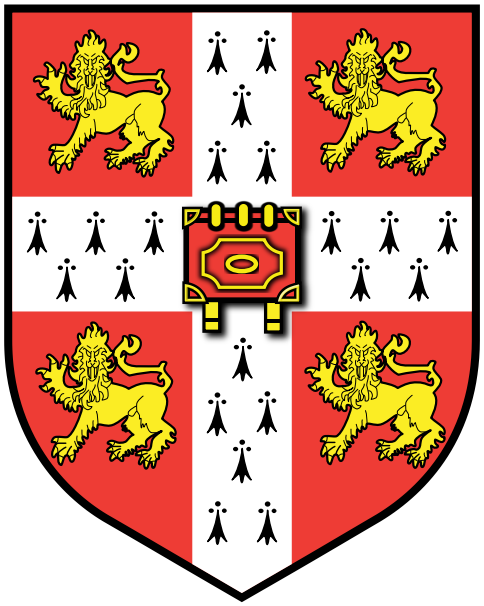
Investigating the role of the Polycomb Repressive Complex 2 in human in vitro pancreatic differentiation



Charlene Fabian

This thesis is submitted for the degree of Doctor of Philosophy (Ph.D.)

Darwin College | Babraham Institute

University of Cambridge

January 2021

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text.

I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

As stipulated by the Degree Committee for the Faculty of Biology, this dissertation does not exceed 60,000 words in length (excluding figures, photographs, tables appendices and bibliography).

Summary

Investigating the role of the Polycomb Repressive Complex 2 in human in vitro pancreatic differentiation.

The potential to use human pluripotent stem cells in regenerative medicine is an idea that continues to excite and captivate much of the science community, decades after the initial discovery of human embryonic stem cells. This interest has grown with the discovery of induced human pluripotent stem cells, combined with the continual improvement of differentiation protocols. These protocols allow the formation of an ever-increasing pool of mature and functional cell types within an *in vitro* laboratory environment. However, a number of important questions still exist in terms of the production of these cells, including their functionality and how closely these cells resemble the *in vivo* counterparts. One such area still to be explored is the epigenetic aspect of the *in vitro* derived cells. Beyond this understanding, the potential to apply epigenetic modulation to improve differentiation programmes, and/or the final cell product, is an important area of research.

In this dissertation, I use the *in vitro* production of human pancreatic endocrine cells as a model to study histone methylation changes, and associated transcription, of differentiating cells produced in culture from a pluripotent stem cell population. I describe the dynamic epigenetic and transcriptional changes that occur in a bulk population of human pluripotent cells developing through a number of intermediary cell types to form pancreatic endocrine cells. This analysis was then used to study the differences between *in vitro* end-cell populations and *in vivo* derived cells, highlighting the large variation in gene expression and persistent H3K27me3 modifications present in the *in vitro* cells. I show that manipulation of Polycomb protein-mediated H3K27me3 levels in differentiation at specific time-points has a variable effect on end-cell population, with an apparently negative effect early in differentiation, but with potentially a more positive effect later in the differentiation. Lastly, I was able to introduce a reversible knockdown system within the human pluripotent cells, which will allow us to study the functional role of Polycomb-proteins in the context of *in vitro* differentiation. In summary, my work has established a role of aberrant histone modifications in limiting *in vitro* differentiation capabilities and provides a new framework for manipulating epigenetic processes to improve differentiation outcomes that may be relevant for improving targeted cell production.

Charlene Fabian - January 2021

Acknowledgements

This PhD was supported through the Medical Research Council, and took place at the Babraham Institute, both of which I am wholly thankful to for giving me this opportunity. The contribution of Babraham facilities was invaluable, specific thanks to the Babraham Sequencing and Babraham Bioinformatic facilities.

Thank you also to those that I worked with in DefiniGEN, without them and their pancreatic differentiation expertise, this project would not have been possible. Thank you to Filipa Soares for her help, guidance, training, and all of the support she provided. Also thank you to Kalpesh Jhaveri, Carlos Gil, Katarzyna Tilgner and all of the other team members who gave me so much help in the lab.

I want to thank all members of the Rugg-Gunn lab that I have been luckily enough to work with throughout this time. Clara Novo and Amanda Collier, who taught me so much and had such patient as I started in this whole new world of PhD research. Thank you to Andrew Malcolm, Kate Maskalenka, Stephen Bevan for all your help and support, and a specific thank you to Adam Bendall, who was so helpful in both the lab experiments and during some of the more stressful times. A very special thank you to Claudia Semprich and Katarzyna Wojdyla, both of whom discussed scientific ideas, provided me with novel insights, and took their time to further this thesis, but most importantly became genuine close friends. Lastly, I would like to thank my supervisor, Dr Peter Rugg-Gunn, who was the perfect supervisor throughout my PhD. Teaching me much more than just how to be competent in the lab, he was always supportive, understanding, and willing to push me to perform the best I can.

And to those outside of the lab who supported me through it all, I cannot express my forever gratitude, Dominic Reeve, I could not have done it without you.

Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Meaning
AAVS1	AAVS1 locus (PPR1R12C gene on human chr. 19)
AID	Auxin inducible degron
bp	basepairs
BSA	Bovine serum albumin
Bsr	Blasticidin resistance gene
CAG	CMV enhancer, chicken β-actin, and rabbit β-globin hybrid promoter
cDNA	Complementary DNA
ChIP	Chromatin immunoprecipitation
COMPASS	Complex proteins associated with set1
CpG	Cytosine-phosphate-Guanine dinucleotide
CPI	Complete EDTA-free protease inhibitor
CRISPR	Clustered regularly interspaced short palindromic repeat
CRISPRi	CRISPR interference
CS	Carnegie stage
DAPI	4'6-Diamidino-2'-phenylindole dihydrochloride
DE	Definitive endoderm
DEG	Differentially expressed genes
DMSO	Dimethyl sulfoxide
DPBS-/-	Dulbecco’s PBS without calcium and magnesium
EDTA	Ethylenediaminetetraacetic acid
EED	Embryonic ectoderm development
EZH	Enhancer of Zeste Homolog 2
FGF	Fibroblast growth factor

GCDR	Gentle cell dissociation reagent
GFP	Green fluorescent protein
GO	Gene Ontology
gRNA	Guide RNA
H2AK119ub	Histone 2a lysine 119 ubiquitination
H3K4me	Histone 3 lysine 4 methylation
H3K27ac	Histone 3 lysine 27 acetylation
H3K27me	Histone 3 lysine 27 methylation
hPSC(s)	Human pluripotent stem cell(s)
HBSS	Hank’s balanced salt solution
Hygro	Hygromycin resistance gene
iPSC(s)	Induced pluripotent stem cell(s)
JMJC	jumonji domain containing
KRAB	Krüppel associated box
LoxP	Bacteriophage P1 Cre recombinase recognition site
mAID	Mini-auxin inducible degron
mESC(s)	Mouse embryonic stem cell(s)
MLL	Mixed-lineage leukemia
OsTIR1	Oryza sativa transport inhibitor response 1
PBS	Phosphate-buffered saline
PCA	Principal component analysis
PCR	Polymerase chain reaction
PC1	Principal component 1
PC2	Principal component 2
PRC1	Polycomb repressive complex 1
PRC2	Polycomb repressive complex 2

Puro	Puromycin resistance gene
ROCKi	p160-Rho-associated coiled-coil kinase inhibitor
RPKM	Reads per kilobase of transcript per million reads
RT-qPCR	Reverse transcription quantitative real time PCR
TAT-CRE	TAT-CRE Recombinase
TBS	Tris-buffered Saline
TBS-T	Tris-buffered Saline (TBS) with 0.1% Tween
TSS	Transcriptional start site
SCF	Skp1, Cullin and F-box complex
SETD1A/1B	SET domain containing 1A/1B
WT	Wild-type

Chapter 1

Introduction

1.1 Human Pluripotent Stem Cells

1.1.1 Human Embryonic Stem Cells

Stem cells were first discovered in the 1960's when transplantation of mouse bone marrow proved that some cells possess a self-renewing capacity (McCulloch and Till, 1960; Becker, McCulloch and Till, 1963). The idea that a cell contains the capacity to produce both further stem cells and specialised cell types opened up the field of stem cell research. The first mouse embryonic stem cells (mESC) were identified in 1981 (Evans and Kaufman, 1981; Martin, 1981). These cells were successfully isolated from a blastocyst stage embryo, and through mouse teratoma assays they were shown to be pluripotent, with the ability to form cell types of all three germ layers (Evans & Kaufman, 1981). The first isolation of human embryonic stem cells (hESC) occurred in the laboratory of James Thomson in 1998 (Thomson *et al.*, 1998). Thomson and his group were able to derive cells from human blastocysts which were capable of maintaining an undifferentiated culture and had the potential to form all three embryonic germ layers. These cells were considered to be pluripotent stem cells in all aspects that it was possible to test. Of the original five hESC lines formed at this time many continue to be maintained in culture and are still used for experiments today (Thomson et al., 1998). The discovery of these cells opened up many exciting prospects, including their application in studying human development, modelling disease and a role for them in regenerative medicine (Maury *et al.*, 2012; Avior, Sagi and Benvenisty, 2016).

Although offering great promise for many applications, hESCs have limitations in both ethical and practical aspects. The use of human embryos, which are required for hESC derivation, is still a prominent debate when using hESCs for research purposes. Indeed, in some countries, such as Italy, the derivation of hESCs is not permitted (Palazzani, 2011). In practical terms, these cells also have limitations due to the scarcity of embryos from which the cells are derived. Although a number of established hESC lines exist, isolating cells from donors

of a large genetic variety and particularly those with rare genetic disease is challenging, due to the scarcity of these cell sources. Similarly, utilising the cells for downstream applications such as regenerative medicine is at present, unlikely on a large scale as patients can undergo an immune response unless sufficiently immunologically matched (Bradley, Bolton and Pedersen, 2002; Liu *et al.*, 2017).

1.1.2 Induced Pluripotent Stem Cells

Many of the limitations related to hESCs seemed to be resolved after an astonishing discovery by Shinya Yamanaka and Kazutoshi Takahashi in 2006. Through a number of elegant experiments, these researchers were able to demonstrate the production of ESC-like cells from differentiated fibroblasts, first in mice, and followed closely by human fibroblasts (Takahashi and Yamanaka, 2006; Takahashi *et al.*, 2007; Yu *et al.*, 2007). This finding exemplified work performed years previously by John Gurdon, in which he demonstrated the ability of differentiated Xenopus laevis cells to reverse developmental potential when transplanted into an enucleated egg. This egg implanted with the donor nucleus was able to produce a viable animal and this discovery initiated years of work in the cloning field (Gurdon, 1962). The importance of these findings was demonstrated when both Gurdon and Yamanaka were awarded the Nobel prize for their work.

In their attempts to produce a pluripotent-like state from a mature cell type Yamanaka and colleagues selected, previously identified factors that were thought to be important for ESCs phenotype. Overexpression of these factors in dermal fibroblast cells was able to produce cells with pluripotent characteristics. Further experiments reduced & refined the number of factors necessary to reprogramme the cells to just four transgenes – *OCT4, SOX2, KLF4* and *C-MYC* (OSKM). With the addition of only these four transgenes, cells could be reprogrammed from fibroblasts into cells that resembled hESCs in terms of self-renewal capacity, differentiation ability, transcriptional expression and epigenetic profile. These cells were termed induced pluripotent stem cells (iPSCs) (Takahashi *et al.*, 2007). Almost in parallel another study demonstrated the successful production of human iPSCs, using a similar technique with small variations compared to that of the Yamanaka lab (Yu *et al.*, 2007). Since these initial studies, the production of iPSCs has been further optimised in many aspects and their ability to replace hESCs has been utilised for many applications.

1.1.3 Applications of Pluripotent Stem Cells

Human PSCs have generated much excitement within biomedical research due to the promise they hold in a number of applications. This includes the ability to study early human development, as the hPSCs are representative of early type cells and can be studied in terms of cell differentiation and also disease progression. The ability to create cells of any type also offers promise in using the cells to study human disease and potential treatments, and also offer promise in regenerative medicine in therapies which can utilise the cells directly to cure pathologies (Singh *et al.*, 2015; Avior, Sagi and Benvenisty, 2016; Doss and Sachinidis, 2019). Although these applications are possible, a number of limitations still exist which inhibit the full utilisation of hPSCs for these purposes. These limitations include problems such as viral vectors and mutagenesis that can occur in the production of iPSCs (Singh *et al.*, 2015). Also, the ability to produce any cell types from hPSCs is technically possible but this requires efficient differentiation protocols to produce the desired cell types, many of which have yet to be determined (Doss and Sachinidis, 2019). Therefore, the hPSCs possess huge potential for numerous applications but a number of these are still to be further developed before they can be fully utilised.

The use of hPSCs to study disease has become an intensive research area within recent years due to the ability to easily create human iPSCs directly from patients with a wide range of pathologies. Through the expansion and differentiation of human iPSCs, these cells can be used as an inexhaustible source to produce any cell type as required for studying a disease. This can therefore potentially serve to study both the development of disease and investigate the molecular mechanisms causing the disease and its phenotype. Producing large banks of cells derived from many donors may also assist in the production of new therapeutics by ensuring the treatments are suitable and effective in a large population of patients. Finally, the ability to use hPSCs themselves as therapies to correct disease phenotypes is being studied, in particular utilising the hPSCs to produce new tissue which can correct a patient's diseased cells or tissue. With the ability to create vast numbers of iPSCs from any human donor now possible, the limiting factor in a number of these downstream applications is producing the differentiated cell type of interest. Although robust and efficient culture protocols exist to produce a number of cell types, many current cell types cannot yet be derived *in vitro* (Burridge *et al.*, 2012; Tabar and Studer, 2014). One cell type that has a previously elucidated differentiation protocol and is offering great promise in this field is the formation of pancreatic cell types.

1.2 Pancreatic differentiation

The pancreas fulfils multiple purposes in maintaining the homeostasis of an organism. Firstly, the organ has a role in assisting digestion through an exocrine mechanism by secreting numerous digestive enzymes such as lipases, carbohydrases, and amylases. This exocrine region accounts for more than 90–95% of the pancreatic mass, and contains the acinar cells, responsible for digestive enzyme secretion, as well as ductal cells that transport these enzymes into the duodenum (Murtaugh and Melton, 2003; Guo and Hebrok, 2009). The organ also has an essential function in maintaining euglycemia within an organism, functioning through an endocrine department, which consists of small regions of cells referred to as the islets of Langerhans (Murtaugh and Melton, 2003; Guo and Hebrok, 2009).

These islet substructures are small globular clusters composed of heterogeneous cell populations, dispersed throughout the exocrine tissue. The structures are composed predominantly of five distinct endocrine cell types - the α, β, δ, ε, and γ cells. Each of these cell types will secrete their associated hormone in response to stimuli, to maintain the body homeostasis. The α-cells are responsible for glucagon secretion in response to decreasing levels of glucose in the blood, and the β-cells will secrete insulin in response to increased levels of blood glucose. Somatostatin is released from δ-cells, ghrelin from ε-cells and pancreatic polypeptide from the γ-cells. These islet cells together make up less than 2% of the total pancreas mass (Murtaugh and Melton, 2003; Collombat *et al.*, 2006). The individual cell-type composition of human islets is approximately 50-60% β-cells, 30-45% α-cells, less than 10% δ and γ-cells, and less than 1% ε-cells (Cabrera *et al.*, 2006), however, this composition of cell types varies greatly among individuals. Within the rodent pancreas, and to a lesser extent in the human, the islets will form a specific architecture with β-cells at the centre, surrounded by the other endocrine cell types (Murtaugh and Melton, 2003; Cabrera *et al.*, 2006).

Due to the functional importance of this organ, the cellular diversity it possesses, and the commonality of diseases related to the organ, the pancreas offers an attractive model system for both human organ development and regenerative therapy treatments. Each of the cell types found within the pancreas initiate from the same pool of progenitor cells but terminate in cell types with different adult functions. Studying the formation of mature pancreatic tissue is interesting for developmental biology as it can demonstrate highly specific developmental pathways, that occur in tandem but have opposing outcomes. Therefore, studying the

development of each of these cell types provides an excellent model for human cell differentiation and functional specialisation.

Moreover, the ability to produce functional pancreatic endocrine cells is one of the most promising fields of regenerative medicine. Type I diabetes, caused by the autoimmune-mediated destruction of insulin-producing β-cells, results in individuals lacking the ability to produce insulin, leading to potentially lethal hyperglycaemia with no medical intervention. This disease has a high global burden with millions of sufferers world-wide, and a prevalence estimated to be approximately 15 per 100,000 people in the European population (Mobasseri *et al.*, 2020). The current therapy to treat Type I diabetics dates back to the 1920s and involves the introduction of exogenous insulin to control the patient's blood-glucose levels (Banting *et al.*, 1922). With improper control of blood glucose levels, the patient can suffer both life-altering and life-threatening consequences (American Diabetes Association, 2010).

Importantly, a cure for type 1 diabetes is available using whole-pancreas transplants, a technique this was first proposed as an effective treatment in the 1970s, with successful treatment in rats (Ballinger and Lacy, 1972). This treatment was first applied in a human patient in 1989 and has continued to be used in patients suffering serious complications associated with the disease (Scharp *et al.*, 1990; Shapiro *et al.*, 2000). Since then, it has also been shown that islet transplantation is sufficient to produce insulin independence in diabetic patients (Shapiro *et al.*, 2000). The majority of patients did not retain this insulin independence in the long-term, due to decreased insulin secretion form the transplanted cells, but after transplantation the blood-glucose levels were easier to regulate with exogenous insulin injections (Ryan, Bigam and Shapiro, 2006). This treatment offered promise for a diabetic cure based on islet transplant upon further refinement of the approach. However, one of the most prominent issues in expanding this as a treatment option arises due to the lack of available pancreatic organs or islets for transplantation (Sneddon *et al.*, 2018). Therefore, alternative sources of islet cells are needed to allow realistic progression of this treatment option. It is for this reason that the ability to generate mature pancreatic endocrine cells *in vitro* has become a promising focus within the regenerative medicine field (Bonner-Weir and Weir, 2005; Sneddon *et al.*, 2018).

Although it is well established that pancreatic β-cells are responsible for the secretion of insulin, the mechanism for which cells undergo hormone secretion is still contested. It is known that glucose will stimulate the β-cells to secrete insulin in a process termed glucose-stimulated

insulin secretion (GSIS) (Dean and Matthews, 1968), which in the body acts to maintain the homeostasis, but due the complexity of the environment within the body, it has been difficult to definitely determine how this result in hormone secretion although a leading model is general accepted (Figure 1.1). In this model glucose enters the β-cells through a glucose transporter, and is metabolised in the cytoplasm, before being transported into the mitochondria. This reaction leads to the generation of ATP, stimulating the ATP-sensitive K^+ channels (Ashcroft, Harrison and Ashcroft, 1984; Cook and Hales, 1984). In β-cells the K^+ channels are the primary determinant of the membrane potential, therefore closure of these results in membrane depolarization. The membrane depolarization opens voltage-dependent Ca^{2+} channels, causing Ca^{2+} influx and elevation of cytosolic free Ca^{2+} concentration resulting in a rapid increases in the rate of C-peptide exocytosis, the insulin precursor molecule (Inagaki, Gonoi and Seino, 1997). Although GSIS is modulated by a number of further factors, such as non-glucose nutrients, hormones and neural inputs thereby adding complexity to this model that is not yet fully understood.

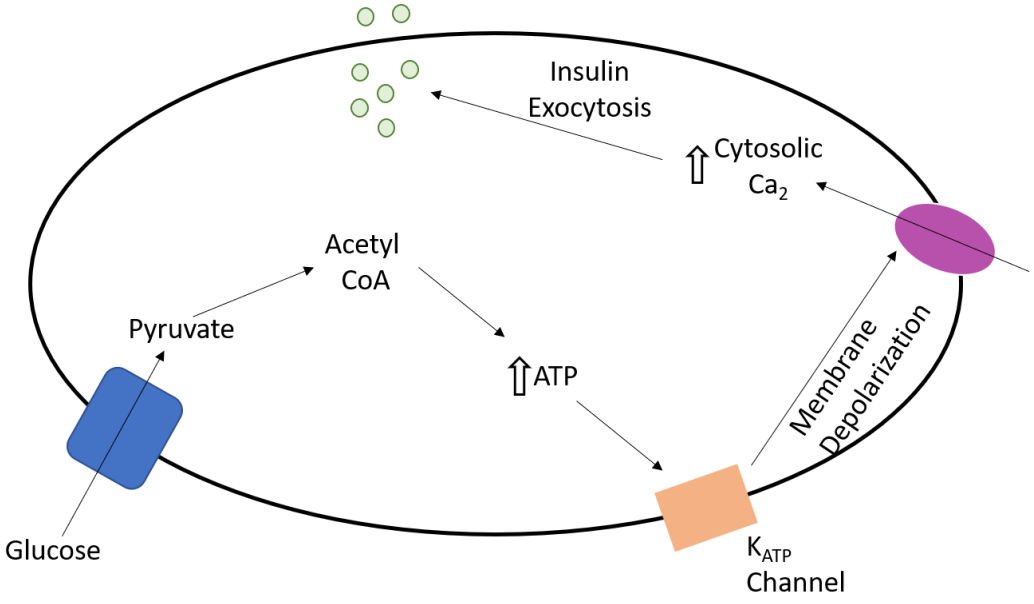


Figure 1.1 Model of Cellular Insulin Secretion. The above mechanism is the accepted model for which a cell will respond to increased glucose levels with insulin release.

1.2.1 Pancreatic Embryogenesis

The initial production and optimisation of pancreatic differentiation protocols from hPSCs has utilised knowledge obtained from mimicking pancreatic development. Therefore, an in-depth understanding of the developmental processes and detailed knowledge of the cells derived during this time is beneficial in helping to produce efficient pancreatic-like cells *in vitro*. However, studies using human embryos to study organ development are highly restricted by both ethical implications and tissue availability. Due to this, much of early human development, including pancreatic organ formation relies heavily on data obtained in other animal species, and in this circumstance much of the knowledge formed is based on mouse and rat development (Petersen *et al.*, 2018).

In both mouse and human, the pancreas will form at the foregut-midgut boundary, with the initial pancreas formation beginning approximately 25-27 days post-fertilisation (dpf) in humans, and at mouse embryonic day of development (E) 7.5 (Jennings *et al.*, 2013). Pancreatic specification into the dorsal pancreatic bud then occurs. In the mouse, the formation of early pancreatic cells is initiated by exclusion of Sonic Hedgehog (SHH) from the cells, thereby allowing the expression of the Pancreatic and Duodenal Homeobox 1 (PDX1) transcription factor, which is critical for pancreatic development (Hebrok, Kim and Melton, 1998). In human development, it is hypothesised that a similar process occurs with SHH detected at approximately 25-27 dpf, and PDX1 present by 29-31 dpf (Jennings *et al.*, 2013). By day 30-33 dpf in humans, both dorsal and ventral pancreatic buds exist, with expression of the growth factors SOX9, PDX1 and GATA4 (Piper *et al.*, 2004; Jennings *et al.*, 2013). At this developmental timepoint there is also microlumen formation, which are structures that later give rise to the luminal network of the exocrine compartment of the pancreas (Piper *et al.*, 2004). After this point, the developing pancreas will then undergo a large expansion of proliferative progenitors, with no clear cell specification until approximately 45-47 dpf. At this point, central cells form ductal-like structures, with lower levels of GATA4, and more peripheral cells will cluster and are SOX9 / GATA4 / NKX6.1 positive (Jennings *et al.*, 2013).

The endocrine specification for human pancreas is predicted to occur at approximately 49 - 52 dpf, at which time NEUROGENIN3 (NGN3) expression increases rapidly (Lyttle *et al.*, 2008). At approximately this time of increased NGN3 expression, the first insulin positive cells are detectable. These early foetal β-cells are the first of the islet cell-types to appear in human development (Piper *et al.*, 2004; Lyttle *et al.*, 2008). By 10 weeks post conception (wpc) the

insulin positive cells have undergone vascularisation, with glucagon, somatostatin and pancreatic polypeptide positive cells present in the foetal islets by 12-13 wpc (Piper *et al.*, 2004; Jennings *et al.*, 2013). Within the highly NGN3-expressing endocrine cells, the expression of SOX9 is lost but pancreatic ductal cells will maintain the expression of SOX9. After 35 wpc *NGN3* expression is no longer detected in the human foetus (Lyttle *et al.*, 2008). Due to the necessity of the *NGN3* transcription factor in endocrine differentiation, this indicates that endocrine cell specification had occurred and the resulting formation of final cell populations in the pancreas were due to proliferation and apoptosis of the existing populations (Jennings *et al.*, 2013).

1.2.2 Key transcription factors in pancreatic development

From both human and mouse studies a number of transcription factors (TFs) have been identified that play key roles in the formation of the pancreas (reviewed in Chakrabarti and Mirmira, 2003; Conrad, Stein and Hunter, 2014). Determining these TFs is important for studying developmental programs and also for determining the success of *in vitro* differentiation protocols. The identification of key TFs will provide a framework for expression patterns in the differentiating populations and allows these factors to be used as robust markers of the cell types present.

One of the key TFs in human and mouse pancreatic formation is the PDX1 homeobox factor. In mice, *Pdx1* transcripts label the dorsal and ventral epithelial buds of the posterior foregut around E9.5, which gives rise to the pancreatic organ (Pan and Wright, 2011). In Pdx1-null mutant mice, these buds initially form but quickly regress, resulting in complete pancreatic agenesis, severe hyperglycaemia, and death within a few days of birth (Jonsson *et al.*, 1994; Offield *et al.*, 1996). Similarly, *PDX1* is present early in the developing human pancreas, observed in the human dorsal and ventral foregut from around 4 wpc (Piper *et al.*, 2004; Jennings *et al.*, 2013). In human patients who possess a homozygous or compound heterozygous mutations in PDX1, an agenesis phenotype is observed, similar to that seen in Pdx1-deficient mice (Stoffers *et al.*, 1997; Schwitzgebel *et al.*, 2003). As the expression of *PDX1* is detected early in pancreatic development, it is suspected to be important for the development of all pancreatic cell types. However, studies demonstrate that some expression of this factor persists in both mouse and human differentiated β-cells, with little or no expression in other mature pancreatic cells (Gao *et al.*, 2014). The high levels of PDX1 and insulin (INS) co-expression suggest the PDX1 is also critical in specifying and maintaining β-cell fate (Lyttle *et*

al., 2008; Jennings *et al.*, 2013), with point mutations in the PDX1 gene demonstrating negative effects on the development and function of human β-cells (Wang *et al.*, 2019).

NKX6.1 is another factor important in the developing pancreas, also first observed in the developing pancreas approximately 4 wpc (Lyttle *et al.*, 2008; Jennings *et al.*, 2013). This TF is initially observed in the multipotent progenitor cell types and is hypothesised to be important initially for specifying the endocrine cell fate in the developing pancreas (Schaffer *et al.*, 2010; Jennings *et al.*, 2013). By weeks 14 - 16 of development the expression of NKX6.1 is restricted to the β-cells (Brissova *et al.*, 2005; Jennings *et al.*, 2013). Nkx6.1 expression is similar in developing rodent pancreatic cells, before gradually becoming restricted to β-cells. Nkx6.1 expression in the developing pancreatic rodent cells has been shown to be essential for the formation of β-cells (Sosa-Pineda *et al.*, 1997; Henseleit *et al.*, 2005). Demonstrating this, inactivation of Nkx6.1 specifically in endocrine precursors or in β-cell caused the conversion of β-cells to alternative endocrine lineages, and Nkx6.1-null mice had severely reduced number of β-cells (Henseleit *et al.*, 2005; Schaffer *et al.*, 2013).

Similarly, a TF that is important for endocrine progenitor formation in both human and mouse development is NGN3 (Neourogenin3, also known as Neurog3). This factor is important in allowing differentiation of endocrine progenitors into hormone-expressing islet cells with each NGN3-positive cell producing only one islet cell type (Gradwohl *et al.*, 2000). In the development of the human pancreas, NGN3 is first observed at approximately 8 weeks, with increased expression until around 11 weeks at which point maximal expression is observed, with a gradual loss after endocrine lineage has been specified (Lyttle *et al.*, 2008; Jennings *et al.*, 2013). The *NGN3* expression precedes a large number of endocrine TFs, including *ISL1, MAFB, NKX2.2,* and *PAX6*, and also a set of downstream factors such as *Arx* and *NeuroD1* which rely on NGN3 expression (Jeon *et al.*, 2009). Perturbation of Ngn3 in developing mice will result in a diabetic phenotype with mortality observed shortly after birth due to a complete absence of endocrine cells (Johansson *et al.*, 2007). Similarly, in humans, a rare NGN3 mutation resulted in no detectable islets, although low levels of C-peptide were present, however this still resulted in permanent neonatal diabetes (Rubio-Cabezas *et al.*, 2011).

A number of additional factors are also both important and restricted in their expression patterns within endocrine and mature β-cells. For example, MAF BZIP Transcription Factor A (MafA) was a factor uncovered in mice to be specifically expressed in the β-cells of the endocrine compartment (Matsuoka *et al.*, 2003; Wang *et al.*, 2007). This factor was crucial for

regulation of insulin and Glut2 in β-cells, both of which are necessary to allow glucose-responsive function of the cells (Hang and Stein, 2011). MAFA is also expressed specifically in late human β-cells, and therefore assumed to have similar functions (Dai *et al.*, 2012). NeuroD, a basic helix-loop-helix TF, is also essential in the formation of pancreatic β-cells to achieve and maintain functional maturity (Gu *et al.*, 2010). In the developing human pancreas NEUROD1 is expressed in all endocrine cell types of adult islets, including mature β-cells and is first observed at approximately week 15 of the development (Lyttle *et al.*, 2008; Jennings *et al.*, 2013). NEUROD1 has the ability to bind the *INS* promoter, and humans with mutations in the *NEUROD1* gene are predisposed to developing maturity onset diabetes of the young (Malecki *et al.*, 1999). In mouse, NeuroD-null mutants have poorly differentiated α and β-cells with few islets formed, and almost no β-cells are present resulting in the death of the mice from severe diabetes shortly after birth (Naya *et al.*, 1997). Lastly, the Insulin gene enhancer protein (ISL-1) is another important factor for β-cell development, first expressed in the foetal pancreas at approximately 8 - 10 weeks (Lyttle *et al.*, 2008; Jeon *et al.*, 2009). Mice with endocrine Isl1 mutations become diabetic, showing impaired islet cell maturation and reduced postnatal islet mass expansion, with MafA determined as a direct downstream target of this TF (Du *et al.*, 2009). The identification of a patient with a ISL1 mutation determines a likely association with type 2 diabetes, thereby demonstrating a role of the ISL1 in functional human β-cells (Shimomura *et al.*, 2000).

1.2.3 *In vitro* pancreatic differentiation

The *in vitro* pancreatic differentiation field has developed and progressed mainly by forming protocols mimicking pancreatic development based on animal models and the limited knowledge of human development. Although key papers such as those by Piper et al, 2004 and Jennings et al, 2013, demonstrate that the development of cells through the stages from distal foregut endoderm, specific to pancreatic endoderm which then forms into a pancreatic bud containing multipotent progenitors, which can give rise to the final pancreatic cells as seen in other species (Piper *et al.*, 2004; Lyttle *et al.*, 2008; Jennings *et al.*, 2013, 2017). Similarly, papers studying the human pancreatic development have determine that many of TF found to be of importance in other species pancreatic development will play a role of similar function and importance in the human development, including PDX1, NGN3, NKX6.1 and MAFA (Piper *et al.*, 2004; Lyttle *et al.*, 2008; Dai *et al.*, 2012; Jennings *et al.*, 2013). However, due to the difficulties in obtaining developing human tissue to study pancreatic organogenesis, *in vitro* differentiation has been used as a model of the development.

The first stage of hPSC to pancreatic differentiation is to produce definitive endoderm (DE) (Figure 1.2) capable of giving rise to numerous cell types including intestine, lung, liver, and pancreas. The production of these cell types from hPSCs can occur at high levels of efficiency within approximately 3 days of protocol initiation. Development of DE *in vitro* can occur solely with the introduction of extracellular factors, such as Activin/Nodal, thereby producing cells which have the expression of mature endodermal markers such as SOX17, CXCR4 and FOXA2 (D'Amour *et al.*, 2005; Vallier *et al.*, 2009; Teo *et al.*, 2011). It is also possible, under certain culture conditions, to maintain the DE population in a self-renewing state, allowing expansion of this progenitor cell population (Cheng *et al.*, 2012).

After establishment of the DE population, cells can be further differentiated to replicate cells of a primitive-foregut type, after approximately 2 days in culture (Figure 1.2) This is driven by the removal of Activin and the addition of factors to stimulate FGF signalling such as FGF7/10 (D'Amour *et al.*, 2005; Cheng *et al.*, 2012). The addition of these factors alone is sufficient to form cells which are PDX1 expressing downstream, however the addition of retinoic acid is necessary to produce hormone-expressing endocrine like-cells later in the differentiation (D'Amour *et al.*, 2005; Cho *et al.*, 2012). The protocol is further optimised with the addition of BMP inhibitors, as activation of this pathway can drive cells towards a hepatocyte lineage (Gouon-Evans *et al.*, 2006). Differentiating cells may also then be exposed to SHH inhibitors, based on the observations that SHH is absent in the developing pancreatic endoderm of chick embryos and restriction of *Shh* expression in the developing endoderm was able to induce pancreatic cell specification (Hebrok, Kim and Melton, 1998). Maintenance of cells in similar culture conditions will result in the formation of posterior foregut-like cells after approximately 2 days of further differentiation, with pancreatic endoderm formed between 9 and 13 days post-induction of differentiation, depending on the specifics of factor concentration and timings within the protocol (D'Amour *et al.*, 2005; Cho *et al.*, 2012; Bruin *et al.*, 2014; Russ *et al.*, 2015).

The main focus of this field has been to produce mature endocrine cells, found within the islets of the pancreas. Therefore, the pancreatic endoderm cells formed will need to be further directed towards the endocrine lineage by first producing an endocrine progenitor state. The production of these progenitors focuses on promoting the expression of NGN3, due to its importance in the development of endocrine cells (Gradwohl *et al.*, 2000; Johansson *et al.*, 2007). It is not clear how to directly promote NGN3 expression during *in vitro* differentiation (Petersen *et al.*, 2018), but typically most protocols will use a combination of Notch inhibition and/or ALK5 inhibition to promote the production of the progenitor cells (Cho *et al.*, 2012;

Pagliuca *et al.*, 2014; Rezania *et al.*, 2014). These pancreatic endocrine progenitors are successful in producing mature, insulin-secreting pancreatic endocrine cells after transplantation into a mouse, and excitingly, these cells possess the ability to rescue a diabetes phenotype in the transplanted mice (Kroon *et al.*, 2008; Rezania *et al.*, 2012; Bruin *et al.*, 2013). It has also been shown that through continued *in vitro* differentiation, typically maintaining the cells in similar media composition to that required to form endocrine progenitors, the cells can then produce hormone-secreting pancreatic endocrine cells (D'Amour *et al.*, 2005; Cho *et al.*, 2012). Recent 3D suspension culture systems particularly show promise in promoting the production of these mature cells, containing responsive β-cells (Pagliuca *et al.*, 2014; Rezania *et al.*, 2014; Russ *et al.*, 2015; Nair *et al.*, 2019; Velazco-Cruz *et al.*, 2019). However, as discussed in the next section, these *in vitro* derived pancreatic endocrine cells have persistent drawbacks, which limit their ability to be used in place of the *in vivo* formed islets.

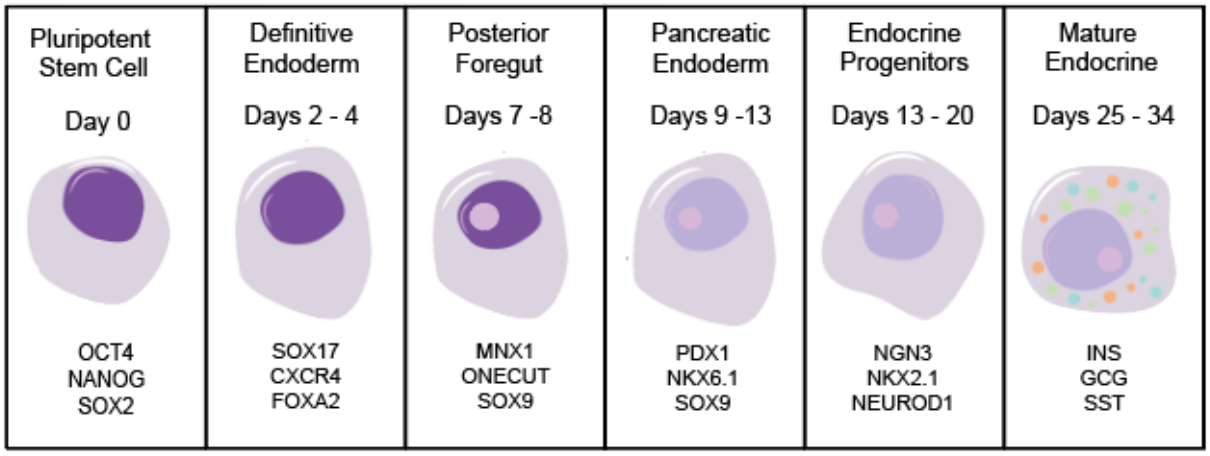


Figure 1.2 Human *in vitro* pancreatic differentiation. The key stages that a cell is estimated to form during *in vitro* hPSC to pancreatic differentiation, with the days at which different protocols report their presence, and key factors associated with the cell types.

1.2.4 Limitations in current pancreatic differentiation protocols

The ability to successfully produce differentiated cells *in vitro* that are similar to pancreatic endocrine cells has offered much excitement in terms of regenerative medicine abilities. These cells express the factors associated with mature pancreatic endocrine cells, with the three major hormones expressed from the three main cell types - α, β and γ cells – that are

present in adult pancreatic islets (D'Amour *et al.*, 2005; Pagliuca *et al.*, 2014; Russ *et al.*, 2015; Nair *et al.*, 2019). The ability to produce hormone secreting cells is an exciting discovery, and with more research the cells are continuously being improved in terms of functionality. However, early studies that could produce insulin expressing cells *in vitro*, observed a number of deficiencies in the cells of the obtained population, including lack of glucose responsiveness, little insulin production, and frequent co-expression of pancreatic hormones (D'Amour *et al.*, 2005; Basford *et al.*, 2012). Since this discovery, the continual optimisation of differentiation protocols has allowed further improvement to the *in vitro* derived cells, however a number of limitations still persist in the final-stage cell types.

One of the main problems that has continually been observed in the *in vitro* derived endocrine populations is the persistence of polyhormonal cell types. Commonly, the presence of C-peptide in a cell will also overlap with the expression of other pancreatic hormones, typically glucagon, and somatostatin (D'Amour *et al.*, 2005; Rezania *et al.*, 2014; Hrvatin *et al.*, 2014). However, in the adult pancreas, each endocrine cell type is responsible for secretion of a single hormone (Polak *et al.*, 2000; Riedel *et al.*, 2012). It is therefore still promising that hormone expressing cells can be obtained, but the presence of polyhormonal is likely to indicate the existence of a more immature, or only partially functional endocrine cells population. Although interestingly, these cells do show the capability to develop into monohormonal cells after being grafted into a mouse, showing that these cells do have further developmental potential in certain, currently undefined, conditions (Kroon *et al.*, 2008; Rezania *et al.*, 2012). Thereby, these cells when produced *in vitro* do not fully resemble cells from an adult islet but are partially representative and with further analysis of the developing cells, protocols may be further optimised to further progress these cells.

Further evidence for the inability to form fully mature and functional endocrine cells *in vitro* is suggested by the reduced levels of insulin expression from the end-cell type, particularly in response to glucose stimulation (Basford *et al.*, 2012; Hrvatin *et al.*, 2014). Under resting conditions, adult islets will secret low levels of C-peptide, the insulin precursor. This is also observed from the *in vitro* derived pancreatic cells, but the levels of insulin transcription and C-peptide secretion are lower in this cell population compared to adult islets (Basford *et al.*, 2012; Hrvatin *et al.*, 2014). It is unclear if this is due to intrinsic differences within the insulin production and/or secretion from the cells. It could also be caused by the presence of fewer β-cells in the *in vitro* derived population, or by a difference in the distribution of cell types, or by the failure to differentiate all of the cells in culture to a mature cell type. With further improvement and

optimisation on those existing differentiation protocols, it has been possible to produce cells with some glucose responsiveness in terms of increased insulin secretion, however a direct comparison between the functionality of these cells and the adult islets is yet to be shown (Pagliuca *et al.*, 2014; Rezania *et al.*, 2014; Russ *et al.*, 2015).

Another problem observed within the cells produced *in vitro* is the production of cells with a suspected transcriptional signature of foetal pancreatic cells, compared with that of the adult islet cells. In a study by Hrvatin et al, 2014, microarray gene expression analysis was undertaken on β-cells isolated from both foetal and adult islets, as well as *in vitro* differentiated cells isolated based on insulin expression. Correlation clustering of the cells demonstrated that the *in vitro* derived β-cells from iPSC and ESC lines were transcriptionally more similar to foetal than adult β-cells. The similarity was so high between the foetal and *in vitro* cells that the correlation between these samples was not significantly different from the correlation between the biological replicates of the adult β-cell samples (Hrvatin *et al.*, 2014). The foetal transcriptional programme may also underlie other problems within the final-stage *in vitro* cells, for example, human foetal islets and β-cells will undergo an insulin secretion response to glucose stimulation, but at a reduced level compared to adult cells (Otonkoski *et al.*, 1988; Rorsman *et al.*, 1989). There is also evidence from multiple studies that has identified foetal islets with co-expression of other pancreatic hormones in insulin expressing cells, including glucagon and somatostatin (de Krijger *et al.*, 1992; Polak *et al.*, 2000; Riedel *et al.*, 2012). Therefore, the persistence of a foetal transcriptome within the *in vitro* derived cells, in place of a mature adult transcriptome, could be hindering the ability of the derived cells to properly function.

Although these problems persist throughout differentiation protocols, it is worth noting how much protocols have progressed, since the initial ability to form insulin expressing cells was achieved *in vitro*. In recent years, studies have produced cells capable of glucose-responsive insulin secretion and have the ability to rapidly reverse diabetes within mice *in vivo* (Rezania *et al.*, 2014; Russ *et al.*, 2015). Importantly, these culture systems use chemically defined media that are devoid of animal products, thereby allowing the ability to utilise the cells downstream in human therapies. The progress in this field has been so great that clinical trials are ongoing, testing the safety and efficacy of hESC-derived pancreatic progenitor cells in patients with type I diabetes (https://clinicaltrials.gov, identifiers: NCT03163511). Therefore, although limitations exist within the *in vitro* pancreatic differentiation, this is an exciting area of regenerative medicine, which is consistently progressing knowledge in the field.

1.3. Histone Methylation Modifications

Within a cell, the extracellular signalling will converge with intracellular factors which in turn control the transcriptional programs within a cell. One of the major intrinsic mechanisms of controlling transcription of an individual cell is through the structure of the chromatin thereby controlling the accessibility of the genome. Modifications which can affect the chromatin organisation and therefore control the genes which are able to be transcribed are termed epigenetic mechanisms. These mechanism are known to be important in a multitude of cellular functions including: acquiring cell identity, stabilising the cells transcriptional network in steady-state conditions, and through-out cell proliferation (reviewed Margueron and Reinberg, 2010; Chen and Dent, 2013; Lawrence, Daujat and Schneider, 2016).

Within the cell nucleus DNA is found to exist wrapped around histone proteins, forming highly repetitive nucleosome subunits, which are the basic building blocks of chromatin. Each nucleosome is comprised of approximately 146 base pairs of DNA wrapped around a protein octamer that contains two copies of each histone protein: H2A, H2B, H3, and H4 (Luger *et al.*, 1997). The histone protein H1 is also present and will interact with both the nucleosome and linker DNA to stabilise the organisation of chromatin (Harshman *et al*., 2013). Both the DNA and histone proteins can be transiently modified in ways which will affect the association between them, thereby controlling the accessibility of the DNA as a method of controlling gene expression. These editions can occur either on the DNA directly, most typically through DNA methylation, or the tails of the histone proteins are subject to numerous covalent modifications.

The possible modifications that occur in the histone tails include methylation, acetylation, phosphorylation, and ubiquitination, among others. These modifications can occur at numerous sites, and in varying combination within the histone tails (Bannister and Kouzarides, 2011; Zhang, Cooper and Brockdorff, 2015). The presence of different modifications are thought to control accessibility to DNA through modulation of chromatin structure and by recruitment of effector proteins (Margueron and Reinberg, 2010; Voigt and Reinberg, 2011). Although numerous studies have been undertaken to determine the consequence of the different modifications, the direct consequence of each modification and the mechanisms of how they bring about effects on gene expression remain unclear.

1.3.1 Histone 3 Lysine 27 methylation

Histone 3 lysine 27 trimethylation (H3K27me3) is one of the most extensively studied histone modifications to date. This modification is associated with maintaining transcriptional repression and has been observed throughout evolution where it is present as a transcriptional repressor in various organisms (Croce and Helin, 2013). This histone modification is catalysed by Polycomb repressive complex 2 (PRC2), through the Enhancer of Zeste Homolog 1 and 2 (EZH1 and EZH2) enzymes (see section 1.4.1) (Cao *et al.*, 2002; Czermin *et al.*, 2002; Kuzmichev *et al.*, 2002; Müller *et al.*, 2002). To introduce this post-translational modification, EZH1/2 use S-adenosyl-methionine (SAM), the substrate for all known histone methyltransferases, as a methyl donor to create mono-, di- or tri- methylation at the lysine 27 residue of the H3 histone protein, (Cao *et al.*, 2002; Kuzmichev *et al.*, 2002; Müller *et al.*, 2002; Fan *et al.*, 2015). The existence of H3K27me1 and H3K27me2 are prominent within the cells and are proposed to play separate functions from the H3K27me3. Interestingly, H3K27me1 was observed to be enriched at actively transcribed regions (Vakoc *et al.*, 2006; Cui *et al.*, 2008; Ferrari *et al.*, 2014; Lee *et al.*, 2015). H3K27me2 is largely dispersed throughout the genome, and present in large chromatin domains. This modifications is typically found in the intergenic, non-transcribed regions, during which it is estimated to act in a protective manner from other histone modifications (Ferrari *et al.*, 2014; Lee *et al.*, 2015; Lavarone, Barbieri and Pasini, 2019). Although these histone modifications play a role in the cell it is the trimethylation that is the most extensively studied.

The presence of H3K27me3 is typically considered a repressive modification due to its association and enrichment within facultative heterochromatin regions of the genome (Cao *et al.*, 2002; Kuzmichev *et al.*, 2002; Müller *et al.*, 2002). This modification is highly associated with the promoter region of a gene, which is the focus of this thesis, although it can also be present at enhancer regions (Rada-Iglesias *et al.*, 2011; Zentner *et al*., 2011), repetitive regions (Leeb *et al.*, 2010), and it will coat the inactive X-chromosome (Plath *et al.*, 2003; Silva *et al.*, 2003). Numerous studies utilising next-generation sequencing data demonstrate that H3K27me3 modified promoters will strongly correlate with repressed or lowly expressed genes within the cell (Boyer *et al.*, 2006; Lee *et al.*, 2006; Barski *et al.*, 2007). The H3K27me3 is not a permanent mark once deposited, as it can be passively removed or through the activation of demethylases. In the case of H3K27 di and tri-methylation the jumonji domain containing 3 (JMJC3) or the JMJC protein ubiquitously transcribed tetratricopeptide repeat, X chromosome (UTX) are responsible for this function (Agger *et al.*, 2007; De Santa *et al.*, 2007; Lan *et al.*, 2007; Min *et*

al., 2007). Similarly, H3K27me3 can be lost passively, with gradual dilution at the modified region as cells divide if the modification is not actively maintained by PRC2 (Hong *et al.*, 2007).

One of the most interesting points within the epigenetic field is ongoing discussion behind the mechanism of action of H3K27me3 in gene silencing, the modification is a hallmark of facultative chromatin, but a direct role of H3K27me3 in inducing gene silencing is not clear. The classic hierarchal model focused on the interplay between Polycomb Repressive Complex 1 (PRC1) and PRC2, in which PRC2 is recruited by the CpG islands to CpG-rich gene promoters thereby depositing H3K27me3 (Ku *et al.*, 2008; Lynch *et al.*, 2012; Jermann *et al.*, 2014).The presence of H3K27me3 at these CpG promoters will then act as a recognition site for the chromobox (CBX) within PRC1 (Blackledge, Rose and Klose, 2015; Wiles and Selker, 2017). After recruitment of PRC1, this complex will catalyse histone 2A lysine 119 monoubiquitylation (H2AK119ub) at these regions resulting in chromatin compaction and silencing of the underlying gene (Min, Zhang and Xu, 2003; Wang *et al.*, 2004; Eskeland *et al.*, 2010). The H3K27me3 will also act as a recruiter and an activator for the PRC2 complex itself, through an aromatic cage region in the one the complex proteins thereby inducing further H£K27me3 deposition and spreading, further generating the facultative chromatin regions associated with the Polycomb repressed genes (Hansen *et al.*, 2008; Margueron *et al.*, 2009; Oksuz *et al.*, 2018). I support of this model mammalian PRC1 and PRC2 have highly similar and overlapping binding patterns (Boyer *et al.*, 2006; Bracken *et al.*, 2006) , and in *Drosophila* loss of the PRC2 enzyme component, E(z) the *EZH1/2* ortholog abolished PRC1-binding (Rastelli, Chan and Pirrotta, 1993).

However, further studies have uncovered results that will challenge this model of PRC2/H3K27me3 recruitment of PRC1/H2AK119ub and subsequent gene silencing. This includes the observation that H3K27me3 will be associated with gene repression in organisms that lack PRC1 equivalents (Jamieson *et al.*, 2013), suggesting that H3K27me3, may have some, as of yet uncovered silencing mechanism in these organisms, which may also occur in cells with PRC1 present. To further support a functional role for H3K27me3 outside of PRC1, it is known that the PRC1 complex does not localize to all H3K27me3 marked repressed targets, and that the PRC1 enzymatic activity is not required for all repression (Ku *et al.*, 2008; Pengelly *et al.*, 2015). Similarly, it is now accepted that PRC1/H2AK11ub will in many instances recruit PRC2 to target loci as opposed to the original proposed hierarchy of PRC2 followed by PRC1 (Blackledge *et al.*, 2014; Cooper *et al.*, 2014). Although the silencing method of this modification remains poorly understood, new studies are continuously uncovering potential mechanisms

behind H3K27me3 and its role in silencing, but it is unlikely this will involve only a single mode of action and a single outcome, instead it is expected to be a complex system dependent on numerous other factors beyond the presence of H3K27me3 at a promoter.

1.3.2 Histone 3 Lysine 4 methylation

Histone 3 lysine 4 trimethylation (H3K4me3) is also a well-studied and highly prevalent histone modification found in numerous cells. This modification is deposited through the action of a large group of heterogenous proteins termed Trithorax proteins, which exist in complexes with numerous subunits. These complexes will function through the action of one of several methyltransferase enzymes; the SET domain containing 1A or 1B (SETD1A/1B) or mixed lineage leukemia 1 to 4 (MLL1-4) (Milne *et al.*, 2002; Santos-Rosa *et al.*, 2002; Lee and Skalnik, 2005; Lee *et al.*, 2007; P. Wang *et al.*, 2009). These enzymes form complexes termed COMPASS (complex proteins associated with set1), with all of variations of COMPASS complexes contain a number of core proteins – WDR5, ASH2, RBBP5, and DPY30 – which are essential for the function of the enzyme complex (Schneider *et al.*, 2005; Dehé *et al.*, 2006; Dou *et al.*, 2006). Also, found in the COMPASS complexes are a number of other accessory proteins, the choice of which is dependent on the enzyme present in the complex (Miller *et al.*, 2001; Krogan *et al.*, 2002; Schneider *et al.*, 2005; Vermeulen *et al.*, 2010). These variations of the complex confer specialisation in terms of the genomic regions targeted for methylation, with the SET containing complexes associated with more global H3K4me3 methylation, and the MLL containing complex associated with the methylation of more specific sites, such as gene promoters in mammals (Wu *et al.*, 2008; Denissov *et al.*, 2014; Sze *et al.*, 2020). Within these different forms more specific role can also be ascribed, for example the MLL2 methyltransferase has specifically been identified as that responsible for depositing H3K4me3 within the bivalently marked regions, although these regions are not the sole target of the enzyme complex (Hu *et al.*, 2013; Denissov *et al.*, 2014). The H3K4 methylation modification can be actively removed, though lysine specific demethylase and JMJC demethylase protein families (Lee *et al.*, 2005; Christensen *et al.*, 2007; Iwase *et al.*, 2007).

These complexes are responsible for depositing all three possible histone 3 lysine 4 methylation marks - H3K4me1, and H3K4me2 and H3K4me3. All three of which are found to be enriched at the gene transcription start site (TSS), however the H3K4me3 is more strongly associated with highly transcribed genes than either the mono or di methylated (Barski *et al.*, 2007; Soares *et al.*, 2017). Similarly, trimethylation is more closely aligned to the TSS, with the

modification transitioning from tri to di and then mono-methylation with increasing distance from the TSS (Barski *et al.*, 2007; Heintzman *et al.*, 2007; Soares *et al.*, 2017). H3K4me3 is recognised by specific protein domains within a variety of complexes including the PHD, chromo, tudor and WD40 domains (Yun *et al.*, 2011). These recognition domains are contained within a number of important proteins including the transcription factor II D (TFIID), SAGA and nucleosome remodelling factor/bromodomain PHD finger transcription factor (NuRF/BPTF) complexes (Wysocka *et al.*, 2006; Vermeulen *et al.*, 2007; van Ingen *et al.*, 2008). These complexes are all involved in transcriptional activation and, therefore H3K4me3 is commonly thought of as a modification indicative of transcriptional activation (Wysocka *et al.*, 2006; Vermeulen *et al.*, 2007; van Ingen *et al.*, 2008). The H3K4me3 has also been shown to actively induce transcription of some specific gene subsets including p53-dependant DNA-damage inducible genes through TAF3 (Lauberth *et al.*, 2013) and ectopic targeted genes (Clouaire *et al.*, 2012; Cano-Rodriguez *et al.*, 2016) Furthermore, the presence of H3K4me3 is strongly associated with transcription of genes (Bernstein *et al.*, 2002; Santos-Rosa *et al.*, 2002; Barski *et al.*, 2007; Guenther *et al.*, 2007). Although the association of H3K4me3 with active genes exists, numerous loss of function studies demonstrated minimal effects on transcription indicating H3K4me3 is not necessary for gene activation (Clouaire *et al.*, 2012; Hödl and Basler, 2012; Howe *et al.*, 2017). Therefore, the direct role of H3K4me3 on gene activation is still unclear however, the presence of this modification in association with H3K27me3 playing an important role in controlling gene expression.

1.3.3 Further Prominent Epigenetic Modifications

Aside from H3K27me3 and H3K4me3, a large number of other histone modifications exist on all of the histone proteins. Many of these modifications are not well described and have, as of yet, undefined functions in the cell. Despite this, a number of prominent and important modifications do have roles that have been determined. One such modification is the histone 3 lysine 27 acetylation (H3K27ac), which is associated with the TSSs of actively transcribed genes (Wang *et al.*, 2009; Bonn *et al.*, 2012). This modification is also highly associated with active enhancer regions and is often used to classify these regions as such (Creyghton *et al.*, 2010; Rada-Iglesias *et al.*, 2011; Bonn *et al.*, 2012). H3K4me1 is also highly associated with enhancer regions and, as is the case of H3K27ac, the presence of H3K4me1 is often used to distinguish enhancer regions within the genome. The presence of H3K4me1, however, does not indicate an active enhancer, unless in the presence of H3K27ac. In the absence of H3K27ac, H3K4me1 may indicate an enhancer region which is in a poised state. Enhancers in this poised

but non-active state may also have H3K27me3 present (Creyghton *et al.*, 2010; Rada-Iglesias *et al.*, 2011; Bonn *et al.*, 2012).

Trimethylation of histone 3 lysine 36 (H3K36me3) is another prominent modification. This modification is associated with the gene bodies of actively transcribed regions, with an enrichment in the decondensed regions (Bannister *et al.*, 2005; Barski *et al.*, 2007). This modification has also been associated with a number of other functions in the cell including alternative splicing, DNA repair and recombination (Wagner and Carpenter, 2012).

Trimethylation at histone 3 lysine 9 (H3K9me3) is also an important histone modification, implicated in chromatin silencing. This modification is enriched particularly at pericentric heterochromatin, in which it is believed to induce a self-reinforcing feedback loop involving HP1, a heterochromatin adaptor protein, maintenance via the UHRF1 protein, and the recruitment of de novo DNA methyltransferases (Fuks *et al.*, 2003; Lehnertz *et al.*, 2003). In addition, transcriptional repression associated with H3K9me3 involves the recruitment of DNA methylating enzymes and HP1 to the promoters of repressed genes (Yearim *et al.*, 2015). Another modification associated with repressed chromatin is H2AK119ub, which is mediated by the PRC1, and as previously mentioned, it is typically associated with H3K27me3/PRC2 (Wang *et al.*, 2004; Eskeland *et al.*, 2010; Blackledge *et al.*, 2014), discussed further in section 1.4 .

1.3.3.1 Bivalent Chromatin Domains

Histone modifications are often present in combination with each other. A commonly observed combination of histone marks leads to the formation of bivalent chromatin, which refers to the presence of both H3K4me3 and H3K27me3 at the promoter region of a gene. The existence of these overlapping modification was first identified through two independent studies investigating the histone modifications in mESCs (Azuara *et al.*, 2006; Bernstein *et al.*, 2006). In the study by Bernstein et al, 2006 a large percentage of TSSs modified by H3K27me3 were found to overlap with the genes also possessing the activating H3K4me3 modification. The existence of both modifications at the same gene was confirmed by sequential ChIP Real-Time PCR experiments. Interestingly, those bivalently modified genes were found to express RNA, at low levels, and they were enriched for developmental genes (Bernstein *et al.*, 2006). Supporting these findings, a parallel study by Azura and colleagues, 2006 also demonstrated the repressive H3K27me3 present at genes that possessed the active H3K4me3 and H3K9ac modifications (Azuara *et al.*, 2006). From these studies, the existence of bivalent domains was separately discovered and both studies showed the existence of these genes prominently in mESCs.

Shortly after their discovery in mESCs, whole genome analysis of hESCs indicated that a number of promoters also exist in this bivalent state. Two studies undertook a similar analysis, with one determining that approximately 16% of annotated genes (Pan *et al.*, 2007) had both H3K4me3 and H3K27me3 present at promoters in hESCs, with the second study indicating approximately 10% of all annotated genes are bivalent (Zhao *et al.*, 2007). In both studies the vast majority of H3K27me3 modified promoters also had H3K4me3 present, although a large number of promoters were also marked solely by the H3K4me3 modification. These bivalent genes had a strong overlap with the genes classified as bivalent in mESCs, including a similar enrichment in developmental genes (Pan *et al.*, 2007; Zhao *et al.*, 2007). Within reprogrammed iPSCs, the genes modified by H3K4me3 and H3K27me3 are found to be an almost exact overlap with those genes similarly modified in hESCs, including the existence of a large number of bivalently modified genes (Guenther *et al.*, 2007; Mikkelsen *et al.*, 2007).

It was hypothesised that bivalent chromatin domains may contribute to the developmental plasticity of pluripotent cells. The presence of both activating and repressive modifications was believed to maintain a gene in a poised state that was transcriptionally repressed but responsive to differentiation signals (Azuara *et al.*, 2006; Bernstein *et al.*, 2006). The enrichment of developmental TFs within the bivalently marked genes supports the theory that the presence of both modifications is important to control the expression of the developmental genes in these cell types, in the pluripotent cells and the resulting differentiating cells. Importantly, it was observed upon differentiation that genes will tend to resolve this bivalency, with genes losing H3K27me3 observed to increase in expression and those genes which lose H3K4me3 will become silenced (Bernstein *et al.*, 2006; Mikkelsen *et al.*, 2007; Pan *et al.*, 2007; Zhao *et al.*, 2007; Mohn *et al.*, 2008). Although some bivalency continues to be observed in lineage committed cells, this is at a lower number of genes compared to ESCs, and this retention will typically be in associated genes of different lineages, as opposed to the more global distribution in ESCs (Bernstein *et al.*, 2006; Mikkelsen *et al.*, 2007). Similarly, inducing loss of H3K27me3 such as in *Eed* depleted mESC, will cause the upregulation of lineage markers from these cells (Azuara *et al.*, 2006). This demonstrates that bivalent chromatin likely acts to maintain these important developmental genes in a responsive state to signals in the PSCs.

The mechanism of how bivalent chromatin is established and how it maintains the gene in a poised state is currently unclear. One of the key questions that remains in regard to bivalency is how the modified histone protein are organised within the cells. This could occur in

a number of different ways as indicated in Figure 1.3. The first of these is the heterogeneous population model, where the bivalency observed is not due to overlapping modifications at a single locus but caused by cells within a population containing opposing histone modifications. Although this would produce the pattern of bivalency observed in ChIP-sequencing experiments, and may sometimes be responsible for observed bivalency, multiple sequential ChIP experiments have demonstrated 'true' bivalency of both modifications present at the same single gene (Bernstein *et al.*, 2006; Roh *et al.*, 2006; Pan *et al.*, 2007; De Gobbi *et al.*, 2011; Mas *et al.*, 2018; Grzybowski *et al.*, 2019). However, with the knowledge that H3K4me3 and H3K27me3 are present at a single gene it is not clear which of the further three possibilities as indicated in Figure 1.3 exist. Some of the strongest evidence in favour of any of these confirmations, comes from a study which combined mononucleosome ChIP with mass spectrometry based quantitative profiling for histone modifications (Voigt *et al.*, 2012). Within this study, it was indicated that both H3K27me3 and H3K4me3 are present on a single mononucleosome, and the data suggest that these different opposing modifications predominantly exist asymmetrically - on the separate sister histone proteins within the nucleosome. By this mechanism, the presence of bivalent modifications was present at higher levels in ESCs, as compared to mouse embryonic fibroblast cells, which is also observed through conventional ChIP-sequencing methods (Voigt *et al.*, 2012). Further evidence for the presence of truly bivalent asymmetric nucleosomes was present by Shema and colleagues, through a single molecule imaging assay demonstrating H3K27me3 and H3K4me3 on a single nucleosome in ESCs and at a much reduced level in both EBs and fibroblasts (Shema *et al.*, 2016). Both of these suggesting 'true bivalency' with both opposing modifications present on a single histone protein. One key piece of evidence for this bivalency is through numerous sequential ChIP-seq (re-ChIP) experiments, in which one modification is isolated through ChIP, and then retargeted for a second modification ChIP and sequenced. In one study, individually combining ChIP-seq data for H3K27me3 and H3K4me3 in mESCs identified 6817 potentially bivalent genes, and through reChIP-seq experiments we identified 4778 and 5582 regions through H3K4me3-H3K27me3 and H3K27me3-H3K4me3 experiments, respectively (Mas *et al.*, 2018). Further re-ChIP experiments have been undertaken in a large variety of cell types including further studies in mESCs, mouse lung cells, mouse kidney cells (Weiner *et al.*, 2016), human T-cells (Kinkley *et al.*, 2016) and *Drosophila* neural progenitors (Sen *et al.*, 2019). All of these provide strong evidence that both H3K4me3 and H3K27me3 modifications will be found at the same histone protein. The development of further techniques such as the recent advent of

new single cell level analysis methods, including CUT & Tag, (Kaya-Okur *et al.*, 2019) will allow further investigation of this topic.

