seq protocol (Hanna *et al.*, 2018) was adapted for use with *in vitro* cultured TS-Rs26 cells, using 5000 cells per ChIP.

### 2.13.1 Chromatin preparation and MNase digestion

Following culture, TS-Rs26 cells were snap frozen in PBS and stored at -80°C until use for ChIP-seq library preparation. Samples were thawed and 20,000 cells per sample were resuspended in nuclear isolation buffer and nuclei lysed with 0.1% triton/0.1% deoxycholate. The chromatin was then prepared for MNase digestion in 1X MNase Buffer containing 5 % PEG6000 and 2mM DTT in (New England Biolabs, M0247S). MNase digestion was carried out at 21 °C for 7.5 minutes, after which the reaction was stopped with 10 mM EDTA. Chromatin was prepared for ChIP followed by NGS library preparation.

### 2.13.2 ChIP-seq

Protocol adapted from Brind'Amour *et al.*, (2015). The MNase digested chromatin was then pre-cleared by incubation in a 1:1 mix of Protein A and Protein G beads (ThermoFisher Scientific, 10002D, 10004D, respectively) for 3 hours at 4 °C, by rotating at 25 rpm. Prior to ChIP, ~10 % of each sample was removed and stored at -20 °C for generation of reference input libraries. ChIPs were then performed using Protein A and G beads pre-loaded with antibodies specific to H3K4me1, H3K4me2, H3K4me3 or H3K36me3 (Table 2.2.2); samples were rotated at 25 rpm overnight, at 4 °C. Following salt washes (2.1.3), the DNA was eluted from antibody-bound beads ChIP elution buffer (2.1.3). The DNA was then purified with a 1:1.8 ratio of SPRI beads and from this, ChIP-seq libraries were prepared, on ice, using the Microplex preparation Kit v2 (48 indices, Diagenode, C05010014) as per the manufacturers' instructions.

The indexed libraries were then purified followed by quantitation and their quality checked using Bioanalyser 2100 system (Agilent) and KAPA Library Quantification Complete Kit (KAPA Biosystems, KK4824). The libraries were pooled and sequenced on a NextSeq500 HighOutput using 75 bp Single End protocol. Raw fastq data were mapped to the Mus musculus GRCm38 genome assembly using Bowtie 2 by the Bioinformatics facility at Babraham Institute.

## 2.14 Bioinformatic analysis

Raw next generation sequencing (NGS) data was processed and mapped by the Bioinformatics facility at The Babraham Institute.

### 2.14.1 RNA-seq analysis

Visualisation of the RNA-seq data was performed using the RNA-seq quantitation pipeline in the SeqMonk software (http://www.bioinformatics.babraham.ac.uk), normalised according to total read count (RPKM). Differentially expressed (DE) genes were identified using DESeq2 (Love, Huber and Anders, 2014) and by the intensity difference filter in SeqMonk, which identifies genes whose differential expression is an outlier in the distribution of expression differences in the system, with a p-value threshold of $p<0.05$ with the Benjamini-Hochberg method of correction for multiple comparison. Stringently identified genes were identified by both DESeq and intensity difference filter, with less stringent DE genes identified by one method or both.

### 2.14.2 ChIP-seq analysis

Peaks were called on bam files from H3K4me1, H3K4me2 and H3K4me3 ChIP-seq libraries using the MACS2 tool (Zhang *et al.*, 2008; Feng *et al.*, 2012) with an FDR cutoff of 0.05, using grouped input samples for each time point as reference. Consensus peak lists were generated containing all called peaks or WT-only peaks for each histone mark, removing regions with high enrichment in input samples. Reads under MACS2 peaks were quantitated and normalised to largest data store, further normalisation was carried out at the $40^{th}$ and $99^{th}$ percentiles and $20^{th}$ and $90^{th}$ percentiles to account for variation in the ChIP efficiency in H3K4me1/3 and H3K4me2 data, respectively. Differentially enriched peaks were then identified by LIMMA statistical analysis (Smyth, 2004). Active enhancer status in stem cells was defined by co-enrichment of H3K4me1 and H3K27Ac using published data available at accession number GSE42207 in NCBI GEO database.

For visualisation of H3K36me3 data, reads within 1 kb running windows of both H3K36me3 and H3K4me3 data were quantitated and normalised to total read count (RPKM), excluding regions with high read counts in WT input samples. Highly enriched 5 day differentiated regions within the H3K36me3 data were identified as regions with RPKM difference greater than 25 between H3K36me3 3D and 5D samples.

### 2.14.3 Gene ontology and motif enrichment analysis

For identification of enriched terms in the RNA-seq data, enrichment was calculated using Gene Ontology enRIchment anaLysis and visuaLizAtion tool (Gorilla, http://cbl-gorilla.cs.technion.ac.il/, Eden et al., 2009) using all genes with at least 10 reads in at least one time point, as a background.

Terms with a p-value corrected for multiple testing by Benjamini and Hochberg method p<0.05. Motif analysis was performed using the find.MotifGenome.pl tool in the HOMER motif analysis pipeline (http://homer.ucsd.edu/homer/motif). Central 1 kb regions of ChIP-seq peak lists were analysed using randomly selected size-matched regions of mm10 genome as background reference. Statistically enriched motifs were identified with a p-value less than 0.05.

## 2.15 Metaphase spreads

Cells were cultured as described in section 2.3. Cells plated in 10cm dishes were incubated in KaryoMAX (0.1 μg/mL, Gibco, 15212012) for 40 minutes at 37 °C. The cells were then washed in HBSS Buffer (ref) before trypsinisation. Following centrifugation, the cells were resuspended in 0.075 M KCl, vortexing gently, and incubated at 37 °C for 10 minutes. The cells were then washed in fresh Carnoy's fixative (3:1 ratio of methanol:glacial acetic acid) twice before finally being resuspended in 10 mL Carnoy's fixative and being stored at 4 °C for further use.

An aliquot of fixed cells were then dropped onto glass slide from a distance of 2 m and air-dried for 10 minutes. The cells were then stained for DAPI for 10 minutes, covering from light. The slides were then washed twice in distilled water and left to dry. A cover slip was then mounted with 50 % glycerol in PBS and the slide dried.

# 3 Chapter Three

## Impact of loss of KDM1B on TSCs

## 3.1 Introduction

My first goal was to try to identify a target that would meet our criteria for a potential epigenetic regulator that could play a role in TSC *in vitro* differentiation. I began by assessing RNA-seq data that had previously been generated in the Hemberger Lab using two TSC lines, TS-Rs26 and TS-EGFP, that were subjected to a 6-day differentiation time course and analysed for transcriptomic changes that occur during differentiation (data unpublished). Sequencing libraries were generated for each cell line from stem to 6D differentiated cells, in duplicate.

As per our hypothesis, I began by identifying differentially expressed (DE) genes between trophoblast stem cells and their earliest differentiated descendants, i.e. TSCs after one day of *in vitro* differentiation. Comparing the two cell lines separately and then combined, as they produced very similar lists of differentially expressed genes. As my hypothesis was based on the idea that a gene whose expression is induced at the onset of differentiation likely plays a role in directing differentiation and its loss might enhance the stem cell state, I filtered the DE gene list for with a log fold-change (LFC) greater than 2, between stem and 1 day differentiation. This left a list of 1128 up-regulated genes (Figure 1.4a). I then decided to use the GO term, *GO:0040029: epigenetic regulation of gene expression*, and category *chromatin modifier* from the transcription factor database (Zhang *et al.*, 2015) to filter out any genes that would be of particular interest. This produced a list of just 13 genes. I then assessed whether any of these had been previously studied and had a placental phenotype attributed to them *in vivo*. This left me with only one candidate, *Kdm1b* (Circled, Figure 1.4a).

Ciccone et al. saw a maternal effect with embryonic lethality of heterozygous foetuses of *Kdm1b*$^{-/-}$ mothers at mid-gestation, at E10.5 of mouse development, with *severe placental defects and various embryonic abnormalities, including growth impairment, neural tube defects and pericardial oedema*. These phenotypes are likely due to an imprinting defect conferred by loss of KDM1B in the null oocytes (Ciccone *et al.*, 2009). Several maternal differentially methylated regions (DMRs) were unmethylated in the KDM1B null oocytes, which resulted in incorrect expression of the associated genes and thus failure of the foetus. Mid-gestational lethality is well known for often be caused by a placental defect (Copp, 1995; Perez-Garcia *et al.*, 2018). Also, the *Kdm1b* maternal null phenotype phenocopies that of other maternal effect genes involved in setting the oocyte-specific DNA methylation marks, notably *Dnmt3l* (Arima *et al.*, 2006). The fact that *Kdm1b*$^{-/-}$ mice grew to full term and matured to fertile adults, does not necessarily mean that there was not a placental phenotype with this mouse.

Given its expression is induced in TS-Rs26 and TS-EGFP cells at the onset of differentiation by withdrawal of stem-promoting factors, FGF4 and TGF$\beta$, I hypothesise that KDM1B plays a role in the trophoblast. This epigenetic modifier is known to associate with active transcriptional machinery and

to demethylate H3K4me2/1, thereby effecting gene expression. This chapter will determine the temporal changes to transcription of *Kdm1b* and expression of KDM1B by TSCs during differentiation *in vitro*. Also assessed will be the effect of loss of KDM1B on TS-Rs26 cells' stem state and during *in vitro* differentiation. This will be done by quantitating the expression of certain trophoblast marker genes and *Kdm1b*'s family member, *Lsd1*, as well as assessing the amount of KDM1B-associated histone modifications and finally, cellular proliferation.

## 3.2 Results

### 3.2.1 Assessment of KDM1B expression in TS-Rs26 cells

Having identified *Kdm1b* as my gene of interest, I decided to confirm the dynamics of *Kdm1b* expression during *in vitro* TSC differentiation to corroborate the RNA-seq data (Figure 3.1a). Initially, I measured *Kdm1b* transcription by RT-qPCR over a five day differentiation time course, starting from the stem cell state (Figure 3.1b). RT-qPCR data demonstrated that *Kdm1b* transcription increases gradually over the course of differentiation, reaching a peak at 3 days (3D) differentiation that is maintained to 5D. Comparing this to the protein levels of KDM1B measured by western blot (Figure 3.1c), the amount of bulk protein continued to increase throughout the differentiation time course.

When assessing the populations of cells expressing KDM1B by immunofluorescence (Figure 3.1d), however, I saw that even in stem conditions, the cells positive for KDM1B (KDM1B$^+$) were potentially more highly differentiated, being larger than other cells in the dish. Figure 3.1e shows higher magnification of these large KDM1B$^+$ cells at 1D differentiation. It is well known that TSCs display considerable levels of heterogeneity even in stem cell culture conditions, with some cells undergoing spontaneous differentiation. Throughout the course of differentiation, the proportion of KDM1B$^+$ cells increased until all cells were staining positive for KDM1B at 5D, albeit to differing extents. Moreover, my results showed that the majority of the KDM1B protein was localized to the nucleus, in line with the function of this protein as a chromatin modifier.

The western blot data combined with the immunofluorescence analysis suggest that the low levels of KDM1B transcription and protein at the bulk level in stem conditions, are likely originating from spontaneously differentiating cells. The up-regulation of KDM1B throughout differentiation is reflected in the immunofluorescence images as the proportion of KDM1B$^+$ cells increase, in line with the advancing differentiation of trophoblast. The variation in the staining of KDM1B at 5D, may represent cells at different stages of differentiation, or different trophoblast cell subtypes. Overall, however, my data show that KDM1B becomes ubiquitously expressed by differentiated trophoblast cells *in vitro*.

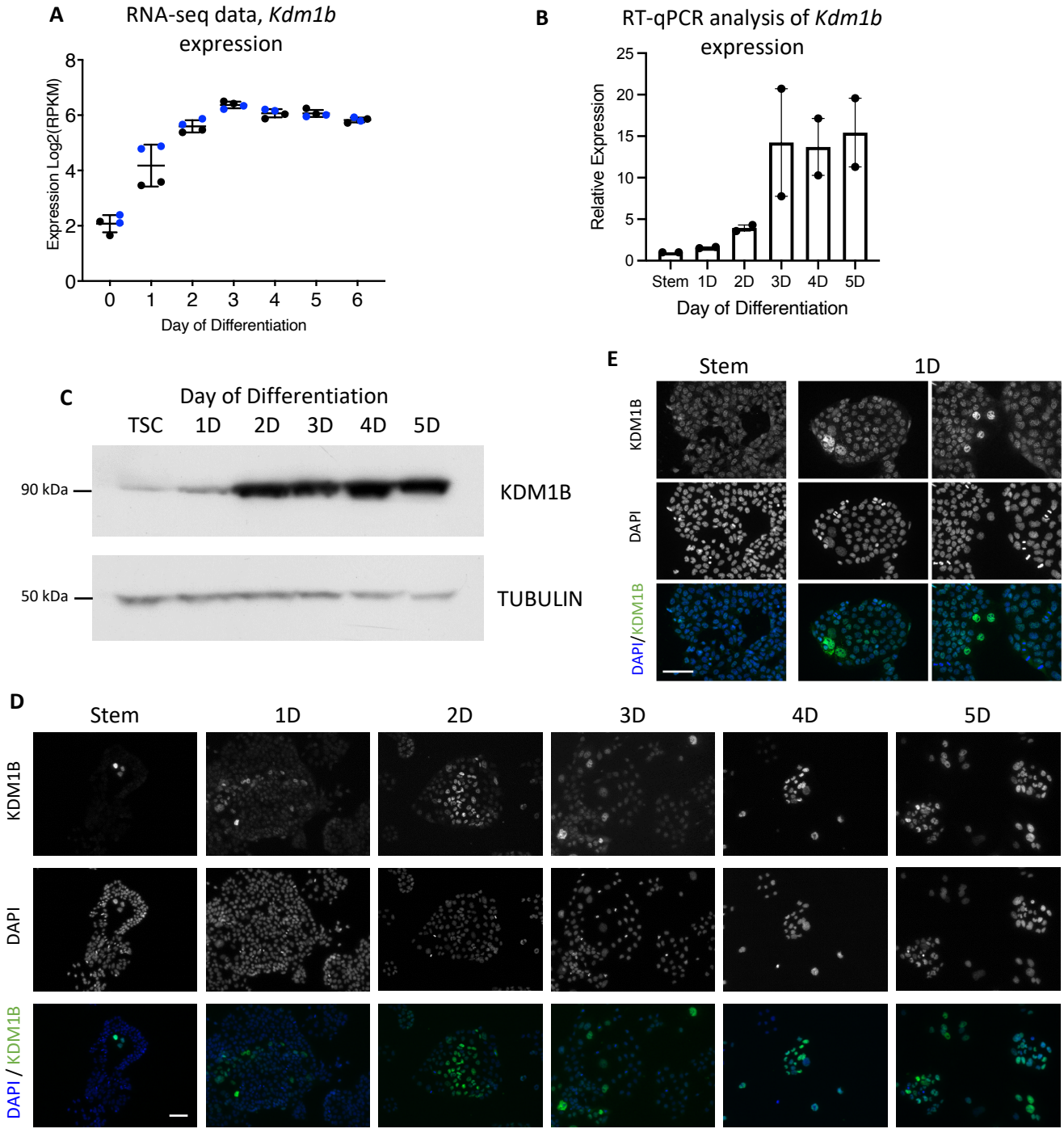



**Figure 3.1 Expression of *Kdm1b* throughout in vitro differentation**

**A**. Expression of *Kdm1b* during *in vitro* differentiation in TS-Rs26 (black) and TS-EGFP (blue) cells (Expression is reported as Log2(RPM)) by RNA-seq analysis from Dominika. **B** RT-qPCR analysis of *Kdm1b* transcription relative to *Sdha* in stem to 5D differentiated cells. **C.i** Western blot analysis showed an increase in KDM1B throughout differentiation *in vitro*, quantified relative to loading control, TUBULIN in **C.ii**. **D.** Immunofluorescence showed an increase in the proportion of KDM1B$^+$ cells (green) throughout *in vitro* differentiation, peaking at 5D where all cells appeared to be KDM1B$^+$ to some extent. DNA was stained with DAPI (blue). Scale bar is 200 µm. **E.** Immunofluorescence of stem and 1D differentiated cells showing large KDM1B$^+$ cells. Scale bar is 100 µm.

### 3.2.2 Generation of *Kdm1b* knockout clones

In order to test the hypothesis that a factor whose expression was induced at the onset of differentiation would have a role to play in the maintenance of the stem cell state, or to affect differentiation trajectories or speed of differentiation, I generated CRISPR Cas-9-mediated knockout TSC clones of *Kdm1b*. In total, six gRNAs were designed: two in each of introns 4, 5 and 10. These were designed such that when paired, exon 5 or exons 5-10 would be excised from the *Kdm1b* gene (Figure 3.2a). These cleavage sites were designed to induce a frameshift mutation and thus introduction of a premature stop codon into the remainder of the *Kdm1b* open reading frame; the deletion of exon 5 ablates two protein-coding isoforms and that of exons 5-10 deletion produces a full knockout of all three *Kdm1b* isoforms.

An initial first-pass test was performed in mouse ESCs, as they are easier to transfect than TSCs, in order to test whether the gRNAs successfully induced cutting of the *Kdm1b* gene. Intron 4 gRNAa did not induce cutting in the bulk samples, and hence was discarded for experimental use; as such four CRISPR test groups were performed, each using intron 4 gRNAb combined with one of the gRNAs in introns 5 or 10. These groups were designated C1, C2, C3 and C4 (Table 2.4.3). Control samples were transfected with the CRISPR Cas9-EGFP plasmid, without gRNAs.

Twenty-four hours after transfection, cells expressing high amounts of GFP (and therefore, Cas9) were single cell sorted into 96 well plates and expanded. Approximately 6 % of clones retrieved were confirmed knockouts. Intriguingly, these all originated from the C3 and C4 groups in which *Kdm1b* was cleaved in intron 4 and intron 10 and no mutant clones of the smaller deletion allele were obtained. Two diagnostic PCRs were performed to identify these *Kdm1b*$^{-/-}$ clonal TSC lines (Figure 3.2b). PCR primer pair 1 amplified a region of exon 4, with which a PCR product would be expected in WT clones and no product would be expected in *Kdm1b*$^{-/-}$ cells. PCR primer pair 2 flanked exons 5 and 10, therefore amplifying across the deleted region in knockouts but giving no product in control cells.

Figure 3.2c illustrates the results of these two diagnostic PCRs. The control clones in lanes 1-5 are verified WT for *Kdm1b*, having a PCR product of the predicted size 760 bp for primer pair 1 and no PCR product for primer pair 2 (Figure 3.2c.i and ii, respectively). All potential knockouts show loss of exon 5, having no PCR product with primer pair 1 (Figure 3.2c.i). Clones B6, C9 and E12 produced a PCR product of the predicted size, 440 bp, with primer pair 2, indicating successful cleavage and repair in the expected manner, of *Kdm1b*. Clones E9 and G6, however, showed no PCR product with either primer pair. Instead, these latter clones likely have larger deletions of the targeted region that also erases the primer binding sites. Western blot analysis provided the ultimate proof that all five clones were null for KDM1B (Figure 3.2d). Thus, clones C9, E9, E12 and G6 from CRISPR test group C3 and clone B6 from group C4 as well as control clones, A6, B4, D4, F2 and G7, were taken for further analysis.

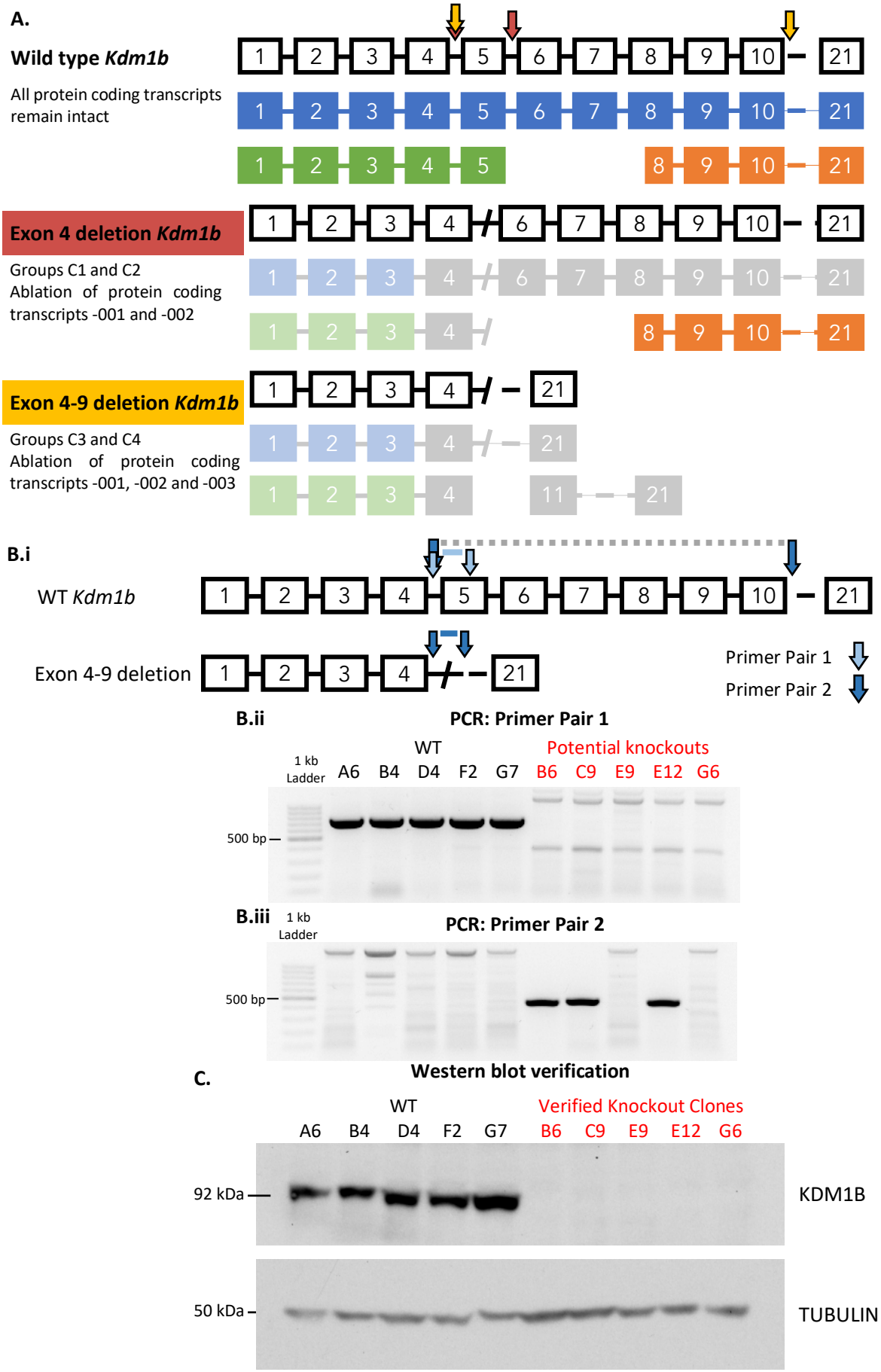

A.
Wild type Kdm1b
All protein coding transcripts remain intact
Exon 4 deletion Kdm1b
Groups C1 and C2
Ablation of protein coding transcripts -001 and -002
Exon 4-9 deletion Kdm1b
Groups C3 and C4
Ablation of protein coding transcripts -001, -002 and -003
B.i
WT Kdm1b
Exon 4-9 deletion
Primer Pair 1
Primer Pair 2
B.ii
PCR: Primer Pair 1
WT
Potential knockouts
1 kb Ladder
A6 B4 D4 F2 G7 B6 C9 E9 E12 G6
500 bp
B.iii
1 kb Ladder
PCR: Primer Pair 2
500 bp
C.
Western blot verification
WT
Verified Knockout Clones
A6 B4 D4 F2 G7 B6 C9 E9 E12 G6
92 kDa
KDM1B
50 kDa
TUBULIN

**Figure 3.2 Generation of *Kdm1b*$^{-/-}$ clones by CRISPR Cas9-mediated cleavage and removal of exon(s)**

**A**. Schematic representation of the *Kdm1b* gene, illustrating the two knockouts attempted and the protein-coding transcripts each deletion would ablate. **B.i** Schematic representation of PCR primer pairs 1 and 2 as they pertain to WT and C3/C4 null *Kdm1b*. **B.ii** PCR using Primer Pair 1 which flank exon 5, a product of 760 bp is expected and observed in all five WT clones. **B.iii** PCR with Primer Pair 2, which flank exons 5-10; a 440 bp PCR product is expected and observed in three of five KO clones. **C** Western blot analysis showing complete ablation of KDM1B in all five knockout clones. TUBULIN is used as a loading control.

### 3.2.3 Effect of loss of KDM1B on the expression of trophoblast lineage markers

In order to assess how loss of KDM1B affects differentiation of TSCs cells *in vitro*, I performed RT-qPCR analysis for several stem cell and differentiation marker genes in TSCs and after 1D-3D of differentiation (Figure 3.3a). Given that *Kdm1b* transcription was rapidly induced at the onset of differentiation (Figures 1.4, 3.1), i.e. withdrawal of FGF4 and TGFβ from the culture medium, I collected RNA from stem cells and from the first three days of *in vitro* differentiation, termed 1D, 2D and 3D. The hypothesis being that if loss of a rapidly induced differentiation marker is important for maintaining the stem state or in directing differentiation potential, it was likely to have a fast-acting effect on marker gene expression.

Stem cell markers such as *Cdx2* (not significant), *Eomes, Elf5, Il17rd* and *Egr1* were expressed at significantly higher levels in *Kdm1b*$^{-/-}$ TSCs in the stem cell state compared to their WT control counterparts. Upon induction of differentiation, these genes were downregulated as expected. Interestingly, the only stem cell marker gene to show the opposite pattern was *Esrrb*, whose mean expression in *Kdm1b*$^{-/-}$ clones was 0.62 times that of WT vector-only controls (Figure 3.3a.i).

Further differences in the behaviour of *Kdm1b*$^{-/-}$ TSCs were observed upon differentiation. As such the early syncytiotrophoblast (SynT) marker, *Gcm1*, was significantly upregulated in 1D differentiated cells, by 2.7 times in *Kdm1b*$^{-/-}$ cells compared to WT. The levels of transcription of *Gcm1* tended to remain increased throughout the time course albeit not to significance (Figure 3.3b.i). Two markers of terminally differentiated SynT cells, *Syna* and *Synb*, however, showed no significant difference between WT and *Kdm1b*$^{-/-}$ clones, up to 3D differentiation (Figure 3.3b.ii, .iii). This could indicate that *Kdm1b*$^{-/-}$ TSCs stall in their SynT differentiation trajectory at the very early stages, or the detection of expression differences of late SynT markers may only become obvious beyond 3D of differentiation. Markers of trophoblast giant cell (TGC) differentiation, *Tpbpa*, *Ascl2*, *Pl2* and *Plf* did not show significant expression differences in *Kdm1b*$^{-/-}$ cells compared to WT controls (Figure 3.3c).

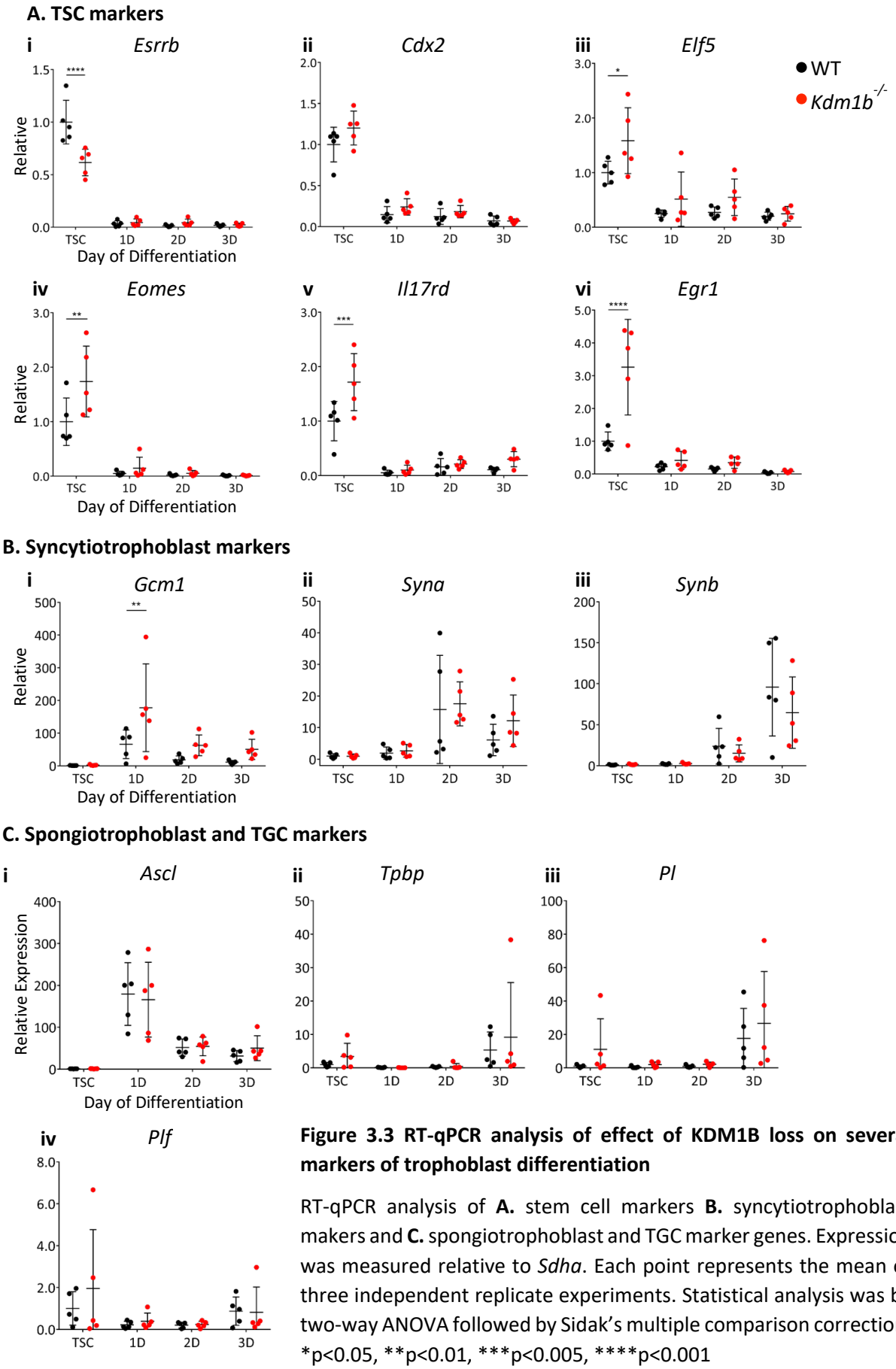



**Figure 3.3 RT-qPCR analysis of effect of KDM1B loss on several markers of trophoblast differentiation**

RT-qPCR analysis of **A.** stem cell markers **B.** syncytiotrophoblast makers and **C.** spongiotrophoblast and TGC marker genes. Expression was measured relative to *Sdha*. Each point represents the mean of three independent replicate experiments. Statistical analysis was by two-way ANOVA followed by Sidak's multiple comparison correction. *p<0.05, **p<0.01, ***p<0.005, ****p<0.001

The fact that *Kdm1b*$^{-/-}$ clones appeared to have lower *Esrrb* expression compared to WT controls when all other stem cell markers tested showed increased expression to some extent, is intriguing and may be related to the known interaction between KDM1B and ESRRB. Previous studies of the KDM1B family member, LSD1, in trophoblast differentiation shows that loss of *Lsd1* results in reduction in the barrier to trophoblast differentiation, with LSD1 interacting directly with ESRRB (Zhu *et al.*, 2014; Latos *et al.*, 2015). When tested by RT-qPCR, *Lsd1* appeared to be more highly transcribed in *Kdm1b*$^{-/-}$ cells (Figure 3.4a) both in stem cells and at 1D differentiation. Stem KDM1B null cells had relative mean expression of *Lsd1* 4.3 times higher than WT stem cells and at 1D, 2.3 times that of WT cells, at 4.8 compared to 1.7. *Lsd1* is highly expressed in TSCs, with Log2(RPKM) of 4.7, with expression maintained during differentiation (data not shown). Further, upon western blot analysis of stem cells, the five knockout clones showed increased LSD1 protein relative to TUBULIN, compared to WT clones (Figure 3.4b). All three isoforms bound by LSD1-Ab were used in the quantitation.

This could point to a compensatory upregulation of LSD1 in *Kdm1b*$^{-/-}$ cells. It is interesting that the increase in LSD1 expression was observed in stem cells, given how few cells appear to be KDM1B positive in stem cell conditions. It is possible that the increase in LSD1 might be the driver for the apparent decrease in *Esrrb* transcript levels in stem *Kdm1b*$^{-/-}$ cells. I did not investigate protein levels at later time points of differentiation, which would be interesting to probe, to assess if this increase in LSD1 persists or even increases as more cells become KDM1B$^{+}$.

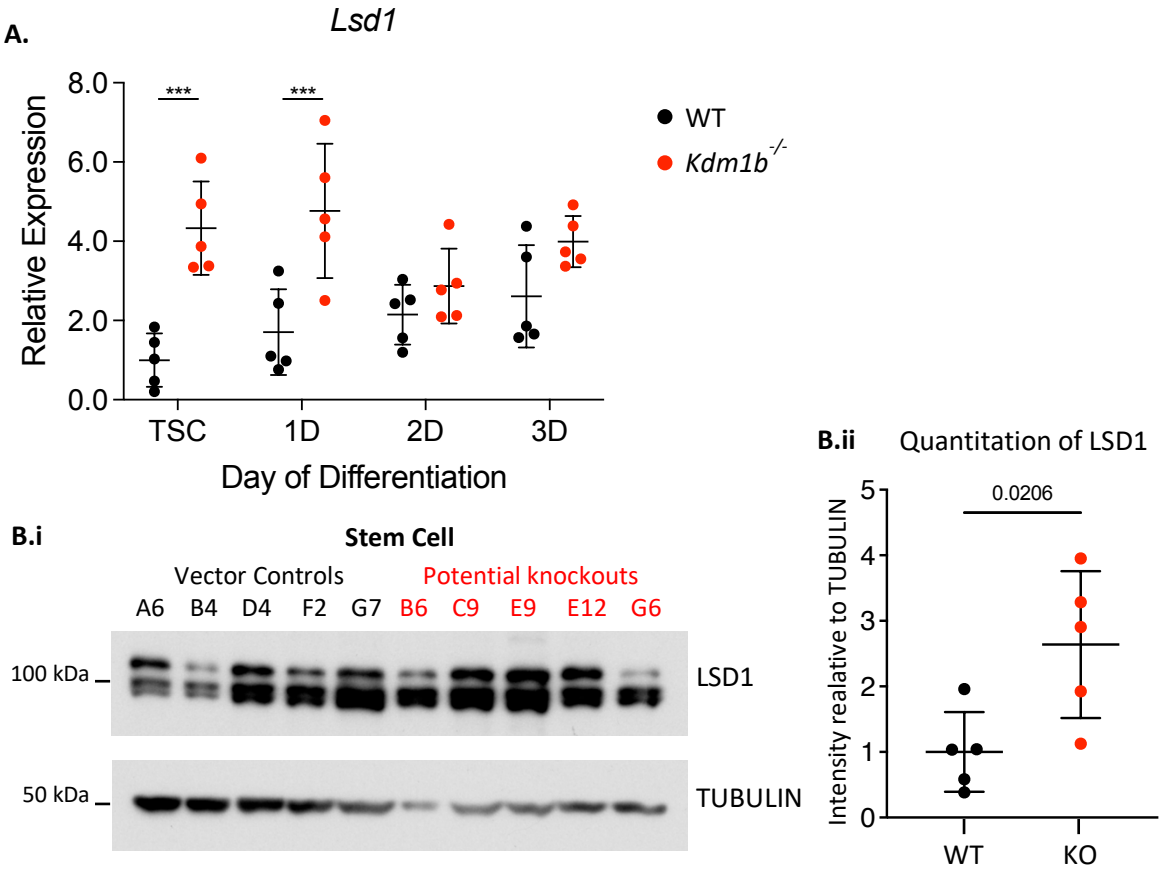


**Figure 3.4 Loss of KDM1B leads to an increase in LSD1 expression**

RT-qPCR analsysis of *Lsd1* transcription throughout a 4-day time course of TSC differentiation *in vitro.* Expression is measured relative to *Sdha* and each point represents the mean of three independent experiments. Statistical analysis was by two-way ANOVA followed by Sidak's multiple comparison correction. *p<0.05, **p<0.01, ***p<0.005. **B.i** Western blot analysis of LSD1 both WT and *Kdm1b*$^{-/-}$ stem cells with TUBULIN as a loading control. **B.ii** Quantification of LSD1 (all three isoforms detected) intensity measured relative to TUBULIN. Statistical analysis was by two-tailed Welch's t-test, p-value is quoted.

### 3.2.4 Changes to histone modifications with loss of KDM1B

Western blot analysis for several histone modifications was performed in stem and 2D differentiated cells to assess whether loss of KDM1B affected the global levels of these modifications. The chromatin marks tested were H3K4me3/2/1 (Figure 3.5a), H3K9me3/2 (Figure 3.5b) and H3K36me3 (Figure 3.5c). Testing of H3K4 methylation status was performed to assess if changes to the direct substrates of KDM1B, H3K4me1 and H3K4me2, could be detected on a bulk level, and if there would be any subsequent changes to H3K4me3. H3K9 methylation changes were assessed in order to ascertain whether KDM1B targets this substrate in TS-Rs26 cells. Finally, H3K36me3 was assessed to see if loss of KDM1B from complex with elongating RNA PolII (RNAP2), had any impact on the bulk levels of this mark *in vitro*.

In stem cell conditions, there was no difference in the relative amount of H3K4me1 in KDM1B null cells compared to WT (Figure 3.5a.i). At 2D differentiation, however, loss of KDM1B tended to lead to a modest increase in the relative amount of H3K4me1, with a mean intensity approximately 25 % higher in knockouts compared to WT cells (Figure 3.5a.ii). This increase was not statistically significant ($p$=0.0504), however an increase to this extent on the bulk level, is certainly biologically relevant and occurs in the expected direction: with loss of a demethylase preceding increase of its substrate mark. It would certainly be important to identify in which genomic loci the accumulation of H3K4me1 is occurring, if indeed it is not global. This will be addressed in Chapter 4.

Conversely, there were no changes in global levels of H3K4me3 or H3K4me2 relative to total H3 in KDM1B null cells in stem or 2D differentiated cells (Figure 3.5b and c). Perhaps unsurprisingly, there were also no significant changes to levels of H3K9me2 or H3K9me3 (Figure 3.5d and 3.5e, respectively) relative to total H3 in stem or 2D differentiated cells upon loss of KDM1B.

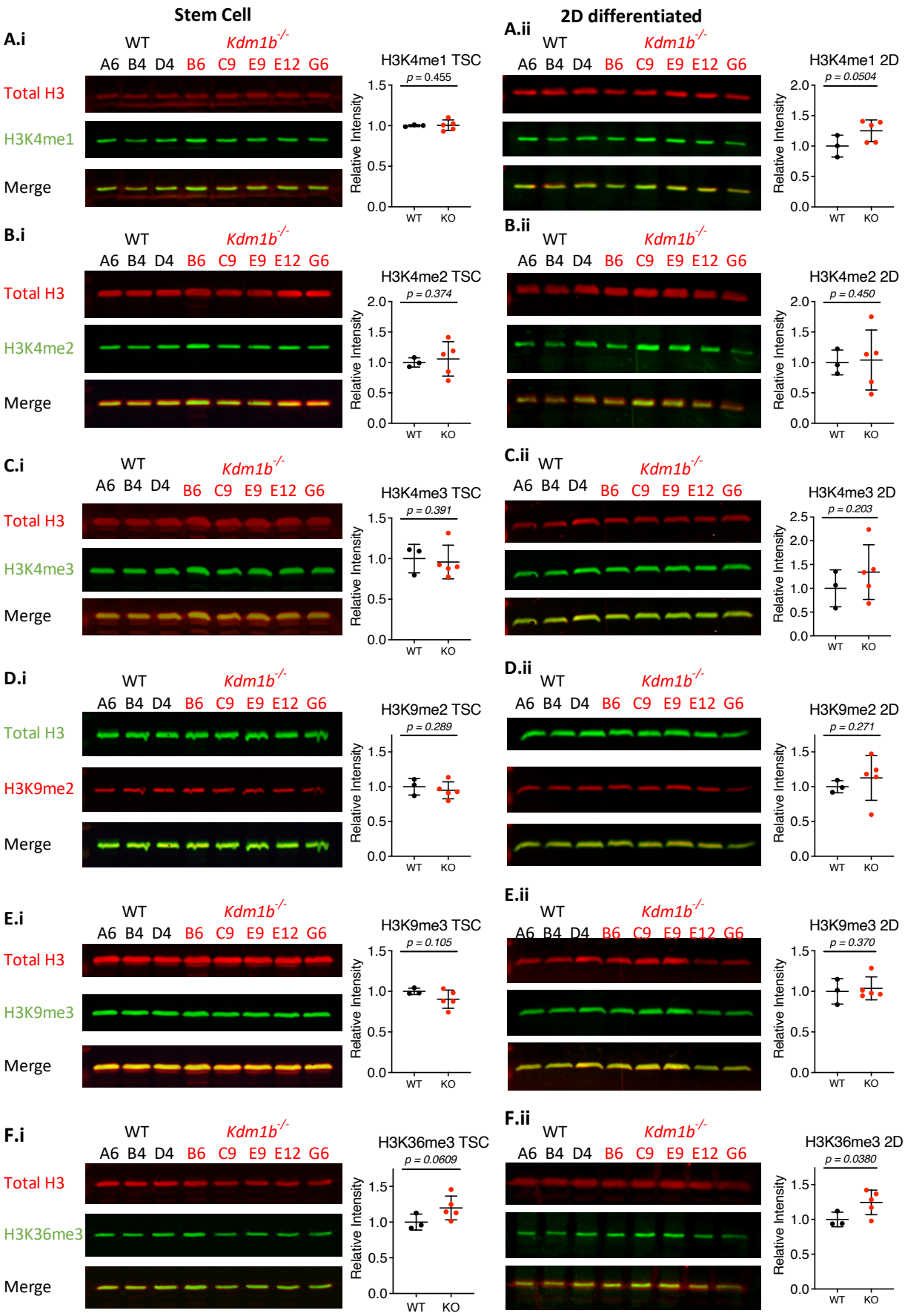

Stem Cell
2D differentiated
A.i
A.ii
WT
Kdm1b-/-
A6 B4 D4 B6 C9 E9 E12 G6
Total H3
H3K4me1
Merge
H3K4me1 TSC
p = 0.455
H3K4me1 2D
p = 0.0504
B.i
B.ii
H3K4me2
H3K4me2 TSC
p = 0.374
H3K4me2 2D
p = 0.450
C.i
C.ii
H3K4me3
H3K4me3 TSC
p = 0.391
H3K4me3 2D
p = 0.203
D.i
D.ii
H3K9me2
H3K9me2 TSC
p = 0.289
H3K9me2 2D
p = 0.271
E.i
E.ii
H3K9me3
H3K9me3 TSC
p = 0.105
H3K9me3 2D
p = 0.370
F.i
F.ii
H3K36me3
H3K36me3 TSC
p = 0.0609
H3K36me3 2D
p = 0.0380
Relative Intensity
WT
KO

**Figure 3.5 Histone western blots assessing how loss of KDM1B effects bulk levels of certain histone modifications**

Bulk levels of **A**. H3K4me1, **B**. H3K4me2, **C**. H3K4me3, **D**. H3K9me2, **E**. H3K9me3 and **F.** H3K36me3 in stem (**i**) and 2D differentiated cells (**ii**). Total H3 was probed in the red channel and each histone modification in the green channel for all blots except H3K9me2 in which these were reversed. Merge shows both red and green channels. Plotted are the intensities of each histone modification relative to each corresponding total H3 signal and then normalized to WT samples. Statistical analysis was by one-tailed t-test, *p* values are quoted.