How these bivalent modifications have the ability to tightly control gene expression is also still not clearly defined. As previously discussed, the presence of H3K4me3 is typically associated with active transcription (Bernstein *et al.*, 2002; Santos-Rosa *et al.*, 2002; Barski *et al.*, 2007; Guenther *et al.*, 2007) and H3K27me3 with repressed genes (Boyer *et al.*, 2006; Lee *et al.*, 2006; Barski *et al.*, 2007), and therefore could explain the low expression from these bivalent modified genes in the pluripotent cells (Bernstein *et al.*, 2006; Mikkelsen *et al.*, 2007; Thalheim *et al.*, 2017). Recent work has indicated that it is through the presence of the H3K4me3 that the genes are protected from active repression (Douillet *et al.*, 2020). To further support this hypothesis it has been shown that the presence of H3K4me3 will protect H3 from PRC2 binding through allosteric inhibition (Schmitges *et al.*, 2011). Importantly, this inhibition was found to only occur when nucleosomes are symmetrically H3K4me3 modified, the presence of asymmetric H3K4me3 still allowed PRC2-mediated methylation (Voigt *et al.*, 2012), thereby allowing bivalent nucleosome to form. It has been estimated that the natural state of the histones is to retain H3K4me3, and a delay in recruitment of H3K27me3 upon DNA replication allows gene promoters to resolve bivalency (Petruk, Cai, *et al.*, 2017; Petruk, Mariani, *et al.*, 2017). In terms of maintaining developmental plasticity, that is receptive to dynamic changes, it has also been seen that poised RNA polymerase II will commonly be associated with the bivalently modified promoters, paused downstream of the TSS (Guenther *et al.*, 2007; Ferrai *et al.*, 2017). The open 3D chromatin structure within pluripotent cells may also contribute to this poised state of the cell, and the bivalency contributes to the open architecture (Petruk, Mariani, *et al.*, 2017; Mas *et al.*, 2018). It is likely a combination of factors contributes to maintain the poised state of the pluripotent cells, and similarly allows specification upon differentiation.

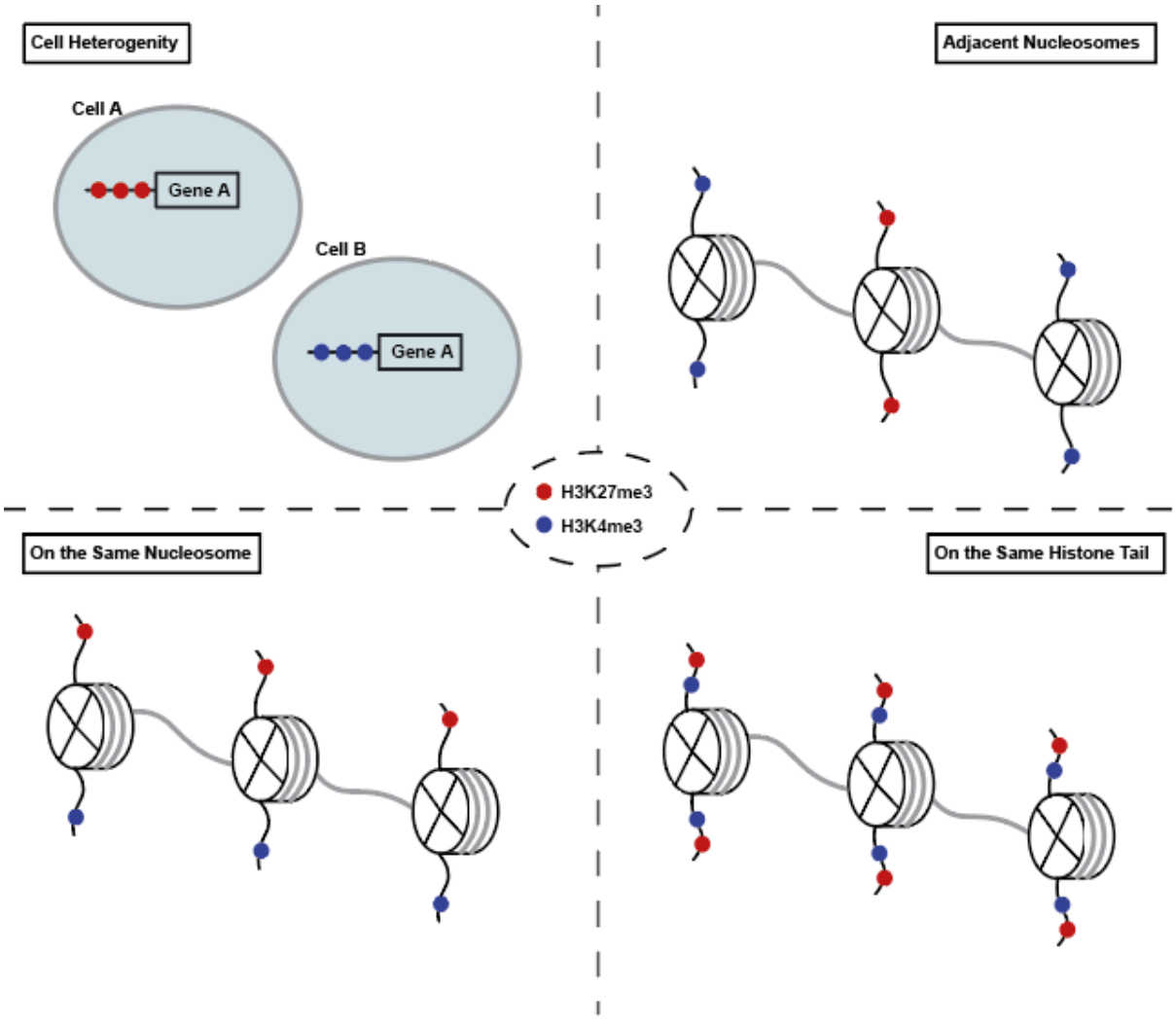


Figure 1.3 Hypothesised organisations of histone modifications in bivalency. Four proposed modification arrangements are indicated, all of which could be observed as bivalent modified genes when analysing cell populations, each coloured circle represents a modification present on the histone tail.

1.3.4.2 Mutually Exclusive Domains

As well as bivalency in which H3K27me3 and H3K4me3 are proposed to modify the same loci, a number of epigenetic modifications are found to exist in a mutually exclusive manner within the cell. This exclusivity, among other variations, can exist between an activating and a repressive modification, an example of which is the exclusivity in the H3K27me3 and H3K27ac modification. As these modifications exist at the same histone lysine residue, they must be mutually exclusive and the switch from the activating acetyl state to the repressive methylation modification by the nucleosome remodelling and deacetylation complex (NuRD) has

been observed in ESCs (Reynolds *et al.*, 2012). H3K27me3 is also proposed to exist in a mutually exclusive manner to H3K36me3, which is enriched in gene bodies. This exclusivity was first observed through large-scale genomics data from the human ENCODE project, which demonstrated little/no overlap between H3K27me3 and H3K36me3 modifications (Dunham *et al.*, 2012; Voigt *et al.*, 2012). This mutual exclusivity may be caused by an intrinsic activity of the PRC2 complex itself. It was demonstrated that the PRC2 accessory subunit, PHF19, can bind to H3K36me3, via its Tudor domain. This protein can also recruit the H3K36me3 histone demethylase, resulting in the loss of H3K36me3 and gain of H3K27me3 and gene silencing (Ballaré *et al.*, 2012; Brien *et al.*, 2012). Similarly, the presence of H3K36me3 and H3K36me2 has also been shown to be inhibitory of PRC2 in an allosteric manner (Schmitges *et al.*, 2011; Yuan *et al.*, 2011), thought to be responsible for the very low levels of overlapping H3K27me3 and H3K36me3 found in these experiments, below 0.1% of histones (Schmitges *et al.*, 2011).

Mutual exclusion is not only observed between repressive and active modifications but can also exist between those epigenetic modifications that have similar functional outcomes. For example, several studies have observed H3K9me3 and H3K27me3 present in a mutually exclusive manner (Peters *et al.*, 2003; Pauler *et al.*, 2009; Cooper *et al.*, 2014; Saksouk *et al.*, 2014). Moreover, it has been demonstrated that loss of SUV39H1/H2 function, the methyltransferase responsible for H3K9me3 addition, resulted in increased H3K27me3 levels, suggesting H3K9me3 may inhibit the addition of H3K27me3 at the same location (Cooper *et al.*, 2014; Saksouk *et al.*, 2014). However, a number of studies also suggest H3K9me3 and H3K27me3 may overlap on some genes, in a variety of cell types (Bilodeau *et al.*, 2009; Hawkins *et al.*, 2010). Similarly, in DNA methyltransferase knockout cells, H3K27me3 is observed to overlap with H3K9me3 modified sites (Saksouk *et al.*, 2014). Lastly, mass spectrometry analysis of HeLa nucleosomes demonstrated some limited overlap with H3K9me3 (Voigt *et al.*, 2012). It is therefore unclear if the H3K27me3 and H3K9me3 modifications are genuinely mutually exclusive, as initially believed, but may be partially inhibitory, or only present in some cell types.

1.4 Polycomb Group Proteins

Polycomb group proteins were first identified in 1947 (Lewis, 1947), and the importance of this protein family for the segmentation of the Drosophila embryo body was discovered in 1978 (Lewis, 1978). The Polycomb proteins were then identified as transcriptional repressors,

initially in Drosophila embryos. Further genetic experiments confirmed the ability of these proteins to undertake transcriptional silencing and maintain genes in a repressed state in both Drosophila and mammalian cells (Bunker and Kingston, 1994; Müller, 1995). Since these initial discoveries, a large amount of research has been undertaken into the Polycomb group proteins in a variety of organisms from plants to mammals (Blackledge, Rose and Klose, 2015; Del Prete et al., 2015; Schuettengruber et al., 2017). This includes the identification of two separate and distinct Polycomb repressive complexes. The first complex identified was found to contain the RING1A/1B component and is involved of transcriptional repression. This first identified complex of Polycomb proteins was named Polycomb Repressive Complex 1 (PRC1) (Reijnen *et al.*, 1995; Satijn *et al.*, 1997; Shao *et al.*, 1999). Further analysis of the Polycomb proteins then identified that the Drosophila Polycomb proteins – ESC and EZ interact together independently of the PRC1 complex. The mammalian homologues, EED and EZH2, were shown to do the same, indicating the presence of a second Polycomb Repressive Complex - PRC2 (Sewalt *et al.*, 1998; Ng *et al.*, 2000).

The PRC1 complex acts as a multi-protein chromatin modifying enzyme containing a RING1B/1A catalytic component that is able ubiquitinate lysine 119 on histone H2A (Wang *et al.*, 2004; Gao *et al.*, 2012). This complex has numerous forms, based on its with various associated subunits (Figure 1.4), however none of the proteins within the complexes are common between the PRC1 and PRC2. The numerous PRC1 forms, are typically separated into the canonical and non-canonical complex groups. The PCGF2 and PCGF4 can form canonical complex, due to the presence of CBX chromodomain protein (Fischle *et al.*, 2003; Wang *et al.*, 2004; Gao *et al.*, 2012). The non-CBX containing PRC1 complexes – which can contain any of the six PCGF proteins are referred to as the ‘non-canonical’ or ‘variant’ PRC1 complexes. The presence of H2AK119ub is associated with gene repression, likely though an ability to induce chromatin compaction and inhibition or RNA II polymerase elongation (Min, Zhang and Xu, 2003; Wang *et al.*, 2004; Eskeland *et al.*, 2010). The interplay between PRC1 and PRC2 has long been established, with the initial hierarchical models demonstrating the recruitment of PRC1 to the H3K27me3 modified regions (Cao *et al.*, 2002; Czermin *et al.*, 2002; Fischle *et al.*, 2003; Min, Zhang and Xu, 2003; Wang *et al.*, 2004). However, more recent investigation of the PRC1 and PRC2 interplay has also demonstrated an increasing role of the H2AK119ub in recruiting the PRC2 complex (Blackledge *et al.*, 2014, 2020; Cooper *et al.*, 2014; Kalb *et al.*, 2014; Almeida *et al.*, 2017)

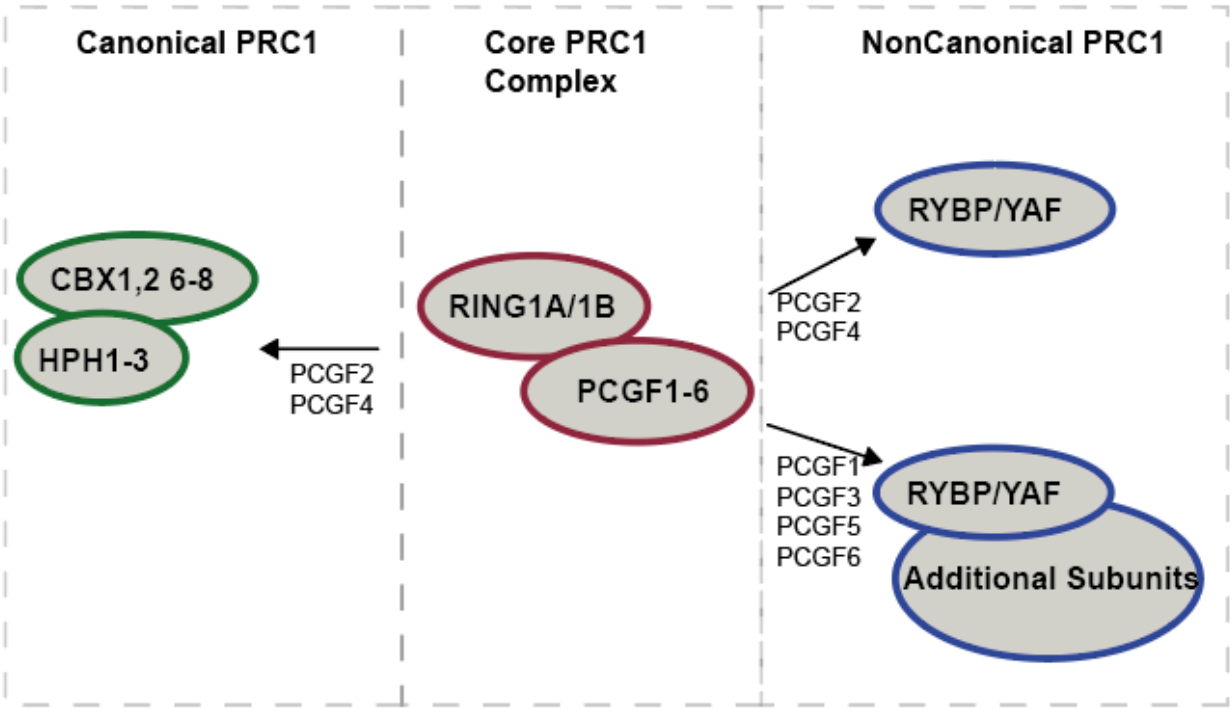


Figure 1.4 The Composition of PRC1 Complexes. The PRC1 complex can be composed of various subunits but will always consist of a minimum of a RING1A/1B, one of six PCGF proteins, and at least one other subunit, as indicated. Numerous additional subunits can be incorporated into the noncanonical complexes with PCGF1,3,5 and 6, forming unique complex but due to the large variations these are not indicated.

1.4.1 Polycomb Repressive Complex 2 Core Subunits

PRC2 is also a multi-protein chromatin modifying enzyme complex, which is responsible for catalysing of mono-, di- or tri- methylation to histone 3 lysine 27 (Cao *et al.*, 2002; Czermin *et al.*, 2002; Kuzmichev *et al.*, 2002; Müller *et al.*, 2002). Mammalian PRC2 is composed of a core of functionally essential proteins, and a number of accessory proteins, Figure 1.5. One of the two possible enzyme subunits of PRC2, EZH1 or EZH2 will be present with two further core components: embryonic ectoderm development (EED), suppressor of zeste (SUZ12) (Czermin *et al.*, 2002; Müller *et al.*, 2002; Margueron *et al.*, 2008). The CAF1 histone-binding proteins RBBP4 and RBBP7 are necessary and sufficient for the methyltransferase ability of PRC2 (Cao and Zhang, 2004; Nekrasov *et al.*, 2005). The importance and contribution of these components to the complex has been largely studied

through perturbation and biochemical studies which are commonly undertaken in mESCs (Margueron and Reinberg, 2011; Laugesen and Helin, 2014; van Mierlo *et al.*, 2019).

The catalytic component of PRC2 relies on the activity of the SET-domain containing methyltransferase subunits EZH1 and EZH2 (Cao *et al.*, 2002; Czermin *et al.*, 2002; Kuzmichev *et al.*, 2002; Müller *et al.*, 2002). In mammals, these two genes can both form part of a functional PRC2 complex, where they are present in a mutually exclusive manner. Evidence appears to show the two proteins have differential expression patterns, with EZH1 expressed in more terminally differentiated cells and EZH2 enriched in cells that are actively dividing, and therefore EZH2 is the main catalytic subunit of PRC2 in embryonic development and has a higher catalytic activity (Margueron *et al.*, 2008; Lee *et al.*, 2018). When *Ezh2* null mice were generated, the resulting embryos were non-viable, with a failure to develop shortly after implantation or incomplete gastrulation (O'Carroll *et al.*, 2001). Interestingly knockout of *Ezh2* in mESCs had little effect on the undifferentiated cells, including on morphology, self-renewal and proliferation, with only a small change in gene expression. However, upon differentiation the *Ezh2* null cells demonstrated highly impaired differentiation and proliferation (Chamberlain, Yee and Magnuson, 2008; Shen *et al.*, 2008). Similarly, EZH2 knockout hESCs had a much stronger phenotype compared to mouse cells, with loss of proliferation and self-renewal observed in these mutant cells, along with a loss of differentiation potential (Collinson *et al.*, 2016; Shan *et al.*, 2017). These studies demonstrate the key role of the EZH catalytic component in the PRC2 function in both pluripotency and lineage specification.

Although the EZH catalytic component is essential for the function of the PRC2, the complex can only undertake methyltransferase activity when complexed further core PRC2 subunits. It is proposed that EED functions with PRC2 by binding to histone methylated lysine's and stimulating the methyltransferase activity of the complex. Crystal structure analysis of EED demonstrated an aromatic cage assembly formed from WD40 repeats of the protein. Three of the amino acids within these repeats - Phenylalanine97, Tyrosine148 and Tyrosine365 contact the trimethylation lysine residue of the histone proteins and are thought to be important in propagation of the H3K27me3 at silencing regions (Margueron *et al.*, 2009). Mutation of either Phenylalanine97 or Tyrosine365 was observed to decrease global H3K27me3 levels in mESCs and the resulting cells could not undergo normal differentiation, however, unlike full EED knockout cells, PRC2 could still form (Oksuz *et al.*, 2018). EED is therefore likely to function through stabilising the complex and maintain H3K27me3 at repressed regions of the genome.

The core SUZ12 protein is also essential in PRC2 through a DNA binding function, and a complex stabilisation function. Similar to the phenotype seen in *Ezh2* and *Eed* null mice, *Suz12* mutant mouse embryos are non-viable, with lethality observed shortly after implantation. These KO embryos have largely decreased levels of H3K27me2 and H3K27me3 (Pasini *et al.*, 2004). The SUZ12 protein has two functional domains, a zinc finger motif and a VEFS domain, which is responsible for the interaction with EZH2 (Yamamoto *et al.*, 2004). Expression of a SUZ12 VEFS domain within *Suz12* knockout mESCs was enough to stabilise PRC2 and restore global methylation levels, but expression of this domain did result in an aberrant methylation pattern compared to wild-type cells. In contrast, expression of SUZ12 with a deleted VEFS domain demonstrated protein binding at the normal wild-type SUZ12 bound regions (Højfeldt *et al.*, 2018). This suggests SUZ12 plays a role in PRC2 through a stabilisation and binding to DNA regions, essential for the normal function of the complex.

The final core component, RBBP4 or 7 (also known as RBAP48 and RBAP46, respectively), are two highly similar proteins that can bind histones (Cao *et al.*, 2002; Murzina *et al.*, 2008). These proteins are incorporated into multiple complexes and therefore it has been difficult to deduce their direct role solely in PRC2 (Zhang *et al.*, 2013; Vizán *et al.*, 2015). It is hypothesised that these core proteins are not required for the full methylase activity under all circumstances. However, RBBP4/7 are expected to be functionally important for nucleosome binding through the presence of a WD40 domain within both forms of the protein (Kuzmichev *et al.*, 2002; Nekrasov, Wild and Müller, 2005; Murzina *et al.*, 2008; Schmitges *et al.*, 2011).

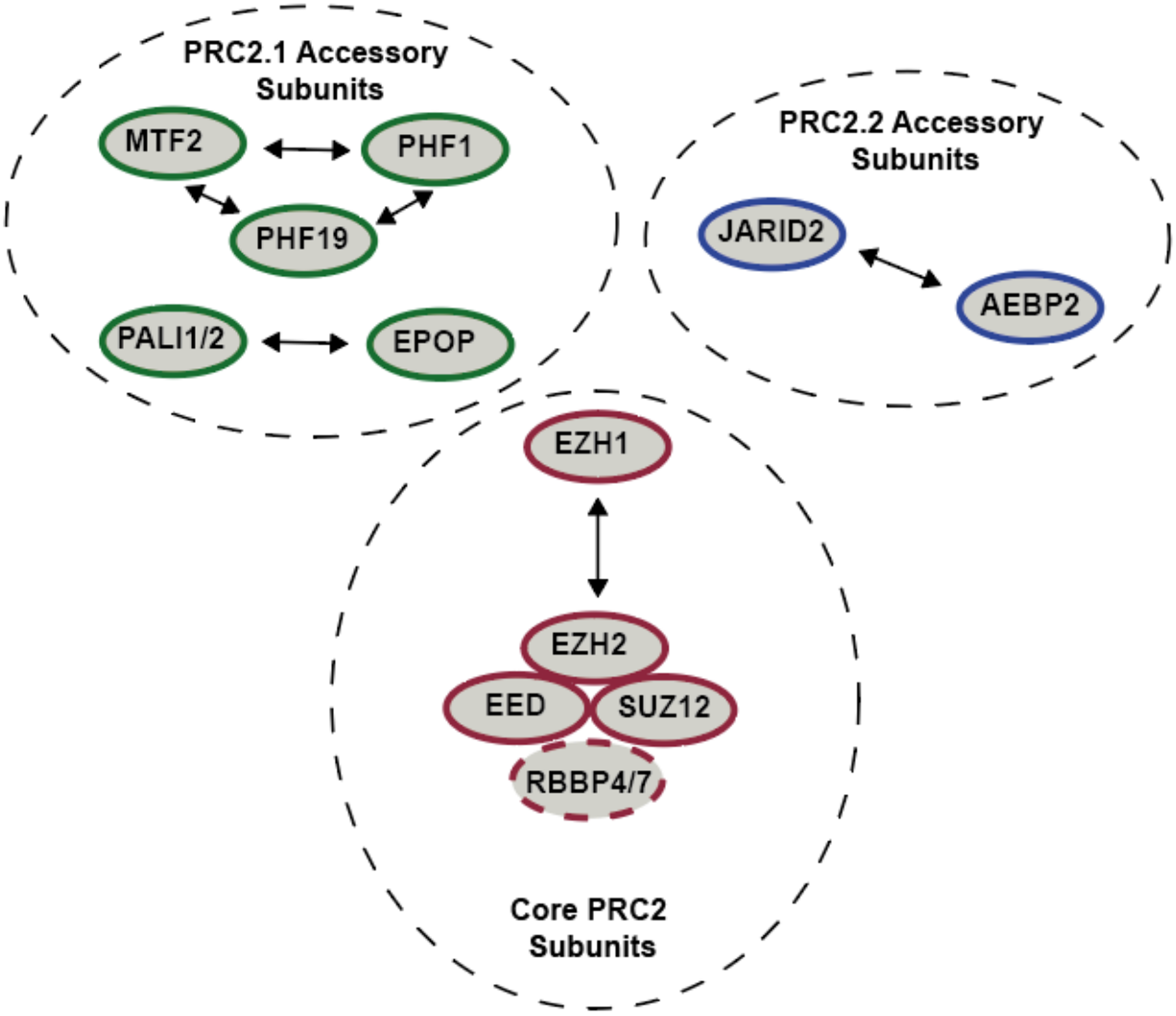


Figure 1.5 The core and accessory subunits of PRC2. PRC2 can exist as either PRC2.1 or PRC2.2, depending on the subunits which bind to the core components of the complex. Mutually exclusive binding within the complexes are indicated by double headed arrows.

1.4.2 Polycomb Repressive Complex 2 Additional Subunits

Along with the core components, a number of cofactors have been identified to associate with the PRC2, which act to modulate both target binding and enzymatic activity of the complex. A number of proteins transiently interact with the complex but many more form more stable interactions with the complex and are co-purified. However, unlike the previously mentioned core components, these auxiliary subunits are not essential for the methyltransferase

activity of PRC2, instead the exact functions of many of these cofactors are still unknown, with the significance of their PRC2 interaction unclear (Margueron and Reinberg, 2011; Kloet *et al.*, 2016; Oliviero *et al.*, 2016; van Mierlo *et al.*, 2019). The association of these accessory subunits occurs in two different configurations called PRC2.1 and PRC2.2, shown in Figure 1.5, and the contribution of these two different forms of the complex are also still to be further investigated (Hauri *et al.*, 2016; van Mierlo *et al.*, 2019).

PRC2.1 is composed of numerous accessory subunits which will interact with the core enzyme units, in a mutually exclusive manner. When forming PRC2.1, one of either PHF1, MTF2 or PH19 (also referred to as Polycomb-like 1, 2 and 3, respectively), interact with the core components in a mutually exclusive manner (Smits *et al.*, 2013; Oliviero *et al.*, 2016). All three proteins contain a Tudor domain, two PHD fingers, an extended homology domain, and a conserved C-terminal domain. The Tudor domains bind preferentially at H3K36me3 chromatin regions, and through a separate domain the proteins bind unmethylated CpG DNA, thereby linking PRC2 to chromatin (Li *et al.*, 2017). MTF2 is thought to be of particular importance in PRC2 recruitment to the DNA and both PHF1 and PHF19 are shown to increase the catalytic activity of the methyltransferase (Sarma *et al.*, 2008; Hunkapiller *et al.*, 2012; Li *et al.*, 2017). In mESCs, these three accessory proteins are not at equimolar levels, but instead MTF2 is at approximately ten times higher levels compared to the other proteins, and therefore the majority of PRC2.1 in mESCs contain this subunit (Smits *et al.*, 2013; Kloet *et al.*, 2016; Oliviero *et al.*, 2016).

In combination with one of the Polycomb-like proteins, PRC2.1 will contain either EPOP (C17orf96) or PALI1/2 (C10ORF12), in a mutually exclusive manner (Smits *et al.*, 2013; Kloet *et al.*, 2016; Oliviero *et al.*, 2016). EPOP is hypothesised not to have an effect on H3K27me3, but instead links Elongin BC, a positive regulator of RNA Polymerase II, to PRC2 bound chromatin allowing low levels of transcription from the bound region (Beringer *et al.*, 2016). When not bound by EPOP, PRC2.1 contains PALI1/2, and in mESCs that are undergoing differentiation the PRC2.1 composition shifts from EPOP to PALI1/2 (Kloet *et al.*, 2016). The role of PALI1/2 in binding PRC2.1 is unclear, however the depletion of PALI1 from mESCs causes only a small decrease in H3K27me3 levels, but it is thought the proteins may link transcriptional corepressors to PRC2 (Conway *et al.*, 2018).

PRC2.2 is composed of the core PRC2 subunits interacting with JARID2, a catalytically inactive member of the *Jumonji* histone demethylase family, or the Adipocyte enhancer-binding

protein 2 (AEBP2) protein (Kim, Kang and Kim, 2009; Peng *et al.*, 2009; Shen *et al.*, 2009; Pasini *et al.*, 2010; Hauri *et al.*, 2016; Kloet *et al.*, 2016; Oliviero *et al.*, 2016) The exact role that these proteins play in terms of the PRC2 complex function is still being investigated, however histone methyltransferase assays showed that the incorporation of these two cofactors greatly increases the enzyme activity of PRC2 (Son *et al.*, 2013). JARID2 is expected to promote activity by acting as a scaffold and causing an activating conformational change in PRC2.2 (Justin *et al.*, 2016). This accessory protein is also believed to assist in targeting PRC2.2 to chromatin, with mESCs depleted of JARID2 showing decreased PRC2.2 enrichment at chromatin (Peng *et al.*, 2009). The binding of Jarid2 to chromatin could be partially explained by a ubiquitin interaction motif at the amino-terminus of Jarid2 that facilitates localization to H2AK119u1 (Cooper *et al.*, 2016), .role of AEBP2 in PRC2.2 is less defined, but it is hypothesised that it may increase the stability of PRC2.2, as this been demonstrated in structural analysis of the complex (Kasinath *et al.*, 2018). This protein may also be important for recruitment to chromatin, as AEBP2 has affinity for binding to nucleosomes, particularly those modified with H2AK119ub (Kalb *et al.*, 2014). The importance of the PRC2.2 subunits in recruitment through H2AK119ub was demonstrated through the loss of H2AK119ub1, that induced a rapid displacement of PRC2 activity, specifically PRC2.2 activity compared to PRC2.1, and a loss of H3K27me3 deposition (Tamburri *et al.*, 2020)

1.4.3 Targeting of H3K27me3 modifications

The process of targeting PRC2, and subsequent H3K27me3 modification, to regions of the genome is still not completely understood. Although no clear mechanism is yet determined in mammalian cells, in Drosophila cells PRC2 targeting is achieved through specific regions of DNA, which are entitled Polycomb repressive elements (PREs). These consensus sequences bind PRC2 through interaction with other DNA binding factors. The deletion or manipulation of the sequence of the PRE results in the site-specific loss of H3K27me3 (Coleman and Struhl, 2017; Laprell *et al*., 2017). A similar consensus recruiter sequence like PREs has not been observed, to date, in other organisms including vertebrates (Schuettengruber *et al.*, 2017). However, genome-wide studies in mammalian cells have identified a strong association between Polycomb binding and CpG Islands (CGIs). Furthermore, artificial GC-rich elements have shown in numerous independent experiments that they are sufficient for recruitment of Polycomb proteins, strongly suggesting that CGIs may act as PREs in mammalian cells (Mendenhall *et al.*, 2010; Lynch *et al.*, 2012; Jermann *et al.*, 2014; Riising *et al.*, 2014). These CGI elements were also shown to be sufficient to recruit trithorax/H3K4me3, and data indicates

that a bivalent state was the default for CpG-rich DNA sequences (Wachter *et al.*, 2014). Within the cell environment the PRC2 can be seen to also induce de novo methylation at CGI regions, this has observed from knockout/loss of function experiments in key PRC2 components, after which do novo H3K27me3 can be established upon reintroduction of the necessary proteins (Højfeldt *et al.*, 2018; Oksuz *et al.*, 2018; Lavarone, Barbieri and Pasini, 2019). Although the CGI content is important in PRC2 recruitment, numerous other factors control the deposition of H3K27me3, an example of which is transcription. It has been demonstrated that the presence of active transcription will inhibit H3K27me3 modification and transcriptional inhibition results in ectopic PRC2 recruitment in mESCs (Riising *et al.*, 2014). One of the strongest factors in controlling PRC2 recruitment and H3K27me3 deposition, and potentially how transcription can have an effect is through other histone modifications present.

The interplay between H3K27me3 and active marks, most specifically H3K4me3 at these regions can cause a multitude of outcomes in terms of PRC2 binding and the presence of H3K27me3 modifications. The presence of H3K36me3 is known to act in a mutually exclusive manner to H3K27me3, therefore the presence of this even with CG-rich regions will likely inhibit the ability of PRC2 and thereby protecting the marked region from repressive modifications and maintaining an active transcriptional state (Schmitges *et al.*, 2011; Yuan *et al.*, 2011; Dunham *et al.*, 2012; Voigt *et al.*, 2012). In the case of H3K4me3, if both histone proteins are modified by the H3K4me3 this will also protect from H3K27me3 addition of PRC2 by allosteric mechanisms maintain a modified gene in an active state (Schmitges *et al.*, 2011; Petruk, Mariani, *et al.*, 2017)). However, the presence of asymmetric H3K4me3 does not protect from H3K27me3, therefore in the promoter regions of CG-rich promoters can gain both H3K27me3 in the presence of H3K4me3 or gain H3K4me3 in the presence of H3K27me3, thereby preventing further modifications while retaining the gene in this poised, or partially active state (Voigt *et al.*, 2012; Douillet *et al.*, 2020).

Although the CG-rich promoters regions have been shown to be sufficient in recruiting H3K27me3 deposition by PRC2, within the cell de novo modification will rarely occur and the majority of H3K27me3 deposition is during the cell cycle (Reverón-Gómez *et al.*, 2018), in which the Polycomb protein will act to maintain methylation and silencing. As previously discussed, the PRC1 will, in at least some instances, work in sequence with the PRC2, recruited by the H3K27me3 to deposit H2AK119ub and therefore bring about chromatin and gene silencing. The presence of H2AK199ub will stabilise the binding of the PRC2 to the modified region resulting in H3K27me3 addition, most likely through the Jarid2 component of PRC2 (Blackledge *et al.*,

2014; Cooper *et al.*, 2016; Fursova *et al.*, 2019), further reinforcing the repressive modification present at this region. The H3K27me3 modification can also induces self-reinforcement through an allosteric positive feedback loop through the WD40 domain of EED (Margueron *et al.*, 2009; Poepsel, Kasinath and Nogales, 2018; Perino *et al.*, 2020). This allows spreading of the H3K27me3 (Oksuz *et al.*, 2018) throughout the regions marked by the Polycomb modifications, forming large condensed regions of chromatin, retaining it in the repressive state and protecting from transcription and other histone modifications.

However, Polycomb proteins are only located at ~30% of mammalian GCIs in ESCs with these bound regions correspond to repressed promoters (Ku *et al.*, 2008; Riising *et al.*, 2014), but some evidence suggest sequence alone is not responsible for recruitment of PRC1/2 in mammalian cells. Furthermore, due to the dynamic nature of gene transcription in changing cellular conditions further mechanisms are thought to assist in directing PRC2 targeting. A second proposed action of PRC2 recruitment is through interactions with RNA molecules. This was first proposed as the targeting of Polycomb to the inactive X chromosome is known to involve the long non-coding RNA, XIST. The recruitment of PRC2 was demonstrated to occur in associate with long non-coding RNAs. Although the role of these RNAs in recruitment is still unclear, and the relevance of the association between RNAs and PRC2 in recruitment is still disputed (Zhao *et al.*, 2008; Davidovich and Cech, 2015; Portoso *et al.*, 2017; X. Wang, Goodrich, *et al.*, 2017). Evidence also exists for the ability of short RNAs to function in PRC2 recruitment, with RNA transcribed from H3K27me3 marked genes forming stem-loops which will associate with PRC2. The addition of RNase A, or a chemical inhibitor of transcription, or the use of an RNA-binding-defective mutant, were all observed to disrupt PRC2 chromatin occupancy and localization genome wide (Kanhere *et al.*, 2010; Long *et al.*, 2020). However, Polycomb proteins and RNA interactions have also been observed to inhibit PRC2 recruitment to chromatin (Beltran *et al.*, 2016; X. Wang, Paucek, *et al.*, 2017; Zhang *et al.*, 2019), therefore the role of RNA in PRC2 binding is currently unclear, with an apparent opposing effect depending on the binding RNA and other factors present.

1.5 The Role of Polycomb Repressive Complex 2 in Pluripotency and Differentiation

1.5.1 The role of PRC2 in pluripotency

PRC2 has an essential role in early mouse development, which can be seen from knockout studies in developing mouse embryos. Mouse embryos deficient for *Eed, Suz12* or *Ezh2,* are non-viable, as these embryos are not able to progress past early development. Similarly, it was originally not possible to isolate mESCs from the knockout embryos (Faust *et al.*, 1998; O'Carroll *et al.*, 2001; Pasini *et al.*, 2004). This suggests that PRC2 plays a role in maintaining pluripotency of mESCs (Boyer *et al.*, 2006). However, some studies have also demonstrated that knockout of these core complex subunits does not appear to have a significant effect on previously derived mESCs. For example, depletion of *Suz12* in mESCs caused a large decrease in H3K27me3 levels, and upregulation of some differentiation genes, but the mESCs appeared phenotypically normal and could be isolated from blastocysts (Pasini *et al.*, 2007). Similarly, through the creation of *Eed* null mouse ESC, Chamberlain and colleagues, demonstrated higher expression of developmental factors, but no obvious effect of the knockout in the pluripotent stem cells (Chamberlain, Yee and Magnuson, 2008). These apparent contradictory studies raise questions for how important the PRC2 complex is for maintaining pluripotency in mESCs. This could be explained by a role for the Polycomb proteins in preparing the ESCs for lineage commitment, by controlling expression of a key set of developmental genes which are responsive downstream from the pluripotency release and therefore does not directly affect pluripotency but what occurs after the exit from this state.

Contrasting the non-essential role of PRC2 in undifferentiated mESCs, this epigenetic complex is proposed to play a more active role in maintaining primed hPSCs (Collinson *et al.*, 2016; Moody *et al.*, 2017; Shan *et al.*, 2017). Generation of hPSCs with loss of either *EZH2, SUZ12* or *EED*, resulted in severely compromised hPSCs populations. The cells were capable of maintenance in culture for a number of passages, but the resulting populations had restricted growth compared to wild-type cells, and knockout cells underwent gradual differentiation when maintained in pluripotent culture conditions (Collinson *et al.*, 2016; Moody *et al.*, 2017; Shan *et al.*, 2017). Transcriptional analysis of three key pluripotency factors, *OCT4, SOX2 and NANOG* in the knockout cells demonstrated a strong decrease in these factors, demonstrating the loss of stemness from the cells. Similarly, the *EZH2, SUZ12* or *EED* knockout hPSC lines demonstrated an upregulation of lineage specific factors, with a particular bias towards

mesoderm and endoderm transcript upregulation (Collinson *et al.*, 2016; Shan *et al.*, 2017). This would suggest that within hPSCs PRC2 plays an important role in maintaining pluripotency, with loss of the complex resulting in upregulation of developmental genes normally repressed in these cells.

1.5.2 PRC2 in Early Lineage Specification

Although loss of PRC2 from mESCs did not have a clear effect on cell pluripotency, the loss of PRC2 activity did affect the ability of mESCs and hESCs to undergo efficient exit from pluripotency into early differentiation (Pasini *et al.*, 2004; Boyer *et al.*, 2006; Collinson, Collier, *et al.*, 2016; Moody *et al.*, 2017; Shan *et al.*, 2017). PRC2 was predicted to play a role in differentiation, likely through a function in maintaining the low expression, but responsiveness of bivalent marked promoters to differentiation signals within pluripotent cells. The enrichment of bivalent chromatin at developmental genes and the observation that they commonly resolve into single modified histone upon differentiation (Bernstein *et al.*, 2006; Mikkelsen *et al.*, 2007), further supports this role of PRC2 in gene repression in pluripotency and allowing proper activation of lineage factors in differentiation.

To determine if the function of PRC2 is important for the specification of all three lineages, studies utilised loss of function analysis for both human and mouse pluripotent cells. Firstly, it was shown that *Suz12* null mESCs could not differentiate into neuronal cell types, and these cells could also not form proper embryoid bodies comparable to wild-type cells. The *Suz12*$^{-/-}$ cells were also incapable of sufficiently silencing pluripotent factors and could not form teratomas in immunodeficient mice (Pasini *et al.*, 2007). Similarly, differentiation and chimera generation from *Eed* null mESCs demonstrated an inability to differentiate into mature cells. Although these pluripotent cells were able to contribute to chimera formation, there is a skewing in the cell types formed, and will have minimal contribution to mesodermal cells (Chamberlain *et al.*, 2008). Human PSCs with PRC2 loss of function also demonstrated deficiencies in their ability to differentiate. In these cells, there was a strong defect in forming ectoderm, although there may be some ability to initially form cells in all three lineages the ability to outgrow these cells is not maintained as they are further differentiated and expanded (Collinson *et al.*, 2016; Shan *et al.*, 2017). Therefore, PRC2 loss of function plays a role in the early lineage specification, through the mis-expression of lineage-specifying factors. To date different studies indicate an importance for varying lineages, to different extents (Collinson *et al.*, 2016; Shan *et al.*, 2017) that may be explained through signalling molecules within different culture conditions,

but these studies do show an inefficient differentiation capacity as a consequence of loss of PRC2 function.

1.5.3 PRC2 and Epigenetic Contributions to Pancreatic Differentiation

As well as a role in pluripotency and early lineage specification, PRC2 has also been proposed to function in the development and maintenance of mature cells, including pancreatic cell types (Dumasia and Pethe, 2020). The PRC2 will play a role in early lineage determination, and the first stage within pancreatic endocrine formation is to produce DE cells. The differentiation to DE is controlled through activin/nodal signalling, including a SMAD2-mediated eviction of EZH2 from the promoters of mesoendoderm genes such as MIXL, T, NODAL, thereby promoting the induction of these genes and subsequent cell differentiation (Wang *et al.*, 2017). Specifically, the induction of DE induces the rapid loss of H3K27me3 from numerous gene bivalently modified in the pluripotent cells including endoderm genes such as *SOX17, EOMES, MIXL1* (Xie *et al.*, 2014). Therefore, the activity of PRC2 and, importantly, the removal of PRC2 from endoderm genes demonstrates a role for this complex in early pancreatic differentiation.

As cells are differentiated further into pancreatic cell types, PRC2 activity is again observed to have a role in the progression of the cells. Similar to the function of PRC2 in forming DE, where selected removal of the complex enables cell differentiation, the same is required for pancreatic progenitor formation. At this development stage, EZH2 controls the decision of the cells to enter either a pancreatic or liver fate. Loss of *Ezh2*, through a conditional allele, in developing mice causes an increase in the early forming pancreatic compartment and a resulting decrease in the liver bud. This suggests that PRC2 acts to restrain pancreatic specification in the endoderm of developing mouse cells (Xu *et al.*, 2011). This could potentially be through the presence of H3K27me3 at important factors involved in pancreatic progenitor specification. A study by Xu et al, 2014, which analysed H3K27me3 in cells from sequential stages of a developing mouse pancreas, determined the presence of H3K27me3 at the *Pdx1* gene in endoderm cells, a factor known to be essential in pancreatic cell differentiation. Depletion of Ezh2 from the developing pancreatic cells through a *Pdx1*-Cre model, caused an increase in the number of endocrine cells, as indicated through increased in the number of NGN3 positive cells (Xu *et al.*, 2014). Similarly, a second study determined that the loss of H3K27me3 is important for the transition of cells from a *Ngn3* low to *Ngn3* high state, which is indicative of endocrine formation. Associated with the removal of H3K27me3 was a large

increase in the transcription of TFs important for endocrine specification including *Neurod1, Rfx6* and *Insm1* (Yu *et al.*, 2018). The *Ngn3*, and an associated enhancer of *Ngn3*, are marked by H3K27me3 in non-pancreatic mouse cell types, thereby inhibiting the differentiation of cells towards a pancreatic endocrine phenotype (Van Arensbergen *et al.*, 2017; Yu *et al.*, 2018). Therefore, PRC2 is likely to act in developing cells to restrict the formation of pancreatic progenitors/endocrine cells, through the presence of H3K27me3 at key developmental factors in pancreatic development.

Epigenetic factors, particularly PRC2, are also important in maintaining the stability of mature islet cells. This was demonstrated through experiments that disrupted PRC2 components specifically in β-cells through creation of a β-cell specific *Eed* KO mice. These mice did not have an obvious phenotype during development, and were born healthy, however islets from 25-week-old β cell *Eed* KO mice had almost no insulin positive cells, with the majority of the islet regions lacking any hormone expressing cells. The islet cells from these mice had a strong decrease in the expression of numerous endocrine/β-cell markers including *Pdx1, MafA, Nkx2.2*, and *Nkx6.1* (Lu *et al.*, 2018). Through the transcriptional analysis of human cells obtained from diabetic individuals, this study also suggests a role of PRC2/H3K27me3 in maintaining β-cell integrity, with loss of PRC2 associated with cell dedifferentiation (Lu *et al.*, 2018). Interestingly, this suggests that PRC2 may function during pancreatic development to restrict pancreatic endocrine formation, and once formed, PRC2 is necessary to retain the identity of the mature endocrine cells.

1.6 Thesis Overview

From previous work on generating human pancreatic cells *in vitro*, it is clear that the production of hormone producing cells is possible. However, several problems continue to persist at the conclusion of most differentiation protocols when the cells are compared to those derived from an adult pancreas. Investigating the final-stage cells that are produced and their differentiation status is likely to advance this important area of research, with relevance for cell therapy. Studies have indicated the importance of PRC2 throughout pancreatic development, functioning in both the restriction of the developing endocrine cells and in safeguarding the stability of the mature cell types. However, much of this work has focused on developing mouse cells, and the applicability of this system to human *in vitro* development is still to be fully established. We hypothesise that PRC2, through H3K27me3, might be restricting the *in vitro* differentiation of human pancreatic cells through the transcriptional repression of key pancreatic factors. Studying H3K27me3 throughout pancreatic differentiation, and the response of cells to PRC2 perturbation during this process, is important to establish whether PRC2 has a role in controlling the development and function of human pancreatic cell types.

The specific aims of this thesis are as follows:

- To profile the genome-wide localisation of histone modifications – H3K27me3 and H3K4me3 – and the associated transcriptional dynamics, during the *in vitro* differentiation of human pancreatic cells.

- To compare the transcriptional and epigenetic programmes between *in vitro* derived and *in vivo* developed human pancreatic cell types.

- To establish the role of PRC2 during human pancreatic cell differentiation including in the final-stage mature endocrine cells.

- To develop new cellular tools that will enable the efficient and controllable investigation of epigenetic modifiers in pluripotency and differentiation

Chapter 2

Material and Methods

2.1 Cell Culture

2.1.1 Human Pluripotent Stem Cell Culture

For this study the following cell lines were used: human ESC line H9 (WA09; obtained from WiCell Research Institute) (Thomson, 1998), the human iPSC line HDF derived from fibroblasts by reprogramming (obtained from DefiniGEN) and the WTB iPSC CRISPRi Gen1B line (Miyaoka *et al.*, 2014; Mandegar *et al.*, 2016). H9 cells were used with authorisation from the UK Stem Cell Bank Steering Committee.

Cells were cultured at 37°C in humidified incubators at either 5% CO_2; 5% O_2 or 5% CO_2; 21% O_2. Human PSCs were maintained under feeder-free conditions on plates coated with 5µg/ml vitronectin (ThermoFisher; A14700) in TeSR-E8 (StemCell Technologies; 05990) or mTeSR (StemCell Technologies; 85850) (Chen *et al.*, 2011) with daily media changes.

To passage the cells, plates were first coated with vitronectin in Dulbecco's PBS without calcium and magnesium (PBS) (Sigma-Aldrich; D8537) for at least 1 hour at room temperature. The solution was removed by aspiration and media was added to the plate before cell seeding. Cells were passaged every 5-10 days at a 1:8 - 1:20 ratio with 0.5mM EDTA (Sigma-Aldrich; D8537) in PBS with an incubation for 5 mins at room temperature. After replacement of GCDR with media, cells were manually dissociated from the culture vessels, collected and transferred to freshly coated vitronectin plates.

To cryopreserve cells, approximately $5x10^5$ primed hPSCs were suspended in freezing media comprising of complete TeSR-E8 or mTeSR1 medium containing 20% Knockout Serum (KSR) (Thermofisher Scientific; 10828028), 10µM Y-27632, a p160-Rho-associated coiled-coil kinase inhibitor (ROCKi) (Cell Guidance Systems; SM02) and 10% DMSO (Sigma-Aldrich; D2650), transferred to cryovials. Cells were frozen slowly using a Mr Frosty™ (Thermofisher Scientific) then transfer to liquid nitrogen for long term storage.

The UNC1999 (Cayman Chemicals; 14621) and GSK343 (Sigma-Aldrich; SML0766) were reconstituted in sterile DMSO (Sigma-Aldrich; D2650), to 10mM concentration. Once dissolved compounds were stored at -20°C for short term and -80°C for long term, avoiding multiple freeze thawing.

2.1.2 Pancreatic Differentiation

Pancreatic differentiation was undertaken based on the initial protocol by Cho et al, 2012, with changes as described in (Bertero *et al.*, 2016). All pancreatic differentiation experiments were performed in collaboration with Dr Filipa Soares and Dr Katarzyna Tilgner. The media compositions used according to Bertero et al, 2016 are described in Table 2.1. For the first 14 days of differentiation, media was changed on a daily basis, and at day 14 media was changed once every 72 hrs for the remainder of the differentiation protocol.

Table 2.1 Media composition used in pancreatic differentiation.

Media Type	Component	Final Concentration	Supplier
CDM-PVA	IMDM:F-12	-	Thermofisher Scientific; 1244053
	Chemically Defined Concentrated Lipids	1%	Gibco; 11905031
	1-thioglycerol	450mM	Sigma-Aldrich; M1753
	Transferrin	15µg/ml	Sigma-Aldrich; T3309
	Penicillin-Streptomycin	1%	Sigma-Aldrich; P4333
	Polyvinyl alcohol	1mg/ml	Sigma-Aldrich; 341584
Advanced-BSA	Advanced F12/DMEM	-	Thermofisher Scientific; 12634028
	BSA	5mg/ml	Sigma-Aldrich; A9418
	L-glutamine	20µM	Sigma-Aldrich; G8540
	Penicillin-Streptomycin	1%	Sigma-Aldrich; P4333
Day 0 → Day 1	CDM-PVA	-	

Media	Activin-A	100ng/ml	Sigma-Aldrich; SPR3003
	bFGF	80ng/ml	Sigma-Aldrich; GF446
	LY294002	10µM	Sigma-Aldrich; L9908
	BMP4	10ng/ml	Sigma-Aldrich; SPR3016
	CHIR 99021	3µM	Sigma-Aldrich; SML1046
Day 1 → Day 2 Media	CDM-PVA	-	
	Activin-A	100ng/ml	Sigma-Aldrich; SPR3003
	bFGF	80ng/ml	Sigma-Aldrich; GF446
	LY294002	10µM	Sigma-Aldrich; L9908
	BMP4	10ng/ml	Sigma-Aldrich; SPR3016
Day 2 → Day 3 Media	RPMI	-	Thermofisher Scientific; 21875042
	Activin-A	100ng/ml	Sigma-Aldrich; SPR3003
	bFGF	80ng/ml	Sigma-Aldrich; GF446
	B-27 Supplement	2%	Thermofisher Scientific; 17504044
Day 4 → Day 6 Media	Advanced-BSA	-	-
	Retinoic acid	3µM	Sigma-Aldrich; R2625
	FGF10	50ng/ml	Sigma-Aldrich; F8924
	Noggin	150ng/ml	Sigma-Aldrich; SPR4675
	SB-431542	10µM	Sigma-Aldrich; S4317
	Ascorbic acid	0.25mM	Sigma-Aldrich; A4544
Day 7 → Day 8 Media	Advanced BSA	-	-
	Retinoic acid	1µM	Sigma-Aldrich; R2625

	FGF10	50ng/ml	Sigma-Aldrich; F8924
	Noggin	150ng/ml	Sigma-Aldrich; SPR4675
	KAAD-cyclopamine	0.25µM	
	Ascorbic acid	0.25mM	Sigma-Aldrich; A4544
Day 9 → Day 13 Media	Advanced BSA	-	
	Retinoic acid	0.1µM	Sigma-Aldrich; R2625
	FGF10	50ng/ml	Sigma-Aldrich; F8924
	KAAD-cyclopamine	0.25µM	Sigma-Aldrich; 239804
	Ascorbic acid	0.25mM	Sigma-Aldrich; A4544
Day 14 → Day 19 Media	Advanced BSA	-	
	Retinoic acid	0.1µM	Sigma-Aldrich; R2625
	B-27 Supplement	1%	Thermofisher Scientific; 17504044
	DAPT	1µM	Sigma-Aldrich; D5942
	6-Bnz-cAMP*	0.1mM	Sigma-Aldrich; B4560
	Alk5i II	10µM	Cell Guidance System; SM09
	Glucose	25mM	Sigma-Aldrich; G8769
Day 20 → Day 27 Media	Advanced BSA	-	
	Retinoic acid	0.1µM	Sigma-Aldrich; R2625
	B-27 Supplement	1%	Thermofisher Scientific; 17504044
	KAAD-cyclopamine	0.25µM	Sigma-Aldrich; 239804
	Alk5i II	10µM	Cell Guidance System; SM09

*First 72 hrs only

2.1.3 Glucose Stimulated Insulin Secretion

Complete Hank's Balanced Salt Solution (HBSS) medium was prepared with the addition of the compounds listed in Table 2.2, and pH was adjusted to 7.4. To create a low (1.67mM) and high (16.7mM) glucose medium, glucose (Sigma-Alrich; G8644) was dissolved in complete HBSS medium.

Cells were washed with low glucose medium and then pre-incubated for 1 hour at 37°C in low glucose medium. To perform the Glucose Stimulated Insulin Secretion (GSIS) assay, the medium was removed from pre-incubated cells, thses were then washed with low glucose medium and incubated in fresh low glucose medium for 1 hour at 37°C. After stimulation, the cell medium was collected and frozen for storage at -80°C for later analyses. Cells were then washed as before with low glucose medium and then incubated with high glucose medium for 1 hour at 37°C. Cell medium was collected as before and frozen for storage -80°C. Stimulated cells were dissociated by treating with Accutase (ThermoFisher Scientific; A1110501) for 5 mins at 37°C. Cells were washed with PBS, pelleted by centrifugation and frozen at -80°C.

Table 2.2. Complete HBSS medium for glucose stimulation assay

Component	**Final Concentration**	**Supplier; Catalogue Number**
HBSS	-	Thermofisher Scientific; 14175046
HEPES	9mM	Sigma-Aldrich; H0887
Sodium Chloride	113.3mM	Sigma-Aldrich; 31434
Sodium phosphate dibasic	0.63mM	Sigma-Aldrich; S5136
Potassium Chloride	5.37mM	Sigma-Aldrich; P5405
Magnesium sulphate hepta-hydrate	0.81mM	Sigma-Aldrich; M5921
Potassium phosphate monobasic	0.44mM	Sigma-Aldrich; P5655
Sodium bicarbonate	4.17mM	Sigma-Aldrich; S5761
Calcium chloride dihydrate	1.5mM	Sigma-Aldrich; C7902
Bovine Serum Albumin	30mM	Sigma-Aldrich; A7906

2.2 Gene Expression Analysis

2.2.1 RNA Isolation and cDNA conversion

For Reverse Transcription Quantitative Real Time Polymerase Chain Reaction (RT-qPCR), total RNA was extracted using the RNeasy Mini kit (QIAGEN; 74106), following the manufacturer's instructions. A maximum of 1µg RNA was then converted to cDNA using a QuantiTect Reverse Transcription Kit (QIAGEN; 205311), according to manufacturer's instructions. The cDNA was measured using a Nanodrop 2000 and diluted to 50ng/µl.

RNA for sequencing was extracted using TRI Reagent (Sigma-Alrich; T9424). Cell pellets were resuspended in 1ml of TRI Reagent and RNA was isolated by adding 200µl chloroform and vortexing for 10 seconds. Samples were then incubated for 2 mins at room temperature, and centrifuged for 15 mins at 4°C at 12,000xg. The aqueous phase was moved to a new tube and precipitation with 500µl of 100% isopropanol was undertaken at -20°C for 1 hour or overnight, with addition of 50µg GlycoBlue Coprecipitant (ThermoFisher Scientific; AM9515) to improve yield. Samples were then centrifuged for 15 mins at 12,000xg at room temperature. The resulting RNA pellet was washed with 75% ethanol and air dried before resuspension in RNase-free water.

2.2.2 Reverse Transcription Quantitative Real Time Polymerase Chain Reaction

RT-qPCR was performed using JumpStart SYBR Green ReadyMix (Sigma-Aldrich; S4438) in the following reaction mixture: 6µl SYBR Green, 0.24µl 10mM forward primer, 0.24µl 10mM reverse primer, 5µl DNA at 50ng/µl concentration, 0.52µl water. RT-qPCR was performed on a Bio-Rad CFX96 RT-qPCR machine. Unless stated, all RT-qPCR experiments were always performed in technical triplicate. Results were analysed using the δ - δ cycle threshold with *GAPDH* used as a housekeeping gene.

Table 2.3. Primer sequences for RT-qPCR analysis

Gene	**Primer**	**Sequence**
CXCR4	FOR	CCCACCATCTACTCCATCATCT
	REV	TGTACTTGTCCGTCATGCTTCT

FOXA2	FOR	GTCCGACTGGAGCAGCTACTAT
	REV	ATGTACGTGTTCATGCCGTTC
GAPDH	FOR	CGCTGAGTACGTCGTGGAGT
	REV	GGGCAGAGATGATGACCCTTT
GATA4	FOR	TCCCTCTTCCCTCCTCAAAT
	REV	TCAGCGTGTAAAGGCATCTG
GCG	FOR	GAATTCATTGCTTGGCTGGTGA
	REV	TGATCTGGATTTCTCCTCTGTGTCT
HAND1	FOR	AAGCAAGCGGAAAAGGGAGT
	REV	GGTGCGCCCTTTAATCCTCT
INS	FOR	GAACGAGGCTTCTTCTACACAC
	REV	GTTCCACAATGCCACGCTTC
NANOG	FOR	TCCAGCAGATGCAAGAACTCTC
	REV	GGTTCTGGAACCAGGTCTTCAC
NEUROG3	FOR	GACTCAAACGCTGCGCATAG
	REV	GAGACTGGGGAGTAGAGGGA
NESTIN	FOR	CACCTCAAGATGTCCCTCAGC
	REV	GAAAGCTGAGGGAAGTCTTGGAG
NKX2.2	FOR	AACCCCTTCTACGACAGCAG
	REV	GTCTCCTTGTCATTGTCCGGT
PAX6	FOR	GGTTGGTATCCGGGGACTTC
	REV	CGTTGGAACTGATGGAGTTGGT
	FOR	GGATGAAGTCTACCAAAGCTCAC

PDX1	REV	GCCGTGAGATGTACTTGTTGAAT
POUF51 (OCT4)	FOR	GGATATACACAGGCCGATGTGG
	REV	ATGGTCGTTTGGCTGAATACCT
SOX2	FOR	AACCAGCGCATGGACAGTTAC
	REV	GTTCATGTAGGTCTGCGAGCTG
SOX17	FOR	CAGAATCCAGACCTGCACAAC
	REV	CTCTGCCTCCTCCACGAAG
SST	FOR	CTCCGTCAGTTTCTGCAGAA
	REV	CGGGTTTGAGTTAGCAGATCTCT
TBX3	FOR	ATCGTGTCAAGCTCACCAACA
	REV	GCTGCCCGAACTAGGTGTATG

2.2.3 mRNA-Sequencing Library Preparation

RNA-sequencing libraries were produced from 500ng of total RNA using the NEBNext Ultra RNA Library Prep Kit for Illumina (NEB; E7530), with the Poly(A) mRNA Magnetic Isolation Module (NEB; E7490), according to manufacturer's instructions. Library indexing was undertaken using NEBNext Multiplex Oligos for Illumina (Index Primers Set 1 and Set2) (NEB; E7335 and E7500) according to the Library Preparation Kit. Library concentration and fragment size was determined using an Agilent Bioanalyzer 2100 and KAPA Library Quantification Kit (KAPA Biosystems; K4828). Samples were sequenced at the Babraham Institute Sequencing Facility on an Illumina NextSeq500 instrument as 75bp single-end libraries.

2.2.4 RNA-Sequencing Analysis

Sequenced RNA reads were trimmed using trim galore v0.4.2 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) using default parameters to remove the Illumina adaptor sequences. Reads were mapped to the human GRCh38 genome

assembly using bowtie2 v2.3.4 and converted to BAM files that were imported to Seqmonk v1.45 (http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/) for downstream analysis.

Raw read counts were calculated using the RNA-sequencing quantitation pipeline on the Ensembl v70 gene set using directional counts. Initial quality control of the samples was undertaken using FastQC (Wingett and Andrews, 2018), further analysis was then undertaken to ensure enrichment of reads within exons, genomic DNA contamination was not present and read distribution was similar between samples. Three replicates were merged and differentially expressed genes were determined using the DESeq2 (Love et al., 2014) implementation in SeqMonk. Regularised log transformation was applied prior to visualization to correct for library size and variance among counts. Data was analysed in Seqmonk v1.45 and RStudio v1.2.5. Gene ontology (GO) analysis was carried out using enrichR (Kuleshov *et al.*, 2016). Principal Component Analysis was undertaken in RStudio using the top 2000 most variable genes between samples. The differentially expressed genes as calculated by DESeq2 were imported into iDEP for K-means clustering analysis (Ge, Son and Yao, 2018). Keygene analysis was undertaken using the Keygene we application using the fetal training set (Roost *et al.*, 2015).

Human pancreatic data was obtained from Fadista et al, 2014 (accession no; GSE50398). Donor sample numbers 3, 8, 14, 16, 23 and 24 were randomly selected to cover a range of donor sex and age. Mapped sequence files were imported into Seqmonk v1.45 as paired end 101bp reads (Fadista et al., 2014). Fetal samples of CS 16 -18 were obtained from Cebola et al, 2015 (accession no; E-MTAB-1990) and CS 13 data samples were obtained from Jennings et al, 2017 (accession no; E-MTAB-3259) (Cebola *et al.*, 2015; Jennings *et al.*, 2017).