Finally, H3K36me3 was significantly increased relative to total H3 in 2D differentiated *Kdm1b*$^{-/-}$ cells, compared to WT by about 24 %. One theory as to the role of KDM1B in the complex with elongating RNA PolII is that the complex is erasing 'open' chromatin marks such as H3K4 methylation and writing 'closed' marks such as H3K9 methylation, preventing spurious transcriptional initiation within the gene body rather than at the promoter. The increase in H3K36me3 seen by western blot may indicate that spurious transcription in the *Kdm1b* knockout cells was occurring, or that a more general increase in transcription was occurring.

Another side-effect of the loss of KDM1B from the elongating RNA PolII complex could be a pausing proximal to the promoter, leading to an increase in the relative amount of initiating (phospho-Ser5) RNA PolII relative to elongating RNA PolII (phospho-Ser2). In order to assess whether there were any changes to the phosphorylation status of RNAP2 in KDM1B null cells, I performed western blot analysis probing for total RNA PolII, phospho-Ser5 RNA PolII and phospho-Ser2 RNA PolII in stem and 2D differentiated cells (Figure 3.6a and 3.6b, respectively).

Overall, there were no significant changes to the relative proportion of elongating or initiating RNA PolII upon loss of KDM1B in stem and 2D cells. However, it appears that the amount of RNA PolII being probed, varied between cells. It is currently unclear what the biological significance of such varying RNA PolII levels might be, and indeed whether this is a biologically relevant observation or a technical failure of the protein extraction. However, since no overt differences were observed, this point was not further pursued.

### 3.2.5 There is wide variation in the proliferation rate of TSCs

While culturing the cells, an observational finding was that there was large variation in the proliferation rates of the clones, with some clones requiring more frequent passaging than others. In order to ascertain whether TSC proliferation rates were impacted by loss of KDM1B, I performed a cell viability assay (PrestoBlue, Lall *et al.*, 2013; Figure 3.7a), and a straightforward cell counting assay (Figure 3.7b).

I also analysed the cell cycle progression of the samples throughout differentiation by propidium iodide staining (Figure 3.7c).

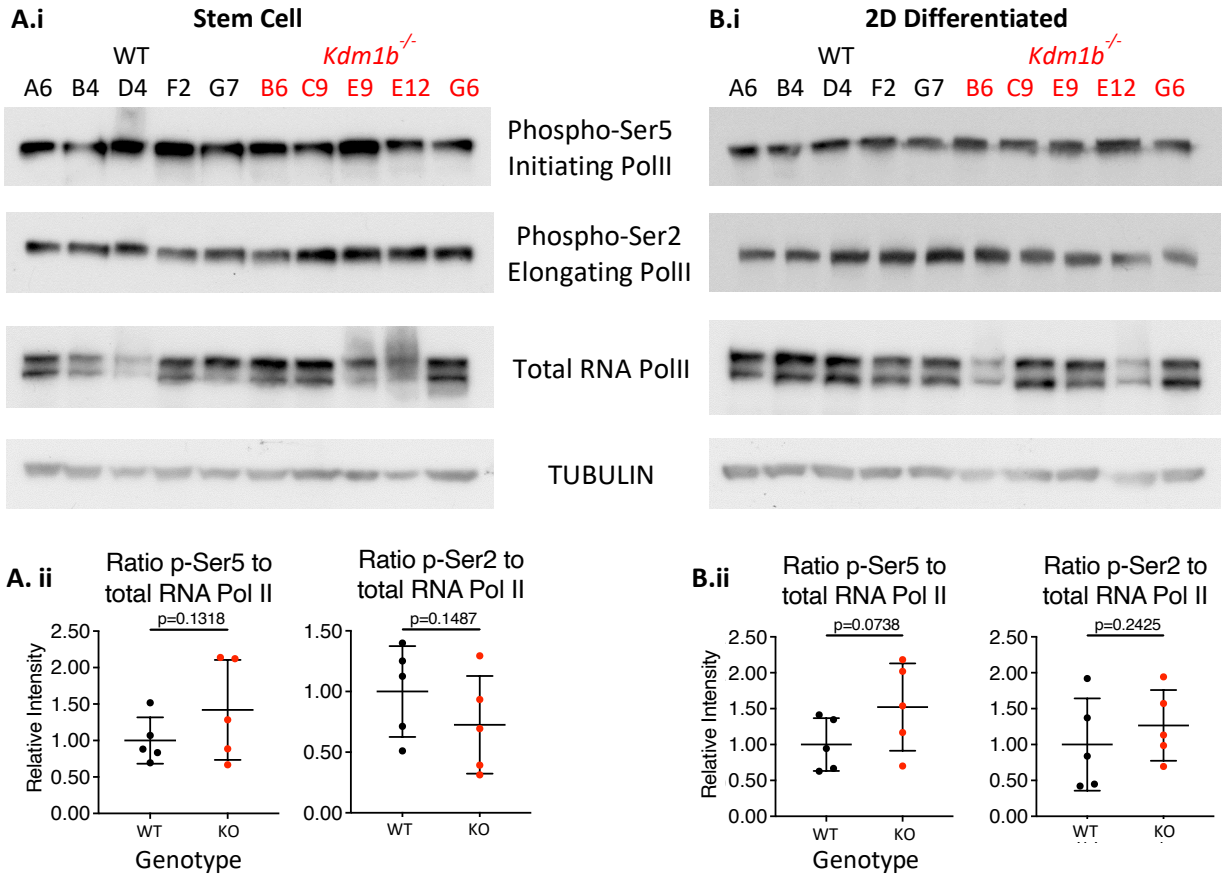



**Figure 3.6 Western blot analysis of different phosphorylation states of RNA PolII in stem and 2D differentiated cells**

Bulk levels of initiating, elongating and total RNA PolII were probed in **A**. stem and **B**. 2D differentiated cells. A representative blot for TUBULIN is shown for stem and 2D differentiated samples. The relative intensity for each RNA PolII state was quantitated relative to its corresponding TUBULIN. The ratio of phosphorylated to total RNA PolII was then calculated and these, normalized to WT, are plotted in **A.ii** and **B.ii** for stem and 2D samples, respectively. Statistical analysis was by Welch's two-tailed t-test, p-values are quoted.

Figure 3.7a plots the cellular proliferation via oxidative capacity in stem cell (i) and differentiating (ii) conditions. There was large variation in the cell numbers during proliferation, with maximal variation reached at day 2 in both stem cell and differentiation conditions. At day 2, the median cell number in WT cells was around 470,000 compared to 270,000 cells *Kdm1b*$^{-/-}$ cells. The issue with this methodology is that TS-Rs26 cells proliferate very quickly; they reached 95 % confluency at day 2 in stem cell conditions. Indeed, as evidenced by both the stem and differentiating conditions, the cells appeared to exit the log phase of growth between day 2 and day 3.

Bearing the potential issue of overgrowth in mind, I also performed a straight-forward cell counting assay, the results of which are plotted in Figure 3.7b. This plot illustrates that the log phase of growth was maintained to at least 3 days, unlike with the oxidative assay, likely because the cells were plated at a lower confluency in a larger dish. The WT clones showed the broader range in proliferation rate with this assay too, peaking at 3 days with a median of 1.1 million cells. There was no significant difference in the cell numbers at any point in the proliferation assay. I hypothesise there is potential that if left over a longer period of time, the WT would overtake the *Kdm1b*$^{-/-}$ cells, but with the short assays used here, this is not conclusive.

Further to the proliferation rate, I decided to assess the effect of loss of KDM1B on progression through the cell cycle (Figure 3.7c) in stem, 2D and 5D differentiated cells. When comparing WT to *Kdm1b*$^{-/-}$ cells at the same day of differentiation for each cell cycle stage, there were no statistically significant differences.

In stem cell conditions, most of the cells were designated as S-phase with a mean of 70 % for WT and KO clones measured. This proportion decreased throughout differentiation indicating a loss of replicative capacity or slowing of the cell cycle as the cells terminally differentiate. The relative number of cells in G1 and in G2 increased as the cells differentiated *in vitro*. Between stem and 2D differentiated cells, the percentage in G1 increased from 17 % to 32 % in WT clones and from 17 % to 38 % in KDM1B null cells. For both WT clones and *Kdm1b*$^{-/-}$ clones, this decreased at 5D differentiation to 24 % in and 33 %, respectively.

Regarding the proportion of cells in G2, a large increase was seen between 2D and 5D of differentiation in both genotypes, increasing from 12 % to 34 % in WT clones and from 12 % to 25 % in KO clones. For both genotypes the proportion of cells with genomic content >4n, designated >G2, increased starkly at 5D differentiation. This increase was greater in the knockout clones, due to one WT clone having markedly fewer >4n cells at just 13 % (Figure 3.7c.ii).

The increase in the proportion of cells designated >G2 during differentiation, is likely caused by the acquisition of the TGC fate. The other predominant cell fate acquired during *in vitro* trophoblast differentiation, SynT cells, would not be identified as >G2 if remaining intact, rather as doublets (or higher) as these cells are multinucleated. TGCs however, undergo many rounds of endoreduplication becoming >4n as they mature.

Overall, despite a great amount of variability between assay types and individual clones, I conclude that there is no significant difference in proliferation rates between WT and *Kdm1b* TSCs.

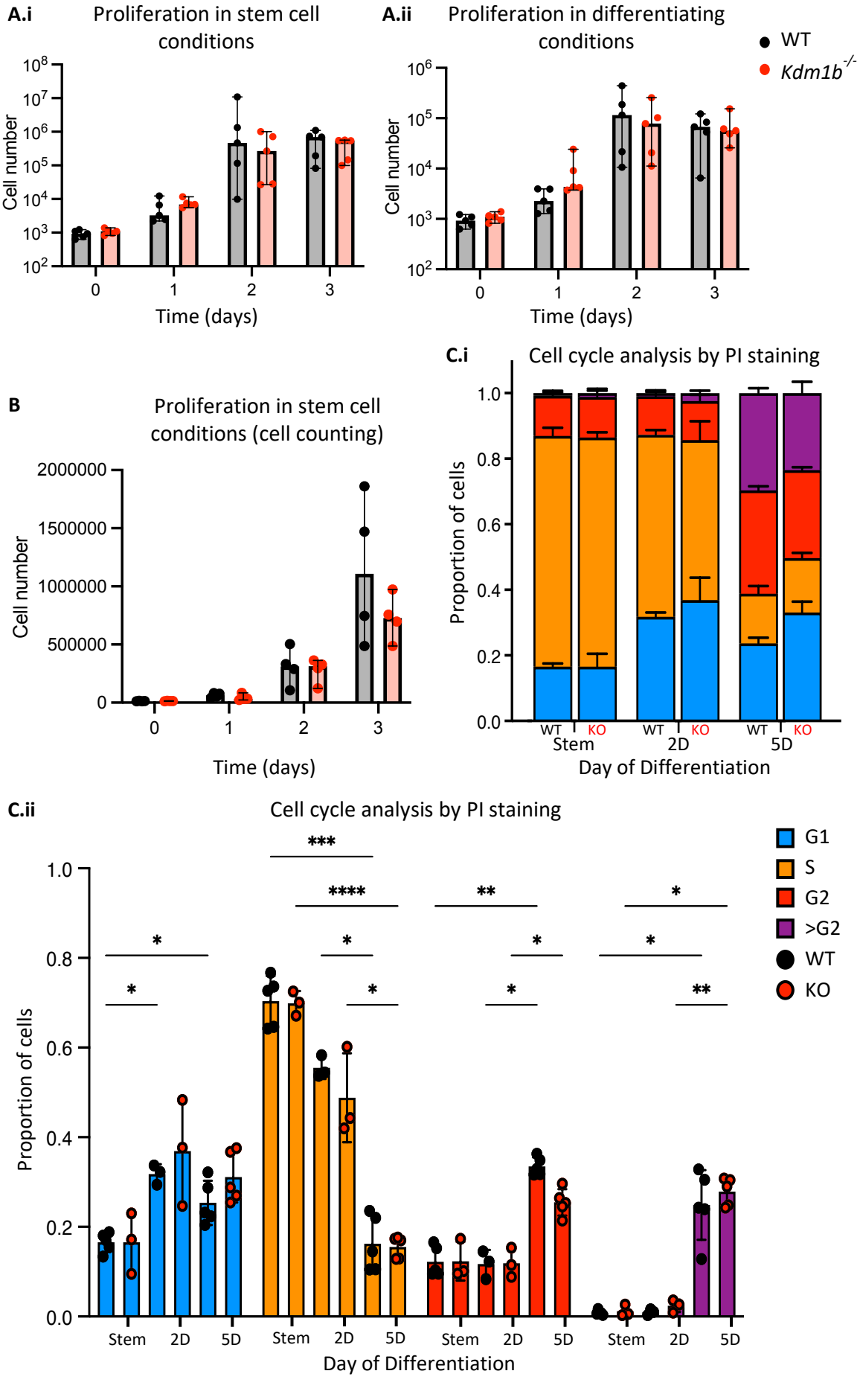

A.i
Proliferation in stem cell conditions
Cell number
Time (days)
A.ii
Proliferation in differentiating conditions
WT
Kdm1b-/-
B
Proliferation in stem cell conditions (cell counting)
C.i
Cell cycle analysis by PI staining
Proportion of cells
Stem
2D
5D
Day of Differentiation
C.ii
G1
S
G2
>G2
WT
KO

**Figure 3.7 Assessing how loss of KDM1B impacts proliferation during *in vitro* culture**

Cellular proliferation was measured by PrestoBlue assay (**A**) in stem (**i**) and differentiating (**ii**) conditions, as well as by directly counting the cells (**B**). Points represent a single clone, bars plotting the median and range. Statistical analysis was by two-way ANOVA followed by Sidak's multiple comparison correction; no significant differences were identified. Progression through the cell cycle was analysed by propidium iodide staining (**C**). (**C.i**) Mean proportion of cells recorded in each cycle stage: G1 (blue), S (orange) and G2 (red) as well as cells designated >G2 (purple) are plotted, with WT and *Kdm1b*$^{-/-}$ for each time point of differentiation measured, error bars are SEM. (**C.ii**) Mean proportion of cells recorded is plotted (bar) for each cell cycle stage, points represent one clone and error bars are SD. Statistical analysis was by mixed-effects analysis followed by Tukey's multiple comparison correction, *p<0.05, **p<0.01, ***p<0.005, ****p<0.001. Only pairwise comparisons that were significant and **within** genotypes are plotted, for simplicity.

## 3.3 Discussion

This chapter has illustrated the expression dynamics of *Kdm1b*, an H3K4me1/2-specific demethylase that was chosen for analysis for its role in TSC self-renewal and differentiation because it was the earliest epigenetic modifier to be considerably upregulated upon onset of TSC differentiation. I corroborated the initial RNA-seq expression data by detailed RT-qPCR and western blot analyses. I then generated *Kdm1b* knockout TSCs and performed an initial wide-spectrum assessment of these cells for potential differences in self-renewal and differentiation behaviour, for global differences in the amounts of relevant histone modifications, and for cellular proliferation rates.

Loss of KDM1B had a conflicting effect on transcription of stem marker genes. I found that *Kdm1b*$^{-/-}$ TSCs exhibit elevated expression levels of multiple stem cell markers, such as *Eomes* and *Elf5*, which could point an enrichment of the stem state. Conversely, the highly sensitive TSC marker, *Esrrb*, was downregulated. *Esrrb* is a marker of proliferative trophoblast stem cells and is one of the most rapidly down-regulated TSC markers (Latos *et al.*, 2015), it's downregulation could point to a reduction in the barrier to differentiation of TSCs upon loss of KDM1B. These seemingly conflicting data could point to a compensatory mechanism whereby expression of other stem markers are induced in response to reduced *Esrrb* expression, potentially to reinforce the TSC TF network. This would be counter to previous reports in which loss of *Esrrb* induced reduced expression of other stem markers, *Elf5*, *Eomes*, *Cdx2* etc (Latos *et al.*, 2015).

Downregulation of *Esrrb* over other TSC marker genes, is potentially significant, pointing to KDM1B playing an activating role in *Esrrb* transcription, possibly via binding to RNAP2 or modulation of trans-acting intragenic enhancers. The simultaneous up-regulation of *Kdm1b* family member, LSD1, in KDM1B null cells, confounds the ability to draw firm conclusions as to the mechanism and

consequences of reduced *Esrrb* expression. ESRRB directly interacts with LSD1 (Latos *et al.*, 2015). Both in stem and differentiating *Kdm1b*$^{-/-}$ cells showed increased LSD1 expression, thus it cannot be ruled out that the changes to *Esrrb* transcription might be a result of increased LSD1 rather than loss of KDM1B. In order to test this, and to rule out any redundancy between KDM1B and LSD1 in trophoblast cells, it would be beneficial to inhibit or knockout LSD1 in *Kdm1b*$^{-/-}$ cells.

Further, the early syncytiotrophoblast marker, *Gcm1* was expressed at higher levels at the onset of differentiation without subsequent changes to later markers of SynT differentiation. Finally, turning to the commitment of TS-Rs26 cells to the TGC lineage, no markers of these cell types were altered upon loss of KDM1B. In combination these data show that loss of *Kdm1b* induces transcriptional changes that do not just simply align with reduced, increased or biased levels of differentiation. Instead, they point to the possibility that KDM1B has important roles in the fine-tuning of transcription and with the timely modulation (up- or downregulation) of gene activity, in line with KDM1B being a component of the RNA PolIII complex. Also, these results are not dissimilar from the subtle changes to stem marker expression and ~2-fold induction of *Gcm1* observed with *Lsd1*$^{-/-}$ TSCs, and may point to an auxiliary role of KDM1B in enhancer decommissioning (Whyte *et al.*, 2012; Zhu *et al.*, 2014).

Comparing the in vitro differentiation of *Kdm1b*$^{-/-}$ TS-Rs26 cells with in vivo data from *Kdm1b*$^{-/-}$ placentae would be invaluable. As previously stated, *Kdm1b*$^{-/-}$ mice are grossly normal and can become pregnant (D N Ciccone *et al.*, 2009). There is a maternal effect of loss of KDM1B from oocytes, and heterozygous embryos from *Kdm1b*$^{-/-}$ mothers, arrest and die at E10.5. The only reference to the trophoblast of these animals was that *Kdm1b*$^{+/-mat}$ embryos had “severe placental defects” (Ciccone *et al.*, 2009). More in-depth characterisation of these placentae including transcriptional assays, histological staining and gross structural measurements would be invaluable in ascertaining what role KDM1B plays in the placenta. Also vital, would be KDM1B staining of WT placentae at various gestational time points, such as E9, E11 and E18.5 to ascertain which trophoblast cell types and placental compartments KDM1B plays a role in developing *in vivo*.

Contrastingly to the case of reduced *Esrrb* expression, the increase in expression of *Gmc1*, is unlikely to be attributable to the increase in LSD1. Zhu et al (2014) demonstrated that loss of LSD1 induces over-expression of *Gcm1*, thus an increase in LSD1 would be unlikely to have the same effect. It is interesting that both LSD1 and KDM1B have now been shown to induce expression of *Gcm1* during *in vitro* TSC differentiation. This could point to a key regulatory region, possibly an enhancer, being modulated by both proteins. One way of testing this possibility would be performing ChIP-qPCR of the *Gcm1* gene and surrounding region, targeting both KDM1B and LSD1 proteins, during differentiation.

As to my assessment of global epigenetic changes induced by loss of KDM1B, I observed that during differentiation of *Kdm1b*$^{-/-}$ TSCs the amount of H3K4me1 tended to increase ($p$=0.05), pointing to an accumulation of the mark during differentiation specifically in the *Kdm1b*$^{-/-}$ cells compared to wild type. Since KDM1B is an H3K4me1/2 demethylase, this is an expected finding, which highlights that the enzymatic activity of KDM1B has significant impact on the trophoblast's chromatin state. The fact that the tendency to accumulate H3K4me1 was evident only upon 2D differentiation is also in line with the expression dynamics of *Kdm1b*/KDM1B, that increase profoundly as TSCs exit the stem cell state. Extending the histone western blot time course to a later point in differentiation, such as 5D, would be necessary to prove that loss of KDM1B leads to an increase in global H3K4me1 by accumulation of this mark.

Also observed was a similar increase in H3K36me3 at 2D differentiation. This, combined with no significant changes to the ratio of RNAP2 in the initiating state (phosphorylated Ser5) to the elongating state (phospho-Ser2) points to increased transcription in the *Kdm1b*$^{-/-}$ cells. One method to quantitate this would be straightforward measurement of the amount of RNA extracted from a defined number of TS-Rs26 cells of each clone. This however would be confounded by mRNA stability and does not indicate relative level of nascent transcripts produced. One such assay is nuclear run-on (NRO), which allows direct quantitation of nascent transcripts in nuclear isolates based on the incorporation of biochemically labelled nucleotides (Khraiwesh, 2011).

# 4 Chapter Four

## Effects of KDM1B ablation on trophoblast transcriptome and epigenome

## 4.1 Introduction

In chapter 3, I illustrated that loss of KDM1B leads to an increase in the expression of the early syncytiotrophoblast marker, *Gcm1*, at early stages of differentiation. Further, these cells also appeared to exhibit de-regulated expression levels of key trophoblast stem cell transcription factors, with *Esrrb* being down-regulated and others being up-regulated. In order to gauge the widespread changes to gene expression that occur as a result of loss of KDM1B, I performed RNA-seq analyses for each of the five WT and *Kdm1b*$^{-/-}$ clonal TSC lines, with samples taken from stem, 1D, 3D and 5D differentiated cells. Thus, transcriptomic analysis was performed across five biological replicates per genotype at each time point.

Additionally, *Kdm1b*$^{-/-}$ cells had increased amount of global H3K36me3 relative to WT, as assessed by western blot, in addition to H3K4me1 which although not statistically significant, represented a biologically relevant accumulation of this mark following loss of KDM1B. This is an indicator of widespread epigenetic changes that occur as a consequence of KDM1B ablation. Therefore, I decided to perform ChIP-seq experiments to ascertain locus-specific changes that occurred due to the loss of KDM1B. I performed ChIP-seq on stem, 3D and 5D differentiated cells, using four clones from each genotype, probing the histone modifications: H3K4me1, H3K4me2, H3K4me3 and H3K36me3.

## 4.2 Results

### 4.2.1 RNA-seq data quality control and quantification

Before differentially expressed genes could be identified between my WT and *Kdm1b*$^{-/-}$ clonal lines, I performed some quality control checks on the datasets using the SeqMonk software. Overall, all the datasets appeared to be of sound quality with approximately 95 % reads mapping to mRNA, of which almost all mapped to exons. There was little-to-no signal from rRNA or tRNA genes and the libraries were correctly identified as opposing strand-specific with very little signal on the sense strand. The size of the libraries did vary rather substantially, however, they were around the expected number of reads with a median library size of 17.5 million reads. There was no consistency in terms of the genotype of the largest and smallest libraries, so I felt it safe to down-sample the four largest libraries such that the majority of the libraries would be within 50 % of the largest. Therefore, knockout samples B6 stem, B6 1D, C9 stem and WT sample G7 3D were reimported, down-sampling with a target of 20 million reads. Upon re-plotting the RNA-seq QC plot, I saw that the distribution in library sizes now ranged from 100-50 % with only two libraries more than 2 times smaller than the largest sample: *Kdm1b*$^{-/-}$ clones, E9 1D and G6 5D (Figure 4.1a).

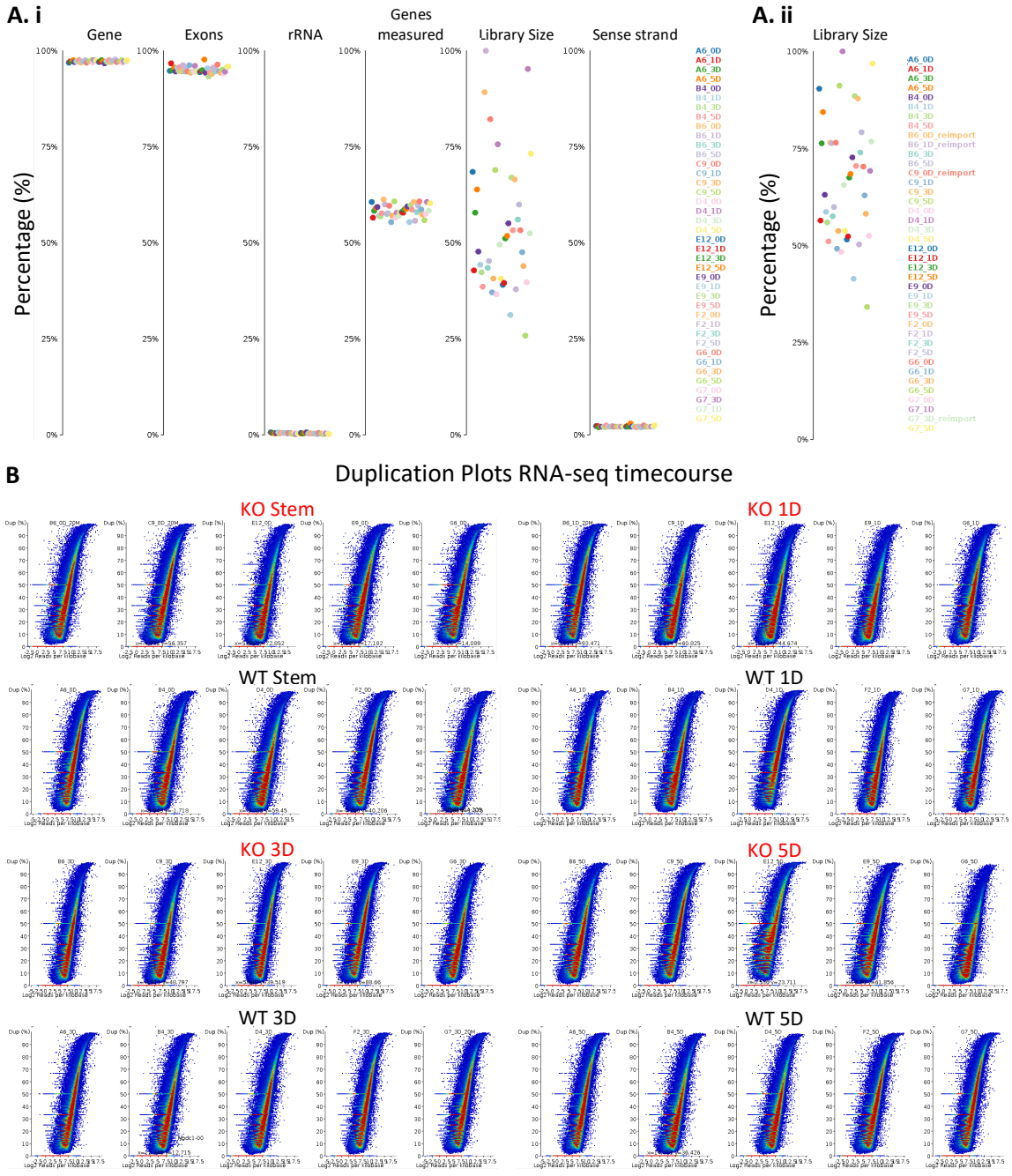



**Figure 4.1 Quality control of RNA-seq libraries**

**A.** RNA QC plot for RNA-seq libraires before (**i**) and after (**ii**) down-sampling libraries B6 stem, B6 1D, C9 stem and G7 3D to 20M reads. From left to right, plotting proportion of reads in genes, exons, and rRNA, percentage of genes with reads mapping, overall relative library size and proportion of reads mapping to the sense strand. These libraries are opposing-strand specific. The broad range of relative library size was adjusted by reimporting the four largest libraries at a read count of 20M reads. **B.** Duplication plots of all RNA-seq libraries show smooth distribution as regions of low read density have low duplication and high read density, high duplication.

There were no signs of increased duplication in any of the samples, with all duplication plots showing a continuous distribution with high density exons having higher levels of duplication and low-density exons, low levels of duplication (Figure 4.1b). Therefore, there was no need to account for duplication in the quantitation. Further, there was very little DNA contamination with samples ranging from 0.04 to <0.01 % DNA, therefore DNA contamination was also not used to normalise the quantitation of mRNA expression.

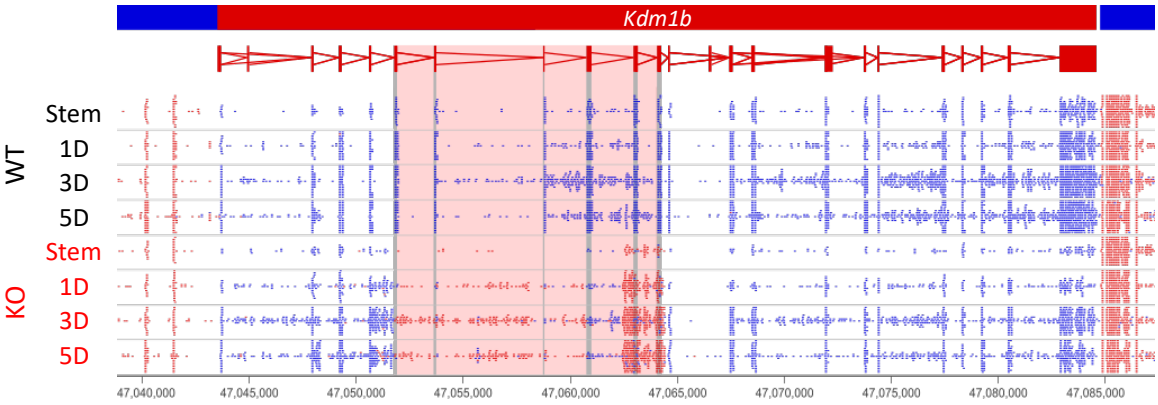
### 4.2.2 Loss of KDM1B had little effect on gene expression

My first port of call was look at the *Kdm1b* locus, to see how the excision of 5 exons affected transcription during differentiation. Figure 4.2a demonstrates that, as expected, transcription of the locus was reduced at all time points in KDM1B null clones. Figure 4.2b demonstrates that loss of KDM1B does not dramatically affect the expression of other lysine demethylases or histone methyltransferases. Indeed, only *Kdm1b* stands out as having striking differences in normalised expression values between WT and KDM1B null TSCs. It is interesting that the transcription of *Lsd1* (*Kdm1a*) was not at all impacted, according to the RNA-seq, although the gene showed distinct up-regulation by RT-qPCR and western blot (Figure 3.4).

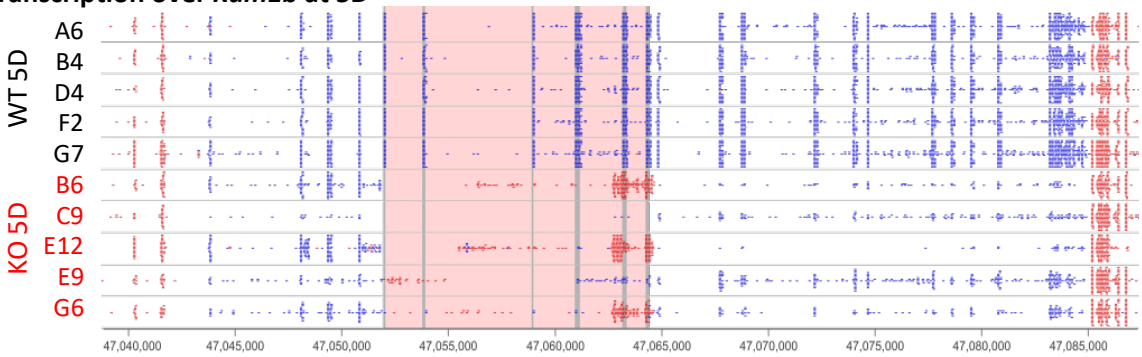
Bearing this in mind, I decided to assess whether the changes to *Esrrb* and *Gcm1* seen by RT-qPCR upon loss of KDM1B were recapitulated by RNA-seq. Figure 4.3 illustrates the log2RPM values for each of the stem, syncytiotrophoblast and giant cell marker genes tested in chapter 3. Broadly speaking, the RNA-seq data did not reflect the transcriptional de-regulation previously observed by RT-qPCR, except that *Gcm1* showed consistently higher expression in KDM1B null cells, although not significant.

To further assess the RNA-seq data, I performed differential gene expression analysis using both DESeq2 and an intensity difference analysis ($p<0.05$) to assess what effect loss of KDM1B had on TSCs. Very few genes were identified as differentially expressed by either method when comparing WT and knockout groups at each time point (Figure 4.4a). This is reflected in the principal component analysis (PCA), which did not separate the samples based on genotype (Figure 4.4b). The PCA analysis, however, did identify the divergence of the transcriptional profiles of cells along the differentiation time course, with PC1 (40 %) separating the samples based on day of differentiation. PC2 (12 %) separated 1D and 3D samples from stem and 5D samples, potentially identifying a higher degree of similarity between 1D and 3D cells. This is reflected in the lower number of differentially expressed genes identified between these groups. It could also reflect two waves of large changes that occur: first immediately at the point when TSCs exit the stem state (i.e. stem vs 1D), and second as terminally differentiated populations begin to arise (3D vs 5D). This will be further explored in Chapter 5.

**A.i Transcription over *Kdm1b* during *in vitro* differentiation**




**A.ii Transcription over *Kdm1b* at 5D**




**B.i Histone lysine methyltransferases**

**ii. Histone lysine demethylases**

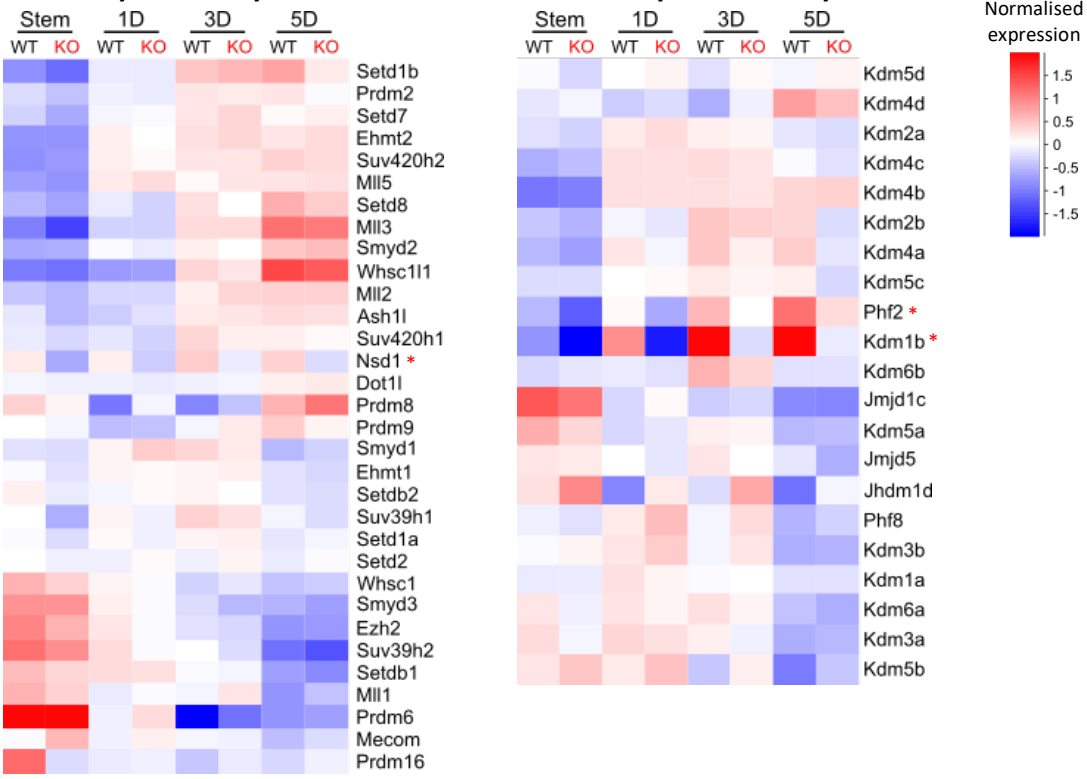



**Figure 4.2 Transcription of Kdm1b and other histone demethylases and methyltransferases during in vitro differentiation**

**A.i** The raw reads over the Kdm1b gene throughout in vitro differentiation in WT and *Kdm1b*$^{-/-}$ cells, **ii** shows transcription of each clone at 5D differentiation. Highlighted in red is the excised region of

*Kdm1b* and in grey are the deleted exons **B.** Heatmap showing normalised expression of histone lysine methyltransferases and demethylases across the differentiation timecourse. Red asterisk denotes those KMT and KDM enzymes that are on chromosome 13.

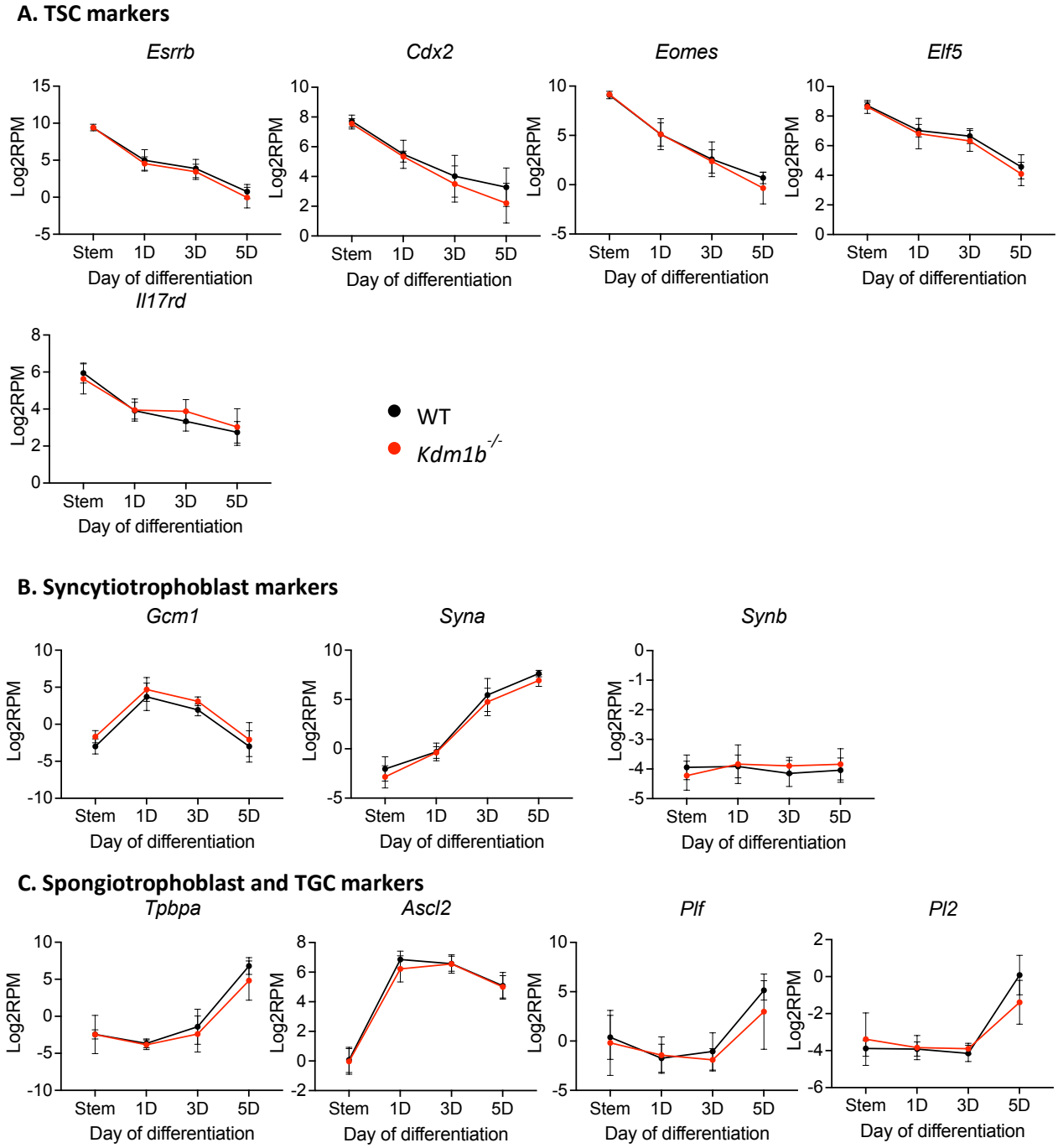



**Figure 4.3 LogRPKM values of key markers of stem cell, syncytiotrophoblast and TGCs**

Line graphs showing transcriptomic data of the expression of key markers of trophoblast stem cells (**A**), Syncytiotrophoblast (**B**) and spongiotrophoblast and TGCs (**C**) over the differentiation timecourse, previously assessed by RT-qPCR.