2.3 Histone Modification Analysis

2.3.1 Chromatin Immunoprecipitation

During Chromatin Immunoprecipitation (ChIP), unless stated otherwise, all steps were carried out on ice or at 4°C, and pre-cooled buffers, in Table 2.4, were used. For histone ChIP, 500,000 cells were harvested using Accutase, washed with PBS and frozen at -80°C. Upon thawing, cells were washed with PBS and pelleted by centrifugation at 300xg for 5 mins. Pellets

were resuspended in 95µl Digestion Buffer with 1X Complete EDTA-free Protease Inhibitor (CPI) (Roche; 11873580001). To perform a spike-in normalised ChIP, 200,000 drosophila cells were added at this stage to each individual sample. The following protocol was used for both non-normalised and normalised ChIP experiments. Chromatin was digested using micrococcal nuclease (NEB; M0247) at 37°C; the time (6 – 9 mins) and concentration of enzyme (~200 U/µL) treatment was optimised for each cell type , to obtain primarily mononucleosomes. 10µl of Stop Buffer was added to inactivate the reaction, and chromatin was sonicated for 1 minute, high power on a Diagenode Bioruptor. Chromatin was diluted in 100µl RIPA-IP Buffer, pH adjusted and supplemented with 1X CPI, and centrifuged at 12,000xg at 4°C for 15 mins. Supernatant was transferred to a new tube and 10% was removed as an input sample. The remaining sample was precleared by addition of 50µl prewashed Protein A and G Dynabeads (Life Technologies; 10001D; 10003D), with rotation for 1 hour at 4°C. Dynabeads were removed and chromatin samples were diluted and split into two tubes, with 100µl in each tube. ChIP was performed overnight with rotation at 4°C using 1µg H3K27me3 (Cell Signalling Technology: 9733) and 0.5µg H3K4me3 (Abcam; ab8580). 10µl of prewashed A and G Dynabeads were added to samples and incubated for 4-5 hrs with rotation at 4°C. Beads were separated from supernatant using a magnet and washed 5 times with RIPA buffer supplemented with 1X CPI, with 5 mins incubation on rotation at 4°C between washes, and one wash using LiCl Buffer supplemented with 1X CPI. Beads on the magnet were also washed once with TE buffer. Beads were resuspended in 100µl of TE Buffer with 50µg Proteinase K and incubated at 55°C for 1 hour with shaking. ChIP and Input DNA was extracted using a Zymo Genomic DNA Clean and Concentrator kit, according to the manufacturer's instructions, and eluted in 30µl of TE Buffer. To quantify the DNA obtained from each ChIP, samples were measured on a Qubit using a dsDNA HS Assay Kit (ThermoFisher Scientific; Q33230).

Table 2.4 Buffers for chromatin immunoprecipitation experiments.

Buffer	**Component**	**Final Concentration**	**Supplier; Catalogue Number**
Digestion Buffer	Tris-Hydrochloride pH 8	50mM	ThermoFisher Scientific; 15568025
	Calcium Chloride	1mM	SigmaAldrich; C1016
	Triton X-100	0.2%	SigmaAldrich; T8787

	Sodium Butyrate	5mM	SigmaAldrich; B5887
Stop Buffer	Tris-Hydrochloride pH 8	110mM	ThermoFisher Scientific; 15568025
	EDTA	55mM	Invitrogen; 10135423
RIPA-IP Buffer	Sodium Chloride	280mM	SigmaAldrich; S5150
	Triton X-100	1.8%	SigmaAldrich; T8787
	SDS	0.2%	Biorad; 1610418
	Sodium Deoxycholate	0.2%	SigmaAldrich; D6750
	EGTA	5mM	SigmaAldrich; E3889
	Sodium Butyrate	5mM	SigmaAldrich; B5887
RIPA Buffer	Sodium Chloride	140mM	SigmaAldrich; S5150
	Tris-Hydrochloride pH 8	10mM	ThermoFisher Scientific; 15568025
	Triton X-100	1%	SigmaAldrich; T8787
	SDS	0.1%	Biorad; 1610418
	Sodium Deoxycholate	0.1%	SigmaAldrich; D6750
	EDTA	1mM	Invitrogen; 10135423
	Sodium Butyrate	5mM	SigmaAldrich; B5887
LiCl Buffer	Lithium Chloride	250mM	SigmaAldrich; L7026
	Tris-Hydrochloride pH 8	10mM	ThermoFisher Scientific; 15568025
	NP-40	0.5%	SigmaAldrich; 74385

	Sodium Deoxycholate	0.5%	SigmaAldrich; D6750
	EDTA	1mM	Invitrogen; 10135423
TE Buffer	Tris-Hydrochloride pH 8	10mM	ThermoFisher Scientific; 15568025
	EDTA	1mM	Invitrogen; 10135423

2.3.2 Chromatin Immunoprecipitation Library Preparation

To produce indexed ChIP-sequencing libraries, a NEBNext Ultra II DNA Library Prep Kit (NEB; E7645) was used according to manufacturer's protocol, and without size selection or pre-amplification. Indexing was undertaken using NEBNext Multiplex Oligos for Illumina Set 1 (NEB; E7335) and Set 2 (NEB; E7500). Agencourt AMPure XP beads (Beckman Coulter; A63881) were used for DNA clean-up and size selection. Library concentration and fragment size was determined using an Agilent Bioanalyzer 2100 and KAPA Library Quantification Kit (KAPA Biosystems; K4828). Samples were sequenced at the Babraham Institute Sequencing Facility on an Illumina NextSeq500 instrument as 75bp single-end libraries.

2.3.3 Chromatin Immunoprecipitation Data Analysis

Sequenced RNA reads were trimmed using trim galore v0.4.2 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) using default parameters to remove the Illumina adaptor sequences. Reads were mapped to the human GRCh38 genome assembly using bowtie2 v2.3.4, and converted to BAM files that were imported to Seqmonkv1.45 (http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/) with reads extended by 200bp.

Peak calling for ChIP-Sequencing data was undertaken by MACS peak calling within Seqmonk, using stage matched inputs with a p-value cut-off of 10^{-9}, FDR filtering was used to select peaks with a more stringent cutoff. To analyse histone marks at promoters, ChIP peaks were assigned to a window +/- 500bp centred on annotated TSSs with peaks then associated to the closest gene, to a maximum of 1kb. Wiggle plot genome tracks were created in Seqmonk with each window of analysis selected based on the gene length. Trend plots were created

using Seqmonk programme, with feature probe created +/- 2 kb around the TSSs. Publiclly available ChIP datasets were downloaded and imported into Seqmonk for analyses as described above for Bhandare et al, 2010 data sets (accession no; E-MTAB-191) (Bhandare *et al.*, 2010).

For spike-in ChIP analysis, reads were mapped to the human GRCh38 genome assembly, and Drosophila BDGP6 genome assembly, using bowtie2 v2.3.4. To calibrate using the spike-in genome, data was normalised as previously described in van Mierlo et al, 2019 and Fursova et al, 2019. For each sample normalisation against input, a downsampling factor was used. This factor was calculated using the formula below (Fursova *et al.*, 2019; Guido van Mierlo *et al.*, 2019):

$$\text{Downsampling factor} = \alpha \times \frac{1}{N(\text{spike-in in ChIP})} \times \frac{N(\text{spike-in in input})}{N(\text{human in input})}$$

To calculate the **α** coefficient, the largest drosophila to human ratio sample was set as 1, this was then used in all further calculations to determine the downsampling factor, where N equals the number of reads, as indicated in the equation. Downsampling was applied to all samples before analysis by using the downsampling factor as calculated to normalise the read count in each sample. The read count alignment numbers of both genomes, can be found in Appendix III.

2.3 Immunofluorescence Microscopy

Adherent cells were fixed by aspirating the culture media then immediately adding 500µl of 4% paraformaldehyde (VWR; 43368.9M) and incubating for 20 min at 4°C. Cells were washed three times with 500µl PBS. Cells were incubated with 500µl of PBS-T (0.1% Triton X-100 in PBS) containing 10% donkey serum (AbD Serotec; C06SB) to block nonspecific binding. Cells were incubated in blocking solution for 20 min at room temperature. Blocking solution was removed and cells were then incubated overnight at 4°C with 250µl of primary antibodies; glucagon (Santa-Cruz; SC-7780), somatostatin (Dako; A0566), c-peptide (Acris Anitbodies; BM270S) diluted in PBST containing 1% donkey serum, at 1:100, 1:200 and 1:300 respectively.

Cells were then washed 3 times with PBST to remove unbound antibodies, followed by incubation with 300 µl of fluorophore-conjugated secondary antibodies; donkey anti-rabbit Alexa Fluor 488 (Invitrogen; A-21206), donkey anti-mouse Alexa Fluor 647 (Invitrogen; A-31571), donkey anti-goat Alexa Fluor 555 (Invitrogen; A-21432) diluted 1:1000 in PBST containing 1% donkey serum. After an incubation for 1 hour at room temperature, unbound antibodies were removed by 3 washes in DPBS. After the final wash DAPI was added at a final dilution of 1 µg/ml and incubated for 10 min at room temperature. Cells were washed once more before being imaged on a Nikon AR-1 confocal microscope a the Babraham Institute Imaging Facility and processed using FIJI software.

2.4 Flow Cytometry

Cells were washed with PBS, followed by incubation with Accutase for 5 mins at 37°C to create single cell suspension. Cells were washed with PBS containing 5% FBS (flow buffer), and then passed through a 30µM cell strainer. DAPI was added to the cell suspension in flow buffer to the concentration of 0.2µg/ml (Sigma-Alrich; D9542). Data was obtained on a BDLSRFortessa at the Babraham Institute Flow Core and data was analysed on FlowJo v10.6.1 software.

2.5 Protein Analysis

2.5.1 Histone Extraction

To isolate histones for protein analysis, cells were washed with PBS and a single cell solution was generated by Accutase treatment for 5 mins at 37°C. Cells were washed with PBS and pellets were snap frozen in liquid nitrogen and stored at -80°C until required.

Unless specified, all steps were undertaken on ice or at 4°C to preserve protein samples, and all solutions were ice-cold before use. Frozen cell pellets were thawed at 37°C and washed with 1ml of room temperature PBS supplemented with 1xCPI. Cells were centrifuged for 5 mins at 300xg and the supernatant was removed. The pellet was resuspended in 1ml of PBS supplemented with 1xCPI and left on ice for 10 mins. Cells were centrifuged for 5

mins at 300xg and the supernatant was removed. The pellet was resuspended in 1ml of 0.2M sulphuric acid and left on ice for 30 mins. Samples were then centrifuged for 2 mins at 12,000xg at 4°C. The supernatant was transferred to a fresh tube and 333µl of trichloroacetic acid (SigmaAldrich; T0699) was added. Samples were left on ice for 30 mins, and centrifuged for 10 mins at 12,000 xg at 4°C. Pellets were washed with 1ml of acetone and left on ice for 10 mins, and samples were then centrifuged for 10 mins at 12,000xg at 4°C. The acetone wash was repeated, and all supernatant fully removed after centrifugation. Histone proteins were then resuspended in 100mM Tris-HCl (ph8) supplemented with 1xCPI and left to dissolve overnight at 4°C with rotation. Dissolved samples were then centrifuged for 10 mins at 12,000xg at 4°C, and the supernatant was transferred to a fresh tube and frozen at -80°C until required.

2.5.2 Bicinchoninic acid assay

To determine the protein concentration of extracted samples, the Pierce BCA Protein Assay Kit (ThermoFisher Scientific; 23227) was used. Standards for measurement were prepared using the BSA provided, ranging from 25µg - 2000µg/ml, and samples to be measured were diluted at 3 different dilutions, to bring the measurements within the range of the standards. The working reagent was prepared by mixing 50 parts BCA reagent A with 1 part BCA reagent B, and 200µl was added to 25µl of standard or sample in a 96 well microplate. The plate was covered and left to incubate at 37°C for 30 mins with shaking. The reading for each standard and sample were then taken using a PHERAstar FS Plate reader, and sample concentrations calculated using a standard curve produced.

2.5.3 Western Blot

To extract whole cell proteins for western blot, approximately 1 million cells were resuspended in 50µl RIPA buffer (1% NP-40, 0.5% Sodium Deoxycholate, 15mM NaCl, 50mM Tris pH8, 0.1% SDS, 1xCPI). Cells were agitated at 4°C for a minimum of 30 mins before centrifugation at 12,000xg for 15 mins at 4°C. The supernatant was transferred to remove cell debris and quantified by BCA assay, and the remaining protein was frozen at -80°C until required. After thawing, samples were incubated with 5x loading dye (10% SDS, 500mM Tris-HCl pH6.8, 250mM B-mercaptoethanol, 30% Glycerol, 0.5% bromophenol blue dye) and boiled for 4 mins at 90°C. Proteins were separated by electrophoresis in SDS-polyacrylamide gels (10-15%) and transferred to PVDF membrane (Millipore; IPVH00010). Protein membranes were blocked in 1X Tris-Buffered Saline (TBS) with 0.1% Tween (TBS-T) and 5% milk for 2-3 hrs at

room temperature with rotation. Membranes were transferred to 5% milk TBS-T for hybridisation overnight at 4°C with the following primary antibodies; β-actin at 1:10,000 (Sigma-Aldrich; A5441), H3K27me3 at 1:1,000 (Millipore; 07-449), H2B at 1:5,000 (Active Motif; 39210). Membranes were washed 3 times for 10 mins in TBS-T with 5% milk, then incubated at room temperature for 1 hour with the following secondary antibodies: donkey anti-rabbit immunoglobulin DyLight 800 (Invitrogen; SA5-10044) at 1:10000 and donkey anti-mouse immunoglobulin DyLight 680 (Invitrogen; SA5-10170) at 1:10000. Detection was performed using a Licor Odyssey® CLx instrument.

2.5.4 C-peptide ELISA

To measure the C-peptide release from cells, a C-peptide ELISA kit (Mercodia; 10-1136-01) was used according to manufacturer's instructions. In advance of starting, samples obtained from glucose stimulation (Section 2.1.3) were defrosted and warmed to room temperature. Enzyme conjugate and wash buffer were prepared to a 1x solution. As a negative control, freshly prepared HBSS media was used. 25µl of calibrator, control or sample was added to each well, followed by 50µl of assay buffer. The plate was incubated on a shaker for 1 hour at room temperature. After incubation, wells were washed by discarding liquid, adding 350µl wash buffer to each well, for a total of 6 washes. 100µl of 1x enzyme conjugate was added to the washed wells and the plate was incubated on a shaker for 1 hour at room temperature. After incubation, the wells were washed as above for a total of 6 washes, and 200µl of substrate TMB was added. The plate was incubated for 15 mins at room temperature, after which 50µl of stop buffer was added, and the plate was briefly placed on to a shaker to ensure mixing. The plate was read using a Microplate Absorbance Spectrophotometer and C-peptide concentration of samples was calculated using the calibration curve samples provided.

2.6 Molecular Cloning

All molecular cloning was carried out using standard cloning techniques with restriction enzymes obtained from NEB or ThermoFisher and T4 DNA ligase from Promega. Gibson cloning was undertaken using the Gibson Assembly Master Mix (NEB; E2611) using primers designed using the NEBuilder® Assembly Tool (https://nebuilder.neb.com). If possible, for later applications, plasmids were dephosphorylated using Calf Intestinal Alkaline Phosphatase (ThermoFisher; 18009019). All oligonucleotides were designed using Primer3

(http://primer3.ut.ee) and were ordered from Sigma-Aldrich. To extract DNA fragments from agarose gels the Monarch DNA Gel Extraction Kit (NEB; T1020) was used, and the Monarch PCR and DNA Cleanup Kit (NEB; T1030) was used to purify DNA PCR products. To purify plasmids, QIAGEN Plasmid Mini Kit (Qiagen; 12125), Midi Kit (Qiagen; 12145) or Maxi Kit (Qiagen; 12162) were used, depending on the concentration and volume required. All plasmids were subcloned using Subcloning Efficiency DH5α Competent Cells (Invitrogen; 18265017) and bacterial stocks were produced with 25% glycerol for storage at -80°C.

2.6.1 Plasmid Production

Plasmid maps for all plasmids produced during this project can be found in the Appendix I, and details on how specific plasmids were created is described below.

Table 2.5. Reagents used to create plasmids.

Plasmid Name	Source	Used to produce
AAV-CAGGS - EGFP	A gift from Rudolf Jaenisch (Addgene plasmid #22212)	AAV _ OsTIR1 _ Puro
CMV-OsTIR1-PURO	A gift from Masato Kanemaki (Addgene plasmid #72834)	AAV _ OsTIR1 _ Puro
mAID-Hygro	A gift from Masato Kanemaki (Addgene plasmid #72825)	EED _ mAID _ Hygro
mAID-Bsr	A gift from Masato Kanemaki (Addgene plasmid #72826)	EED _ mAID _ Bsr
mAID-mCherry2-Hygro	A gift from Masato Kanemaki (Addgene plasmid #72831)	EED _mAID _ mCherry2 _Hygro
mAID-mCherry-Bsr	A gift from Masato Kanemaki (Addgene plasmid #72832)	EED _mAID _ mCherry2 _Bsr
pgRNA-CKB	A gift from Bruce Conklin (Addgene plasmid #73501)	EED1_ CRISPRi EED6_ CRISPRi EED7_ CRISPRi EED9_ CRISPRi
pCR2.1-TOPO	Invitrogen; 451641	EED_gRNA_1 EED_gRNA_2
pGEM-T Easy	Promega; A1360	EED_mAID_Hygro EED_mAID_Bsr EED_mAID_mCherry2_Hygro

		EED_mAID_mCherry2_Bsr

2.6.1.1 AAV_CAG_OsTIR1 Plasmid

The CMV-OsTIR1-Puro plasmid (gifted from Masato Kanemaki, see table 2.5) was cleaved with MluI (Thermofisher; FD0564) and XbaI (Thermofisher; FD0684) restriction enzymes. The cleaved plasmid was run on an agarose gel to separate the OsTIR1 sequence from the plasmid backbone; this fragment was isolated using the Monarch DNA Gel Extraction Kit (NEB; T1020), as per manufacturer's instruction. The AAV-CAGGS-EGFP was also cleaved with MluI (Thermofisher; FD0564) and XbaI (Thermofisher; FD0684) restriction enzymes and dephosphorylated using Calf Intestinal Alkaline Phosphatase (ThermoFisher; 18009019). The OsTIR1 insert was ligated with the linearised AAV-CAGGS Vector with T4 DNA ligase from (Promega; M1801) according to manufacturer's instructions with an overnight ligation. The ligated plasmid was transformed into DH5α Competent Cells (Invitrogen; 18265017), as per manufacturer's instructions, and plated overnight on Luria broth agar plates containing 100 µg/ml ampicillin. Single colonies were grown overnight in Luria broth supplemented with 100 µg/ml ampicillin before isolation with QIAGEN Plasmid Mini Kit (Qiagen; 12125) according to manufacturer's instructions. Screening of colonies was undertaken using XbaI and MluI to determine those with OsTIR1 and plasmid sequences were confirmed by Sanger sequencing by Genewiz, Inc. The resulting plasmid map is shown in appendix 1.

2.6.1.2 EED_mAID_Hygro/Bsr and EED_mAID_mCherry2_Hygro/Bsr

To create the homology arms required for the EED targeting vectors, a DNA template was ordered from Integrated DNA Technologies, Inc. This template contained a 491 bp region complementary to upstream of the gRNA target site and a 502 bp region complementary to downstream the target site, with a 6 bp BamHI recognition site between the regions (genome maps available Appendix I). The DNA template was integrated into a pGEM® T-Easy Vector (Promega; A1360) as per the manufacturer's instructions. The resulting plasmid was transformed into DH5α Competent Cells (Invitrogen; 18265017), as per manufacturer's instructions, and plated overnight on Luria broth agar plates containing 100 µg/ml ampicillin. Single colonies were grown overnight in Luria broth supplemented with 100 µg/ml ampicillin before isolation with QIAGEN Plasmid Mini Kit (Qiagen; 12125) according to manufacturer's instructions and the plasmid sequence was confirmed by Sanger sequencing by Genewiz, Inc. Once the correct sequence was confirmed, a larger stock of the plasmid was produced by

further bacterial overnight growth in Luria broth supplemented with 100 µg/ml ampicillin with plasmid isolation using QIAGEN Maxi Kit (Qiagen; 12162).

The inserts required from mAID-Bsr, mAID-Hygro, mAID-mCherry-Bsr and mAID-mCherry-Hygro were removed from the original plasmids (gifted from Masato Kanemaki, see table 2.5) by cleavage with BamHI restriction enzyme (Thermofisher; FD0054), and the mAID-mCherry2-Bsr was further cleaved with SsiI restriction enzyme (ThermoFisher; FD1794) and the mAID-Hygro was further cleaved with Cfr42I restriction enzyme (ThermoFisher; ER0201) to cut the plasmid backbone. The cleaved plasmid was run on an agarose gel to resolve the mAID sequence insert from the plasmid backbone; these fragments were isolated using the Monarch DNA Gel Extraction Kit, as per manufacturer's instruction. The pGEM T-easy vector with the EED insert was also cleaved with BamHI (Thermofisher; FD0054) and dephosphorylated using Calf Intestinal Alkaline Phosphatase. The linearised plasmid was extracted from the digest reaction using Monarch PCR and DNA Cleanup (NEB; T1030). The four individual mAID containing inserts were ligated with the linearised EED pGEM T-Easy Vector with T4 DNA ligase (Promega; M1801), according to manufacturer's instruction, with an overnight ligation. The resulting plasmid was transformed into DH5α Competent Cells (Invitrogen; 18265017), as per manufacturer's instructions, and plated overnight on Luria broth agar plates containing 100 µg/ml ampicillin. Single colonies were grown overnight in Luria broth supplemented with 100 µg/ml ampicillin before isolation with QIAGEN Plasmid Mini Kit (Qiagen; 12125) according to manufacturer's instructions and the plasmid sequence was confirmed by Sanger sequencing by Genewiz, Inc. The resulting plasmid maps from all four plasmids formed can be found in appendix I.

2.6.2.3 Guide RNA Plasmids

The guide RNAs (gRNAs) for AID targeting (table 2.6) were designed using http://crispr.mit.edu (Hsu *et al.*, 2013). gRNA sequences were incorporated into a U6 target gRNA expression vector and were synthesized through Integrated DNA Technologies as a Gblock Gene Fragment. The Gblock was sub-cloned into the pGEM-T Easy vector (Promega; A1360), according to manufacturer's instructions.

The gRNA plasmids were cloned into Subcloning Efficiency DH5α Competent Cells (Invitrogen; 18265017). For this, 50µl of competent bacteria were defrosted on ice and 1µl of plasmid mix was added to the bacterial solution. This mix was left for 30 mins on ice and then heat-shocked at 42°C for 30 seconds. Heat-shocked bacteria was left on ice for 2 mins, and

200µl of S.O.C media (Thermofisher; 15544034) was added and bacteria was incubated at 37°C for 1 hour with shaking. LB agar plates with ampicillin (100µg/ml) plates were coated with X-gal to a concentration of 20µg/ml and dried, then bacteria were spread and left to grow overnight on plates at 37°C. The following day white colonies were picked and inoculated into 3ml of LB broth with ampicillin (100µg/ml) for overnight growth. Plasmids were isolated using QIAprep Spin Miniprep Kit (Qiagen; 27106) according to manufacturer's instructions. Plasmid sequences were confirmed by sequencing.

To produce gRNAs for use in CRISPRi, sequences were selected from Gilbert et al, 2014 and 4 gRNAs that cover the broadest region close to the TSS sequence were selected, table 2.6. Selected gRNA sequences in sense and antisense orientations were ordered as DNA oligos from SigmaAldrich with the addition of a 4 nucleotide overhang “TTGG” to the 5 prime end of the forward primer and a 4 nucleotide overhang of “AAAC” was added to the 5 prime end of the reverse primer. The complementary DNA oligos were annealed and the pgRNA-CKB vector (Addgene; 73501) was linearised using BsmBI (NEB; R0580) digestion. The gRNA oligos were ligated into the linearised vector using T4 DNA Ligase (Invitrogen; 15224017) according to manufacturer's protocol, with overnight ligation at room temperature. The resulting ligated plasmid was transformed into bacteria and gRNA sequence was confirmed as described above, without bacterial blue/white screening.

Table2.6. gRNA sequences used in gene targeting systems

Target Gene		gRNA Target Sequence	System
EED	EED_gRNA_1	GTCGATCCCAGCGCCAAATAC	AID
	EED_gRNA_2	GTGATGCCAGTATTTGGCGCT	AID
	EED1_CRISPRi	TTGGGGGCTACACCACCAATGTCCC	CRISPRi
	EED6_CRISPRi	TTGGGCTGAAACGTCTTTGGAAGG	CRISPRi
	EED7_CRISPRi	TTGGGGGAGCGAAAGTCTGTGGAA	CRISPRi
	EED9_CRISPRi	TTGGGCATGCACAGCTAAAGCAAG	CRISPRi

2.7 Gene Targeting

2.7.1 Nucleofection

The Human Stem Cell Nucleofector kit (Lonza; VPH-5002) was used for all nucleofections. Cells were pretreated with 10μM Y-27632 (Cell Guidance Systems; SM02-1) for a minimum of 2 hrs before experiments, to increase cell survival. Plated cells were washed with PBS and colonies dissociated into single cells using Accutase (Gibco; A1110501) for 5 mins at 37°C. Cells were washed twice using PBS and cell number was determined using a haemocytometer. Two million cells were resuspended in 100μl of nucleofection reagent and with 10μg of plasmid DNA. Cells were nucleofected using the Amaxa Nucleofector 2b Device on setting B-016, and immediately after nucleofection the cells were plated onto vitronectin-coated plates in mTesR medium supplemented with 10μM Y27632 for the first 24 hrs growth. After the first media change cells were maintained in standard mTeSR medium, until 48 hrs post-nucleofection, after which they were analysed by flow cytometry or antibiotic treatment was undertaken. After the designated time of selection (3-9 days, antibiotic dependant), individual clones were allowed to expand for 5-10 days. Single colonies were picked into 96 well plates and expanded.

2.7.2 Transfection

Transfection was undertaken using Genejuice Transfection Reagent (Sigma-Alridch; 70967), according to manufacturer's instructions. On the day before transfection, cell colonies were dissociated into single cells using Accutase for 5 mins at 37°C. Cells were washed in media and plated at 1-3 x 10^5 per 6 well plate (or at an equivalent density in other formats) in mTeSR with 10μM Y-27632. After 12-24 hrs, 3-6μl Genejuice transfection reagent was added to 100μl of serum-free medium, and vortexed to mix. After 5 mins incubation at room temperature, a total of 1-3μg plasmid DNA was added to the serum-free medium and the solution was pipetted 6 times to mix, then incubated for a further 15 mins at room temperature. The media was changed to fresh mTeSR with 10μM Y-27632, and the transfection reagent was added dropwise to the medium. After 24 hrs, cells were transferred to standard mTeSR medium, and flow cytometry analyses or antibiotic treatment was undertaken at 48 hrs post-transfection. After the designated time of selection (3-9 days, antibiotic dependant), individual clones were allowed to expand for 5-10 days. Single colonies were picked into 96 well plates and expanded.

2.7.3 Polymerase Chain Reaction (PCR) Genotyping

The DNA to be used for PCR genotyping was extracted using either a Monarch Genomic DNA Purification Kit (NEB; T3010) or using yolk sac lysis buffer (0.5M KCl, 1M Tris pH8, 1M $MgCl_2$, 10% NP-40, 10% Tween20, plus 1.5µg of Proteinase K freshly added). For extraction using yolk sac lysis buffer, cells from a minimum of 1 well of a 96 well plate were suspended in the buffer. The cells were lysed overnight at 55°C followed by 95°C for 15 mins to inactivate the enzyme. PCRs were performed using 50-100ng DNA from the genomic extraction kit or using 5µl of yolk sac lysed DNA, in a 25µl reaction using either Q5 High-Fidelity DNA Polymerase (NEB; M0491) or HotStarTaq Polymerase (Qiagen; 203203). Product amplification was confirmed using agarose gel electrophoresis.

Table 2.7. Primer sequences used in genotyping.

Primer Target	**Direction**	**Primer Sequence**	**Enzyme Used**
Puromycin Transgene	FOR	GAGCTGCAAGAACTCTTCCTCA	Hotstar
	REV	GCTCGTAGAAGGGGAGGTTG	
EED Wild Type Gene	FOR	CCTCCAATGCTTTAGAACCTGG	Q5 HiFi
	REV	GTGTGAGACAGTTAGAACAGT	
EED mAID Transgene	FOR	CCTCCAATGCTTTAGAACCTGG	Q5 HiFi
	REV	AGGTATCAATGGACGGAGCA	
TIR1 Transgene	FOR	CGTGGACACTCTGGCTAAGA	Q5 HiFi
	REV	GCAATGTGACTCTGAAGGGC	
AAVS1 External (Right)	FOR	TTACCGCCTTTGAGTGAGCT	Q5 HiFi
	REV	AGATGACCGAGTACAAGCCC	
AAVS1 External (Backbone)	FOR	GCCTACATACCTCGCTCTGC	Hotstar
	REV	GCCTACATACCTCGCTCTGC	

EED Hygromycin Transgene	FOR	CCTCCAATGCTTTAGAACCTGG	Q5 HiFi
	REV	ATAGGTCAGGCTCTCGCTGA	

Chapter 3

Transcriptional and epigenetic dynamics of in vitro-derived human pancreatic cell development

3.1 Background

The ability to generate hormone producing pancreatic endocrine cells from hPSCs (D'Amour *et al.*, 2006; Jiang *et al.*, 2007; Shim *et al.*, 2007; Rezania *et al.*, 2014b; Russ *et al.*, 2015b), is one of the most well-defined *in vitro* differentiation protocols currently available. The *in vitro* pancreatic differentiation process involves the stepwise addition of signalling molecules to induce successive stages of cell populations, based on the formation of cell types that would normally be formed during the development of pancreatic tissue *in vivo*, ending with production of hormone secreting pancreatic endocrine cells. Through these progressive stages the cells transition from a broad pluripotent state, capable of producing multiple lineages, gradually becoming more restricted and eventually producing a specialised pancreatic endocrine cell type. Since the publication of initial differentiation protocols (D'Amour *et al.*, 2006), many studies have used these experimental conditions to further study the end cell types that are produced and have also continued to add complexity and improvements to the differentiation protocol (Shahjalal., 2018). However, to date, limitations still exist in all protocol variations to varying degrees, including low cell yield, polyhormonal expression, limited glucose response and a partially incorrect transcriptional signature (Kroon, Martinson, *et al.*, 2008; Schulz *et al.*, 2012; Pagliuca *et al.*, 2014). To determine how the cells derived by *in vitro* differentiation differ from their *in vivo* counterparts, and potential ways to improve the pancreatic endocrine cells formed *in vitro*, this study focuses on characterising the cells formed throughout *in vitro* pancreatic differentiation.

An important goal of the *in vitro* differentiation protocol is to produce cells that imitate the pancreatic endocrine cells normally found within the human body. The pancreatic tissue of a

human adult has a number of functions, including both exocrine and endocrine responses. The endocrine functions are carried out by the cells that reside within a defined structure within the pancreas, termed the islets of Langerhans. These islets are made up primarily of three cell types, α cells, β cells and δ cells, capable of secreting the hormones, glucagon, insulin and somatostatin, respectively (Murtaugh and Melton, 2003; Jennings *et al.*, 2013). Producing endocrine pancreatic cell types, representative of human islets, is a major aim for the regenerative medicine field, as *in vivo* derived human islets are in scarcity for applications such as toxicology testing and transplantation. Therefore, establishing a differentiation protocol capable of producing cells highly representative of those found in adult tissue is fundamentally required for pancreatic regenerative medicine research.

A specific area of *in vitro* differentiation that may benefit from further study is the role played by epigenetics during the formation of target cell types. Most differentiation protocols take advantage of iPSCs, which are seen to have virtually the same epigenome as hESC, although there is some indication of an 'epigenetic memory' in the iPSC derived cells (Vaskova., 2013), although it is thought that this 'epigenetic memory', that is reflective of the originator cell type used for cell reprogramming, may be lost after prolonged passaging (Polo *et al.*, 2010). As the epigenetics within a cell are important for directing differentiation, maintaining a cells identity and responding to physiological conditions (Mohn and Schübeler, 2009; Chen and Dent, 2013; Dambacher, De Almeida and Schotta, 2013), it would be hypothesised that during *in vitro* differentiation, the developing cell types should gradual obtain an epigenome which resemble the target cell type, as opposed to retaining the epigenome of hPSCs. One of the most common, and well-described epigenetic modifications observed, is the methylation of histone proteins, specifically tri-methylation of histone lysine 27 (H3K27me3) and histone lysine 4 (H3K4me3), which when found in combination, are typically associated with pluripotent cell types (Harikumar and Meshorer, 2015). Studying how these modifications change throughout the pancreatic differentiation protocol, in association with transcriptional changes, may further the knowledge of the *in vitro* differentiation process.

3.2 Hypothesis and Aims.

As previously introduced, the ability to differentiate human pluripotent cells into functional cell types is a key aim in the regenerative medicine field. In the case of some specific cell types, such as pancreatic endocrine cells and neuronal cell types, the field is tantalisingly close to utilising this technology for applications in human health (Senior and Pettus, 2019). However, the cells produced *in vitro* continue to possess a number of caveats, and therefore the use of these cells as a direct replacement to *in vivo* derived cells remains unattainable. To address this, further novel approaches are required to study and mature current *in vitro* derived cells.

In this chapter, the epigenetic and transcriptional interplay was investigated as hPSCs differentiate through multiple intermediate stages, eventually producing mature pancreatic endocrine cells *in vitro*, with a focus on the repressive histone modification H3K27me3 and the activation mark, H3K4me3. It was hypothesised that the epigenetic signature of these cells may contribute to differences that exist between the *in vitro* and *in vivo* pancreatic endocrine cells.

To investigate this hypothesis this chapter has three aims:

1. Confirm the production of pancreatic endocrine-like cells from the *in vitro* differentiation of human iPSCs.

2. Study the transcriptional dynamics, and associate this with each histone modification state in progressive cell populations, to determine epigenetics transition in the developing cells.

3. Undertake an in-depth comparison analysis between, *in vitro* derived cells compared to *in vivo* derived pancreatic cell types, at both the transcriptional and histone modification data, to indicate strengths of the current differentiation and potential areas for improvement.

3.3 Results

3.3.1 Confirming the production of pancreatic-like cells through isolation of intermediates.

To study the transcriptional and epigenetic dynamics during *in vitro* pancreatic differentiation, hPSCs were differentiated over a 27-day period. This was achieved according to the protocol described in section 2.1.2, which was based on previously published protocols where pancreatic endocrine-like cells were successfully produced *in vitro* (Kroon, Martinson, *et al.*, 2008; Schulz *et al.*, 2012; Pagliuca *et al.*, 2014a; Russ *et al.*, 2015; Bertero *et al.*, 2016). When initially developing differentiation protocols, researchers utilised human and mouse developmental data to determine signalling molecules important in the formation of successive cell stages which would give rise to pancreatic lineage restricted cell types. For this study, I isolated samples at multiple stages, which are representative of the predicted cell populations, seen in the developmental data, that would form throughout the differentiation protocol. At these selected timepoints, capturing several progressive cell states, samples were collected for RNA and histone modifications analysis from cell populations which are undergoing continuous differentiation. These samples were then processed to generate RNA-sequencing and H3K4me3 and H3K27me3 ChIP-sequencing data for each of these stages, through replicates obtained from independent differentiations (Figure 3.1). Read counts for all sequencing libraries are shown in appendix II.

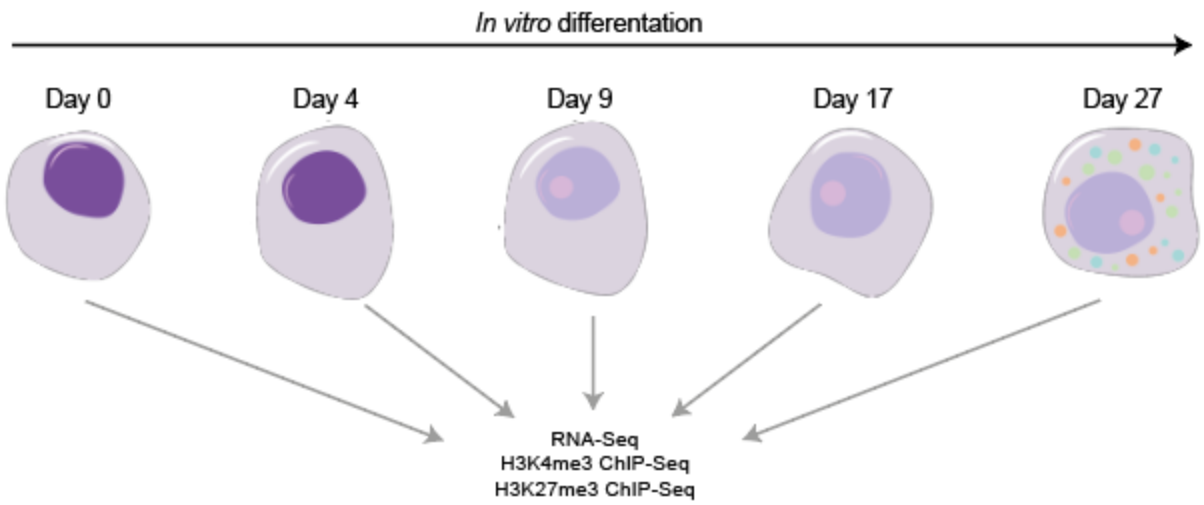


Figure 3.1 Experimental schematic to study transcriptional and epigenetic dynamics during *in vitro* differentiation. RNA and chromatin were isolated from samples during a continuous differentiation at the time points indicated, from three separate differentiations.

The differentiation protocol used in this study was elongated and had added small molecule addition, according to the DefiniGEN OptiDIFF pancreatic differentiation system, compared to the first published protocols. Therefore, the first step was to confirm the presence of expected cell types within the populations at the isolated time points. Using previously described markers (D'Amour *et al.*, 2005; Cho *et al.*, 2012; Jennings *et al.*, 2013), that can indicate the presence of specific cell types, all five isolated time points were analysed for the expected cell populations using RNA-sequencing expression data (Figure 3.2). High expression of pluripotency (*OCT4, NANOG and SOX2*) and endoderm (*SOX17, CXCR4 and GATA4*) markers confirmed undifferentiated iPSCs are present at day0 and definitive endoderm cell types at day4 of differentiation (Figure 3.2). Although there are no previously well described pancreatic endoderm markers, specific to only this cell type, the cells isolated at day9 had reduced expression of *OCT4* and *NANOG* but retained moderate expression of endoderm markers. This analysis also revealed expression of *PDX1*, a pancreatic progenitor marker, which together with endoderm marker expression is suggestive of a pancreatic endoderm stage. By day17, all three pancreatic progenitor markers examined (*NEUROG3, NKX2.2 and PDX1*) were clearly expressed, with some expression of pancreatic hormone genes, demonstrating the presence of pancreatic progenitor cell type within this population. By day27 the strong induction of pancreatic hormone genes and the small decrease in progenitor expression, particularly in

NEUROG3 expression, indicates the maturation into more endocrine pancreatic-like cells (Figure 3.2).

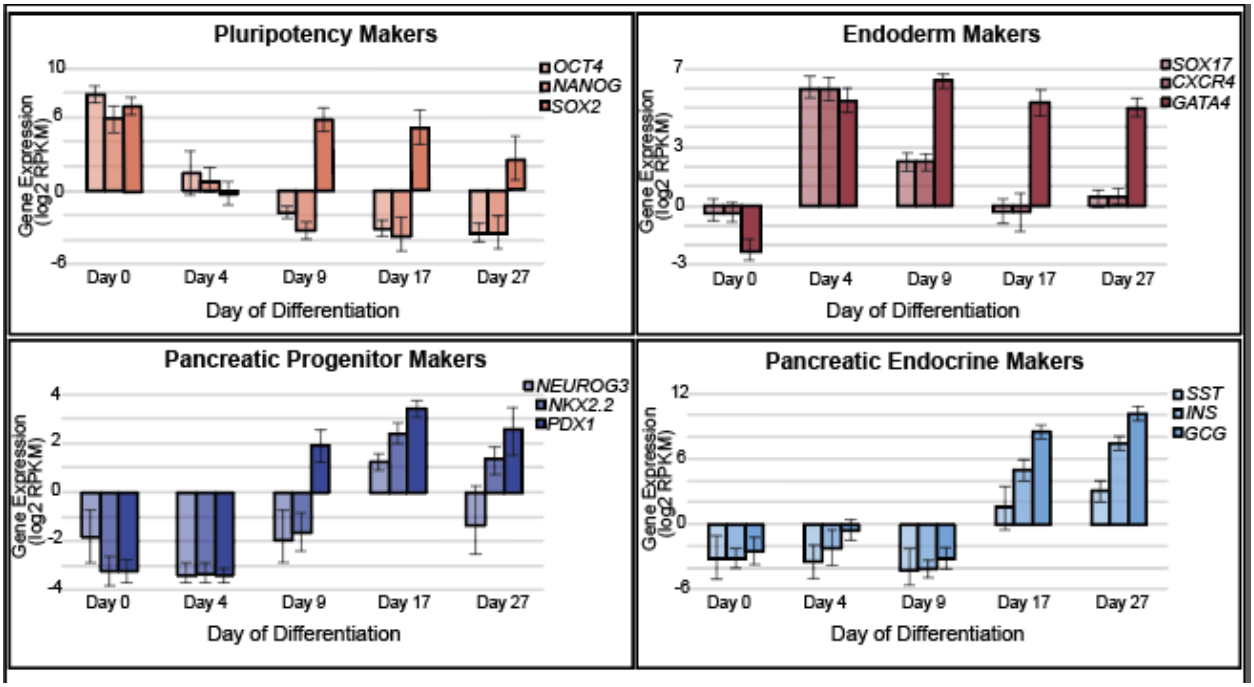


Figure 3.2 Expression of known cell markers determines the cell types present in the differentiating population. RNA-sequencing analysis of cell type specific marker genes at 5 different time points (day0, day4, day9, day17, day27) shown. The mean of 3 biological replicates is shown. Error bars indicate the standard deviation.

To study the formation of the progressive cell types it is important to confirm the isolated time points are transcriptionally distinct, beyond the markers previously known, therefore an analysis of the top 2000 most variably expressed genes throughout the *in vitro* differentiation protocol was undertaken. The transcription of the three replicates for each of five isolated populations, was analysed by a principal component analysis (PCA), which could be used to indicate the cell populations showing the highest similarities and differences between the cell types (Figure 3.3). This revealed a level of variation between individual replicates at the different time points; earlier differentiation time points had little transcriptionally variability, lying close on the PCA plot, and later stages had more variability on the replicate level (Figure 3.3a). Some

variability between replicates may be expected at later stages, due to the presence of multiple cell types expected to be formed by the differentiation, as cells progress to produce human islet–like hormone secreting cells, which is a heterogeneous population (Brissova *et al.*, 2005). However, this also indicates a known problem with *in vitro* differentiation protocols, in terms of the variability that occurs between differentiation replicates. When analysing the different time points, it was observed that the iPSC and DE cell types cluster closely on principal component 1. The pancreatic lineage cells however cluster away from these earlier timepoints on the principal component 1, which accounts for 56% of the variability. The three later pancreatic cell types then have a close but distinctive spread along both principal components, suggesting there may be a distinctive change in transcription between day4 and day9 populations. Analysis of the principal component loading (Figure 3.3b) supports the gradual formation of pancreatic-like cells through the five stages, with enrichment of pluripotency genes, such as *NANOG* and *OCT4* at day0, and pancreatic markers, such as *INS* and *SST* contributing to the separation of cell types at the three later time points. The isolated time points therefore have distinctive transcriptional signatures which can separate the earlier multipotent cells and the later lineage constrained pancreatic cells.

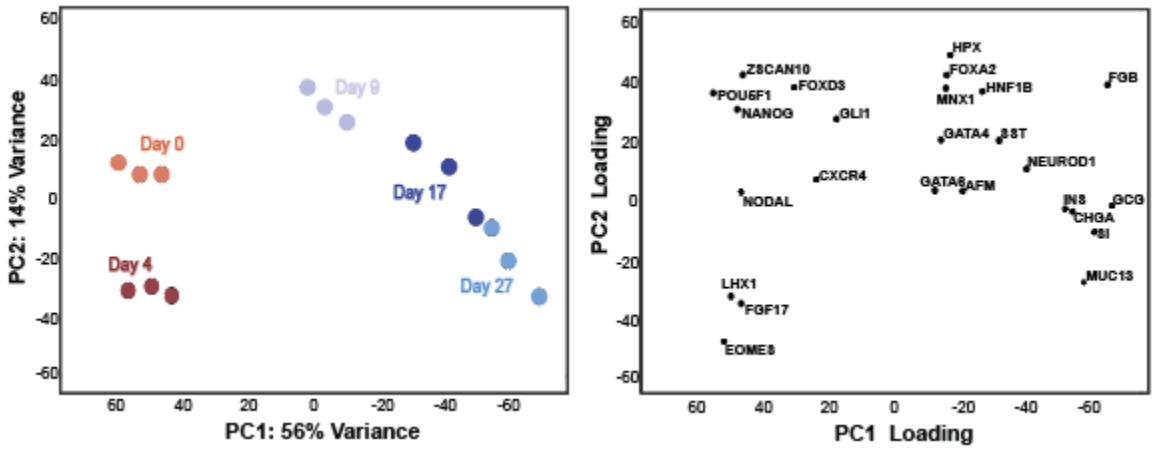


Figure 3.3 Comparative transcriptional analysis of *in vitro* differentiated cells. Principal component analysis plot for RNA-sequencing data of the top 2000 most variable genes as determined by DESeq2 between the five isolated time points of *in vitro* pancreatic differentiation (left). The contribution of the most variable genes plus selected differentiation marker genes to the first two principle components (right).

Although it is clear that the cells are progressing in their transcriptional signature, and the upregulation of at least some pancreatic markers is occurring, a gene signature analysis was undertaken to ensure differentiation is specific to the formation of pancreatic-like cells, and not due to spontaneous differentiation from pluripotent cell type. To that end, publicly available RNA-sequencing *in vivo* data of three different mature cell types, was obtained from the human tissue atlas (Uhlen et al., 2015., Human Protein Atlas avaliable from http://www.proteinatlas.org), and the expression of the genes associated with these cell types was then analysed in the *in vitro* differentiating cells (Figure 3.4). The comparison to the pancreatic signature was compared along with both skin and kidney transcriptional signatures, as this demonstrates the signature associated with a cell type of a similar lineage, in the kidney cells, and one of a more distant lineage cell, in skin cells. These two cell types were specifically selected due to a similar number of enriched genes in their transcriptional signature. The transcriptional signature of the cells was defined as genes which are expressed at levels at least four-fold higher expression in the isolated tissue, as compared to any other tissue. Analysis of these pancreatic enriched genes demonstrated an expression gain throughout the *in vitro* differentiation. The strongest increase of pancreatic-associated genes was observed between day4 and day9 when cells progress from a broader endoderm progenitor towards a more specialised pancreatic endoderm cell. In contrast, when measuring gene expression associated with other lineages a decrease in the overall expression, and the average expression was observed during differentiation (Figure 3.4b). The average expression of genes in all three lineage enriched categories was very low at day0, within the hPSCs population but only in pancreatic tissue enriched genes did this expression increase into positive values, as opposed to the decrease to further low levels seen in both kidney and skin enriched genes. This clearly demonstrates the specificity of the protocol used here in producing mature pancreatic endocrine cell types.

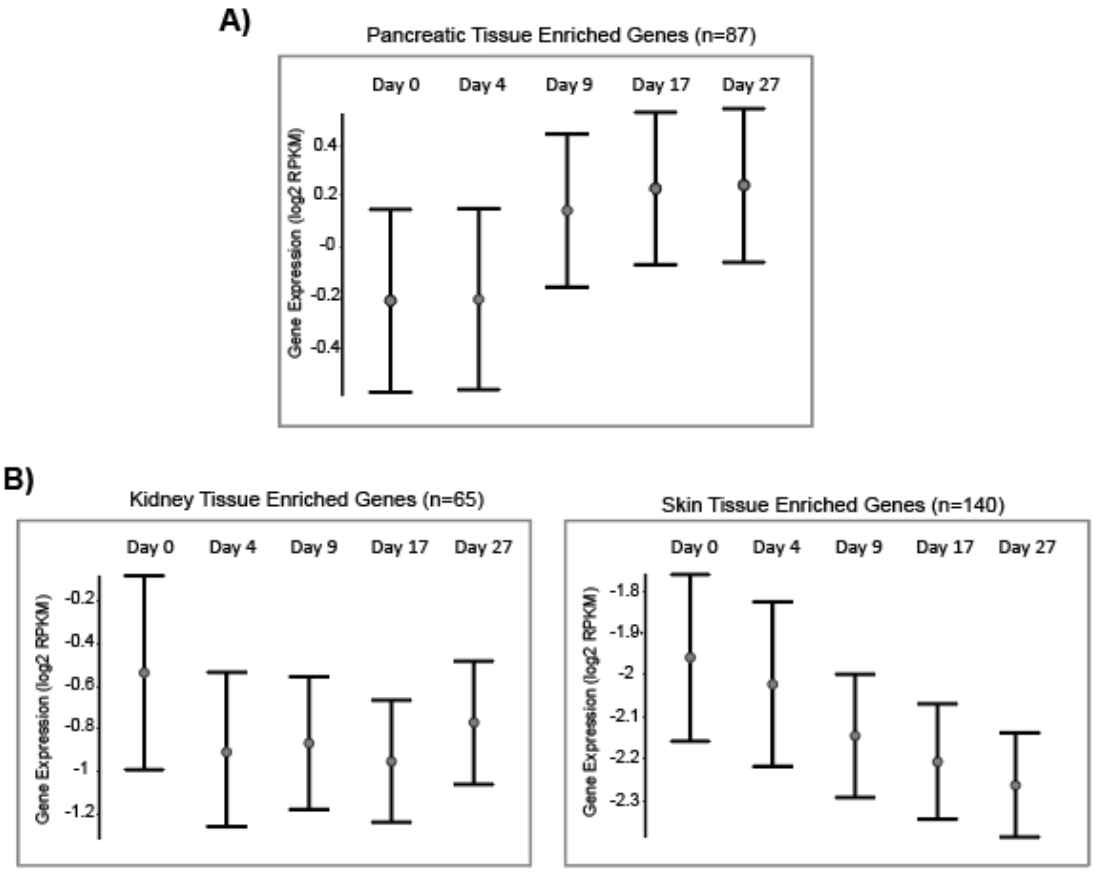


Figure 3.4 Pancreatic Lineage Specificity of *in vitro* Differentiated Pancreatic Cells. Lineage specific gene expression, as defined by the tissue cell atlas, at varying time points during *in vitro* pancreatic cell differentiation; (A) pancreatic, (B) kidney and skin, representative of other lineages. Shown are the mean quantification of all lineage specific genes at the different time points (represented by a filled circle) with standard error confidence intervals (represented by whiskers). Transcriptional expression is log transformed and averaged over three replicates.

With the specificity of the protocol shown, it was important to confirm the ability of the protocol to produce functional endocrine cells *in vitro* supporting the transcriptional data which demonstrate the formation of pancreatic endocrine-like cells. From marker expression analysis (Figure 3.2) it was observed that day27 cells are capable of expressing pancreatic endocrine hormones. A key function of pancreatic endocrine cells in the physiological setting is to secrete endogenously produced hormones in response to various stimuli. Therefore, to test the functional maturity of the cells, hormone production was analysed at protein levels by immunofluorescence (Figure 3.5). All three pancreatic hormones measured - c-peptide (the

insulin precursor), glucagon and somatostatin - are clearly expressed in the in vitro differentiated pancreatic cells at day27, indicating the production of β-cells, α-cells and δ-cells, respectively. The quantification of the percentage of hormone producing cells is challenging, due to the high density of the layer in which the cells grow. However, it is apparent that only a limited number of cells within the population are hormone producing, and as previously reported (D'Amour *et al.*, 2005; Schulz *et al.*, 2012) a number of these hormone expressing cells appear to be polyhormonal (Figure 3.5, arrow heads). From this it can be concluded pancreatic endocrine-like cells, capable of hormone section are produced during our protocol but these cells possess some of the previously described weakness commonly observed in in vitro pancreatic differentiation (Kroon, Laura A. Martinson, *et al.*, 2008; Schulz *et al.*, 2012; Pagliuca *et al.*, 2014), and therefore could benefit from further analysis into their development.

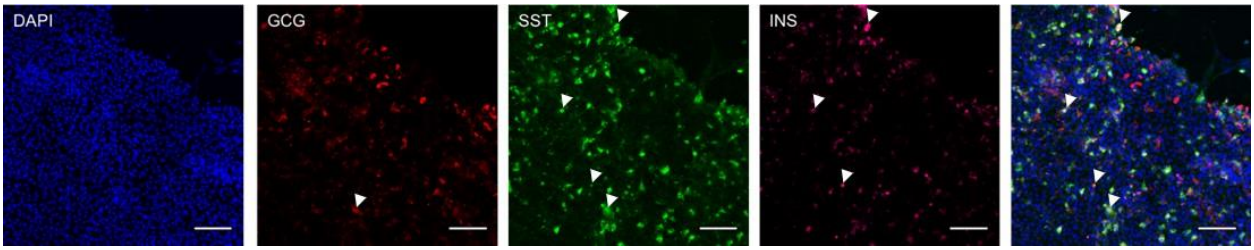


Figure 3.5 Expression of hormones from *in vitro* Differentiated Pancreatic Cells.
Immunofluorescence microscopy of day27 *in vitro* derived pancreatic endocrine like cells. Arrows indicate some of the identified regions of hormone overlap, suggesting polyhormonal cell expression. Scale bars, 100 μm.

3.3.2 Transcriptional dynamics during *in vitro* pancreatic differentiation.

To further improve a differentiation protocol, it is important to have an in-depth understanding of the cell populations produced during the prolonged *in vitro* differentiation at both a transcriptional and epigenetic level. Therefore, to further increase our knowledge of the formation of pancreatic endocrine cells, the transcriptional dynamics of the *in vitro* differentiating pancreatic cells were first analysed. To determine the levels with which gene expression changes can be detected between time points, individual gene loci were selected for analysis. Firstly, genes with a hypothesised expression pattern were selected, for example, during *in vitro* pancreatic differentiation, pluripotency gene expression is expected to decrease and eventually become lost as cells progress into a mature and restricted cell lineage. In contrast genes

associated with pancreatic cell expression are predicted to increase over time, from initially little or no expression. One such pluripotency factor is *ZSCAN10* (Zhang *et al.*, 2006) which clearly shows a dramatic transcriptional downregulation between day0 to day4 (Figure 3.6). In contrast the pancreatic progenitor marker *PDX1* (Leonard *et al.*, 1993) begins to show some expression in the day9 pancreatic endoderm population, with high expression at day17 and into day27, and no expression at earlier time points (Figure 3.6). Importantly, analysis of individual loci can also indicate more irregular or unexpected expression patterns. *FGF17* is one such gene, which shows very specific expression on day4 only with no levels of detection in either day0 or day9 (Figure 3.6). Similarly, the pancreatic gene *MNX1* (Harrison., 1999) began to show some expression by day4, although very low, earlier than expected for a pancreatic associated gene (Figure 3.6). The RNA-sequencing data generated can determine dynamic gene expression and can therefore be used to interrogate more global patterns of gene expression.

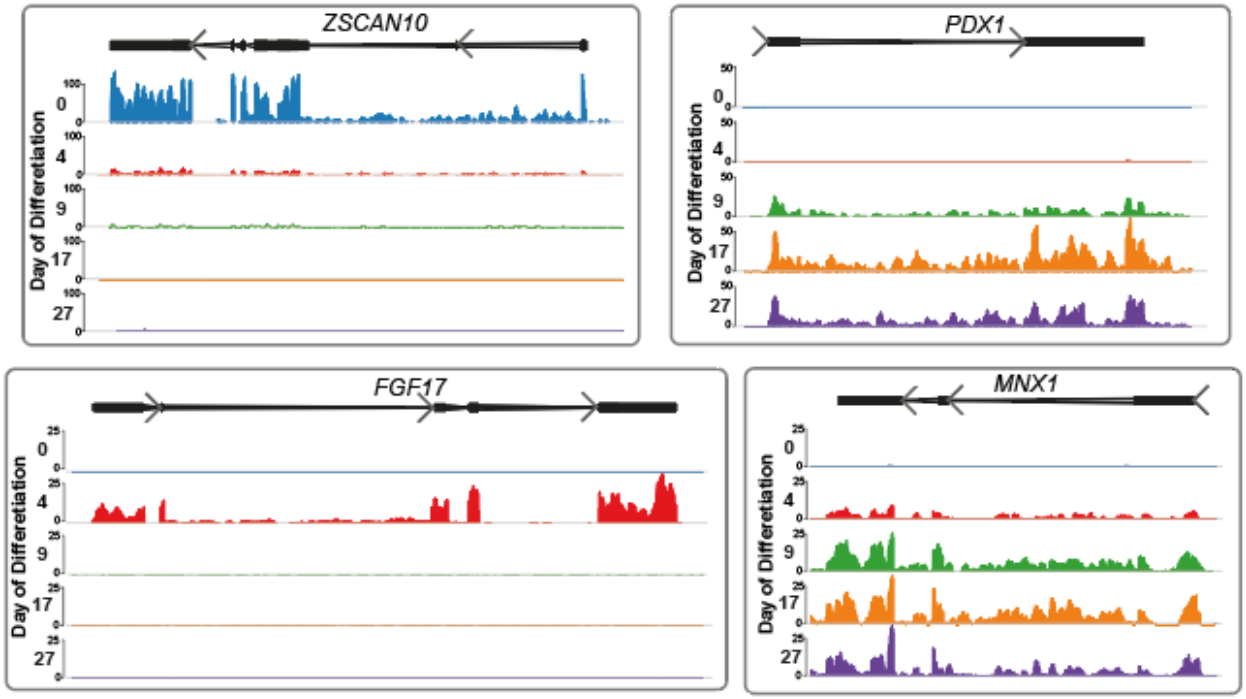


Figure 3.6 Examples of dynamic gene expression profile during differentiation. RNA-sequencing tracks for genes *ZSCAN10*, *PDX1*, *FGF17* and *MNX1*, on day 0, 4, 9, 17 and 27 of *in vitro* pancreatic differentiation. Each data track is an average of 3 biological replicates. The y-axis shows normalised read counts, scaled for each gene.

As the RNA-sequencing data can demonstrate patterns of differential gene expression between the cell types, analysis at a more global level could isolate patterns of expression in the transcriptional differences between the cell types. The first step in determining global transcriptional differences was to identify the differentially expressed genes between each individual time point using DEseq2 (Figure 3.7). These differentially expressed genes were then separated into either downregulated or upregulated, if they had decreased or increased, between an earlier and then subsequent time point during *in vitro* pancreatic differentiation. Their expression levels in all other time points were not taken into consideration when determining differentially expressed genes in order to only focus on neighbouring time points (4 possible combinations shown in Figure 3.7b). This identified a total of 5,868 genes which are differentially expressed between at least two consecutive time points during differentiation (Figure 3.7). The largest number of differentially expressed genes occurred in the transition of the two earliest time points, day0 to day4, and day4 to day9. In terms of the numbers of genes upregulated, the largest change was seen as cell transition from day4 and day9 (2,066 genes upregulated), with the next largest upregulation seen between day0 to day4 transition (1,298 gene upregulated). In contrast, the number of differentially expressed genes is much lower at the transition between the later time points. The transition from day9 to day17 showed an upregulation of 901 genes and the smallest number of genes upregulated, 306 genes, was seen between day17 to day27 (Figure 3.7a). For the number of genes downregulated, the largest change was seen in day 0 to 4, with 1,359 genes, similar to the number of genes downregulated between day4 to day9 (1,320 genes). However, day9 to day17 had only 607 genes downregulated, and this was decreased even further to only 91 genes downregulated between day17 and day27. This indicates that more transcriptional dynamics occur at the earlier time points and the numbers of genes which change in expression between time points decrease when the cells are more differentiated into a mature lineage. When the expression of these differentially expressed genes is viewed within all other time points of differentiation, no clear patterns of expression throughout the differentiation are observed (Figure 3.7b). This is most clear in genes differentially expressed at the earlier time points where a large number of genes have been selected for differential analysis, as a result a large spread in gene expression is seen at non-comparison timepoints, with the average expression of genes similar between all other time points. Due to the variation in the number of differentially expressed genes and the lack of clear pattern during differentiation, all differentially expressed genes were taken forward, to be analysed for specific expression dynamics.

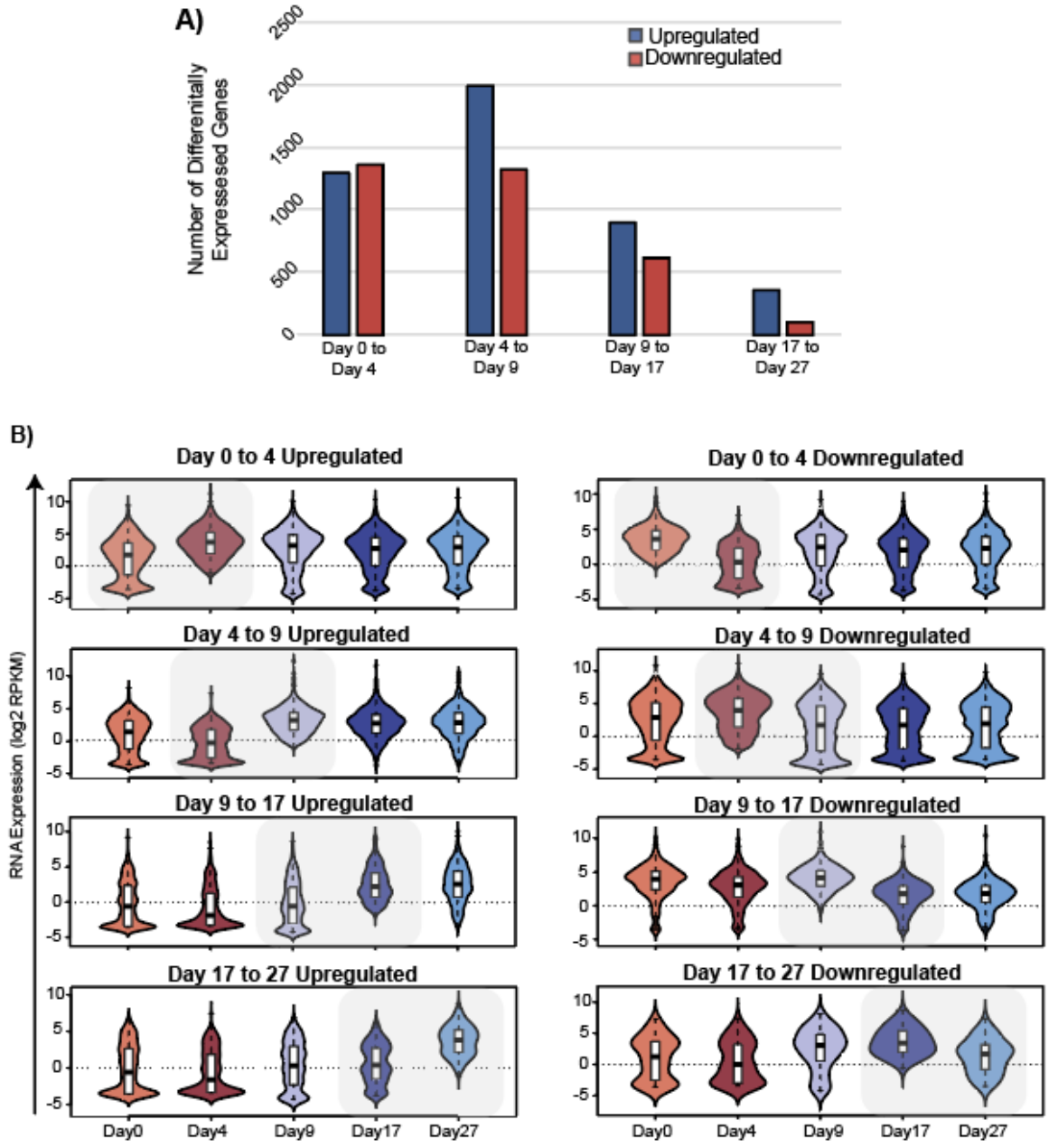


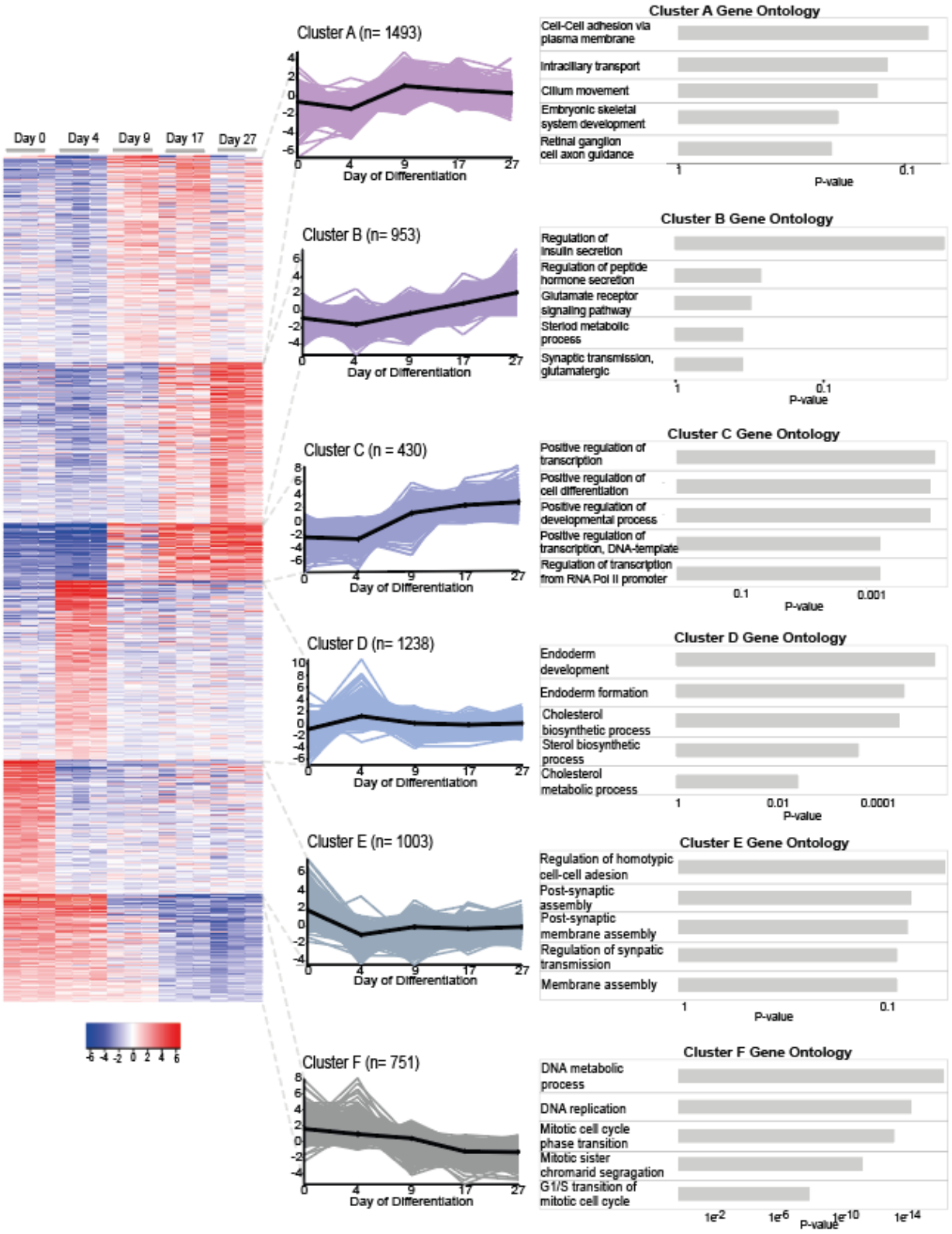
Figure 3.7 Differentially Expressed Genes between consecutive stages of Pancreatic Cell Differentiation. (a) The number of differentially expressed genes between each of the time point comparisons, split into upregulated (blue) and downregulated (red). (b) Combined violin and box plots demonstrating the average expression of genes which were deemed as differentially expressed between consecutive time points as determined by DESeq2, p-value < 0.05. The genes are separated by the time points at which they are differentially expressed, as labelled, and highlighted, and the expression of these genes is shown for all other time points in the differentiation. Data obtained from 3 biological replicates.

Using the set of differentially expressed genes determined above, the next step was to find distinctive expression patterns in order to study global transcriptional dynamics during *in vitro* pancreatic differentiation. Therefore, k-means clustering of the 5,868 differentially expressed genes was used to group the genes by specific expression patterns (Figure 3.8). From this analysis, curating the data into 6 distinct clusters was found to be optimal because it ensured each cluster would have a minimum of 200 genes and a pattern of gene expression that is distinct from any other cluster. Additional clusters did not capture further expression dynamics but instead separated clusters by the magnitude of expression change.

For clusters A, B and C, gene expression generally increases over the duration of the *in vitro* pancreatic differentiation. Of these increasing clusters, A contained the largest number of genes, 1,493 genes, and had only a moderate increase in gene expression, mainly occurring between day4 and day9, with a varied group of gene ontology (GO) terms, none of which reached significance, when analysed by EnrichR (Chen et al., 2013; Kuleshov et al., 2016). Genes within cluster B had a more defined increase in expression throughout differentiation, with a gradual increase from day4 onwards (Figure 3.8). In cluster C, an increase occurs sharply from day 4 to day9 with a small continual rise until day27 (Figure 3.8). Interrogation of GO categories for cluster B showed a strong and significant enrichment for insulin secretion ($p=0.03$), indicating that genes in this cluster are key in the differentiation protocol for forming mature pancreatic endocrine cells (Figure 3.8). Interestingly, the GO terms enriched in cluster C, which had a sharp increase in expression before cluster B, are associated with transcriptional and developmental regulators, therefore these genes may be important for the earlier stages in promoting differentiation (Figure 3.8). Cluster D expression pattern shows an increase in gene transcription from day 0 to day 4, followed by a loss of expression into day9, indicating these genes are important solely for the formation of the definitive endoderm cell population at day4 (Figure 3.8). Supporting this pattern of expression, the most enriched GO terms are endoderm development and endoderm formation.

The E and F clusters contain genes which decrease throughout the differentiation with more genes in cluster E compared to cluster F (1,003 and 751, respectively). In cluster E a clear loss of expression occurs after day0 into day4, and these genes continue with little or no expression throughout subsequent differentiation stages (Figure 3.8). In contrast, genes in cluster F show a more gradual loss of expression during differentiation with a reduction in expression typically occurring around the day9 time point (Figure 3.8). The genes in cluster F are highly enriched for cell cycle associated genes (Figure 3.8). This finding is consistent with a

decrease in the autonomous expansion and division as they exit the early progenitor stage into mature cell types (Becker et al, 2006). This transcriptional analysis demonstrated different patterns of increasing gene expression during the differentiation, with differentiation associated genes beginning to be expressed at day9 and genes associated with pancreatic endocrine becoming expressed by day17. Interestingly, genes which decrease in expression during differentiation appears to occur in two waves, the first after cells leave day0 pluripotent, and the second as cells differentiate into the pancreatic lineage, with loss occurring mainly between day4 and day17. These defined gene sets will be particularly valuable when integrated with ChIP-sequencing profiles to determine the contribution of H3K27me3 and H3K4me3 in gene expression dynamics over this prolonged *in vitro* differentiation.



(Legend over page)

Figure 3.8 K-means Clustering of Differentially Expressed Genes to Determine Expression Patterns. Using K-means clustering, all genes differentially expressed between consecutive timepoints were separated into 6 clusters based in expression patterns throughout pancreatic *in vitro* differentiation. Heatmaps indicates the individual gene expression, as indicated by the key, normalised between the overall gene expression. The line graphs represent the average expression pattern of all gene contained within the cluster at each of the five days. Isolated for analysis. The top 5 GO terms for each cluster was determined using EnrichR, genes contributing to the GO category are shown.

3.3.3 Associating transcriptional and histone modification changes during *in vitro* pancreatic differentiation.

To investigate the response of histone modifications, in combination with transcriptional data, as cells differentiate *in vitro* from hPSCs to pancreatic endocrine-like cells, H3K4me3 and H3K27me3 ChIP-sequencing libraries were produced from three independent differentiation replicates. To interrogate the histone modification during differentiation, they were first studied to determine if these selected modifications are present at individual gene loci, as would be expected from the transcriptional expression of the gene. As discussed in section 1.3.1, the H3K4me3 is typically assigned as an activating modification, and the H3K27me3 a repressive modification (Harikumar and Meshorer, 2015). For this analysis tracks from ChIP-sequencing of both H3K27me3 and H3K4me3, and RNA-sequencing were analysed for selected loci. Three genes were selected from the previous differential expression analysis and are therefore known to have varying expression between the differentiating cell types. As these genes change in expression level it would be expected, the histone modifications would also change between at least one time point to reflect this transcriptional change (Figure 3.9). In all the genes selected, either the H3K4me3 or the H3K27me3 modification is present at the gene loci for at least one of the developmental time points. Both *FOXA2* and *FGB* are genes that increase in expression during differentiation, and both loci have H3K27me3 initially present at day0. This repressive modification is lost at approximately day 4/day 9, which coincides with the beginning of transcription from the genes. An increase in the active mark H3K4me3 is also detected at these locations in parallel with the increasing transcription. In contrast, *NANOG*, which is highly expressed initially, followed by a gradual downregulation during differentiation, has high H3K4me3 levels at day0. These presence of the H3K4me3 modification is then are strongly reduced by day 4. The H3K27me3 mark was observed at the gene at day17 and day27, associated with the repressive state of the gene. These examples clearly indicate that the H3K27me3 and H3K4me3 histone modifications during *in vitro* differentiation correlate with

gene expression as expected. Although, studying individual gene loci clearly confirms the expected association with transcription it does not indicate a role for the histone modification controlling the cell transcriptional network as a cell progresses through *in vitro* differentiation and therefore the next step is to analyse histone modifications at a global level.

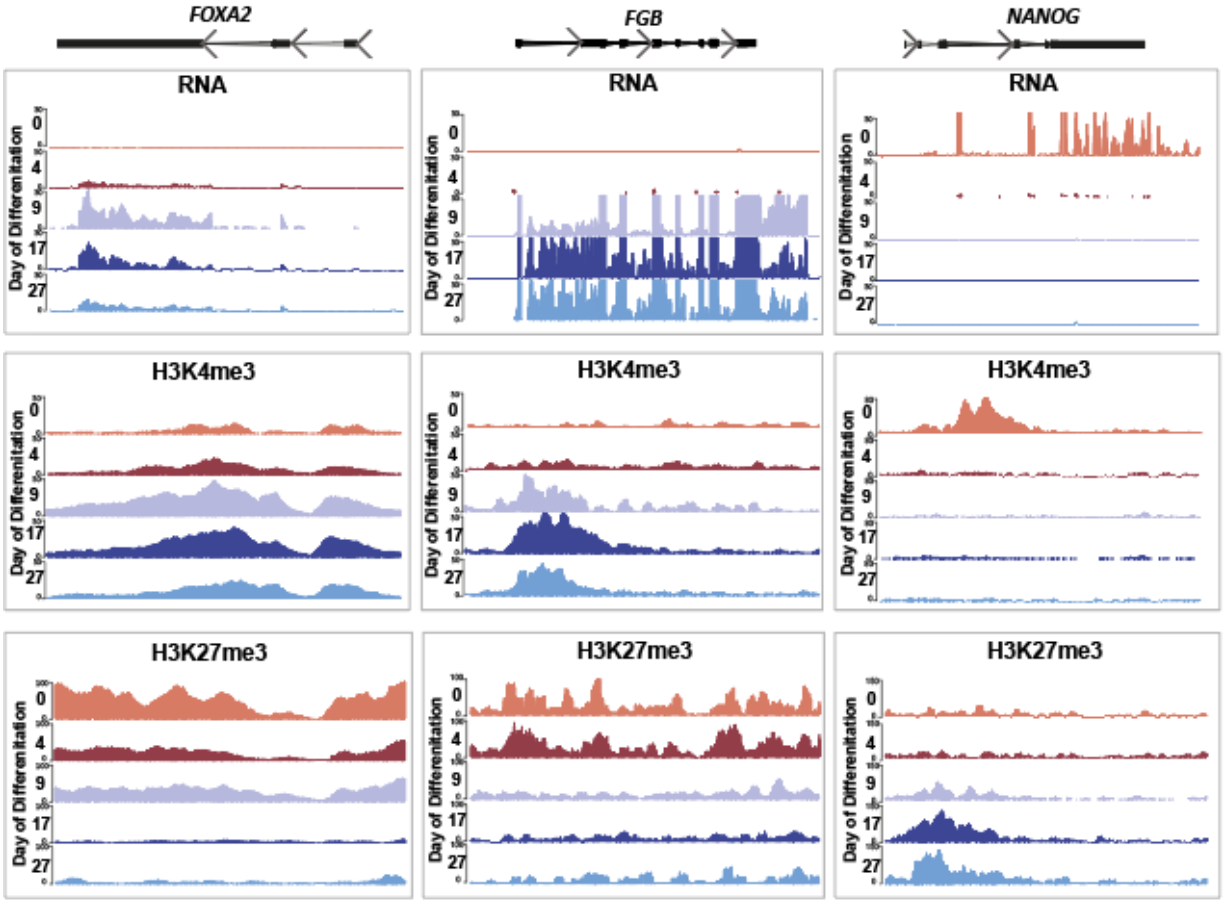
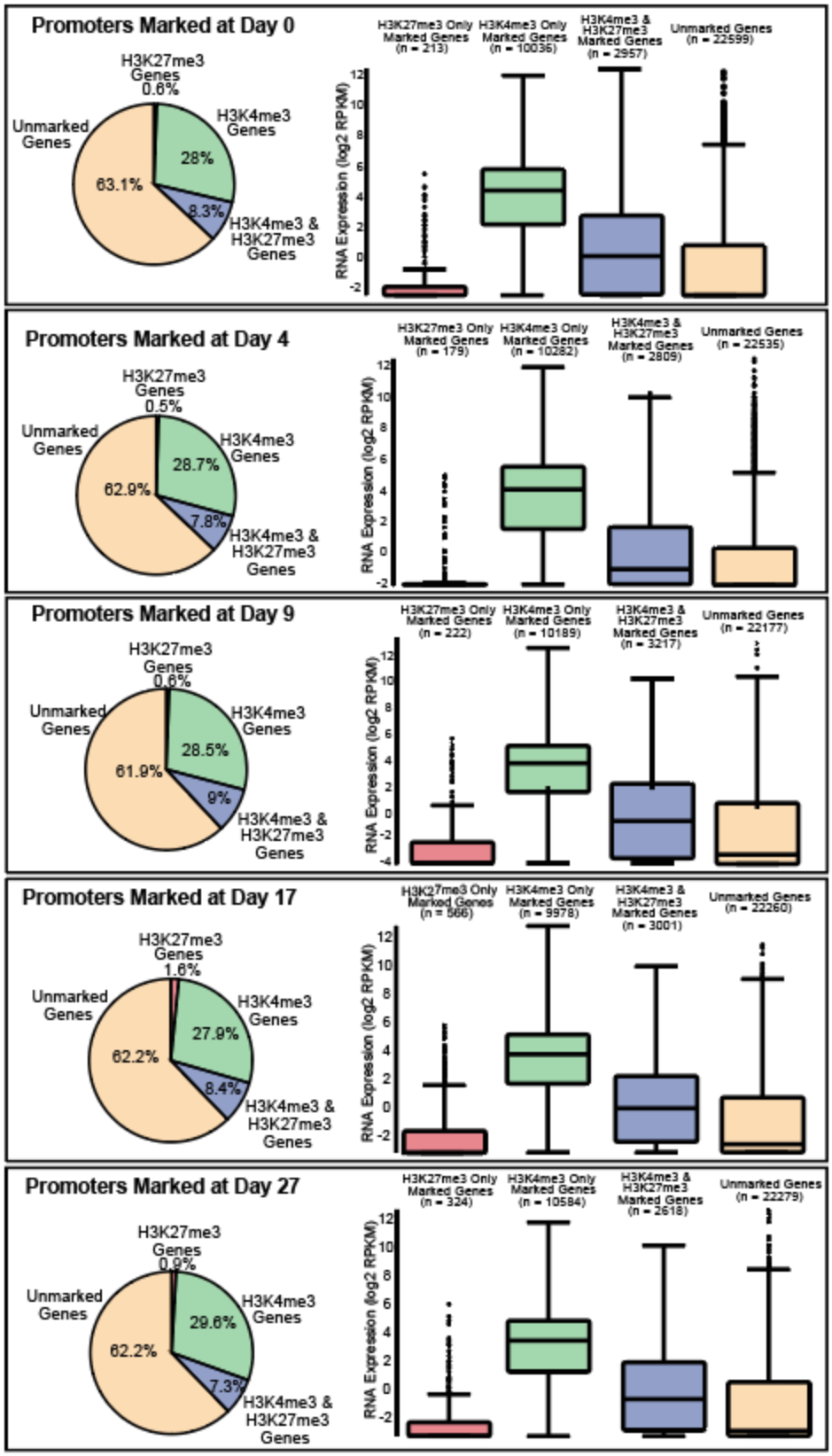


Figure 3.9 Examples of histone modifications associated with differentially transcription during differentiation. ChIP-sequencing and RNA-sequencing tracks for the genes *FOXA2, FGB* and *NANOG*. Plots were produced using running window probes with read count corrected quantification. Data is an average of 3 biological replicates.

To determine how the histone modifications respond in the developing cell population during differentiation, their global distribution at transcriptional start sites (TSS) +/- 500 bp of gene promoters, as determined by MACS peak-calling, in each of the isolated time points was analysed (Figure 3.10). Corroborating our individual loci data, and as would be predicted, gene promoters marked by H3K4me3 only are, on average, the most highly expressed genes at all time points. The H3K27me3 only marked gene promoters are amongst the lowest expressed genes with average expression levels that are lower than H3K4me3 marked, unmarked and H3K27me3 and H3K4me3 marked, in keeping with the repressive effect of H3K27me3 modifications. The genes with both H3K4me3 and H3K27me3 present at promoters represent an intermediate expression level, although it is unclear if the presence of both modifications is due to bivalent marking of a single gene or if two separate cell populations are marked by each modification. The intermediate expression level does not clarify between these scenarios as both situations would lead to this average expression pattern in a mixed population (Figure 3.10).

The distribution of histone modifications was similar throughout all five time points, with promoters unmarked by either of the histone modifications forming the largest group, between 61.9 - 63.1% of TSS, although many of these genes are likely to be modified by other histone modifications not analysed in this study (Figure 3.10). The H3K4me3 only modification contributes to slightly more than a quarter of all TSS, with 28% (10,036 TSS) modified at day 0 and 29.6% (10,584 TSS) modified at day 27. The H3K27me3 only modified genes comprised the smallest group with a maximum of 566 genes (1.6% total TSS) at day 17 and as little as 179 (0.5% total TSS) genes classified as H3K27me3 only at day 4. The most surprising result from this global data analysis was the lack of variation in the percentage of genes co-marked by H3K4me3 and H3K27me3 throughout differentiation. The presence of both modifications at the same gene was measured at a minimum of 7.3% of TSS at day 27 and a maximum of 9% at day 9. This means that the H3K27me3 modification either alone, or in combination with H3K4me3, persisted at a minimum of 8.3% of all gene promoters and at day 27 H3K27me3 was present at 8.4% genes (Figure 3.10). H3K27me3-marked loci are known to be abundant in undifferentiated hPSCs, typically in combination with H3K4me3 (Pan *et al.*, 2007; Zhao *et al.*, 2007) to maintain genes in a poised state. The persistence of a similar percentage of H3K27me3-modified genes could suggest that very little of the initial H3K27me3 modification is lost throughout the 27-day *in vitro* differentiation.



(Legend over page)

Figure 3.10 (over page) The global percentage of gene promoters marked by histone modifications throughout the stages of differentiation. Pie charts demonstrate the % of TSS +/- 500bp that have the presence of histone modifications by ChIP-sequencing at each timepoint. Genes were divided into 4 histone modified categories and the gene expression in each category is demonstrated by box whisker plots, with the central line representing the median, boxes demonstrating the 25th and 75th interquartile and whiskers show the median plus/minus the interquartile range multiplied by 2. Individual points that fall outside this range are shown as individual points. Data is an average of 3 biological replicates.

Although the histone modifications were distributed at TSS, at similar levels throughout the differentiation timepoints it was unclear if this was caused by a persistence of modifications at the same genes or a shift to a different set of genes. Analysing the H3K27me3 modifications demonstrates a large majority of the genes that are H3K27me3 marked at later time points are the same genes that initially possessed H3K27me3 at day0. For example, at day27, 72% (2,118 genes) that were marked by H3K27me3, were also marked at the initiation of the differentiation at day0. For H3K4me3, an even higher percentage of those genes modified at later time points are classified as H3K4me3 marked at day0, with 85.6% (11,272 genes) of H3K4me3 modified genes at day27 also being marked by H3K4me3 at day 0 (Figure 3.11). The majority of genes that are modified at day27 will be marked by these histone modifications due to the retention of the marks, as opposed to de novo formation throughout differentiation.

The acquisition of histone modifications during differentiation appears to contribute only a minimal proportion to the overall modified number, this analysis can indicate which time points are associated with the de novo appearance of histone modifications at genes. In the case of the H3K27me3 modification, cell populations succeeding day0, will have approximately 14% of newly modified genes contribute to the total marked genes for their respective time points at days4, 9 and 17. At day27 however only 0.7% of previously unmodified genes gain H3K27me3. Of the genes which gain H3K27me3 at day4, these genes maintain H3K27me3 into day27 in almost at 9.7% of total H3K27me3 modified in the final cell type, with 8.7% of day27 H3K27me3 modified genes formed at day9 and 8.9% of day27 H3K27me3 modified gaining the modification at day17. This indicates that during *in vitro* pancreatic differentiation approximately 824 of the day27 H3K27me3 modified genes are acquired in the developing cell populations, although little formation occurs between day 17 and day 27.

Differences are also seen with the appearance of H3K4me3 at newly marked genes. For this modification, the only time point to significantly gain H3K4me3 at a large percentage of genes, is day 4, with 12.9% of genes (1,694 genes) at day 4 newly marked by H3K4me3. These day 4 formed genes retain a contribution of about 12% to H3K4me3 modified genes in later time points. It was previously observed that a large set of genes are transiently expressed at day4 (Figure 3.8), which may contribute to the large percentage of genes gaining novel H3K4me3 modifications at this timepoint. In contrast only 2.5% of genes (329 genes) H3K4me3 modified at day 9 are novel, and only 0.1% (17 genes at day 17 and 11 genes at day 27) in both day 17 and day 27. This analysis reveals that the majority of gene marked by either modification in the day27 pancreatic cells is most likely due to retention of the modification from day0 iPSC.

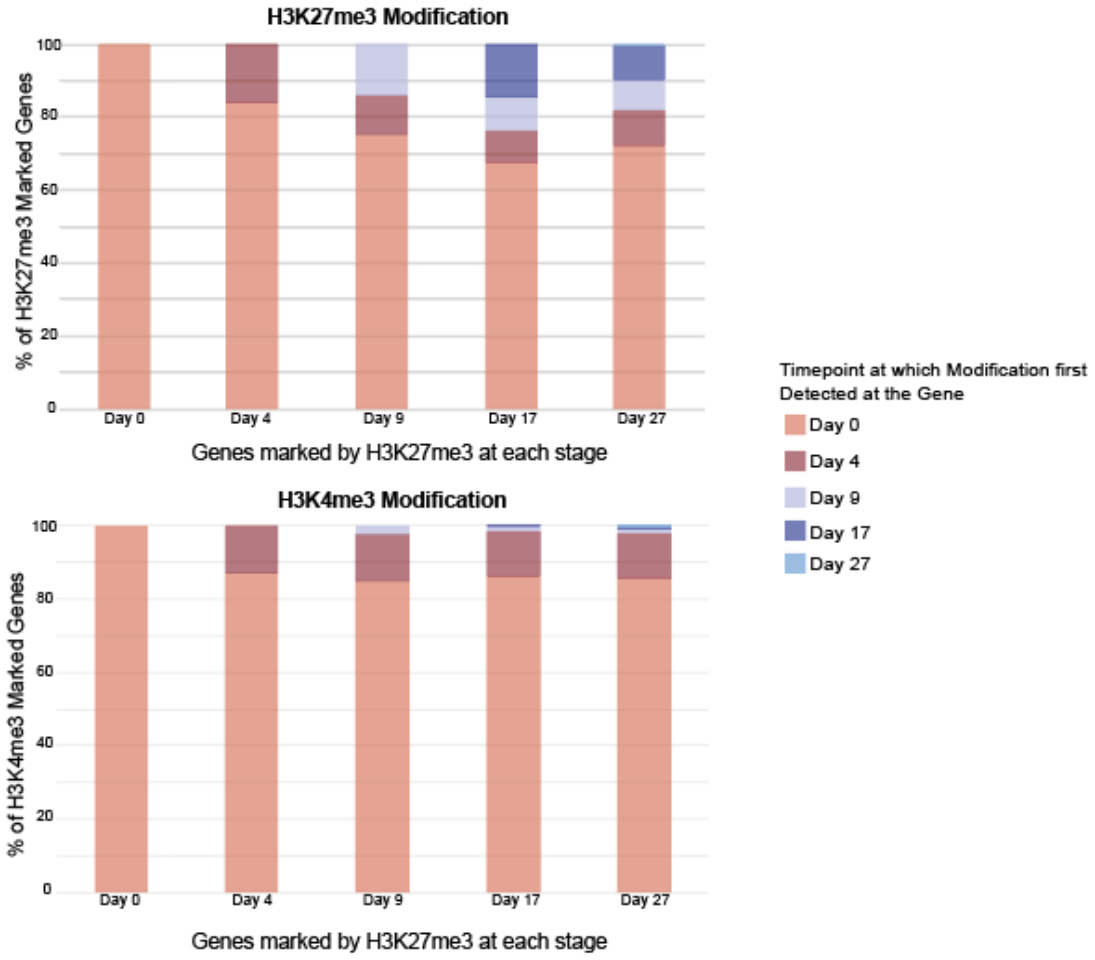


Figure 3.11 The formation of H3K27me3 modifications throughout *in vitro* differentiation. H3K27me3 marked genes at each timepoint, separated by the time point at when that gene was first classified as H3K27me3 modified. This was calculated for all H3K27me3 marked genes throughout the 5 stages of differentiation. All data is averaged from 3 biological replicates.

To study the loss and gain of histone modifications throughout differentiation at genes with changing transcription, differentially expressed genes (as determined by k-means clustering presented, Figure 3.8) were associated with the histone modifications present at TSSs. ChIP-sequencing data categorised each gene as either unmarked, H3K27me3 only, H3K4me3 only or both H3K27me3 and H3K4me3. To study the interplay between histone modification changes and transcriptional dynamics during *in vitro* differentiation, clusters with increasing transcription over the 27 days were first analysed (Figure 3.13). These clusters - A, B and C - all show an increase in the percentage of genes marked by H3K4me3 only. In cluster A, with a moderate increase in gene expression occurring mainly between day4 and day9, the percentage of genes with H3K4me3, regardless of other modifications, increases from 44.1% at day 0 to 51.3% by day27, meaning only 7.2% of genes in this cluster gained H3K4me3. This H3K4me3 increase did not occur predominantly between day4 and 9, in parallel with increased transcription, but an increase in 2.1% genes was observed from day 0 to 4, preceding transcriptional upregulation, followed by a 0.9% increase from day4 to 9. This was then followed by an increase of 2.1% between both day9 to day17 and day17 to day27, after the largest transcriptional increase was observed.

Clusters B and C have more continuous increases in gene expression throughout differentiation as opposed to the single time point increase seen in cluster A (Figure 3.12). In cluster B, 11% of genes gain H3K4me3 only, with over 10% of these genes changing between day9 into day17 and day17 into day27, although increasing transcription begins at day4. Similarly, in cluster C, transcription increases throughout differentiation with the strongest upregulation between day4 and day9, but the largest increase in H3K4me3 modifications occurs between day9 to day17 when 13.8% of genes transition to H3K4me3 only modified. In cluster C, a total of 21.5% of genes gain H3K4me3 from day0 to day27, but only 1.7% of these genes change between day 4 and day 9 (Figure 3.13). This indicates that the transition of genes to the H3K4me3 only category does not occur at a single point in relation to the increase in transcription but occurs at timepoints both preceding and following transcriptional upregulation.

With this increase in percentage of H3K4me3 only, there was almost no change in the percentage of either unmarked or H3K27me3 only genes within the clusters. Instead, with increasing percentages of H3K4me3 only, a concomitant decrease occurs in the percentage of genes marked by H3K4me3 and H3K27me3, for all three clusters (Figure 3.12). This suggests that genes which gain transcription throughout differentiation do so by losing the H3K27me3 modification gene with H3K4me3 already present, as opposed to gaining novel H3K4me3,

which supports the analysis showing little novel modifications on previously unmodified genes, Figure 3.12. This is the situation that would be expected as cell resolve bivalency as they differentiate from a pluripotent state (Meissner et al., 2008; Wen., 2009). Interestingly, in genes that increase in transcription, over 10% of genes in each cluster maintain H3K27me3 solely or in combination with H3K4me3. This persistence of genes with both H3K27me3 and H3K4me3 is seen in both cluster B and C at day27 but would be expected to be lost from genes as they transition to a transcriptionally active state. This suggests a persistence of H3K27me3 in some cells, even at genes upregulating their transcription in differentiation.

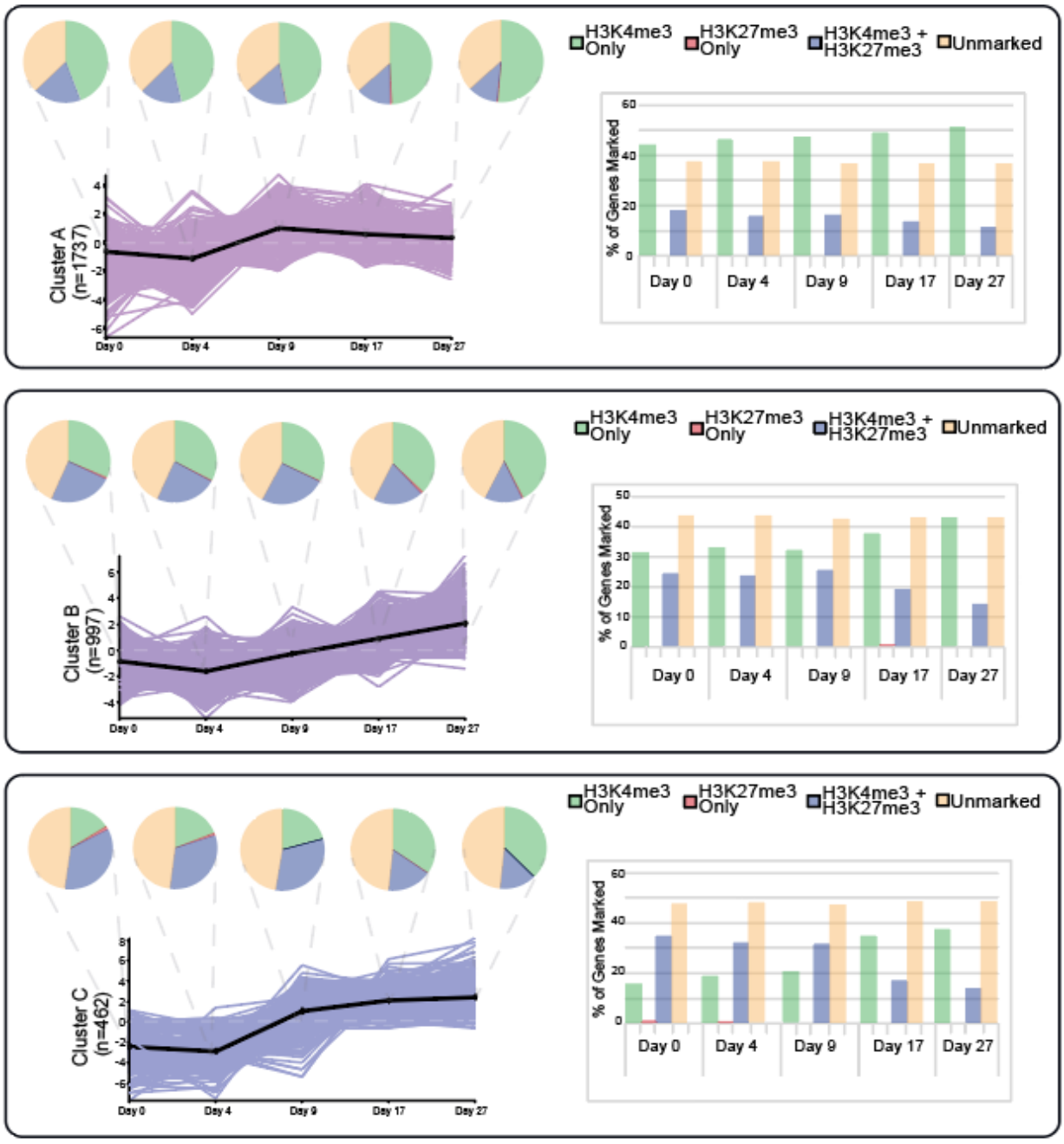


Figure 3.12 Histone modifications associated with genes that increase in expression during differentiation. Pre-defined gene sets with increasing expression were split into categories based on the histone modification present at the gene promoters. This is represented as a percentage of the total gene set at each of the 5 differentiating timepoints. All data is from an average of 3 independent differentiation.

To determine if the histone modifications are gained and lost in a similar manner at genes which decrease over time, pre-defined gene clusters - E and F - with decreasing expression throughout the 27 days of differentiation were assigned into the four histone modification categories as in Figure 3.12. Note that cluster D was not analysed here because it has a transcriptional increase at a single time point. In cluster E, a sharp decrease in gene expression was observed between day 0 and day 4 and this loss of expression was relatively stable until day 27. As expected, the percentage of genes modified with only H3K4me3 was decreased at genes in cluster E from 49.5% at day 0 to 39% at day 27. Although the transcriptional decrease occurred mainly between day 0 and 4, the H3K4me3 only modified group decreased by 4.7% between day 4 and day 9 and further decreased by 4.3% between day 9 to day 17, it was then observed that 3.2% of genes gained H3K4me3 only modification into day 27. In cluster F, a more gradual, and continuous decrease in gene expression is observed, with a 14.6% overall decrease in the H3K4me3 only modified category between day 0 to day 27. Interestingly, in this transcriptionally decreasing cluster, day 0 into day 4 and day 17 to day 27 sees an increase in H3K4me3 only marked genes by 3.2% and 2.4% respectively. This same gene set sees 7.2% and 12.9% of genes lose the activating H3K4me3 only modification between day 4 into day 9 and day 9 into day 17, respectively. This suggests that the histone modifications studied do not precede the transcriptional changes during *in vitro* differentiation and are only partially indicative of the transcriptional change occurring in the cell population.

For the genes which are becoming downregulated, it was hypothesised that the H3K4me3 would be lost as the genes transition from active to repressed, with the genes then becoming either unmarked, as other epigenetic mechanisms that were not analysed are deposited, or becoming modified by H3K27me3 only, if this modification was previously present. However, in both clusters, very little or no increase in the percentage of unmarked genes were seen (Figure 3.13) instead the loss in the percentage of H3K4me3 only resulted in a similar increase of the H3K4me3 and H3K27me3 marked promoters. As previously discussed, the presence of H3K4me3 and H3K27me3 in a mixed population is unclear as to whether these marks exist at the same promoter in a single cell or if we are studying the aggregate signal from different cells. In either of these circumstances, the presence of both H3K4me3 and H3K27me3 was unexpected. The presence of both marks at a single gene would result in bivalent chromatin, which is typically associated with undifferentiated hPSCs (Pan *et al.*, 2007; Zhao *et*

al., 2007). Alternatively, different cells in the same population with opposing modifications could indicate an inability to efficiently silence genes during *in vitro* differentiation. Therefore, it appears that the histone modifications during this differentiation do not necessarily occur before transcriptional changes, and the persistence of H3K4me3 and H3K27me3 at gene both increasing and decreasing in transcription may indicate problems with insufficient resolution of histone modifications during *in vitro* differentiation.

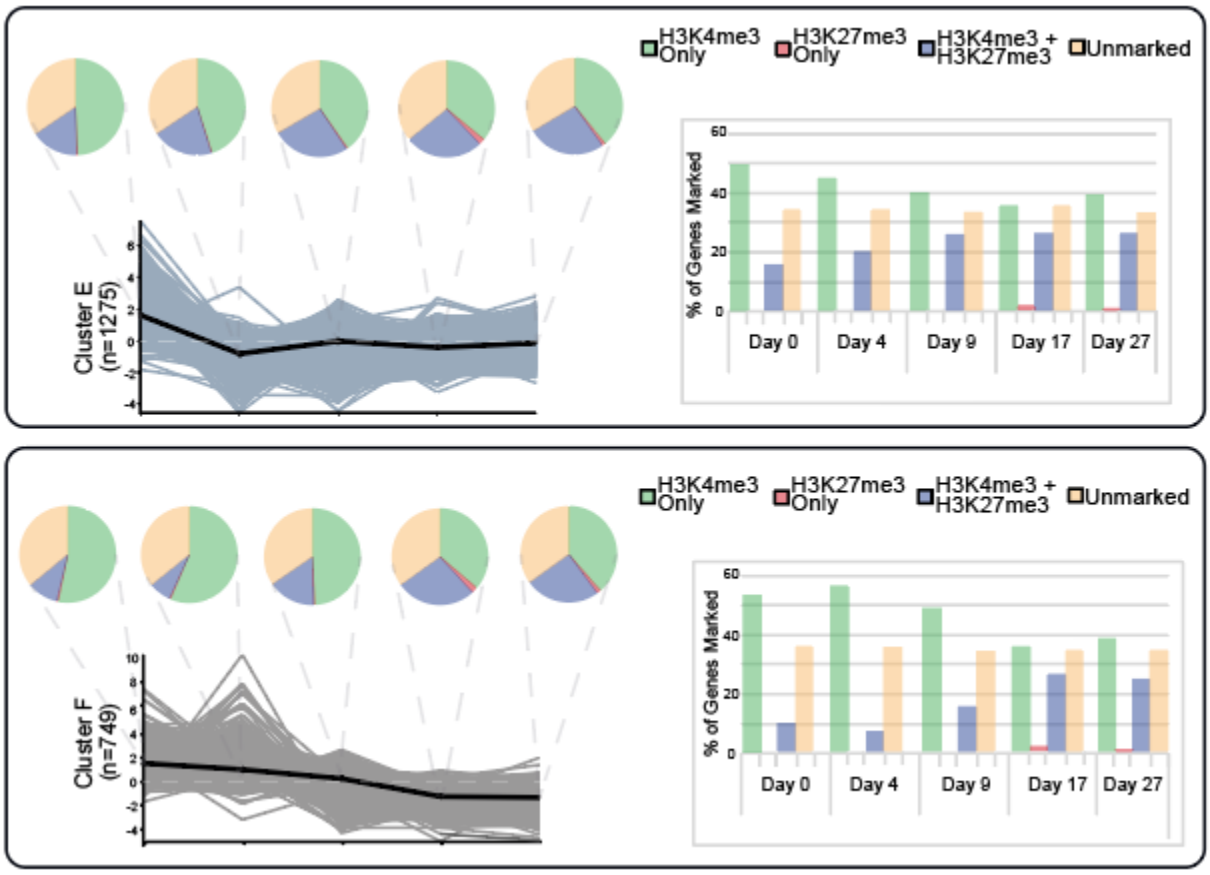


Figure 3.13 Histone modifications associated with genes that decrease in expression during differentiation. Gene sets with increasing patterns of expression were split into categories based on the histone modification present at the gene promoters in the cluster based on ChIP-sequencing. This was represented as a percentage of the total gene set at each of the 5 differentiating timepoints. All data is from an average of 3 independent differentiation.

3.3.4 Comparing transcriptional and epigenetic profiles between in vitro derived pancreatic cells and in vivo cells.

Along with improving our understanding of the contribution of histone modifications to prolonged human *in vitro* differentiation, by studying the *in vitro* derived pancreatic endocrine cells in detail, improvements to the current differentiation protocol could be uncovered which may produce higher quality final cell types. Current protocols produce functional pancreatic-like cells that closely mimic the developmental processes; however, the resultant cells suffer from several drawbacks, including polyhormonal expression and foetal-like transcriptomes (D'Amour *et al.*, 2005; Siniša Hrvatin *et al.*, 2014; Veres *et al.*, 2019). To improve the outcome of these differentiation events, it is important to understand the limitations of the *in vitro* derived cells by comparing them to the desired *in vivo* counterparts - adult pancreatic islet cells. The first stage towards investigating any differences between the cells is to investigate how they compare transcriptionally. To examine this, publicly available datasets of adult and foetal pancreatic cells transcriptomes were compared to my previously generated RNA sequencing timecourse data. For this analysis, adult islet data from Fadista et al, 2014 was used in which 5 donors from this data set were selected from a large sample of data sets to cover a breadth of donor age and gender. The foetal data is a compilation of data sets from 2 separate studies. The first by Cebola et al, 2015, where human embryonic pancreatic buds at Carnegie Stage (CS) 16 - 18 were isolated by dissection, and the second by Jennings et al, 2017 where the human embryonic dorsal pancreatic structures at CS 12 – 14 were removed using laser dissection (Fadista *et al.*, 2014; Cebola *et al.*, 2015; Jennings *et al.*, 2017).

Previous studies have suggested that *in vitro* derived pancreatic cells transcriptionally resemble developing foetal pancreatic cell types more closely compared to fully mature adult pancreatic islets (Siniša Hrvatin *et al.*, 2014). A global, unbiased comparison of *in vivo* derived pancreatic cells to the time points of a differentiating *in vitro* pancreatic cells was used to determine if this holds true for the protocol used in this current study (Figure 3.15). This analysis clearly demonstrates that the day 27 *in vitro* pancreatic endocrine cells cluster closer to foetal tissue than to adult cells, based on the transcription of the top 2000 most variable genes between all samples. By including the *in vitro* differentiation intermediates, the gradual emergence of pancreatic phenotype can be observed. Interestingly the *in vitro* derived cells are located, on both principal components 1 and 2, at an intermediate point in between the two foetal pancreatic developmental time points, CS 12 - 14 and CS 16 - 18. With only a small

number of in vivo data sets available, and the difficulties in obtaining such samples, this intermediate positioning is an interesting result but is challenging to determine if this has true biologically meaning, as other technical factors could contribute to transcriptional differences, particular with the foetal sample preparation (Figure 3.15). While demonstrating the strong similarity to foetal-like pancreatic tissue, this analysis also demonstrates the inability of the protocol to generate the desired, mature adult-like pancreatic cells by day27, based on the final transcriptional signature of the cells. The adult pancreatic transcriptional clusters away from all other data point on principal component 1, which contributes the most (47%) to the variability between the samples. With such a separation between the adult islet cells and all other samples, it was important to confirm that this was a biological effect and had no confounding technical variability that could result in this large separation. To overcome this, an adult islet sample was obtained (from Lonza Cell Biosource) and was taken from RNA isolation to sequencing, in parallel with the processing of *in vitro* samples. This revealed little contribution of sample preparation to the differences between *in vitro* generated pancreatic endocrine cells and the adult islet cells, implying that the differences observed are due to the transcriptional differences between the cells, as opposed to technical sample preparation/sequencing. It can be concluded that the *in vitro* pancreatic cells produced by this study more closely resemble the transcriptional signature of foetal pancreatic cells compared to adult pancreas.

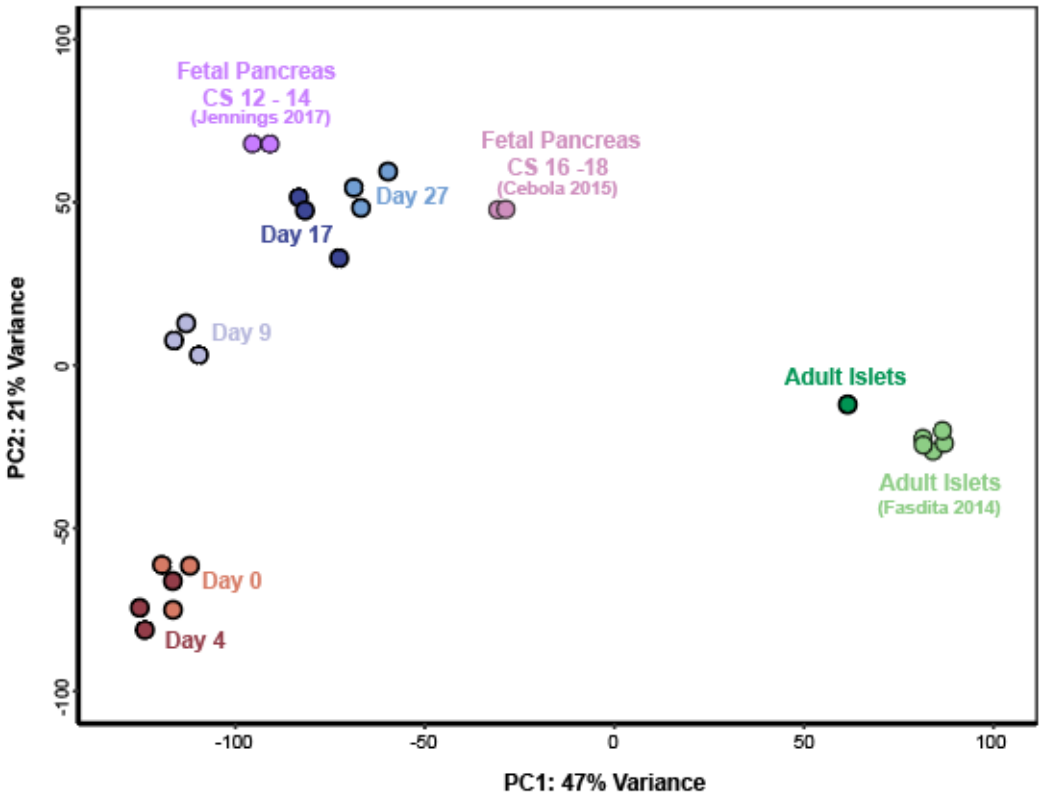


Figure 3.14 Comparative transcriptional analysis of *in vitro* and *in vivo* derived sources of pancreatic differentiated cells. Principal component analysis plot from RNA-sequencing data of the top 2000 most variable genes as determined by DESeq2. Included in the analysis are the five isolated time points of *in vitro* pancreatic differentiation, an adult islet sample, and Carnegie staged (CS) foetal and adult pancreatic data from publicly available data sources as indicated.

Although the *in vitro* derived cells cluster closer to foetal tissues based on a broad transcriptome comparison, the expression levels of key pancreatic hormone genes was also analysed as an indication of the functional performance of day 27 *in vitro* cells in comparison to adult and foetal cells. The expression of the three key endocrine hormones - *insulin*, *glucagon,* and *somatostatin* - which are expressed by the three main cells which constitute the pancreatic endocrine cell types, were compared between the different sources of pancreatic cells (Figure 3.14). As expected, the transcriptional expression of all three endocrine hormones is significantly less in the *in vitro* derived pancreatic compared with the adult sourced pancreatic islets. However, both *INS* and *GCG* have significantly higher expression in the *in vitro* day 27

cells compared with foetal pancreatic cells. This indicates that whereas *in vitro* pancreatic cells are not representative of adult transcriptional levels, they do have expression levels of key genes that are distinct from the foetal cells, with an intermediate level of expression between foetal and adult pancreatic tissues (Figure 3.14). In contrast the expression of *somatostatin* in *in vitro* derived cells was more similar to foetal pancreatic cells, as compared to adult islets. The PCA therefore identified a similar transcriptome profiles between the *in vitro* and foetal cells, but the pancreatic hormones suggest that the *in vitro* derived cells may be at a more mid-point between the adult and foetal cell types, in terms of pancreatic transcript expression.

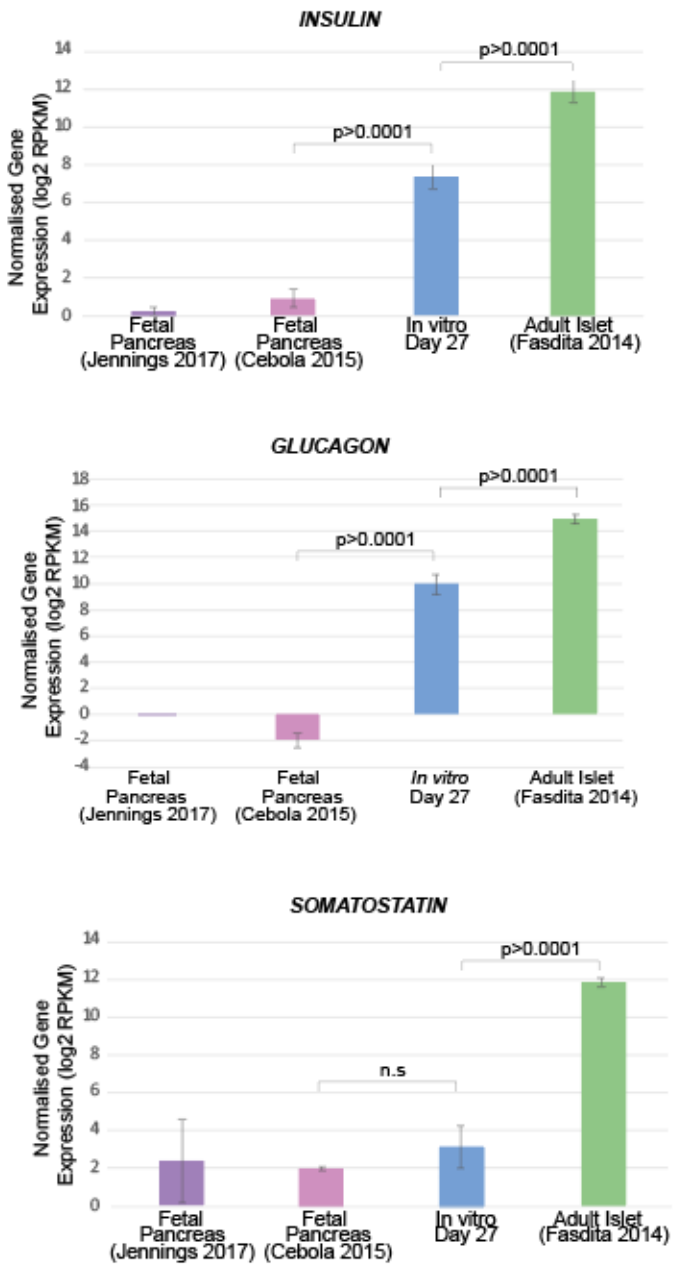


Figure 3.15 Intermediate expression of pancreatic hormone genes from *in vitro* derived cells compared to foetal and adult pancreatic cells. The normalised read counts for 3 selected genes, pancreatic endocrine genes – *INSULIN, SOMATOSTATIN* and *GLUCAGON* pancreatic endocrine progenitor gene are shown for all 4 cell types. Data averaged from a minimum of n=2, error bars indicate ±SEM, and were compared using one-way ANOVA followed by Bonferroni's multiple comparison test.

To better determine the differences that exist between foetal and adult cells compared to *in vitro* derived cells, differential gene expression analysis was performed (Figure 3.17). Using a false discovery rate of 0.01 and a minimum fold change of 2, differentially expressed genes between foetal, adult and day27 *in vitro*-derived cells were determined for each pancreatic cell source. For this analysis, only the Cebola et al 2014, CS 16 - 18 data sets was used to represent the foetal cell type, as raw data was unavailable for the Jennings et al, 2017 data, and this is required for the pipeline analysis. As predicted from the prior global analysis, there are fewer differentially expressed genes between the *in vitro* derived cells and foetal cells as compared to the number of differentially expressed between either *in vitro* derived cells and adult cells or foetal cells and adult cells (Figure 3.17a). Because the aim of the *in vitro* pancreatic differentiation protocol presented here was to generate adult-like pancreatic cell types, the differentially expressed genes between *in vitro* and adult cells were analysed further. Of the 3,874 genes which are down regulated between the *in vitro* and the adult samples, 1,983 of these down regulated genes (52%) are also downregulated when comparing foetal and adult pancreatic cells. Similarly, of the 4,705 genes which are transcriptionally upregulated in the *in vitro* derived cells over the adult islet cells, 2,451 (52%) of these same genes are also upregulated in foetal over adult pancreatic cells (Figure 3.17b). This suggests that approximately half of the genes that are differentially expressed between the *in vitro* cells and adult cells can be accounted for by the foetal-like transcriptional profile of the *in vitro* derived cells. To investigate the possible causes of the differentially expressed genes between the *in vitro* and adult islet cells, the histone modification profiles were next compared between the cell types.

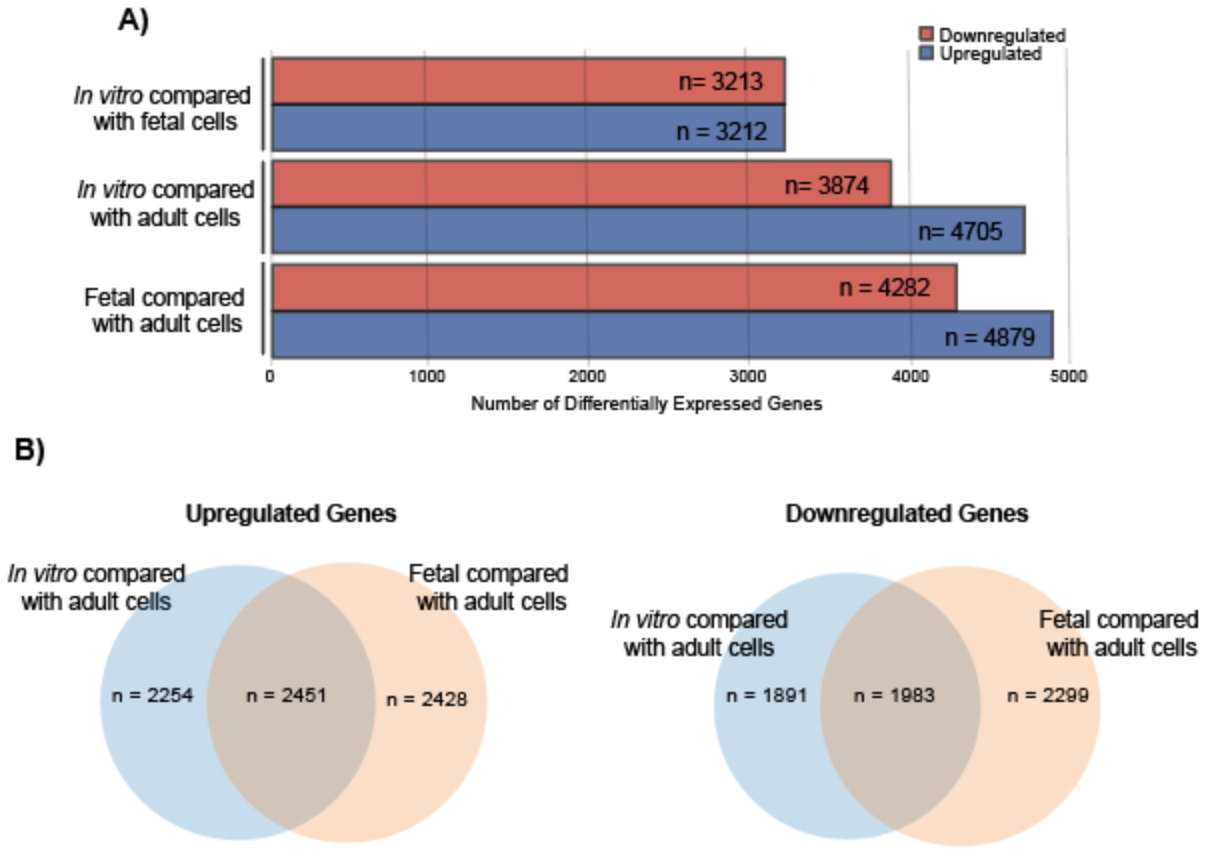


Figure 3.16 Determining differentially expressed genes between *in vitro* and *in vivo* derived pancreatic cells. a) The number of differentially expressed genes between each cell type as determined by DESeq2, with >2-fold change, and a false discovery rate <0.01. Differentially expressed genes were separated into upregulated and downregulated for downstream analysis. b) Overlap between genes differentially expressed in day 27 *in vitro* cells compared with adult cells and foetal isolated pancreatic cells compared with adult cells, separated into up and downregulated genes.

As my previous analysis had indicated an unexpected persistence of H3K27me3 in the day 27 *in vitro* derived pancreatic cells, histone modification profiles between the cell types were studied to determine if this may contribute to the transcriptional variation between the cells. To investigate this hypothesis, data was obtained from a study by Bhandare et al, 2010, in which adult islets were isolated from four separate donors and ChIP-sequencing for H3K27me3 and H3K4me3 was undertaken in these cell types. The data was then processed using the same settings used in the previous ChIP-sequencing analysis of *in vitro* pancreatic cells, to identify the

gene promoters that are marked by H3K27me3 and H3K4me3 in adult islets (Figure 3.17). This analysis demonstrated that the distribution of H3K4me3 only modified promoters were similar between adult islets (10,458 TSS H3K4me3 only) and *in vitro* derived cells (10,584 TSS H3K4me3 only). Similarly, the number of H3K27me3 only modified promoters is similar between the cell types, with 324 and 271 H3K27me3 only marked genes in *in vitro* and adult islets, respectively (Figure 3.17). Differences were observed in the number of genes that are categorised as unmarked, or marked by both H3K4me3 and H3K27me3, between the cell types. In adult islets, only 90 genes are classified as containing both H3K4m3 and H3K27me3, in contrast *in vitro* derived pancreatic cells have 2618 genes with both modifications at the gene promoter. The analysis presented within the Bhandare et al, 2010 study supports these findings, as they concluded that H3K4me3 and H2K27me3 were almost mutually exclusive in adult islet cells (Bhandare *et al.*, 2010). There is also 7.8% fewer TSS that were classified as unmarked in the *in vitro* derived cells compared with adult cells, which likely accounts for the 7.05% more TSS that were classified as H3K4me3 and H3K27me3 modified in the *in vitro* cells compared to adult islets (Figure 3.18). Importantly for this analysis, adult islets are composed of a heterogeneous population of cells, which would also be expected in the day27, *in vitro* derived cells. Therefore, regardless of whether the H3K4me3 and H3K27me3 modified genes that are detected in the *in vitro* are caused by differences in marks between cell types within a mixed population or by both of the marks being present at a gene in the same cell, the levels of H3K27me3 in the *in vitro* cells are not observed in adult islets.

Although global analysis can indicate similarities and differences in the number of genes with each modification, it is less informative for determining how the modifications can vary at specific genes between the samples. To examine this, genes were assigned to one of the four categories based on histone modifications for adult islets and day27 *in vitro* cells – unmarked, H3K4me3 only, H3K27me3 only or H3K4me3 and H3K27me3. These gene lists were then analysed for the histone modification present in the other cell type to determine if the genes are marked by the same or different histone modification (Figure 3.17). This demonstrated that within categories that had a similar distribution between the cell types, there is still a variation in the genes which are marked by these modifications. For example, 21% of genes that are marked by H3K4me3 only in adult islets are marked by both H3K27me3 and H3K4me3 in the *in vitro* derived cells (2,246 genes, Figure 3.18). Similarly, of the 10,584 gene marked with H3K4me3 only in the *in vitro* derived cells, 2,842 genes (26.9%) of these same genes are unmarked in adult islet cells. Furthermore, this analysis was then used to determine which gene

modifications are different due to the difference in the percentage of unmarked and H3K4me3 and H3K27me3 co-modified genes between the cells. When the H3K4me3 and H3K27me3 co-modified genes in the *in vitro* cells are analysed in adult islets, only 3.5% (90) of the same genes are also marked by both H3K4me3 and H3K27me3. Instead, nearly all (87.9%; 2,301) of these gene are marked by H3K4me3 only, and 8.6% (224) genes are modified with H3K27me3 alone. Of the genes which are categorised as unmarked in adult islet cells, 11% (n=2748) of these are marked by H3K4me3 only at their promoters in *in vitro* cells. This detailed comparison of histone modifications, therefore, indicates both a difference in the distribution of histone modifications between adult islets and *in vitro* derived pancreatic cells, as well as in the number of genes which are differentially modified by activating and repressive histone modifications between the cells.