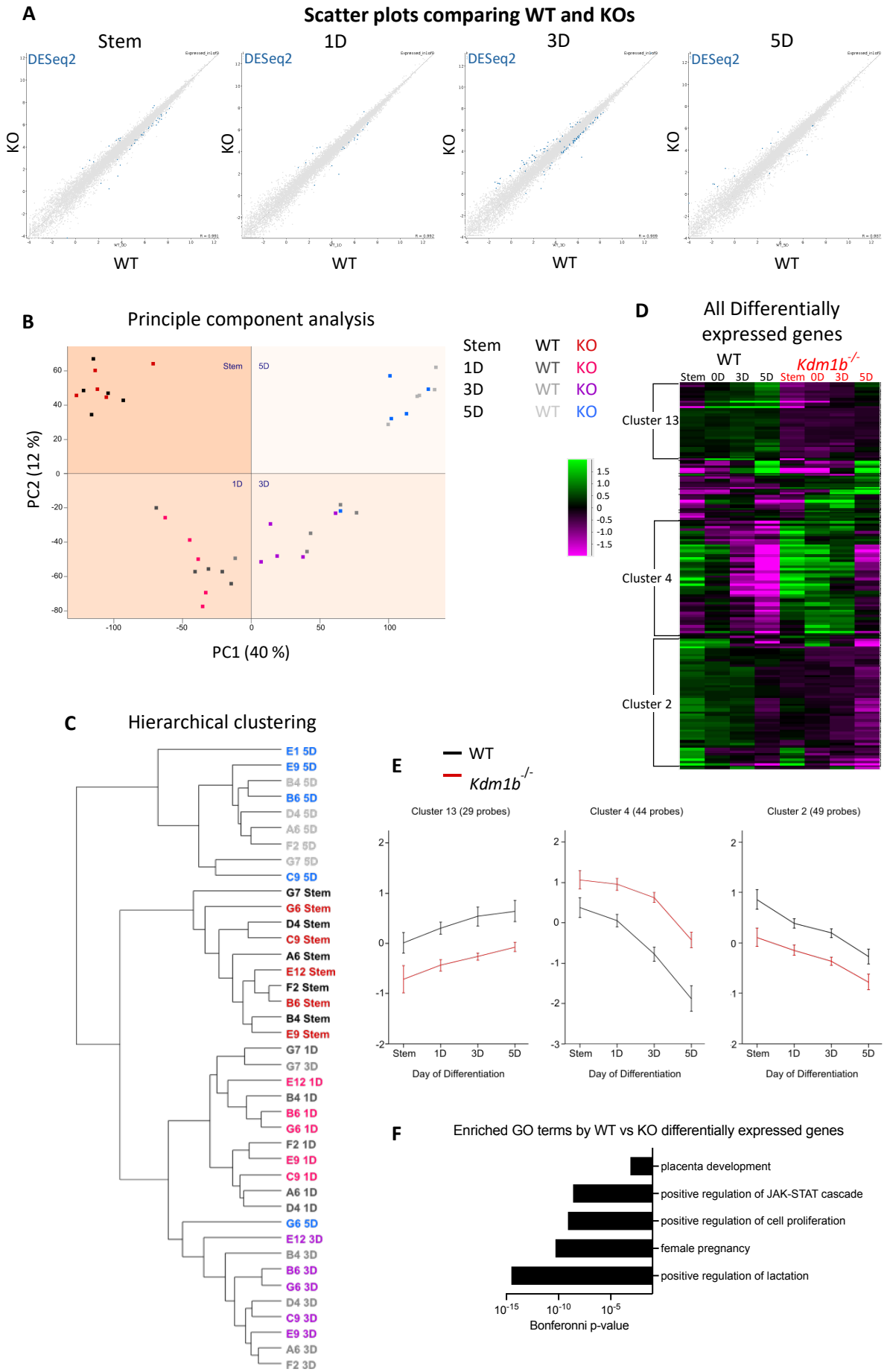

A
Scatter plots comparing WT and KOs
Stem
1D
3D
5D
DESeq2
KO
WT
B
Principle component analysis
PC2 (12 %)
PC1 (40 %)
Stem
5D
1D
3D
Stem WT KO
1D WT KO
3D WT KO
5D WT KO
1.5
1.0
0.5
0
-0.5
-1.0
-1.5
D
All Differentially expressed genes
WT
Kdm1b-/-
Stem 0D 3D 5D Stem 0D 3D 5D
Cluster 13
Cluster 4
Cluster 2
C
Hierarchical clustering
E1 5D
E9 5D
B4 5D
B6 5D
D4 5D
A6 5D
F2 5D
G7 5D
C9 5D
G7 Stem
G6 Stem
D4 Stem
C9 Stem
A6 Stem
E12 Stem
F2 Stem
B6 Stem
B4 Stem
E9 Stem
G7 1D
G7 3D
E12 1D
B4 1D
B6 1D
G6 1D
F2 1D
E9 1D
C9 1D
A6 1D
D4 1D
G6 5D
E12 3D
B4 3D
B6 3D
G6 3D
D4 3D
C9 3D
E9 3D
A6 3D
F2 3D
E
WT
Kdm1b-/-
Cluster 13 (29 probes)
Cluster 4 (44 probes)
Cluster 2 (49 probes)
Stem 1D 3D 5D
Day of Differentiation
F
Enriched GO terms by WT vs KO differentially expressed genes
placenta development
positive regulation of JAK-STAT cascade
positive regulation of cell proliferation
female pregnancy
positive regulation of lactation
10-15 10-10 10-5
Bonferonni p-value

**Figure 4.4 RNA-seq analysis shows little impact of loss of KDM1B on *in vitro* TS-Rs26 differentiation**

**A.** Scatter plots comparing WT and Kdm1b-/- cells at (**i**) stem, (**ii**) 1D, (**iii**) 3D and (**iv**) 5D Differentiation highlighting differentially expressed genes. DESeq2 p<0.05. **B.** PCA plot of RNA-seq libraries based on expressed genes (>10 reads). Each quadrant separates the samples on day of differentiation. Red arrows highlight WT sample G7 3D and *Kdm1b*$^{-/-}$ sample G6 5D **C.** Hierarchical clustering shows separation of RNA-seq libraries by day of differentiation. **D.** Heatmap showing normalized expression of all differentially expressed genes (DESeq2 p<0.05) with hierarchical clusters highlighted, R>0.7, more than 10 probes per cluster. **E.** Line graphs summarising normalised expression of genes in each cluster identified by hierarchical clustering of DESeq2 (p<0.05) genes. **F.** Statistically significant GO terms identified with all DESeq2-identified genes.

Hierarchical clustering of the RNA-seq data also did not separate the KDM1B null cells from the WT clones, showing clustering by day of differentiation only (Figure 4.4c). Interestingly, two of the samples clustered with different days: clone G7 at 3D differentiation clustered more closely to the 1D samples, and 5D differentiated clone G6 clustered with the 3D samples. This was also reflected in the PCA plot.

All differentially expressed genes identified by DESeq2 analysis: cumulative pair-wise comparisons between WT and *Kdm1b*$^{-/-}$ at each time point (i.e comparing WT stem with KDM1B null stem samples) are plotted by heatmap in Figure 4.4d. Of the few genes identified as differentially expressed at each of the four timepoints between WT and *Kdm1b*$^{-/-}$ cells, most of them appeared to be consistently down- or up-regulated, respectively, throughout the time course (Figure 4.4d/e). Too few differentially expressed genes were identified when running pairwise comparisons at each day individually to assess them with GO term analysis. I therefore consolidated the lists to see if any biological processes or molecular pathways could be identified as overrepresented (Figure 4.4f). The compiled list of DE genes identified by DESeq2 analysis did not reveal any meaningful enrichment terms. However, the compiled list of DE genes identified by Intensity Difference (p<0.05) did yield some applicable GO terms, such as placental development and female pregnancy. These were almost entirely driven by the Prolactin (*Prl*) genes, which appear in two large clusters on chromosome 13.

#### 4.2.2.1 The role of Chromosome 13 during trophoblast differentiation

When assessing the genomic distribution of the differentially expressed genes identified, I noticed that there was a high bias in the data for genes on chromosome 13 (Figure 4.5a). Parietal trophoblast giant cells, while undergoing endoreduplication, do not copy their entire genome equivalently (Hannibal *et al.*, 2014; Hannibal and Baker, 2016). Several large regions of the genome appear underrepresented and five smaller regions appear overrepresented. Four of the five over-represented regions appear on chromosome 13, with the fifth on chromosome 6. As the cells become more highly differentiated, the expression of the genes within these overrepresented regions increases (Hannibal and Baker, 2016).

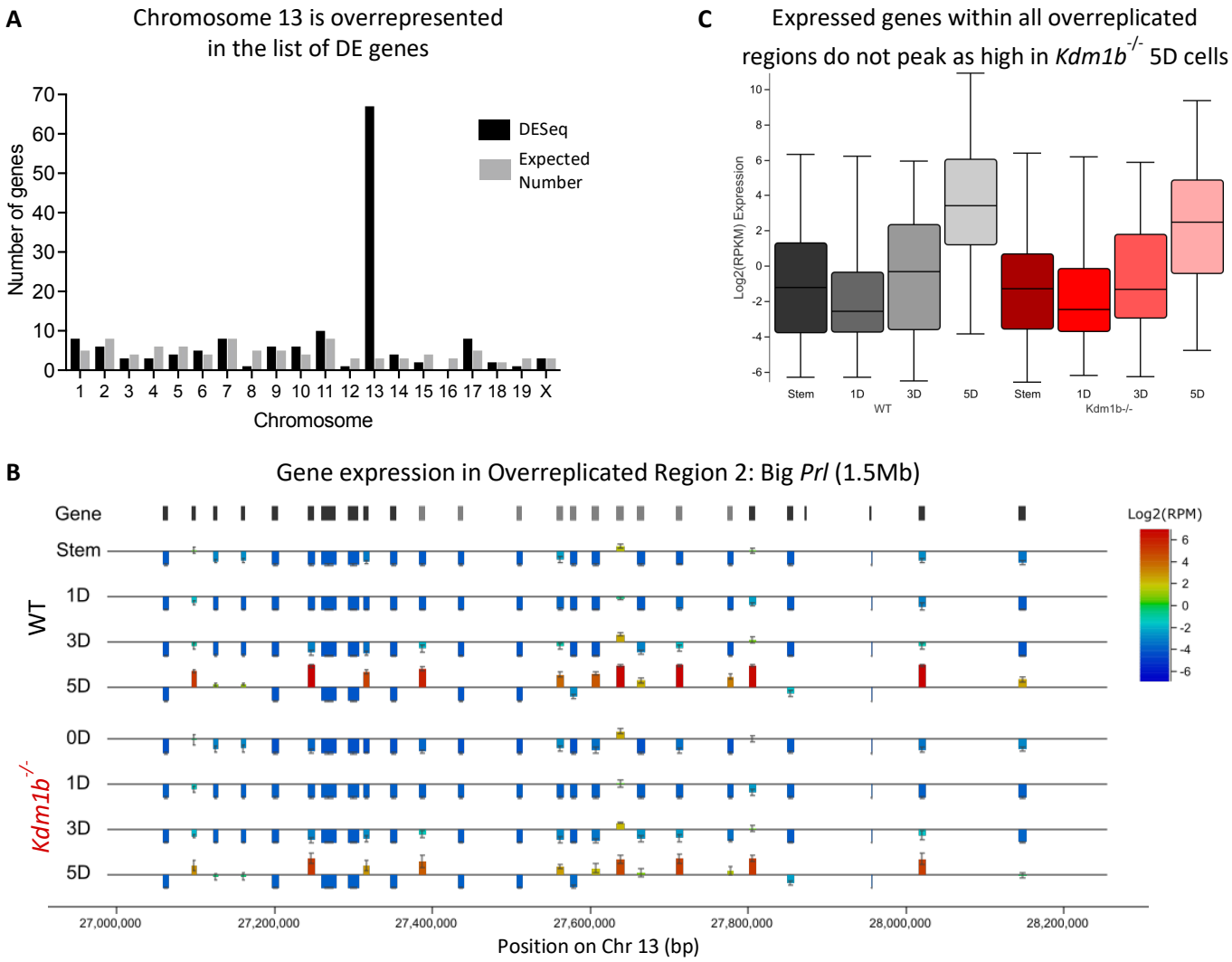



**Figure 4.5 Chromosome 13 contains key placental genes and are overrepresented in differentially expressed genes (WT v *Kdm1b*$^{-/-}$)**

(**A**) Bar plot of the number of genes on each chromosome of the differentially expressed genes (DESeq p<0.05, black) compared to the expected umber based on number of expressed genes on each chromosome (grey). (**B**) Chromosome view of the Hannibal and Baker (2016) "Big *Prl*" region. Gene expression is log2(RPM), error bars are SEM. (**C**) Box and whisker plot of gene expression of those genes within these overreplicated regions. (**D**) Overlap of differentially expressed genes identified by DESeq or Intensity difference and genes within Hannibal and Baker's overreplicated regions. (**E**) Box and whisker plots of those DE genes within over-replicated regions.

While TGCs at 5D differentiation *in vitro* represent an earlier time point than those studied by Hannibal et al, I observed an upregulation of genes in these regions throughout the differentiation time course (Figure 4.5b,c).

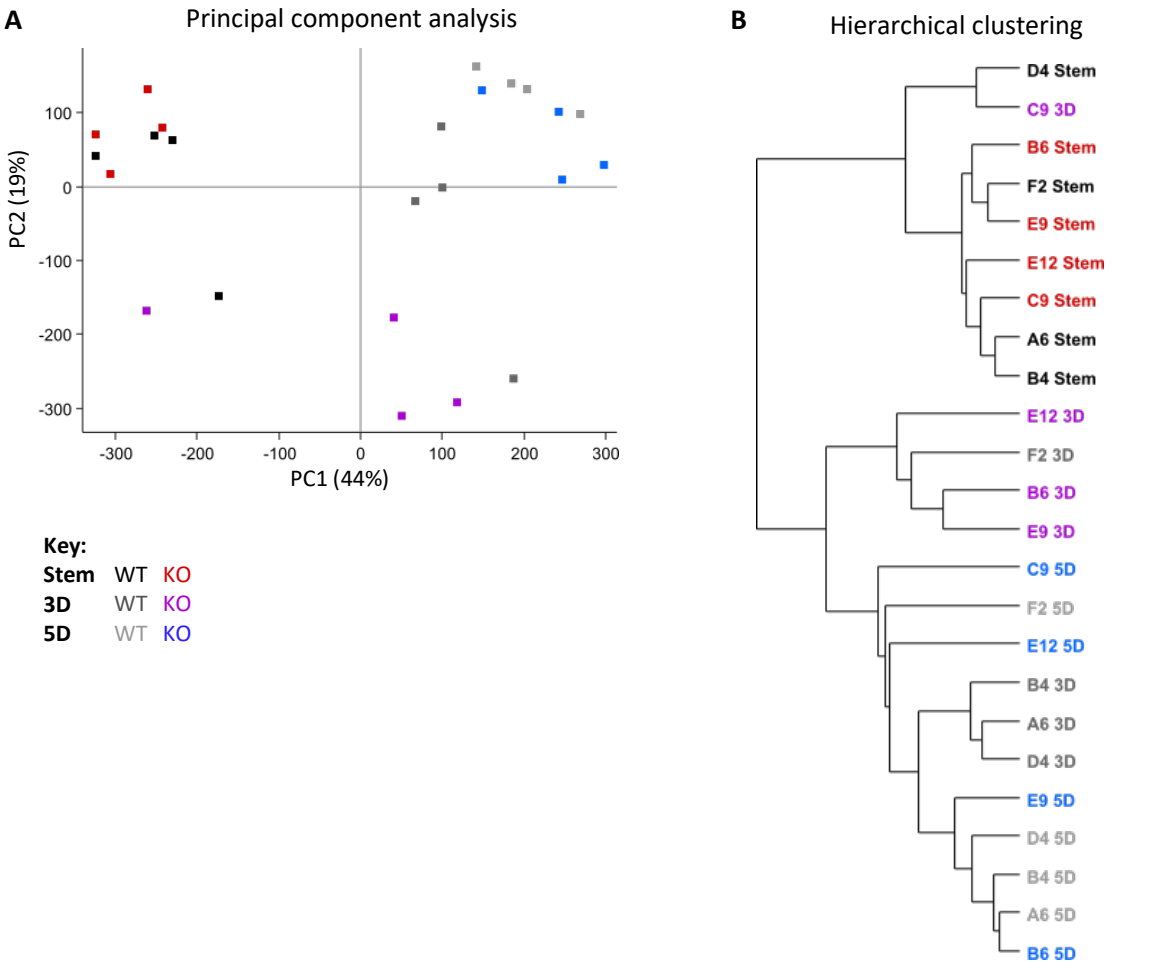
None of the differentially expressed genes identified by DESeq2 ($p<0.05$) were found in these overreplicated regions, however several DE genes identified by the intensity difference filter ($p<0.05$), were (19 out of 123 down-regulated genes, Figure 4.5d). Box whisker plots depicting the mean expression levels of these 19 genes in WT versus KO TSCs across the differentiation time course are shown in Figure 4.5e; whilst these genes are up-regulated in *Kdm1b*$^{-/-}$ cells, they did not reach the same level of expression at 5D compared to WT. This could potentially be attributed to KDM1B regulating transcription and even potentially, the subsequent over-replication of these regions. This will be revisited in section 4.2.8.

### 4.2.3 H3K4 monomethylation is affected by loss of KDM1B at 5D differentiation

As KDM1B is a histone lysine demethylase and I had observed global increase in its substrate mark, H3K4me1, as well as in H3K36me3, by western blot, I performed ChIP-seq in stem, 3D and 5D differentiated trophoblast cells, targeting H3K4me3 and H3K4me2 in addition to these two marks. Libraries of four biological replicates per genotype were generated at each time point: clones A6, B4, D4 and F2 were used for the WT ChIP-seq analysis and the *Kdm1b*$^{-/-}$ clones used were B6, C9, E9 and E12. One day differentiated cells were not used as western blot analysis showed a tendancy towards accumulation of H3K4me1 after 2D differentiation. I hypothesised that changes to chromatin architecture were more likely to become evident later in differentiation, when more cells become positive for KDM1B (Figure 3.1). Therefore, I decided to prioritise later time points in differentiation, performing ChIP-seq analysis on 5D differentiated cells rather than 1D.

All ChIP-seq samples showed reads mapping to the sense and opposing strands at a rate of 50 %, indicating no bias in the data, as would be expected from chromatin enrichment. Input samples showed some regions of enrichment as well as areas of significant depletion, which is expected as certain regions of the genome are more accessible and therefore over-represented in ChIP-seq libraries and some are harder to ChIP or map to the genome due to heterochromatic or repetitive regions, respectively. To begin, I focused on the H3K4me1, H3K4me2 and H3K4me3 ChIP-seq data because KDM1B acts directly on H3K4me1 and H3K4me2-marked chromatin, which in turn could impact H3K4me3 peaks. Also, these histone modifications routinely produce sharp peaks in ChIP-seq assays facilitating the analysis of the resultant data. Peak calling was performed with MACS2 and the corresponding input sample was used for reference; regions which were highly enriched in the input samples were then excluded and peaks within 1kb of each other were combined.

Firstly, I focussed on H3K4me1 data, as this mark was globally enriched in the KDM1B null cells. PCA was performed using the consolidated list of high-confidence H3K4me1 peaks, which separated stem, 3D and 5D differentiated samples, relatively well (Figure 4.6a). Broadly speaking, the stem and 5D samples clustered closely together within their groups, while the 3D intermediately differentiated samples were more disparate. Hierarchical clustering of the samples showed similar tight clustering of the stem cell samples from 3D and 5D samples, with less distinction between the latter (Figure 4.6b). This suggests that H3K4me1 did not change as distinctly as the transcriptome during this phase of *in vitro* differentiation.



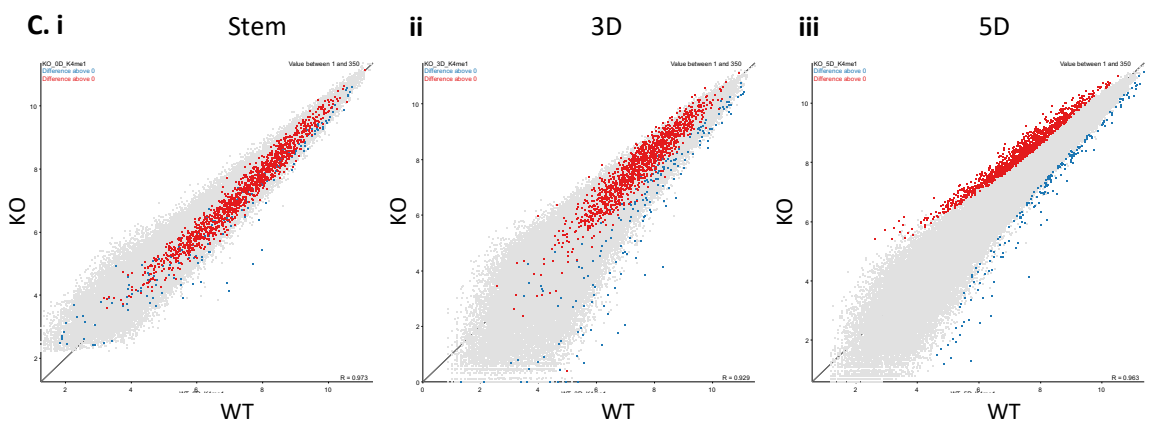



**Figure 4.6 There are around 1500 differentially enriched H3K4me1 peaks at 5D differentiation**

(**A**) Principal component analysis of the 24 ChIP-seq libraries probing H3K4me1. WT clones are in black coded by day of differentiation and *Kdm1b*$^{-/-}$ clones in red. The same colour-coding is used in (**B**) Hierarchical clustering of the H3K4me1 libraries. (**C**) Scatter plots of H3K4me1 peaks comparing WT and KDM1B null cells in **i.** stem, **ii.** 3D and **iii.** 5D differentiation. Highlighted in red are those differentially enriched peaks (LIMMA p<0.05) which are more enriched in *Kdm1b*$^{-/-}$ cells at 5D and in blue, those peaks which are down at 5D

In order to assess what impact loss of KDM1B had on H3K4me1 across the genome, I performed LIMMA statistics (p<0.05) on the consolidated list of peaks, calling differentially enriched peaks between the WT and *Kdm1b*$^{-/-}$ replicate sets at each time point: stem, 3D and 5D differentiation. Interestingly, differentially enriched peaks were only found at 5D differentiation (Figure 4.6c). The vast majority of these peaks (1330/1525, 87 %) were more highly enriched in the KDM1B null cells compared to WT. This was expected as the loss of a histone demethylase would likely result in more of its substrate mark residing in the genome.

The progressive enrichment of these 5D-specific H3K4me1 peaks over differentiation can be nicely visualized in Scatter plots comparing WT and KO cells. In stem cell conditions, the majority of these peaks were equally distributed in the stem cell samples; at 3D differentiation the peaks start to migrate away from the y=x line, becoming increasingly divergent from the WT cells (Figure 4.6ci, ii). The gradual nature of the accumulation might suggest that these regions in WT cells are cyclically methylated and de-methylated, and loss of KDM1B led to a steady gain of methylation. Thus, the onset of global H3K4me1 differences observed at 2D by western blot becomes manifest as highly reproducible (n=4 biological replicates per time point), locus-specific enrichment of this histone mark by 5D of differentiation.

### 4.2.4 Differentially enriched H3K4me1 peaks mostly localize to intragenic regions

Next, I assessed whether these differentially enriched H3K4me1 peaks (Figure 4.7a) were co-marked with H3K27ac, a mark of active enhancers. For this analysis, I used publicly available H3K27ac ChIP-seq data derived from TSCs in the stem cells state (Chuong *et al.*, 2013). Indeed, the differentially enriched peaks were more likely to coincide with H3K27ac compared to all H3K4me1 peaks (53 % compared to 18 %). The majority of peaks which gained H3K4me1 signal in the knockouts coincided with H3K27ac signal (56 %), whereas for peaks which lost signal in 5D *Kdm1b*$^{-/-}$ cells, the majority were not co-marked with H3K27ac (61 %). Further, differentially enriched peaks that coincided with H3K27ac had a higher

mean H3K4me1 enrichment than peaks not marked with H3K27ac, in groups of peaks both gaining and losing H3K4me1 signal in KDM1B null cells (Figure 4.7b).

KDM1B has been shown to localise to the gene body of actively transcribed genes (Fang *et al.*, 2010), so I assessed the genomic distribution of the differentially enriched H3K4me1 peaks compared to the global distribution of this mark, relative to genes. Differentially enriched H3K4me1 peaks were more likely to be intragenic (inner donut) compared to all H3K4me1 peaks (outer donut, Figure 4.7c.i). Further, the peaks which gained methylation in KDM1B null cells (Figure 4.7c.ii) were almost entirely found to be overlapping or within 5 kb of genes, with less than 1 % further than 5 kb from genes, suggesting that these peaks play a regulatory role in gene expression. The fact that loss of KDM1B resulted in this enrichment of overwhelmingly intragenic H3K4me1 peaks, confirms that at least in these regions, KDM1B is likely to be acting intragenically, possibly with elongating RNA PolII, to remove methylation on lysine 4 of histone H3.

Conversely, Figure 4.7c.iii demonstrates that the peaks that lost methylation in the *Kdm1b*$^{-/-}$ cells, were more likely to be intergenic compared to all H3K4me1 peaks. This pointed to either a distinct mode of action regulating these peaks; or possibly tied to what was previously seen with differentially expressed genes whose expression was reduced in *Kdm1b*$^{-/-}$ cells: being significantly enriched on chromosome 13. Figure 4.7d illustrates the relative chromosomal distribution of the H3K4me1 peaks that gained (red) or lost (blue) methylation in *Kdm1b*$^{-/-}$ cells, normalised to the expected distribution. There was a striking enrichment for chromosome 13 in H3K4me1 peaks that lost methylation in knockout cells. When looking genome-wide at the location of differentially enriched H3K4me1 peaks (Figure 4.7e), I noticed that in addition to chromosome 13, the 5' end of chromosome 2 was also enriched for peaks losing methylation (highlighted, blue). These peaks also seemed to be clustered closely together and when looking solely at these 40 regions, I found that they were more likely to be located within genes compared to all other 200 down-regulated H3K4me1 peaks (Figure 4.7f).

The majority of these intragenic chromosome 2 peaks that lose methylation in 5D differentiated cells overlapped three genes: *Tshz2*, *Gm14266* and *Cdh4* (Figure 4.8). In the case of *Tshz2* (Figure 4.8b), there appeared to be a total loss of H3K4me1 peaks across this gene in *Kdm1b*$^{-/-}$ cells. For peaks located within *Cdh4* (4.8c), there was reduced signal in KDM1B null stem cells, which was maintained throughout differentiation, indicating a failure to accumulate H3K4me1 methylation in these regions. Interestingly, the regions upstream of both *Tshz2* and *Cdh4* appeared to have universally low H3K4me1 signal. One hypothesis for this observation is that loss of KDM1B resulted in the loss of a boundary between hetero- and euchromatin in these locations. For the *Tshz2* region, this was potentially happening with higher penetrance compared to the *Cdh4* region. The *Cdh4* locus (Figure 4.8c) showed maintenance of a low-level of H3K4me1 in 5D differentiated cells.

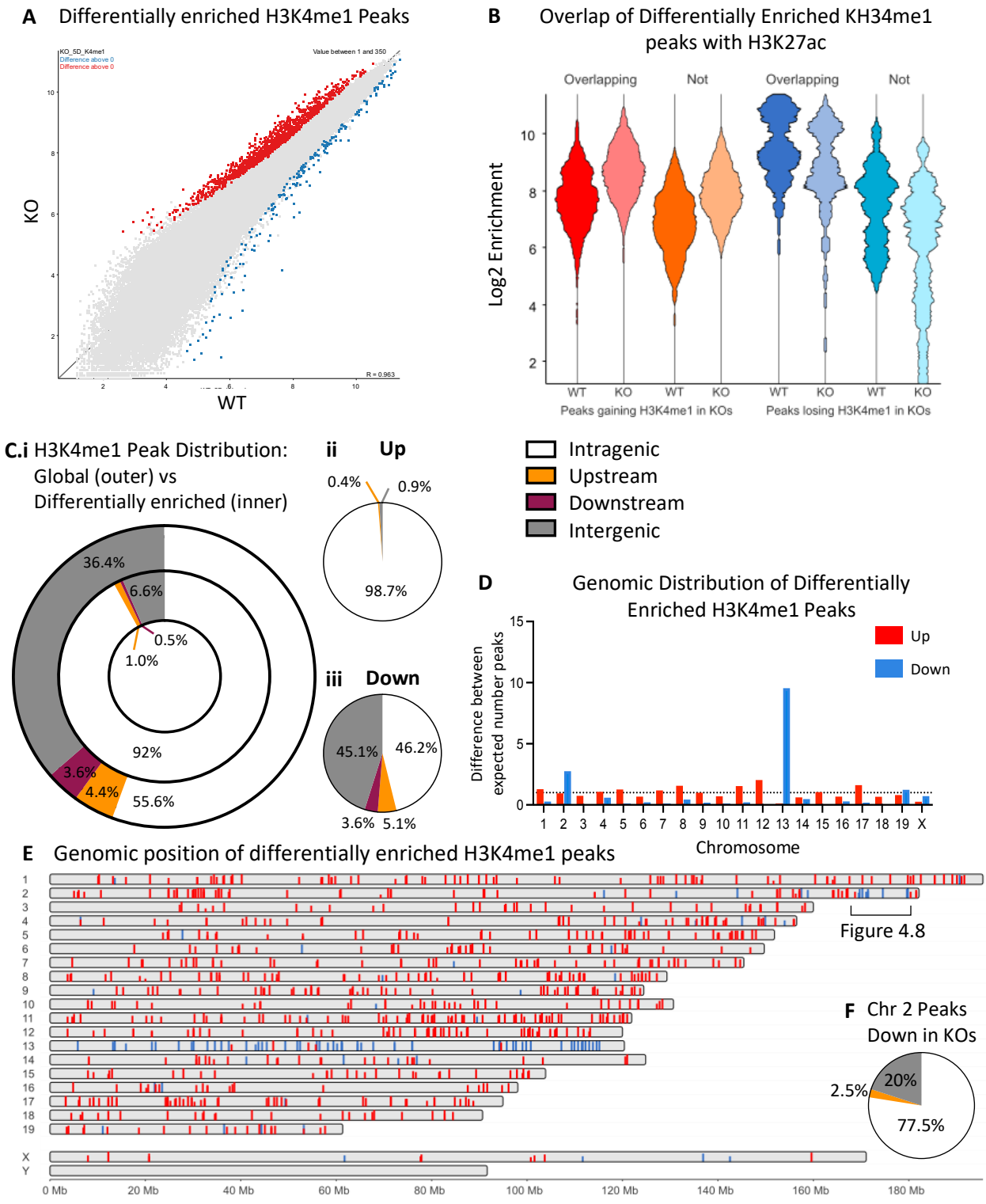



**Figure 4.7 Distinguishing features of H3K4me1 peaks gaining/losing methylation in KDM1B null 5D differentiated cells**

**A**. Scatter plot highlighting DE H3K4me1 peaks (statistical test was LIMMA p<0.05) highlighting peaks gaining (red) or losing (blue) methylation in 5D differentiated KDM1B null cells compared to WT. **B**. Bean plots showing log2 enrichment of differentially enriched peaks. Plotted are peaks gaining or losing methylation split by whether they overlap H3K27Ac peaks. **C**. Genomic distribution of differentially enriched peaks relative to genes. **C.i** Donut plot of all (outer) and DE (inner) H3K4me1 peaks. **C.ii-iii** Pie charts plotting those DE peaks separated by whether they gain (**ii**) or lose (**iii**) signal

in knockouts. **D**. Bar chart plotting chromosomal distribution of DE peaks gaining (red) losing (blue) methylation at 5D. Line at y=1 represents no enrichment. **E**. Genome view of DE peaks gaining (red) and losing (blue) signal at 5D differentiation. **F**. Pie chart showing distribution of peaks losing methylation located on chromosome 2 only.

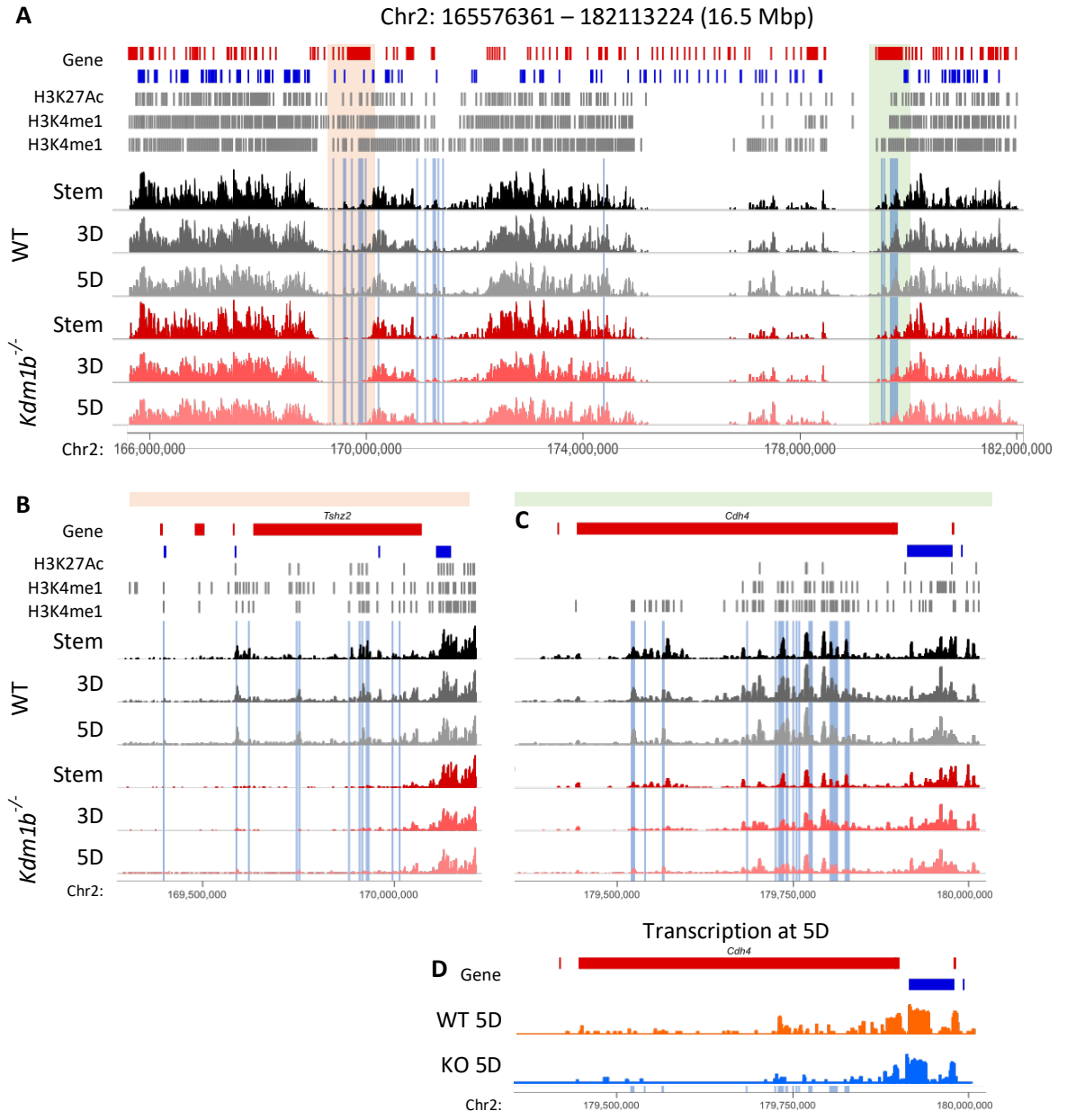



**Figure 4.8 chromosome view of 16.5Mbp region of chromosome 2 enriched in intragenic down-regulated H3K4me1 peaks**

**A**. Seqmonk tracks showing H3K4me1 ChIP-seq peaks during differentiation across the 16.5Mb region of chromosome 2. The orange region contains *Tshz2* (**B**) and the green region contains *Cdh4* (**C**). Highlighted in blue are the downregulated H3K4me1 peaks in this region. Annotation tracks are Gene (sense in red, antisense in blue), H3K27ac & H3K4me1 (Chuong *et al* 2013) and my consensus H3K4me1 MACS2-called peaks. **D.** Wiggle plot showing mean transcription within the *Cdh4* region in **C** at 5D in WT cells (orange) and KDM1B null cells (blue).

Indeed, expression of *Cdh4* increased in WT 5D differentiated cells to Log2RPM value of 1.66 compared with just 0.06 in *Kdm1b*$^{-/-}$ 5D differentiated cells (wiggle plot, Figure 4.8d). Conversely, both *Tshz2* and *Gm14266* were not expressed at all (fewer than 10 reads across all samples, data not shown). The question of why sites that lose H3K4me1 were so enriched on chromosome 13 will be addressed in section 4.2.8.

### 4.2.5 Upregulated H3K4me1 peaks were enriched for TSC TF binding sites

Next, I decided to further characterise the H3K4me1 peaks that gain methylation in *Kdm1b*$^{-/-}$ TSCs (Figure 4.9). The mean log2 enrichment of DE H3K4me1 peaks gaining methylation in *Kdm1b*$^{-/-}$ cells decreased slightly during WT in vitro differentiation from 7.6 to 7.2 and increased from 7.5 to 8.3 in *Kdm1b*$^{-/-}$ cells (Figure 4.9a). Hierarchical clustering of these peaks demonstrated four patterns of enrichment (accounting for >99.9 % of peaks; Figure 4.9b). Just under half (44 %) of these peaks formed three distinct clusters. Cluster 1 (Figure 4.9c.i, 215 peaks) showed an accumulation of methylation between stem and 3D differentiation in KDM1B null cells, which was maintained to 5D. Comparatively, the WT enrichment of these peaks decreased between 3D and 5D differentiation. Cluster 2 included 268 peaks and showed no accumulation of H3K4me1 over the course of differentiation in *Kdm1b*$^{-/-}$ cells (Figure 4.9c.ii). For both clusters, the relative gain in enrichment appeared to be due to a failure to demethylate these peaks, rather than increased deposition of monomethylation.

The smallest group, cluster 4 (Figure 4.9c.iii, 103 peaks), showed a relative increase in signal between 3D and 5D differentiation in *Kdm1b*$^{-/-}$ cells only, with average enrichment not changing during WT differentiation. The largest cluster, cluster5, containing 743 peaks (56 % peaks), showed an accumulation of methylation in both WT and *Kdm1b*$^{-/-}$ cells across differentiation, but this gain in methylation was far more pronounced in the KO cells (Figure 4.9c. iv). Clusters 4 and 5 represent peaks that accumulated methylation in KDM1B null cells, suggesting that these regions are cyclically methylated and de-methylated, and loss of KDM1B leads to a steady gain of methylation.

I also decided to run HOMER analysis on these peaks, using 1kb regions, to assess whether there were any over-represented binding motifs present. Interestingly, the top 10 hits contained motifs for several highly expressed trophoblast-specific transcription factors such as GATA3, TFAP2C and NR4A1 (Figure 4.9d).

As almost 99 % of these peaks were intragenic, I decided to look at expression of the genes containing the upregulated peaks. Almost all of the expressed genes (>10 reads in at least one replicate set) that contained one or more of the upregulated H3K4me1 peaks at 5D differentiation, were highly expressed across all samples throughout the time course (Figure 4.9e). Figure 4.9f shows that there was very little

difference in expression of these genes, between WT and *Kdm1b*$^{-/-}$ cells. Broadly speaking, there were three patterns of gene expression across the time course: 1. genes that lost gene expression during differentiation, 2. those with a temporary increase in expression and 3. those whose expression increased throughout differentiation. Despite there being apparently little-to-no consequence of the peaks gaining methylation in KDM1B null cells on gene expression, I decided to see whether there were any biological processes enriched in this gene list.

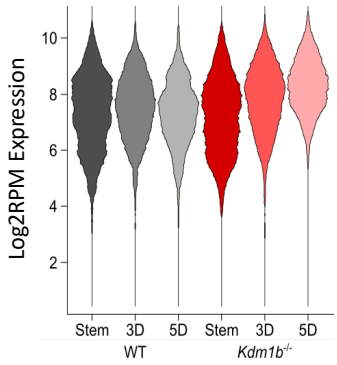
Figure 4.9g illustrates that GO terms such as 'regulation of chromosome organisation', 'DNA repair' and 'cell cycle' were among the most highly enriched. This might be indicative of DNA damage which may impact the cell cycle. In addition, there were several GO terms focussed on ATP metabolism, and histone H3-K4 methylation. Despite there not being any outright changes to gene expression of histone methyltransferases or demethylases, DE H3K4me1 peaks were found within some of these genes, suggesting that KDM1B plays a role in the expression or in the regulatory networks centred on these enzymes. With an extended time course of differentiation, and presumably even higher accumulation of H3K4me1 in these regions, expression of these genes might become altered.

### 4.2.6 Changes to H3K4 dimethylation in the absence of KDM1B mirrors changes to monomethylated H3-K4

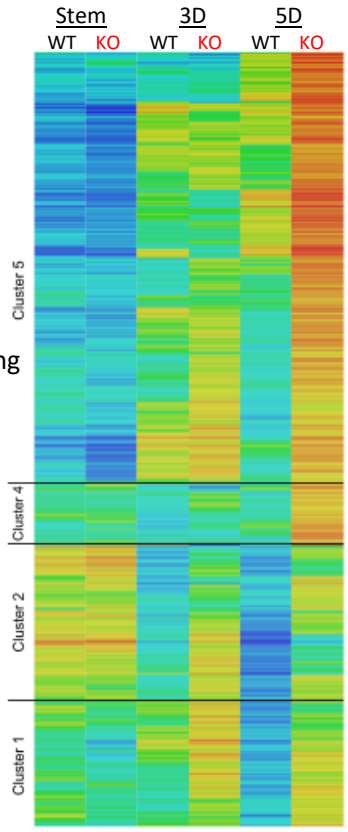
As well as H3K4me1 peaks, there were differentially enriched H3K4me2 peaks between WT and *Kdm1b*$^{-/-}$ 5D differentiated cells. Having consolidated the list of MACS peak called H3K4me2 peaks, combining those within 1 kb of each other, I performed PCA and hierarchical clustering to assess whether KDM1B null cells could be separated from WT cells at any time point on the basis of H3K4me2 enrichment (Figure 4.10a, b). As with the RNA-seq and H3K4me1 ChIP-seq data, the samples did not split by genotype, but the cells' differentiation state was well identified. When calling differentially enriched peaks by LIMMA ($p<0.05$), I noticed a similar pattern to the DE H3K4me1 peaks: there was a gradual shift in the peaks away from equal enrichment in stem cells to 3D differentiated cells, which reached significance at 5D (Figure 4.10c).

Also similar to monomethylated H3K4 peaks, the differentially enriched H3K4me2 peaks identified at 5D were highly enriched over genes (Figure 4.10d.i). Those peaks with increased signal in *Kdm1b*$^{-/-}$ cells (ii) were almost entirely intragenic, compared to those more highly enriched in WT cells (iii), whose distribution was much closer to that of the global H3K4me2 peak distribution.

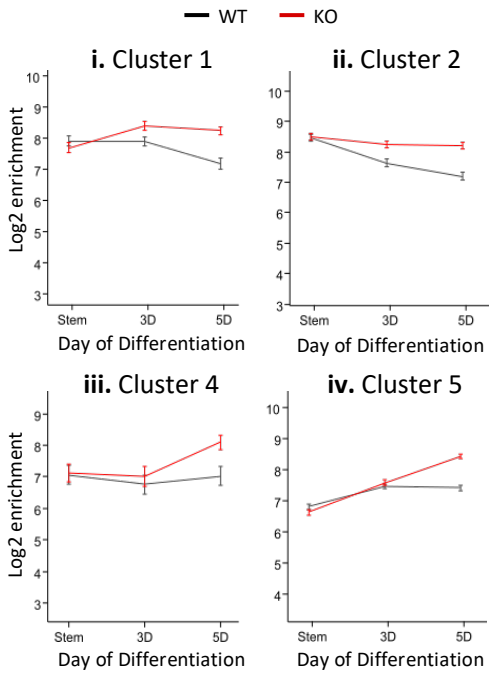
**A** Enrichment of H3K4me1 peaks gaining methylation in KOs



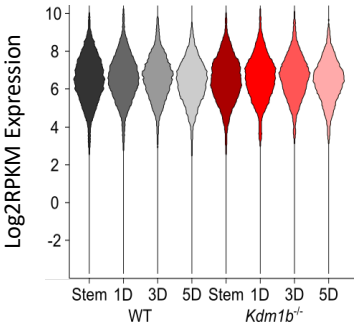

**B** Normalised enrichment of up-regulated Peaks



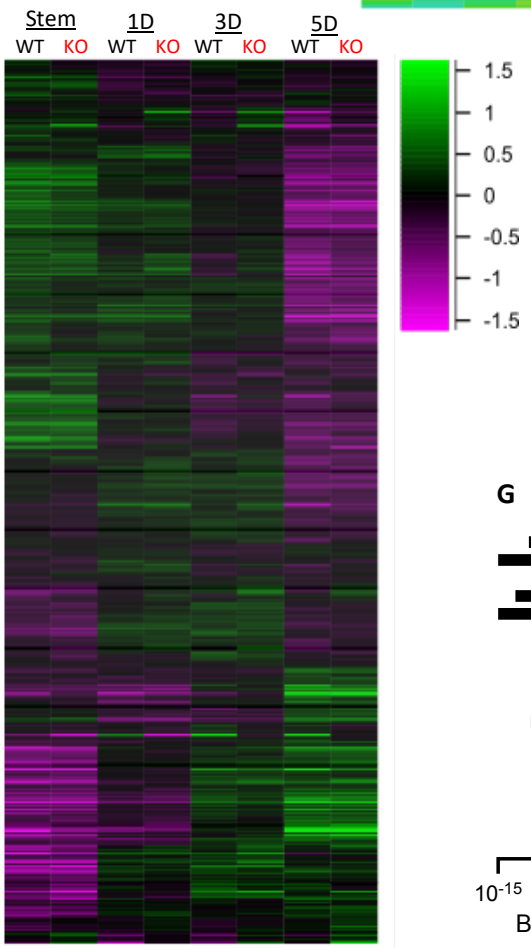

**C** Mean enrichment of DE H3K4me1 Peaks




**E** Expression of Genes containing upregulated H3K4me1 Peaks




**F** Normalised Expression




**D**

| ChIP *gene* | Log P-value | Q-value (Benjamini) | Motif |
|---|---|---|---|
| MEF2b *Mef2b* | -1.14E+01 | 0.0051 | |
| NR4A1 *Nr4a1* | -9.01E+00 | 0.0268 | |
| SCL *Tal1* | -8.06E+00 | 0.046 | |
| GATA3 *Gata3* | -8.03E+00 | 0.046 | |
| TFAP2C *Tfap2c* | -6.88E+00 | 0.0901 | |
| KLF4 *Klf4* | -6.58E+00 | 0.1017 | |
| PU.1 *Sfpi1* | -6.54E+00 | 0.1017 | |
| MYOD *Myod1* | -6.51E+00 | 0.1017 | |
| TFAP4 *tfap4* | -6.38E+00 | 0.1017 | |
| TEAD1 *tead1* | -5.71E+00 | 0.1454 | |

**G** Enriched GO terms

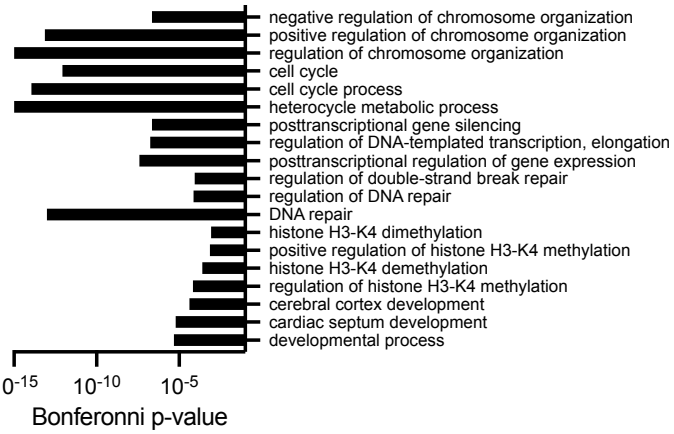


**Figure 4.9 H3K4me1 peaks gaining methylation in *Kdm1b*$^{-/-}$ 5D differentiated cells are enriched for trophoblast-specific transcription factor motifs.**

**A**. Bean plot showing enrichment of the differentially enriched peaks, showing global increase at 5D in *Kdm1b*$^{-/-}$ cells (red). **B**. Heatmap showing normalized enrichment of these peaks during differentiation, comparing WT and KDM1B null cells. Four clusters were identified (**C**) and the mean Log2 enrichment for each cluster is plotted. **D**. HOMER motif analysis on these upregulated regions identified binding motifs of several trophoblast-specific transcription factors. **E**. Bean plots of log2RPKM expression of genes overlapping upregulated DE H3K4me1 peaks show stable, high levels of expression throughout differentiation in WT and *Kdm1b*$^{-/-}$ cells. **F**. Heatmap of normalised expression of these genes showing no overall impact of gaining H3K4me1 signal on gene expression. **G**. GO term analysis of these genes included terms for DNA repair, synthesis, cell cycle and chromosome organization. Benjamini and Hochberg p-value of selected GO terms is plotted.

Figure 4.11a highlights those H3K4me2 DE peaks which directly overlap H3Kme1 DE peaks. The bean plots (a.ii) highlight that sites uniquely modified by H3K4me2 were more highly enriched in both WT and *Kdm1b*$^{-/-}$ cells compared to H3K4me1/2-dual peaks. There was no distinction in these groups as to enrichment for H3K27ac (data not shown), so it is not enhancer status that is responsible for this difference. I suspect it might be a knock-on effect of the accumulating H3K4me1 in and around these peaks, which led to increased H3K4me2 as a result. HOMER motif enrichment analysis on the 200 H3K4me2 peaks which gain methylation in the KDM1B null 5D cells (Figure 4.11b) produced far fewer hits than observed for H3K4me1, likely due to the fact that it was such a small list. The most enriched motif belonged to ZEB1, a factor best known for its role in epithelial-mesenchymal transition (Figure 4.11b; Peinado, Olmeda and Cano, 2007). ZEB1 has also been shown to interact with LSD1 as a component of the CoREST transcriptionally repressive complex (Wang *et al.*, 2007).

Of the down-regulated peaks, there was some overlap with the *Cdh4* and *Tshz* genes, with two H3K4me2 peaks having decreased enrichment at 5D within each gene; the *Cdh4* region is shown in Figure 4.11c. Similar to H3K4me1, these regions showed a fair amount of variation in the extent of the loss of ChIP-seq signal in each clone.

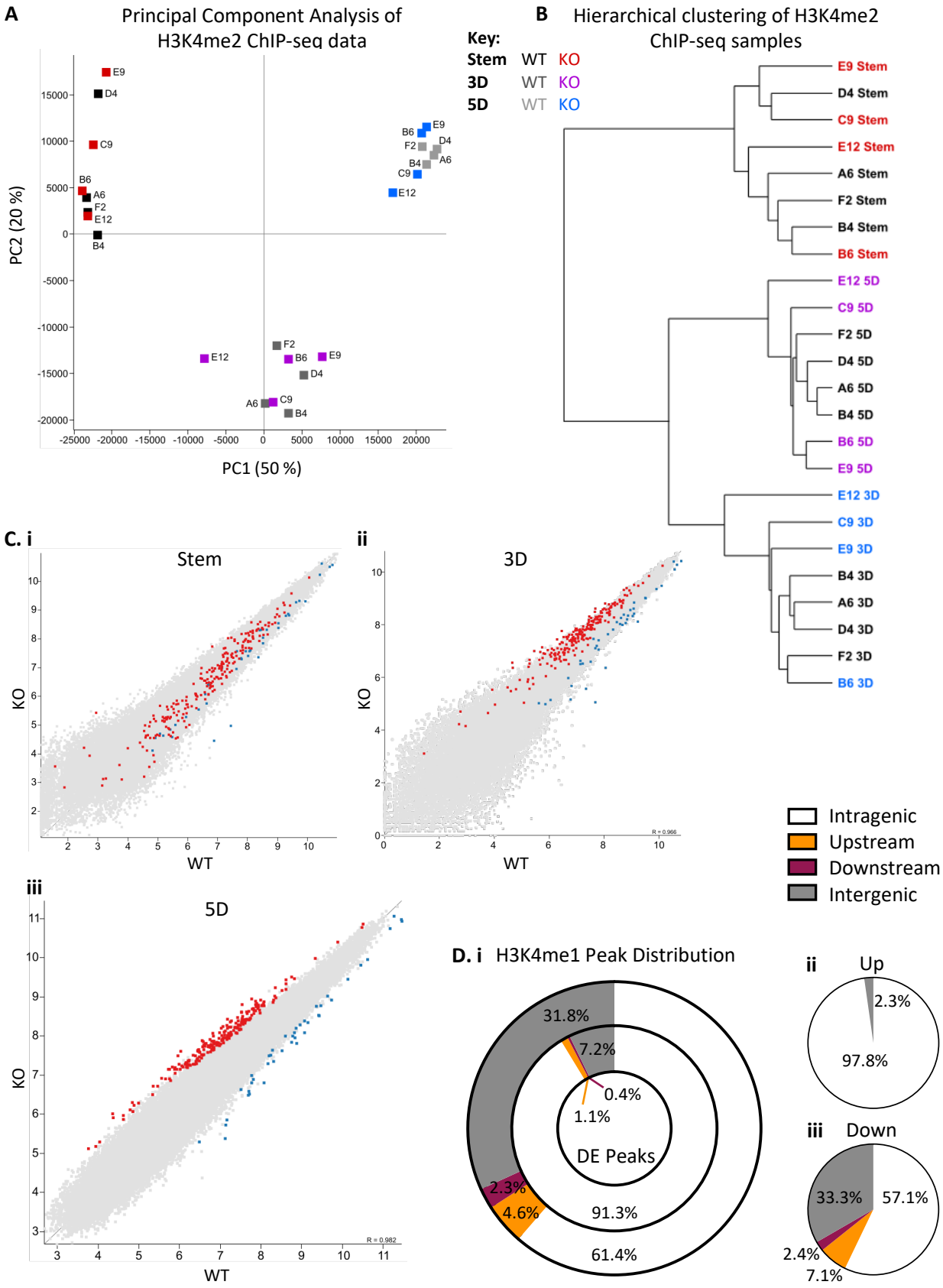
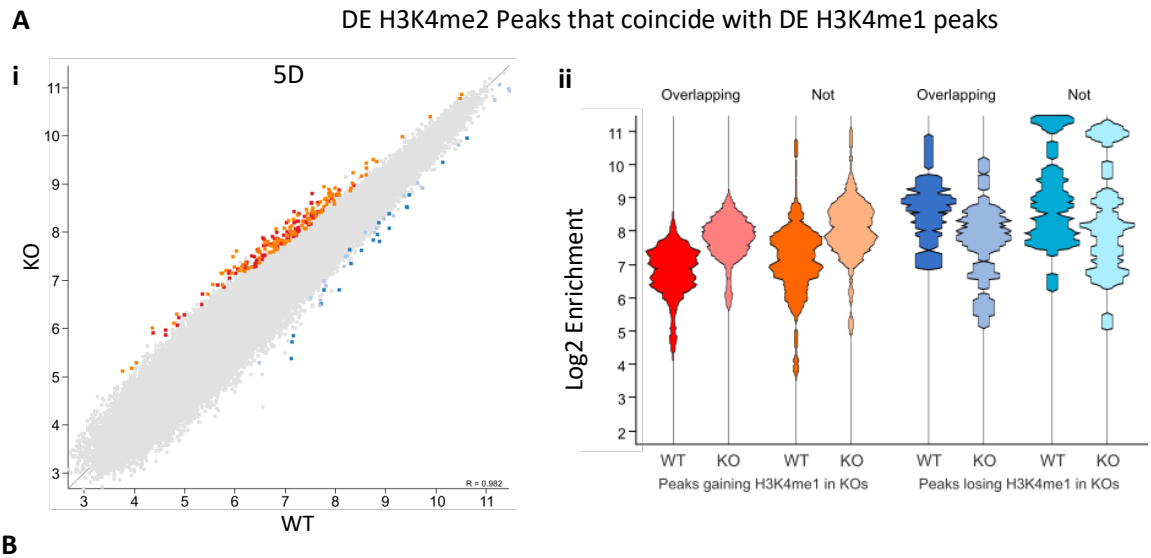

A
Principal Component Analysis of H3K4me2 ChIP-seq data
PC2 (20 %)
PC1 (50 %)
B
Hierarchical clustering of H3K4me2 ChIP-seq samples
Key:
Stem WT KO
3D WT KO
5D WT KO
E9 Stem
D4 Stem
C9 Stem
E12 Stem
A6 Stem
F2 Stem
B4 Stem
B6 Stem
E12 5D
C9 5D
F2 5D
D4 5D
A6 5D
B4 5D
B6 5D
E9 5D
E12 3D
C9 3D
E9 3D
B4 3D
A6 3D
D4 3D
F2 3D
B6 3D
C. i
Stem
ii
3D
iii
5D
KO
WT
Intragenic
Upstream
Downstream
Intergenic
D. i H3K4me1 Peak Distribution
31.8%
7.2%
0.4%
1.1%
DE Peaks
2.3%
4.6%
91.3%
61.4%
ii Up
2.3%
97.8%
iii Down
33.3%
57.1%
2.4%
7.1%

**Figure 4.10 Differentially enriched H3K4me2 peaks also show significant enrichment within gene bodies**

**A.** PCA plot for H3K4me2 libraries **B.** Hierarchical clustering of H3K4me2 peaks, highlighted in red are *Kdm1b*$^{-/-}$ cells and in black are WT. **C.** Scatter plots comparing WT and KDM1B null cells in i. stem, ii. 3D and iii. 5D differentiated cells, highlighting those DE peaks (LIMMA p<0.05) upregulated in knockouts in red and downregulated, in blue. **D.i** Donut plot showing global distribution of H3K4me2 peaks (outer) compared to DE H3K4me2 peaks (inner). Pie charts highlights those peaks gaining (**D.ii**) methylation in KDM1B null cells and those losing (**D.iii**) methylation in the knockouts.

**A**




**B**

| ChIP gene | log P-pvalue | q-value (Benjamini) | Motif |
|---|---|---|---|
| ZEB1 *Zeb1* | -8.34E+00 | 0.1043 | |
| P53 | -7.31E+00 | 0.1043 | |
| P73 | -6.41E+00 | 0.1794 | |

**C**

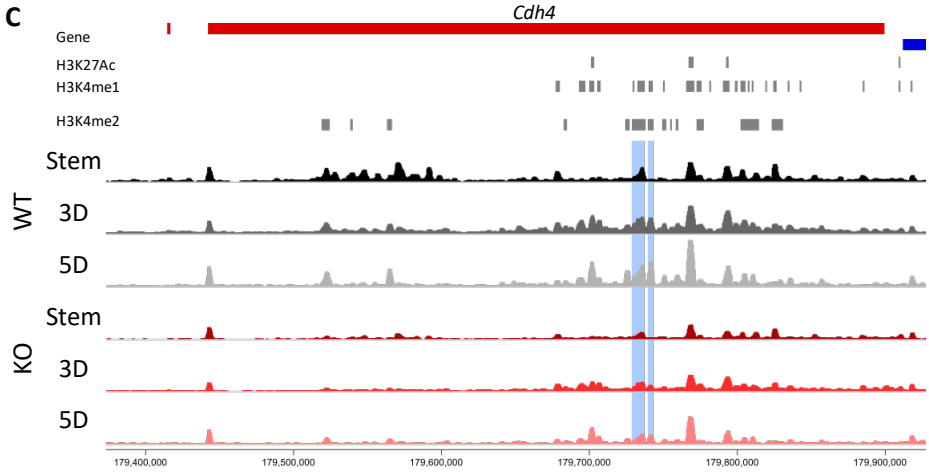


**Figure 4.11 Differentially enriched H3K4me2 peaks seem to behave similarly to DE H3K4me1 peaks**

**A.i** Scatter plot highlighting DE H3K4me2 peaks (LIMMA p<0.05) at 5D differentiation. **A.ii** Bean plots showing log2 enrichment of the DE peaks at 5D, WT on the bottom with *Kdm1b*$^{-/-}$ overlayed on top. Peaks that gained methylation in the knockouts are highlighted in red and orange and those that lost methylation are highlighted in blue. Peaks in orange and light blue did not coincide with a DE H3K4me1 peak. **B.** Results of HOMER Motif analysis of the highly genic DE peaks gaining methylation. **C.** Chromosome view of *Cdh4* highlighting general reduction in H3K4me2 in KDM1B null cells. Annotation tracks are genes (sense strand in red and antisense strand in blue), H3K27ac and H3K4me1 peaks from Chuong et al. 2013 and my consensus list of H3K4me2 peaks. Highlighted in blue are the differentially enriched H3K4me2 peaks (LIMMA p<0.05).

### 4.2.7 Genome-wide H3K4me3 and H3K36me3 profiles were unaffected by *Kdm1b* deletion

Analysis of my ChIP-seq datasets for H3K4me3 and H3K36me3 did not identify any differentially enriched peaks between WT and *Kdm1b*$^{-/-}$ TSCs at any time point. Figure 4.12 depicts the data obtained or H3K4me3 and shows that, once again, PCA separated the samples by day of differentiation (a), but not by *Kdm1b* genotype, as did hierarchical clustering (12b). Figure 4.12c shows scatter plots comparing WT and KDM1B null cells at stem, 3D and 5D differentiation, highlighting the lack of divergence in enrichment patterns.

Finally, I turned to the H3K36me3 ChIP-seq data. This histone modification is known to mark the gene body of actively transcribed genes, accumulating towards the 3' end. My data shows this distribution well in stem and 3D samples of both WT and *Kdm1b*$^{-/-}$ cells. However, in 5D samples, I saw a dramatic shift in the distribution (Figure 4.13a). Due to this profound re-distribution, PCA and hierarchical clustering of the H3K36me3 ChIP-seq data only separated the 5D samples from the stem and 3D samples (Figure 4.13b, c). PC1 was the main axis on which 5D samples were separated from stem and 3D conditions. PC2 appears to somewhat distinguish the stem from the 3D and 5D samples, placing the stem cell libraries in the upper two quadrants, however this was by no means absolute. Despite this considerable reshuffle in H3K36me3 distribution between stem/3D and 5D samples, no significant differences in enrichment were identified between WT and *Kdm1b*$^{-/-}$ cells.

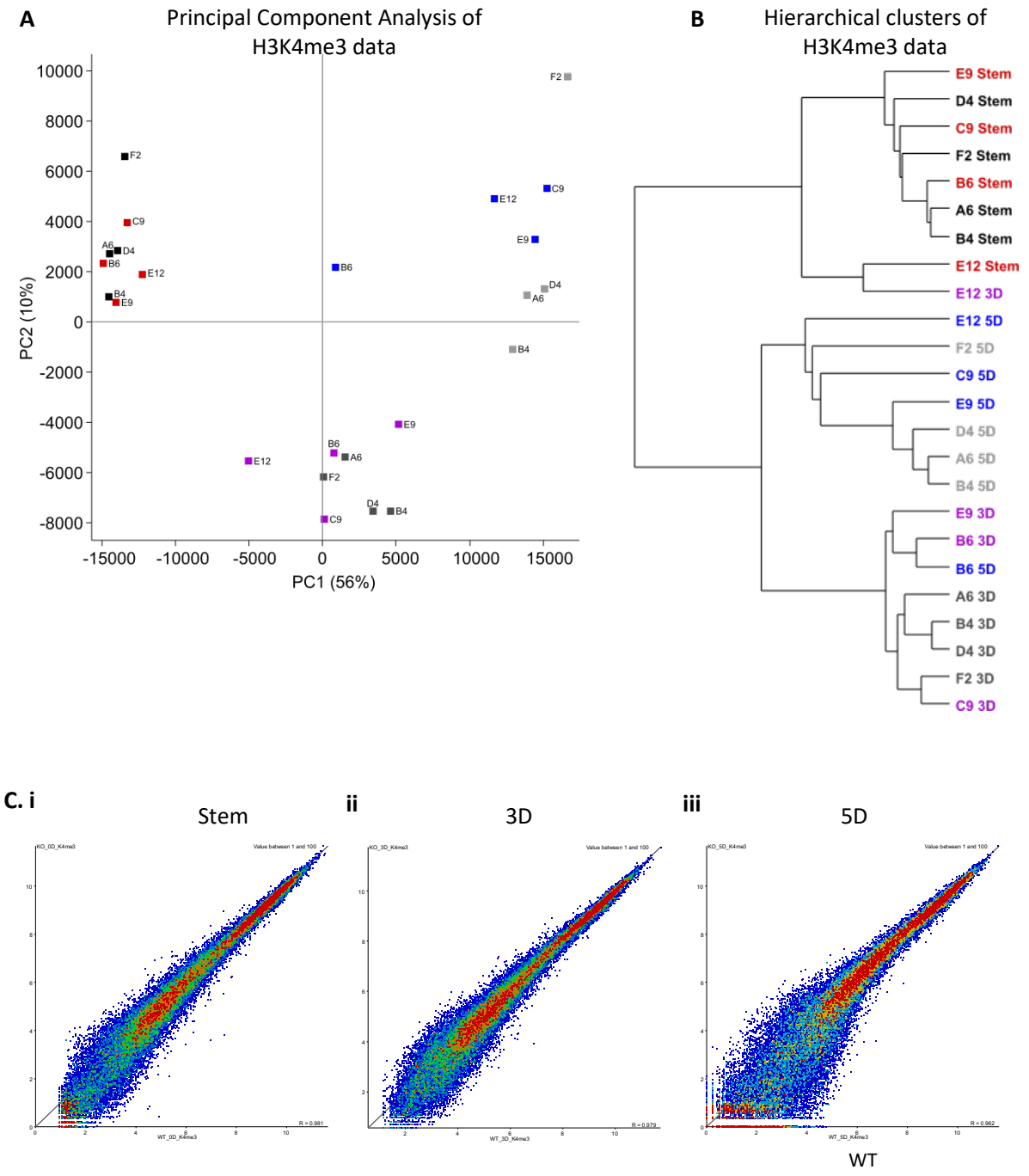
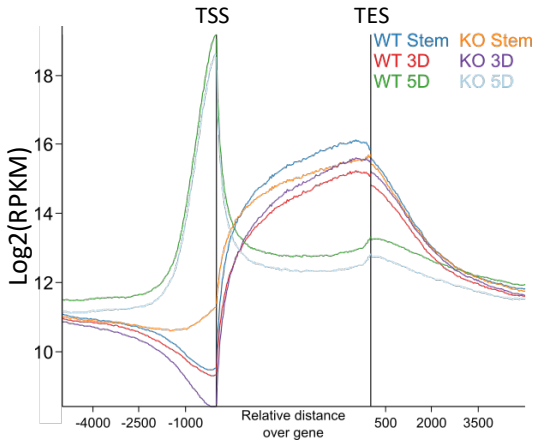


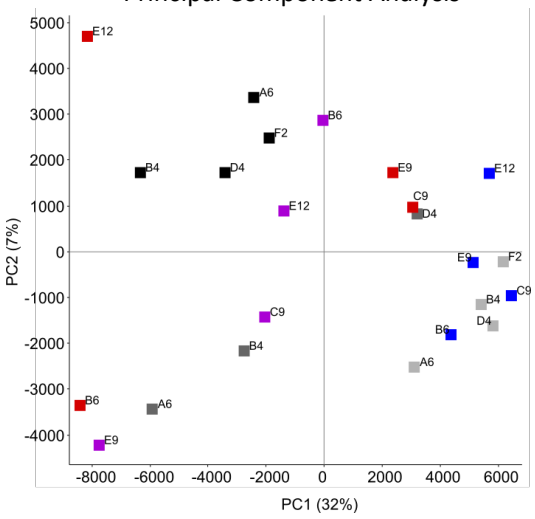

**Figure 4.12 H3K4me3 MACS2 called peaks were not differentially enriched at any time point between WT and KDM1B null cells**

**A**. PCA plot showing separation of H3K4me3 libraries by day of differentiation, but not genotype. This is mirrored by hierarchical clustering (**B**). **C.** Scatter plots comparing WT and KDM1B data at each time point, which showed no differentially enriched peaks.

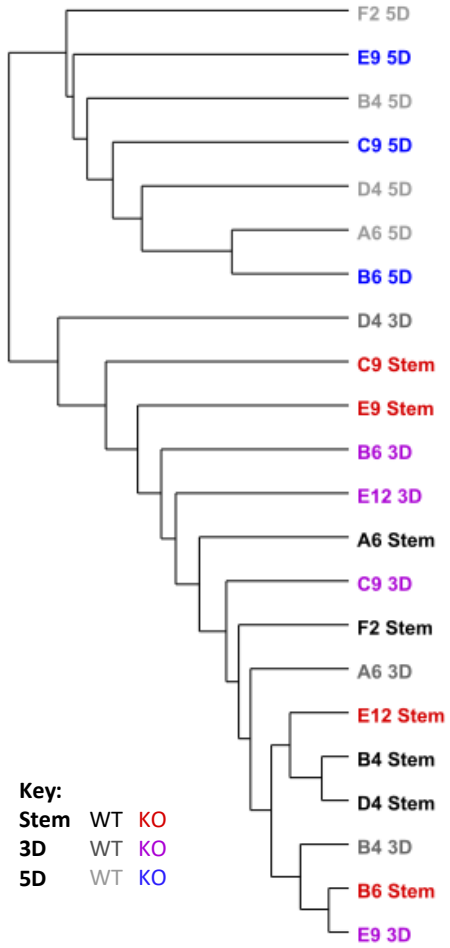
**A** Trend of 1kb probes over genes




**B** Principal Component Analysis




**C** Hierarchical clusters of 1kb running windows




**D** Scatter plot of 1kb probes at 5D differentiation

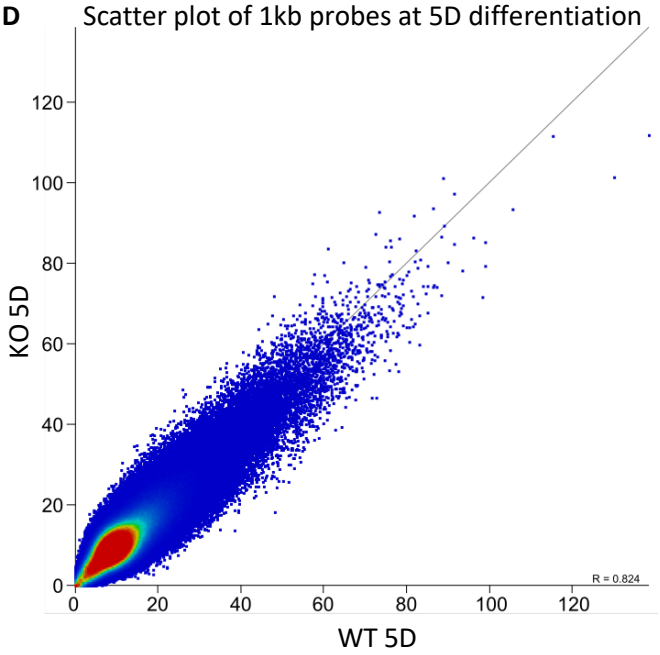


**Figure 4.13 H3K36me3 undergoes redistribution at 5D, but loss of KDM1B did not affect this mark *in vitro***

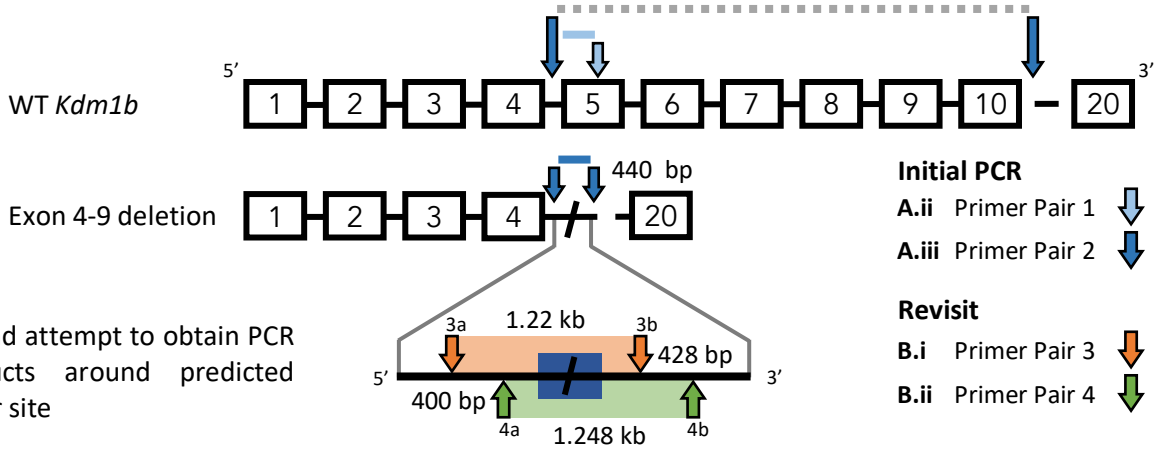
**A.** Probe trend plot demonstrates a striking redistribution in H3K36me3 reads at 5D differentiation in trophoblast in vitro. This occurs in both WT and *Kdm1b*$^{-/-}$ cells. **B.** PCA plot showing separation of 5D H3K36me3 libraries at 5D of differentiation, but less separation between stem and 3D libraries. This was mirrored by **C.** hierarchical clustering. **D.** Scatter plot comparing WT and KDM1B null data at 5D.

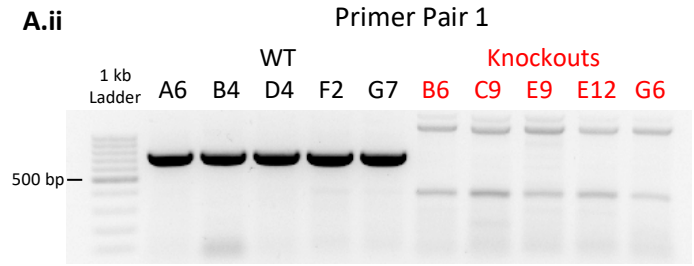
### 4.2.8 Chromosome 13

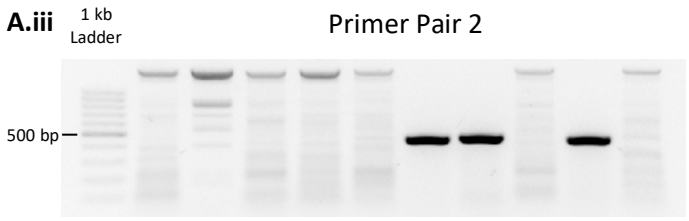
There was increased representation of chromosome 13 in both DE genes and differentially enriched H3K4me1/2 peaks. Firstly, what is vital to reiterate is that *Kdm1b* resides on chromosome 13. So, my initial concern was that there might have been a larger deletion than intended as a result of the CRISPR Cas9 knockout strategy that impacted the stability of the chromosome. Indeed, the initial diagnostic PCRs I performed were inconclusive for two of the five *Kdm1b*$^{-/-}$ clones that were later verified by western blot (Figure 3.3, 4.14). Therefore, my first point of action was to expand the area around the recombined join of *Kdm1b* introns 5 and 10 that was amplified by PCR (Figure 4.14a.i). The two new pairs of primers were designed to be off-set from the centre to provide more information as to which open end, any exonuclease activity occurred. Figure 4.14b. demonstrates that clone E9, while having increased DNA damage at the dsDNA breaks induced by Cas9 compared with clones B6, C9 and E12, was successfully repaired as PCR products were formed with both primer pairs 3 and 4. On the other hand, PCR was only successful with primer pair 4 for clone G6, with the much reduced PCR product of 650-700 bp. This suggests that there was relatively extensive DNA damage following cutting in intron 3 by CRISPR Cas9, such the PCR primer 4a binding site was disrupted or entirely lost in clone G6. The primer 4b binding site was likely preserved as it is downstream of the primer 3b, with which PCR was successful.

These results confirm that the CRISPR-Cas9 cut-induced non-homologous end-joining repair of the *Kdm1b* locus was successful in all five knockout clones, however it also potentially suggests that this only occurred once as all clones only have one PCR product. One possible reason for this is allele-specific sequence differences preventing annealing of the primers. However, as I have used several primer pairs, this seems unlikely. Indeed, an even larger deletion may have occurred on the second *Kdm1b* allele preventing primer binding, but this, too seems highly unlikely to have occurred in all five clones. Another possibility is that a larger translocation event occurred in the second allele. The final possibility is that these clones only have one copy of *Kdm1b*, and are perhaps monosomic for chromosome 13, despite the fact that all WT clones that were clonally expanded in parallel from the same starting pool of TS-Rs26 cells clearly did not have this problem.

**A.i** Initial diagnostic PCR test to identify possible *Kdm1*$^{b-/-}$ clones



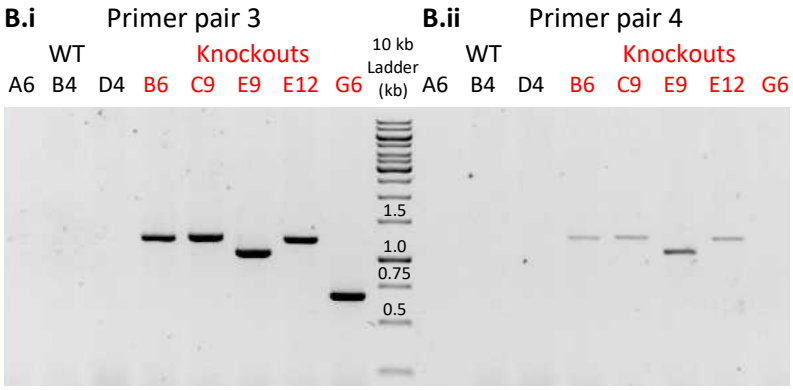




**Figure 4.14 PCR around the recombined join of *Kdm1b* shows successful repair of the dsDNA break**

**A.i** Schematic describing the initial diagnostic PCR strategy compared with the second attempt with PCR primer pairs 3 and 4. **A.ii-iii** The results previously shown in chapter 3: PCR products with primer pair 1 (**ii**), and primer pair 2 (**iii**). **B.** shows the PCR products with primer pair 3 (**i**) and 4 (**ii**).

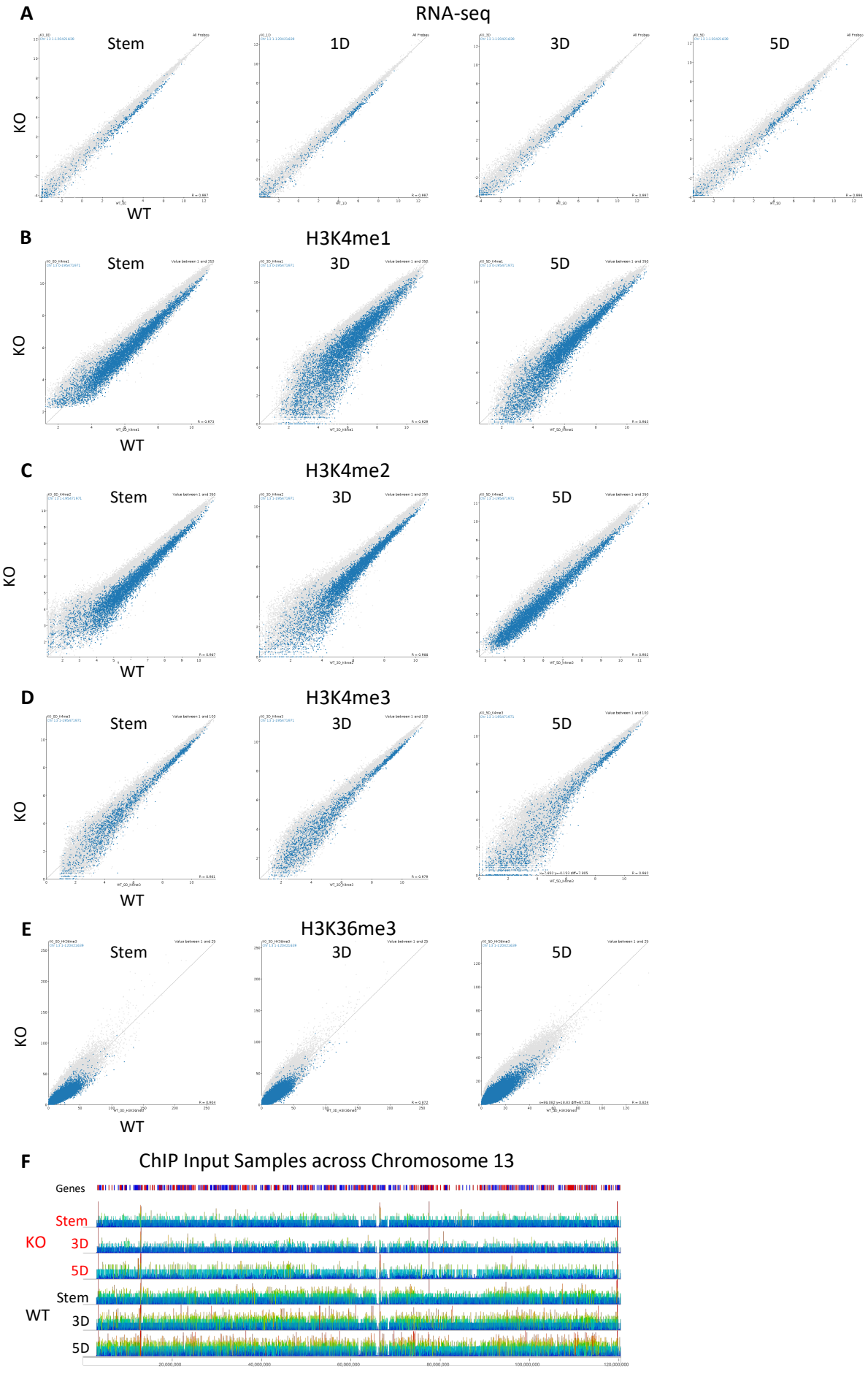

A
RNA-seq
Stem
1D
3D
5D
KO
WT
B
H3K4me1
Stem
3D
5D
KO
WT
C
H3K4me2
Stem
3D
5D
KO
WT
D
H3K4me3
Stem
3D
5D
KO
WT
E
H3K36me3
Stem
3D
5D
KO
WT
F
ChIP Input Samples across Chromosome 13
Genes
Stem
KO
3D
5D
Stem
WT
3D
5D

**Figure 4.15 Loss of chromosome 13 in all sequencing data towards the WT**

Scatter plots for **A** RNA-seq, and ChIP-seq H3K4me1 (**B**), H3K4me2 (**C**), H3K4me3 (**D**) and H3K36me3 (**E**), highlighting all probes on chromosome 13. **F.** ChIP input samples with KDM1B null samples showing reduced reads across chromosome 13 (probes are 1 kb running windows Log2RPKM).

When highlighting chromosome 13 in all sequencing datasets, the entire chromosome was underrepresented, in terms of reads, at all time points when comparing WT and *Kdm1b*$^{-/-}$ replicate sets (Figure 4.15). As a result of this, I decided to utilise the input control libraries of the ChIP-seq experiments to try to assess ploidy in the clones. After partitioning the genome into 25 kb bins, I normalised the read count to the average for each clone, to produce a ratio indicative of ploidy. As TS-Rs26 cells are male, I was expecting autosomes to have a ratio of 1 and chromosomes X and Y to have a ratio of 0.5. Plotted in Figure 4.16a-c are the box whisker plots for each chromosome at each time point. WT clones (i) and *Kdm1b*$^{-/-}$ clones (ii) show similar noise around the balanced average of 1 throughout differentiation. As evidenced by these plots, chromosome 13 (green) was consistently under-represented in the *Kdm1b*$^{-/-}$ samples, throughout *in vitro* differentiation. Chromosome 13 had a median ratio of 0.75 compared to 1.0 in WT cells. Figure 4.16b demonstrates that copy number of chromosomes 1, 13 and X appeared to be stable throughout differentiation, remaining at a ratio of 1.0, 0.75 and 0.5, respectively, in *Kdm1b*$^{-/-}$ cells.

If the TS-Rs26 cells were diploid, a ratio of 0.75 would mean the *Kdm1b*$^{-/-}$ cells had a median number of copies of chromosome 13 of 1.5 (instead of 2). As the whole length of the chromosome had fewer reads in the *Kdm1b*$^{-/-}$ samples compared to the corresponding WT time point, I concluded that the 'half' was not coming from a second copy of one region of the chromosome. One potential explanation was that the KDM1B null cells were heterogeneous for chromosome 13, with some cells having one copy and other cells having two copies. However, a simpler explanation is that TS-Rs26 cells are in fact tetraploid and the *Kdm1b*$^{-/-}$ cells contain three copies of chromosome 13. In order to assess this, I decided to perform metaphase spreads of all ten clones, as well as the starting population of TS-Rs26 cells from which my clones were derived.