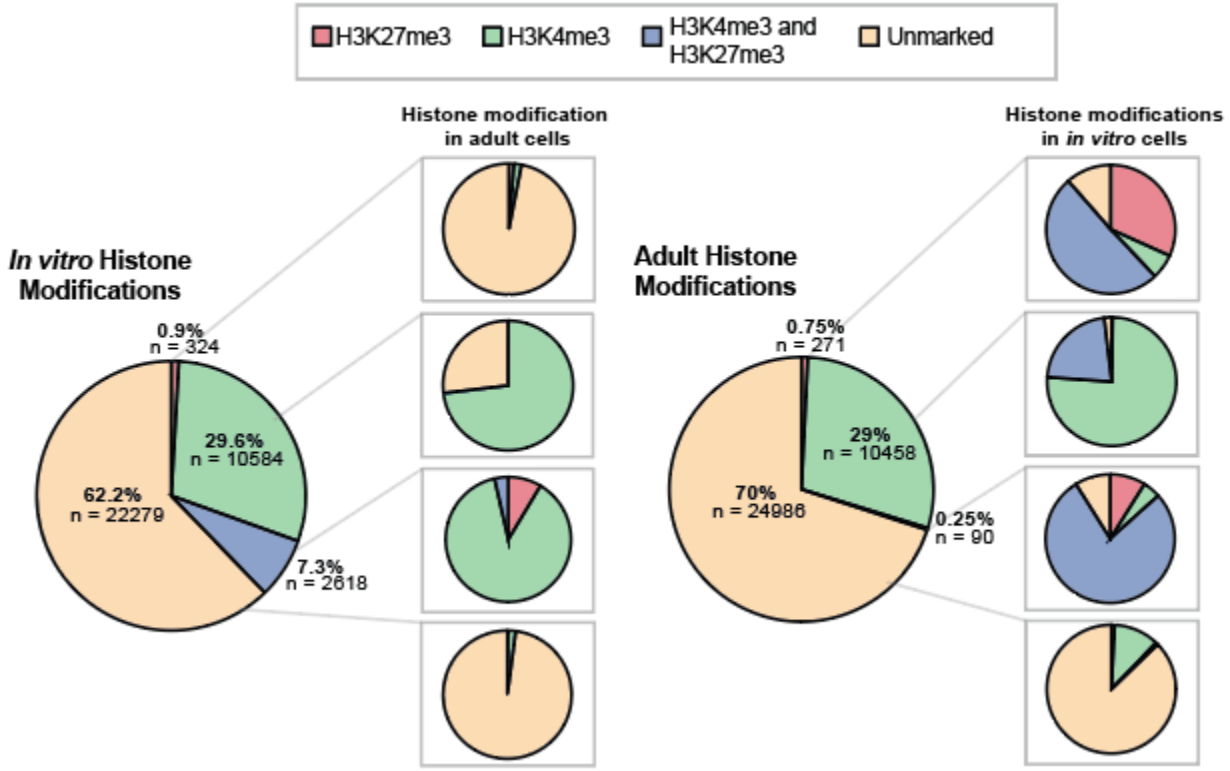


Figure 3.17 Comparison of H3K4me3 and H3K27me3 histone modifications between day 27 *in vitro* pancreatic-like cells and *in vivo* adult islet cells. The total percentage of all gene promoters marked by different histone modifications are shown. After determining the gene promoters marked by the different modifications in adult islets the same genes were analysed in *in vitro* cells to determine the histone modifications present at the gene promoters in these cells.

To investigate if these different histone modifications contribute to the differentially expressed genes which exist between the cell type, the transcriptional data and the histone ChIP-sequencing data were integrated for analysis. The genes that had previously been determined to be differentially expressed between day27 *in vitro* cells compared to the adult islets, in Figure 3.17 were then examined in further detail (Figure 3.18). The lists of 4,705 and 3,874 genes which are upregulated and downregulated in the *in vitro* cells compared with adult islets, respectively, were used. The histone modifications associated with the promoter of these genes in both adult islets and *in vitro* derived cells was determined, and the genes were categorised as before, into either H3K4me3 only, H3K27me3 only, H3K4me3 and H3K27me3 co-modified or unmarked (Figure 3.18). In addition, the same categorisation was performed for 3,000 randomly selected genes. A random gene list was included to provide an unbiased view of the expected percentage of histone modifications present in the genome, and therefore also allows comparison if the percentage of gene with histone modifications in the selected gene lists is necessarily higher or lower than the average coverage of the gene promoters.

A higher percentage of genes (53.4%) which are transcriptionally upregulated in *in vitro* derived cells compared with adult islets have H3K4me3 present at their promoter in day27 cells, as compared to the percentage of random genes (37.1%). In adult islets, these same genes have a lower-than-expected percentage that are modified by H3K4me3 (24.9%), compared to random genes (35.3%). This finding suggests that the presence of H3K4me3 at genes in the *in vitro* cells could contribute to the transcriptional upregulation of these genes compared with adult islets.

For the genes which are downregulated in *in vitro* derived cells compared with adult islets, the opposite pattern in H3K4me3 modifications is observed. The *in vitro* cells have a slightly lower than expected percentage marked by H3K4me3 (31.3% compared to 37% random genes) and adult islets have a higher percentage than expected (52.4% compared to 35.3% random genes) (Figure 3.19). Very few genes are modified by H3K4me3 and H3K27me3 in adult islets in any category (Figure 3.18). However, in the *in vitro* derived cells, we observe an enrichment of genes marked by both H3K4me3 and H3K27me3, which are downregulated in *in vitro* cells over adult islets. The H3K4me3 and H3K27me3 modification is present at 19.1% of downregulated genes, compared to only 6.3% of random genes, and 7.8% of upregulated genes. This may also suggest that the presence of H3K4me3 and H3K27me3 contributes to the downregulation of some genes in *in vitro* compared with adult. Therefore, the aberrant presence

of H3K4me3, and H3K27me3 and H3K4me3, is associated with genes that are differentially expressed between adult islets and *in vitro* derived pancreatic cells

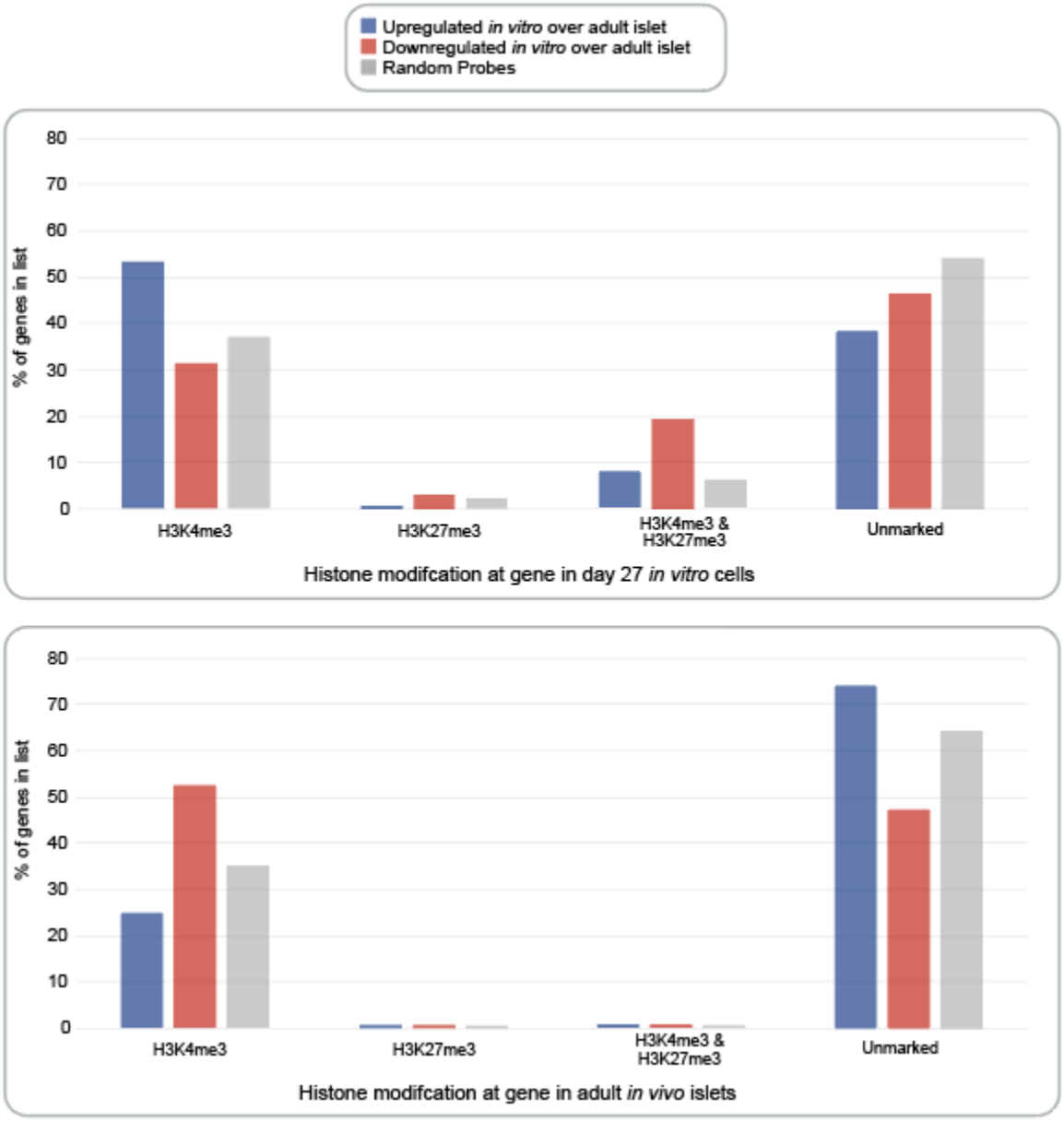


Figure 3.18 Enrichment of H3K4me3 and H3K27me3 modification at genes downregulated in *in vitro* derived cells over adult derived islets. The percentage of genes in three categories - upregulated, downregulated and random genes, marked by each of the 4 histone modification categories, in day 27 *in vitro* cells (top panel) and adults *in vivo* islets (bottom panel).

To determine if the presence of H3K4me3 and H3K27me3 at gene promoters could be implicated in the downregulation of genes that are important for the development and function of the adult pancreas, the *MAFA* gene was selected for individual locus analysis (Figure 3.20). The *MAFA* gene is produced exclusively within β-cells of the pancreatic islet (Aramata et al, 2007) and represents an example of a transcription factor which is critical for both β-cell development and maturation (Zhu et al, 2017). Interestingly, this key gene is significantly downregulated in the *in vitro* derived pancreas compared with adult islet, with very low expression of this gene in day 27 *in vitro* differentiated cells or in isolated foetal cells (Figure 3.20). Examining the ChIP-sequencing tracks for H3K4me3 and H3K27me3 throughout differentiation, both modifications are present over this gene promoter region at day 0 and persist throughout all subsequent timepoints. In contrast, the ChIP-sequencing tracks for adult islets demonstrates the presence of the H3K4me3 modification at the *MAFA* promoter but the H3K27me3 is not detected at the gene in the adult *in vivo* islet cells. This demonstrates an example of an important pancreatic regulator which fails to be upregulated during *in vitro* pancreatic differentiation, potentially due to the aberrant presence of H3K27me3 at the promoter of this gene.

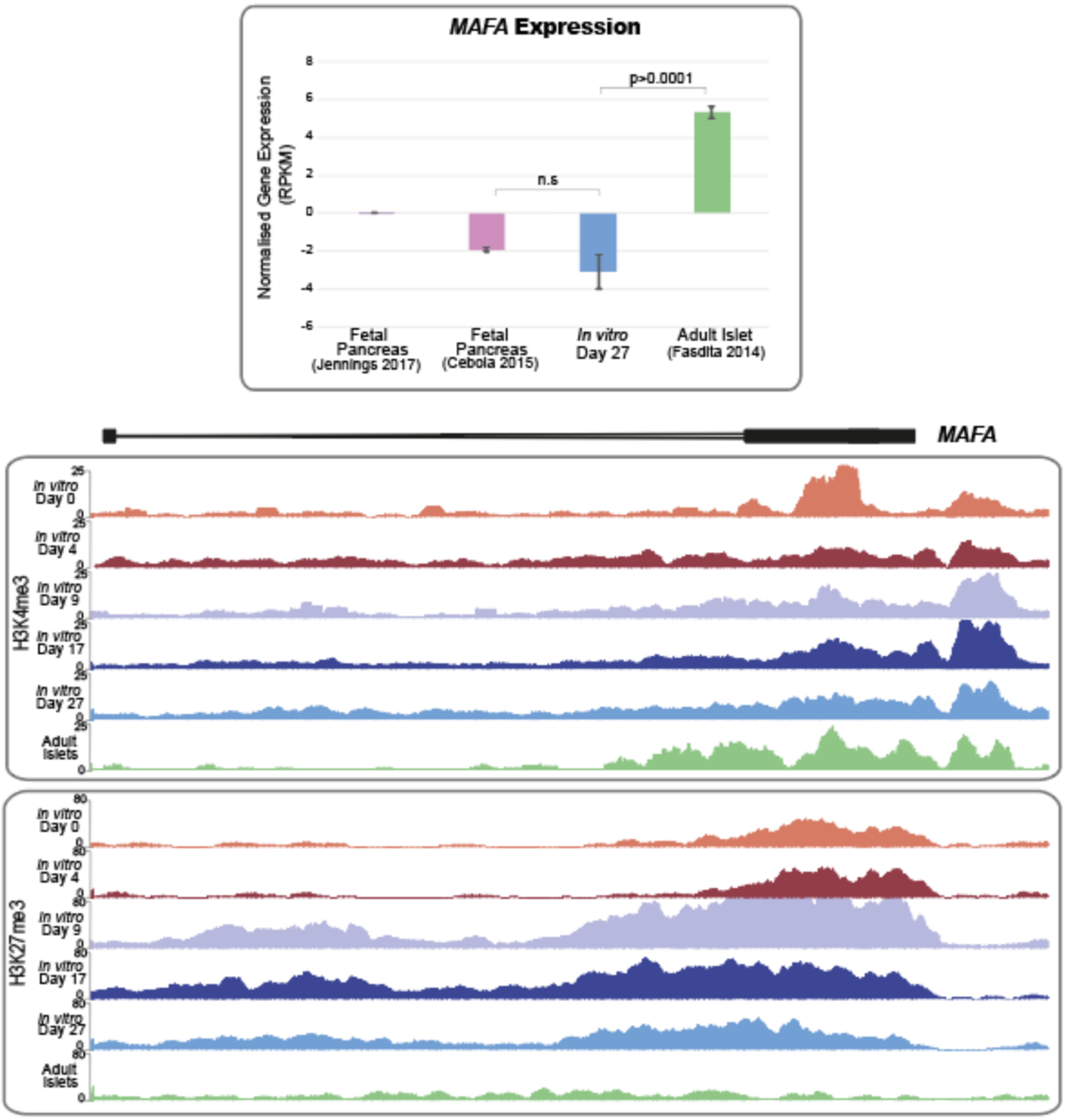


Figure 3.19 Examples of H3K27me3 and H3K4me3 modifications at pancreatic genes, which are differentially expressed between *in vitro* and *in vivo* cells. The RNA-sequencing expression of *MAFA* genes from different pancreatic sources of cells. ChIP-sequencing tracks for the genes during 5 stages of *in vitro* pancreatic differentiation and *in vivo* adult islets. Plots were produced using running window probes with read count corrected quantification. Data is an average of 3 replicates for *in vitro* differentiation and 4 *in vivo* adult islets donors.

The presence of H3K27me3 at gene promoters is predicted to have an effect on the transcriptional differences between adult islets and *in vitro* derived pancreatic cells, but gene promoters are not the only regulatory element in controlling levels of transcription. The presence of H3K27me3 at enhancer elements is implicated in hESCs with maintaining enhancers in a poised state, which is associated with inactive gene expression (Rada-Iglesias *et al.*, 2011). As the *in vitro* derived pancreatic cells appear to have significantly more promoter H3K27me3 compared to adult islets, the possible retention of aberrant H3K27me3 at active pancreatic enhancers was therefore analysed. Active enhancers that were defined by Ferrer et al, 2019 were used for this analysis. To classify enhancers, H3K27ac, H3K4me1 and mediator binding was used and, based on the levels of each of these features, enhancers were split into class I (13,635 regions), which had the strongest signal of all three features, class II (22,767 regions) which has mid-level of binding and class III (9,281 regions) which has the presence of all three features but at the lowest levels (Miguel-Escalada *et al.*, 2019). The number of designated enhancer regions which overlap with H3K27me3 modifications were then determined in both adult islets and *in vitro* cells (Figure 3.20a). This demonstrates that a substantial number of active pancreatic enhancer regions were marked by H3K27me3 in the *in vitro* cells, in all of the three enhancer categories. The number of enhancers with H3K27me3 present is much higher in *in vitro* cells as compared to the same enhancers within the adult islets, although it is a relatively small proportion of the total active enhancers modified by H3K27me3 in both cell types. As the total number of H3K27me3 modifications in the *in vitro* cells is considerably higher than in adult islets, the number of enhancers with H3K27me3 overlap was normalised to the total number of H3K27me3 modifications detected in both cell types (Figure 3.20b). With the normalisation applied, an enrichment of H3K27me3 overlapping with enhancers is still observed in the *in vitro* derived cells compared to adult islet cells. This indicates that there is a persistence of H3K27me3 at a proportion of enhancers in the day27 *in vitro* derived cells which may maintain the enhancers in the poised state, as opposed to an active enhancer state mainly observed in the adult islets, which may limit transcription of target genes.

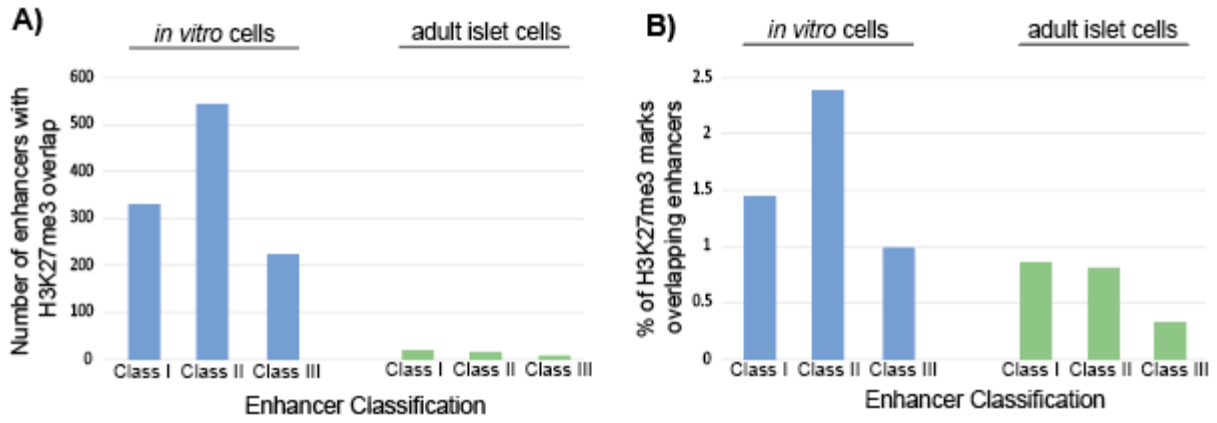


Figure 3.20 H3K27me3 overlaps active pancreatic enhancers in adult islets. Enhancer classifications are based on H3K4me1, H3K27ac and mediator levels as designated by Ferrer et al, 2019. a) The number of enhancers in each enhancer class that are overlapped by H3K27me3 in adult islets and *in vitro* derived pancreatic cells. b) The % of all H3K27me3 marks in adult islets and *in vitro* derived pancreatic cells which overlaps with enhancer regions in each class.

To examine if the presence of H3K27me3 at enhancers could be sufficient to have an effect on transcription of nearby genes, examples of enhancers that are modified differently between *in vitro* pancreatic derived cells and adult islets were selected for analyses (Figure 3.21). Enhancers can be located in a number of regions, some in close proximity and others distal from their target gene(s) (Bulger and Groudine, 2011). In Figure 3.21a, the enhancer regions identified are approximately 10-20kb from the nearest gene, *NKX2.2*. H3K27me3 modification in the *in vitro* derived pancreatic cells is present at much higher levels over both the class I and class II enhancer regions, compared to adult islets H3K27me3 and the input tracks of the matching samples. A similar, high enrichment of H3K27me3 in the *in vitro* derived pancreatic cells is also observed at the more proximal enhancers surrounding the *NEUROGENIN-3* endocrine pancreatic gene. H3K27me3 signal is present over the gene and extends past the gene border to also mark the enhancers that are approximately 2-10kb either side of the gene. These data exemplify the presence of H3K27me3 in *in vitro* derived cells but not in adult islets at regions that are designated as active enhancers in adult islets.

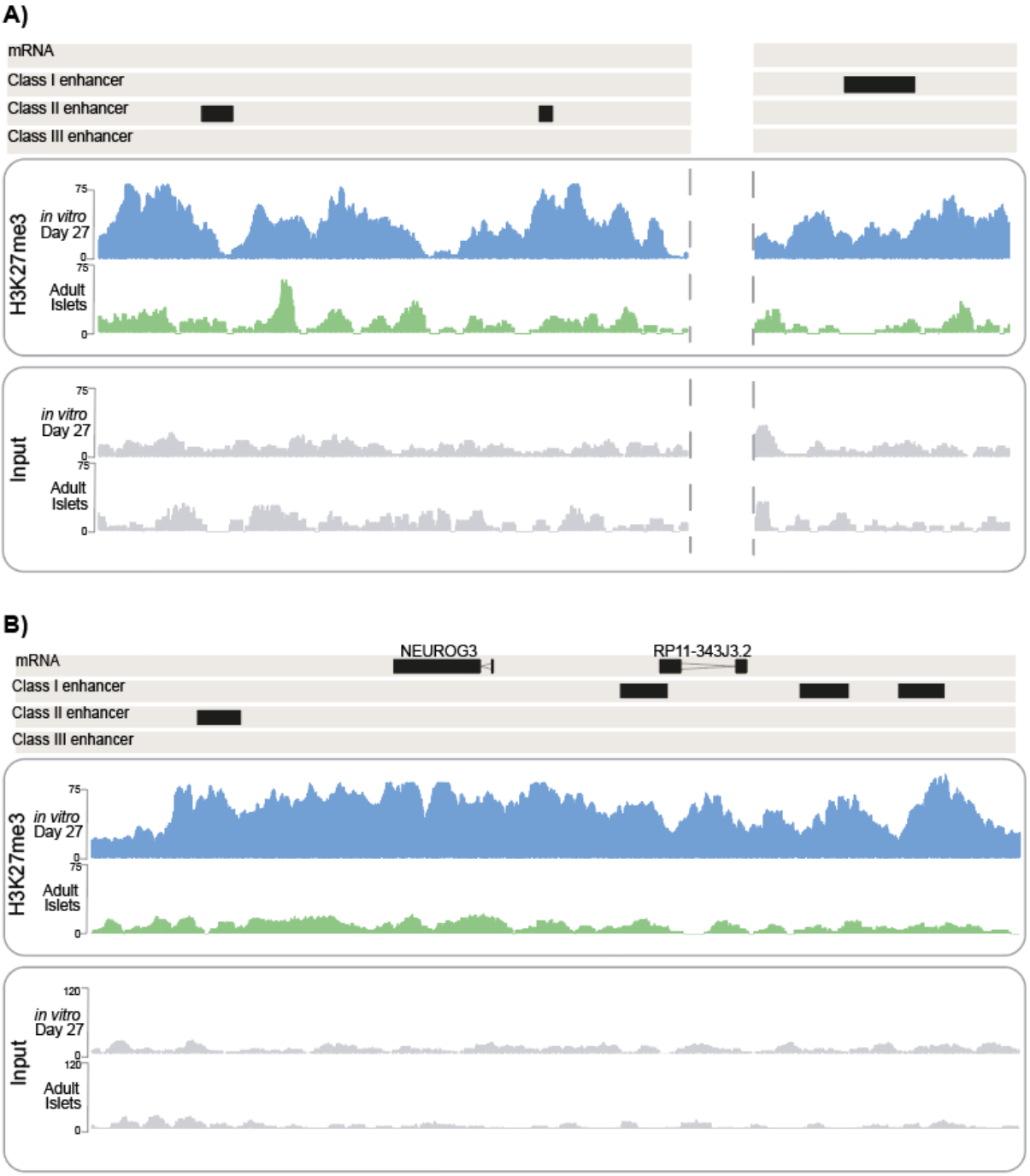


Figure 3.21 Examples of H3K27me3 overlap with active pancreatic enhancers in adult islets and *in vitro* derived pancreatic cells. ChIP-sequencing tracks for H3K27me and input of day 27 *in vitro* derived pancreatic cells and adult islets over enhancer regions, identified as described in Figure 3.20. Plots were produced using running window probes with read count corrected quantification. Data is an average of 3 replicates for *in vitro* differentiation and 4 *in vivo* adult islets donors.

3.4 Discussion and Conclusions

With efficient *in vitro* pancreatic differentiation protocols available which can produce cell types representative of mature pancreatic-like cells, these protocols offer great promise in the regenerative medicine field. However, a number of problems still exist in these cells formed and therefore further investigation into the final cell type provided and their *in vitro* development are required. In this chapter I undertake an *in vitro* differentiation of hPSCs to form pancreatic-like cells, isolating key stages of the differentiation to study the dynamics of the emergence of final cell population. From this, transcriptional dynamics associated with the differentiating cells were isolated and used to study when the cell will undergo the largest transcriptional changes. Associating histone modifications with these transcriptional changes highlighted unusual patterns of H3K4me3 and H3K27me3 persistence at some genes in the final cell types. The comparison of the day27 *in vitro* cells with adult islets cells further identified unusual patterns of histone modifications, which may partially contribute to the differentially expression of genes that exists between these two cell types.

3.4.1 Studying histone modifications dynamics during *in vitro* pancreatic differentiation.

The protocol in this, and previous studies, uses defined media composition based on signalling in human and mouse *in vivo* development, to differentiate a hPSC population into the desired mature cell type (D'Amour *et al.*, 2005; Jiang *et al.*, 2007; Shim *et al.*, 2007; Pagliuca *et al.*, 2014; Rezania *et al.*, 2014; Russ *et al.*, 2015). During this *in vitro* differentiation, a definitive endoderm population is present by day 4, a pancreatic endoderm population formed by day 9, while the presence of an endocrine progenitor population observed by day17 and by day27 the cells formed are capable of producing pancreatic hormones, indicating mature endocrine cell types are present. However, as previously reported with *in vitro* pancreatic differentiation protocols, the cells produced at day 27, were polyhormonal, had significantly less gene expression of pancreatic hormone genes than adult *in vivo* derived islets and transcriptionally closer resembled foetal pancreatic cells compared to adult (Kroon, Laura A. Martinson, *et al.*, 2008; Schulz *et al.*, 2012; Pagliuca *et al.*, 2014; Bertero *et al.*, 2016). Therefore, the isolation of intermediary timepoints during differentiation allowed the study of normal transcriptional dynamics that occur in a prolonged *in vitro* differentiation protocol, determining areas of improvement in the current protocol.

The H3K4me3 and H3K27me3 modifications analysed in this study are used as markers of active and repressive chromatin, respectively (Allis & Jenuwein, 2016; Hyun., 2017). To determine when the H3K27me3 or H3K4me3 histone modifications are both gained or lost during *in vitro* differentiation, in respect to genes undergo dynamic transcription, these with histone marks were associated with selected genes. The histone modifications at these genes in the different cell populations was examined before and after expression changes. Interestingly, when analysing a transcriptional dynamic gene set, histone modifications did not predominantly change at any specific time point in relation to the gene expression changes. Instead, the predicted histone modification changes would be obtained gradually throughout the differentiation, including time points after which gene expression changes had already occurred. This contrasts with an earlier study, which have observed a potential pre-patterning by the repressive histone modification H3K27me3 or the activating histone modification H3K4me3 in progressively forming cell populations. This study utilising the liver/pancreatic lineage choice from endoderm progenitors, suggested a role of histone modifications, along with other factors in pre-patterning chromatin to assist in lineage differentiation. Although this pre-patterning was not considered essential for the resulting cell type to be formed, removal of these factors did cause a reduction of specific populations (Xu *et al.*, 2011). As this pre-patterning did not appear to have a significant effect on the cells in which it was observed it is unclear if it is important to the progress of the cells, and therefore if it should be present in our cells, which was not seen.

3.4.2 The persistence of the H3K27me3 modification during *in vitro* pancreatic differentiation.

In our analysis when attempting to associate transcriptional dynamics with histone modifications, we observed the gain of H3K27me3 at the promoters of genes which are being downregulated. In clusters E and F, in which gene sets are downregulated, a number of the genes transitioned from H3K4me3 only to H3K27me3 and H3K4me3 modified. Although H3K27me3 is a repressive modification, which could account for the gain of the modification at promoters of genes being transcriptionally silenced, almost no increase is observed in unmarked genes. It had previously been hypothesised that resolving bivalent genes into the repressed state will result in the addition of DNA methylation or H3K9me2 (Meissner *et al.*, 2008; Wen *et al.*, 2009). Both of these epigenetic modifications are associated with repressed genes in mature differentiated cells, as opposed to the H3K27me3 which is typically associated with ‘poised’ genes in the pluripotent state. If the genes that underwent loss of expression

during the *in vitro* differentiation were to be efficiently silenced by these hypothesised repressive modifications then the genes would be expected to transition from the active H3K4me3 to the unmarked category, as other epigenetic modifications, which were not analysed in this study, were added. Instead, apparent gain of H3K27me3, and maintenance of H3K4me3 is an unexpected result. However, as previously mentioned, the heterogeneity of the differentiating cell population, makes it unclear if these opposing modifications are present at a single locus in the same cell or if the presence of both modifications is caused by a population, where different cells in the population will have a single modification at the same gene, see Figure 1.3. Overall, both situations would produce the same histone modification patterns when observing a population, but these would have different consequences for the genes marked. The presence of both modifications in a single cell, to form a bivalent marked histone, will lead to the gene being maintained in a poised state, in which transcription is low but still possible from the gene, and the gene can respond quickly to cues to induce activation or silencing (Bernstein *et al.*, 2006; Mikkelsen *et al.*, 2007; Pan *et al.*, 2007; Zhao *et al.*, 2007). The second possibility is the presence of activating H3K4me3 marks in some cells and the H3K27me3 modification at the same gene promoter but in different cells. The consequence for this would be that some cells continue to express the gene in those with the active mark, but those that had gained the H3K27me3 would silence this expression. As the genes in the downregulated categories are moving from an expressed to silent state, this may suggest that the genes are being inefficiently silenced in the population, regardless of the situation which leads to the resulting H3K4me3 and H3K27me3 patterning observed. With the presence of a heterogeneous population, it is not possible to distinguish between these possibilities, with the current data available, although further experiments utilising single cell analysis may be able determine which of these combinations is accurate.

The existence of heterogeneity of the cell populations is a limitation throughout this study as it could produce confounding results in analysing cells at a population level, but this can be overcome in some analyses through different means. Firstly, the heterogeneity in the population will generally increase as we progress in the differentiation time points. At initial time points, such as day 0, a relatively homogenous population of pluripotent cells is present, but throughout the 27 days of differentiation a more heterogenous population of cells develops, with heterogeneity likely increasing at each time point analysed. As a result, the 2,957 genes determined as H3K27me3 and H3K4me3 modified at day0 are likely to be bivalent, with similar numbers of bivalent marked genes determined in previous studies (Mikkelsen *et al.*, 2007; Pan

et al., 2007; Zhao *et al.*, 2007). From this, it is not possible to determine if the H3K4me3 and H3K27me3 marked genes that persist in all stages of differentiation, at similar percentages, are bivalently marked in a single cell. However, a large percentage, over 70%, of the genes modified by H3K27me3 in later time points also have this modification present at day 0. Of all H3K27me3 modified genes at day27, approximately 90% of these are present at genes also marked by H3K4me3, which is similar to that seen at day 0, with approximately 93% of all H3K27me3 modified genes existing in the bivalent state. Therefore, a large number of the genes H3K27me3 modified at day27, have the persistence of this mark from day 0, at which point most genes are bivalently modified, and with a high H3K4me3 retention also seen at genes from day 0, it may be hypothesised that this bivalency is at least partially maintained. Although experimentally not confirmed, this data may suggest a percentage of genes exist well into day 27 of differentiation with bivalent promoter's present. Another way in which the problem with heterogeneity is limited, is through the use of adult islets as the *in vivo* comparison, as opposed to using any sorted cell population or single cell data. Although the heterogeneity may confound the ability to draw conclusions from the differentiating cells, the heterogeneity seen in the day27 cells should be comparable to that in the adult islets, as both will have multiple cell types present. The use of adult islets data from multiple donors will contribute additional heterogeneity in the comparison samples, as donor to donor variation can be significant in terms of cell type composition of islets, for example the percentage of β-cells has been reported to range between 28% and 75% (Brissova *et al.*, 2005). Therefore, differences determined *in vitro,* and *in vivo* samples are likely to be of biological importance, as both are equally heterogeneous and these differences that arise are not caused by comparing a heterogeneous population with a homogenous population.

3.4.3 A potential role of the H3K27me3 modification in limiting maturation of *in vitro* derived pancreatic cells.

When the day27 *in vitro* cells and adult islets are compared, the number of genes with H3K27me3 modification is drastically different, with many more H3K27me3 modified genes in day27 cells. The majority of these genes with H3K27me3 also had H3K4me3 present, something that is observed at very few genes at adult islets. This would suggest that a potential problem within the *in vitro* differentiation occurs due to the inability to resolve the H3K27me3 modification of progenitor cell types. An earlier study by Xie et al, 2014 supports these observations, as in this study the authors grafted *in vitro* differentiated pancreatic endocrine

progenitor stage cells into a mouse using semipermeable microcapsules, after 20 - 22 weeks these grafted cells were isolated and analysed. After engraftment, the cells were transcriptionally, more similar to isolated adult islets over fully *in vitro* derived pancreatic endocrine cells. When transcriptionally comparing the *in vitro* derived cells with the *in vivo* matured cells, it was noted that a large percentage (48%), of the genes that are more highly expressed in the *in vivo* matured cells continue to be marked by H3K27me3 in the *in vitro* derived cells. A number of these genes highly expressed *in vivo* matured cells also did not gain H3K4me3 in *in vitro* derived endocrine cells. This study also showed, that during *in vitro* pancreatic differentiation from hPSCs to the pancreatic endoderm the percentage of bivalent genes, those marked by H3K4me3 and H3K27me3, had very little variation (Xie *et al.*, 2014). The authors therefore concluded that the retention of H3K27me3 at gene promoter during *in vitro* differentiation may be partially responsible for the inaccurate transcriptional signature observed in *in vitro* pancreatic endocrine cells, with a similar hypothesis suggested by the data in this analysis.

The H3K27me3 modification does not only appear to persist at the gene promoters, but higher levels are also observed at enhancer regions as compared to *in vivo* adult islets. Enhancers are defined by the presence of p300, BRG1 and H3K4me1 and then can be classified as active by the presence of H3K27ac or poised due to the presence of H3K27me3 (Rada-Iglesias *et al.*, 2011). A previous study focusing on enhancer formation in human *in vitro* differentiations demonstrated that enhancers will typically originate as unmarked chromatin, before transitioning to a poised state, which is associated with the developmental competency of the cells. With continued differentiation these enhancer regions will then gain H3K27ac (Wang *et al.*, 2015), that due to the presence on the same lysine as H3K27me3, these modifications must exist in a mutually exclusive manner (Zhang, Cooper, & Brockdorff, 2015). Therefore, the existence of the H3K27me3 at a number of enhancers identified as active in adult islets cells, may indicate that these cells have formed a poised state of the enhancers, showing the developmental competency of the cells, but have failed to resolve the H3K27me3 modification to transition into active enhancer state. However, this persistence of H3K27me3 does only occur at a small percentage of active islet enhancers. It would therefore be interesting to integrate this with other data sets such as Hi-C data to determine if these enhancers in the adult islets will interact with genes important in the function of the cell. The inability to activate enhancers, in combination with the repressive H3K27me3 modification at gene promoters could

further contribute to the inability of the *in vitro* derived pancreatic cells to replicate the transcriptional signature of the adult pancreatic islet cells.

It is known that the *in vitro* derived cells are not comparable in terms of functionality to adult islets, and therefore it could be hypothesised that the difference in the transcriptional signature between *in vitro* derived cells and adult islet cells are the basis for this. The data obtained in this study and previous studies (Xie *et al.*, 2014), indicate that H3K27me3 persistence at gene promoters and enhancers in *in vitro* derived pancreatic cells could contribute to the different transcriptional signature seen when compared to adult islets. An important example, demonstrated in this study, is the expression of *MAFA* gene, which is a gene known to be important in the mature β cells (Zhu et al., 2017). It was shown that the promoter of this gene maintains H3K27me3, from day0 into day27, and the gene maintains a repressed state in all the cell types. This contrasts the adult islet cells which have no H3K27me3 present at the *MAFA* promoter, but the gene was active in these cell types. The reasons for the aberrant histone modifications and therefore the cause of this inaccurate transcriptional signature, are unknown but could be due to a number of underlying reasons. When the *in vitro* derived cells are compared to both foetal and adult islets, in both this present and in previous studies, the *in vitro* derived cells will, transcriptionally, more closely align to foetal pancreatic cells (Siniša Hrvatin *et al.*, 2014; Veres *et al.*, 2019). Therefore, it could be hypothesised that aberrant histone modifications maintain the end cell type in a foetal phenotype, which are suspected of being functionally different from adult islets, with little to no insulin response to glucose stimulation (Otonkoski *et al.*, 1988; Rorsman *et al.*, 1989; Stolovich-Rain *et al.*, 2015). Although ChIP-sequencing data was not available for foetal pancreatic cells, a study has attempted to investigate the number of histone modifications present in other foetal cell types (Yan *et al.*, 2016). When interrogating the histone modifications in foetal brain, heart, and liver of 12 weeks gestation human embryo, they determined that the global genomes of the foetal tissues will have more H3K27me3 peaks than adult tissues. Utilising sequential ChIP-sequencing, it was demonstrated that all three foetal tissue types will have a number (minimum of 780) of bivalently marked promoters. Although pancreatic foetal tissue was not included in this study, three different foetal cell types appeared to show an enrichment for H3K27me3 compared to adult cells, and all had a presence of several H3K27me3 and H3K4me3 marked promoters. Therefore, the continued presence of high H3K27me3, particularly at sites also possessing H3K4me3 may indicate that the histone modifications are contributing to a foetal

phenotype which is restricting the differentiation of the *in vitro* cells to mature adult pancreatic-like cells.

Another hypothesis that could link the aberrant histone modifications to transcriptionally and functionally immature, is due to the presence of progenitor cells in the population. It is known that partially lineage restricted progenitor cells will maintain bivalent modifications (Cui et al., 2008; Roh., 2006). Therefore, the continuous presence of the H3K27me3 modification could be caused by the existence of a progenitor population at day27 that has not yet terminally differentiated. Evidence for the presence of this progenitor population can be seen in the day 27 differentiated cells, when analysing protein level hormone expression by immunofluorescence. Only a small percentage of the cell are positive for the expression of pancreatic hormones, suggesting the other cells are not yet at the mature stage of hormone secretion. Transcriptionally the cells will also continue to express genes associated with pancreatic progenitors, although at lower levels than the day 17 time point at which pancreatic progenitors are expected. Therefore, the aberrant histone modifications and partial functionality of the *in vitro* pancreatic cells may be caused by an inefficiency with the differentiation protocol which causes the existence of a significant progenitor population within the day 27 cells.

The pancreatic endocrine cells produced *in vitro* are known to be distinctive from the endocrine cells obtained from adult islets. This study supports previous data in demonstrating both, the foetal transcriptional signature of the *in vitro* derived cells and the presence of aberrant histone modifications, which are potentially detrimental to the maturation of the pancreatic differentiation. These observations may, in part explain the functional immaturity of the *in vitro* derived cells, including the existence of polyhormonal cells and reduced insulin secretion abilities. It could also be hypothesised from this data that the perturbation of histone modifications, focusing most strongly on the H3K27me3 repressive modification, may positively benefit the production of pancreatic islet like cells through *in vitro* differentiation.

Chapter 4

The Effect of PRC2 Inhibition on *in vitro* Differentiating Human Pancreatic Cells.

4.1 Background

As observed in this thesis and in previous studies (Xie *et al.*, 2014), the presence of aberrant H3K27me3 histone modifications in pancreatic endocrine cells may restrict the form and function of mature cell types. When pancreatic endocrine cells are produced *in vitro*, they have a transcriptional signature more similar to foetal pancreatic tissue compared to adult cells (Figure 3.14). This difference may partially underlie the functional defects observed in the day 27 derived cells. H3K27me3 is added to histones through the activity of a single complex, PRC2 (Pengelly *et al.*, 2013). This complex is composed of multiple core and accessory subunits, described in detail in section 1.3.3. All forms of the PRC2 complex rely on the same catalytic component - the EZH protein, which has two homologues, EZH1 and EZH2 (G. van Mierlo *et al.*, 2019; Yu *et al.*, 2019). The EZH1 is typically associated with more terminally differentiated cells whereas EZH2 is enriched in actively dividing cells and has a higher catalytic activity, EZH2 is therefore the main catalytic subunit of PRC2 in embryonic development (Margueron *et al.*, 2008; Lee *et al.*, 2018). Through the inhibition of this catalytic EZH component of PRC2 it is possible to directly affect H3K27me3 levels in a cell.

A previous study targeted PRC2 to determine whether decreasing H3K27me3 through EZH inhibition affected pancreatic differentiation (Xu *et al.*, 2014). Xu and colleagues first used a conditional *Ezh2* mouse model to deplete *Ezh2* at the pancreatic progenitor stage in developing mouse pancreas. The genetic depletion of *Ezh2* led to an enhanced endocrine cell induction and a greater number of β cells produced. To determine if a similar result can be achieved using small molecule inhibitors targeting Ezh2, 3-deazaneplanocin A (DNZnep) a methyl donor pathway inhibitor that affects Ezh2 activity, and GSK-126, an Ezh2-specific inhibitor, were also used by Xu and colleagues. Inhibitor treatment of *in vitro* culture of *ex* vivo embryos, indicated that PRC2 inhibition can cause an increase in the pancreatic progenitor and endocrine progenitor population in mouse embryo explants. Lastly, this same study used

DNZnep and GSK-126 to determine if the inhibition of EZH2 during human *in vitro* pancreatic differentiation had an effect on the cell produced (Fiskus *et al.*, 2009; McCabe *et al.*, 2012). In this study, a self-renewing endodermal progenitor starting cell population (Cheng *et al.*, 2012) was utilised, which can be efficiently differentiated to pancreatic endocrine-like cells, during which Xu et al inhibitor treated cells at various points during differentiation. The results indicated that EZH2 inhibition during early timepoints in differentiation have a small inhibitory effect on the formation of C-peptide positive cells. In contrast, inhibiting EZH2 during later timepoints led to an increase in the C-peptide positive population at the end of differentiation and increased *NGN3* expression immediately after inhibition (Xu *et al.*, 2014).

The study by Xu and colleagues produced very interesting proof-of-principle data and can be considered as further evidence that the presence of H3K27me3 in the differentiated population may act negatively against pancreatic endocrine production and function. However, many important questions remain unanswered. Although Xu et al. demonstrated an increase in the number of cells that are C-peptide positive, little analysis was undertaken on the final cell types produced after EZH2 inhibition. It is therefore unclear if increased C-peptide positive cells would cause increased functionality in terms of secreted hormone levels and the glucose responsiveness. It is also not clear the effect these changes might have on other pancreatic endocrine cell types that are present in the differentiated population. This study also did not analyse the effect of transient EZH2 inhibition on global transcription, or on the effect that a decrease in global H3K27me3 might have on histone modification levels at the end point of cell differentiation. Therefore, I chose to undertake an in-depth analysis of the effects of PRC2 inhibition during the *in vitro* differentiation of pancreatic endocrine cells.

4.2 Hypothesis and specific aims

As previously discussed, the *in vitro* differentiated human pancreatic cells retain a foetal pancreatic phenotype. The results obtained in Chapter 3 allude to a possible role of histone modifications in contributing to this phenotype. It remains unclear whether the histone modification differences are a cause or consequence of the foetal transcriptional signature associated with the *in vitro* derived cells.

I therefore chose to analyse the effects of H3K27me3 loss through controlled PRC2 inhibition on *in vitro* differentiating human pancreatic cells. Through the use of small molecule inhibitors, the catalytic EZH component of the PRC2 complex can be used to decrease H3K27me3 at specific timepoints during the development of pancreatic endocrine cells. It is hypothesised that this epigenetic modulation may release the transcriptional repression of key pancreatic genes, and thereby promote cells to further maturation towards an improved pancreatic endocrine phenotype.

Specific aims:

1. To optimise EZH2 inhibition in the differentiating cell population.
2. To determine if PRC2 inhibition has a functional effect on the *in vitro* differentiated human pancreatic cells.
3. To analyse the transcriptional effect of PRC2 inhibition during pancreatic *in vitro* differentiation.
4. To analyse the effect of PRC2 inhibition on histone modifications during a prolonged *in vitro* differentiation protocol.

4.3 Results

4.3.1 Use of small molecule inhibitors to decrease PRC2 activity in hPSCs.

In order to define the consequence of H3K27me3 loss in differentiating pancreatic cells, small molecule inhibitors specifically designed to target the PRC2 complex were utilised, which should decrease H3K27me3 deposition within a cell. Because the catalytic activity of PRC2 has been associated with cancers (Kim and Roberts, 2016) several small molecule inhibitors targeting EZH1 and EZH2 have been designed. I chose to use GSK343 and UNC1999 , which target EZH2 only, and EZH2 and EZH1, respectively. GSK343 will specifically target only EZH2, with a more selective inhibition of EZH2 over EZH1 (over 50-fold) by acting in competition with the cofactor S-Adenosyl methionine, which is required for methyl group transfer (Verma *et al.*, 2012). UNC1999 inhibits EZH1 and EZH2 activity by competitively binding with the S-Adenosyl methionine cofactor, and this inhibitor only has ~10-fold greater specificity for EZH2 over EZH1, and is therefore considered an efficient inhibitor for both EZH proteins (Konze *et al.*, 2013). As these inhibitors were originally designed for therapeutic treatment in human cancer cells, initial experiments were designed to optimise their use in reducing H3K27me3 levels in cells that are used in pancreatic differentiation.

Initial experiments utilised undifferentiated HDF hPSCs to determine the maximal concentration of both UNC1999 and GSK343 that could be applied while still allowing cell growth and survival (Figure 4.1). In all experiments using the small molecule inhibitors, the addition of DMSO is used as a control comparison. Both inhibitors are resuspended in DMSO, and to ensure that DMSO does not have a direct effect on the cells, it is added at the same dilution as the highest small molecule inhibitor. Inhibitors were initially trialled at four different concentrations – 2, 5, 10 and 20 µM, to determine the highest concentration at which GSK343 and UNC1999 will not cause excessive cell death. Most concentrations tested were comparable to DMSO controls in terms of cell number and morphology, however very little cell survival was observed with 10 and 20 µM UNC1999 and with 20 µM GSK343 (data not shown). It is known that the genetic loss of PRC2 in hPSCs will cause a loss of self-renewal and a decrease in cell number (Collinson, Amanda J Collier, *et al.*, 2016; Shan *et al.*, 2017), suggesting that 20 µM GSK343 and 10 µM UNC1999 have inhibited PRC2 to levels which can no longer support the cell growth and survival. As there was very little cell survival under these conditions, not enough material could be obtained for analysis these cells. The small molecule inhibitors act by blocking EZH protein, through competition with the S-adenosyl-L-methionine cofactor, therefore no

difference will be seen in EZH1/2 protein levels, and the best read-out for the activity of the inhibitors is the resulting changes in H3K27me3 levels. PRC2 is the only identified molecule capable of catalysing trimethylation of H3K27 (Margueron and Reinberg, 2011) and H3K27me3 levels should therefore reflect PRC2 activity. In the treatment conditions which had sufficient cell survival for analysis, histones were isolated and the levels of H3K27me3 were analysed using western blot (Figure 4.1a). As expected, the higher concentrations of inhibitor resulted in lower levels of H3K27me3 in the cell populations. At 2µM and 5µM concentrations of both inhibitor the an almost complete reduction is observed in the levels of H3K27me3.

To determine if the H3K27me3 reduction observed after small molecule addition is sufficient to cause a loss of gene repression, the expression of specific genes was analysed. In hPSCs, H3K27me3 is typically present at developmental genes, including genes which are involved in the development of all three lineages (Mikkelsen *et al.*, 2007; Pan *et al.*, 2007; Zhao *et al.*, 2007). Genes which would normally be modified by H3K27me3 and transcriptionally repressed in hPSCs were selected for analysis after small molecule inhibition. The addition of either GSK343 or UNC1999 caused the loss of gene repression in hPSCs at all concentrations tested (Figure 4.1b). The higher the concentration of inhibitors used, typically caused a higher upregulation in expression for both inhibitors. This demonstrated that the level of H3K27me3 decrease observed in the cells was sufficient to remove the gene repression. The six genes selected for analysis are representative of genes important for all three lineages. The gene upregulation demonstrates that de-repression is likely to be global, with all six genes demonstrating an increased expression, however the levels of upregulation are different between the genes. It can be seen that the increase of expression in *GATA4* and *FOXA2* are at much higher fold levels as compared with the other four genes analysed. Therefore, GSK343 and UNC1999 applied at 5 µM are capable of decreasing H3K27me3 and thereby cause gene de-repression, to varying levels, in hPSCs.

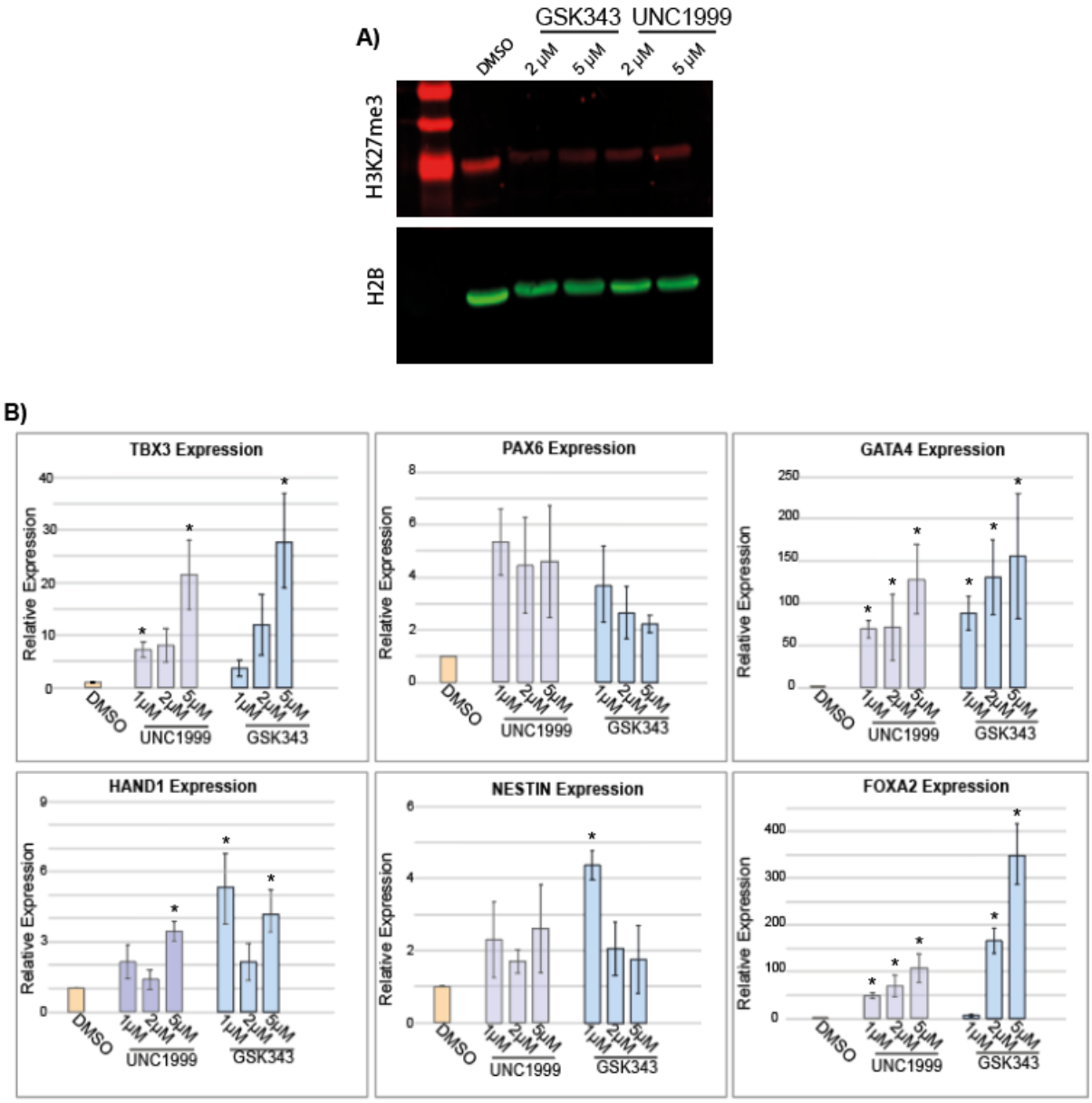


Figure 4.1 Loss of H3K27me3 and gene repression in hPSCs with the addition of PRC2 small molecule inhibitors. A) Histone blot analysis of H3K27me3 and H2B levels after 72 hrs of small molecule addition. Image representative of two separate experiments. B) RT-qPCR expression of lineage markers from hPSCs after 72 hrs treatment with EZH inhibitors UNC1999 and GSK343 at varying concentrations or a DMSO control. Expression was calculated as relative expression to DMSO control. Data shows the mean of 3 biological replicates ± SD, * indicates a p-value ≤0.05 when compared with DMSO by one-way ANOVA with post-hoc Tukey HSD.

When optimising the concentration of GSK343 and UNC1999 treatment in hPSCs, cells were treated for 72 hrs before analysis, but it is unclear if this is the most optimal time of treatment to cause maximal inhibition. The 72-hour treatment was selected because hPSCs have a self-renewal time of approximately 15 -16 hrs (Becker et al, 2006) and would therefore allow 4 -5 divisions during this time. Divisions are important for the functional outcome of the small molecules, as H3K27me3 is not actively removed but new modifications will not be added following cell division, which will cause the decreasing H3K27me3 levels. The 72-hour treatment caused loss of H3K27me3 to levels in which the gene repression was no longer maintained. To determine the difference with varying treatment lengths – 48, 72 and 96 hrs of treatment were analysed for gene de-repression with both GSK343 and UNC1999 (Figure 4.2). The addition of either inhibitor caused gene de-repression after 48 hrs of treatment, for all six genes analysed, for both GSK343 at 10 µM and UNC1999 at 5 µM. Similar, or slightly higher levels of de-repression, is seen with 72- and 96-hrs treatment for all six genes, therefore a minimum inhibition time of 48 hrs is suitable for de-repression with the same levels sustained until at least 96 hrs of inhibitor treatment.

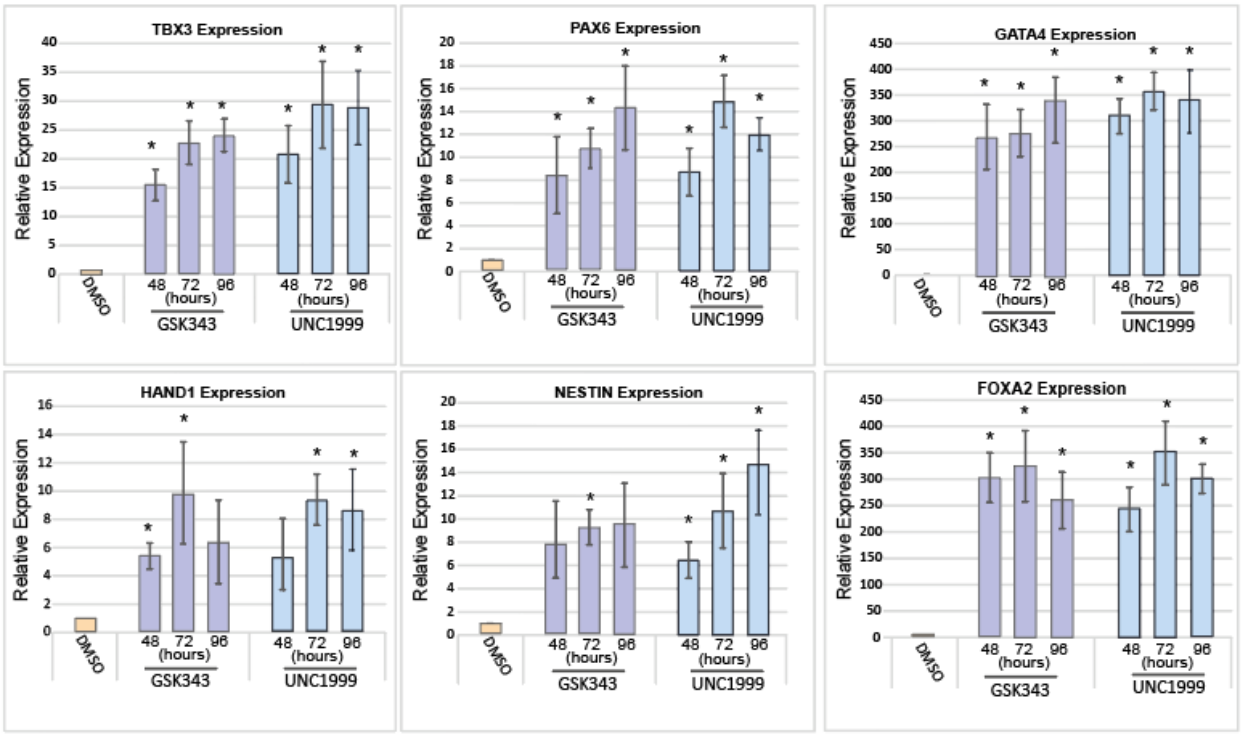


Figure 4.2 Gene repression by small molecule inhibitors is effective by 72 hrs. RT-qPCR expression of lineage markers in hPSCs after 48-, 72- and 96-hrs treatment with PRC2 inhibitors UNC1999 and GSK343 or a DMSO control. Expression was calculated as relative expression to DMSO control. Data shows the mean of 3 biological replicates ± SD, * indicates a p-value ≤0.05 when compared with DMSO by one-way ANOVA with post-hoc Tukey HSD.

4.3.2 Use of small molecule inhibitors targeting PRC2 in differentiating pancreatic cells.

After confirmation of the ability of PRC2 small molecule inhibitors to remove H3K27me3-associated gene repression in hPSCs, the next step was to determine the expression of PRC2 during the *in vitro* differentiation. A previous study (Pethe, Nagvenkar and Bhartiya, 2014), has reported changes in levels of both transcript and protein of PRC1 and PRC2 components during *in vitro* pancreatic differentiation. To determine if this occurs during the differentiation used in this study, and therefore if potential stages during differentiation at which PRC2 may have more effect, the subunits of the PRC2 complex, both core and variable, were analysed throughout human pancreatic differentiation using the previously obtained RNA-sequencing data (Figure 4.3). PRC2 is formed of multiple core and accessory subunits, described in chapter 1.3.3, which

can be analysed throughout the 27 days to determine timepoints at which their expression changes, indicating possible points where the complex may be of more or less importance. Analysis of the core components demonstrated that there is very little or no change of the core subunits between the timepoints. Similarly, expression of almost all accessory subunits, regardless of their association with PRC2.1 or PRC2.2, is consistent throughout the *in vitro* pancreatic differentiation. The only exceptions to this pattern are *JARID2*, which decreases in expression from day 0 to later timepoints, and *EPOP,* which becomes downregulated from day 17 to day 27. In the case of *JARID2*, the expression is still retained at high levels and is therefore unlikely to change complex formation between these timepoints, based on the gene expression although this does not indicate any changes in complex formation at the protein level (Figure 4.3). The expression of *EPOP,* an accessory subunits of PRC2.1 (G. van Mierlo *et al.*, 2019), changes throughout differentiation, and no change is observed in other subunits of this complex, therefore it is unclear if this would have any effect on the complex formation. By studying the PRC2 subunit gene expression, it is unclear if there is any point during *in vitro* pancreatic differentiation that this complex plays a more or less important role in the cell population.

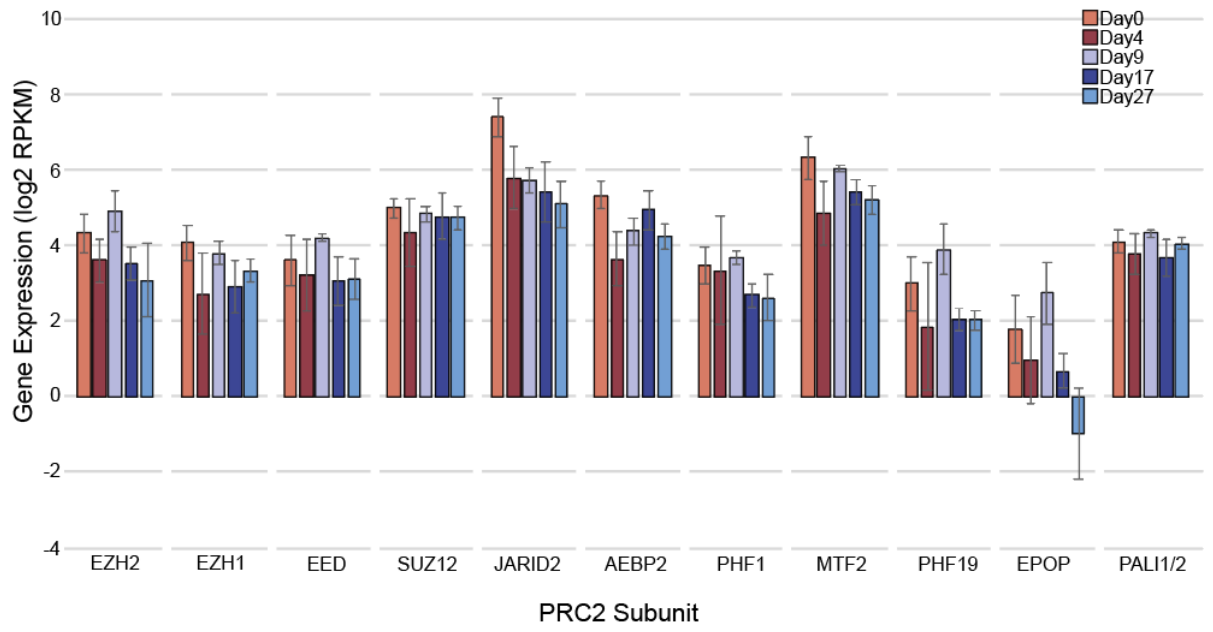


Figure 4.3 Expression of PRC2 components throughout *in vitro* pancreatic differentiation. mRNA expression levels of PRC2 core and accessory subunits at 5 timepoints in differentiating *in vitro* pancreatic cells. Data show the mean of 3 biological replicates ±SD.