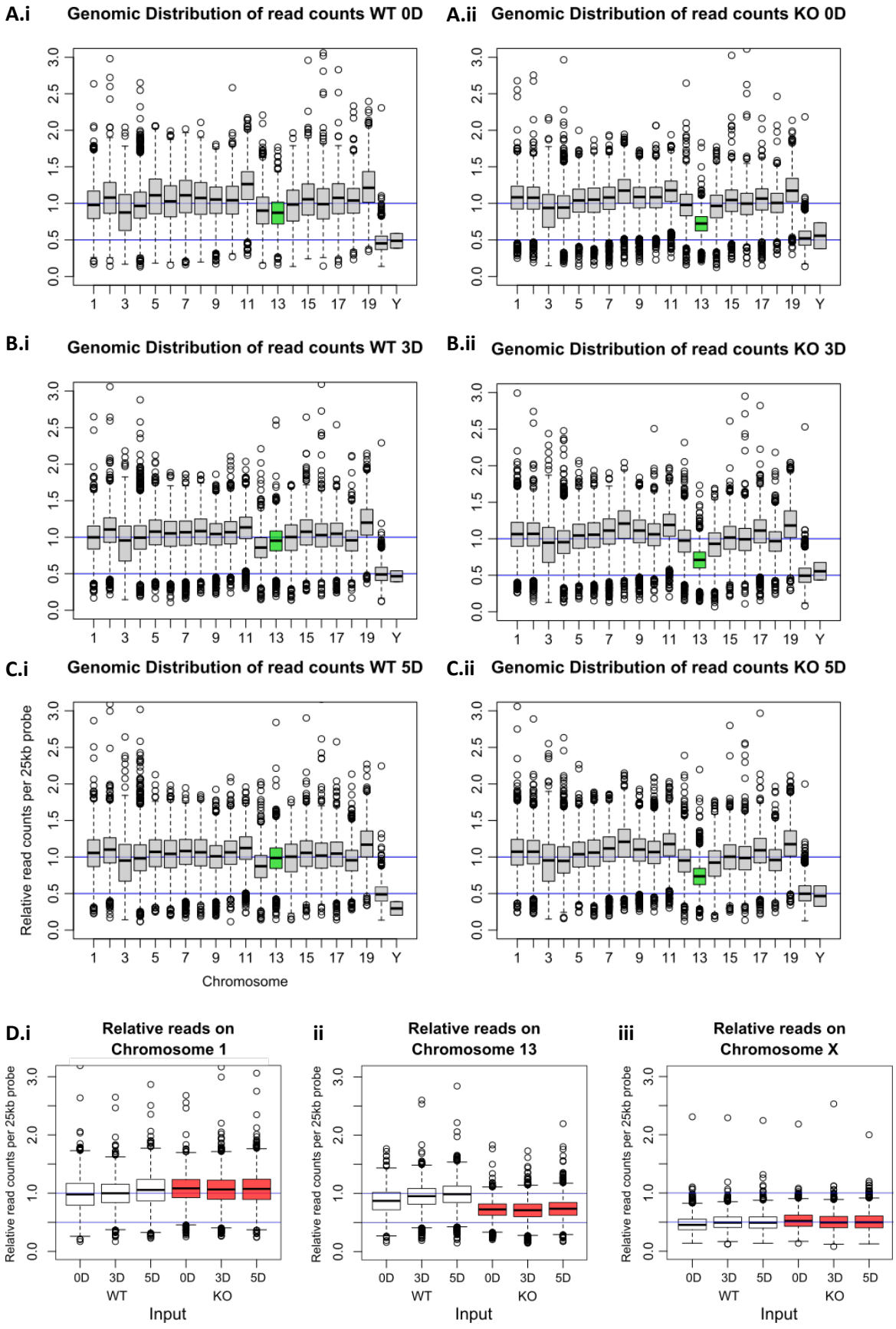

A.i
Genomic Distribution of read counts WT 0D
A.ii
Genomic Distribution of read counts KO 0D
B.i
Genomic Distribution of read counts WT 3D
B.ii
Genomic Distribution of read counts KO 3D
C.i
Genomic Distribution of read counts WT 5D
C.ii
Genomic Distribution of read counts KO 5D
Relative read counts per 25kb probe
Chromosome
D.i
Relative reads on Chromosome 1
ii
Relative reads on Chromosome 13
iii
Relative reads on Chromosome X
0D 3D 5D 0D 3D 5D
WT KO
Input

**Figure 4.16 Genomic distribution of reads in ChIP-seq input libraries.**

Graphs showing reads per 25kb bins relative to the genome average for each sample, these are plotted as box and whicker plots for each chromosome: **A.** Stem, **B.** 3D and **C.** 5D, WT (**i**) and *Kdm1b*$^{-/-}$ (**ii**) replicate sets. **D**. Box and whisker plots showing relative reads in chromosomes 1 (**i**), 13 (**ii**) and X (**iii**) across genotype and time course.

### 4.2.9 TS-Rs26 cells were triploid and KDM1B null cells show increased variation in chromosome number

Figure 4.17a shows representative images of metaphase spreads for all 10 clones and the original population of TS-Rs26 cells for comparison. What is immediately evident is that the cells contain far more than the expected 40 chromosomes. Figure 17b plots the mean chromosome number in the five cells measured, for each clone. Interestingly, the TS-Rs26 cells showed very low variation compared to the clonal lines. This indicates that the process of single cell sorting and expansion stressed the cells and led to variation in ploidy.

The mean number of chromosomes in the starting population of TS-Rs26 cells was 60, which is denoted by the black line. The green dashed lines at y= 0.3 and 0.67 indicate 20 and 40 chromosomes, respectively and the blue and red lines at y = 0.5 and 2 indicate 30 and 120 chromosomes, respectively. There was vast variation in the chromosome numbers recorded. This is in part because I was unable to achieve maximal spreading of the chromosomes, so the counts are not as accurate as I would have liked.

Overall, the *Kdm1b*$^{-/-}$ clones had a significantly higher number of chromosomes by 2-way ANOVA with Giesser-Greenhouse correction $p<0.05$, however, this was driven by clone C9, which appeared to be tetraploid relative to the starting population (6n). Although, three of the other *Kdm1b*$^{-/-}$ clones did also contain cells with very high chromatid number. One of the mitotic G7 cells recorded also had over 120 chromatids, indicating that this is something that can occur in WT TS-Rs26 cells, although seemingly loss of KDM1B, or CRISPR itself, increased its likelihood.

The main goal of this experiment was to ascertain the ploidy of the TS-Rs26 cells in order to gauge the likely copy number of chromosome 13 in the *Kdm1b*$^{-/-}$ clones. This, in my opinion, has been successful, and it appears that the cells are triploid. The ratio of 0.75 (Figure 4.16) still is hard to interpret as it would correspond to 2.25 copies of chromosome 13. However, it might indicate that the *Kdm1b*$^{-/-}$ cells were heterogeneous for chromosome 13 copy number. What is harder to ascertain is whether it is loss of KDM1B that induces this apparent instability of chromosome 13, or CRISPR Cas9-induced DSBs.

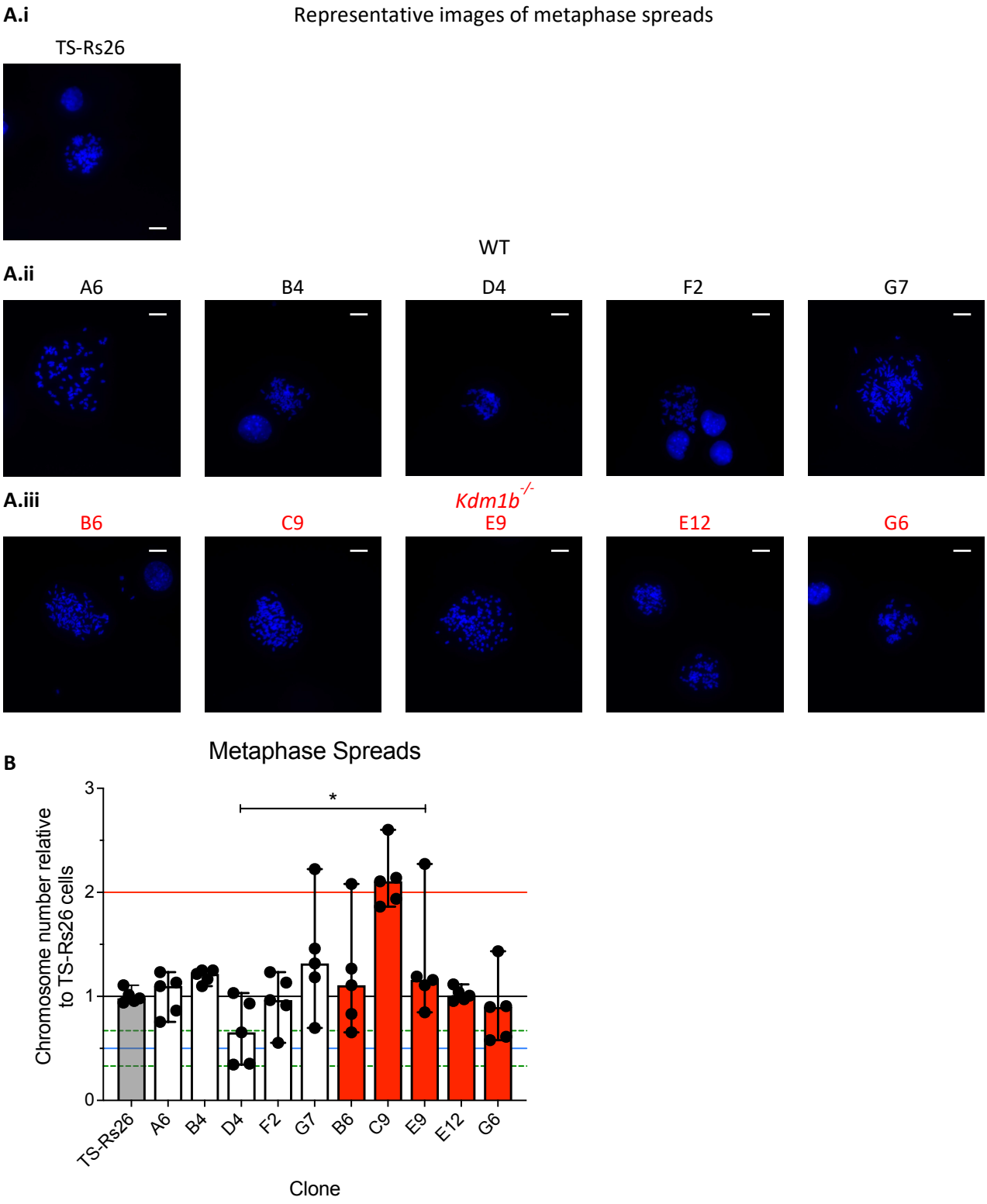



**Figure 4.17 *Kdm1b*$^{-/-}$ clones showed increased variation in chromosome number**

**A.** Representative images of metaphase spreads performed on the starting population of TS-RS26 cells **(i)** and each WT **(ii)** and *Kdm1b*$^{-/-}$ **(iii)** clone derived for this study. Scale bars are 10 μm. **B.** Sister chromatids of five mitotic cells were counted for each sample and these are plotted relative to the TS-Rs26 sample. Bars show median with range indicated, individual cells represented by points. Statistical analysis was by ANOVA comparing WT and KO clones, followed by Giesser-Greenhouse correction *p<0.05.

In addition to cells consistently having more than 40 chromosomes, when looking at the genomic distribution of reads in the input sequencing data, there were several regions with depleted signal. This included a 40 Mbp region at the 5' end of chromosome 3, a 3 Mbp region at the 3' end of chromosome 9 and a 5.5 Mbp region of chromosome 11, which has been previously noted (Senner *et al.*, 2020). The fact that these regions were consistently under-represented across all clones, suggests that one or more copies of these regions are deleted in TS-Rs26 cells. This would need to be considered for any future studies in these cells involving genes or cis-regulatory elements within these regions.

## 4.3 Discussion

In this Chapter, I describe an in-depth analysis of *Kdm1b*$^{-/-}$ TS-Rs26 cells' transcriptomic and epigenomic changes across differentiation to 5D. I assessed four histone modifications that are of key relevance in the context of KDM1B biology, namely H3K4me1, H3K4me2, H3K4me3 and H3K36me3.

Overall, the RNA-seq analysis detected surprisingly few transcriptionally de-regulated genes. Even genes previously identified as differentially expressed by RT-qPCR and, in the case of LSD1, additionally by western blot, were not identified as DE in the RNA-seq data. One possible explanation for these discrepancies is that RT-qPCR is more sensitive than RNA-seq as it is designed to amplify a single cDNA per reaction compared to measuring the expression of thousands of genes with transcriptomic analysis. With the expression of *Esrrb*, particularly, I saw an approximate 40 % reduction in expression, which would correspond to a difference in Log2RPKM of around 1, which was not evident in the RNA-seq data. However, looking at the H3K36me3 data, I noticed a decreased signal in the KDM1B null stem cells compared to WT, potentially indicating less transcription in these clones (data not shown).

In terms of the epigenomic rearrangements that take place as a function of TSC differentiation and *Kdm1b* loss, my data reveal a progressive gain of H3K4me1 and H3K4me2 peaks in the KO cells. This nicely corroborates my earlier western blot data that detected a first, subtle accumulation of H3K4me1 levels on the global scale. The fact that these differentially methylated H3K4me1 and H3K4me2 peaks were only robustly identified at 5D makes sense. KDM1B expression levels continue to increase across the 5D differentiation window (Figure 3.1), which explains why any effects on the chromatin landscape are most significant at this time point. According to Fang et al.'s study (2010), KDM1B acts in complex with elongating RNA PolII to demethylate H3K4 within the gene body of actively transcribed genes to prevent spurious transcription. My data indicate that KDM1B likely forms the same complex in TS-Rs26 cells, as over 98 % of differentially enriched H3K4me1 peaks and over 97 % of differentially enriched H3K4me2 peaks, which gained methylation in knockout clones, were found within the gene body of actively transcribed genes. The fact that the increase in methylation level of these potential intragenic

enhancers did not coincide with changes in expression of their host genes, does not preclude them from being regulatory regions. Indeed, previous studies have demonstrated that intragenic enhancers can regulate transcription of either their host gene or nearby genes (Birnbaum *et al.*, 2012; Cinghu *et al.*, 2017). Further, relative level of H3K4me1 signal does not necessarily correlate with activating power of enhancers, but rather their TF occupancy and long-range contacts (Dogan *et al.*, 2015; Dorighi *et al.*, 2017).

In order to ascertain whether these peaks are acting as enhancers, functional studies would have to be performed. Traditional methods of excising potential regulatory region(s) and assessing knock-on effects on the transcriptome are problematic when looking at intragenic enhancers, as their excision could disrupt expression of the host gene, regardless of the enhancer activity. Similarly, utilisation of epigenetic editing methods such as CRISPR interference (CRISPRi, Larson et al., 2013), would also likely effect host gene expression regardless of enhancer activity; either by spreading of the repressive histone modifications introduced, or by steric hindrance of RNA PolII progression through the gene body due to catalytically dead Cas9 binding the enhancer (Qi *et al.*, 2013).

An alternative method to probe enhancer activity could be utilisation of a luciferase reporter assay to ascertain whether these regions have transcriptionally activating characteristics. Another method of identifying whether these regions are likely enhancers, would be to assess levels of other epigenetic marks, including H3K27ac during *in vitro* differentiation. Yang et al., (2019) demonstrated that intragenic enhancers in the human genome are enriched for H3K79me2 and depleted for H2A.Z compared to intergenic enhancers. A similar pattern may be present in the mouse and would help to distinguish active intragenic enhancers. Another characteristic of active enhancers is bi-directional transcriptional initiation and the production of enhancer RNA (eRNA, Kim et al., 2010). Performing total RNA-seq on TSCs in stem conditions and at 5D differentiation could enable identification of any enhancer-specific transcripts originating from these differentially enriched H3K4me1 peaks. However, these experiments would not provide insight as to which genes these potential enhancers were regulating in the trophoblast.

Extending the time course and performing further ChIP-seq and RNA-seq experiments could uncover any downstream effects of the differentially enriched H3K4me1 and H3K4me2 peaks I have identified. It is important to acknowledge that *Kdm1b*$^{-/-}$ mice were grossly normal (D N Ciccone *et al.*, 2009), so any effect of loss of KDM1B on the developing trophoblast are unlikely to be catastrophic. The fact that the majority of the differentially enriched H3K4me1 peaks showed reducing signal during WT differentiation, it is possible that the sustained signal in *Kdm1b*$^{-/-}$ cells could coincide with retained transcriptional activation at later time points in differentiation.

It is certainly interesting that there were several differentially methylated H3K4me2 peaks that did not overlap with, nor were even within the same gene as any differentially enriched H3K4me1 peaks. It would be interesting to extend the time course for a longer time period, to assess how these regions change during further differentiation.

Given the western blot analysis indicating a global accumulation of H3K36me3 and to some extent, H3K4me1, the use of a spike-in of DNA would have been useful in normalising the amount of histone modifications being captured for each ChIP-seq experiment (Bonhoure *et al.*, 2014; Orlando *et al.*, 2014; Egan *et al.*, 2016). This would have enabled a more quantitative assessment of each histone modification in the *Kdm1b*$^{-/-}$ cells compared to WT. As executed herein, these data can be compared between genotype insofar as an increase in signal at one locus can be quantitated relative to all reads, however, not to the absolute level of histone modification captured. On the one hand, being able to normalise the changes in modification abundance would have provided further insight as to the effect of loss of KDM1B on the steady-state levels of the histone modifications assessed. However, with the appropriate normalisation performed herein, these data still offer invaluable insight into the locus-specific changes to the epigenome during *in vitro* differentiation.

It is tempting to draw parallels between the loss of H3K4me1/2 within the *Tshz2* and *Cdh4* genes on chromosome 2, and studies in drosophila and *S. pombe*, which identified LSD1 paralogues being directly involved in hetero-/euchromatic boundary formation (Lan *et al.*, 2007; Rudolph *et al.*, 2007). Lan et al. demonstrated that *spLsd1* directly acts on H3K9 methylation and its loss leads to spreading of heterochromatin in *S. pombe*. Contrastingly, Rudolph et al. found that in drosophila, SU(VAR)3-9-dependent demethylation of H3K4 was necessary for the spread of heterochromatin. Indeed, in the regions upstream of both *Tshz2* and *Cdh4*, there are WT H3K9me3 peaks, indicating that KDM1B may play a role at these loci in preventing the spread of heterochromatin.

I also assessed H3K4me3 profiles but observed no changes between WT and KO cells across differentiation. This was not unexpected as tri-methylated lysine is not a substrate of KDM1B. Further, as there was very little impact on gene expression, I was not expecting a dramatic change to H3K4me3 patterns, a mark of active promoters. Finally, my H3K36me3 genomic occupancy profiles data did not reveal any differences between the genotypes, but did reveal a striking shift in localization away from the enrichment across gene bodies to a sharp peak over the transcriptional start sites at 5D of differentiation. This is an unexpected and utterly surprising finding, the significance of which remains unclear at the moment.

In addition to the transcriptomic and epigenomic analysis, I explored in some detail the peculiar finding of a chromosomal imbalance of chromosome 13 that is specific to the *Kdm1b* KO cells and appears to

drive many of the transcriptional and chromatin changes observed. In the course of my analysis addressing this phenomenon, I determined that the TS-Rs26 starting cell population is polyploid, although the count of 5 chromosomal spreads per cell line is somewhat limited to make firm conclusions on this point. Yet it is clear that single cell sorting and expansion introduces a far greater variability of ploidy in TSCs. The acquisition of a polyploid state through endoreduplication is a characteristic feature of trophoblast. As such, polyploidization as well as the uneven over- or underreplication of chromosomal regions is well tolerated in differentiated trophoblast; it is indeed a required feature that endows specific trophoblast cell types with their unique functions. By comparison, polyploidization and/or chromosomal imbalances lead to apoptosis in most other cell types, notably in ESCs, as the developmental stem cell counterpart of TSCs.

It also seems clear that targeting *Kdm1b* introduces a still higher level of chromosomal variability than single cell cloning alone. This "*Kdm1b*-effect" specifically affects the gene's host chromosome, 13. This could result from the targeted gene itself, *Kdm1b*, the other genomic elements contained within the deleted region, or with the gRNAs used in the nature of off-target effects (Hsu *et al.*, 2013; Le *et al.*, 2013; Mali *et al.*, 2013)(Fu *et al.*, 2013; Pattanayak *et al.*, 2013). Such chromosomal aberrations were not noted in many previous CRISPR-Cas9-driven gene deletion strategies in TSCs, targeting for example genes such as *Bap1*, *Plet1*, *Tet1*, *Tet2*, *Nubpl*, *Crb2* and numerous others (Murray, Sienerth and Hemberger, 2016; Chrysanthou *et al.*, 2018; Perez-Garcia *et al.*, 2018), although it is important to note that this may not have been assessed. Chromosome 13 is of particular relevance for trophoblast as it contains multiple trophoblast-specific gene clusters that have arisen by repeated gene duplication; these entail the Prolactin, Cathepsin, and Serpin clusters (Deussing *et al.*, 2002; Kaiserman *et al.*, 2002; Wiemers *et al.*, 2003). Many of these genes are uniquely expressed in trophoblast (Jackson-Grusby *et al.*, 1988; Orwig *et al.*, 1997; Soares *et al.*, 1998; Deussing *et al.*, 2002). As such, chromosome 13 is already predisposed to large-scale rearrangements that specifically affect genes important to the trophoblast lineage. The targeting of *Kdm1*b, itself located on chromosome 13 but not in proximity to any of these clusters, appears to exacerbate this effect.

# 5 Chapter Five

## Integrated transcriptomic and epigenomic analysis of WT trophoblast differentiation

## 5.1 Introduction

The main focus of this work has been on the effect of the loss of KDM1B on TSCs and differentiation *in vitro*. This chapter will focus on the wild-type setting, with the aim of integrating transcriptomic and epigenomic data to identify novel features of trophoblast differentiation. The fact that I have data at several time points during in vitro differentiation meant that I was able to assess dynamic changes to these layers, and link certain aspects both between omics and across time points of differentiation, analysis that has not previously been done in *in vitro* differentiating TSCs.

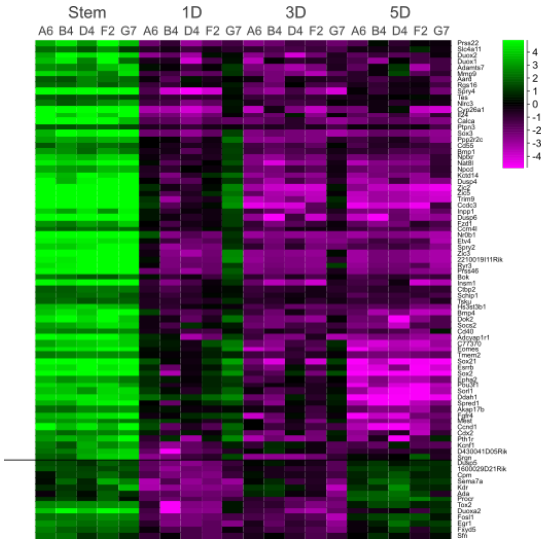
## 5.2 Results

### 5.2.1 Early markers of trophoblast stem cell differentiation were robustly down-regulated as the cells differentiated *in vitro*

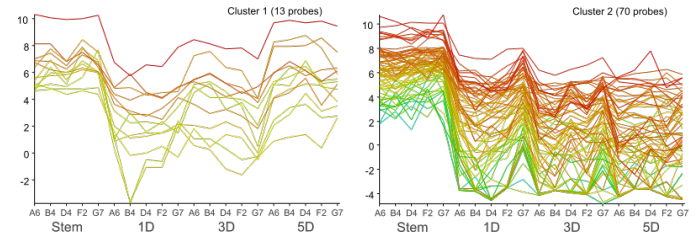
When identifying differentially expressed genes across the time course of differentiation from stem cells to 1D, 3D and 5D differentiated cells, I decided to prioritise the step-wise comparisons. Interestingly, I noticed that the number of genes identified as differentially expressed between stepwise time points, by DESeq2, were very different. Between stem and 1D differentiation, there were 4925 DE genes and between 3D and 5D, there were 5311 DE genes, compared to just 1968 genes differentially enriched between 1D and 3D differentiated cells. This implies that two waves of large transcriptional changes occur, firstly as the cells exited multipotency and secondly between 3D and 5D differentiation, corresponding to the emergence of terminally-differentiated TGCs. Between 1D and 3D differentiation, many SynT genes become highly expressed and the cells begin undergoing syncytialisation between these time points *in vitro*. Additionally, cells of the spongiotrophoblast are specialising before differentiating to TGCs. The smaller list of DE genes between 1D and 3D differentiated cells suggests that similar pathways of genes are involved in regulating early SynT formation and spongiotrophoblast specification. Larger transcriptional changes are required to exit multipotency and for later stages of differentiation, i.e between 3D and 5D compared to the intermediate 1D-3D time points.

Heatmaps of differentially expressed genes identified by DESeq and Intensity difference ($p<0.05$) with a difference in log2RPM greater than 2, are plotted in Figure 5.1 comparing stem and 1D differentiating cells. Of the genes downregulated between stem and 1D differentiation, two patterns of subsequent expression were identified. Cluster 2 genes included canonical TSC TFs, *Esrrb*, *Cdx2* and *Eomes*, and were robustly down-regulated throughout the differentiation time course. Conversely, cluster 1 genes were temporarily downregulated, with expression increasing between 3D and 5D differentiation (5a.ii)

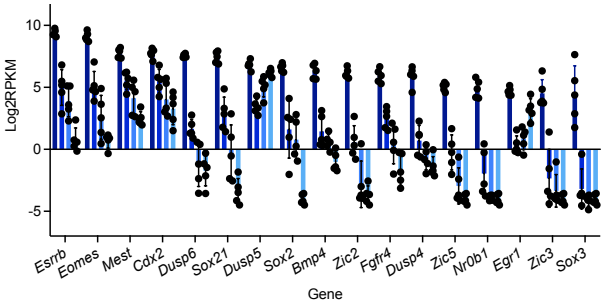
**A.i** Downregulated



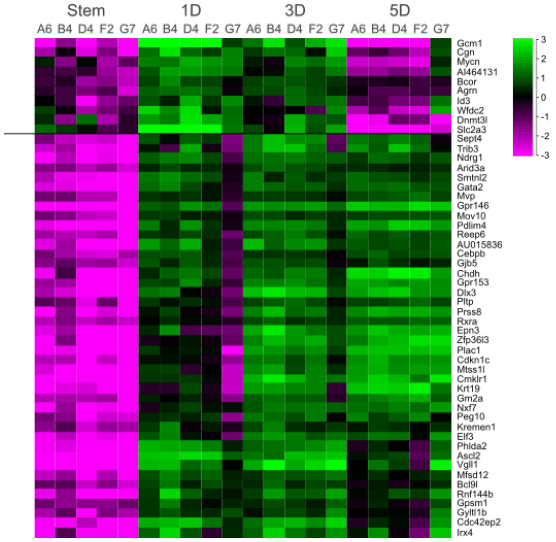

**A.ii** Log2RPM Expression



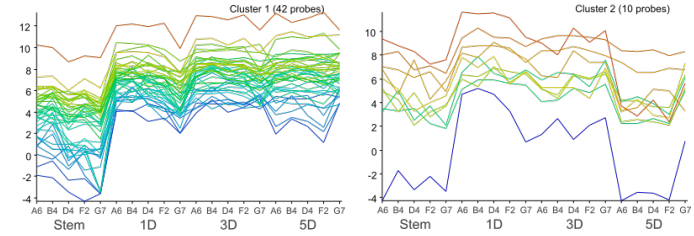

**A.iii** Expression of select differentiation markers



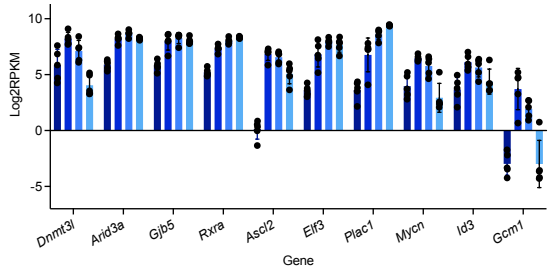

**B.i** Stem vs 1D Upregulated Genes: Difference greater than Log2




**B.ii** Log2RPM Expression




**B.iii** Expression of select differentiation markers




**C** Significantly enriched Biological Processes

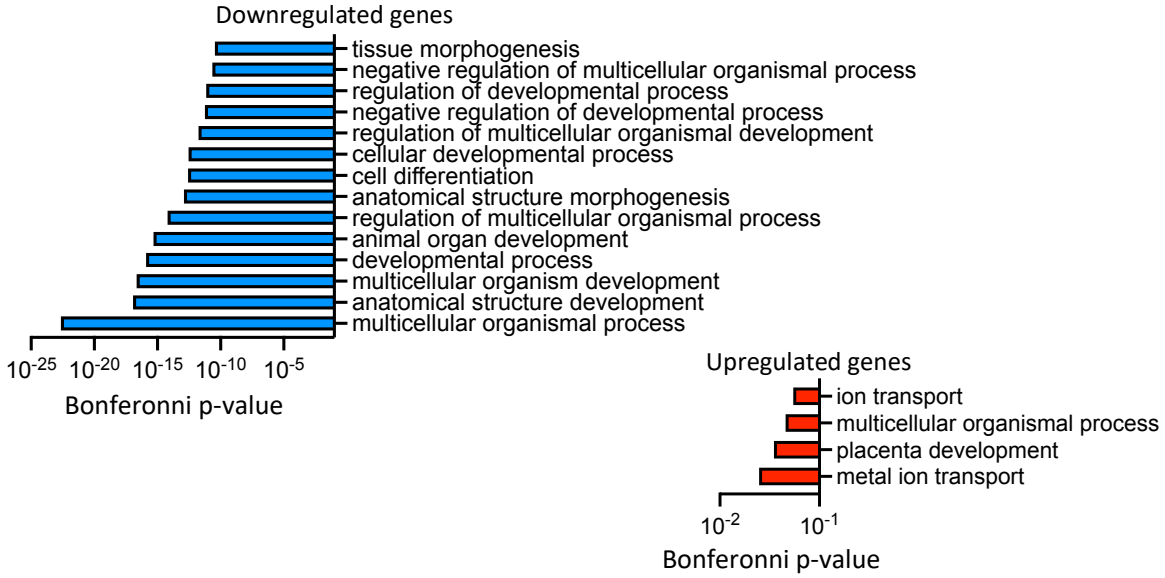


**Figure 5.1 Transcriptomic changes between stem and 1D differentiation reflected early differentiation of trophoblast**

**A.i** Heatmap showing normalised expression of down-regulated stringently identified DE genes (DESeq2 and intensity difference p<0.05). **A.ii** Line graph showing expression of down-regulated genes in cluster 1 and cluster 2 identified by hierarchical clustering (R>0.7). **A.iii** Bar graph showing expression (Log2RPKM) of down-regulated stem cell markers. **B.i** Heatmap showing normalised expression of up-regulated stringently identified DE genes (DESeq2 and intensity difference p<0.05). **B.ii** Line graphs showing expression of genes in cluster 1 and cluster 2 identified by hierarchical clustering (R>0.7). B.**iii** Bar graph showing up-regulated expression (Log2RPKM) of early differentiation markers. **C.** Go terms (Biological Process) enriched in DE gene lists between stem and 1D differentiated cells.

and included *Egr1* an important factor in placental vascularisation (Guo *et al.*, 2014; Chen *et al.*, 2021), and *Dusp5*, which has been shown to play a role in trophoblast migration and invasion in human placentation (Zhang et al., 2021, Figure 5.1a.iii).

Those genes up-regulated as the cells exit multipotency (Figure 5.1b) also showed two patterns of expression. Expression of Cluster 1 genes such as the spongiotrophoblast driver, *Ascl2* and essential trophoblast TFs *Rxra*, *Arid3a* and *Gata2* were maintained throughout differentiation (Figure 5.1b.ii and iii). Cluster 2 showed transient induction of gene expression and included syncytiotrophoblast TF *Gcm1*, as well as *Dnmt3l*. These genes were all drivers of the significantly enriched GO term (biological process), *placental development* (Figure 5.1c).

Interestingly, wild-type clone G7 shows global delay in down-regulation of stem markers and up-regulation of differentiation markers (Figure 5.1a and b). Indeed, this clone appeared to express stem cell genes more highly and suppressed expression of early differentiation genes in stem conditions, more effectively than other clones, suggesting this clone was inhabiting a somewhat enhanced stem cell state compared to other clones. This clone could be an interesting candidate for more in-depth characterisation in the future.

### 5.2.2 During differentiation the majority of differentially expressed genes are up-regulated

Very few differentially expressed genes identified by DESeq2 and/or intensity difference (p<0.05) were found to be downregulated between 1D and 3D differentiation (Figure 5.2a, b). This could mean that the majority of the gene networks that are required to be silenced for *in vitro* differentiation were sufficiently silenced between stem and 1D differentiation. This imbalance could also be due to the missing 2D time point in which there might be transient silencing of genes whose expression was then re-activated at 3D. These down-regulated genes include transcription factors, *Id2* and *Tbx3*, both of which showed transient up-regulation at 1D differentiation.

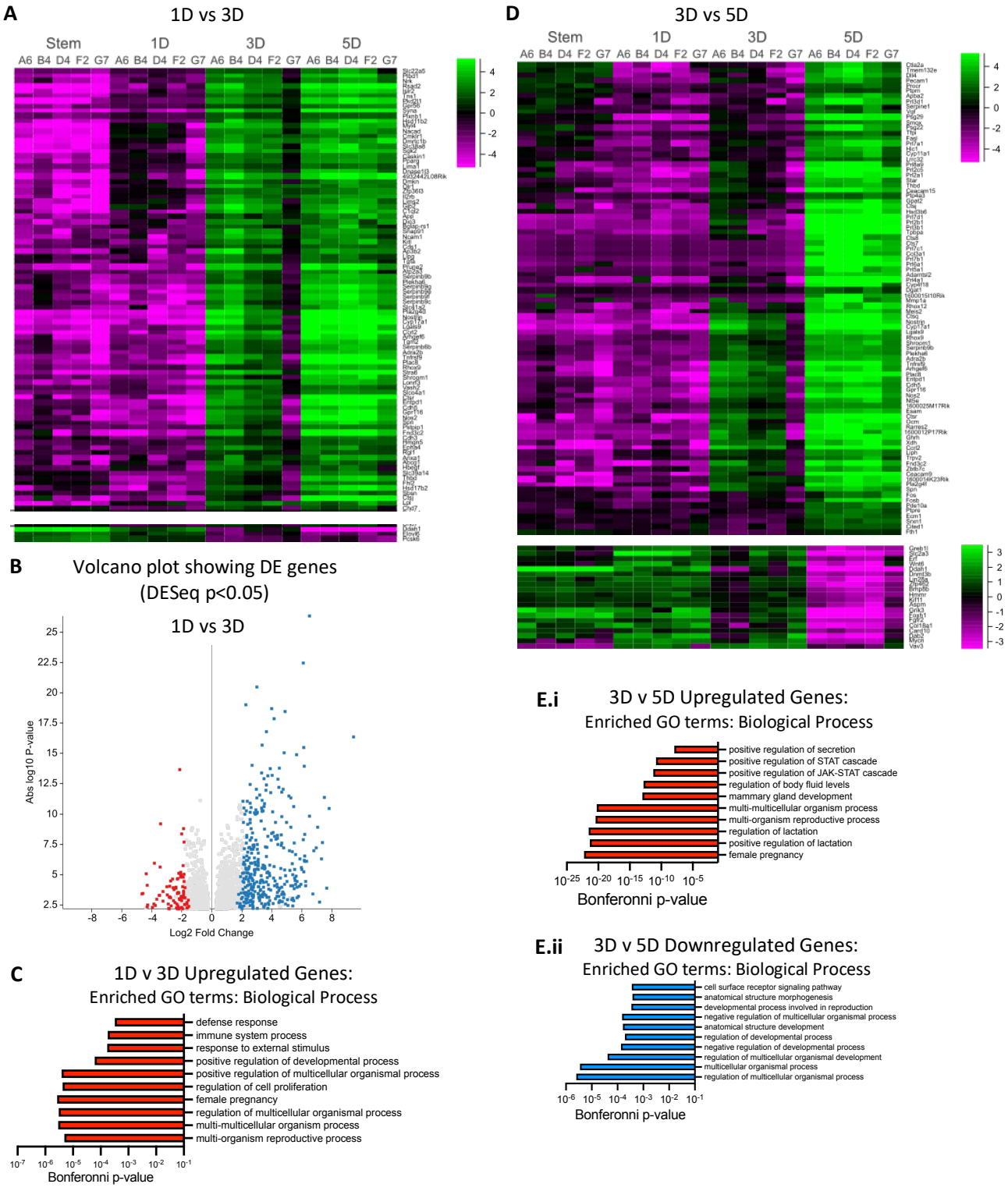



**Figure 5.2 Gene expression during differentiation**

**A.** Heatmap showing normalized expression of differentially expressed genes between 1D and 3D differentiated cells identified by DESeq2 and intensity difference filter (p<0.05). **B.** Volcano plot showing up- and down-regulated genes between 1D and 3D differentiated cells (DESeq2, p<0.05). **C.** GO terms (biological process) enriched by up-regulated genes at 3D. **D.** Heatmaps of normalized expression of differentially expressed genes between 3D and 5D differentiated cells identified by DESeq2 and intensity difference filter (p<0.05). **E.** GO terms (biological process) enriched by up-regulated genes at 5D (**i**) and down-regulated at 5D (**ii**).

Conversely, there were many up-regulated genes between 1D and 3D differentiation. Contrastingly to those patterns of expression identified with the stem vs 1D comparisons, all up-regulated genes at 3D continued to be expressed to 5D. This indicates that transcriptional networks established at 3D were maintained during later stages of differentiation. Up-regulated genes at 3D included several TGC-specific genes such as *Plac8* as well as several prolactin, serpin and cathepsin genes present in the over-replicated TGC regions on chromosome 13. Also included were *Syna* and *Gjb2*, which are expressed in the syncytiotrophoblast. Together these data confirm that there was differentiation towards spongiotrophoblast, TGC and syncytiotrophoblast lineages during unsupervised differentiation. Similarly to DE genes between stem and 1D differentiation, clone G7 appeared to be delayed in its up-regulation of DE genes between 1D and 3D.

Looking at the GO terms (biological pathway) enriched by the up-regulated DE genes between 1D and 3D differentiation, one of the most significantly enriched pathways was *Female pregnancy*, which was driven by the prolactin genes up-regulated at this time point (Figure 5.2c).

Finally, when comparing 3D and 5D differentiated cells, there were four times as many up-regulated genes as down-regulated genes (DESeq2 and intensity difference, $p<0.05$, Figure 5.2d). Many of the up-regulated genes were again located in the overamplified TGC regions, with several *Prl* genes being induced between 3D and 5D differentiation. The sinusoidal TGC (sTGC) marker, *Ctsq* was also induced. *Female pregnancy* was once again one of the most enriched biological pathways by up-regulated 5D genes; in addition to this was, *regulation of lactation* and *regulation of body fluid levels*, all of which were mainly driven by the up-regulated *Prl* genes (Figure 5.2e.i).

Conversely, biological pathways enriched by down-regulated genes at 5D differentiation, (Benjamini and Hochberg p-value<0.005), included *regulation of multicellular organismal development*, *regulation of developmental proces*s and *cell fate commitment* (Figure 5.2e.ii). Genes downregulated at this time point included *Fgfr2* and *Mycn*. Signalling via the FGFR2 receptor maintains TSCs in the stem cell state, via binding of the exogenously provided ligand, FGF4 and its down-regulation is vital for differentiation. The reduction in the expression of N-Myc at 5D, after transient expression between 1D and 3D differentiation, could indicate the need for proliferation in precursor cell types in order to continue along differentiation trajectories *in vitro*. Its down-regulation at 5D potentially confirms the presence of populations of terminally differentiated, non-proliferative cell types.

There were certain differences between WT and *Kdm1b*$^{-/-}$ cells when calling differentially expressed genes between each time point of *in vitro* differentiation. With each step-wise comparison of the transcriptome during differentiation, there were between 25-40 % of WT DE genes that were not identified in the stepwise comparisons of *Kdm1b*$^{-/-}$ replicate sets. Due to the very low numbers of

differentially expressed genes by either DESeq2 or intensity difference between WT and KDM1B null cells, this implies that these genes were delayed in their up- or down-regulation at each of these time points, rather than out-right dysregulated. Indeed, Figure 5.3 plots normalised expression of these WT-only DE genes and shows a slight delay in the up- or down-regulation of these genes in KDM1B null clones. The expression of these genes showed very similar heterogeneity in both WT and KDM1B null clones, but their mean expression in the knockout clones was sufficiently different for these not to be identified by DESeq and/or intensity difference.

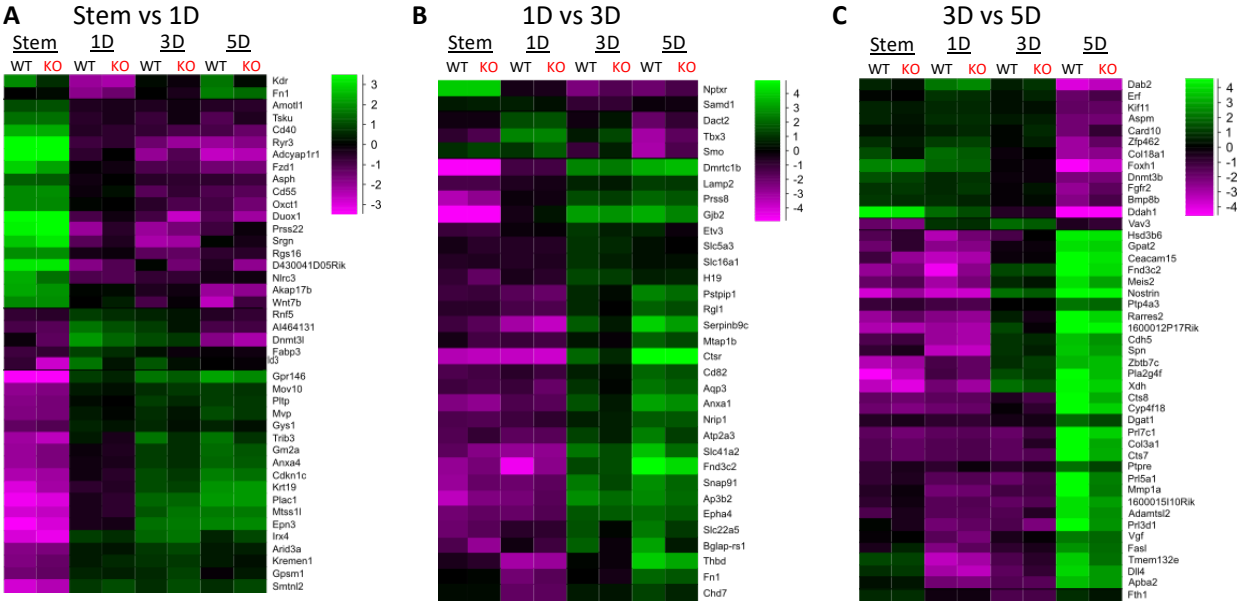



**Figure 5.3 KDM1B null cells appear to show slightly delayed regulation of gene expression**

Heatmaps showing DE genes not identified when comparing *Kdm1b*$^{-/-}$ cells at each stepwise comparison: **A.** Stem vs 1D, **B.** 1D vs 3D and **C.** 3D vs 5D, demonstrating a delay in modulating gene expression

### 5.2.3 Active stem cell enhancers containing differentially enriched H3K4me1 peaks are associated with differentially expressed genes

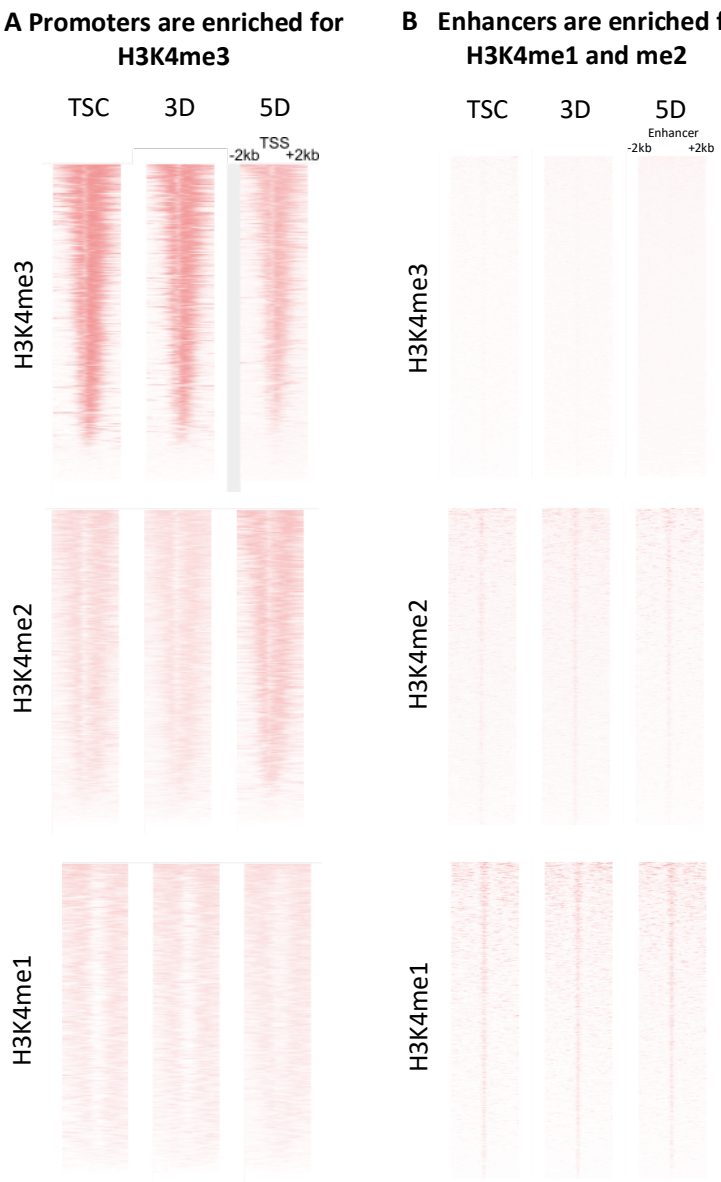
Having ascertained that the predicted genes were being up- and down-regulated as expected during trophoblast differentiation, I wanted to try to link enhancer function with gene expression. I began by defining enhancers as H3K4me1 peaks at least 5kb from an active TSS (mRNA with more than 10 reads) with no associated H3K4me3 peaks. This reduced the list of 105,171 H3K4me1 peaks to 75,720 potential enhancers. Figure 5.4a shows the strong enrichment of H3K4me1 and H3K4me2 signal within these enhancers and the absence of enrichment of H3K4me3. Using previously published ChIP-seq

data for H3K27ac produced in stem cells, I identified active stem enhancers. I was interested in this group of active enhancers, because their H3K27ac status meant they were more likely to have a detectable impact on gene expression. HOMER motif analysis of these probable active TSC enhancers showed high representation of two motifs: the first AP-1 family of TFs and the second pertaining to trophoblast-specific TFs, TFAP2C and TFAP2A (Figure 5.4b).

Using LIMMA statistical analysis to identify differentially enriched H3K4me1 peaks, I compiled a list of just over 1,700 up-regulated (Figure 5.5a) and 2,100 (Figure 5.5b) down-regulated active stem enhancers between stem and 3D differentiation. Interestingly, down-regulated enhancers showed a distribution similar to that of all enhancers (Figure 5.5a.ii), whereas active stem enhancers whose H3K4me1 signal increased between stem and 3D differentiation, had a higher tendency to be intragenic (Figure 5.5b.ii).

I linked these differentially enriched H3K4me1 enhancers to their nearest gene, with a cut-off of 100 kb. When comparing these lists to DE genes between stem and 3D differentiated cells identified by DESeq2, I noticed some overlap. Figure 5.5c highlights those DE genes linked to one or more differentially enriched enhancer. In both groups, there were more genes whose expression changed in the same direction as H3K4me1 signal, i.e increased gene expression linked to enhancers with increased H3K4me1 ChIP signal. Further, these genes tended to have a higher fold-change in expression, compared to genes whose expression was anticorrelated with change in H3K4me1 signal. Nearly half of the genes linked to differentially H3K4-monomethylated active stem enhancers at 3D, were also differentially expressed.

This analysis was then repeated for differentially enriched H3K4me1 peaks between 3D and 5D differentiation (Figure 5.6). Those enhancer peaks which were marked by H3K27ac in stem cells and showed decreased H3K4me1 signal between 3D and 5D differentiation (Figure 5.6a), had two main patterns of enrichment. Most of these peaks were consistently down-regulated between stem, 3D and 5D time points. Given the decrease in H3K4me1 signal, it is possible that these peaks do not maintain their H3K27ac status to 5D differentiation, thus these enhancers may become inactive whilst not losing H3K4me1 signal altogether. The rest of the peaks showed a transient increase at 3D in H3K4me1 ChIP signal. Compared to those DE enhancers between stem and 3D differentiation that lost H3K4me1, these peaks showed slightly more change to their genomic locations compared to all enhancers (Figure 5.6a.ii).




**C** HOMER Motif analysis of H3K27ac-marked Enhancers

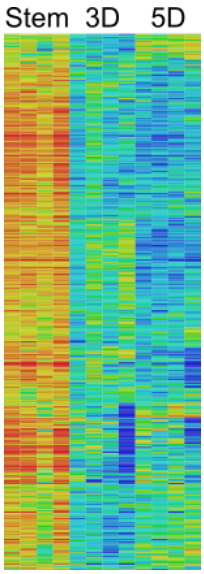
| ChIP *gene* | log P-pvalue | Motif |
|---|---|---|
| JUNC *Jun* | -3.44E+02 | |
| FOSL2 *Fosl2* | -3.23E+02 | |
| FRA2 *Fosl2* | -2.67E+02 | |
| JUNB *Junb* | -2.57E+02 | |
| FRA1 *Fosl1* | -2.50E+02 | |
| BATF *Batf* | -2.38E+02 | |
| ATF3 *Atf3* | -2.36E+02 | |
| TFAP2C *Tfap2c* | -2.18E+02 | |
| PU.1 *Sfpi1* | -2.02E+02 | |
| TFAP2A *Tfap2a* | -1.81E+02 | |

**Figure 5.4 ChIP-seq data showing enrichment of H3K4me3 around TSS and H3K4me1 in enhancers**

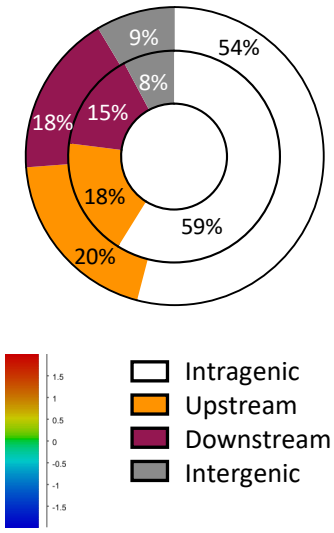
**A.** Aligned probe plots showing Log2RPKM of H3K4m3, H3K4me2 and H3K4me1 data 2kb either side of the TSS of expressed genes (>10 reads). **B.** Aligned probe plots showing Log2RPKM of H3K4m3, H3K4me2 and H3K4me1 data 2kb either side of defined enhancers. **C.** Top 10 results of HOMER motif analysis of H3K27ac-marked enhancers.

**TSC Active Enhancers differentially enriched for H3K4me1 at 3D**

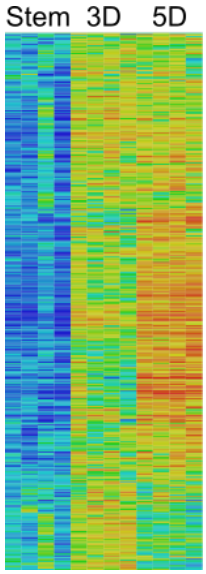
**A.i** Active stem Enhancers losing H3K4me1 at 3D



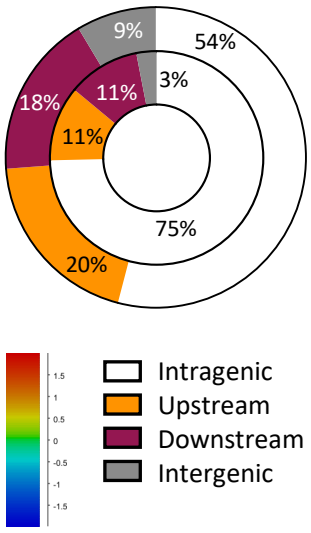

**A.ii** Genomic Distribution



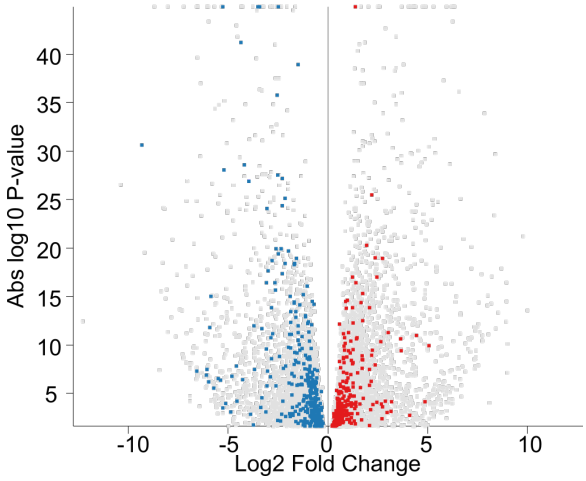

**B.i** Active stem Enhancers gaining H3K4me1 at 3D



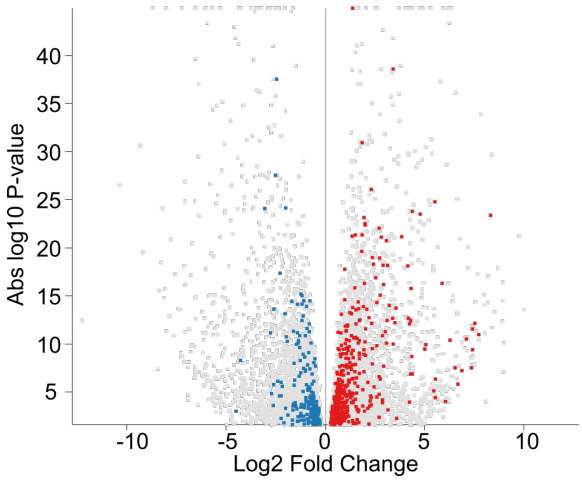

**B.ii** Genomic Distribution




**C.i** Genes linked to enhancers that lost H3K4me1 at 3D




**C.ii** Genes linked to enhancers that gained H3K4me1 at 3D




**D** Many DE genes identified between stem and 3D were associated with a differentially enriched active stem enhancer

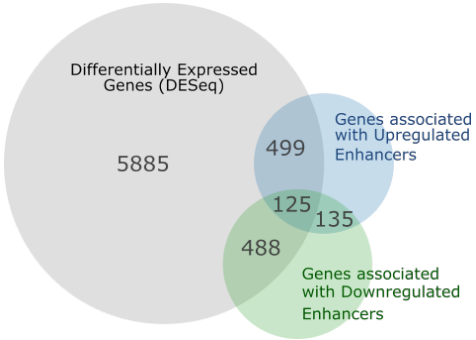


**Figure 5.5 Differentially enriched enhancers between stem and 3D that are marked with H3K27ac in stem cells link to DE genes**

**A.i** Heatmap of differentially enriched H3K27ac-marked enhancers that lost H3K4me1 signal at 3D; **A.ii** genomic distribution of down-regulated peaks relative to all enhancers. B.i Heatmap of differentially enriched H3K27ac-marked enhancers that gained H3K4me1 signal at 3D; **A.ii** genomic distribution of up-regulated peaks relative to all enhancers. **C.** Volcano plot of DE genes between stem and 3D (DESeq2 $p<0.05$) highlighting genes associated (<100 kb) of a differentially enriched enhancers **i.** down-regulated or **ii.** up-regulated between stem and 3D. **D.** Venn diagram showing overlap between DE genes and genes associated with differentially enriched enhancers.

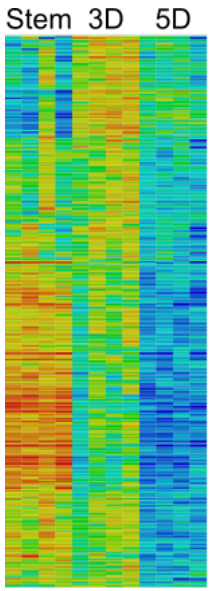
The H3K27ac-marked H3K4me1 peaks that gained signal between 3D and 5D showed a similar pattern to those losing methylation: most enhancers gained methylation gradually over the time course, with a subset of peaks showing transient down-regulation between stem and 3D, which returned to stem cell levels at 5D. In my opinion, it is a relatively safe assumption that active stem enhancers which gained H3K4me1 signal throughout the differentiation time course retained their H3K27ac status and were likely still active enhancers. Those which showed transient loss of H3K4me1 signal, however, might have been silenced during differentiation to an intermediate cell type and the return of H3K4me1 signal could point to the reactivation of these enhancers in one or more terminally differentiated cell types. Figure 5.6bii illustrates that these enhancers had a very similar genomic distribution to those peaks upregulated at 3D differentiation, with 77 % being within gene bodies compared to the global average of 54 %.

Also similar to those H3K27ac-marked peaks which were differentially enriched at 3D, those downregulated at 5D had most overlap with genes whose expression was also downregulated between 3D and 5D differentiation in the transcriptomic data, by DESeq2 (Figure 5.6ci). Similarly, those peaks gaining H3K4me1 signal at 5D were linked with more genes whose expression increased between 3D and 5D (Figure 5.6cii).

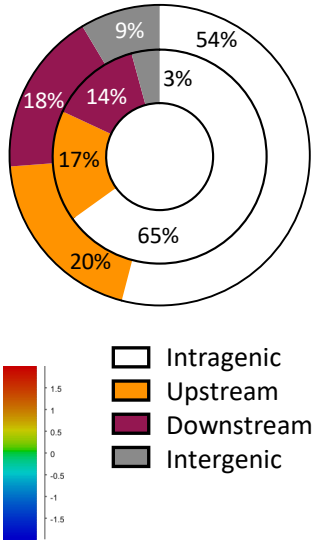
It is interesting that the relatively small number of differentially methylated peaks had considerable overlap in their linked genes, with 260 genes being associated with peaks that both gained and lost H3K4me1 signal at 3D and 83 at 5D (Figure 5.5d and 5.6d, respectively). Linking enhancers to genes via transcriptomic analysis has its limits. It is likely that many of these active stem enhancers do not regulate expression of the closest gene, even the intragenic enhancers, as evidenced by these apparent conflicting changes in H3K4me1 enrichment. Further, the fact that many of the linked genes changed expression in the opposite direction to that expected based on H3K4me1 enrichment, also suggests that these genes were not under the regulation of the linked enhancer.

**TSC Active Enhancers differentially enriched for H3K4me1 at 5D**

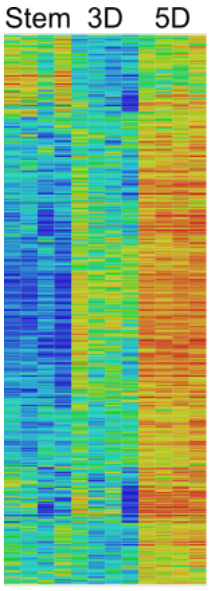
**A.i Active stem Enhancers losing H3K4me1 at 5D**



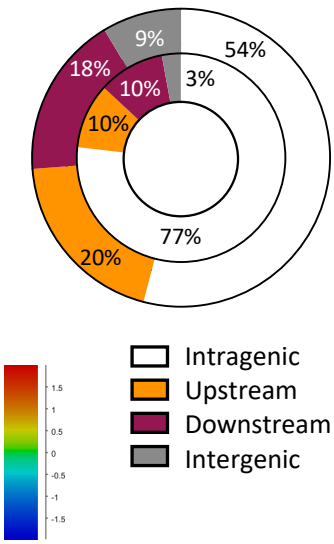

**A.ii** Genomic Distribution



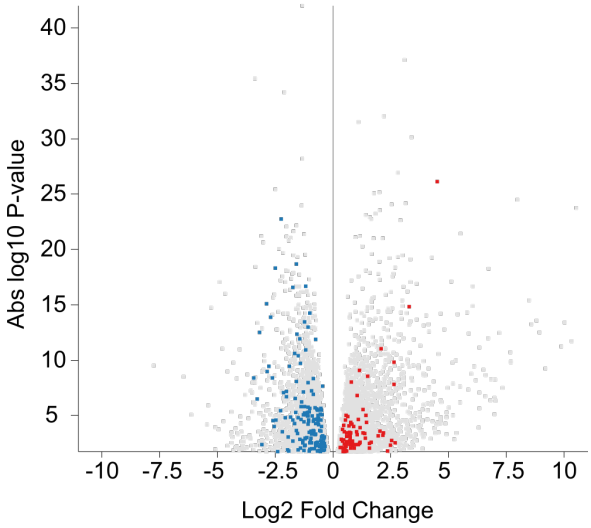

**B.i Active stem Enhancers gaining H3K4me1 at 5D**



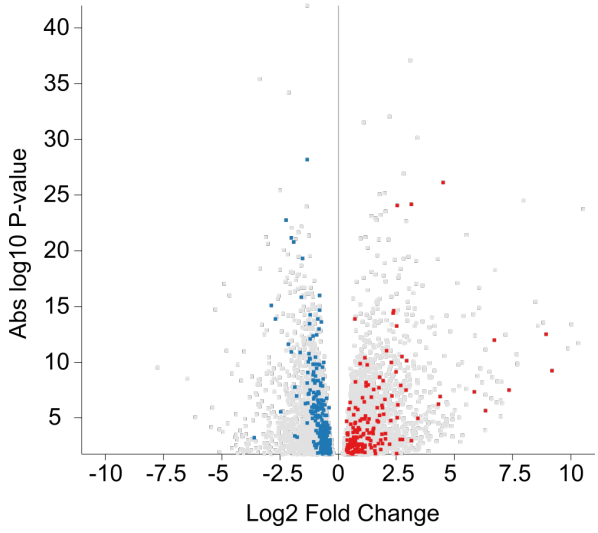

**B.ii** Genomic Distribution




**C.i** Genes linked to enhancers losing H3K4me1 at 5D




**C.ii** Genes linked to enhancers gaining H3K4me1 at 5D




**D** Many DE genes identified between 3D and 5D differentiation were associated with a differentially enriched H3K27Ac-marked enhancer

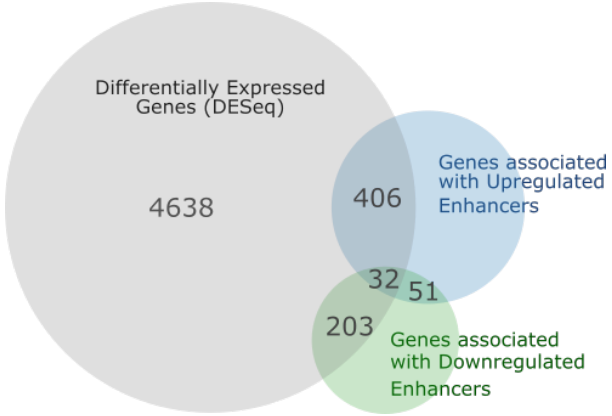


**Figure 5.6 Differentially enriched enhancers between 3D and 5D that are marked with H3K27ac in stem cells link to DE genes**

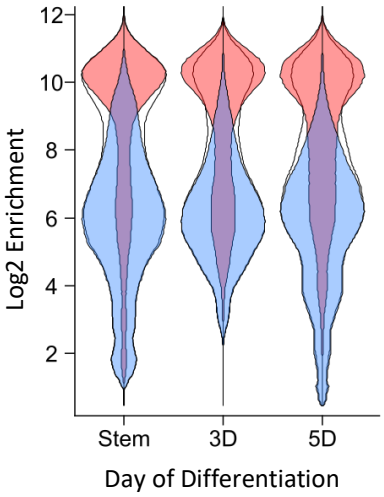
**A.i** Heatmap of differentially enriched H3K27ac-marked enhancers that lost H3K4me1 signal at 5D; **A.ii** genomic distribution of down-regulated peaks relative to all enhancers. B.i Heatmap of differentially enriched H3K27ac-marked enhancers that gained H3K4me1 signal at 5D; **A.ii** genomic distribution of up-regulated peaks relative to all enhancers. **C.** Volcano plot of DE genes between 3D and 5D (DESeq2 p<0.05) highlighting genes associated (<100 kb) of a differentially enriched enhancers **i.** down-regulated or **ii.** up-regulated between 3D and 5D. **D.** Venn diagram showing overlap between DE genes and genes associated with differentially enriched enhancers.

### 5.2.4 Identifying distal trophoblast-specific promoters

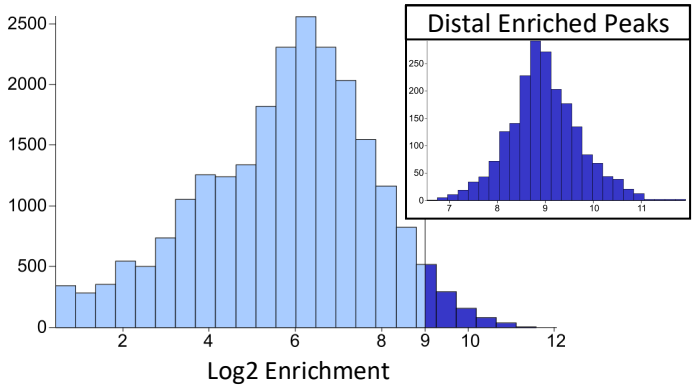
In addition to being able to identify trophoblast-specific enhancers, having both transcriptomic and H3Kme3 ChIP-seq data allows for the possibility of identifying trophoblast-specific promoters and transcripts. Plotting the enrichment of all H3K4me3 peaks, I found the predicted binomial distribution, showing peaks with 'medium' (Log2~6) and 'high' (Log2~9) enrichment. Splitting all H3K4me3 peaks based on their proximity to the TSS of expressed mRNAs (>10 reads), I found that peaks which fell within promoters tended to be more highly enriched, compared to those peaks more distal to the TSS (Figure 5.7a). There was, however, a subpopulation of distal peaks with high enrichment (Log2 >9, Figure 5.7b). I postulated that these peaks might correspond to distal promoters, potentially indicating trophoblast-specific transcripts. I assessed the coincidence of these distal H3K4me3 peaks with annotated lists of transposable elements of the ERV, LINE and SINE families of retrotransposons. There was no significant increase in the incidents of these transposable elements within the distal enriched H3K4me3 peaks. Although, peaks overlapping one or more of these elements maintained a higher enrichment across the time course of differentiation, compared to those not (data not shown).

Having identified these potential distal promoters, I decided to filter the list of expressed genes for those that did not have an H3K4me3 peak over their annotated TSS (649 genes) and then filtered for those without promoter-associated H3K4me3 peak (within 5kb of TSS), which left around 450 genes (corresponding to 1.7 %, Figure 5.7c). Annotating the enriched distal H3K4me3 peaks with the closest downstream expressed gene, I found several examples of genes in both groups. Looking at the RNA-seq data, I did indeed find what look like trophoblast-specific transcripts with evidence of splicing, whose TSS overlapped with these highly enriched distal peaks; the example *Glce* gene locus is shown in Figure 5.7d.

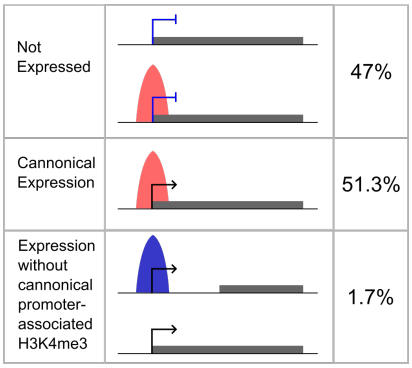
**A** Binomial Distribution of H3K4me3 Peaks



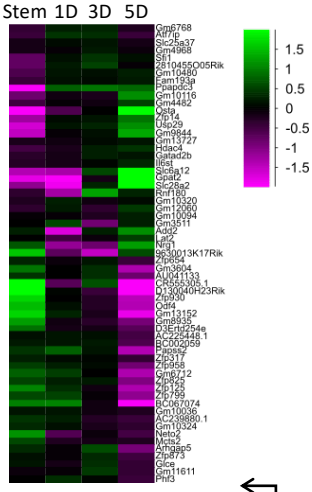

**B** H3K4me3 Peaks Distal to Promoters



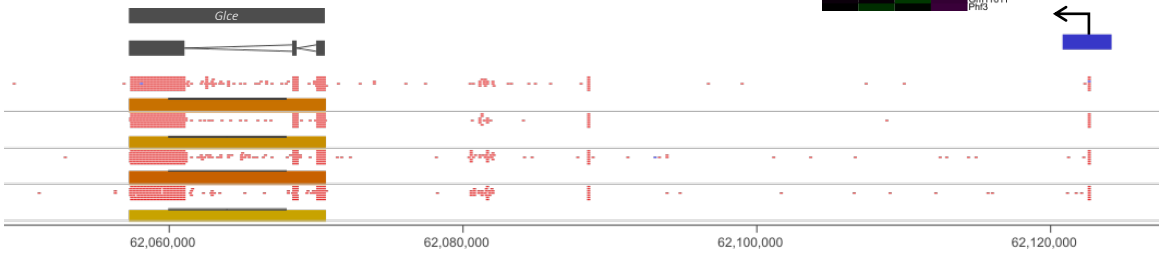

**C**




**E** Expression of Genes linked to a Distal Promoter




**D** *Glce* locus showing transcription from distal putative promoter




**F** HOMER Motif Analysis

| ChIP *gene* | log P-pvalue | q-value (Benjamini) | Motif |
|---|---|---|---|
| KLF14 *Klf14* | -1.52E+01 | 0.0001 |  |
| MAZ *Maz* | -1.46E+01 | 0.0001 |  |
| DLX1 *Dlx1* | -1.39E+01 | 0.0001 |  |
| MYOD *Myod1* | -1.38E+01 | 0.0001 |  |
| NKX6-1 *Nkx6-1* | -1.36E+01 | 0.0001 |  |
| MYOG *Myog* | -1.34E+01 | 0.0001 |  |
| ATOH1 *Atoh1* | -1.23E+01 | 0.0003 |  |
| TCF12 *Tcf12* | -1.19E+01 | 0.0004 |  |
| ISL1 *Isl1* | -1.19E+01 | 0.0004 |  |
| TFAP4 *Tfap4* | -1.15E+01 | 0.0004 |  |

**Figure 5.7 Several hundred genes show expression without canonical H3K4me3 peak at the TSS/promoter**

**A.** Bean plot showing all H3K4me3 peaks (MACS2), highlighted in red are peaks over TSS of actively transcribed genes and in blue are peaks not associated with active promoters. **B.** H3K4me3 peaks not associated with active promoters, highly enriched peaks are highlighted in dark blue (inset). **C.** Schematic showing percentage of genes within different groups based on expression (>10 reads) and H3K4me3 promoter status. **D.** Example locus *Glce* gene showing evidence of distal H3K4me3 peak and a novel TSS around 40 kb upstream of the annotated gene. **E.** Heatmap showing normalized expression of expressed genes associated with a distal enriched H3K4me3 peak. **F.** Top 10 hits from HOMER motif analysis of distal enriched H3K4me3 peaks.

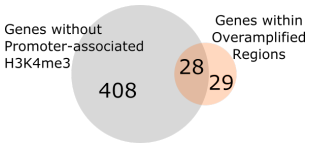
Figure 5.7e plots those genes linked to a distal promoter, over differentiation, showing that half of these genes were upregulated and half, downregulated between stem and 5D differentiation. This indicates that these distal promoters could be active throughout *in vitro* differentiation and might not be regulated by the core trophoblast stem transcription factor network. Performing HOMER motif analysis on all potential distal promoter peaks, I found the imprinted TF, KLF14, as the top hit. Loss of KLF14 results in increased placental weight and it acts antagonistically with *Mest* to regulate placental size, in mice (Koppes *et al.*, 2019). Also, in the top motif hits were TCF12 and TFAP4. TCF12 is highly expressed throughout differentiation (data not shown) and is a binding partner of ASCL2, which is required for successful differentiation towards spongiotrophoblast lineage (Bogutz et al., 2018). These results are encouraging and provide examples of both stem cell-specific and differentiation-dependent TF binding motifs within these distal promoters.

### 5.2.5 Transcriptional induction of genes within TGC-specific amplified regions was not coupled with promoter-associated H3K4me3

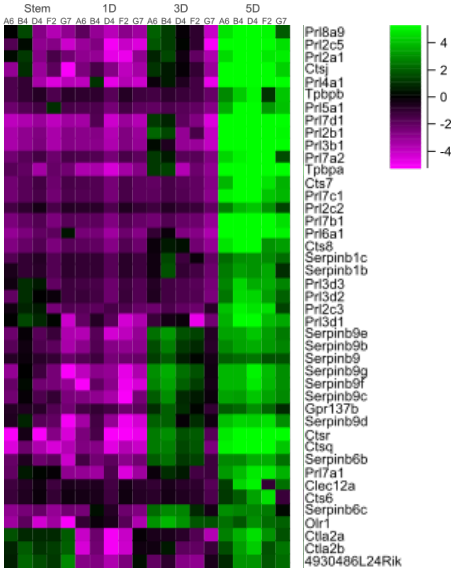
As well as those genes whose expression I was able to link to an enriched distal H3K4me3 peak, also represented were genes without an apparent associated H3K4me3 peak. This list contained several TGC-associated genes located in the expanded regions that had been reported (Hannibal and Baker, 2016). As mentioned previously, these genes are located within regions of the genome which were over-replicated during endoreduplication of TGCs. These regions arose from gene duplications and occupy five loci, with four being found on chromosome 13; two regions containing prolactin genes, one the cathepsins and the final encoding the serpins. Of the relatively small list of expressed genes without an associated H3K4me3 peak, 27 genes were found to be within these regions (Figure 5.8a). Expression of many of these genes was robustly induced at either 3D or 5D differentiation (Figure 5.8b).

**A** Expressed Genes within TGC overamplified regions, with no associated H3K4me3 Promoter peak

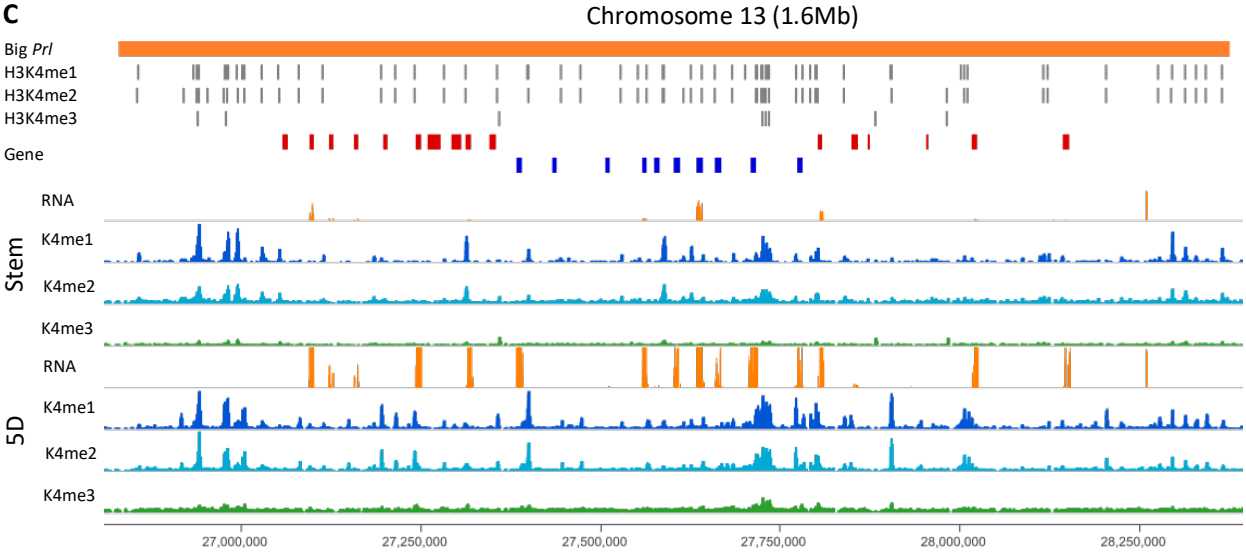
| Gene | Chr | 5D-Stem |
|---|---|---|
| *Gm5884* | 6 | -0.5 |
| *Clec12a* | 6 | 3.2 |
| *Prl2c3* | 13 | 4.6 |
| *Prl2c2* | 13 | 4.0 |
| *Prl3d1* | 13 | 4.8 |
| *Prl3d2* | 13 | 3.9 |
| *Prl3d3* | 13 | 3.9 |
| *Prl3b1* | 13 | 11.0 |
| *Prl6a1* | 13 | 8.1 |
| *Prl2b1* | 13 | 9.9 |
| *Prl8a9* | 13 | 6.3 |
| *Prl7b1* | 13 | 8.3 |
| *Prl7a1* | 13 | 5.0 |
| *Prl7a2* | 13 | 6.2 |
| *Prl7d1* | 13 | 11.5 |
| *Prl7c1* | 13 | 7.4 |
| *Prl2a1* | 13 | 6.3 |
| *Prl4a1* | 13 | 10.0 |
| *Prl5a1* | 13 | 5.9 |
| *Serpinb1c* | 13 | 3.9 |
| *4930486L24Rik* | 13 | 0.5 |
| *Ctsq* | 13 | 9.6 |
| *Cts6* | 13 | 2.6 |
| *Cts7* | 13 | 7.3 |
| *Gm12060* | 13 | 0.8 |
| *CT573034.1* | 13 | -3.6 |
| *Zfp934* | 13 | -2.0 |
| *Gm3604* | 13 | -2.3 |




**B** Normalised Expression of overreplicated TGC regions




**C**




**D**

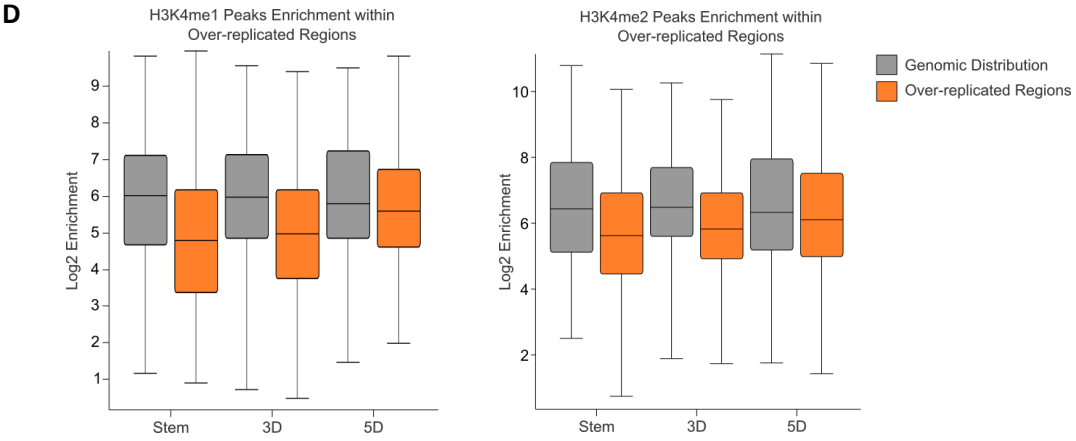


**Figure 5.8 Induction of gene expression in overreplicated TGC regions is not associated with gain in promoter-associated H3K4me3**

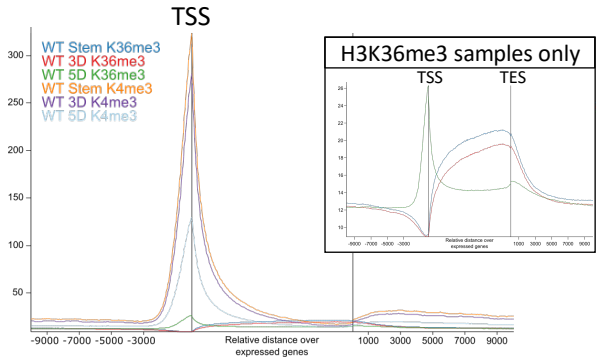
**A.** Many expressed genes without a promoter-associated H3K4me3 peak were located within over-replicated TGC-specific regions identified by Hannibal and Baker (2016). **B.** Heatmap plotting normalized expression of genes within the TGC-specific amplified regions not associated with H3K4me3 peak in the promoter. **C.** 1.6Mbp Big Prl region of chromosome 13 showing induction of gene expression at 5D differentiation and very little H3K4me3 signal. **D.** H3K4me1 and H3K4me2 peaks within the over-amplified regions show slight under-enrichment compared to global levels throughout the differentiation time course.

Looking at the raw H3K4me3 data, these regions were indeed depleted for H3K4me3 ('Big Prl' region is provided as an example, Figure 5.8c). This is significant as it suggests that expression of these genes was associated with some other chromatin mark. The increase in transcripts without concurrent accumulation of H3K4me3 could also be a product of endoreduplication, as more copies of these genes could correlate to more transcript without necessarily coinciding with changes to chromatin. Looking at the other H3-K4 data, I noted that H3K4me1 and H3K4me2 were also somewhat depleted in these regions compared to the surrounding chromatin, at 5D (Figure 5.8d). Their enrichment, however, was not depleted to the same extent as that of H3K4me3.

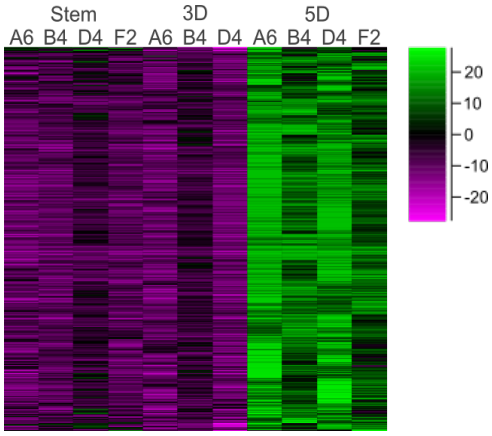
### 5.2.6 Redistribution of H3K36me3 is linked to H3K4 trimethylation and gene expression

Having noticed an apparent redistribution of H3K36me3 to the TSS at 5D (Figure 4.13), I decided to compare this to H3K4me3. Figure 5.9a illustrates the similarity in the distribution over actively transcribed genes, of H3K4me3 and 5D H3K36me3 signal, however the H3K36me3 samples showed much less enrichment. Figure 5.9b shows the comparison of 3D and 5D H3K36me3 samples. The group of enriched probes at 5D (red) showed a slight bias towards highly enriched H3K4me3 regions but did populate the entire distribution (Figure 5.9biii). Comparing 3D and 5D H3K4me3 samples here, illustrates that there was less enrichment in 5D H3K4me3 samples, which might be biological, but could also be technical. When looking at H3K4me3 peaks and assessing differentially enriched peaks during differentiation, I corrected for this by normalising to the $40^{th}$ and $99^{th}$ percentiles.

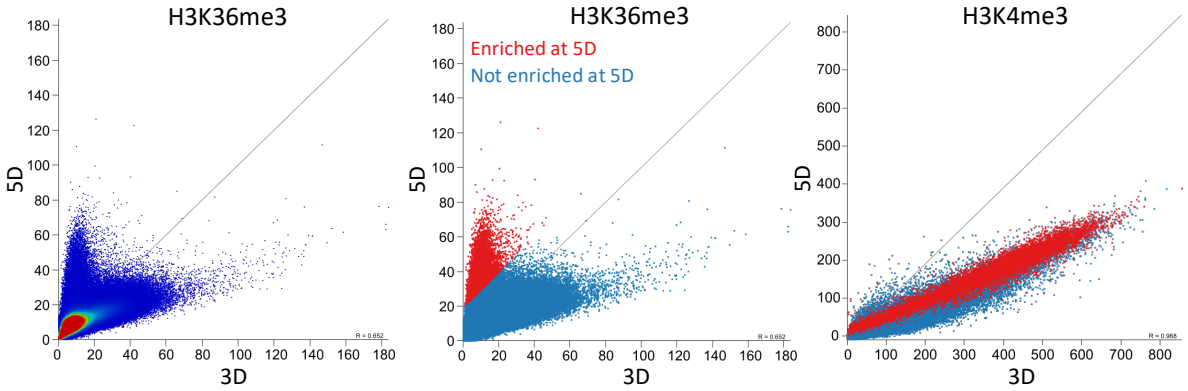
**A** Redistributed H3K36me3 signal at 5D mimics H3K4me3 pattern but is much less enriched.