As there is no timepoint in which PRC2 subunit expression was significantly changed throughout differentiation, previously published data was used to determine the eventual inhibitor treatment protocol (Figure 4.4). In the first four days of differentiation, cells transition from iPSC to definitive endoderm. This is a highly efficient and well-established process with the vast majority of cells expressing the desired cell markers (D'Amour *et al.*, 2006). Adding EZH inhibitors at this stage may negatively affect the formation of the endoderm lineage, as it is likely to cause non-specific upregulation of multiple cell lineages and would therefore compromise the efficiency of endoderm differentiation. Similarly, once endoderm is formed, these cells still retain the ability to differentiate into multiple different endoderm lineages, such as hepatic-like cells (Teng Ang et al, 2018). Therefore, the addition of PRC2 inhibitors may negatively affect the formation of pancreatic cell types, if added before pancreatic progenitors are formed. In agreement with this, a study by Xu et al 2014, that added EZH2 inhibitors to *in vitro* differentiating pancreatic cells, demonstrated that earlier inhibition negatively affected the percentage of insulin positive cells. Although the differentiation in Xu et al, 2014 does not associate directly with the protocol used in this study, in terms of the formation of different cell populations, treatment with DNZep before the expression of *NGN3,* negatively affected the number of C-peptide positive cells at the final timepoint of differentiation. When cells were treated at a later timepoint, once *NGN3* was expressed, the percentage of C-peptide positive cells was increased at the final day of differentiation and also showed increased *NGN3* expression immediately after inhibition. This study also created mouse *in vivo* knockouts of *Ezh2* through a *Pdx1*-Cre gene. Loss of *Ezh2* in correlation with increased *Pdx1* expression showed an increase in the pancreatic progenitor population, as marked by *Ngn3* expression. This also resulted in an increase in β cell mass, as measured by insulin immunostaining, in the pancreas of mice (Xu *et al.*, 2014). This suggests that addition of PRC2 inhibitors would be most effective after the formation of pancreatic progenitors, and after the expression of NGN3 and in association with PDX1 expression. The *in vitro* protocol used in our study, observed *PDX1* by day 9, and *NGN3* expression at day 17 of differentiation (Figure 3.2). As the PRC2 inhibitors are effective over 72 hrs of cell growth, it was decided that inhibitors will be added from day 14. In addition, the differentiation protocol changes at this timepoint from daily media changes to media changes every 72 hrs. To coordinate with the 72-hour media changes, four timepoints of inhibitor addition were tested in this study, from day 14 to day 26, with daily addition of PRC2 inhibitor over these 72 hrs periods (Figure 4.4).

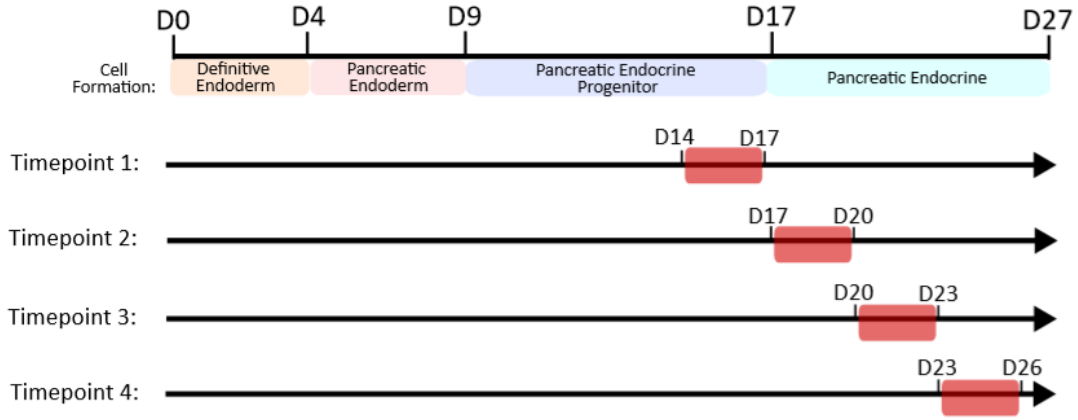


Figure 4.4 Schematic demonstrating the experimental strategy for inhibition of EZH proteins during *in vitro* human pancreatic differentiation.

After selecting the experimental strategy, the ability of GSK343 and UNC1999 to induce the loss of H3K27me3 and associated gene derepression in differentiating pancreatic cells needed to be confirmed (Figure 4.5). It was quickly determined that the concentrations optimised in hPSCs did not have the same level of efficacy in the differentiating pancreatic cells, with no cell death occurring, as previously seen following inhibitor addition (data not shown). This difference was hypothesised to be a result of the varying density of cells between the cultures. The hPSCs typically grow at fairly low confluence, however early in the differentiation protocol the cell population reaches a high level of confluency. With more cells present in a large continuous layer it was predicted the treatment would not have the same effectiveness. The levels of PRC2 inhibitors were therefore increased to 20 µM GSK343 and 10 µM UNC1999. A small amount of cell death was observed in the treated cell populations, similar to the levels observed in hPSCs with lower inhibitor treatment.

Cells undergoing pancreatic differentiation were treated with either GSK343 or UNC1999, as summarised in Figure 4.4, and samples were collected immediately after each window of inhibition to confirm the efficiency of the EZH inhibition. After isolation of histone proteins from the treated cell populations, the levels of H3K27me3 were analysed to determine if PRC2 activity was reduced (Figure 4.5a). Note that insufficient material was obtained from timepoint 2 for histone blot analysis. As expected, only a very low level of H3K27me3 was

detected in all timepoints analysed and for both inhibitors, demonstrating these concentrations of inhibitors used are sufficient to reduce PRC2 activity in the differentiating cells.

To determine if the reduction in H3K27me3 is sufficient to remove gene repression, specific genes were analysed for their expression levels between DMSO control treatment and GSK343 and UNC1999 treatment in timepoint 1 and timepoint 4 (Figure 4.5b). In hPSCs, a number of genes that are maintained in a repressed state by H3K27me3 are known, however genes that are repressed by H3K27me3 throughout multiple timepoints in differentiation are less defined. Using H3K27me3 ChIP and RNA-sequencing obtained from *in vitro* differentiating cells (Chapter 3), a number of genes were selected that had both H3K27me3 at their promoters and did not have active transcription. Although the transcriptional upregulation was variable between timepoints, all timepoints showed some degree of gene de-repression (Figure 4.5b). The relatively low level of gene upregulation may be due to a decreased effect of inhibitors on differentiating cells, which may be caused by a slower dividing population, therefore the loss of H3K27me3 occurs slower, but it could also be caused by the presence of a non-synchronised population, and therefore the upregulation of the same genes may be less coordinated. Regardless of the levels of upregulation, the addition of 20 μM GSK343 and 10 μM UNC1999 caused a reduction in H3K27me3 and an upregulation in genes normally marked by H3K27me3. Demonstrating the ability of the small molecule inhibitors to have an effect on cells throughout the pancreatic differentiation.

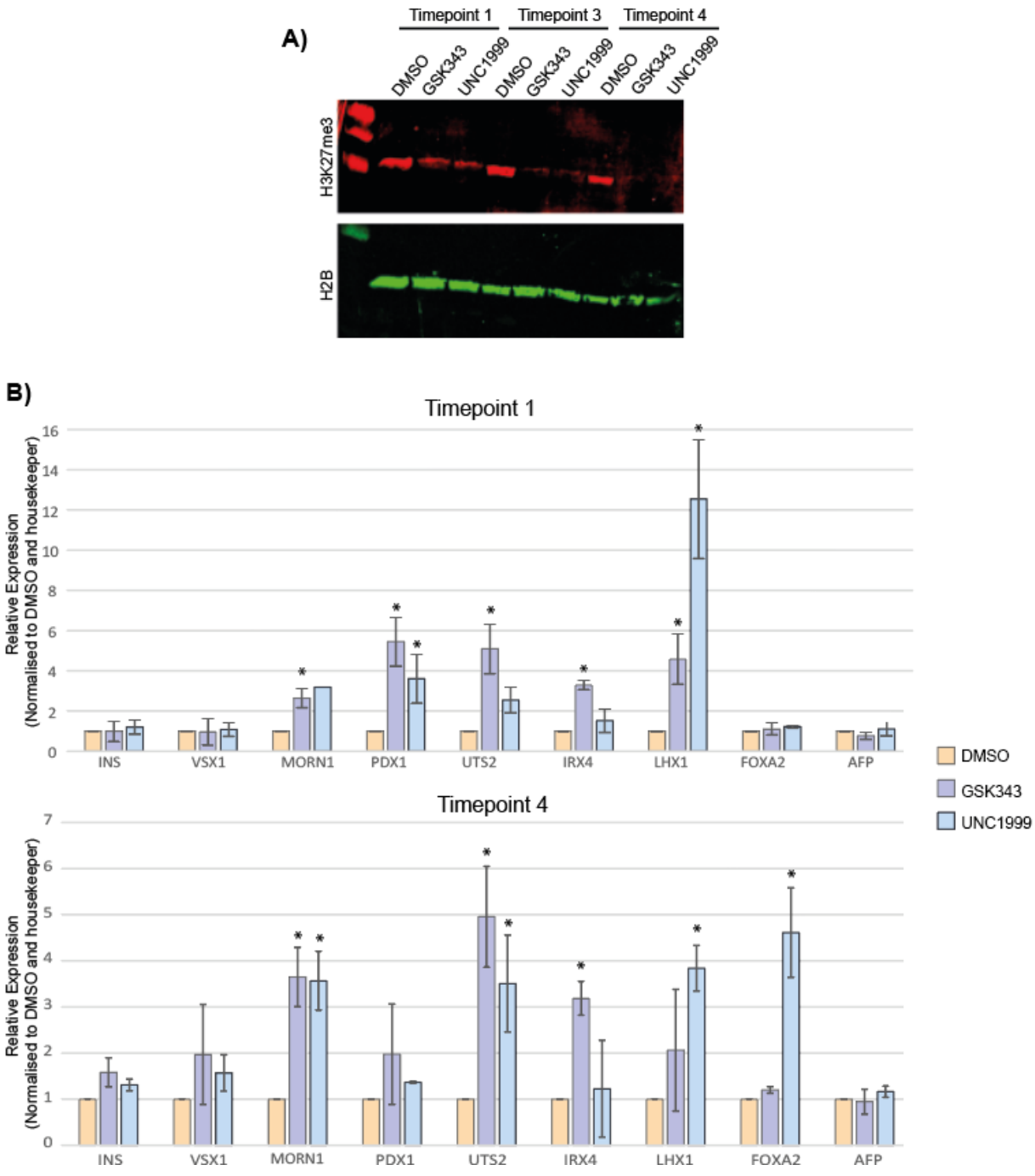


Figure 4.5 Loss of H3K27me3 modification and gene repression in differentiating pancreatic cells using small molecule inhibitors. A) Histone blot analysis of H3K27me3 and H2B levels after 72 hrs of small molecule addition, at three timepoints during differentiation. B) RT-qPCR expression of H3K27me3 marked genes after 72 hrs treatment with EZH inhibitors UNC1999 and GSK343 or a DMSO control. Expression was calculated as relative expression to DMSO control. Data shows the mean of 2 biological replicates ± SD, * indicates a p-value ≤0.05 when compared with DMSO by one-way ANOVA with post-hoc Tukey HSD.

4.3.3 Functional differences in day 27 pancreatic cells population in response to PRC2 inhibition.

Once the optimal use of small molecule inhibitors had been determined, the effect that PRC2 perturbation has on the day 27 *in vitro* pancreatic population was analysed, after perturbation at various point during differentiation. The first aspect analysed was the proportion of hormones expressing cells between the treatment types, however difficulties in preparing cells for flow cytometry analysis, and the inability to efficiently quantify cell population by immunocytochemistry, meant that a suitable population distribution analysis could not be undertaken. Therefore, the cells were analysed to determine the gene expression of these pancreatic hormones in the final cell population. The overall expression in the cell populations for the three pancreatic hormones was analysed in both DMSO control and the stage matched EZH inhibited treated conditions (Figure 4.6). This demonstrated that PRC2 inhibition did not lead to increased expression of pancreatic hormones at any of the different stages. Very little difference in gene expression levels of either insulin, glucagon, or somatostatin was observed between all timepoints and treatment types. The one exception is UNC1999 applied at timepoint 1, in which a small decrease is observed for all three hormones, with a clear reduction on both somatostatin and insulin and a smaller reduction in glucagon expression. A similar reduction in glucagon can also be seen at timepoint 2 of UNC1999 addition but this treatment does not have decreased levels of somatostatin or insulin (Figure 4.6). As pancreatic hormone expression may be representative of the number of cells expressing these hormones, it could be hypothesised that there is little to no change in the number of hormone-expressing cells between the treated cultures. It also demonstrated that the addition of EZH inhibition at later timepoints in differentiation had little to no effect on the gene expression in the pancreatic endocrine cells, and with a potentially negative effect when added earlier.

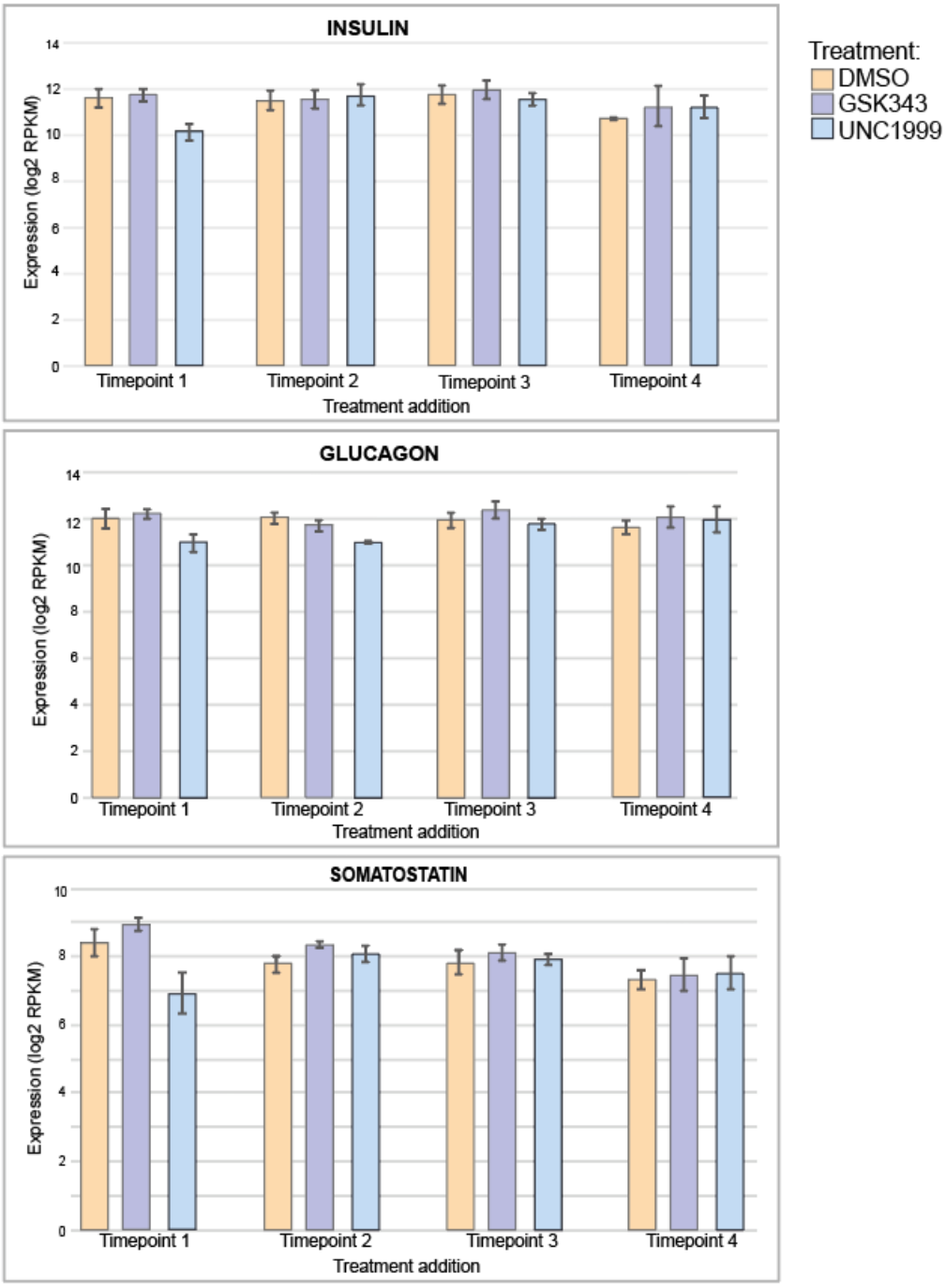


Figure 4.6 EZH inhibition does not affect the gene expression of pancreatic hormones in any treatment conditions. mRNA expression levels of pancreatic hormones from day 27 of *in vitro* derived pancreatic cells after treatment using PRC2 small molecules as indicated. Data show the mean of 3 biological replicates ±SD.

The *in vitro* derived pancreatic cells are known to differ from adult cells in terms of global transcriptome, the levels of expression of pancreatic hormones and the responsiveness of the cell population to repeated glucose stimulation. Although there was little effect on the gene expression of pancreatic hormones, the levels of C-peptide secretion in response to glucose stimulation may still differ between the treated cell populations. It has previously been demonstrated that glucose stimulation of *in vitro* derived pancreatic cells will cause an increase in C-peptide secretion from the cells (Xu *et al.*, 2014), simulating the response in adult pancreatic cells which occurs both in the pancreas *in vivo* and in isolated islets. Unfortunately, in the assay carried out in this study to analyse glucose release, the glucose stimulation failed, a common problem with monolayer *in vitro* derived pancreatic culture, which resulted in neither the DMSO nor inhibitor treated cells increasing the levels of C-peptide secretion in response to high glucose addition. However, measuring the C-peptide release under low glucose addition could indicate the normal C-peptide secretion from the resting final cell populations, as the pancreatic β cell will continuously secretes insulin with a basal secretion under non-stimulatory conditions (low glucose) and increased secretion when stimulated (high glucose) (Leibiger, Leibiger and Berggren, 2008). C-peptide protein levels were therefore measured by ELISA in low glucose conditions and their contribution to the total protein in the supernatant of each population was calculated (Figure 4.7). This demonstrates that both PRC2 inhibition, and addition of DMSO produce cell population with variability in terms of the secretion of C-peptide from the *in vitro* derived cells. In the first replicate tested, the addition of DMSO at progressive timepoints, appeared to increase the levels of C-peptide released. Although the C-peptide release from DMSO control increases with later timepoints, the UNC1999 inhibited populations are seen to have higher C-peptide secretion at timepoints 1, 3 and 4 than the matched DMSO control, in replicate 1. However, in differentiation replicate 2, no increase is seen between treatments and timepoints with the similar levels of C-peptide seen between DMSO and UNC1999 at all addition 4 timepoints (Figure 4.9). Due to the variability between the replicates, it is difficult to determine if the PRC2 inhibition has any effect on the ability of the cell populations to secrete C-peptide. Therefore, the hormone gene expression indicated no change in hormone expressing cell after inhibition. Similarly, C-peptide analysis of the cells did not give a clear indication of a positive effect on the hormone expressing cells in the final population after inhibitor treatment.

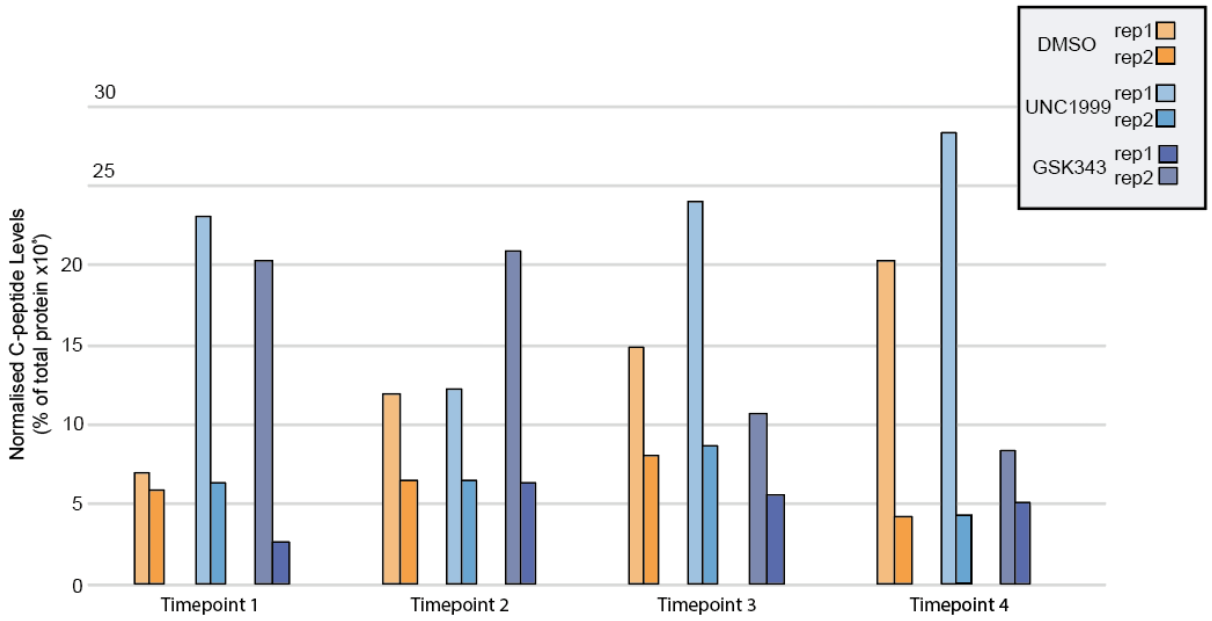


Figure 4.7 Analysis of basal C-peptide secretion from EZH inhibited cell populations. C-peptide levels secreted into the cell media, under low glucose conditions, were measured by ELISA and normalised to the total protein present in each sample.

4.3.4 Global transcriptional effects caused by PRC2 inhibition in *in vitro* differentiating cells.

To understand how the addition of EZH inhibitors will affect *in vitro* differentiating cells as they develop, the global transcriptional changes between inhibited cells and control cell populations were compared through RNA-sequencing. One of the known limitations with long-term *in vitro* differentiation protocols is the variability between replicates. Therefore, the first stage in transcriptional analysis is to determine if the variability between samples is more strongly associated with replicative data or with the timepoints when the inhibitors were added (Figure 4.8) The UNC1999 treated samples cluster by the timepoint at which the PRC2 inhibitors were added. This transcriptional clustering occurs for 11 of the 12 UNC1999 treated samples, demonstrating the time point of inhibitor addition had more transcriptional effect on the final cell population than the differentiation replicate. Interestingly, the GSK343 treated cells do not cluster so clearly by either timepoint of inhibitor addition or replicates. Other than GSK343 timepoint 1, no similarities are seen with the placement of the GSK343 samples on the data similarity tree from global transcription. This suggests that only the addition of UNC1999, will

have a specific effect on the transcription of cells when added during *in vitro* pancreatic differentiation, including timepoints in which cells have time to recover after addition of inhibitors. In contrast clustering of the DMSO samples demonstrated that the replicate samples are more closely correlated to each other, as compared to the timepoint when the inhibitor is added. All four timepoints of DMSO treated differentiation replicate 3 are seen to cluster together, similarly three of the four timepoints in DMSO treated differentiation replicate 1 and replicate 4 will cluster together, the only samples which cluster away from the other replicates are DMSO_timepoint1_replicate2 and DMSO_timepoint2_replicate2. This demonstrates in DMSO treated samples, the differentiation replicate will have more effect on the global transcription than the timepoint at which DMSO is added during *in vitro* pancreatic differentiation, as would be predicted from replication variability, suggesting little consequence of DMSO addition to the differentiating cells.

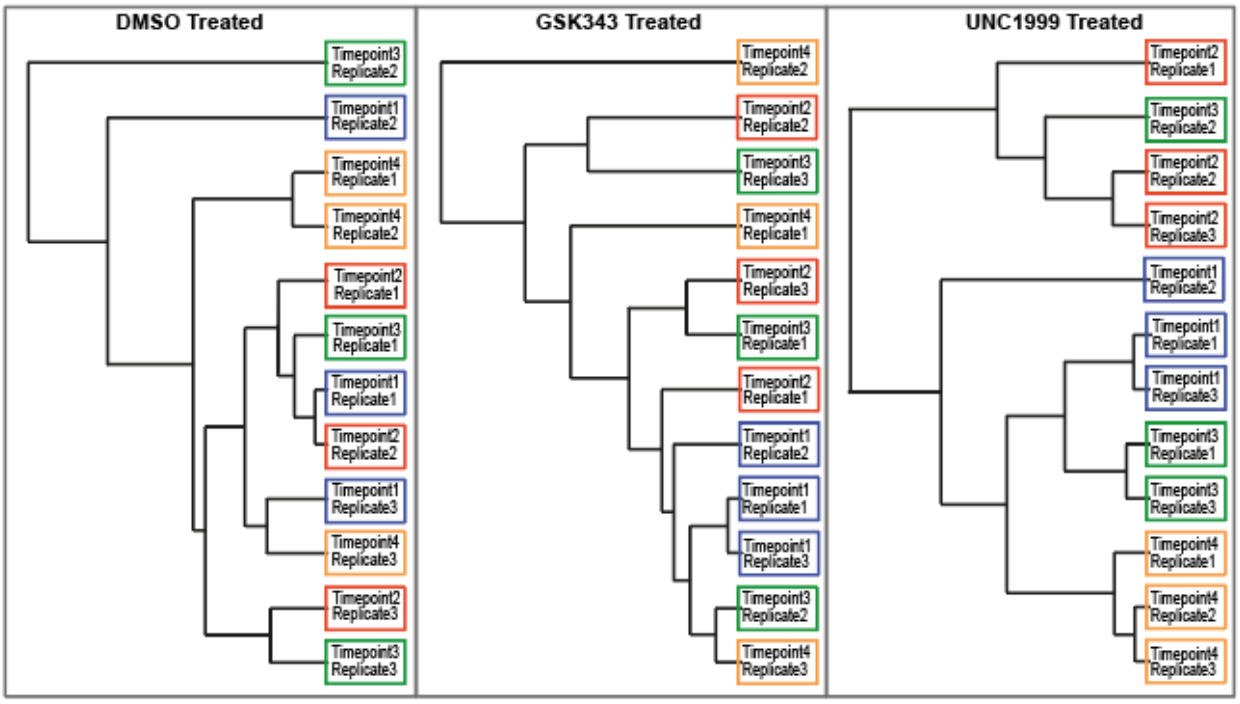


Figure 4.8 Similarity clustering of samples separated by small molecule addition. Samples were clustered for similarity on global gene expression as determined by RNA-sequencing. Separation of samples that had addition of either DMSO, GSK343 and UNC1999 allows a comparison between the effect of differentiation replicate and timepoint addition between samples. Each timepoint of inhibitor addition is indicate by different coloured boxes.

The addition of the UNC1999 inhibitor appeared to have an effect on transcription, but how this change the gene expression of a cell population in comparison with GSK343 or DMSO addition is unclear. To determine the transcriptional changes following small molecule treatment, a principal component analysis (PCA) of the top 2000 most variable genes between the differently treated samples was undertaken. The PCA of each timepoint was examined separately (Figure 4.9). From the PCA, the addition of UNC1999 causes the resulting day 27 population to cluster away from the DMSO treated populations, in three of the four additional timepoints. When analysing timepoint 1 samples (day 14 to 17), the DMSO and GSK343 treated samples are transcriptionally similar, however replicate 1 and replicate 2 of UNC1999 treatment causes a clear separation on principal component 1 (PC1). Replicate 3 of UNC1999 clusters with the DMSO and GSK343 treated populations, which may indicate that this particular replicate sample was not sufficiently inhibited with UNC1999 or the recovery over 10 days of continued differentiation after inhibition was more efficient in this population. For timepoint 2 samples, the treatment of cells with UNC1999 also caused the day 27 population to cluster away the DMSO treated cells on PC1, which contributed to 79% of the variability between the cell populations. Addition of GSK343 at this timepoint also caused a shift in the end-cell population with two of the three replicates spread on PC1 between the DMSO and UNC1999 treated cells. Surprisingly, addition of PRC2 inhibitor at timepoint 3, corresponding to day 20 to 23 of differentiation, resulted in a cell population that was not transcriptional distinct from DMSO treated cells. Treatment with UNC1999 at timepoint 4 caused the end-cell to cluster away from DMSO on PC1, which in this analysis contributed 67% of transcriptional variability between these samples. At this timepoint, one of the GSK343 replicates also clustered away from the DMSO treated cells, closer to the UNC1999 treated cells (Figure 4.9). These analyses therefore indicate that UNC1999 inhibitor treatment at timepoints 1, 2 and 4 causes the formation of a cell populations that are transcriptionally distinct from DMSO treated cells, and the addition of GSK343 causes an intermediary and variable effect on the treated populations.

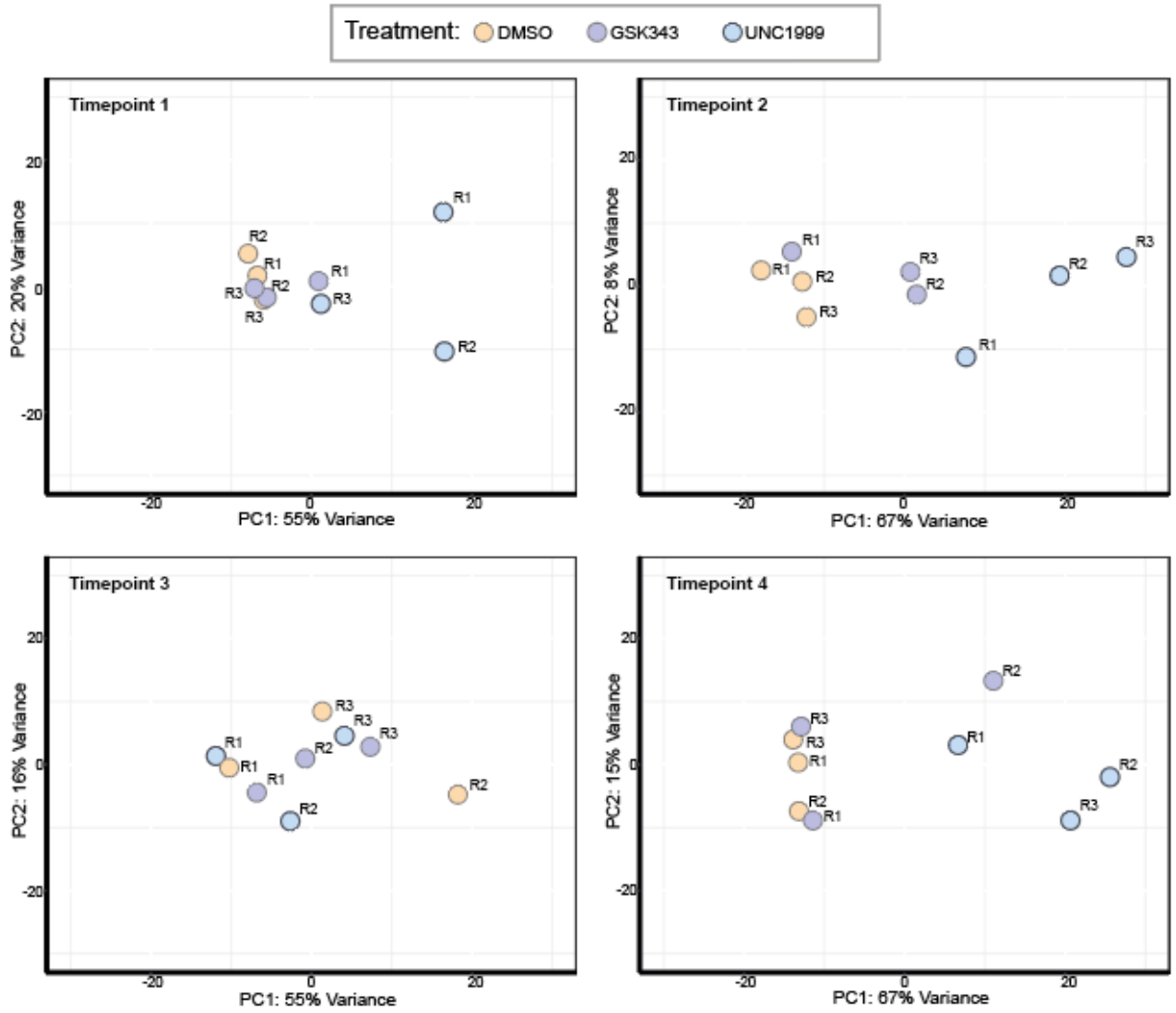


Figure 4.9 Comparative transcriptional analysis of end-cell populations after EZH inhibition during *in vitro* pancreatic differentiation. Principal component analysis of RNA-sequencing gene expression data from day 27 *in vitro* derived pancreatic cells. Samples were treated for 72 hrs with EZH inhibitors, or DMSO control, at 4 different timepoints throughout differentiation. Individual replicates are indicated for each differentiation.

These results establish that the addition of EZH small molecule inhibitors during differentiation produce cells that are transcriptionally distinct from control treated cells. The next focus was to identify the genes that contribute to these differences. The gene expression from cell population formed after treatment with both inhibitors are compared at each timepoint to the DMSO control, to determine differentially expressed genes (Figure 4.10). The UNC1999

treatment led to a much larger and variable number of DEGs. Treatment with UNC1999 caused the upregulation at timepoint 1 of 440 genes, at timepoint 2 of 3409 genes and at timepoint 4 of 1356 genes (Figure 4.10a). Surprisingly, treatment with the PRC2 inhibitors also caused a similar number of genes to be both upregulated and downregulated at timepoint 1 and 2, with slightly more genes downregulated. In timepoint 1, 204 more genes are downregulated over the number of genes upregulated (644 genes downregulated compared to 440 upregulated), similarly timepoint 2 had 596 more genes downregulated over upregulated (3409 genes downregulated compared to 4005 upregulated). Only in timepoint 4, when cells are analysed 24 hrs after inhibition removal are a much larger number of genes seen to be upregulated compared to downregulated (340 genes downregulated compared to 1356 upregulated).

There were no DEGs between GSK343 treated samples and DMSO for timepoints 1, 3 and 4, or for UNC1999 samples versus DMSO at timepoint 3. This lack of differential gene expression supports the data from the previous PCA plots, where very little variability is seen between DMSO control and these treatment timepoints. In the case of GSK343 timepoint2, a small number of genes are differentially expressed, with only 170 genes upregulated and 67 gene downregulated compared to DMSO control (Figure 4.10a). This small number of genes identified as differentially expressed, is likely to be caused by the variability between the replicates, with replicate 2 and replicate 3 differing transcriptionally from DMSO but replicate 1 of GSK343 addition clustering with DMSO.

To determine any genes which are consistently up and downregulated between the timepoints the genes from each of the treatments were analysed for commonality. As only a single treatment for GSK343 had differentially gene expression, this was not included in this analysis (Figure 4.10b). The majority of genes which are higher expressed in UNC1999 inhibited cells are uniquely upregulated to each timepoint, with 81.7%, 93.4% and 87.8% of the genes which are upregulated, specific in timepoint1, timepoint2 and timepoint4, respectively. For downregulated genes, 87%, 96.6% and 75% of these are uniquely downregulated, in timepoints1, timepoint2 and timepoint4, respectively (Figure 4.10b). This demonstrates that very little genes are commonly differentially expressed, in terms of both down and upregulation, between the timepoints. Analysis of upregulated genes, demonstrates that only a single gene (*EFR3B* gene), is upregulated between timepoint 1, 2 and 4, and all timepoints only had 2 genes commonly downregulated (*MME* and *CEACAM6* genes). This demonstrates that the effect of PRC2 inhibition has on gene transcription is specific to the timepoint when inhibition is applied and is different between the timepoints.

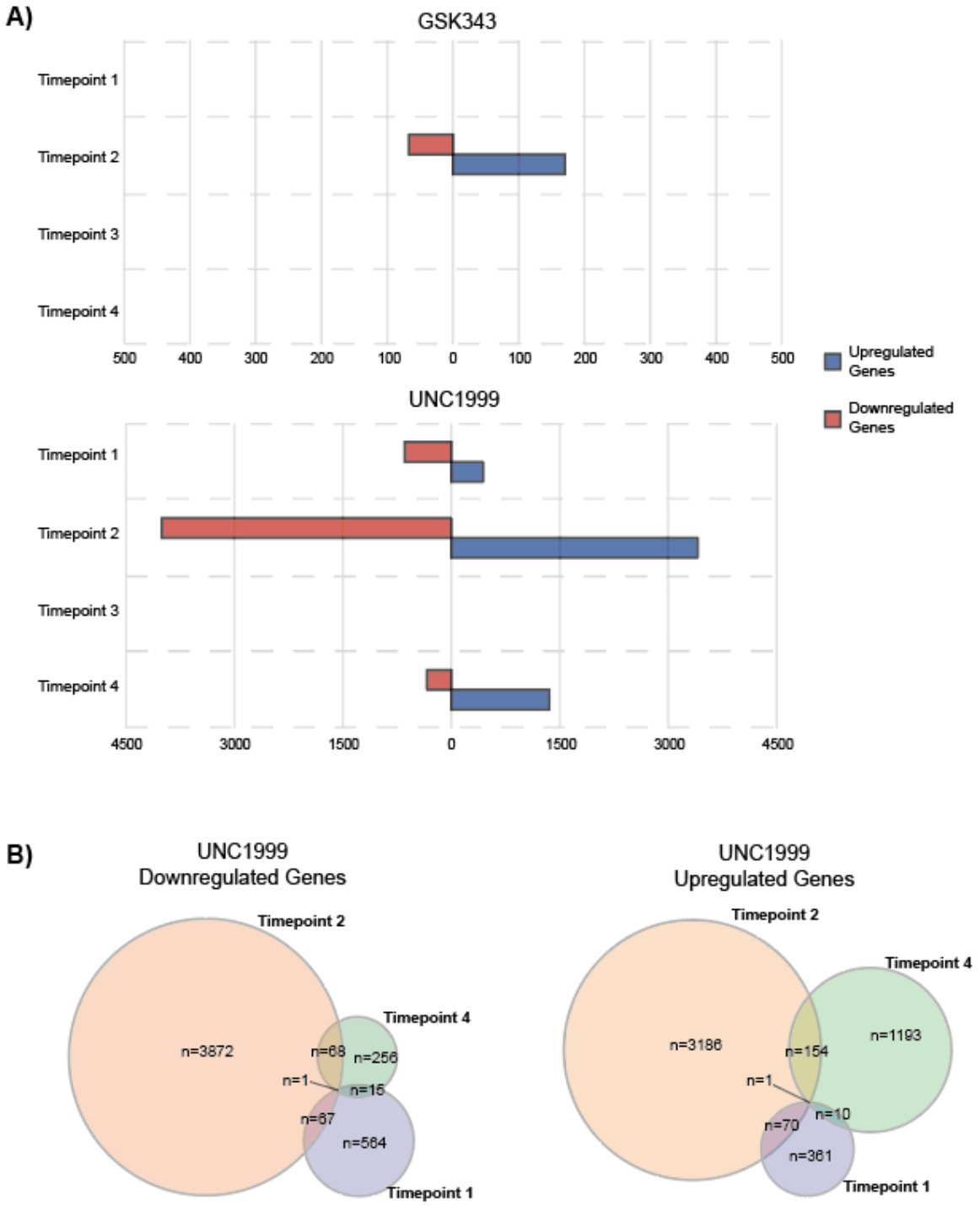


Figure 4.10 The number of differentially expressed genes following EZH2 inhibition is highly variable, depending on the timepoint of treatment. (A) Numbers of differentially expressed genes after EZH inhibition, separated into down and upregulated, as compared with time matched DMSO cells. (B) The overlap of genes which are upregulated and downregulated between the 3 different UNC1999 treated timepoints, compared to DMSO control samples.

4.3.5 Transcriptional signature changes caused by EZH inhibition in pancreatic *in vitro* differentiating cells.

To identify groups of genes enriched in the DEGs of each timepoint and gain further insight into the effect of inhibition of the cells formed, gene ontology using enrichR was undertaken for both upregulated and downregulated genes in UNC1999 inhibited populations (Figure 4.11a). Upregulated genes in both timepoint1 and timepoint2 were enriched for genes that are associated with neuronal cell types. This transcriptional enrichment is associated with distinctive cell types, which may explain the lack of overlap in Figure 4.10, but both are selecting neuronal terms. Interestingly, mouse pancreatic cells and purified β-cells have previously been observed to have high transcriptional similarity to neuronal cell types, as compared with cells of the same endoderm origin such as liver and lung cells (Van Arensbergen *et al.*, 2010). In contrast to the two earlier timepoints, the gene ontologies in genes upregulated at timepoint4 with UNC1999 inhibition, had an enrichment of protein transport, which may be associated with the secretory endocrine cells. The downregulated genes did not produce ontologies which may indicate the loss of any features associated with specific cell types (Figure 4.11a).

The GO analysis indicated there was an increase in the expression of genes that are functionally associated with alternative cell types. Therefore, it was next analysed if this is associated with a loss or gain of pancreatic transcription expression. To determine this, 87 pancreatic enriched genes, as identified with at least four-fold higher expression specific to pancreatic tissue (Figure 3.4), were analysed at day 27 for timepoint 1, 2 and 4 (Figure 4.11b). For both UNC1999 timepoint 1 and timepoint 2, there was a slight decrease in the average expression of these key pancreatic genes when compared with DMSO matched control. Although the overall similarity of the expression levels indicate that a pancreatic phenotype is still largely established. The addition of UNC1999 at timepoint 4 resulted in similar levels of expression of the 87 pancreatic enriched genes between DMSO and UNC1999 samples (Figure 4.11b). Therefore, the addition of EZH inhibitors at earlier timepoints, before day 20, causes changes in the transcriptional signature. This effect is not observed when the inhibitors are applied at a later timepoint.

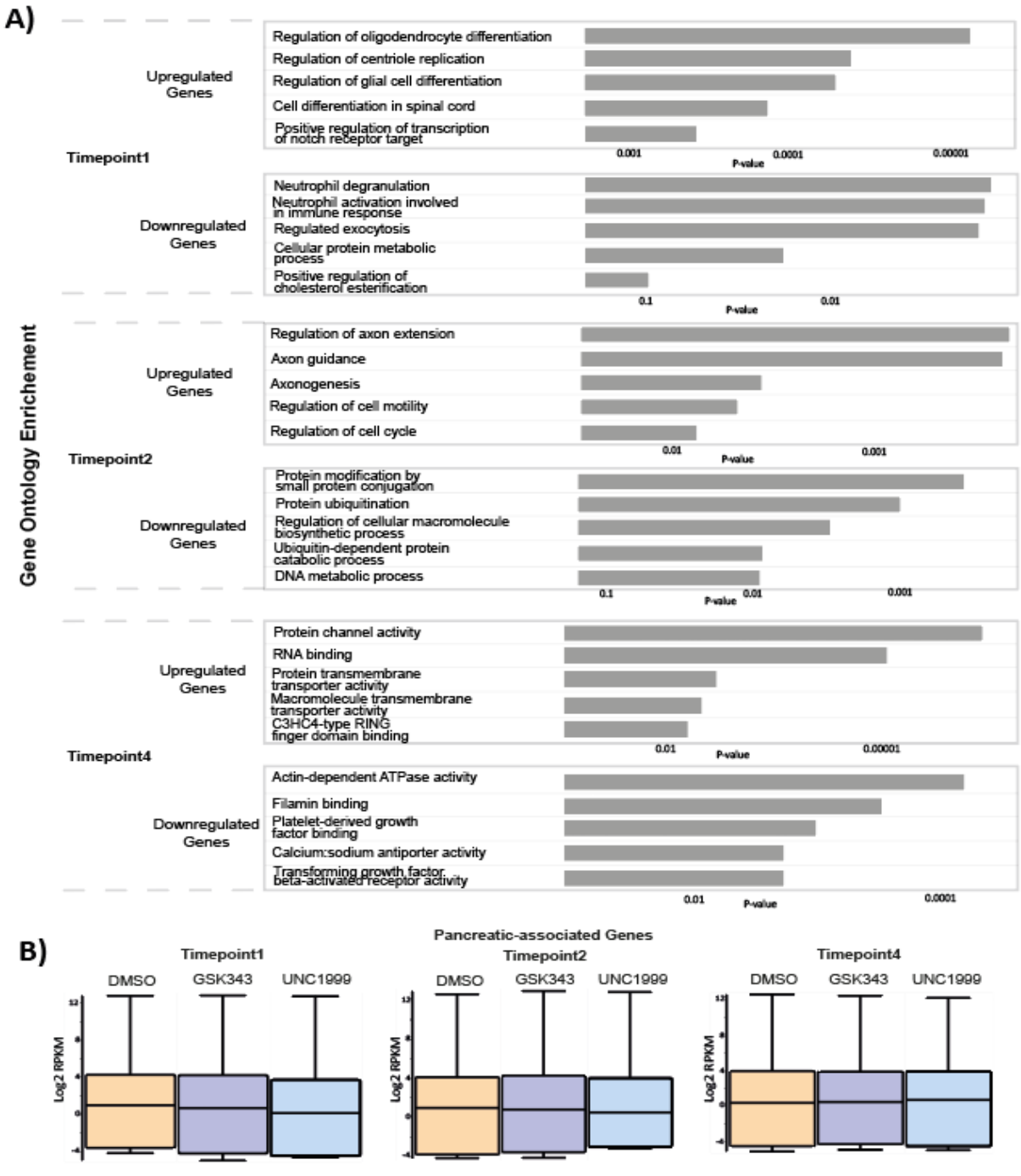


Figure 4.11 Determining changes in the transcriptional signature of end-cell population after PRC2 inhibition by GO term enrichment and by analysis of key pancreatic genes. (A) Top GO enriched terms (as determined by Enrichr) of genes that were differentially expressed between UNC1999 inhibited timepoint1, 2 and 4 and their timepoint matched DMSO control. (B) Boxplots showing the spread of expression values of 87 pancreatic-associated transcripts in the DMSO, GSK343 and UNC1999 treated conditions, indicating the median, the first and third quartile, the minimum and maximum of gene expression.

The transcriptional analysis of DEGs suggested a slight loss of pancreatic-like phenotype in the cells in which PRC2 was inhibited at an early stage of differentiation. To determine if this also indicates a global transcriptional shift towards other cell lineages, an unbiased analysis of the cell's transcriptome was undertaken. To analyse the expression of genes in multiple different lineages, the KeyGene web application was used to probe transcriptional data compared with a human foetal transcriptional atlas (Roost *et al.*, 2015). Using the global transcriptional data from all of the day 27 *in vitro* derived pancreatic samples, and the transcriptional data from *in vivo* derived foetal and adult samples, the lineage association of the cells was tested (Figure 4.12). A fixed training set of foetal samples provided by Roost et al, 2015 was used for this sample comparison, which contains transcriptional signatures of 21 foetal tissues and organs, and the maternal endometrium. As expected, the CS 16 - 18 foetal pancreatic cell samples were most highly correlated with the pancreatic foetal tissue, with very little correlation in any other foetal tissues or organs. Similarly, the adult islet samples were correlated only with pancreatic foetal samples and no other tissue types, with the degree of correlation varying slightly between the donor samples (Figure 4.12). All *in vitro*-derived samples are most highly correlated with the foetal pancreatic signature. In the DMSO control treated samples, little correlation in other tissues is observed for all replicates. For the GSK343 treated samples, most are enriched in the pancreatic lineage, but a small number of samples demonstrate some enrichment for other lineages. The most pronounced of these is timepoint 2, replicates 2 and 3, which demonstrate a small level of association with intestinal, and a low association with the foetal stomach transcriptional signature. Compared to DMSO samples, the UNC1999 samples demonstrate higher enrichment with transcriptional signatures of tissues other than foetal pancreatic. Multiple replicates of the timepoint 1 and timepoint 2 UNC1999 treated samples demonstrate a low, but clear, association with brain and spinal cord transcriptional signature. Furthermore, the UNC1999 timepoint 4 treated samples had a slight association with a number of different lineages, the strongest association after pancreatic tissue is liver, and a low association in some of the replicates is also seen with stomach, spinal cord, heart and kidney (Figure 4.12). Therefore, although the PRC2 inhibited cells differentiate predominantly into pancreatic-like cells, the addition of PRC2 inhibitors between day 14 to 20 causes an upregulation of neuronal/brain related transcripts, as also detected in the GO analysis, In contrast, PRC2 inhibition between days 23 to 26 causes the upregulation of genes associated with numerous other lineages, likely because the inhibitors was removed just prior to the sample collection and therefore the cells have yet to recover sufficient repression of these genes.

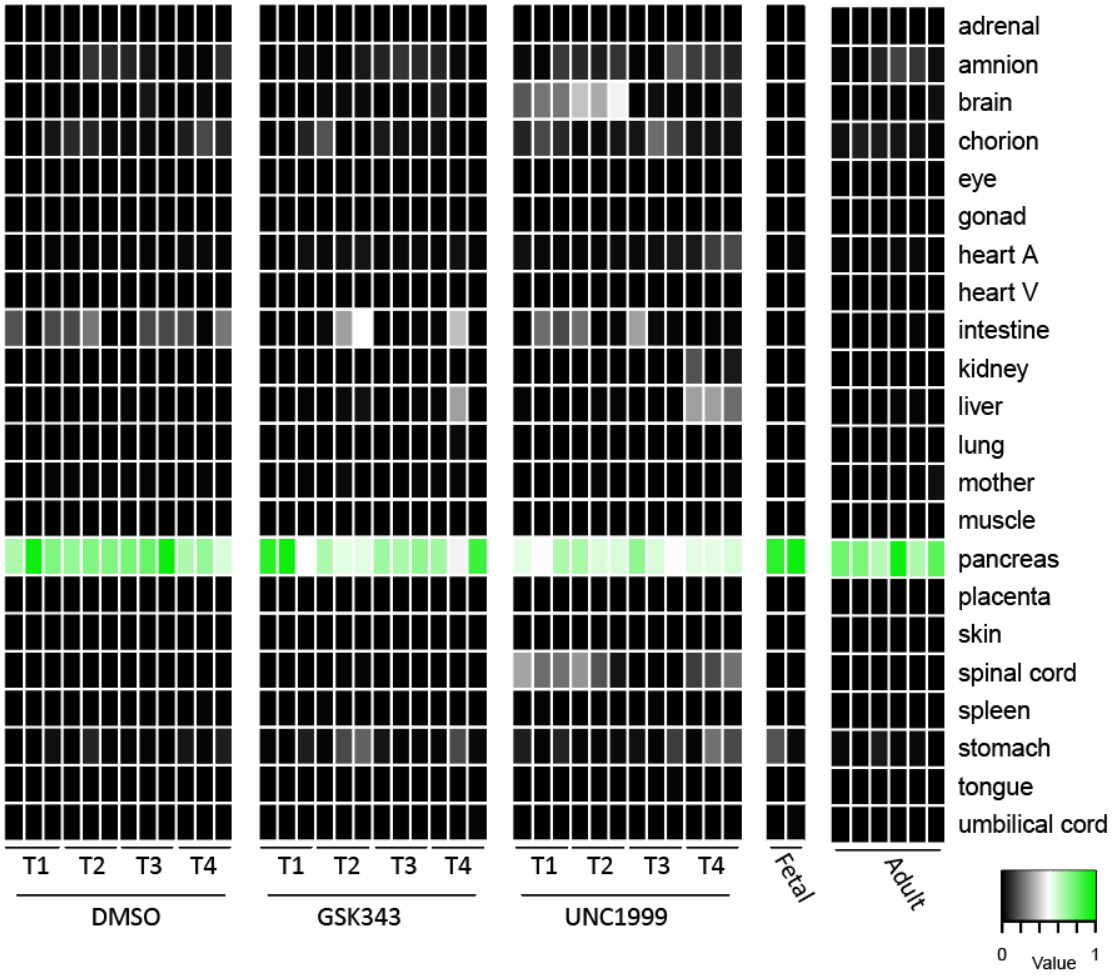
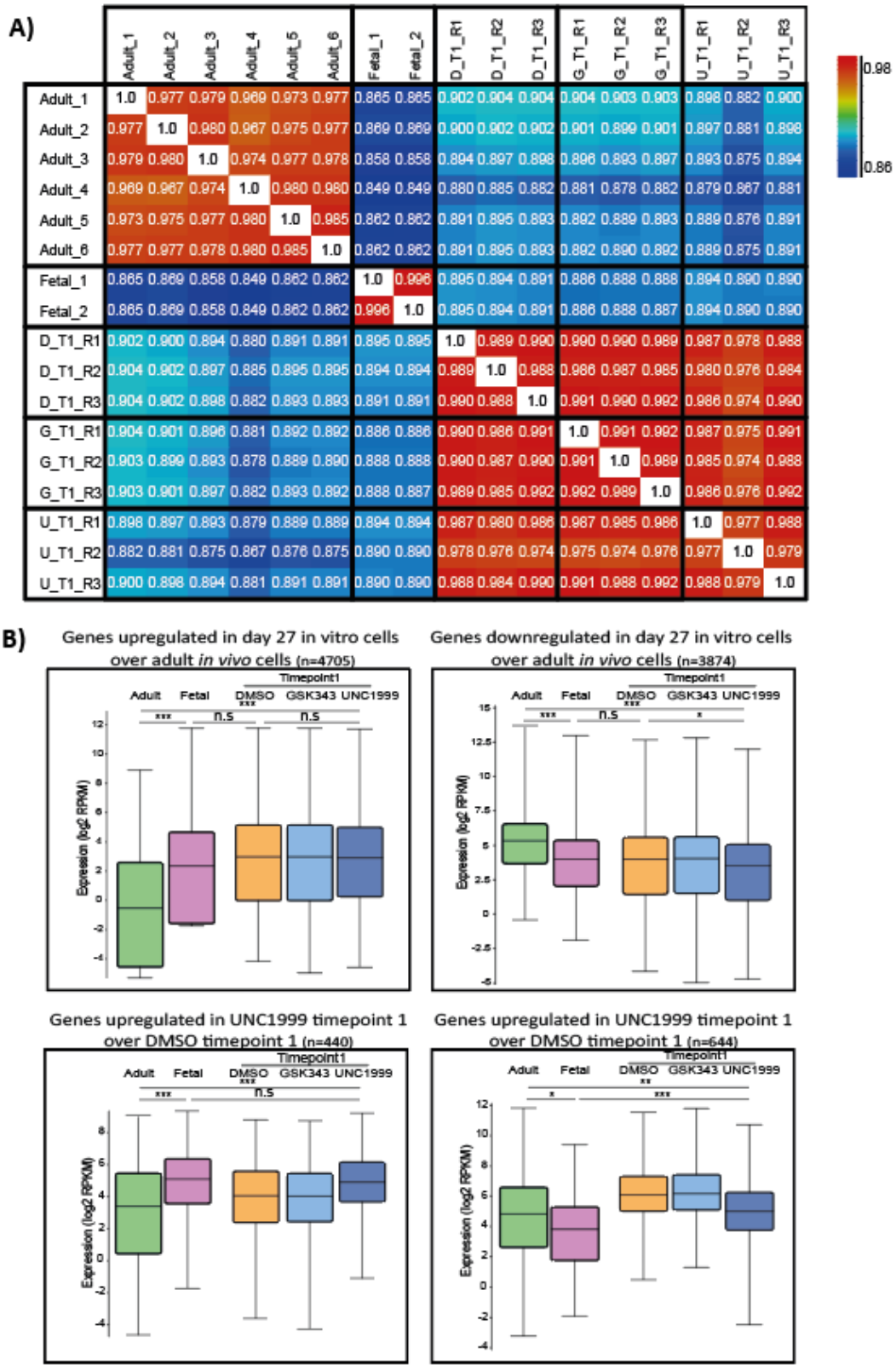


Figure 4.12 Associating foetal transcriptional signatures to end-cell populations after EZH2 inhibition during *in vitro* pancreatic differentiation. Expression from each *in vitro* derived sample was compared with foetal tissues using key genes as indicators of cell type, as described in Roost et al, 2015. The expression of the key genes in each sample is compared to the expression of the genes in each tissue, and a correlation value is calculated. Foetal pancreatic cells and adult islet transcriptional data was included as a positive control.

Enrichment analysis indicates that the EZH inhibited samples have an increasing association with other lineages, however it is unclear if this is because of a loss of pancreatic transcriptional signature or because of upregulation of genes associated with other lineages. Previous analysis of a small number of pancreatic enriched genes (Figure 4.11) indicated changes that early inhibition may negatively affect the pancreatic transcriptional signature, but a more in-depth analysis was now undertaken on each timepoint by direct comparison with *in vivo*

cells. Firstly, for timepoint 1, the correlation was assessed between the transcriptomes of the day 27 *in vitro* differentiated samples compared to either foetal or adult tissues (Figure 4.13). This analysis demonstrated a high correlation between all cell types, whereby the lowest correlation values, which exists between adult and foetal cells, was above 0.84 between these samples. According to this analysis, the transcriptome of all timepoint 1 samples, regardless of addition of inhibitors, correlate to both adult and foetal cells at similar levels.

To determine if the addition of inhibitors at timepoint 1 had a positive effect in changing the gene expression profiles of the *in vitro* derived cells towards an adult *in vivo* pancreatic transcriptional phenotype, the expression of previously identified DEGs was analysed between the samples (Figure 4.13b). Firstly, genes that are differentially expressed between day 27 *in vitro* derived cells (without small molecule addition) and adult islets (Figure 3.16) were used to study the effect of small molecule addition. As previously seen (Figure 3.16), genes that are more highly expressed in *in vitro* cells have higher expression in foetal islet cells compared to adult islet cells, and similarly, genes downregulated in *in vitro* derived cells compared with adult islets also tend to have lower expression in foetal cells. The addition of UNC1999 between days 14 to 17, when compared to DMSO control, showed little transcriptional change in the set of genes that are upregulated in the *in vitro* derived cells. The set of genes upregulated in adult islets had, on average, a small downregulation in cells that were treated with UNC1999, as compared with the DMSO treated sample. This suggests that the addition of UNC1999 is having a slight negative effect on improving the transcriptional signature of the *in vitro* cells towards an adult islet phenotype. Genes upregulated after addition of UNC1999 are typically expressed at higher levels in foetal cells compared with adult cells, and similarly genes downregulated in UNC1999 treated cells typically have lower expression in foetal cells type compared with adult islets (Figure 4.13b). This again indicates that the addition of the UNC1999 inhibitor at day 14 to day 17 does not improve the transcriptional signature of the final cell types to resemble adult islets more closely but may instead direct some gene expression towards the foetal phenotype.



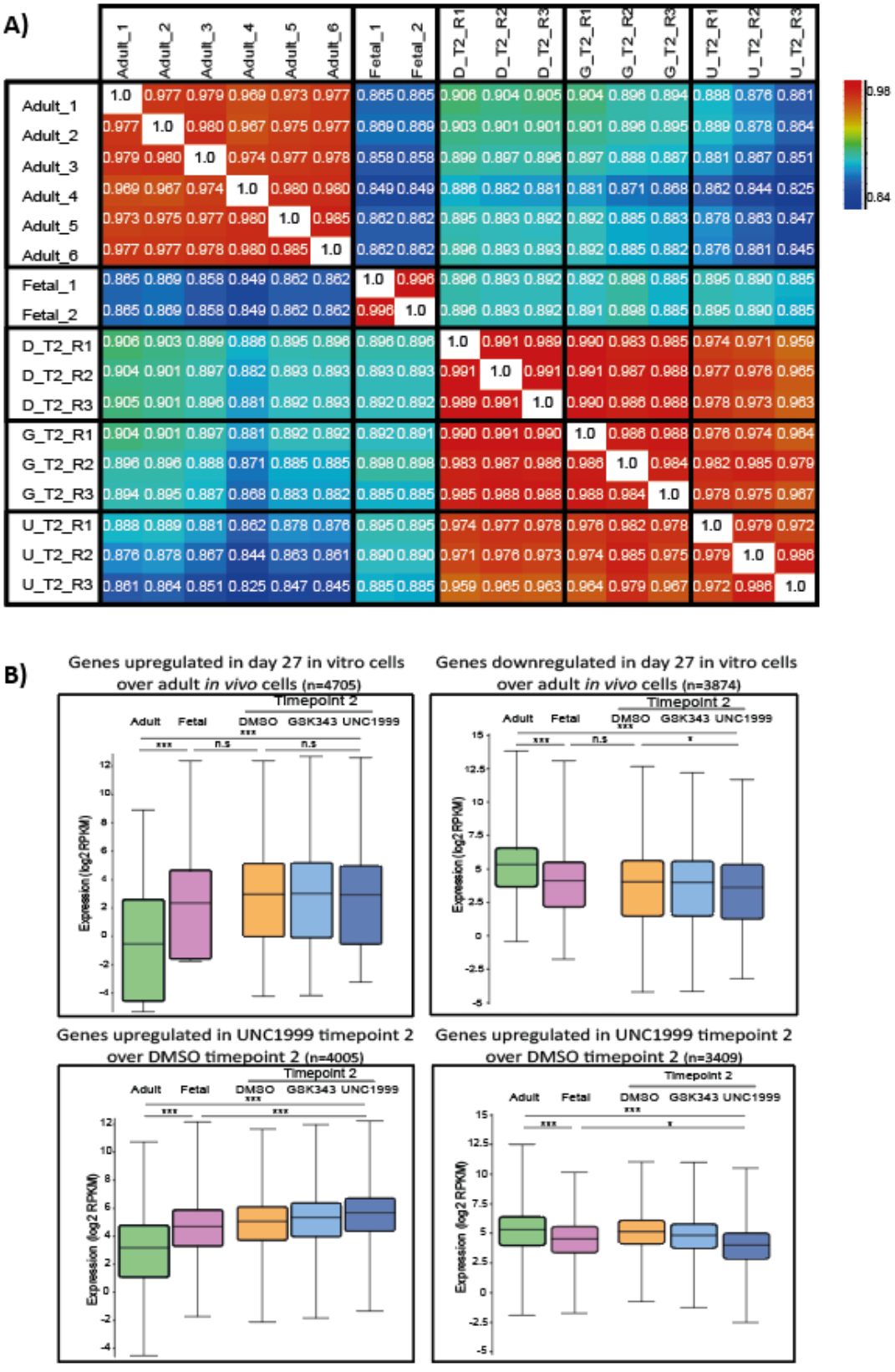
(Legend over page)

Figure 4.13 Comparison of timepoint 1 inhibited *in vitro* derived pancreatic cells compared with *in vivo* derived cells (*overleaf*). **(A)** Pairwise correlation of total transcriptome for all samples including adult islets (Adult_replicate), foetal Carnegie stage 16 - 18 pancreatic buds (foetal_replicate), and *in vitro* derived pancreatic cells treated with DMSO (D_T1_replicate), GSK343 (G_T1_replicate) or UNC1999 (U_T1_replicate) produced Pearson correlation values as indicated for each comparison. **(B)** The expression of genes previously determined to be differentially expressed between day 27 *in vitro* cells and adult *in vivo* cells (figure 3.16), and between DMSO and UNC1999 treated cells at timepoint 1 (figure 4.13), was analysed for each of the samples as indicated in the boxplots showing the median, the first and third quartile, the minimum and maximum of gene expression. Statistics were calculated using an ANOVA with a post hoc Tukey HSD, ns = P >0.05, * = P ≤0.05, ** = P ≤0.01, *** = P ≤0.001.