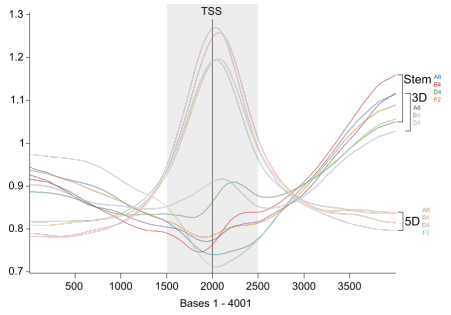

**C** 1kb probes enriched at 5D shows concerted effect in all clones

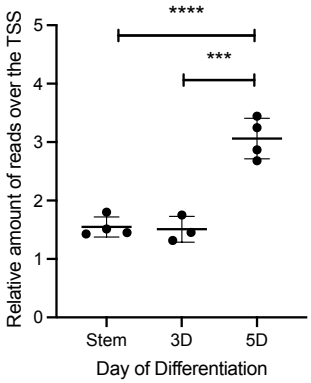



**B** 5D enriched H3K36me3 regions represent entire distribution of H3K4me3 signal

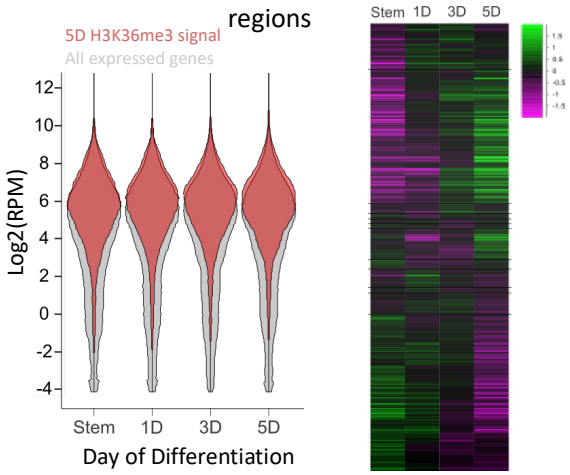



**D** There was a significant shift in H3K36me3 signal to the TSS at 5D differentiation




**E** Genes associated with the 5D enriched regions




**F** H3K4me3 Peaks within which there is a gain in H3K36me3 signal

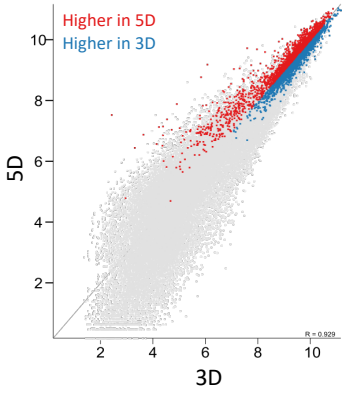


**Figure 5.9 H3K36me3 ChIP signal shows marked redistribution to the TSS at 5D differentiation**

**A.** Probe distribution plot showing reads over the gene body of actively transcribed genes of H3K4me3 and H3K36me3 ChIP-seq data. Inset shows H3K36me3 only. **B.** Scatter plots comparing **i.** 3D and 5D H3K36me3 ChIP-seq data **ii.** showing group of probes highly enriched in H3K36me3 at 5D with very little enrichment at 3D. **iii.** Scatter plot showing the enrichment of the highly enriched 5D H3K36me3 probes between 3D and 5D H3K4me3 data. **C.** Heatmap showing normalised subset of 5D enriched H3K36me3 probes demonstrating concerted upregulation in all clones. **D.** H3K36me3 reads are significantly more enriched in over the TSS (+/-500bp) in 5D samples compared to stem and 3D samples. Statistical analysis was by Tukey HSD test *** $p < 0.005$, **** $p < 0.001$. **E.** Expression of the genes whose TSS gained signal specifically at 5D differentiation: **i.** bean plot Log2RPKM and **ii.** heatmap of normalized expression. **F.** Scatter plot showing comparison of 3D and 5D H3K4me3 peaks, highlighting those which overlap a region that gained H3K36me3 signal.

Figure 5.9c demonstrates that this increase in H3K36me3 signal at 5D differentiation was a concerted change in all four WT clones assessed, albeit to slightly different degrees. Indeed, Figure 5.9d shows that there was significant redistribution of signal to the 500 bp either side of the TSS in 5D H3K36me3 libraries. When looking at gene expression of those genes associated with these enriched regions at 5D, I found that these genes tended to be relatively highly expressed, with little differences in expression throughout differentiation (Figure 5.9e). Further, these regions corresponded to highly enriched H3K4me3 peaks. Those H3K4me3 peaks with less enrichment at 5D that coincided with the redistributed H3K36me3 signal were more highly enriched than in 3D differentiated cells, suggesting that H3K36me3 was redirected to regions gaining H3K4me3 methylation (Figure 5.9f).

## 5.3 Discussion

In this chapter I performed several integrated analyses using the time course of transcriptomic and ChIP-seq data for four histone modifications, H3K4me1, H3K4me2, H3K4me3 and H3K36me3. I confirmed expression of several key markers of TSCs, as well as markers indicating differentiation to the main cell types of the placenta: spongiotrophoblast *Ascl2*, *Tpbpa*; TGC *Prl1*, *Plf* and Syncytiotrophoblast *Gcm1* and *Syna*. These data represent an invaluable resource for the future identification of key players in trophoblast stem cell self-renewal and differentiation.

When assessing the transcriptomic data, I noticed that one clone, G7, showed a significant delay in down-regulating and up-regulating many differentially expressed genes. Indeed, when looking at the chromosome number by metaphase spread in chapter 4, this clone showed the most variation of the

five WT clones characterised. This clone represents an interesting opportunity to assess what changes the process of single cell sorting and expansion had on its apparent resistance to differentiation.

The high overlap between H3K27ac-marked enhancers in stem cells and DE genes identified by DESeq2, throughout differentiation, was exciting to see. It suggests that at least some of these enhancers retain their status and are actively regulating gene expression throughout differentiation. Further validation of some of these regions would be exciting and important to identify links between regulatory regions and their target genes during trophoblast differentiation. The first important step could be to perform sequential ChIP experiments to ascertain the H3K27ac status of these regions, particularly those enhancers whose change in H3K4me1 enrichment was correlated to the change in expression of the linked gene.

By using transcriptomic and ChIP-seq data, I have been able to identify trophoblast-specific transcripts of several genes expressed with different dynamics during differentiation. 5'-RACE could be used to ascertain the existence of these trophoblast-specific isoforms. Moreover, Western blot analysis could be performed to assess whether these novel exons are protein-coding in the trophoblast, as increased molecular weight would confirm this. Further, analysing other transcriptomic datasets could indicate whether these are rodent-specific placental transcripts, or whether they are evolutionarily conserved and present in other species.

Many of the distal H3K4me3 peaks I identified were overlapping one or more transposable element. Expression of ERV elements in the placenta has been well-characterised and is an entire field of research in itself. The distal H3K4me3 peaks identified throughout the time course of development, were no more or less enriched for incidents of retrotransposons. However, I also observed that H3K4me3 enrichment of these distal peaks was maintained throughout differentiation to a higher level, in peaks within which there were one or more retrotransposons present. One explanation for this is that these elements, particularly the endogenous retroviruses, are highly expressed in the placenta and indeed have even been co-opted for placental function (Harris, 1998; Blaise *et al.*, 2003; Chuong, 2018) and thus transcription must be maintained throughout the timecourse, leading to steady levels of H3K4me3 enrichment.

Finally, I tried to answer the question of whether the H3K36me3 redistribution I observed at 5D, is real and not due to material transfer or some other technical error during library preparation. Further, the associated genes showed a range of expression, albeit to a somewhat increased level, throughout differentiation. Of course, the most definitive way of confirming this finding is to re-do the ChIP-seq experiments of this mark during differentiation, to ascertain whether it is reproducible. It would be ideal to be able to do this in a different TSC line, or even using mouse placenta samples, to identify in

which cell type(s) this 'switch' might be occurring. Given the nature of ChIP-seq data, the increased signal around the TSS of actively transcribed genes means that there may be a loss of H3K36me3 signal in the canonical distribution over the gene body. I believe that this mark is still being written in these regions at 5D differentiation, but the addition of reads at the TSS masked this signal. One way to combat this would be to perform directed differentiation of TSCs *in vitro*, towards the synctytiotrophoblast or TGC lineages, in order to obtain a less heterogeneous population of differentiated cells (Maltepe *et al.*, 2005).

# 6 Chapter Six

# Discussion

## 6.1 Summary of thesis findings

The epigenome is a vastly important tool in cellular differentiation, organismal development and lineage commitment. The first cell fate decision made during preimplantation development is the distinction of the trophectoderm from cells of the ICM. TE-derived TSCs are multipotent cells able to give rise to all cell types of the mature placenta, an organ vital for reproduction. Indeed, almost 70 % of reported embryonic lethal mutations are found in conjunction with a malformed placenta and up to a third of these can be rescued in part by return of a functional placenta (Perez-Garcia *et al.*, 2018). Despite being such an important organ, the molecular regulators of trophoblast stem cell self-renewal and differentiation are not completely understood.

In this work I have demonstrated that histone lysine demethylase, KDM1B, plays a role in fine-tuning the transcriptional networks both in maintenance of TSCs and in directing differentiation, becoming ubiquitously expressed during *in vitro* differentiation. Loss of KDM1B leads to accumulation of its substrate modification, H3K4me1, at a subset of intragenic H3K4me1-marked enhancers, without effecting H3K4me3.

Additionally, I have reported increased chromosomal instability of chromosome 13 in KDM1B null TSCs following CRISPR Cas9-targeted gene ablation. Clonal lines developed in the course of this work showed increased variation in ploidy compared to the starting population of TS-Rs26 cells, indicating that the process of single cell clonal expansion has a detrimental effect on genome stability. I believe the data I have shown warrants further investigation to distinguishing the effect of loss of KDM1B from the potential chromosome 13 instability and on-target side-effects of CRISPR Cas9-induced DSBs.

Further, in the context of WT trophoblast differentiation *in vitro*, this thesis provides evidence of distal H3K4me3-marked potential promoters from which several examples of trophoblast-specific transcripts were identified. Further, integrated analysis of enhancer marks and differential gene expression provide evidence that stem H3K27ac-marked enhancers play a role in modulating gene expression as late in differentiation as 5D. Finally, I have identified a striking redistribution of H3K36me3 away from the gene body of transcribed genes, to the TSS of a subset of expressed genes in 5D differentiated cells.

## 6.2 TS-Rs26 cells' ploidy and their future use in research

Adverse effects of prolonged culture of mammalian cell lines *in vitro* have been well documented and include genetic and epigenetic drift, senescence as well as developmental, morphological and metabolic changes (Hughes *et al.*, 2007; Rebuzzini *et al.*, 2016). The findings in this thesis demonstrate

that TS-Rs26 cells have significant genetic drift but maintain developmental plasticity and capacity to differentiate to several terminal cell types of the placenta, with high reproducibility. Further, the trophoblast has been demonstrated to be much more tolerant of aneuploidies compared to the epiblast (Kalousek and Dill, 1983; Kalousek, 1994; Kevin *et al.*, 2001; Starostik, Sosina and McCoy, 2020) and indeed TGCs undergo many rounds of endoreduplication for their functionality (Simmons, Fortier and Cross, 2007).

There were not enough cells measured during the metaphase spread experiment, nor was the spread of the chromatids consistently achieved to a high enough degree to definitively ascertain the median chromatid number in metaphase TS-Rs26 cells (Bates, 2011). However, I am confident that TS-Rs26 cells are not diploid. This indicates that a genome duplication event has occurred and that this was more favourable to culture conditions such that polyploid cells have outcompeted diploid cells. This would be in line with many studies of ESCs' response to prolonged *in vitro* culture (Draper *et al.*, 2004; Lund *et al.*, 2012). The selection for aneuploid or polyploid cells *in vitro* is counter to *in vivo* studies of the effect of cellular polyploidy: that it restricts cells' proliferative capacity (Gorla, Malhi and Gupta, 2001; Wilkinson *et al.*, 2019). However, this could be explained by polyploid cells' increased resistance to apoptosis and genotoxic stress (Mehrotra *et al.*, 2008).

The mechanism by which polyploidisation of the TS-Rs26 cell line occurred, is relevant as this could inform our understanding of the adaptations *in vitro* culture has on cellular biology. The mechanism of genome duplication would also provide an insight into what (if any) cell cycle checkpoints failed in this cell line. As previously described, polyploid, terminally differentiated cell types of the mouse placenta, include TGCs and SynT cells, which arise by endoreduplication and cell fusion, respectively (Reviewed, Latos and Hemberger, 2016). I believe the mechanism which gave rise to the polyploidy identified in stem TS-Rs26 cells is distinct from either of these processes. The endoreduplication employed by TGCs is a post-mitotic process occurring in terminally-differentiated cells, but TS-Rs26 cells retain their developmental plasticity and are highly proliferative (Chapter 3). Further, TS-Rs26 cells retained in stem culture conditions are not multinucleated, retaining cell membranes, thus syncytialisation to give rise to polyploidy by way of multinucleation, is unlikely to be the cause. Therefore, it is probable that another mechanism of genome duplication has occurred to give rise to polyploid TS-Rs26 cells.

In mammals, polyploidisation in somatic tissues occurs in several cell types including, cardiac myocytes, arterial smooth muscle cells, megakaryocytes and hepatocytes (Feinendegen *et al.*, 1962; Brodsky and Uryvaeva, 1977; Owens and Schwartz, 1983; Kudryavtsev *et al.*, 1993). Megakaryocytes enter endomitosis in which the cells cycle between phases of synthesis by entering the early stages of mitosis, before aborting at anaphase A to re-enter the cell cycle at a gap-phase (Vitrat *et al.*, 1998). Conversely, mature hepatocytes utilise acytokinetic mitosis (complete cell cycle except cytokinesis) to

form diploid, binuclear cells which then undergo a complete cell cycle, producing tetraploid daughter cells (Guidotti *et al.*, 2003; Margall-Ducos *et al.*, 2007). This cycle repeats giving rise to livers constituting cells tetraploid and higher at a rate of up to 90 % in mice and of around 30 % in humans (Kudryavtsev *et al.*, 1993; Duncan *et al.*, 2010; Bou-Nader *et al.*, 2020).

Polyploidy is also prevalent in cancer. It is estimated that up to 30 % of human solid tumours are polyploid (Cancer Genome Atlas, https://www.cancer.gov/tcga; Bielski *et al.*, 2018) and indeed that polyploidisation is one of the key drivers of carcinogenesis (Galipeau *et al.*, 1996; Maley *et al.*, 2006; Olaharski *et al.*, 2006). Polyploid giant cancer cells (PGCCs) are thought to form in response to chemotherapeutic stress in order to evade treatment and promote metastasis (Zhang *et al.*, 2014).

The route to polyploidisation taken by TS-Rs26 cells is relevant to their future use in research. Each mechanism of genome duplication is born of failure or evasion at distinct cell cycle checkpoints (Zimmet and Ravid, 2000). Further analysis is required to ascertain the frequency at which genome duplication occurs in trophoblast stem cells and indeed, whether this is a phenomenon unique to the TS-Rs26 cell line.

The fact that not a single cell of the starting TS-Rs26 population recorded had the expected 40 chromatids, indicates that, if present, diploid cells are unlikely to constitute the majority of the stem population. Cells with fewer than 58 chromatids were identified in most clones derived from this population (Figure 4.17). Assuming such cells are more populous in the clonal lines, this suggests that the process of single cell sorting and clonal expansion applies a selective pressure that favours the survival of existing cells with fewer chromatids. Alternatively, cells with fewer chromatids could represent TS-Rs26 cells that have undergone a reversal in polyploidy.

There have been several routes of depolyploidisation identified in mammalian cells. Ploidy contraction in WT murine hepatocytes occurs by mitosis involving tripolar spindles, i.e three-way cytokinesis, giving rise to three daughter cells with genetic material distributed in a 4:2:2 ratio (Duncan *et al.*, 2010). Additionally, double mitosis can also occur in polyploid hepatocytes, in which two bipolar spindles form in parallel in the same cell; this can result in four daughter cells containing equal genetic material. Reduction in ploidy is a phenomenon that induces aneuploidy (Duncan *et al.*, 2010). Depolyploidisation of PGCCs can occur by multipolar mitosis without S-phase, sometimes called neosis (Sundaram *et al.*, 2004; Erenpreisa *et al.*, 2005, 2011). Daughter cells following such cell divisions are enriched for NANOG and OCT4, suggesting this process is somehow renewing the tumour microenvironment (Erenpreisa *et al.*, 2011). The presence of TS-Rs26 cells with different numbers of chromatids suggests that ploidy reduction could be occurring. In-depth study of mitosis and cell cycle-associated protein expression in TS-Rs26 cells would be pertinent to elucidate the mechanism by which this is occurring.

The mechanism by which the TS-Rs26 cells expand and, potentially, contract their genomes, is distinct from the post-mitotic endoreduplication employed by differentiated TGCs. These stem cells are proliferative (Chapter 3) and maintain their ability to differentiate (Chapters 3, 4 and 5). As such, I would argue that the TS-Rs26 line retain its' undeniable utility for the *in vitro* study of murine placentation. Further, as these cells have been widely disseminated and used in countless studies, their continued use will allow for direct comparison to these historical publications.

I believe it would be beneficial to future studies of murine TSCs and placentation, to perform comprehensive karyotyping of this and all trophoblast stem cell lines. Particularly if there is an intention to use CRISPR Cas9 to perform genome editing to study gene function. This would allow rapid identification of correctly edited clonal lines with the desired genetic alterations in the absence of any superfluous on- or off-target mutations. It is very difficult to retrospectively identify which findings are resulting from ablation of the target gene and which are from additional CRISPR-induced genomic alterations. Such comprehensive cataloguing of aneuploidy and polyploidy has been performed for both ESC and cancer cell lines (Spits *et al.*, 2008; International Stem Cell Initiative *et al.*, 2011; Taapken *et al.*, 2011; Y. Liu *et al.*, 2019). In my opinion, identifying and reporting changes to karyotype and ploidy in cell lines used will only strengthen the value and expand the significance of the scientific studies which do this, benefiting everyone. A comprehensive analysis of TS-Rs26 cell lines from across many laboratories that have likely drifted significantly from those first derived in the Rossant Laboratory (Tanaka *et al.*, 1998) would not only be fascinating but also, I believe, represent an incredibly useful resource for the field of placental biology.

Utilisation of other trophoblast stem cell lines and indeed deriving new TSCs from blastocysts (Tanaka *et al.*, 1998; Tanaka, 2006) is a valid alternative to continued/future use of TS-Rs26 cells. This approach is particularly desirable when performing *in vitro* studies complementary to a mouse mutant such that the derived TSC line will have an identical genotype to the mouse strain. Additionally, newly derived TSC lines will potentially recapitulate the *in vivo* setting more accurately as these cells will not have been subject to the stresses of long term culture (Hughes *et al.*, 2007).

## 6.3 Loss of KDM1B exacerbates genome instability

The diagnostic PCRs I utilised to identify potential KDM1B null TSCs, produced single bands in all five knockout clones. This indicates that in all five clones, a single copy of *Kdm1b* was repaired such that both primer binding sites were not ablated. Clone C9 had the largest PCR product with all three primer pairs tested, suggesting that repair occurred following minimal DNA digestion. It is possible that more than one allele of *Kdm1b* was correctly repaired in this clone. However, for the other clones, the single

PCR product likely does correspond to a single *Kdm1b* allele as it would be highly unlikely that two or more copies would be repaired in the same imperfect way, as to correspond to a single PCR product.

On-target effects of CRISPR Cas9 genome editing have been described to cause genomic deletions and other events up to the Mb scale at high rates (Adikusuma *et al.*, 2018; Cullot *et al.*, 2019; Alanis-Lobato *et al.*, 2021). Similarly, my data suggests that large-scale indels occurred at a very high frequency, as diagnostic PCR products were only indicative of a single repaired *Kdm1b* locus per clone.

It is highly likely that there were a variety of mutations induced in the five KDM1B null clones produced in the course of this work. Single cell low-pass whole-genome sequencing (scLP-WGS, Dong et al., 2016) could be used to identify larger genomic rearrangements or indels that occurred following CRISPR/Cas9 editing of *Kdm1b*. Alternatively, chromosomal microarray analysis (CMA) could be utilised to identify copy number variations in KDM1B null clones, at a lower resolution (Pinkel *et al.*, 1998). As the only chromosome to show consistently reduced coverage in knockout clones was chromosome 13, FISH-based approaches could also be used to identify site-specific alterations in the knockouts (Jentsch *et al.*, 2001). Such techniques would also enable identification of any chromosomal translocations in the TS-Rs26 cell line.

In order to combat these side-effects target effects, could be the use of engineered Cas9 'nickase' proteins with reduced endonuclease activity such than single-strand nicks are introduced at the target site, rather than DSBs (Trevino and Zhang, 2014). This, in combination with evoking the cells' innate homology-directed repair (HDR) pathway, could be used to reduce these large-scale chromosomal deletions (Reviewed Liu *et al.*, 2019).

I believe it is important to bear in mind that previous studies utilising CRISPR/Cas9 to ablate gene expression, have not reported chromosomal instability of the genes' host chromosome (Chrysanthou *et al.*, 2018; Perez-Garcia *et al.*, 2018, 2021) on the scale presented herein. Of course, it is not evident whether karyotyping was performed after CRISPR/Cas9 genome editing. Nevertheless, I believe it would be premature to totally discount loss of KDM1B as a potential factor in the aneuploidy I have observed in the KDM1B null cells. Changes to the epigenetic landscape, particularly reduced pericentric H3K9me3, has been implicated in the development of aneuploidies and carcinogenesis (McManus *et al.*, 2006; Kupershmit *et al.*, 2014). These studies implicate amplified expression of H3K9me3 demethylases, KDM4A, KDM4B and KDM4C in the increase of mitotic chromosomal mis-segregation. Additionally, deletion of *Lsd1* leads to chromosomal instability, aneuploidy and apoptosis in the mouse oocyte (Kim *et al.*, 2015). The mechanism of action of this is the increased prevalence of the LSD1 substrate mark, H3K4me2, which leads to aberrant activation of retrotransposons and DNA damage,

in addition to direct modulation of expression of CDC25B, which is required for resumption of mitosis in the oocyte.

## 6.4 The effect of loss of KDM1B

In order to draw concrete conclusions as to the effect of loss of KDM1B in trophoblast cells, it is necessary to separate the loss of KDM1B from changes to chromosomal composition in knockout clones. One way to assess this would be to use techniques that do not target genomic DNA, such as using siRNA or shRNA targeted to *Kdm1b* transcripts; another could be using CRISPRi targeted to the *Kdm1b* promoter, to ablate protein expression. CRISPRi would need to be considered carefully as the promoter region of KDM1B overlaps considerably with that of expressed gene, *Tpmt*.

An alternative to targeting genomic copies of the *Kdm1b* gene, or the mRNA transcripts, would be using the auxin-inducible degron 2 (AID2, Yesbolatova et al., 2020) system to directly target KDM1B protein for degradation via the ubiquitin-proteasome pathway. This technique inducibly targets the degron-tagged protein of interest for degradation via addition of 5-phenyl-indole-3-acetic acid (5-Ph-IAA) and can be utilised to assess the acute responses to loss of protein. This would be an interesting alternative to other approaches, as it would allow for more in-depth assessment of the role of KDM1B at different stages of TSC differentiation. It could also be used in combination with targeted differentiation protocols to evaluate the effect of loss of KDM1B in different trophoblast cell types and at different stages of differentiation.

When considering the identified changes to intragenic H3K4me1 and H3K4me2 peaks in 5D differentiated KDM1B null cells, there is no evidence that these changes are a result of the comparatively reduced amount of chromosome 13. When assessing changes to expression of histone demethylases and histone methyltransferases in the *Kdm1b*$^{-/-}$ clones, *Nsd1* and *Phf2* both reside on chromosome 13 and thus showed slightly (not statistically significant) reduced expression in knockout cells compared to WT. *Nsd1* is a histone methyltransferase specific to H3K36 and H4K20 (Rayasam *et al.*, 2003) and *Phf2* is a PKA-dependent histone lysine demethylase specific to H3K9me2 (Baba *et al.*, 2011). As neither of these genes have been shown to have activity towards H3K4 methylation, the slightly decreased expression of *Nsd1* and *Phf2* is unlikely to play a role in the KDM1B null intragenic H3K4me1/2 differentially enriched peaks. Additionally, expression of *Lsd1* was significantly increased during differentiation, by RT-qPCR analysis, as was the amount of LSD1 protein in stem conditions, in KDM1B null clones compared to WT. Increased LSD1 would be more likely to result in increased demethylase activity specific to H3K4me1/2 rather than the gain in methylation signal observed at this subset of intragenic peaks. Further, LSD1 has been shown to reside at gene promoters where it

regulates gene expression (Shi et al., 2005; Shi et al., 2004). Taking these into consideration, I believe that changes to the intragenic H3K4me1 and H3K4me2 peaks identified in KDM1B null cells can be attributed to ablation of the protein, rather than a biproduct of the loss of chromosome 13, or increased LSD1 expression.

## 6.5 Redistribution of H3K36me3 is associated with a subset of highly enriched H3K4me3 peaks

At 5D differentiation, a substantial proportion of H3K36me3 signal appeared to redistribute to the TSS of actively transcribed genes. This has not been previously described and much future analysis is needed to ascertain the biological significance of this change in 5D differentiated TS-Rs26 cells. In terms of validating what I have observed, the first experiments would be to perform ChIP-qPCR for H3K36me3 targeting several of these regions, or indeed repeating the H3K36me3 ChIP-seq with fresh reagents. Further, assessment of this mark in primary trophoblast from mouse placentas would also lend confidence to this finding as a feature of differentiated trophoblast. So too would derivation of a new TSC line and subsequent H3K36me3 profiling. As previously outlined, a novel TSC line would be free of the confounding effects of prolonged cell culture evident with TS-Rs26 cells.

The fact that these 5D enriched regions are not seen in the absence of H3K4me3 is potentially highly biologically relevant, but it also might point to a technical error or some material transfer. There are instances where this redistributed H3K36me3 signal coincides with H3K4me3 peaks of reduced enrichment. These peaks show increased H3K4me3 signal relative to 3D differentiated cells potentially indicating contamination with H3K4me3-specific antibody. ChIP with antibody that is not in excess would lead to preferential pull down of the most highly H3K4me3-enriched chromatin (Nelson, Denisenko and Bomsztyk, 2006).

Following independent validation of enrichment of H3K36me3 at the TSS in 5D differentiated cells, it would be interesting to perform knockout studies involving H3K36-specific methyltransferaseor demethylases. This should be performed in both TS-Rs26 cells and other TSC lines (either newly derived or of other origin). The implication of loss of H3K36me3 methyltransferases and/or demethylases on trophoblast potential would be fascinating. Such studies could provide insight as to the ubiquity of this redistribution. For example, if only TGCs harbour H3K36me3-marked TSS, this could be functionally tied to endoreduplication and epigenetic regulation of gene expression in the context of expanded genetic content.

The initial hypothesis I had upon seeing the shift in H3K36me3 at 5D, was that this could reflect increased promoter-proximal stalling of RNA PolII. The distribution of RNA PolII by ChIP-seq across genes which show increased PolII pausing, is very similar to the redistributed H3K36me3 pattern observed at 5D (Day *et al.*, 2016). However, many studies demonstrate that H3K36me3 is deposited within the gene body by SETD2 during productive transcription by elongating RNA PolII CTD p-Ser2 (Edmunds, Mahadevan and Clayton, 2008). This is important when considering that PolII stalled at the promoter is in the p-Ser-5 form, rather than the SETD2-bound processive p-Ser-2 state. Therefore, the increased H3K36me3 around the TSS, prior to conversion of RNA PolII from inititating CTD p-Ser5 to p-Ser2, does not fit. My data could indicate that SETD2 binds initiating RNA PolII in the trophoblast. Another potential explanation could be that there is a switch in the localisation of SETD2, and redistribution of H3K36me3 to the TSS indicates formation of a novel SETD2 protein complex at the promoter of certain expressed genes.

## 6.6 Future directions

KDM1B plays a role in the regulation of intragenic H3K4me1-marked enhancers without resultant change to the transcriptome. Extending the *in vitro* differentiation time course beyond 5D differentiation could provide insights as to the downstream effect of these potential regulatory regions on the transcriptome. Certainly, given that KDM1B null mice are grossly normal, any resultant transcriptomic changes are unlikely to be catastrophic to the trophoblast compartment. However, I think it is important to identify what role these potential regulatory elements play in the trophoblast.

These regions could also be independently validated as KDM1B targets by performing ChIP-qPCR probing for KDM1B binding. If KDM1B is found to directly bind these intragenic enhancers, that would bolster the assertion that KDM1B forms a complex with elongating RNA PolII in the trophoblast and that it demethylates H3K4me2/1 in intragenic enhancers. To confirm their status as cis-acting regulatory elements, I would use CRISPR/Cas9 to remove select of these putative enhancers and follow subsequent expression of the surrounding genes. Aternatively, targeting these regions with CRISPRi to remove activating histone methylation by LSD1 (Kearns *et al.*, 2015) could also validate their putative enhancer status.

Further, elucidation as to the role of KDM1B in different trophoblast cell types, is important to assess. In the first instance, this should be performed by staining WT placentae for KDM1B at various gestational time points to give an overall indication as to its spatial and temporal expression. Next would be assessment of the structure and morphology of KDM1B null placentae. If the EPC population is effected, i.e reduced junctional zone/increased TGC population, it could be interesting to use the

trophoblast-specific Tpbpa-Cre-loxP system (Werling and Schorle, 2002) to assess the role of KDM1B in this population. Identifying the effect of chromosome 13 stability, particularly in the P-TGCs of KDM1B null placentae could provide invaluable insights as to the role of KDM1B on chromosome stability.

When considering possible interplay between or redundancy of *Kdm1b* and *Lsd1* in the trophoblast, performing double knockout in TSCs would be most practical. As loss of *Lsd1* is embryonic lethal (Wang *et al.*, 2009) and *Kdm1b* is maternal-effect lethal (Ciccone *et al.*, 2009), obtaining LSD1/KDM1B dKO embryos would likely be very challenging. Further, as *Lsd1* knockout studies in the trophoblast show reduced proliferative capacity and increased propensity to premature differentiation (Zhu *et al.*, 2014), and *Kdm1b* loss does not itself arrest embryonic development (Ciccone *et al.*, 2009), I don't believe the work involved in deriving a placental-specific conditional double knockout would be worth it.

Finally, it is important to confirm the redistribution I have reported in WT trophoblast at 5D differentiation, of H3K36me3 to the promoter region of expressed genes. This confirmation could be done in a myriad of ways, firstly by repeating the H3K36me3 ChIP and either performing qPCR analysis or NGS to confirm increased signal in these regions compared to earlier time points of trophoblast differentiation. Having confirmed this is true, there are many biological implications this redistribution could have on the transcriptome and other epigenetic modifications.

It would be interesting to perform sequential ChIP on H3K4me3 followed by H3K36me3 to ascertain whether these marks appear at the same loci within cells. If the marks are not enriched on the same nucleosomes, it could suggest that the marks are indicative of distinct chromatin states. Elucidating the down-stream effect on the transcriptome would be highly beneficial in order to identify the biological significance of this redistribution. Further, elucidation of the mechanism by which this redistribution was occurring, would be important. Performing knockdown of SETD2 during trophoblast differentiation, in order to ablate H3K36me3, could provide insights as to the role of the redistribution.

# Bibliography


Adikusuma, F. *et al.* (2018) 'Large deletions induced by Cas9 cleavage', *Nature*, 560(7717), pp. E8–E9. doi: 10.1038/s41586-018-0380-z.

Alanis-Lobato, G. *et al.* (2021) 'Frequent loss of heterozygosity in CRISPR-Cas9–edited early human embryos', *Proceedings of the National Academy of Sciences*, 118(22), p. e2004832117. doi: 10.1073/pnas.2004832117.

Alder, O. *et al.* (2010) 'Ring1B and Suv39h1 delineate distinct chromatin states at bivalent genes during early mouse lineage commitment', *Development*, 137(15), pp. 2483 LP – 2492. Available at: http://dev.biologists.org/content/137/15/2483.abstract.

Arima, T. *et al.* (2006) 'Loss of the maternal imprint in Dnmt3Lmat−/− mice leads to a differentiation defect in the extraembryonic tissue', *Developmental Biology*, 297(2), pp. 361–373. doi: https://doi.org/10.1016/j.ydbio.2006.05.003.

Azuara, V. *et al.* (2006) 'Chromatin signatures of pluripotent cell lines', *Nature Cell Biology*, 8, p. 532. Available at: http://dx.doi.org/10.1038/ncb1403.

Baba, A. *et al.* (2011) 'PKA-dependent regulation of the histone lysine demethylase complex PHF2–ARID5B', *Nature Cell Biology*, 13(6), pp. 668–675. doi: 10.1038/ncb2228.

Bahrampour, S. and Thor, S. (2016) 'Ctr9, a Key Component of the Paf1 Complex, Affects Proliferation and Terminal Differentiation in the Developing Drosophila Nervous System', *G3 Genes|Genomes|Genetics*, 6(10), pp. 3229–3239. doi: 10.1534/g3.116.034231.

Barau, J. *et al.* (2016) 'The DNA methyltransferase DNMT3C protects male germ cells from transposon activity', *Science*, 354(6314), pp. 909 LP – 912. Available at: http://science.sciencemag.org/content/354/6314/909.abstract.

Barker, D. J. P. (2004) 'The Developmental Origins of Adult Disease', *Journal of the American College of Nutrition*, 23(sup6), pp. 588S-595S. doi: 10.1080/07315724.2004.10719428.

Barker, D. J. P. *et al.* (2012) 'Resource allocation in utero and health in later life', *Placenta*, 33, pp. e30–e34. doi: 10.1016/J.PLACENTA.2012.06.009.

Bates, S. E. (2011) 'Classical Cytogenetics: Karyotyping Techniques BT - Human Pluripotent Stem Cells: Methods and Protocols', in Schwartz, P. H. and Wesselschmidt, R. L. (eds). Totowa, NJ: Humana Press, pp. 177–190. doi: 10.1007/978-1-61779-201-4_13.

Bernstein, B. *et al.* (2002) 'Methylation of histone H3 Lys 4 in coding regions of active genes',

*Proceedings of the National Academy of Sciences*, 99(13), pp. 8695–8700. doi: 10.1073/pnas.082249499.

Bernstein, B. E. *et al.* (2006) 'A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells', *Cell*, 125(2), pp. 315–326. doi: 10.1016/J.CELL.2006.02.041.

Bielski, C. M. *et al.* (2018) 'Genome doubling shapes the evolution and prognosis of advanced cancers', *Nature genetics*. 2018/07/16, 50(8), pp. 1189–1195. doi: 10.1038/s41588-018-0165-1.

Birnbaum, R. Y. *et al.* (2012) 'Coding exons function as tissue-specific enhancers of nearby genes', *Genome research*. 2012/03/22, 22(6), pp. 1059–1068. doi: 10.1101/gr.133546.111.

Blaise, S. *et al.* (2003) 'Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution', *Proceedings of the National Academy of Sciences*, 100(22), pp. 13013 LP – 13018. doi: 10.1073/pnas.2132646100.

Blakeley, P. *et al.* (2015) 'Defining the three cell lineages of the human blastocyst by single-cell RNA-seq', *Development (Cambridge, England)*, 142(18), pp. 3151–3165. doi: 10.1242/dev.123547.

Bogutz, A. B. *et al.* (2018) 'Transcription factor ASCL2 is required for development of the glycogen trophoblast cell lineage', *PLOS Genetics*, 14(8), p. e1007587. Available at: https://doi.org/10.1371/journal.pgen.1007587.

Bonhoure, N. *et al.* (2014) 'Quantifying ChIP-seq data: a spiking method providing an internal reference for sample-to-sample normalization', *Genome research*. 2014/04/07, 24(7), pp. 1157–1168. doi: 10.1101/gr.168260.113.

Bou-Nader, M. *et al.* (2020) 'Polyploidy spectrum: a new marker in HCC classification', *Gut*, 69(2), pp. 355 LP – 364. doi: 10.1136/gutjnl-2018-318021.

Bourc'his, D. *et al.* (2001) 'Dnmt3L and the Establishment of Maternal Genomic Imprints', *Science*, 294(5551), pp. 2536 LP – 2539. Available at: http://science.sciencemag.org/content/294/5551/2536.abstract.

Braude, P., Bolton, V. and Moore, S. (1988) 'Human gene expression first occurs between the four- and eight-cell stages of preimplantation development', *Nature*, 332(6163), pp. 459–461. doi: 10.1038/332459a0.

Brind'Amour, J. *et al.* (2015) 'An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations', *Nature Communications*, 6, p. 6033. Available at: https://doi.org/10.1038/ncomms7033.

Brodsky, W. Y. A. and Uryvaeva, I. V (1977) 'Cell Polyploidy: Its Relation to Tissue Growth and

Function', in Bourne, G. H., Danielli, J. F., and Jeon, K. W. B. T.-I. R. of C. (eds). Academic Press, pp. 275–332. doi: https://doi.org/10.1016/S0074-7696(08)60100-X.

Carrozza, M. J. *et al.* (2005) 'Histone H3 Methylation by Set2 Directs Deacetylation of Coding Regions by Rpd3S to Suppress Spurious Intragenic Transcription', *Cell*, 123(4), pp. 581–592. doi: https://doi.org/10.1016/j.cell.2005.10.023.

Chang, C.-W. and Parast, M. M. (2017) 'Human trophoblast stem cells: Real or not real?', *Placenta*, 60, pp. S57–S60. doi: 10.1016/J.PLACENTA.2017.01.003.

Chen, J. *et al.* (2021) 'EGR1 Overexpression Inhibits the Occurrence of Preeclampsia by Binding to MicroRNA-574 Promoter and Upregulating GAB1', *Reproductive Sciences*, 28(4), pp. 1112–1121. doi: 10.1007/s43032-020-00392-3.

Chiu, S.-Y., Maruyama, E. O. and Hsu, W. (2010) 'Derivation of Mouse Trophoblast Stem Cells from Blastocysts', *Journal of Visualized Experiments : JoVE*, (40), p. 1964. doi: 10.3791/1964.

Chosed, R. and Dent, S. Y. R. (2007) 'A Two-Way Street: LSD1 Regulates Chromatin Boundary Formation in S. pombe and Drosophila', *Molecular Cell*, 26(2), pp. 160–162. doi: 10.1016/J.MOLCEL.2007.04.009.

Chrysanthou, S. *et al.* (2018) 'A Critical Role of TET1/2 Proteins in Cell-Cycle Progression of Trophoblast Stem Cells', *Stem Cell Reports*, 10(4), pp. 1355–1368. doi: 10.1016/j.stemcr.2018.02.014.

Chuong, E. B. *et al.* (2013) 'Endogenous retroviruses function as species-specific enhancer elements in the placenta', *Nature genetics*. 2013/02/10, 45(3), pp. 325–329. doi: 10.1038/ng.2553.

Chuong, E. B. (2018) 'The placenta goes viral: Retroviruses control gene expression in pregnancy', *PLOS Biology*, 16(10), p. e3000028. Available at: https://doi.org/10.1371/journal.pbio.3000028.

Ciccone, David N *et al.* (2009) 'KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints', *Nature*, 461, p. 415. Available at: http://dx.doi.org/10.1038/nature08315.

Ciccone, D N *et al.* (2009) 'KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints', *Nature.*, 461. doi: 10.1038/nature08315.

Cinghu, S. *et al.* (2017) 'Intragenic Enhancers Attenuate Host Gene Expression', *Molecular Cell*, 68(1), pp. 104-117.e6. doi: https://doi.org/10.1016/j.molcel.2017.09.010.

Cinkornpumin, J. K. *et al.* (2020) 'Naive Human Embryonic Stem Cells Can Give Rise to Cells with a Trophoblast-like Transcriptome and Methylome', *Stem Cell Reports*, 15(1), pp. 198–213. doi: 10.1016/J.STEMCR.2020.06.003.

Coan, P. M., Ferguson-Smith, A. C. and Burton, G. J. (2005) 'Ultrastructural changes in the

interhaemal membrane and junctional zone of the murine chorioallantoic placenta across gestation', *Journal of Anatomy*, 207(6), pp. 783–796. doi: https://doi.org/10.1111/j.1469-7580.2005.00488.x.

Copp, A. J. (1995) 'Death before birth: clues from gene knockouts and mutations', *Trends in Genetics*, 11(3), pp. 87–93. doi: https://doi.org/10.1016/S0168-9525(00)89008-3.

Creyghton, M. P. *et al.* (2010) 'Histone H3K27ac separates active from poised enhancers and predicts developmental state', *Proceedings of the National Academy of Sciences of the United States of America*. 2010/11/24, 107(50), pp. 21931–21936. doi: 10.1073/pnas.1016071107.

Cullot, G. *et al.* (2019) 'CRISPR-Cas9 genome editing induces megabase-scale chromosomal truncations', *Nature Communications*, 10(1), p. 1136. doi: 10.1038/s41467-019-09006-2.

Dahl, J. A. *et al.* (2010) 'Histone H3 lysine 27 methylation asymmetry on developmentally-regulated promoters distinguish the first two lineages in mouse preimplantation embryos', *PLoS ONE*. doi: 10.1371/journal.pone.0009150.

Damsky, C. H., Fitzgerald, M. L. and Fisher, S. J. (1992) 'Distribution patterns of extracellular matrix components and adhesion receptors are intricately modulated during first trimester cytotrophoblast differentiation along the invasive pathway, in vivo.', *The Journal of Clinical Investigation*, 89(1), pp. 210–222. doi: 10.1172/JCI115565.

Day, D. S. *et al.* (2016) 'Comprehensive analysis of promoter-proximal RNA polymerase II pausing across mammalian cell types', *Genome Biology*, 17(1), p. 120. doi: 10.1186/s13059-016-0984-2.

Deussing, J. *et al.* (2002) 'Identification and Characterization of a Dense Cluster of Placenta-Specific Cysteine Peptidase Genes and Related Genes on Mouse Chromosome 13', *Genomics*, 79(2), pp. 225–240. doi: https://doi.org/10.1006/geno.2002.6696.

Dogan, N. *et al.* (2015) 'Occupancy by key transcription factors is a more accurate predictor of enhancer activity than histone modifications or chromatin accessibility', *Epigenetics & Chromatin*, 8(1), p. 16. doi: 10.1186/s13072-015-0009-5.

Dong, C. *et al.* (2020) 'Derivation of trophoblast stem cells from naïve human pluripotent stem cells', *eLife*, 9. doi: 10.7554/eLife.52504.

Dong, Z. *et al.* (2016) 'Low-pass whole-genome sequencing in clinical cytogenetics: a validated approach', *Genetics in Medicine*, 18(9), pp. 940–948. doi: 10.1038/gim.2015.199.

Dorighi, K. M. *et al.* (2017) 'Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation', *Molecular Cell*, 66(4), pp. 568-576.e4. doi: https://doi.org/10.1016/j.molcel.2017.04.018.

Draper, J. S. *et al.* (2004) 'Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells', *Nature Biotechnology*, 22(1), pp. 53–54. doi: 10.1038/nbt922.

Dudley, D. D. *et al.* (2005) 'Mechanism and Control of V(D)J Recombination versus Class Switch Recombination: Similarities and Differences', in Alt, F. W. B. T.-A. in I. (ed.). Academic Press, pp. 43–112. doi: https://doi.org/10.1016/S0065-2776(04)86002-4.

Duncan, A. W. *et al.* (2010) 'The ploidy conveyor of mature hepatocytes as a source of genetic variation', *Nature*, 467(7316), pp. 707–710. doi: 10.1038/nature09414.

Dyce, J. *et al.* (1987) 'Do trophectoderm and inner cell mass cells in the mouse blastocyst maintain discrete lineages?', *Development*, 100(4), pp. 685 LP – 698. Available at: http://dev.biologists.org/content/100/4/685.abstract.

Eden, E. *et al.* (2009) 'GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists', *BMC Bioinformatics*, 10(1), p. 48. doi: 10.1186/1471-2105-10-48.

Edmunds, J. W., Mahadevan, L. C. and Clayton, A. L. (2008) 'Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation', *The EMBO Journal*, 27(2), pp. 406–420. doi: https://doi.org/10.1038/sj.emboj.7601967.

Egan, B. *et al.* (2016) 'An Alternative Approach to ChIP-Seq Normalization Enables Detection of Genome-Wide Changes in Histone H3 Lysine 27 Trimethylation upon EZH2 Inhibition', *PLOS ONE*, 11(11), p. e0166438. Available at: https://doi.org/10.1371/journal.pone.0166438.

Erenpreisa, J. *et al.* (2005) 'Segregation of genomes in polyploid tumour cells following mitotic catastrophe', *Cell Biology International*, 29(12), pp. 1005–1011. doi: https://doi.org/10.1016/j.cellbi.2005.10.008.

Erenpreisa, J. *et al.* (2011) 'Polyploid tumour cells elicit paradiploid progeny through depolyploidizing divisions and regulated autophagic degradation', *Cell Biology International*, 35(7), pp. 687–695. doi: https://doi.org/10.1042/CBI20100762.

Evans, M. J. and Kaufman, M. H. (1981) 'Establishment in culture of pluripotential cells from mouse embryos', *Nature*, 292, p. 154. Available at: http://dx.doi.org/10.1038/292154a0.

F., J. H., Ittai, B.-P. and P., B. A. (2004) 'Mbd1 Is Recruited to both Methylated and Nonmethylated CpGs via Distinct DNA Binding Domains', *Molecular and Cellular Biology*, 24(8), pp. 3387–3395. doi: 10.1128/MCB.24.8.3387-3395.2004.

Fang, R *et al.* (2010) 'Human LSD2/KDM1b/AOF1 regulates gene transcription by modulating intragenic H3K4me2 methylation', *Mol Cell.*, 39. doi: 10.1016/j.molcel.2010.07.008.

Fang, Rui *et al.* (2010) 'Human LSD2/KDM1b/AOF1 Regulates Gene Transcription by Modulating Intragenic H3K4me2 Methylation', *Molecular Cell*, 39(2), pp. 222–233. doi: 10.1016/J.MOLCEL.2010.07.008.

Fang, R. *et al.* (2013) 'LSD2/KDM1B and Its Cofactor NPAC/GLYR1 Endow a Structural and Molecular Model for Regulation of H3K4 Demethylation', *Molecular Cell*, 49(3), pp. 558–570. doi: 10.1016/J.MOLCEL.2012.11.019.

Feinendegen, L. E. *et al.* (1962) 'Kinetics of Megacaryocyte Proliferation.', *Proceedings of the Society for Experimental Biology and Medicine*, 111(1), pp. 177–182. doi: 10.3181/00379727-111-27738.

Feng, J. *et al.* (2012) 'Identifying ChIP-seq enrichment using MACS', *Nature Protocols*, 7(9), pp. 1728–1740. doi: 10.1038/nprot.2012.101.

Flach, G. *et al.* (1982) 'The transition from maternal to embryonic control in the 2-cell mouse embryo.', *The EMBO Journal*, 1(6), pp. 681–686. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC553268/.

Fu, Y. *et al.* (2013) 'High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells', *Nature Biotechnology*, 31(9), pp. 822–826. doi: 10.1038/nbt.2623.

Galipeau, P. C. *et al.* (1996) '17p (p53) allelic losses, 4N (G2/tetraploid) populations, and progression to aneuploidy in Barrett's esophagus.', *Proceedings of the National Academy of Sciences of the United States of America*, 93(14), pp. 7081–7084. doi: 10.1073/pnas.93.14.7081.

Georgiades, P., Ferguson-Smith, A. C. and Burton, G. J. (2002) 'Comparative Developmental Anatomy of the Murine and Human Definitive Placentae', *Placenta*, 23(1), pp. 3–19. doi: 10.1053/PLAC.2001.0738.

Gorla, G. R., Malhi, H. and Gupta, S. (2001) 'Polyploidy associated with oxidative injury attenuates proliferative potential of cells', *Journal of Cell Science*, 114(16), pp. 2943–2951. doi: 10.1242/jcs.114.16.2943.

Gu, T.-P. *et al.* (2011) 'The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes', *Nature*, 477, p. 606. Available at: http://dx.doi.org/10.1038/nature10443.

Guidotti, J.-E. *et al.* (2003) 'Liver Cell Polyploidization: A Pivotal Role for Binuclear Hepatocytes *', *Journal of Biological Chemistry*, 278(21), pp. 19095–19101. doi: 10.1074/jbc.M300982200.

Guo, B. *et al.* (2014) 'Expression, regulation and function of Egr1 during implantation and decidualization in mice', *Cell cycle (Georgetown, Tex.)*, 13(16), pp. 2626–2640. doi: 10.4161/15384101.2014.943581.

Guo, G. *et al.* (2021) 'Human naive epiblast cells possess unrestricted lineage potential', *Cell Stem Cell*, 28(6), pp. 1040-1056.e6. doi: 10.1016/J.STEM.2021.02.025.

Hack, M., Klein, N. K. and Taylor, H. G. (1995) 'Long-Term Developmental Outcomes of Low Birth Weight Infants', *Source: The Future of Children Low Birth Weight (Spring*, 5(1), pp. 176–196. Available at: http://www.jstor.org/stable/1602514%0Ahttp://about.jstor.org/terms.

Haider, S. *et al.* (2018) 'Self-Renewing Trophoblast Organoids Recapitulate the Developmental Program of the Early Human Placenta', *Stem Cell Reports*. doi: 10.1016/J.STEMCR.2018.07.004.

Hanna, C. W. *et al.* (2018) 'MLL2 conveys transcription-independent H3K4 trimethylation in oocytes', *Nature Structural & Molecular Biology*, 25(1), pp. 73–82. doi: 10.1038/s41594-017-0013-5.

Hannibal, R. L. *et al.* (2014) 'Copy Number Variation Is a Fundamental Aspect of the Placental Genome', *PLOS Genetics*, 10(5), p. e1004290. Available at: https://doi.org/10.1371/journal.pgen.1004290.

Hannibal, R. L. and Baker, J. C. (2016) 'Selective Amplification of the Genome Surrounding Key Placental Genes in Trophoblast Giant Cells', *Current Biology*, 26(2), pp. 230–236. doi: 10.1016/J.CUB.2015.11.060.

Harris, J. R. (1998) 'Placental endogenous retrovirus (ERV): structural, functional, and evolutionary significance', *BioEssays*, 20(4), pp. 307–316. doi: https://doi.org/10.1002/(SICI)1521-1878(199804)20:4<307::AID-BIES7>3.0.CO;2-M.

Heintzman, N. D. *et al.* (2009) 'Histone modifications at human enhancers reflect global cell-type-specific gene expression', *Nature*. 2009/03/18, 459(7243), pp. 108–112. doi: 10.1038/nature07829.

Hemberger, M. *et al.* (2003) 'Differential expression of angiogenic and vasodilatory factors by invasive trophoblast giant cells depending on depth of invasion', *Developmental Dynamics*, 227(2), pp. 185–191. doi: https://doi.org/10.1002/dvdy.10291.

Hemberger, M. *et al.* (2010) 'ELF5-enforced transcriptional networks define an epigenetically regulated trophoblast stem cell compartment in the human placenta', *Human Molecular Genetics*, 19(12), pp. 2456–2467. Available at: http://dx.doi.org/10.1093/hmg/ddq128.

Herzing, L. B. *et al.* (1997) 'Xist has properties of the X-chromosome inactivation centre.', *Nature*, 386(6622), pp. 272–275. doi: 10.1038/386272a0.

Hsu, P. D. *et al.* (2013) 'DNA targeting specificity of RNA-guided Cas9 nucleases', *Nature Biotechnology*, 31(9), pp. 827–832. doi: 10.1038/nbt.2647.

Hughes, M., Dobric, N., Scott, Ian C., *et al.* (2004) 'The Hand1, Stra13 and Gcm1 transcription factors

override FGF signaling to promote terminal differentiation of trophoblast stem cells', *Developmental Biology*, 271(1), pp. 26–37. doi: 10.1016/J.YDBIO.2004.03.029.

Hughes, M., Dobric, N., Scott, Ian C, *et al.* (2004) 'The Hand1, Stra13 and Gcm1 transcription factors override FGF signaling to promote terminal differentiation of trophoblast stem cells', *Developmental Biology*, 271(1), pp. 26–37. doi: https://doi.org/10.1016/j.ydbio.2004.03.029.

Hughes, P. *et al.* (2007) 'The costs of using unauthenticated, over-passaged cell lines: how much more data do we need?', *BioTechniques*, 43(5), pp. 575–586. doi: 10.2144/000112598.

Ilgren, E. B. (1983) 'Control of trophoblastic growth', *Placenta*, 4(3), pp. 307–328. doi: https://doi.org/10.1016/S0143-4004(83)80010-1.

Illingworth, R. *et al.* (2008) 'A Novel CpG Island Set Identifies Tissue-Specific Methylation at Developmental Gene Loci', *PLoS Biology*. Edited by E. T. Liu, 6(1), p. e22. doi: 10.1371/journal.pbio.0060022.

Illingworth, R. S. *et al.* (2010) 'Orphan CpG islands identify numerous conserved promoters in the mammalian genome', *PLoS Genet.*, 6. doi: 10.1371/journal.pgen.1001134.

International Stem Cell Initiative *et al.* (2011) 'Screening ethnically diverse human embryonic stem cells identifies a chromosome 20 minimal amplicon conferring growth advantage', *Nature biotechnology*, 29(12), pp. 1132–1144. doi: 10.1038/nbt.2051.

Io, S. *et al.* (2021) 'Capturing human trophoblast development with naive pluripotent stem cells in vitro', *Cell Stem Cell*, 28(6), pp. 1023-1039.e13. doi: 10.1016/j.stem.2021.03.013.

Jackson-Grusby, L. L. *et al.* (1988) 'Chromosomal Mapping of the Prolactin/Growth Hormone Gene Family in the Mouse*', *Endocrinology*, 122(6), pp. 2462–2466. doi: 10.1210/endo-122-6-2462.

Jäger, U. *et al.* (2000) 'Follicular lymphomas' BCL-2/IgH junctions contain templated nucleotide insertions: novel insights into the mechanism of t(14;18) translocation.', *Blood*, 95(11), pp. 3520–3529.

Jentsch, I. *et al.* (2001) 'Karyotyping mouse chromosomes by multiplex-FISH (M-FISH)', *Chromosome Research*, 9(3), pp. 211–214. doi: 10.1023/A:1016696303479.

Jiang, F. and Doudna, J. A. (2017) 'CRISPR–Cas9 Structures and Mechanisms', *Annual Review of Biophysics*, 46(1), pp. 505–529. doi: 10.1146/annurev-biophys-062215-010822.

Jiang, M. *et al.* (2020) 'SMARCD3 is a potential prognostic marker and therapeutic target in CAFs', *Aging*. 2020/10/28, 12(20), pp. 20835–20861. doi: 10.18632/aging.104102.

Jin, C. *et al.* (2014) 'TET1 is a maintenance DNA demethylase that prevents methylation spreading in

differentiated cells', *Nucleic acids research*. 2014/05/29, 42(11), pp. 6956–6971. doi: 10.1093/nar/gku372.

Jinek, M. *et al.* (2012) 'A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity', *Science*. Available at: http://science.sciencemag.org/content/early/2012/06/27/science.1225829.abstract.

Juch, H. *et al.* (2012) 'HLA class I expression in the human placenta', *Wiener Medizinische Wochenschrift*, 162(9), pp. 196–200. doi: 10.1007/s10354-012-0070-7.

Kaiserman, D. *et al.* (2002) 'Comparison of Human Chromosome 6p25 with Mouse Chromosome 13 Reveals a Greatly Expanded Ov-Serpin Gene Repertoire in the Mouse', *Genomics*, 79(3), pp. 349–362. doi: https://doi.org/10.1006/geno.2002.6716.

Kalousek, D. and Dill, F. (1983) 'Chromosomal Mosaicism Confined to the Placenta in Human Conceptions', *Science*, 221(4611), pp. 665–667. doi: 10.1126/science.6867735.

Kalousek, D. K. (1994) 'Confined placental mosaicism and intrauterine fetal development', *Placenta*, 15(3), pp. 219–230. doi: https://doi.org/10.1016/0143-4004(94)90014-0.

Kearns, N. A. *et al.* (2015) 'Functional annotation of native enhancers with a Cas9–histone demethylase fusion', *Nature Methods*, 12(5), pp. 401–403. doi: 10.1038/nmeth.3325.

Kerenyi, M. A. *et al.* (2013) 'Histone demethylase Lsd1 represses hematopoietic stem and progenitor cell signatures during blood cell maturation', *eLife*. Edited by S. J. Morrison, 2, p. e00633. doi: 10.7554/eLife.00633.

Kevin, E. *et al.* (2001) 'Hybrid vigor, fetal overgrowth, and viability of mice derived by nuclear cloning and tetraploid embryo complementation', *Proceedings of the National Academy of Sciences*, 98(11), pp. 6209–6214. doi: 10.1073/pnas.101118898.

Khraiwesh, B. (2011) 'Using Nuclear Run-On Transcription Assays in RNAi Studies BT - RNAi and Plant Gene Function Analysis: Methods and Protocols', in Kodama, H. and Komamine, A. (eds). Totowa, NJ: Humana Press, pp. 199–209. doi: 10.1007/978-1-61779-123-9_14.

Kim, J. *et al.* (2015) 'LSD1 is essential for oocyte meiotic progression by regulating CDC25B expression in mice', *Nature Communications*, 6(1), p. 10116. doi: 10.1038/ncomms10116.

Kim, T.-K. *et al.* (2010) 'Widespread transcription at neuronal activity-regulated enhancers', *Nature*, 465(7295), pp. 182–187. doi: 10.1038/nature09033.

KLIMAN, H. J. *et al.* (1986) 'Purification, Characterization, and in vitro Differentiation of Cytotrophoblasts from Human Term Placentae*', *Endocrinology*, 118(4), pp. 1567–1582. doi:

10.1210/endo-118-4-1567.

Klug, A. (2010) 'The Discovery of Zinc Fingers and Their Applications in Gene Regulation and Genome Manipulation', *Annual Review of Biochemistry*, 79(1), pp. 213–231. doi: 10.1146/annurev-biochem-010909-095056.

Konermann, S. *et al.* (2015) 'Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex', *Nature*, 517, p. 583. Available at: http://dx.doi.org/10.1038/nature14136.

Koppes, E. *et al.* (2019) 'Klf14 is an imprinted transcription factor that regulates placental growth', *Placenta*, 88, pp. 61–67. doi: https://doi.org/10.1016/j.placenta.2019.09.013.

Krishnan, L., Nguyen, T. and McComb, S. (2013) 'From mice to women: the conundrum of immunity to infection during pregnancy', *Journal of Reproductive Immunology*, 97(1), pp. 62–73. doi: 10.1016/j.jri.2012.10.015.

Kudryavtsev, B. N. *et al.* (1993) 'Human hepatocyte polyploidization kinetics in the course of life cycle', *Virchows Archiv B*, 64(1), p. 387. doi: 10.1007/BF02915139.

Kui, S. V. *et al.* (2000) 'Cloning of a Mammalian Transcriptional Activator That Binds Unmethylated CpG Motifs and Shares a CXXC Domain with DNA Methyltransferase, Human Trithorax, and Methyl-CpG Binding Domain Protein 1', *Molecular and Cellular Biology*, 20(6), pp. 2108–2121. doi: 10.1128/MCB.20.6.2108-2121.2000.

Kupershmit, I. *et al.* (2014) 'KDM4C (GASC1) lysine demethylase is associated with mitotic chromatin and regulates chromosome segregation during mitosis', *Nucleic Acids Research*, 42(10), pp. 6168–6182. doi: 10.1093/nar/gku253.

Lall, N. *et al.* (2013) 'Viability Reagent, PrestoBlue, in Comparison with Other Available Reagents, Utilized in Cytotoxicity and Antimicrobial Assays', *International journal of microbiology*. 2013/04/04, 2013, p. 420601. doi: 10.1155/2013/420601.

Lan, F. *et al.* (2007) 'S. pombe LSD1 Homologs Regulate Heterochromatin Propagation and Euchromatic Gene Transcription', *Molecular Cell*, 26(1), pp. 89–101. doi: https://doi.org/10.1016/j.molcel.2007.02.023.

Lang, U. *et al.* (2003) 'Uterine blood flow—a determinant of fetal growth', *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 110, pp. S55–S61. doi: https://doi.org/10.1016/S0301-2115(03)00173-8.

Larson, M. H. *et al.* (2013) 'CRISPR interference (CRISPRi) for sequence-specific control of gene expression', *Nature protocols*. 2013/10/17, 8(11), pp. 2180–2196. doi: 10.1038/nprot.2013.132.

Latos, P. A. *et al.* (2015) 'Fgf and Esrrb integrate epigenetic and transcriptional networks that regulate self-renewal of trophoblast stem cells', *Nature Communications*, 6(1), p. 7776. doi: 10.1038/ncomms8776.

Latos, P. A. and Hemberger, M. (2016) 'From the stem of the placental tree: trophoblast stem cells and their progeny', *Development*, 143(20), pp. 3650 LP – 3660. Available at: http://dev.biologists.org/content/143/20/3650.abstract.

Le, C. *et al.* (2013) 'Multiplex Genome Engineering Using CRISPR/Cas Systems', *Science*, 339(6121), pp. 819–823. doi: 10.1126/science.1231143.

Lee, C. Q. E. *et al.* (2018) 'Integrin α2 marks a niche of trophoblast progenitor cells in first trimester human placenta', *Development*, 145(16). Available at: http://dev.biologists.org/content/145/16/dev162305.abstract.

Lee, J. T. *et al.* (1996) 'A 450 kb transgene displays properties of the mammalian X-inactivation center.', *Cell*, 86(1), pp. 83–94. doi: 10.1016/s0092-8674(00)80079-3.

Lieber, M. R. (2010) 'The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End-Joining Pathway', *Annual Review of Biochemistry*, 79(1), pp. 181–211. doi: 10.1146/annurev.biochem.052308.093131.

Liu, M. *et al.* (2019) 'Methodologies for Improving HDR Efficiency', *Frontiers in Genetics*, 9. doi: 10.3389/fgene.2018.00691.

Liu, Y. *et al.* (2019) 'Multi-omic measurements of heterogeneity in HeLa cells across laboratories', *Nature Biotechnology*, 37(3), pp. 314–322. doi: 10.1038/s41587-019-0037-y.

Love, M. I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome biology*, 15(12), p. 550. doi: 10.1186/s13059-014-0550-8.

Luger, K. *et al.* (1997) 'Crystal structure of the nucleosome core particle at 2.8 Å resolution', *Nature*, 389(6648), pp. 251–260. doi: 10.1038/38444.

Lund, R. J. *et al.* (2012) 'High-throughput karyotyping of human pluripotent stem cells', *Stem cell research*. 2012/07/11, 9(3), pp. 192–195. doi: 10.1016/j.scr.2012.06.008.

Ma, Y. *et al.* (2002) 'Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination.', *Cell*, 108(6), pp. 781–794. doi: 10.1016/s0092-8674(02)00671-2.

Maley, C. C. *et al.* (2006) 'Genetic clonal diversity predicts progression to esophageal adenocarcinoma.', *Nature genetics*, 38(4), pp. 468–473. doi: 10.1038/ng1768.

Mali, P. *et al.* (2013) 'CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering', *Nature Biotechnology*, 31(9), pp. 833–838. doi: 10.1038/nbt.2675.

Maltepe, E. *et al.* (2005) 'Hypoxia-inducible factor-dependent histone deacetylase activity determines stem cell fate in the placenta', *Development*, 132(15), pp. 3393–3403. doi: 10.1242/dev.01923.

Marahrens, Y. *et al.* (1997) 'Xist-deficient mice are defective in dosage compensation but not spermatogenesis.', *Genes & development*, 11(2), pp. 156–166. doi: 10.1101/gad.11.2.156.

Margall-Ducos, G. *et al.* (2007) 'Liver tetraploidization is controlled by a new process of incomplete cytokinesis', *Journal of Cell Science*, 120(20), pp. 3633–3639. doi: 10.1242/jcs.016907.

McManus, K. J. *et al.* (2006) 'Dynamic Changes in Histone H3 Lysine 9 Methylations: IDENTIFICATION OF A MITOSIS-SPECIFIC FUNCTION FOR DYNAMIC METHYLATION IN CHROMOSOME CONGRESSION AND SEGREGATION* ', *Journal of Biological Chemistry*, 281(13), pp. 8888–8897. doi: 10.1074/jbc.M505323200.

McMaster, M. T. *et al.* (1995) 'Human placental HLA-G expression is restricted to differentiated cytotrophoblasts.', *The Journal of Immunology*, 154(8), pp. 3771 LP – 3778. Available at: http://www.jimmunol.org/content/154/8/3771.abstract.

Mehrotra, S. *et al.* (2008) 'Endocycling cells do not apoptose in response to DNA rereplication genotoxic stress.', *Genes & development*, 22(22), pp. 3158–3171. doi: 10.1101/gad.1710208.

Molla, K. A. and Yang, Y. (2020) 'Predicting CRISPR/Cas9-Induced Mutations for Precise Genome Editing', *Trends in Biotechnology*, 38(2), pp. 136–141. doi: https://doi.org/10.1016/j.tibtech.2019.08.002.

Murray, A., Sienerth, A. R. and Hemberger, M. (2016) 'Plet1 is an epigenetically regulated cell surface protein that provides essential cues to direct trophoblast stem cell differentiation', *Scientific Reports*, 6(1), p. 25112. doi: 10.1038/srep25112.

Nelson, J. D., Denisenko, O. and Bomsztyk, K. (2006) 'Protocol for the fast chromatin immunoprecipitation (ChIP) method', *Nature Protocols*, 1(1), pp. 179–185. doi: 10.1038/nprot.2006.27.

Niakan, K. K. *et al.* (2012) 'Human pre-implantation embryo development', *Development*, 139(5), pp. 829 LP – 841. Available at: http://dev.biologists.org/content/139/5/829.abstract.

Noma, K., Allis, C. D. and Grewal, S. I. (2001) 'Transitions in distinct histone H3 methylation patterns

at the heterochromatin  domain boundaries.', *Science (New York, N.Y.)*, 293(5532), pp. 1150–1155. doi: 10.1126/science.1064150.

Noma, K. and Grewal, S. I. S. (2002) 'Histone H3 lysine 4 methylation is mediated by Set1 and promotes maintenance of active chromatin states in fission yeast', *Proceedings of the National Academy of Sciences of the United States of America*. 2002/08/22, 99 Suppl 4(Suppl 4), pp. 16438–16445. doi: 10.1073/pnas.182436399.

Okae, H. *et al.* (2018) 'Derivation of Human Trophoblast Stem Cells', *Cell Stem Cell*, 22(1), pp. 50-63.e6. doi: 10.1016/j.stem.2017.11.004.

Okano, M. *et al.* (1999) 'DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development', *Cell*, 99(3), pp. 247–257. doi: 10.1016/S0092-8674(00)81656-6.

Olaharski, A. J. *et al.* (2006) 'Tetraploidy and chromosomal instability are early events during cervical carcinogenesis', *Carcinogenesis*, 27(2), pp. 337–343. doi: 10.1093/carcin/bgi218.

Ooi, S. K. T. *et al.* (2007) 'DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA', *Nature*, 448(7154), pp. 714–717. doi: 10.1038/nature05987.

Orlando, D. A. *et al.* (2014) 'Quantitative ChIP-Seq Normalization Reveals Global Modulation of the Epigenome', *Cell Reports*, 9(3), pp. 1163–1170. doi: https://doi.org/10.1016/j.celrep.2014.10.018.

Orwig, K. E. *et al.* (1997) 'Identification and Characterization of a Mouse Homolog for Decidual/Trophoblast Prolactin-Related Protein.', *Endocrinology*, 138 12, pp. 5511–5517.

Owens, G. K. and Schwartz, S. M. (1983) 'Vascular smooth muscle cell hypertrophy and hyperploidy in the Goldblatt hypertensive rat.', *Circulation Research*, 53(4), pp. 491–501. doi: 10.1161/01.RES.53.4.491.

Pan, G. *et al.* (2007) 'Whole-Genome Analysis of Histone H3 Lysine 4 and Lysine 27 Methylation in Human Embryonic Stem Cells', *Cell Stem Cell*, 1(3), pp. 299–312. doi: https://doi.org/10.1016/j.stem.2007.08.003.

Pattanayak, V. *et al.* (2013) 'High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity', *Nature Biotechnology*, 31(9), pp. 839–843. doi: 10.1038/nbt.2673.

Peinado, H., Olmeda, D. and Cano, A. (2007) 'Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype?', *Nature Reviews Cancer*, 7(6), pp. 415–428. doi: 10.1038/nrc2131.

Penny, G. D. *et al.* (1996) 'Requirement for Xist in X chromosome inactivation.', *Nature*, 379(6561), pp. 131–137. doi: 10.1038/379131a0.

Perez-Garcia, V. *et al.* (2018) 'Placentation defects are highly prevalent in embryonic lethal mouse mutants', *Nature*, 555, p. 463. Available at: http://dx.doi.org/10.1038/nature26002.

Perez-Garcia, V. *et al.* (2021) 'BAP1/ASXL complex modulation regulates epithelial-mesenchymal transition during trophoblast differentiation and invasion', *eLife*. Edited by L. Solnica-Krezel, E. E. Morrisey, and L. Solnica-Krezel, 10, p. e63254. doi: 10.7554/eLife.63254.

Perez-Pinera, P. *et al.* (2013) 'RNA-guided gene activation by CRISPR-Cas9-based transcription factors', *Nature methods*, 10(10), pp. 973–976. doi: 10.1038/nmeth.2600.

Pinkel, D. *et al.* (1998) 'High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays', *Nature Genetics*, 20(2), pp. 207–211. doi: 10.1038/2524.

Pokholok, D. K. *et al.* (2005) 'Genome-wide Map of Nucleosome Acetylation and Methylation in Yeast', *Cell*, 122(4), pp. 517–527. doi: 10.1016/J.CELL.2005.06.026.

Qi, L. S. *et al.* (2013) 'Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression', *Cell*, 152(5), pp. 1173–1183. doi: 10.1016/j.cell.2013.02.022.

Rayasam, G. V. *et al.* (2003) 'NSD1 is essential for early post-implantation development and has a catalytically active SET domain', *The EMBO Journal*, 22(12), pp. 3153–3163. doi: https://doi.org/10.1093/emboj/cdg288.

Rebuzzini, P. *et al.* (2016) 'Achilles' heel of pluripotent stem cells: genetic, genomic and epigenetic variations during prolonged culture', *Cellular and Molecular Life Sciences*, 73(13), pp. 2453–2466. doi: 10.1007/s00018-016-2171-8.

Reik, W. and Walter, J. (2001) 'Genomic imprinting: parental influence on the genome', *Nature Reviews Genetics*, 2(1), pp. 21–32. doi: 10.1038/35047554.

Rice, J. C. *et al.* (2003) 'Histone Methyltransferases Direct Different Degrees of Methylation to Define Distinct Chromatin Domains', *Molecular Cell*, 12(6), pp. 1591–1598. doi: https://doi.org/10.1016/S1097-2765(03)00479-9.

Richart, R. (1961) 'Studies of Placental Morphogenesis I. Radioautographic Studies of Human Placenta Utilizing Tritiated Thymidine.', *Proceedings of the Society for Experimental Biology and Medicine*, 106(4), pp. 829–831. doi: 10.3181/00379727-106-26490.

Rodolphe, B. *et al.* (2007) 'CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes', *Science*, 315(5819), pp. 1709–1712. doi: 10.1126/science.1138140.

Rossant, J., Gardner, R. L. and Alexandre, H. L. (1978) 'Investigation of the potency of cells from the postimplantation mouse embryo by blastocyst injection: a preliminary report', *Development*, 48(1), pp. 239–247. doi: 10.1242/dev.48.1.239.

Rudolph, T. *et al.* (2007) 'Heterochromatin Formation in Drosophila Is Initiated through Active Removal of H3K4 Methylation by the LSD1 Homolog SU(VAR)3-3', *Molecular Cell*, 26(1), pp. 103–115. doi: https://doi.org/10.1016/j.molcel.2007.02.025.

Sado, T. *et al.* (2000) 'X Inactivation in the Mouse Embryo Deficient for Dnmt1: Distinct Effect of Hypomethylation on Imprinted and Random X Inactivation', *Developmental Biology*, 225(2), pp. 294–303. doi: https://doi.org/10.1006/dbio.2000.9823.

Samuel, C. A., Allen, W. R. and Steven, D. H. (1974) 'Studies on the Equine Placenta', *Reproducttive Fertility*, 41(2), pp. 441–445. doi: 10.1530/jrf.0.0410441.

San Filippo, J., Sung, P. and Klein, H. (2008) 'Mechanism of Eukaryotic Homologous Recombination', *Annual Review of Biochemistry*, 77(1), pp. 229–257. doi: 10.1146/annurev.biochem.77.061306.125255.

Santos, F. *et al.* (2002) 'Dynamic Reprogramming of DNA Methylation in the Early Mouse Embryo', *Developmental Biology*, 241(1), pp. 172–182. doi: 10.1006/DBIO.2001.0501.

Sauer, B. and Henderson, N. (1988) 'Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1.', *Proceedings of the National Academy of Sciences of the United States of America*, 85(14), pp. 5166–5170. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC281709/.

Schim van der Loeff, I. *et al.* (2021) 'Defective neutrophil development and specific granule deficiency caused by a homozygous splice-site mutation in SMARCD2', *The Journal of allergy and clinical immunology*. 2020/12/03, 147(6), pp. 2381-2385.e2. doi: 10.1016/j.jaci.2020.11.025.

Schrode, N. *et al.* (2013) 'Anatomy of a blastocyst: cell behaviors driving cell fate choice and morphogenesis in the early mouse embryo', *Genesis (New York, N.Y. : 2000)*, 51(4), pp. 219–233. doi: 10.1002/dvg.22368.

Scott, I. C. *et al.* (2000) 'The HAND1 Basic Helix-Loop-Helix Transcription Factor Regulates Trophoblast Differentiation via Multiple Mechanisms', *Molecular and Cellular Biology*, 20(2), pp. 530–541. doi: 10.1128/MCB.20.2.530-541.2000.

Seisenberger, S. *et al.* (2012) 'Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers', *Philos Trans R Soc Lond B Biol Sci.*, 368. doi: 10.1098/rstb.2011.0330.

Senner, C. E. *et al.* (2020) 'TET1 and 5-Hydroxymethylation Preserve the Stem Cell State of Mouse Trophoblast', *Stem Cell Reports*, 15(6), pp. 1301–1316. doi: https://doi.org/10.1016/j.stemcr.2020.04.009.

Shi, Y.-J. *et al.* (2005) 'Regulation of LSD1 Histone Demethylase Activity by Its Associated Factors', *Molecular Cell*, 19(6), pp. 857–864. doi: 10.1016/J.MOLCEL.2005.08.027.

Shi, Yujiang *et al.* (2004) 'Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1', *Cell*, 119(7), pp. 941–953. doi: 10.1016/J.CELL.2004.12.012.

Simmons, D. G. *et al.* (2008) 'Early patterning of the chorion leads to the trilaminar trophoblast cell structure in the placental labyrinth', *Development*, 135(12), pp. 2083–2091. doi: 10.1242/dev.020099.

Simmons, D. G., Fortier, A. L. and Cross, J. C. (2007) 'Diverse subtypes and developmental origins of trophoblast giant cells in the mouse placenta', *Developmental Biology*, 304(2), pp. 567–578. doi: https://doi.org/10.1016/j.ydbio.2007.01.009.

Smyth, G. K. (2004) 'Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments', *Statistical Applications in Genetics and Molecular Biology*, 3(1). doi: doi:10.2202/1544-6115.1027.

Soares, M. J. *et al.* (1998) 'The uteroplacental prolactin family and pregnancy.', *Biology of reproduction*, 58(2), pp. 273–284. doi: 10.1095/biolreprod58.2.273.

Soncin, F. *et al.* (2018) 'Comparative analysis of mouse and human placentae across gestation reveals species-specific regulators of placental development', *Development*. Available at: http://dev.biologists.org/content/early/2018/01/19/dev.156273.abstract.

Spits, C. *et al.* (2008) 'Recurrent chromosomal abnormalities in human embryonic stem cells.', *Nature biotechnology*, 26(12), pp. 1361–1363. doi: 10.1038/nbt.1510.

Sproul, D. and Meehan, R. R. (2013) 'Genomic insights into cancer-associated aberrant CpG island hypermethylation', *Briefings in functional genomics*. 2013/01/21, 12(3), pp. 174–190. doi: 10.1093/bfgp/els063.

Starostik, M. R., Sosina, O. A. and McCoy, R. C. (2020) 'Single-cell analysis of human embryos reveals diverse patterns of aneuploidy and mosaicism', *Genome research*. 2020/07/08, 30(6), pp. 814–825. doi: 10.1101/gr.262774.120.

Sun, N. and Zhao, H. (2013) 'Transcription activator-like effector nucleases (TALENs): A highly efficient and versatile tool for genome editing', *Biotechnology and Bioengineering*, 110(7), pp. 1811–

1821. doi: https://doi.org/10.1002/bit.24890.

Sundaram, M. *et al.* (2004) 'Neosis: A Novel Type of Cell Division in Cancer', *Cancer Biology & Therapy*, 3(2), pp. 207–218. doi: 10.4161/cbt.3.2.663.

Sung, P. and Klein, H. (2006) 'Mechanism of homologous recombination: mediators and helicases take on regulatory functions', *Nature Reviews Molecular Cell Biology*, 7(10), pp. 739–750. doi: 10.1038/nrm2008.

Taapken, S. M. *et al.* (2011) 'Karyotypic abnormalities in human induced pluripotent stem cells and embryonic stem cells', *Nature Biotechnology*, 29(4), pp. 313–314. doi: 10.1038/nbt.1835.

Tanaka, S. *et al.* (1998) 'Promotion of Trophoblast Stem Cell Proliferation by FGF4', *Science*, 282(5396), pp. 2072 LP – 2075. Available at: http://science.sciencemag.org/content/282/5396/2072.abstract.

Tanaka, S. (2006) 'Derivation and Culture of Mouse Trophoblast Stem Cells In Vitro BT - Embryonic Stem Cell Protocols: Volume 1: Isolation and Characterization', in Turksen, K. (ed.). Totowa, NJ: Humana Press, pp. 35–44. doi: 10.1385/1-59745-037-5:35.

Tang, M. *et al.* (2021) 'SIRT7: a sentinel of genome stability', *Open biology*. 2021/06/16, 11(6), p. 210047. doi: 10.1098/rsob.210047.

Tie, F. *et al.* (2009) 'CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing', *Development*, 136(18), pp. 3131–3141. doi: 10.1242/dev.037127.

Trevino, A. E. and Zhang, F. (2014) 'Chapter Eight - Genome Editing Using Cas9 Nickases', in Doudna, J. A. and Sontheimer, E. J. B. T.-M. in E. (eds) *The Use of CRISPR/Cas9, ZFNs, and TALENs in Generating Site-Specific Genome Alterations*. Academic Press, pp. 161–174. doi: https://doi.org/10.1016/B978-0-12-801185-0.00008-8.

Turco, M. Y. *et al.* (2018) 'Trophoblast organoids as a model for maternal–fetal interactions during human placentation', *Nature*, 564(7735), pp. 263–267. doi: 10.1038/s41586-018-0753-3.

Turner, B. M. (2002) 'Cellular Memory and the Histone Code', *Cell*, 111(3), pp. 285–291. doi: https://doi.org/10.1016/S0092-8674(02)01080-2.

Uy, G. D., Downs, K. M. and Gardner, R. L. (2002) 'Inhibition of trophoblast stem cell potential in chorionic ectoderm coincides with occlusion of the ectoplacental cavity in the mouse', *Development*, 129(16), pp. 3913 LP – 3924. Available at: http://dev.biologists.org/content/129/16/3913.abstract.

Vakoc, C. R. *et al.* (2005) 'Histone H3 Lysine 9 Methylation and HP1γ Are Associated with Transcription Elongation through Mammalian Chromatin', *Molecular Cell*, 19(3), pp. 381–391. doi:

https://doi.org/10.1016/j.molcel.2005.06.011.

Vićovac, L. and Aplin, J. D. (1996) 'Epithelial-Mesenchymal Transition during Trophoblast Differentiation', *Cells Tissues Organs*, 156(3), pp. 202–216. doi: 10.1159/000147847.

Vitrat, N. *et al.* (1998) 'Endomitosis of Human Megakaryocytes Are Due to Abortive Mitosis', *Blood*, 91(10), pp. 3711–3723. doi: 10.1182/blood.V91.10.3711.

Wakuda, K. and Yoshida, Y. (1992) 'DNA ploidy and proliferative characteristics of human trophoblasts.', *Acta obstetricia et gynecologica Scandinavica*, 71(1), pp. 12–16. doi: 10.3109/00016349209007940.

Wang, J. *et al.* (2007) 'Opposing LSD1 complexes function in developmental gene activation and repression programmes', *Nature*, 446(7138), pp. 882–887. doi: 10.1038/nature05671.

Wang, J. *et al.* (2009) 'The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation', *Nature Genetics*, 41(1), pp. 125–129. doi: 10.1038/ng.268.

Welzel, N. *et al.* (2001) 'Templated Nucleotide Addition and Immunoglobulin JH-Gene Utilization in t(11;14) Junctions: Implications for the Mechanism of Translocation and the Origin of Mantle Cell Lymphoma1', *Cancer Research*, 61(4), pp. 1629–1636.

Werling, U. and Schorle, H. (2002) 'Transcription Factor Gene AP-2γ Essential for Early Murine Development', *Molecular and Cellular Biology*, 22(9), pp. 3149–3156. doi: 10.1128/MCB.22.9.3149-3156.2002.

Whyte, W. A. *et al.* (2012) 'Enhancer decommissioning by LSD1 during embryonic stem cell differentiation', *Nature*, 482(7384), pp. 221–225. doi: 10.1038/nature10805.

Wiedenheft, B., Sternberg, S. H. and Doudna, J. A. (2012) 'RNA-guided genetic silencing systems in bacteria and archaea', *Nature*, 482(7385), pp. 331–338. doi: 10.1038/nature10886.

Wiemers, D. O. *et al.* (2003) 'The Mouse Prolactin Gene Family Locus', *Endocrinology*, 144(1), pp. 313–325. doi: 10.1210/en.2002-220724.

Wildman, D. E. *et al.* (2006) 'Evolution of the mammalian placenta revealed by phylogenetic analysis', *Proceedings of the National Academy of Sciences of the United States of America*, 103(9), pp. 3203 LP – 3208. Available at: http://www.pnas.org/content/103/9/3203.abstract.

Wilkinson, P. D. *et al.* (2019) 'The Polyploid State Restricts Hepatocyte Proliferation and Liver Regeneration in Mice', *Hepatology*, 69(3), pp. 1242–1258. doi: https://doi.org/10.1002/hep.30286.

Wilson, B. G. *et al.* (2014) 'Residual complexes containing SMARCA2 (BRM) underlie the oncogenic drive of SMARCA4 (BRG1) mutation', *Molecular and cellular biology*. 2014/01/13, 34(6), pp. 1136–

1144. doi: 10.1128/MCB.01372-13.

Wray, G. A. *et al.* (2003) 'The Evolution of Transcriptional Regulation in Eukaryotes', *Molecular Biology and Evolution*, 20(9), pp. 1377–1419. doi: 10.1093/molbev/msg140.

Wu, H. and Zhang, Y. (2014) 'Reversing DNA Methylation: Mechanisms, Genomics, and Biological Functions', *Cell*, 156(0), pp. 45–68. doi: 10.1016/j.cell.2013.12.019.

Yan, C. T. *et al.* (2007) 'IgH class switching and translocations use a robust non-classical end-joining pathway.', *Nature*, 449(7161), pp. 478–482. doi: 10.1038/nature06020.

Yang, Y., Dai, Z. and Dai, X. (2019) 'Insights into active intragenic enhancers', *Biochemical and Biophysical Research Communications*, 515(3), pp. 423–428. doi: https://doi.org/10.1016/j.bbrc.2019.05.160.

Yesbolatova, A. *et al.* (2020) 'The auxin-inducible degron 2 technology provides sharp degradation control in yeast, mammalian cells, and mice', *Nature Communications*, 11(1), p. 5701. doi: 10.1038/s41467-020-19532-z.

Yoder, J. A. *et al.* (1997) 'DNA (cytosine-5)-methyltransferases in mouse cells and tissues. studies with a mechanism-based probe', *Journal of Molecular Biology*, 270(3), pp. 385–395. doi: 10.1006/JMBI.1997.1125.

Zhang, H.-M. *et al.* (2015) 'AnimalTFDB 2.0: a resource for expression, prediction and functional study of animal transcription factors', *Nucleic acids research*. 2014/09/27, 43(Database issue), pp. D76–D81. doi: 10.1093/nar/gku887.

Zhang, S. *et al.* (2014) 'Generation of cancer stem-like cells through the formation of polyploid giant cancer cells', *Oncogene*, 33(1), pp. 116–128. doi: 10.1038/onc.2013.96.

Zhang, S. *et al.* (2021) 'Silencing of AFAP1-AS1 lncRNA impairs cell proliferation and migration by epigenetically promoting DUSP5 expression in pre-eclampsia', *Journal of Cellular Biochemistry*, n/a(n/a). doi: https://doi.org/10.1002/jcb.30072.

Zhang, W., Feng, J. and Li, Q. (2020) 'The replisome guides nucleosome assembly during DNA replication', *Cell & Bioscience*, 10(1), p. 37. doi: 10.1186/s13578-020-00398-z.

Zhang, Y. *et al.* (2008) 'Model-based Analysis of ChIP-Seq (MACS)', *Genome Biology*, 9(9), p. R137. doi: 10.1186/gb-2008-9-9-r137.

Zhu, D. *et al.* (2014) 'Lysine-specific demethylase 1 regulates differentiation onset and migration of trophoblast stem cells', *Nature Communications*, 5, p. 3174. Available at: http://dx.doi.org/10.1038/ncomms4174.

Zimmet, J. and Ravid, K. (2000) 'Polyploidy: Occurrence in nature, mechanisms, and significance for the megakaryocyte-platelet system', *Experimental Hematology*, 28(1), pp. 3–16. doi: https://doi.org/10.1016/S0301-472X(99)00124-1.

Zohdi, V. *et al.* (2012) 'Low Birth Weight due to Intrauterine Growth Restriction and/or Preterm Birth: Effects on Nephron Number and Long-Term Renal Health', *International journal of nephrology*. 2012/08/27, 2012, p. 136942. doi: 10.1155/2012/136942.

Zybina, T. G. *et al.* (2004) 'Genome multiplication of extravillous trophoblast cells in human placenta in the course of differentiation and invasion into endometrium and myometrium. II. Mechanisms of polyploidization', *Tsitologiia*, 46(7), pp. 640–648. Available at: http://europepmc.org/abstract/MED/15473375.