To determine if the addition of EZH inhibitors at timepoint 2 has a similar or opposing effect compared with treatment at timepoint 1, the same comparative analysis with *in vivo* cells was undertaken. Firstly, a correlation matrix was produced comparing each individual sample of the *in vivo* derived adult and foetal cells and the small molecule treated *in vitro* derived cells (Figure 4.17a). Unlike the timepoint 1 treated *in vitro* derived cells, where little difference was observed in the correlation, there is a clear difference in the correlation depending on the small molecule added at timepoint 2. Although high correlation values are observed with all samples, the UNC1999 treated pancreatic cells have a lower correlation than the DMSO samples, when compared with adult islets. Correlation values are reduced for all three replicates of UNC1999 treatment, but there is variability in the samples with the highest level of correlation maintained in replicate 1, with slightly less correlation in replicate 2, and further reduction of correlation in replicate 3. (Figure 4.9). This correlation decrease is specific to the adult pancreatic cells, as no difference in correlation is observed between the UNC1999 and the DMSO samples when compared to foetal cells.

Interestingly, the set of genes differentially expressed between day 27 derived pancreatic cells and adult islets demonstrated little expression change between DMSO and UNC1999 treated cells, for both the up and down regulated gene sets (Figure 4.17b). This suggests that the lower correlation value is due to changes in genes that were not originally identified as being differentially expressed in the day 27 *in vitro* cells, and that the addition of EZH inhibitors between day 17 to day 20 does not improve the transcriptional signature towards adult islet cells. Studying the genes that are differentially expressed between day 27 DMSO and UNC1999 treated cells, the genes which are upregulated after inhibition are typically higher

expressed in foetal cells compared to the expression in adult pancreatic cells, similarly the genes downregulated are typically lower expressed in foetal over adult cell types (Figure 4.17b). This comparative analysis suggests that the addition of PRC2 inhibitors between day 17 to 20 during *in vitro* differentiation causes a reduction in the expression of genes associated with adult islets.

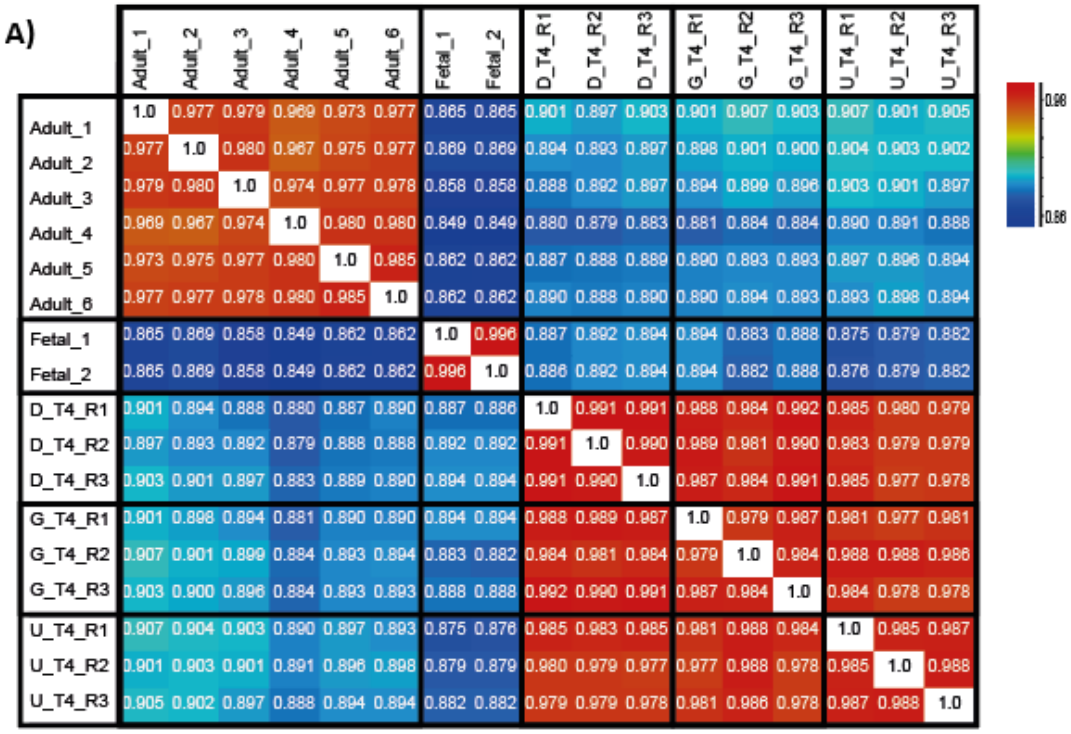


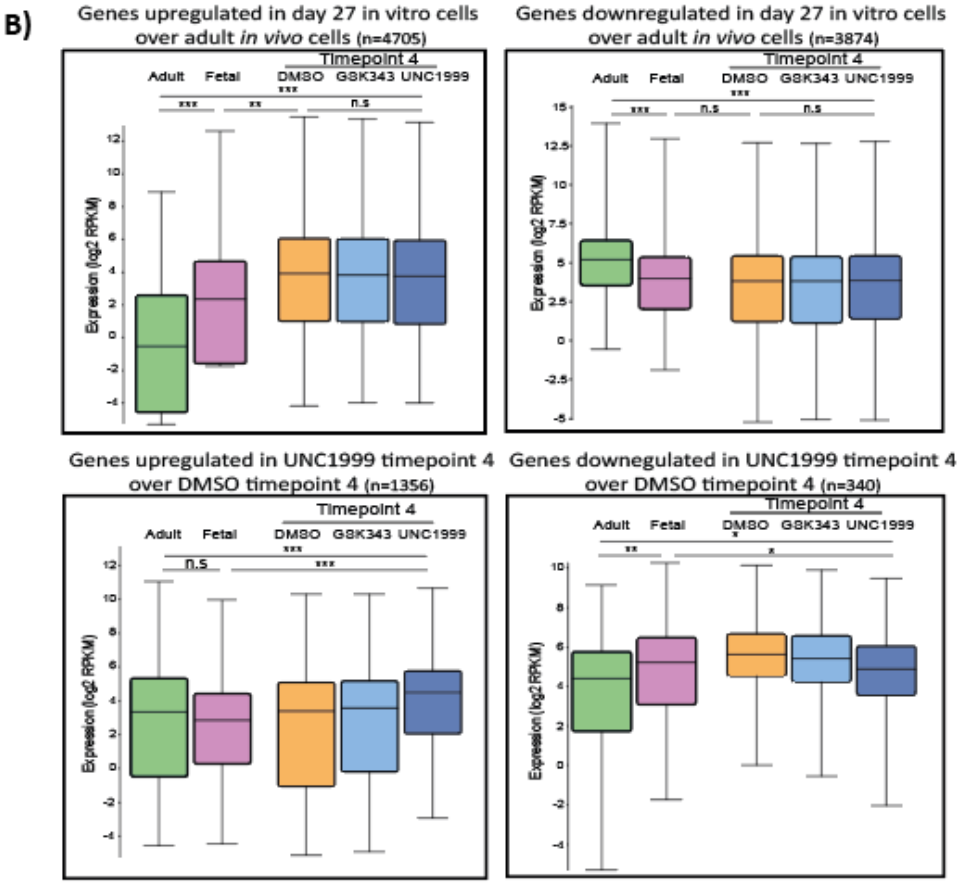
(Legend over page)

Figure 4.14 Comparison of timepoint 2 inhibited *in vitro* derived pancreatic cells with *in vitro* derived cells. (A) Pairwise correlation of total transcriptome for all samples including adult islets (Adult_replicate), foetal Carnegie stage 16 - 18 pancreatic buds (foetal_replicate), and *in vitro* derived pancreatic cells treated with DMSO (D_T2_replicate), GSK343 (G_T2_replicate) or UNC1999 (U_T2_replicate) produced Pearson correlation values as indicated for each comparison. **(B)** The expression of genes previously determined to be differentially expressed, between day 27 *in vitro* cells and adult *in vivo* cells, and differentially expressed genes between DMSO and UNC1999 treated samples, was analysed for each of the samples as indicated in the boxplots showing the median, the first and third quartile, the minimum and maximum of gene expression. Statistics were calculated using an ANOVA with a post hoc Tukey HSD, ns = P >0.05, * = P ≤0.05, ** = P ≤0.01, *** = P ≤0.001.

The comparative analysis was then applied to the day 27 cells obtained from small molecule addition during timepoint 4 of differentiation (Figure 4.15). As with the previous analysis, there was a high correlation between the day 27 *in vitro* derived cells and the foetal and the adult islets. There is little variation between the DMSO and PRC2 inhibited treated cells when compared with the *in vivo* derived cells, although there is a small increase in the correlation value between adult and UNC1999 treated cells as compared with DMSO treated cells. There is also a small decrease in the correlation of the UNC1999 and foetal cells, as compared to the correlation between the DMSO treated cells and foetal tissues (Figure 4.15a).

Genes differentially expressed between day 27 *in vitro* derived pancreatic cells and adult islets show a small difference between DMSO and UNC1999 treated cells (Figure 4.15b). A slight decrease in gene expression in the UNC1999 sample is observed in the day 27 upregulated gene set. Similarly, a small increase is observed in the day 27 downregulated gene set after UNC1999 treatment, compared to DMSO control. When analysing genes that are differentially expressed between day 27 DMSO and UNC1999 treated cells, genes that are upregulated after PRC2 inhibition have, on average, slightly higher expression in adult islets compared to the expression in foetal pancreatic cells. Analysis of genes that are differentially downregulated in the UNC1999 day 27 population sample compared to DMSO have lower expression, on average, in adult islets over foetal cell types (Figure 4.18b). Therefore, overall, the expression pattern of the DEGs in timepoint 4 is the opposite to that observed for timepoint 1 and 2. The upregulated genes in UNC1999 treated are higher in adult islets compared to foetal islets, and genes downregulated in UNC1999 treated are higher in foetal cells over adult cells. The slight increase in correlation towards the adult transcriptome, and the slight decrease in correlation with the foetal samples, suggests a potential positive effect following EZH inhibition at timepoint 4 in promoting the expression of genes associated with adult islets.





(Legend overleaf)

Figure 4.15 Comparison of timepoint 4 inhibited *in vitro* derived pancreatic cells compared with *in vitro* derived cells (*previous page*). (A) Pairwise correlation of total transcriptome for all samples including adult islets (Adult_replicate), foetal Carnegie stage 16 - 18 pancreatic buds (foetal_replicate), and *in vitro* derived pancreatic cells treated with DMSO (D_T4_replicate), GSK343 (G_T4_replicate) or UNC1999 (U_T4_replicate) produced Pearson correlation values as indicated for each comparison. (B) The expression of gene previously determined to be differentially expressed, between day 27 *in vitro* cells and adult *in vivo* cells (figure 3.16), and between DMSO and UNC1999 treated cells at timepoint 4 (figure 4.13), was analysed for each of the samples as indicated in the boxplots showing the median, the first and third quartile, the minimum and maximum of gene expression. Statistics were calculated using an ANOVA with a post hoc Tukey HSD, ns = P >0.05, * = P ≤0.05, ** = P ≤0.01, *** = P ≤0.001.

Comparing the expression of DEGs between *in vitro* derived and *in vivo* cells indicated that late-stage inhibition of PRC2 could have a small positive effect on the differentiating cells in pushing them towards an adult islet transcriptome. To further investigate this, the overlap was examined between the *in vitro* versus *in vivo* DEGs and the DMSO versus UNC1999 DEGs (Figure 4.16). The percentage overlap indicates if the expression changes caused by PRC2 inhibition is favourably affecting genes that are normally differentially expressed between adult islets and day 27 cells, as compared with a randomly selected gene set. This analysis shows that 26% of genes upregulated in the timepoint 1 UNC1999-treated experiments overlap with genes upregulated in the day 27 cells compared to adult pancreas, compared to 10% of genes within the random set. A similar trend was observed for samples in the timepoint 2 experiment. In addition, genes downregulated following UNC1999 treatment within the timepoint 1 and 2 experiments had a greater overlap with the genes that have reduced expression in day 27 compared to adult islet. Excitingly, samples in the timepoint 4 experiment showed the opposite trend. In timepoint 4, the set of genes upregulated following UNC1999 treatment contained a much greater proportion (29.4% vs 11.4% random genes) of genes that are downregulated in day 27 *in vitro* cells compared to *in vivo* cells. There was also a large overlap (35.3%) between the genes downregulated in the timepoint 4 UNC1999 experiment and the genes upregulated in day 27 over adult pancreas. Taken together, these results establish that treatment of late stage differentiating cells with UNC1999 will positively effect genes which are normally DEG between *in vivo* and *in vitro* cells.

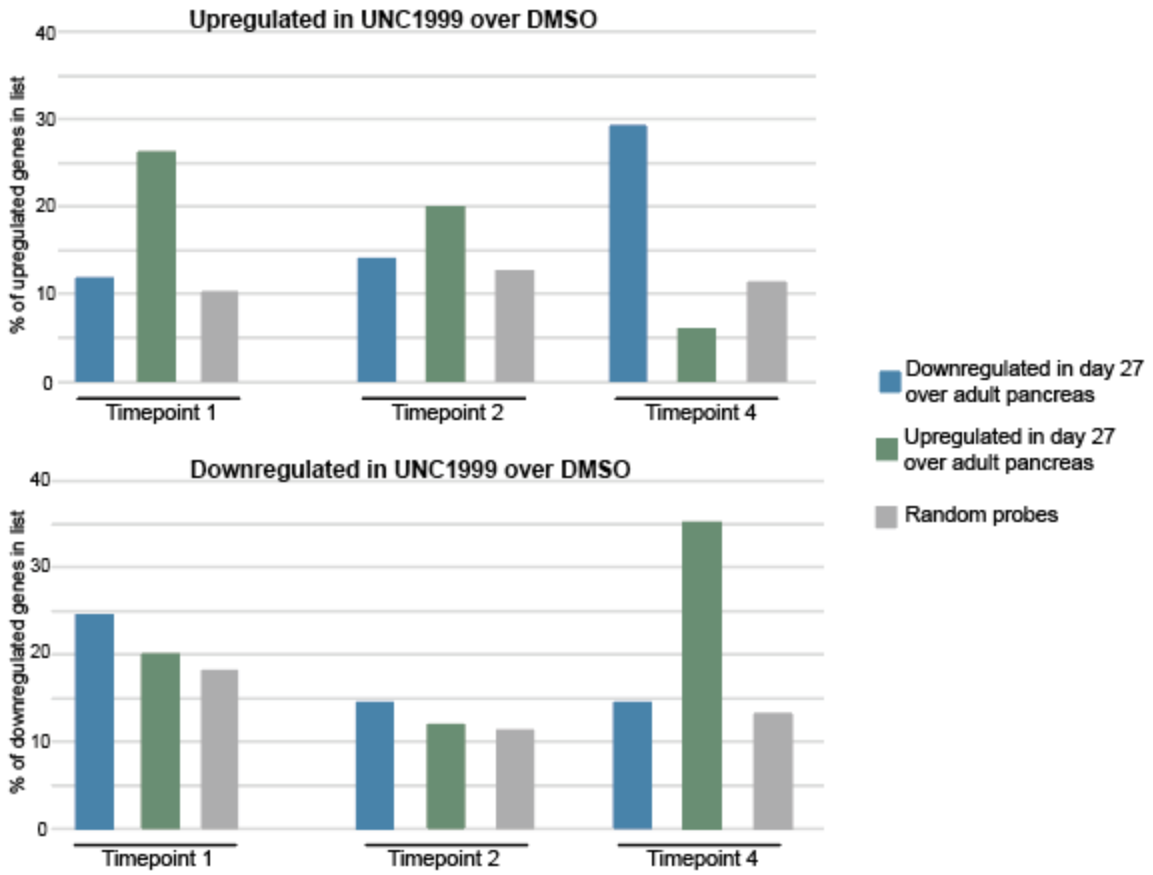


Figure 4.16 Overlap of differentially expressed genes between day 27 in vitro-derived cells versus adult islets, and DMSO versus UNC1999 treated samples. The percentage of overlap in genes that are differentially expressed between DMSO and UNC1999 treated cells in the specified experiment, with genes previously identified as differentially expressed between day 27 *in vitro* cells and adult *in vivo* cells.

4.3.6 The effects on histone methylation after PRC2 inhibition during *in vitro* differentiation

Transcriptional changes are clearly induced in the final-stage cell populations following PRC2 inhibition of *in vitro* differentiating cells, however the effect of this on histone methylation, is currently unknown. The design of the previous inhibition experiments allows the analysis of both a final-stage cell population that is isolated only 24 hrs after the removal of the inhibitor (testing short-term recovery) and also cells that have had multiple days since the withdrawal of the inhibitor (testing long-term recovery).

To analyse the effect of PRC2 inhibition on H3K27me3 levels, calibrated ChIP-sequencing was undertaken for both H3K27me3 and H3K4me3 in the PRC2 inhibited and DMSO treated samples. After normalisation of the samples using spike-in calibration, the genome was divided into 200 bp regions, and the enrichment of histone modifications was quantitated in a 4 kb window centred on all TSS. The levels of H3K27me3 present in UNC1999 and DMSO treated cell populations at each timepoint was determined (Figure 4.17). The GSK343 treated samples had variable transcriptional effects, with samples either more closely representative of DMSO samples or of UNC1999 treated samples, even within the same timepoints. The variability in the GSK343 samples meant it was likely to create unclear results, therefore for the histone modification analysis only DMSO and UNC1999 treated samples were included. In all timepoints, slightly reduced levels of H3K27me3 were seen at TSS in the UNC1999 treated samples compared with DMSO samples. Although lower H3K27me3 levels after UNC1999 treatment is observed in all timepoints, the effect was greatest for timepoints 2 and 4. This suggests that the addition of PRC2 inhibitors affects H3K27me3 levels in both the short-term following inhibitor treatment and after longer recovery times, such as the 7 days between timepoint 2 inhibition and when the samples were collected.

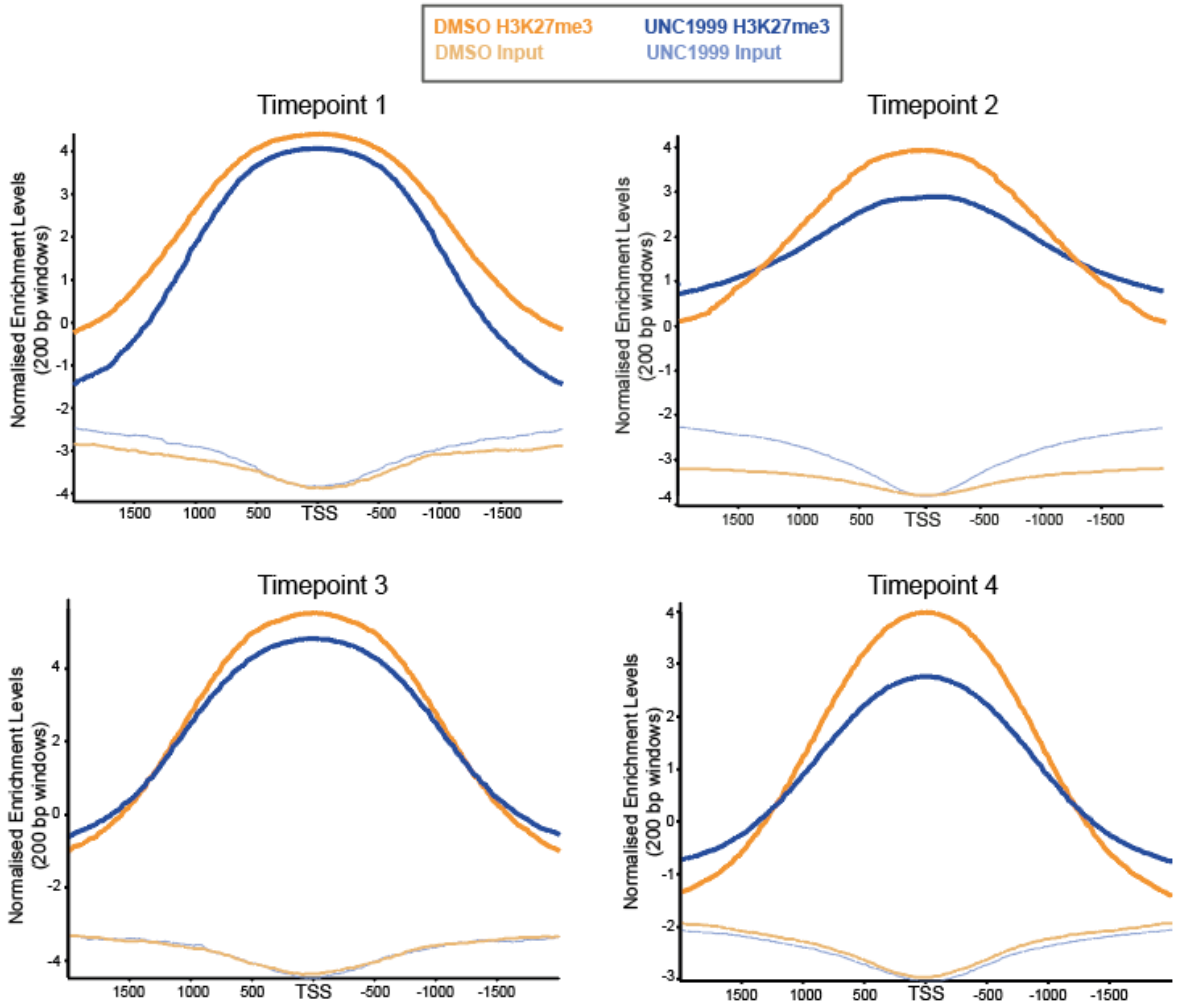


Figure 4.17 Global levels of normalised H3K27me3 over TSS before and after inhibition. The normalised levels of H3K27me3 in 200bp windows were plotted over all TSS +/- 2 kb for DMSO and UNC1999 inhibited. Average of three biological replicates.

The above analysis showed that the levels of H3K27me3 are globally reduced at TSS, however this considered all TSS, regardless of the expected histone modifications, therefore the next step was to determine the histone methylation levels at individual loci. To analyse the effect of PRC2 inhibition, genes were selected that are known to be marked by H3K27me3, either singly, or in combination with H3K4me3 (Figure 4.18). In the DMSO treated samples, both *E2F* and *MAFA* TSS are modified by H3K4me3 and H3K27me3 in the timepoint 2 and timepoint 4 experiments, respectively. Although reduced levels are observed, a small level of H3K27me3 is still present at both genes after inhibition, this is the case both soon after inhibition, as

demonstrated in timepoint 4 and after 7 days of inhibition removal, as seen in timepoint 2. The addition of EZH inhibitors also had an effect on H3K4me3, although this was variable between timepoints, and the gene analysed. In the UNC1999-treated samples, H3K4me3 levels were slightly higher at the E2F TSS in timepoint 2, and slightly lower at the MAFA TSS in timepoint 4 (Figure 4.18). However, other gene categories behaved differently. Genes originally modified by H3K27me3 only in timepoint 2 (XKR4) or in timepoint 4 (GIN1) also lost H3K27me3 following the addition of EZH inhibitors. Interestingly, there was an increase in H3K4me3 levels at these loci after PR2 inhibition, with a small increase at the XKR4 gene in timepoint 2, and a large and clear increase in H3K4me3 at GIN1 in timepoint 4 sample (Figure 4.18).

This individual locus analysis raises the possibility that in response to acute PRC2 inhibition, H3K27me3 levels decrease strongly at both H3K27me3 only and H3K27me3 and H3K4me3 bivalently marked regions. This loss is associated with a corresponding increase in H3K4me3 at regions where the mark was previously not detectable. Regions that were marked bivalently with H3K4me3 show only a small increase in H3K4me3 signal in PRC2 inhibited samples. At timepoint 2, which had multiple days to recover from inhibition, HK27me3 was also decreased in both the H3K27me3 only and H3K27me3 and H3K4me3 modified regions. H3K4me3 levels also increase slightly in both regions examined, however this was not the clear and distinctive increase that was observed at the H3K27me3 only modified locus in timepoint 4.

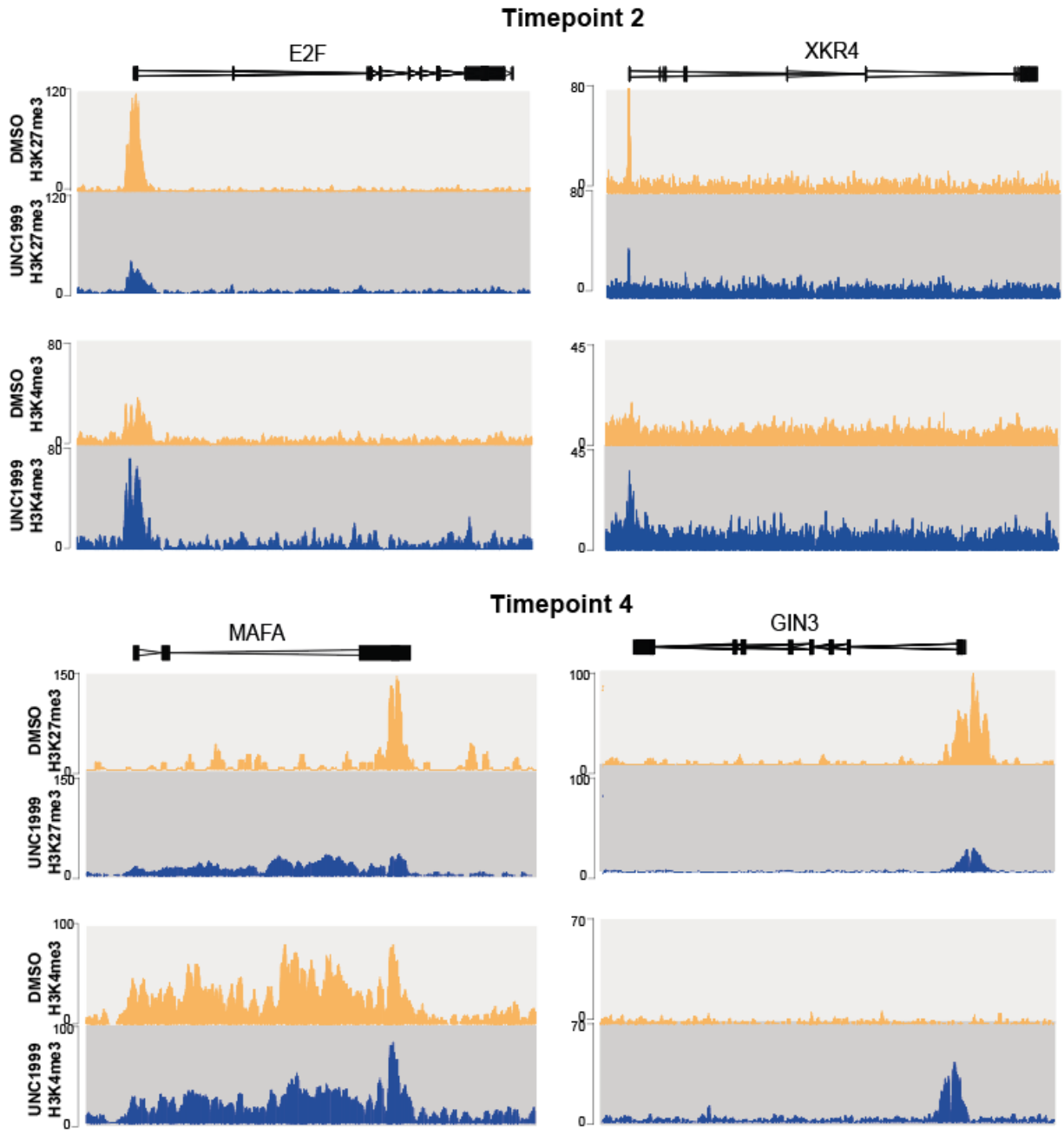


Figure 4.18 Normalised H3K4me3 and H3K27me3 levels at selected individual loci. The normalised levels of both H3K27me3 and H3K4me3 at day 27, over previously identified modified genes for both DMSO and UNC1999 treated samples during timepoint 2 and timepoint 4. Average of three biological replicates.

To determine if these histone modification changes at individual loci had an effect on the transcription, the gene expression of the four genes indicated in figure 4.18 was analysed between DMSO and UNC1999 treated samples within the timepoints (Figure 4.19). The *MAFA* gene is the only gene in which transcription would be considered statistically significant, with a p-value below 0.05 in the four genes studied. The removal of the majority of H3K27me3, and the increased levels of H3K4me3 at the *MAFA* TSS, was sufficient to remove the gene repression on the gene and allowed transcription, of this important pancreatic transcript. However, for each of the other three genes, which initially had expression, the removal of H3K27me3 and gain of H3K4me3 as observed at these TSS, did not cause a significant change in expression. Therefore, the loss of H3K27me3 and gain of H3K4me3 is only sufficient at increasing transcription at specific genes.

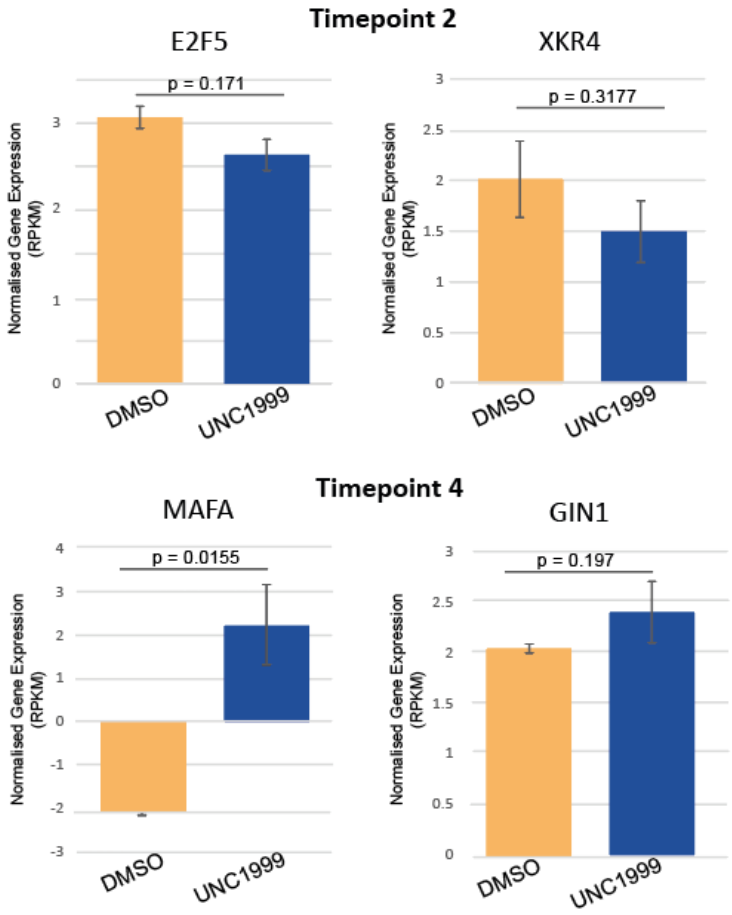


Figure 4.19 Individual gene expression from differentially histone modified loci. The levels of expression from selected genes, which are known to be differentially modified in terms of H3K27me3 and H3K4me3 as seen in figure 4.18. Average of 3 biological replicates, p-value calculated from paired students t-test.

The above observations showed that the levels of H3K27me3 decrease shortly after PRC2 inhibition and also one week after recovery, however the effect of this inhibition more generally on H3K4me3 at these varying timepoints is unclear. To analyse this in detail, genes were categorised as unmarked, H3K27me3-only modified, H3K4me3-only modified, and H3K27me3 and H3K4me3 modified genes, based on MACS peak calling in DMSO-treated samples at timepoint 2 and timepoint 4. The levels of H3K27me3 and H3K4me3 at the genes in each category were quantified to determine how they compare between the PRC2 inhibited and control samples (Figure 4.20). This analysis of timepoint 4 samples demonstrates that EZH inhibition causes a substantial loss of H3K27me3 in both the H3K27me3-only modified genes and also the H3K27me3 and H3K4me3 genes, which is consistent with the single locus level. A reduced level of H3K27me3 was also detected in timepoint 2 for both the H3K27me3-only and the H3K27me3 and H3K4me3 modified genes, however this was not decreased to the very low levels observed in timepoint 4. This suggests the cells might initially undergo a large decrease in H3K27me3 levels following PRC2 inhibition, as the short-term analysis showed, but the H3K27me3 levels subsequently increase over the seven days of recovery that occurred between timepoint 2 inhibition and the isolation of cells for analysis.

A similar pattern is observed with H3K4me3 levels in the H3K27me3-only modified gene set, with a large increase in H3K4me3 measured in timepoint 4, as observed in the single gene analysis. In timepoint 2, with 7 days recovery after PRC2 inhibition, an H3K4me3 increase is also observed in the H3K27me3-only modified samples, however this increase is to a lesser extent than that observed in the timepoint 4 population (Figure 4.20). Little to no differences were observed in H3K27me3 or H3K4me3 in the other gene categories analysed, including in the unmarked and in the H3K4me3-only categories. Together, this analysis shows that the addition of EZH inhibitors to differentiating *in vitro* populations will in the short-term cause a large loss of H3K27me3 at all previously H3K27me3 modified TSS. A resulting increase of H3K4me3 levels is also observed at the H3K27me3-only sites, following the loss of PRC2 activity. After 7 days of recovery from inhibition, the drop in H3K27me3 and the gain in H3K4me3 are detected in similar patterns, but to a lesser extent than the changes observed 24 hrs after inhibitor removal.

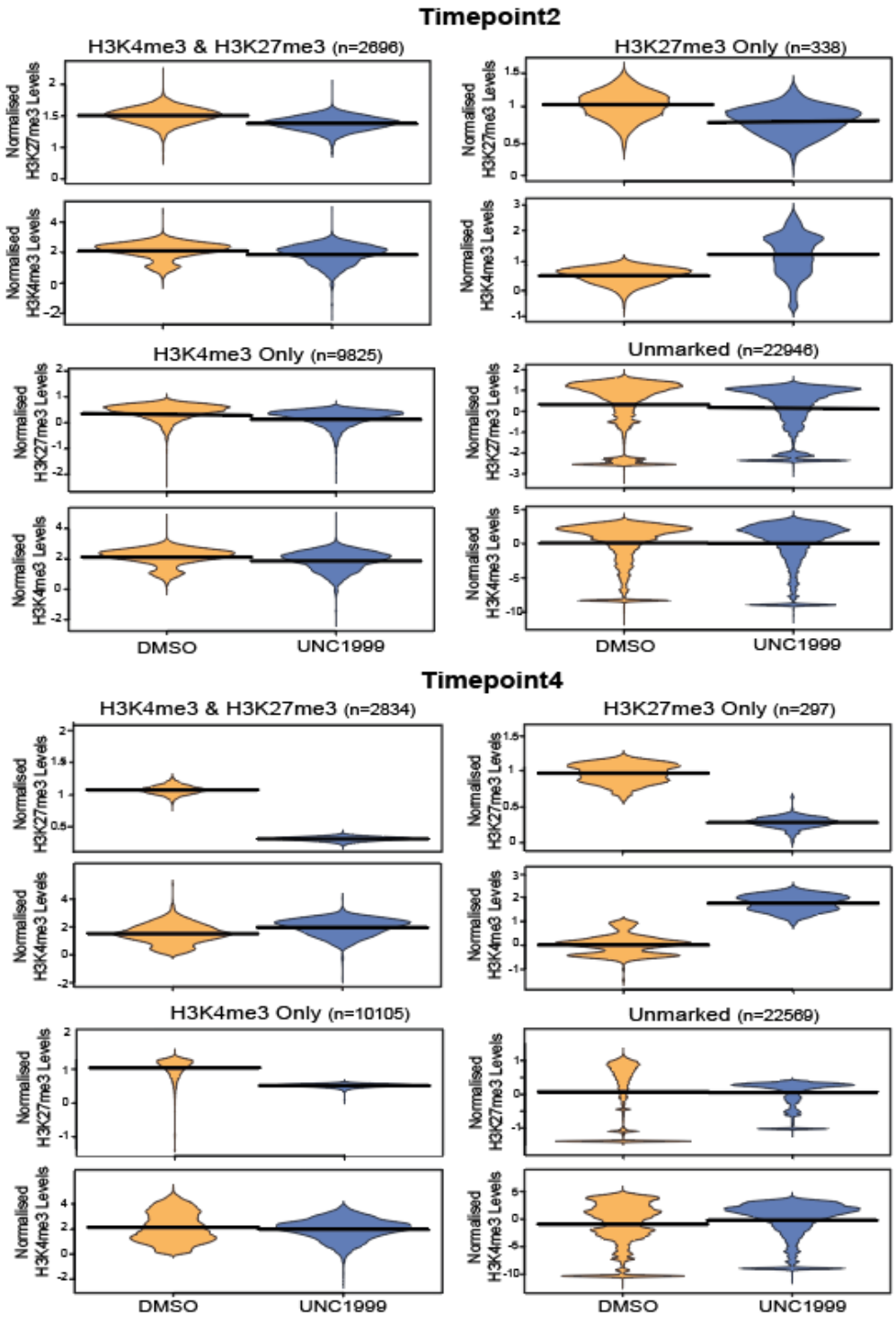


Figure 4.20 Normalised H3K4me3 and H3K27me3 levels at gene sets categorised by their histone modifications. Violin plots shows the normalised levels of H3K27me3 and H3K4me3 at day 27 for DMSO (orange) and UNC1999 (blue) samples, at genes categorised by the histone modification present in the DMSO control, in both timepoint examined. Average levels are indicated by the black line.

From the transcriptional analysis, it is known that a number of genes become differentially expressed after the addition of EZH inhibitors. Therefore, to determine if the changes in histone modifications are associated with the transcriptional changes, both data sets were integrated. Gene expression levels were analysed for genes that were categorised as either H3K27me3-only or H3K4me3 and H3K27me3 in the DMSO samples, as these were the main categories with large and distinctive changes in histone mark levels following PRC2 inhibition (Figure 4.21). This analysis demonstrated that genes with clear histone modification changes do not necessarily have transcriptional changes. Although the earlier transcriptional analysis (Figure 4.10) demonstrated that DEGs are present between the DMSO and UNC1999-treated samples, the absence of overall expression differences in the genes with altered histone modification levels is probably because any gene expression changes are being masked by the larger numbers of genes that are not changing in their expression (Figure 4.20). To confirm this, the histone modification levels of genes that are differentially expressed were compared. This analysis showed there were no histone mark differences between the set of genes that were transcriptionally upregulated or downregulated between the DMSO and UNC1999-treated samples (data not shown), thereby confirming a disconnect between the histone modification changes and the differential gene expression in the cell populations.

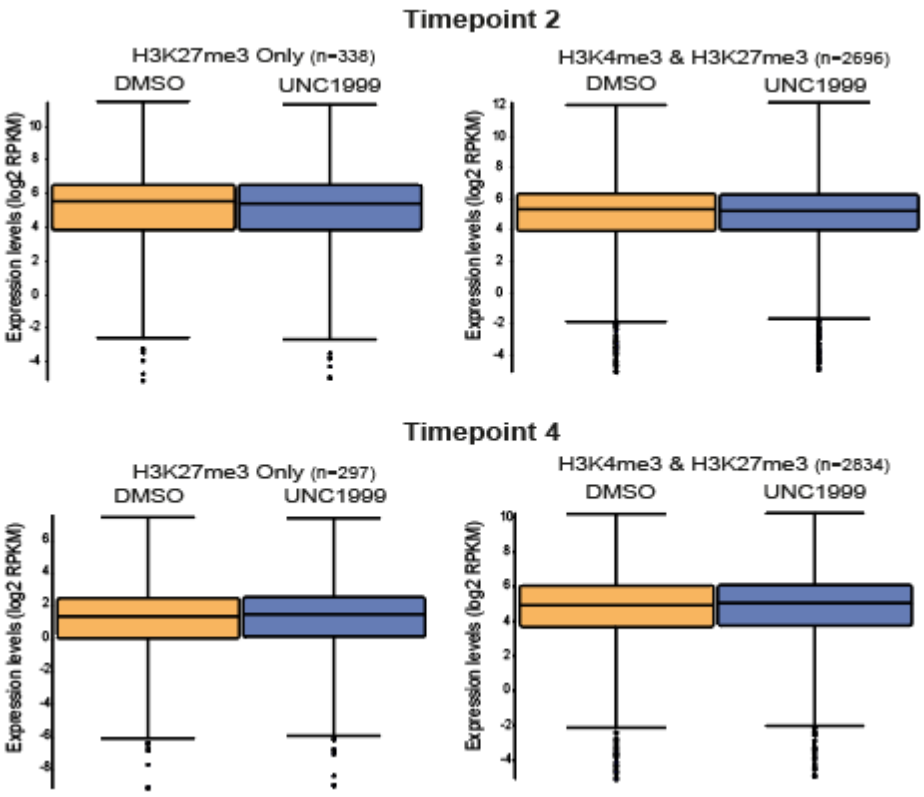


Figure 4.21 Expression levels of gene sets with changing histone modification levels in response to EZH inhibition. The levels of gene expression in DMSO and UNC1999-treated samples are quantified at genes that have detectable differences in histone modifications between DMSO and UNC1999 samples, in timepoint 2 and timepoint 4.

4.4 Discussion

This chapter studied the outcome of inhibiting PRC2 during the *in vitro* differentiation of human pancreatic endocrine cells to determine the effect of inhibition on pancreatic phenotype, the transcriptional signature, and the epigenetic recovery within the cells. To investigate these aspects, cells were inhibited for 72 hrs at different stages of differentiation, ranging from endocrine progenitor cells to mature pancreatic cell types, before removal of inhibition and the continuation of differentiation. Problems with multiple technical problems within assays probing phenotypic changes in the cells made it difficult to draw conclusions. However, the resulting cells, at day 27 of differentiation, were analysed at both a transcriptional and histone modification level by RNA and ChIP-sequencing. Gene expression changes indicated there were varying outcomes in cell phenotype depending on the timepoint when the inhibitor was added. Early inhibition appeared to negatively affect the formation of a pancreatic phenotype, but later addition may have a beneficial effect on the genes normally differentially expressed between *in vitro* and *in vivo* derived cells. The examinations of H3K27me3 and H3K4me3 was used to understand how these histone modifications react to perturbation of PRC2 activity, and their interplay with the changing transcription. The results, which analysed both immediate and delayed recovery to PRC2 inhibition, demonstrated substantial changes in both H3K27me3 and H3K4me3 levels, some of which persist even after seven days post inhibitor treatment. However, the effect that changing histone modification levels had on gene expression changes was variable, with good concordance for some gene loci including key pancreatic regulators, but poor overall concordance when examining large gene sets.

4.4.1 Early-stage PRC2 inhibition alters the transcriptional signature away from a pancreatic cell type.

The effects of PRC2 inhibition were studied at an early stage of cell differentiation at timepoint 1 and timepoint 2, when the cells are transitioning from a pancreatic endocrine progenitor population to the mature pancreatic endocrine cell types at day 27. Following inhibitor treatment, which decreased H3K27me3 levels and led to gene de-repression, cell populations have 10 days and 7 days to recover from the inhibitor in timepoint 1 and timepoint 2, respectively. When analysing these treated cell populations, the GSK343-treated samples were not studied in depth due to the variability seen in these treated cell types. Although some DEGs could be identified when compared with DMSO controls, the variability between replicates as seen in the principal component analysis would likely create confounding results. This variability

is likely contributed to, by the variability of the culture population which occurs in differentiation protocols. Many small factors can affect the time and efficiency of which different cell population will form, and therefore each replicate is likely to have slightly differing population treated with inhibitors which may lead to de-repression of various gene signatures. Due to this variability, which is most prominent in the GSK343 populations, only UNC1999-treated samples were analysed in depth to determine the transcriptional effect of EZH inhibition during pancreatic *in vitro* differentiation.

Importantly, adding UNC1999 at earlier timepoints, resulted in cells that are transcriptionally distinct from DMSO control samples. Analysis of DEGs revealed an upregulation of genes associated with neuronal lineage types in both timepoint 1 and 2 experiments. Further analysis based on a large panel of human foetal tissue transcriptomes confirmed that in comparison to control samples the addition of EZH inhibitors early in pancreatic differentiation resulted in a transcriptional signature that included neuronal cell types and with an associated loss of pancreatic cell phenotype. It is unclear why the inhibition would cause the formation of a neuronal transcriptional phenotype. After removal of the inhibitors the cells continue to be directed through the pancreatic differentiation protocol. However, addition of inhibitors will likely cause broad and unspecific upregulation of genes associated with numerous cell types. Therefore, although the media composition during the early stages of the protocol is designed to direct cells towards pancreatic cell types, the genes upregulated may not be genes which are associated with the pancreatic endoderm/endocrine progenitor cells. The media changes at this time may therefore select for different cell types from this point on, when the initial population is not representative of the cell type that the differentiation conditions were optimised for. The appearance of a neuronal signature could likely be explained by the transcriptional similarity that has been reported between neural and pancreatic cell types. In mice, endocrine cells express many genes in common with neuronal cells and are also electrically excitable (Atouf, Czernichow and Scharfmann, 1997; Van Arensbergen *et al.*, 2010). Interestingly, in mouse islets the localisation of H3K4me3 modifications are also highly comparable between the cell types, with more similarity observed between neuronal and islets, than islets and tissues with a common endoderm origin, such as acinar or liver (Van Arensbergen *et al.*, 2010). Given this similarity, it could be hypothesised that the loss of H3K27me3 following inhibition, together with the presence of H3K4me3 modified loci, could drive the cells partly towards the neuronal cell type, as observed in the differentiating human cells.

When compared to foetal and adult pancreatic data sets, there was no positive benefit observed in terms of directing the transcription of the in vitro derived cells toward an adult phenotype. Instead, the addition of inhibitors may have resulted in a negative effect in improving the endocrine cell types, particularly in the timepoint 2 experiment. Interestingly, the number of DEGs that were upregulated were similar to the number that were downregulated, when timepoint 1 and timepoint 2 samples were compared to DMSO control. With the addition of PRC2 inhibitors, it might be expected that the number of genes upregulated would be higher due to the loss of H3K27me3-associated gene repression, such as that observed in the timepoint 4 experiment. However, the similar number of up and downregulated suggests that the transcriptional effects are no longer a direct result of the de-repression of genes following EZH inhibition, but instead may be due to the downstream effects triggered by the reduction of H3K27me3. Therefore, the addition of EZH inhibitors at timepoints when the cells transition from progenitor to mature pancreatic cells will not positively affect the production of the desired end cell type. This instead causes an increase in genes unrelated to the pancreatic endocrine cells, establishing a new transcriptional signature that does not improve their differentiation outcome.

4.4.2 Later-stage PRC2 inhibition could have a positive effect on pancreatic transcription from differentiating cells.

One interesting result from the transcriptional data is the lack of any effect in the timepoint3 treated cells. When analysing the global transcription through PCA, neither the addition of GSK343 or UNC1999 produced cells that were transcriptionally distinct from the DMSO treated cells. The reason as to why EZH inhibition did not have any effect in these cells after addition at this point is unclear. One possible explanation is that a component of the media composition is in some way restricting or overriding the transcriptional changes that would be caused by the H3K27me3 reduction. However, timepoint 4 samples also had EZH small molecule inhibitors added in the same media composition, and these cells do have transcriptional effects. Instead, these cells may have little discernible DEGs, as they are going through a transition in terms of cell state, or in attempts to correcting gene derepression. The earlier timepoints 1 and 2 have similar numbers of up and down regulated genes, whereas timepoint 4 has a larger number of upregulated compared to downregulated genes. Timepoint 3, which has recovery time less then timepoints 1 and 2, but more than timepoint 4, might be undergoing a transition from an initial large induction of gene upregulation towards a stabilised state where the up and down regulation of genesis more balanced, thereby creating few DEGs between the cells. One other possibility could be that the addition of DMSO at timepoint 3 may

be driving large transcriptional changes. It has been shown that the addition of DMSO to cells can induce epigenetic and transcriptional changes (Nikolaou *et al.*, 2016; Verheijen *et al.*, 2019). However, if the addition of DMSO creates a large transcriptional and/or epigenetic change, this could mask any effect caused the loss of H3K27me3, and it will appear that the addition of the EZH has no effect compared to DMSO control. Regardless of the cause of the minor effect seen in the EZH inhibited cells in the timepoint 3 experiment, with few transcriptional changes observed after small molecule inhibition, these samples were not used for the analysis of the effect of EZH in differentiating pancreatic cells.

When analysing the transcriptional changes in the timepoint 4 inhibited cells, a larger number of genes increased their expression compared to the genes that decreased in expression. This pattern is expected following the removal of a repressive epigenetic modification. The genes identified as differentially expressed are not indicative of a particular cellular or phenotypic change occurring after EZH inhibition. Analysis of GO terms indicates that genes upregulated are enriched for terms such as protein channel activity, and transmembrane transporter activity. Such processes could be associated with endocrine cells that function through membrane secretion, such as cells in the pancreas (Murtaugh and Melton, 2003; Guo and Hebrok, 2009). However, when analysing the enrichment of transcriptional signatures, it was observed that the timepoint 4 UNC1999 inhibited samples have some enrichment overlap with cells that are not pancreatic tissue. These samples, as a result, have lower pancreatic enrichment by Keygene analysis. The enrichment calculation considers all transcripts and therefore the lower pancreatic enrichment may be caused by the upregulation of non-pancreatic cell transcripts, as opposed to a loss of pancreatic phenotype. In support of the maintenance of pancreatic phenotype, very little/no change is observed with pancreatic hormone expression between the UNC1999-treated and DMSO samples. Similarly, little change, or a small increase was observed in the expression of the 87 pancreatic enriched transcripts in UNC1999 timepoint 4 treated samples compared to the DMSO control. Together this data suggests a more general upregulation of repressed genes, as opposed to the targeted upregulation of pancreatic genes.

Although gene upregulation does not appear to be specific to pancreatic transcripts, one of the key genes that was upregulated was *MAFA*. This transcription factor is important for β-cells and has previously been classified as a master regulator of many β-cell genes, potentially playing a role in maintaining β-cell phenotype, metabolism-secretion coupling and proinsulin processing (Matsuoka *et al.*, 2003; Wang *et al.*, 2007). MAFA also combines with other factors to cause the direct conversion of certain pancreatic cell types (Zhou *et al.*, 2008; Lima *et al.*,

2012) and liver cell types (Banga *et al.*, 2012; Cim *et al.*, 2012) towards endocrine cell types in both mouse and rat cells. MAFA has this ability to direct differentiation when expressed in combination with NGN3 and PDX1 transcription factors, both of which are expressed during the *in vitro* pancreatic differentiation. Therefore, the upregulation of this gene in the timepoint 4 experiment could be an exciting prospect in further differentiating the cell towards mature pancreatic β-cells. However, with other, non-pancreatic transcripts also upregulated in the cells, there is the potential that these other factors may act against MAFA and cause undesired transcriptional changes. To determine which of these outcomes will occur with EZH inhibition at timepoint 4, the differentiation protocol should be extended, and the cells analysed after a sufficient time to allow the MAFA upregulation to potentially exert an effect in the cells, in which it is now expressed.

4.4.3 Short and long-term effects on histone modifications after epigenetic perturbation are highly variable.

To analyse the effect of EZH inhibition on histone modifications over a long-term *in vitro* differentiation protocol, H3K4me3 and H3K27me3 modifications from samples inhibited at different timepoints were analysed. Initial analysis of H3K27me3 levels between DMSO and UNC1999-treated samples demonstrated that timepoint 2 and timepoint 4 had the largest effect on the levels of H3K27me3 modifications. The histone modifications at these timepoints were therefore analysed in more detail to determine the effects of epigenetic perturbation during differentiation at an earlier timepoint, with a longer recovery period, and at a later timepoint, with a shorter recovery period.

Analysis of H3K27me3 showed a decrease at TSS globally. Shortly after PRC2 inhibition, there was a strong reduction in H3K27me3 at all previously H3K27me3 modified TSS. This demonstrates the expected response to acute loss of PRC2 activity. However, in the timepoint 2 experiment, although the average level of H3K27me3 is lower in all of the defined gene categories, a complete loss of H3K27me3 is not observed, with normal levels of H3K27me3 present at some TSS. This suggests that during *in vitro* differentiation, there is a memory that helps to restore H3K27me3 levels at regions previously modified by H3K27me3. It is unclear if these regions will have retained H3K27me3 in the presence of the inhibitor, and therefore the increase in the methylation levels are due to positive feedback from the remaining modification present, or the modification will be added *de novo.* Although it is observed in the timepoint 4 samples that some H3K27me3 persists although at much reduced levels. As the

histone modification analysis was undertaken 24 hrs post-inhibition removal, it is likely some levels will also persist in timepoint 2, shortly after inhibition and therefore positive feedback may occur from these regions.

In the timepoint 4 experiment, H3K4me3 levels increased at genes that are typically modified by H3K27me3 only and therefore lack H3K4me3 in control samples. This suggests that the loss of H3K27me3 at these sites following PRC2 inhibition allowed the loci to be targeted quickly and efficiently with H3K4me3. Interestingly, this did not consistently lead to an increase in expression from these newly modified genes, although some genes within this group did increase, however overall there was no specific enrichment in this group of genes compared to the other categories. Genes that were originally marked by H3K27me3 and H3K4me3 also showed some increase in H3K4me3, although this was not to the same levels as for the H3K27me3-only modified genes. Similarly, the expression of genes that gained *de novo* H3K4me3 did not cause a consequential increase in gene expression. This was surprising because the loss of H3K27me3 and gain of H3K4me3 is strongly associated with an increase in transcription (Schuettengruber *et al.*, 2017). This disconnect could be caused by a number os factors including a delay in the responsiveness in the transcription of these genes, as the transcripts are measured only 24 hrs after the final day of inhibition. The histone modification changes may have only very recently occurred and have yet to enforce a measurable gene regulatory effect. However, this response may have been caused due the nature of histone modifications within the cellular environment.

As previously discussed, the general model for bivalency in the cell is that the CpG-rich promoters will recruit both PRC2 and trithorax protein thereby depositing active and repressive modifications with some studies suggesting this is the default state of the in these regions (Wachter *et al.*, 2014). However, in the presence of H3K27me3, this will recruit PRC1 and also self-propagate resulting in spreading of both Polycomb marks, through which the chromatin will become condensed, and the gene silenced (Hansen *et al.*, 2008; Margueron *et al.*, 2009; Oksuz *et al.*, 2018). In contrast, in the presence of H3K4me3, and active transcription, PRC2 is inhibited by this medication and cannot deposit H3K27me3 (Schmitges *et al.*, 2011), and inducing silencing. In the case of bivalency H3K4me3 will still allow the binding of PRC2 to occur by only in an asymmetric manner, thereby creating a balance and holding the gene expression in the poised state associated with bivalency.

Interestingly, perturbation did not apparently cause a disruption in the balance of the histone modifications in this study. Although the H3K27me3 was lost, and at the previously bivalent modified genes, some gain of H3K4me3 was observed, this did not cause a significant increase enough to apparently shift the balance and induce transcription. This agreed with the previous results of global PRC2 knockouts/loss of function studies showing minimal effect on gene transcription (Clouaire *et al.*, 2012; Hödl and Basler, 2012; Howe *et al.*, 2017). Instead, the lack of transcriptional response in timepoint 4 combined with the earlier timepoint 2 data showing recovery of H3K27me3 suggests that an epigenetic inheritance mechanism may have occurred. Within the cell cycle H3K27me3 regions are found to be relatively stable, and almost all H3K27me3 is found to re-established as de novo modifications (Reverón-Gómez *et al.*, 2018). Furthermore, introduction of PRC2 after knockout shows accurate re-establishment (Højfeldt *et al.*, 2018) as observed in this study. This suggests that further mechanisms are maintaining the bivalency within the genome. Numerous other histone modifications exist within the cell which may act to retain the gene expression. Although the PRC2 is inhibited and H3K27me3 is lost the PRC1, which is recruited by H3K27me3 (Blackledge, Rose and Klose, 2015; Wiles and Selker, 2017), is still present and therefore H2AK119ub have been deposited and retained at some level on the previously H3K27me3 modified regions. The H2AK199ub is known to recruit PRC2 and therefore if it is still present then upon reintroduction of functional PRC2, H3K27me3 levels will be restored at the previously modified regions. The lack of transcriptional changes may also contribute to the re-establishment of H3K27me3 at previously modified regions. Active transcription is known mechanism to inhibit the addition of H3K27me3 (Riising *et al.*, 2014) at gene regions, however the presence of CpG-rich promoters in these regions is sufficient to recruit PRC2 (Mendenhall *et al.*, 2010; Lynch *et al.*, 2012; Jermann *et al.*, 2014; Riising *et al.*, 2014). Therefore, if the PRC2 is not blocked from these regions through transcription, then it may be able to re-establish in these silenced regions.

This role of transcription in re-establishment is difficult to determine as one of the most complex questions in studying histone modifications is the inability to determine the cause and consequence of the modification in terms of gene expression. Many previous papers have tried, with numerous PRC2 knockouts/loss of function studies performed, numerous of which see a minimal effect on transcription (Chamberlain, Yee and Magnuson, 2008; Shen *et al.*, 2008; Leeb *et al.*, 2010). This would suggest that H3K27me3 has little if any effect on cells, however inducing transcriptional changes in the cell after loss of H3K27me3 can be seen under some circumstances to completely abrogate the normal transcriptional response (Faust *et al.*, 1998;

O'Carroll *et al.*, 2001; Pasini *et al.*, 2004; Collinson, Amanda J Collier, *et al.*, 2016; Moody *et al.*, 2017). However equally transcription itself appears to be a direct effector of PRC2 mediated H3K27me3 addition, with transcriptional inhibitors resulting in ectopic PRC2 binding (Riising *et al.*, 2014). Therefore, the question is difficult to answer, is the presence of H3K27me3 only caused as the gene was not actively transcribing and marked by activating modifications, allowing the H3K27me3 binding, and to retain the already inactive gene in a further repressed state? Or did the H3K27me3 become bound at the gene first, while some active transcription was still occurring and then the spread of H3K27me3 through PRC2 self-propagation and the PRC1 recruitment (Blackledge *et al.*, 2014; Cooper *et al.*, 2016; Fursova *et al.*, 2019; Margueron *et al.*, 2009; Poepsel, Kasinath and Nogales, 2018; Perino *et al.*, 2020; Oksuz *et al.*, 2018) induce proper silencing of the gene? Global PRC2 knockout experiments do not answer this, as it is difficult to separate out the contribution of other factors such as chromatin confirmation and other histone modifications. This could instead be investigated through targeted epigenome editing further employing techniques such as synthetic constructed promoter-gene DNA sequences. Testing these in both an isolated setting, and then introducing into different cellular system such as E.coli and mESCs may determine if either the cause and consequence of the H3K27me3. Experiments using these thecniques have been trialled before, such as that by O'geen et al, 2017, using a dCas9-Ezh2 tethered construct, showing little effect of inducing H3K27me3 on gene expression levels. However, this study focused on creating tools to use in studies, and few conditions were tested , therefore further application of this tool under more conditions would be informative (O'Geen *et al.*, 2017)

In regards to pancreatic differentiation, an important previous study showed that the inhibition of PRC2 at a late stage of in vitro differentiation increased the levels of C-peptide within the final population culture (Xu *et al.*, 2014). This study implied a positive effect either on the differentiation abilities of the β-cell population or an increase in the proportion of these cells types after PRC2 inhibition (Xu et al., 2014). However, Xu and colleagues also demonstrated that PRC2 inhibition at an earlier stage of differentiation had a small reduction/no change in the C-peptide expression of the final population. Similarly, inhibition at a later stage of differentiation had a potentially positive effect on pancreatic transcripts, including positive changes in gene expression that are normally differentially expressed within the *in vitro* derived cells when compared to adult islet cells. This includes important transcription factors, such as *MAFA*, which may contribute to a positive effect on the production of pancreatic cell *in vitro*. Similarly, ChIP-seq analysis demonstrates that a large shift in histone modifications occurs after the loss of

H3K27me3 inhibition. Earlier inhibition suggests some recovery of these histone modifications will occur but the consequences of this on gene transcription and resulting cell types would be interesting to determine through prolonging the culture of these inhibited cell types.

Chapter 5

Introduction of Inducible Systems to Knockdown EED in Human Pluripotent Cells

5.1 Background

Previous results from this and other studies (Xie *et al.*, 2014; Xu *et al.*, 2014) have indicated a role of the PRC2 complex in the *in vitro* differentiation of pancreatic endocrine cells, in a potentially stages specific manner. To interrogate the role of this epigenetic complex in the developing cells, small molecule inhibitors are used to inhibit the function of the complex. Those small molecule inhibitors currently available to target PRC2 activity function by targeting the EZH1 and EZH2 catalytic components of the PRC2 complex. However, these inhibitors have limitations when used to treat cultured cells to study the function of an epigenetic complex during differentiation. Although these compounds allow an effective means for studying the outcomes of reduced H3K27me3 levels in the differentiation, the PRC2 complex will still be present in the cell, and likely bound to chromatin, thereby having unknown effects at these regions. Loss of any of the PRC2 core components, such as EED will induce destabilisation of the complex (Margueron and Reinberg, 2011; Oksuz *et al.*, 2018; G. van Mierlo *et al.*, 2019), thereby removing complex from the cell, and overcoming this problem. Thereby the development of genetic targeted system could be used to further validate key findings from the experiments utilising small molecule inhibitors. The introduction of genetic inducible loss of function systems could also be beneficial in allowing targeting of the epigenetic modifiers in the cells. The production of small molecules to PRC2, was undertaken due to a associated role of the protein with cancer cells. These inhibitors specifically target the PRC2 complex by acting in competition with the cofactor SAM at the EZH1/2 binding site (Verma *et al.*, 2012; Konze *et al.*, 2013) meaning these are not applicable to any of the other epigenetic modifiers that are likely to function throughout the pancreatic differentiation. The introduction of system designed to *EED* within hPSCs can act as a proof of principle for inducing targeted loss of function, that could

then be applied to other complexes. Thereby, introduction of a complementary genetic loss of function system targeting PRC2 would confirm the role of the complex during *in vitro* differentiation and allow the interrogation of other epigenetic factors in this process.

Deletion of PRC2 components has negative consequences for maintaining cultures of human pluripotent stem cells, including loss of proliferation and self-renewal (Collinson *et al.*, 2016; Shan *et al.*, 2017). The limitation could be overcome through an inducible loss of function system while still allowing depletion of targeted protein. Several different targeting methods are available to create inducible knockouts, this includes an auxin-inducible degron (AID) and a CRISPR interference (CRISPRi) system. Both of which could be used to target EED for depletion in hPSCs. These two systems target differing points in a protein life cycle. The AID system will target translated proteins for degradation, through the use of small degron tags and exogenous expressed proteins, introduced by genome editing into the cells. The degradation of tagged proteins will only occur upon the addition of defined small molecules to the cell media (Nishimura et al, 2009). In contrast, the CRISPRi system acts by blocking mRNA transcription of target genes upon the induction of a repressor protein (Mandegar *et al.*, 2016). The variation in the two targeting mechanisms have associated benefits and drawbacks, and by developing both approaches it would allow the choice downstream about which produce the better targeting methods in different circumstances.

The AID system functions to allow the rapid and reversible depletion of a targeted protein by replicating a process that occurs naturally in plant cells. In plants, cells will use the production of the hormone auxin to induce degradation of the AUX/IAA family of transcriptional repressors, in a time responsive and reversible manner (Nishimura *et al.*, 2009). This degradation system depends on two components specific to the plant system, the expression of a Transport Inhibitor Response 1 (TIR1) protein and the presence of a degron tag on the protein targeted for destruction. The TIR1 protein utilises the E3 ubiquitin ligase machinery to remove target proteins by proteasomal degradation, through ubiquitin addition (Figure 5.1). This system can be manipulated for use in other cell types provided they also express the Skp1, Cullin and F-box (SCF) E3 ligase complex components, required for the addition of ubiquitin (Nishimura et al, 2009). To determine which protein is the target of the TIR1, a short degron sequence must be present. Optimised for use in human cells is a 68 amino acid (7.4 kDa) region of the IAA17 degron tag, referred to as the mini-AID (mAID) (Natsume et al, 2016). Once a protein is tagged, the addition of Indole-3-acetic acid (IAA) to cell media should cause the OsTIR1 protein to bind the mAID sequence and target the tagged protein for proteasomal degradation, thereby

removing the protein specifically and quickly from the cell. Once the IAA is removed from the media and depleted from the cell degradation will cease and protein levels will gradually recover (Nora *et al.*, 2017; Rao *et al.*, 2017).

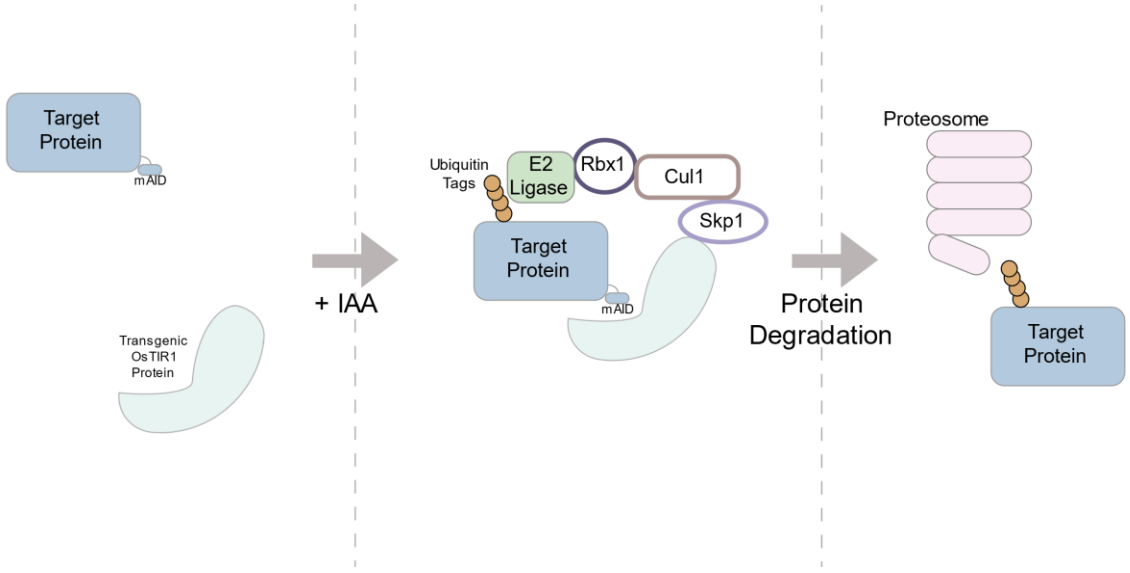


Figure 5.1 Schematic of auxin-inducible degradation system. Demonstrating the destruction of a protein by the endogenous E3 ligase/proteasomal pathway, through the auxin degradation system upon the addition of IAA; OsTIR: *Oryza sativa* transport inhibitor response 1 protein: mAID; mini-auxin inducible degron: IAA; Indole-3-acetic acid

The second system selected to target *EED* for depletion in hPSCs is the CRISPRi system. This utilises a catalytically inactive Cas9 protein fused to a repressive KRAB domain to silence a target gene (Gilbert *et al.*, 2013; Qi *et al.*, 2013). The fused Cas9-KRAB protein is endogenously expressed in a cell through genomic editing and is then directed to specific genomic locations through a short guide RNA (gRNA) sequence. The gRNAs are also endogenously expressed in the cell through the random insertion of a gRNA expression system into the genome under then control of a constitutive promoter (Mandegar *et al.*, 2016). To allow inducibility of the system, the Cas9-KRAB is under the control of TRE3G promoter which is only active in the presence of doxycycline (Figure 5.2). Therefore, only upon addition of doxycycline

to the media, the repressive complex will form and silence a target gene. Upon removal, the Cas9-KRAB protein will gradually be destroyed and the repression will be removed (Mandegar *et al.*, 2016).

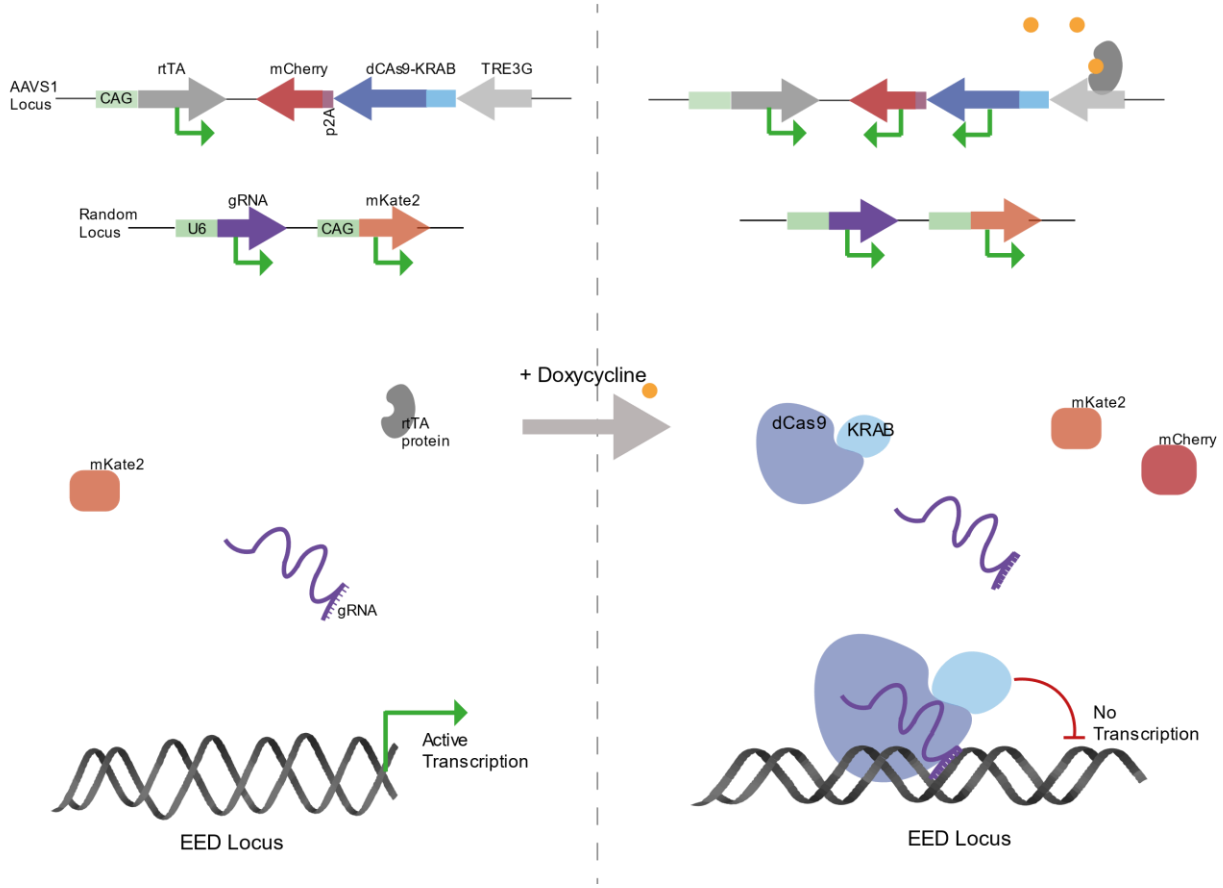


Figure 5.2 Schematic of CRISPR interference gene repression. Under normal conditions a target gene, such as *EED*, will be undergoing active transcription but addition of doxycycline to cell culture media induces activation and formation of a repressive complex at the target gene. gRNA, guide RNA; rtTA, reverse tetracycline-controlled transactivator; dCAS9-KRAB, nucelase-dead Cas9 Krüppel-associated box; TRE3G, TRE3G promoter.

Both knockdown systems work at different levels, with the AID system targeting the protein and the CRISPRi targeting transcription, resulting in functional differences. One key difference between the systems is the time taken for complete removal of the protein from the cell. The CRISPRi system acts to repress at the gene level, therefore it can take days to

observe complete loss of a protein from a cell (Mandegar *et al.*, 2016), as this requires endogenous protein degradation. In contrast, the AID targets the protein directly meaning it can be destroyed within hrs of IAA addition (Natsume et al., 2016; Nora et al., 2017). A weakness with the degron system is due to the presence of the mAID, which may have unknown effects on the target protein, this could include inhibiting function, or changing the protein localisation, and these potential drawbacks are only observed after serial rounds of gene editing. This could potentially lead to the production of a system which results in a non-functional protein after a long and laborious process. One of the major limitations of the auxin inducible system is the basal degradation that may occur in the targeted protein (Li et al., 2019; Zasadzińska et al., 2018), resulting in a knockdown phenotype before initiation of the system. The AID system is also a less preferable system when intended to target multiple genes, as each gene of interest will result in a long process of multiple rounds of gene editing, screening, and functional testing

Therefore, the AID and CRISPRi systems each have difference in terms of the weaknesses and strengths associated with the systems, in terms of introduction to the cells and the mechanisms through which they induce knockdowns. Due to these differences between the systems, both were taken forward in targeting EED in hPSCs, which would then allow a comparison for which is more appropriate in downstream applications. Both have been previously introduced and successfully used in human and mouse pluripotent cells (Mandegar *et al.*, 2016; Nora *et al.*, 2017; Sybirna *et al.*, 2020), with the AID successfully being targeted to Polycomb proteins in mECS (Rhodes *et al.*, 2020). However, the introduction of either system in hPSCs to target the Polycomb proteins has not been reported at this time.

5.2 Hypothesis and Aims.

Human pluripotent cell types are an important cell model in studying the function of complex molecular mechanisms during development, including pancreatic cell differentiation. Attempting to study the function of individual complexes can be difficult without efficient manipulation techniques, as complications can arise in separating cause and consequence of interference in a system. Although small molecule inhibitors offer a suitable mode of perturbation for a number of applications, these reagents have a number of weaknesses. To address these limitations, previously developed targeted depletion systems were selected which would allow manipulation of the PRC2 complex, through EED protein levels, in human pluripotent cells. It was hypothesised that the successful application of an inducible loss of function system would allow the study of the PRC2 and other epigenetic complexes in hPSCs and during differentiation, whilst allowing the untreated cells to function without perturbation before induction.

To investigate this hypothesis this chapter has three aims:

1. Introduce the Auxin-Inducible Degron (AID) system targeting endogenous EED protein into hPSCs.
2. Introduce the CRISPR interference system targeting endogenous *EED* transcription in hPSCs.
3. Compare the ability of the AID and CRISPRi to reversibly deplete the EED protein.

5.3 Results

5.3.1 Design and attempted production of an auxin inducible EED degron system in human pluripotent cells.

To introduce the auxin inducible degron mechanism targeting a selected gene into pluripotent cells, a number of steps are required. The order in which I attempted to introduce all necessary components to target *EED* using the AID system, is demonstrated in Figure 5.3. Firstly, the exogenous *OsTIR1* gene was targeted to the human genomic safe harbour adeno-associated virus integration site 1 (AAVS1), which would induce stable expression with no disruption to normal cell transcription (Papapetrou and Schambach, 2016). The inclusion of homology arms in the OsTIR1 plasmid allows specific targeting to the *AAVS1* site, and the puromycin resistance gene was present between these homology arms, to be used as a positive selection marker for transgene insertion. Secondly, the insertion of the mAID tag at one allele in the endogenous *EED* gene. For targeting, a plasmid containing the mAID short sequence flanked by homology arms was used, which would target the mAID sequence to be inserted directly before the endogenous stop codon sequence of the gene. In the EED-mAID targeting plasmid, an antibiotic resistance marker was included within the homology arms, this induced resistance to hygromycin in the cells after successful insertion. Due to the low targeting efficiency in human pluripotent cells (Lombardo et al., 2007; Steyer et al, 2018), the EED alleles were targeted in sequential subsequent experiments. Firstly, one of the endogenous EED alleles was targeted, selected for positive insertion and then, using a variation of the EED-mAID targeting plasmid with blasticidin resistance, the second allele was targeted (Figure 5.3).

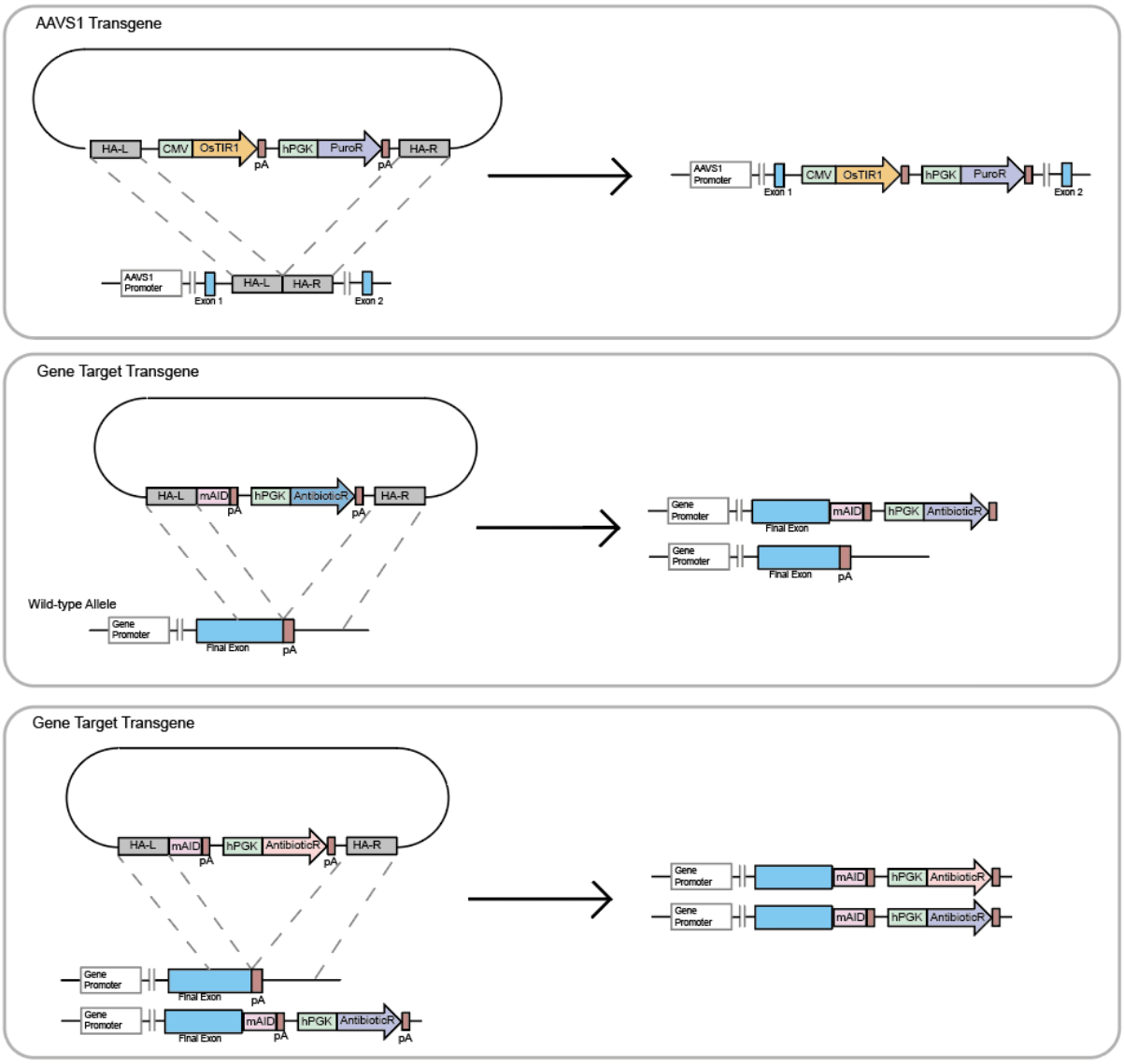


Figure 5.3 Introduction of auxin inducible system components. Schematic demonstrating the steps required to introduce the function auxin inducible targeting system into human pluripotent cells to target a gene of interest. HA-L/HA-R: Homology arm left/right; PGK: phosphoglycerate kinase 1 promoter; CAG: CMV early enhancer, chicken β-actin, and rabbit β-globin hybrid promoter; OsTIR: *Oryza sativa* transport inhibitor response 1 protein mAID; mini-auxin inducible degron; PuroR/HygroR/BlastR: Puromycin/Hygromycin/Blasticidin resistance gene; pA: polyadenylation signal.

To incorporate the mAID sequence to the targeted site at the *EED* gene CRISPR cleavage was utilised. A Cas9 expression plasmid was introduced to the cells to induce cleavage of the genome at target sites, as directed by a gRNA sequence. Nucleofection was initially used to introduce the necessary plasmids into the cells, but a large amount of cell death occurred, and so other techniques were explored to introduce plasmids into human pluripotent cells. One method selected was Genejuice® transfection reagent, a non-lipid based chemical transfection reagent. As the Cas9 expression plasmid has a green fluorescent protein (GFP) gene, the percentage of cells positive for GFP can be used as an indicator of successful transfection of cells (Figure 5.4). The Genejuice® solution to DNA ratio was optimised, indicating that 6 µl of the reagent solution added with 3 µl of DNA was the most efficient concentration of transfection reagent to use (Figure 5.4a). The efficiency for this was only slightly more than 3 µl of solution to 1 µl of DNA, but multiple plasmids are required for each transfection so the condition that would allow higher volumes of DNA insertion was used in downstream experiments. The transfection in either hypoxia or normoxia had very similar effects on efficiency (Figure 5.4b). The transfection of cells in mTeSR™ produced a much higher reproducible transfection compared with TeSR-E8™, higher by approximately 15% (Figure 5.4c). Although this was lower than the efficiency typically obtained when using nucleofection, which is between 25 - 35% (data not shown), the survival of cells after Genejuice® treatment was much higher. As antibiotic selection is used soon after plasmid introduction, and higher cell survival causes higher density, which is beneficial in selection, both nucleofection and transfection were used in parallel to introduce the plasmids into cells.

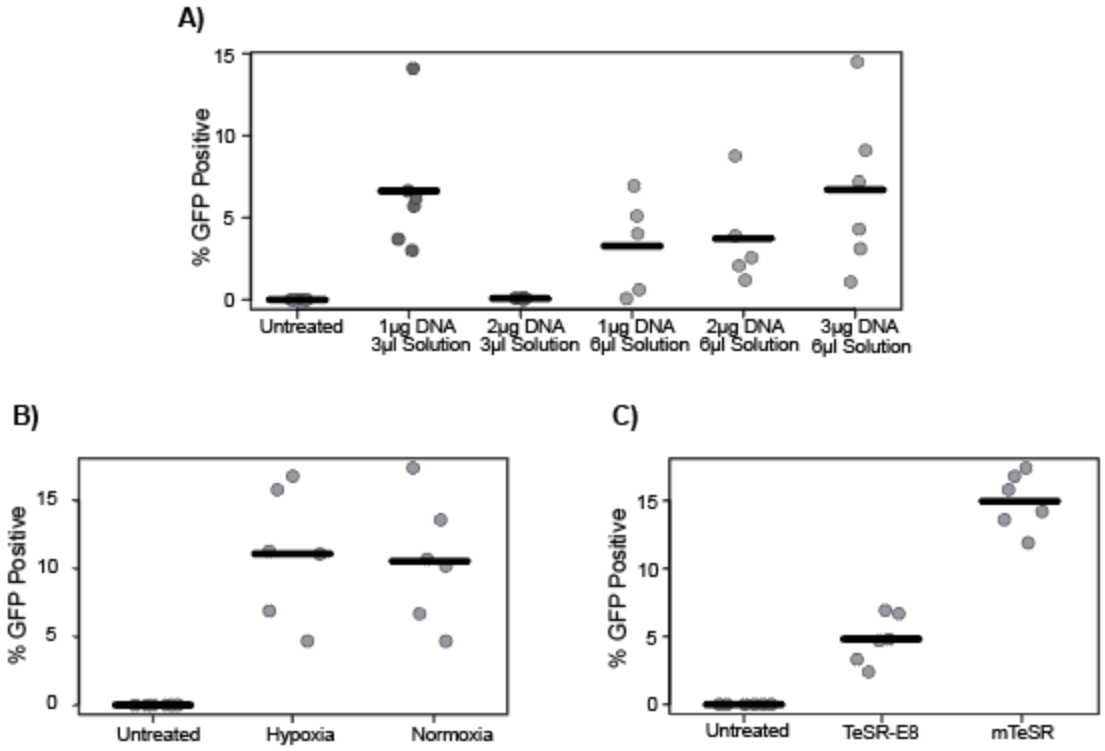


Figure 5.4. Optimisation of Genejuice® transfection reagent in human pluripotent cells. Cells were transfected using Genejuice® transfection reagent under different conditions a) with varying amounts of plasmid DNA and transfection solution b) under hypoxic and normoxic conditions c) with transfection in either TeSR-E8 or mTeSR media. To compare a transfection a Cas9-GFP plasmid was used, the resulting GFP positivity in the population was measured by flow cytometry 48 hrs post-transfection. The percentage of GFP positive cells in each culture condition are indicated by individual dots. Each of condition was repeated three times in both HDF and H9 cell lines.

To undertake positive selection of transfected populations, the OsTIR1 and the two EED-mAID targeting plasmids contained different antibiotic resistance genes, which were used to isolate colonies that have undergone successful insertion of the transgene. To determine the most suitable antibiotic concentrations to use for positive selection, survival curves were produced for puromycin, hygromycin and blasticin in both the HDF iPSC and H9 ESC wild-type lines (Figure 5.5). As the optimal concentration and timings of antibiotics can vary drastically, previously published literature (Schwanke *et al.*, 2014; Mandegar *et al.*, 2016; Steyer *et al.*,

2016) and initial exploratory experiments were used as a guide to determine the timing, in terms of days, and the range of antibiotic concentration to trial. When performing the experiments to determine the antibiotic selection, multiple plating densities were used because it is difficult to determine the density that the cells will be post-nucleofection or transfection, therefore this range helped to identify the concentration that would select resistance regardless of cell numbers (Figure 5.5).

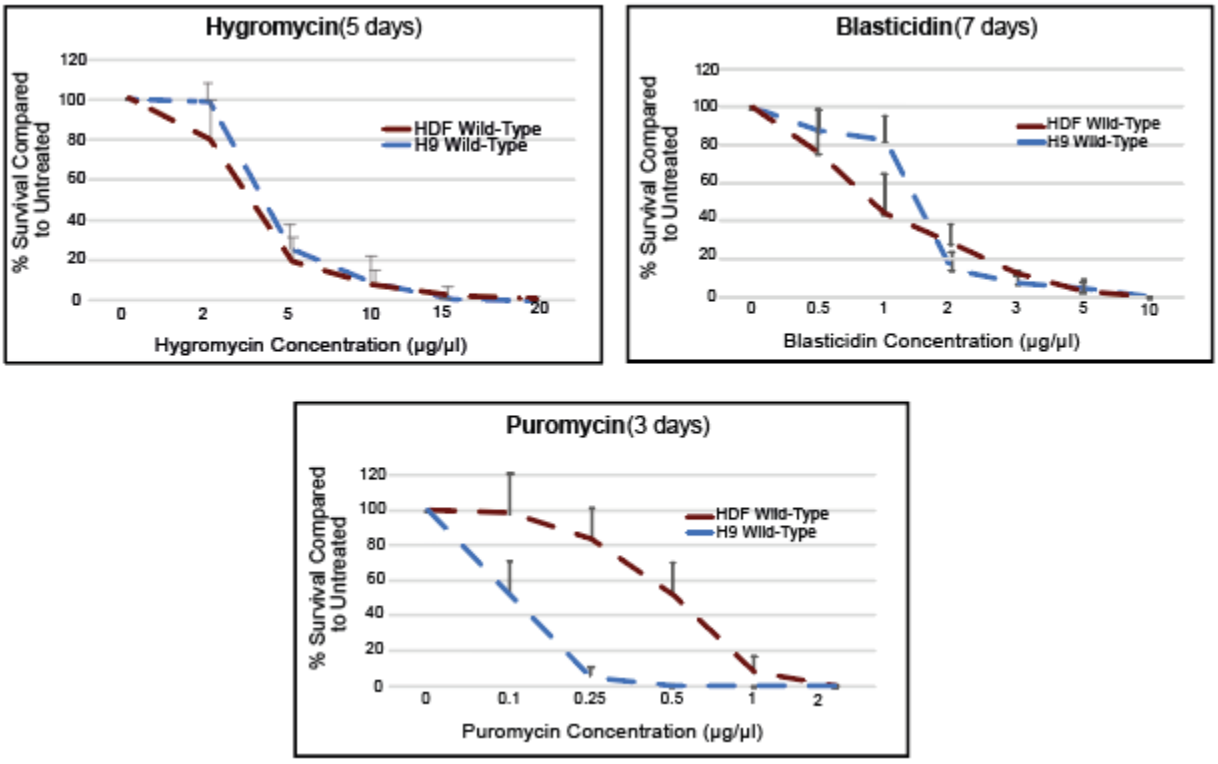


Figure 5.5. Optimisation of antibiotic concentration. Survival curves for a range of concentrations were undertaken for the three antibiotics as indicated. Cells were plated and then treated for the specified day with antibiotic addition to culture media. Cells were then counted and percentage survival was compared using an untreated control. Repeated 3 times, using varying plating densities for HDF (red) and H9 (blue) wild-type cell lines, error bars represent standard deviation.

Next, the OsTIR1 plasmid was introduced into cells by transfection and nucleofection. After 48 hrs, the H9 line was selected with 0.5 µg/ml of puromycin and the HDF line with 1 µg/ml

for 72 hrs (Figure 5.5). Surviving colonies were grown for between 7 and 10 days, and the expanded clones were picked into individual wells. To ensure the stable insertion of the puromycin gene, which would also indicate the stable integration of the OsTIR1 gene, puromycin selection was reapplied to all colonies. Both cell lines were treated with 1 µg/µl puromycin to select for stable transfected cells. The DNA from multiple clones was then isolated and genotyped (Figure 5.6). Initially, genotyping was attempted using primers outside of the homology arms of the AAVS1 target site, however no band was produced from these reactions, for either wild-type or resistant clones. It was hypothesised that the inability to produce a PCR product in wild-type cells may be caused by the highly CG-rich region surrounding the target site, and the use of large homology arms, each over 800bp, requiring a large PCR product. To overcome this an alternative genotyping strategy was used that targeted the puromycin and OsTIR1 regions within the insert to confirm the presence of these genes (Figure 5.6a). All cell clones screened were positive in both PCRs, with no banding for the insert observed in the wild-type cells (Figure 5.6b). It was therefore concluded that targeting of the OsTIR1 to the AAVS1 occurred successfully at a high efficiency when using positive selection. However, this method of genotyping has two limitations: first, positive results could be caused by the presence of non-integrated plasmid within the cells, and second, it does not confirm the integration into the targeted region, as both targeted and random insertion would produce the same banding pattern.

To overcome this, a genotyping strategy was used in which primers will target the OsTIR1 plasmid backbone (Figure 5.7a). By using the plasmid backbone PCR, this would indicate the presence of a non-integrated plasmid or integration at a random site. A clear band was detected in the positive control PCR (plasmid DNA) but no band was observed in any of the cell samples, suggesting that very little or no plasmid remains in the cells when genotyping was undertaken. The lack of any positive banding also indicates that targeted integration of OsTIR1 occurred only at the correct site, as random integration would not utilise the homology arms and therefore would be likely to incorporate some of the backbone, as well as the desired insert. It was concluded from both genotyping experiments (Figure 5.6 and 5.7) that the OsTIR1 transgene was successfully integrated at the desired genomic site in several clones for both the HDF and H9 hPSC lines.

As the cells survive in puromycin selection it is predicted that the inserted transgene is constitutively expressed, however the *OsTIR1* gene is under separate promoter from the puromycin resistance. To ensure the *OsTIR1* gene is also constitutively expressed in our

hPSCs an RT-qPCR was carried out using primer pairs that targeted two distinct regions within the OsTIR1 cDNA (Figure 5.8). Comparison of the wild-type non-targeted line and those that showed positive integration by genotyping, demonstrates the expression of OsTIR1 mRNA within the cells.

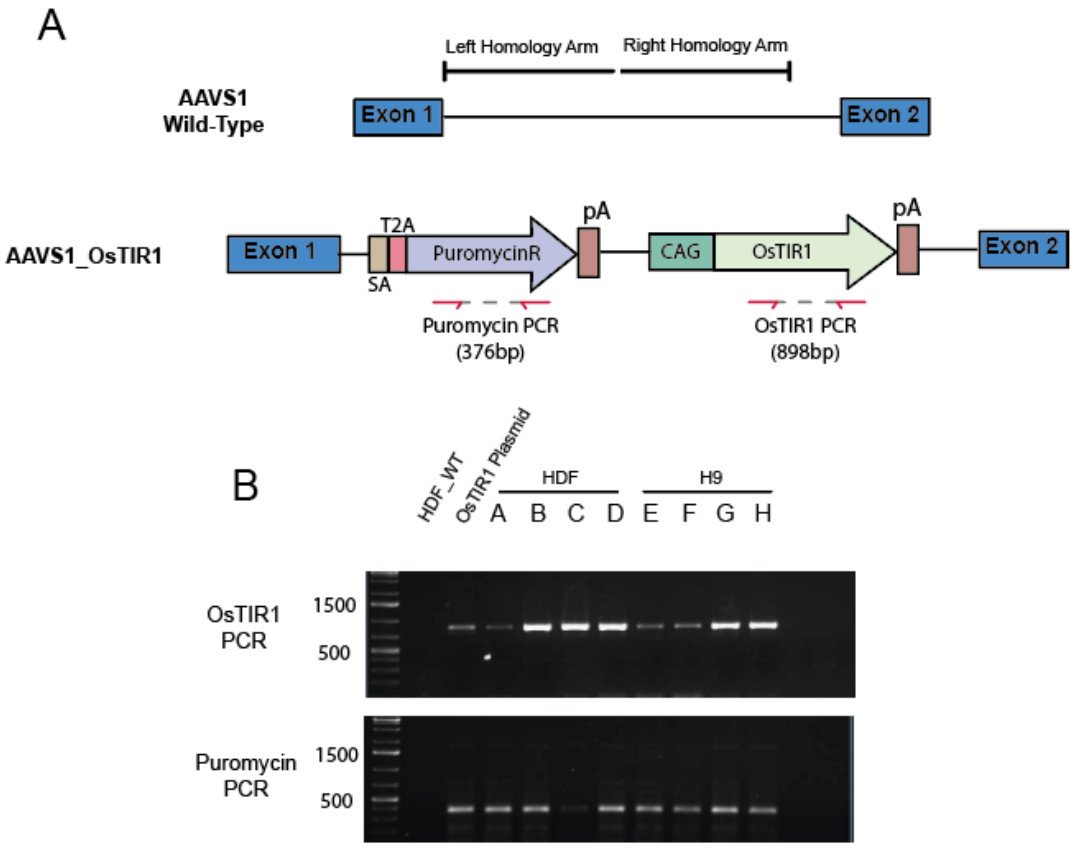


Figure 5.6. Genotyping of OsTIR1 insertion. (A) Schematic demonstrating the successful transgene insertion into the endogenous AAVS1 target site. The PCR genotyping strategy used to determine presence of the transgene is indicated. (B) Genotyping of clones following puromycin selection (lanes A-H). The HDF_WT and OsTIR1 Plasmid DNA was extracted and used as a positive control. SA: splice acceptor; T2A: self-cleaving peptide; 1kb: GeneRuler™ 1 kb Plus DNA Ladder.

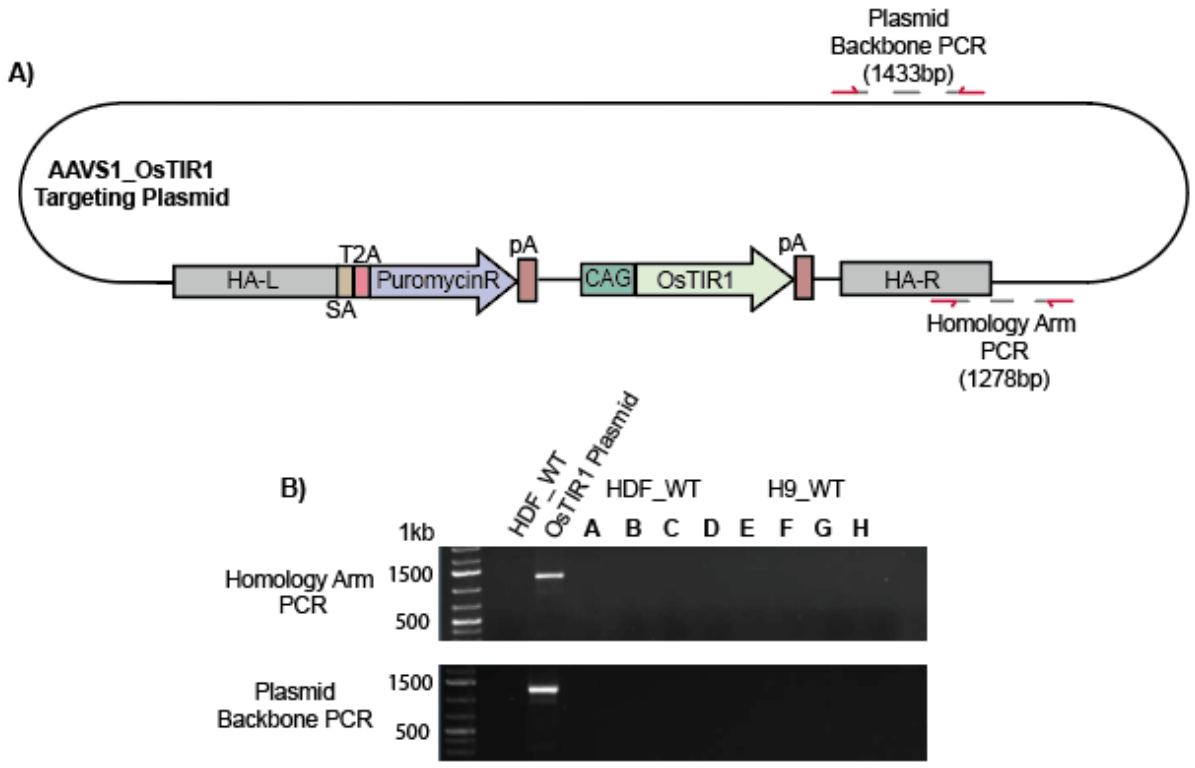


Figure 5.7. Targeted insertion of OsTIR1 transgene. (A) Schematic of AAVS1_OsTIR1 targeting plasmid. The PCR genotyping strategy used to determine presence of the transgene is indicated. (B) Genotyping of clones surviving puromycin selection, HDF_WT and OsTIR1 plasmid DNA were used as a negative and positive control, respectively.

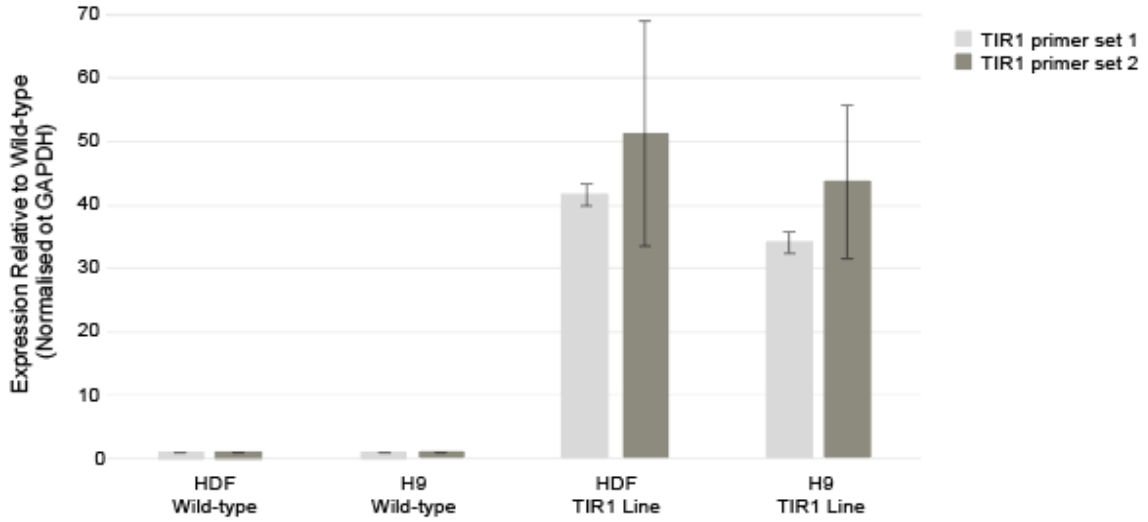


Figure 5.8. Constitutive Expression of OsTIR1. RT-qPCR demonstrates the mRNA expression of OsTIR1 transgene in targeted lines using 2 separate primers sets, n=3, error bars indicate standard error.

After establishing stable cell lines expressing OsTIR1, the next step was to introduce the mAID tag at the EED locus just prior to the stop codon (step 2, Figure 5.3). To introduce the mAID tag at the targeted locus a plasmid containing hygromycin resistance was used thereby allowing positive selection of successfully integrated clones. After selection of cells in 10 µg/ml hygromycin for 5 days, surviving colonies were allowed to grow for between 7 and 10 days and individual colonies were isolated and taken forward. To ensure the stable insertion of the hygromycin resistance gene, hygromycin was reapplied at 10 µg/ml for 5 days. Surviving colonies were then screened for the presence of the transgene by PCR (Figure 5.9). To identify a wild-type EED allele, primers targeting regions outside of the homology arms in the genomic DNA were used. When using these primers on a successfully altered allele, a PCR product would not be produced due to the length of the product (>4000 bp). The mAID PCR will use a primer to target within the mAID sequence and a primer outside of the left homology arm. Therefore, the PCR could not amplify the template plasmid and only cells with the insertion of the mAID tag into the endogenous EED gene should produce a PCR product. Using this screening 5 out of 7 of our hygromycin surviving clones were positive for the mAID insertion. The presence of the wild-type allele in all of these clones indicates that, as predicted due to low efficiency, the clones were all heterozygous.

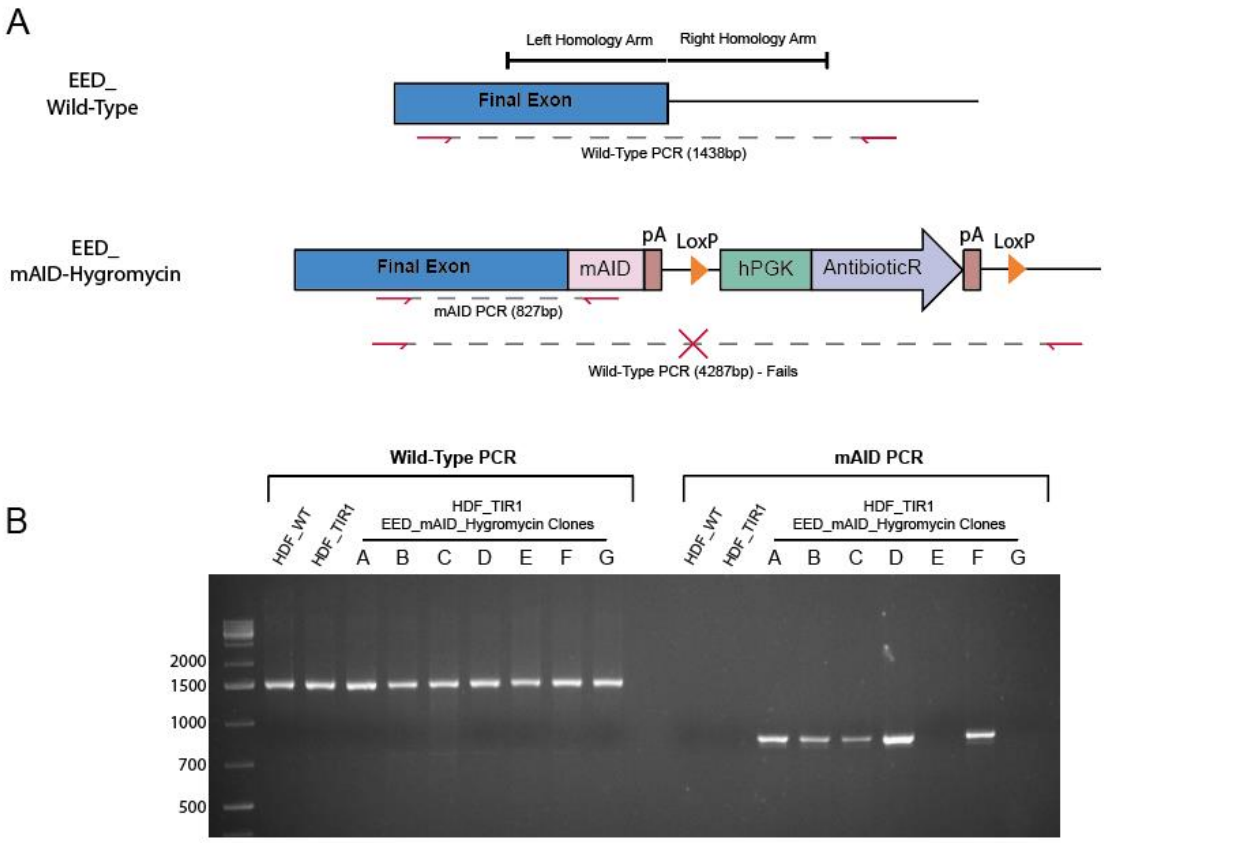


Figure 5.9. Genotyping of mAID_HygroR insertion. (A) Schematic demonstrating the successful transgene insertion into the endogenous EED target site. The PCR genotyping strategy is indicated. (B) Genotyping of hygromycin-resistant clones, HDF_WT and HDF-TIR1 are non-targeted parental lines, that act as negative controls. See Figure 5.3 and 5.6 for abbreviations.

Three of the positive clones were taken forward - HDF_TIR1 EED_mAID_Hygromycin Clones B, C and D - and their PCR products were submitted for sequencing together with an HDF_WT line wild-type control PCR (Figure 5.10). Comparing the wild-type allele in the positive genotyping clones to that of the non-transfected cell line was important to confirm that the clones retain the same sequence. In clones C and D the wild-type allele around the target site remained unchanged, but clone B has a 9 base pair region differing from wild type. The presence of N's within this sequence may indicate that something was wrong with the sequence as opposed to an issue with targeting; however, this clone was not taken forward. One base pair change was observed in the mAID sequence of all three clones, a C to A conversion. This substitution was introduced in the template in order to disrupt the recognition site of the EED

gRNA, to allow retargeting of the cells using the same gRNA sequence without inducing cleavage of the edited allele. A second conversion was also seen in all three clones in the mAID sequence, this was a T to C conversion, as this was present in all three clones it is also likely to be present in the template, and the downstream effects of this are not currently known. The final difference between the predicted sequence and actual sequence is a C to T conversion present in two of the three clones, in the left homology arm. When the sequence is converted into the amino acid sequence this base pair substitution had no effect and encoded the same amino acid sequence produced by both the wild-type allele and the known amino sequence ascribed to EED protein (https://www.uniprot.org/uniprot/O75530). It therefore concluded that two HDF lines that express OsTIR1 mRNA and have a heterozygous EED-mAID gene have been produced, with a retargetable wild-type EED allele (Figure 5.10). This sequencing indicates that the EED gene would still be functional, with the mAID tag in frame at the 3' end of the EED coding sequence, and a stop codon present after the mAID tag.

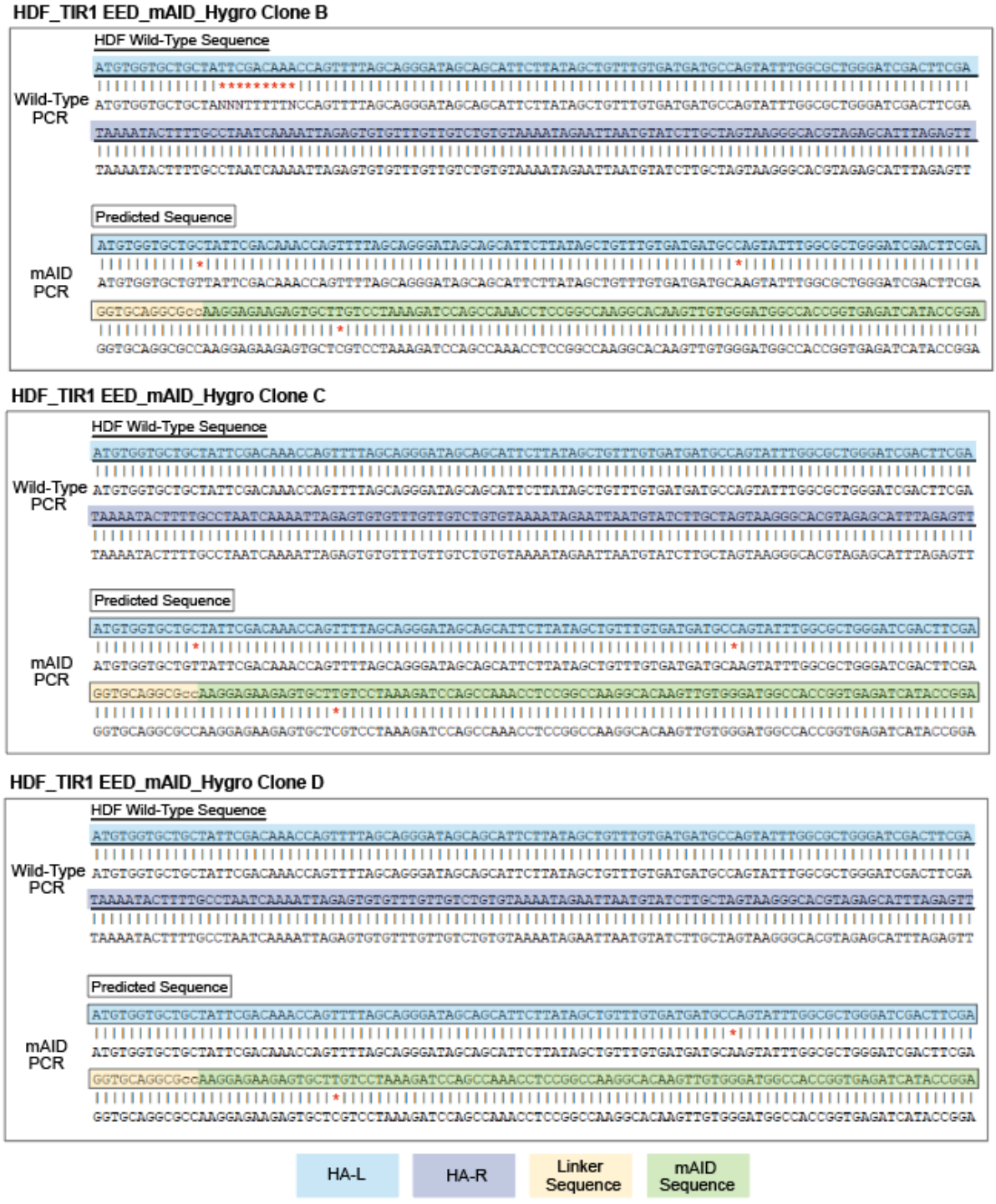


Figure 5.10. Sequencing of EED_mAID_HygroR Clones. Sequencing of PCR products produced in Figure 5.9, for wild-Type PCR the sequenced clones were compared to a non-transfected wild-type sequence and the mAID PCR was compared with a predicted sequence with successful transgene insertion. A connecting line shows the same base pair sequence; * indicates a differing base pair sequence.

The next aim was to target the second allele in our heterozygous clones using the EED_mAID-Blasticidin template plasmid, as outlined in Figure 5.3. After several unsuccessful selection experiments, it was not possible to isolate cells with resistance to the blasticidin antibiotic. One hypothesis for this is a potential inability to produce blasticidin resistance, due to problems with the resistance gene template or a difficulty in finding the necessary antibiotic concentration for selection. It was also proposed that an inability to select for resistant cells could be due to an inability to tag a second allele, as it would in some way negatively affect the function, and this would occur regardless of the selection used. To determine which of these situations is limiting the production of a homozygous knock-in cell line, two further targeting strategies were devised (Figure 5.11). The first of these involved using a template plasmid that would also fuse a fluorescent protein to mAID-tagged EED upon integration (Figure 5.11). As the mCherry2 protein would only be expressed from cells that had a second allele tagged, this would allow us to identify any successfully targeted lines when transfected in the previously isolated heterozygous lines. The second strategy utilises addition of the TAT-CRE Recombinase protein (Peitz et al 2002; Kadari et al, 2014). Due to the presence of LoxP sites within the template, addition of the Recombinase can remove the hygromycin resistance from the tagged allele and allow retargeting using the EED-mAID_Hygro plasmid, which had previous success in tagging the first allele.

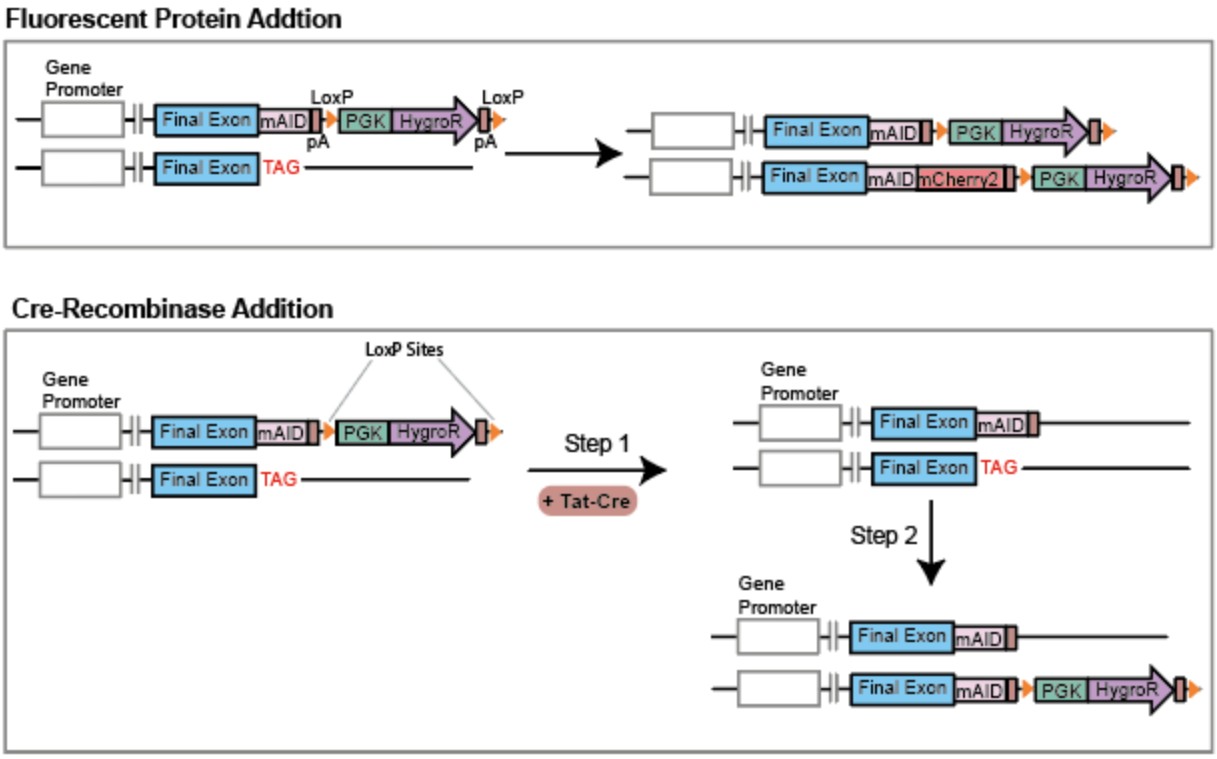


Figure 5.11. Retargeting strategies to create a homozygous mAID tagged cell line. Schematic demonstrating two methods used to introduce a mAID tag onto the second allele in a heterozygous tagged cell line.

The addition of a fluorescent protein can give a quick and easy readout to whether successful transfection had occurred, as the presence of mCherry2 could be measured by flow cytometry. To integrate the mAID tag with the fused mCherry2 protein to the EED site, a template plasmid that contained either Hygromycin resistance or Blasticidin resistance was used. These plasmids were introduced into the HDF_TIR1 EED_mAID_Hygromycin line and an HDF_WT line by nucleofection and lipofection. The transfected cells were anaylsed 48 hrs post-introduction by flow cytometry to measure the levels of mCherry2 expression (Figure 5.12a). Initially, the use of a fluorescent fusion protein to measure integrated clones appeared promising with a small but definitive population of cells demonstrating the expression of mCherry2. This gene is not preceded by a promoter within the targeting plasmid and therefore should only be expressed when integrated into the genome under the endogenous promoter of the target gene. However, the expression of mCherry2 was transient, with the initial expression almost completely depleted in the population after 5 days post-transfection/nucleofection (Figure 5.12b). The reason for this transient expression is unknown, it could be caused by non-specific

expression from the plasmid, silencing of an integrated transgene or negative selection against cells that contain an mAID-mCherry2 tagged allele.

As it is currently unclear as to why the fluorescent fusion protein could not be stably inserted, the focus shifted to utilising TAT-CRE Recombinase protein to excise the LoxP-flanked integrated antibiotic-resistance gene and allow retargeting with the same plasmid. Previous use of the TAT-CRE Recombinase method has been reported in feeder-free human hESCs, therefore these prior studies guided the concentrations used to treat the HDF_TIR1 EED_mAID_Hygromycin line (Nolden *et al.*, 2006, 2007). Higher concentrations of TAT-CRE recombinase will likely induce further LoxP excision however, this higher concentrarion could also cause excessive cell death. To overcome this six different TAT-CRE recombinase treatments were used - 0.5 µM for 5 hrs, 1 µM for 5 hrs, 2 µM for 1 hour, 2 µM for 2 hrs, and 2 µM for 5 hrs. The best readout to determine the efficiency of the TAT-CRE Recombinase in LoxP removal is to measure the susceptibility of treated cells to hygromycin. As the cells that are to be taken forward would now be susceptible to the antibiotic, the cells were first grown for 48 hrs after TAT-CRE Recombinase induced LoxP removal. A bulk population was maintained for all six TAT-CRE recombinase treatments and numerous single cell colony populations were selected from each bulk population. The six different TAT-CRE Recombinase treated bulk populations, and an untreated HDF_TIR1 EED_mAID_Hygromycin control line had 10 µg/ml hygromycin added for 5 days (Figure 5.13). After 5 days of antibiotic selection in the EED_mAID_Hygromycin control line the hPSCs have overgrown (Figure 5.13a), suggesting very little or no death has occurred as expected. Similarly, no/very little cell death was observed in three of the bulk TAT-CRE recombinase treatment conditions - 0.5 µM for 5 hrs, 2 µM for 1 hour and 2 µM for 2 hrs. After antibiotic selection cell counts were undertaken to have a more quantitative measure of how susceptible the cell populations are to antibiotics LoxP excision. This demonstrated that TAT-CRE Recombinase treatment for 2 µM for 2 hrs did appear to have less survival than untreated control, however a large proportion of the cells continued to harbour resistance to hygromycin (Figure 5.13b). In contrast, very little survival is seen in both the 1 µM for 5 hrs and 2 µM for 5 hrs treatments. This supports the observations from population imaging that demonstrates only a small number of cells retain the antibiotic resistance gene within the LoxP sites, indicating that successfully LoxP recombination has occurred in the majority of cells (Figure 5.13). Due to the high efficiency of LoxP recombination, five clonal colonies from both the 2 µM for 2 hrs, and 2 µM for 5 hrs TAT-CRE Recombinase treated samples were taken forward for genotype screening.

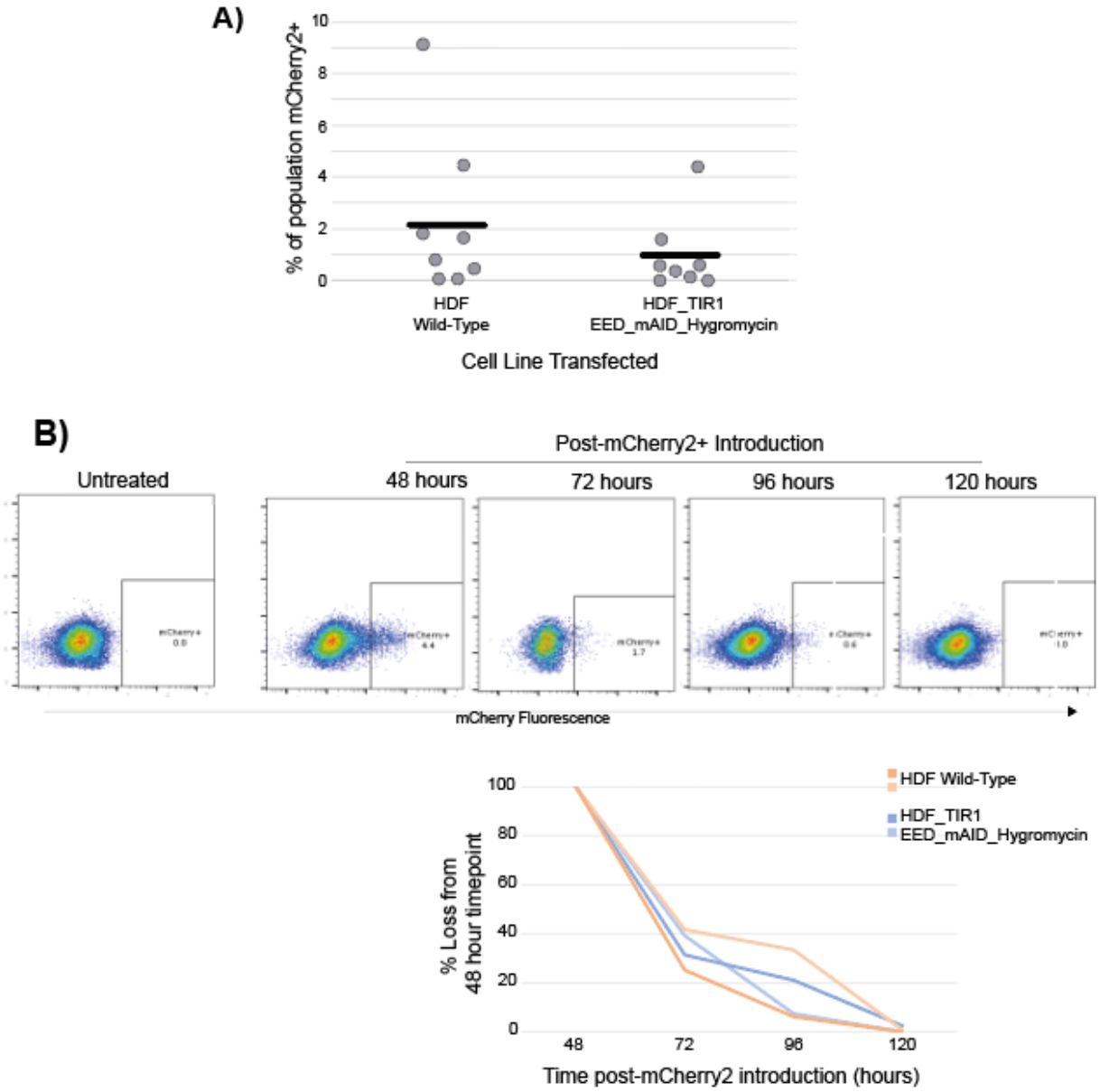


Figure 5.12. Loss of fluorescence expression post-transfection. (A) The percentage of transfected/nucleofected cells that are measured as mCherry2+ by flow cytometry 48 hrs post-introduction. (B) At 48 hrs post-introduction a population of cell was analysed for mCherry2+ expression by flow cytometry and then was split into 3 further populations, which were measured every 24 hrs for mCherry2 expression thereafter. Dot plots show an example of a population undergoing flow analysis every 24 hrs, with a quantification of 4 experiments below. The percentage loss of fluorescence was calculated from the original population calculated at 48 hrs post-introduction.

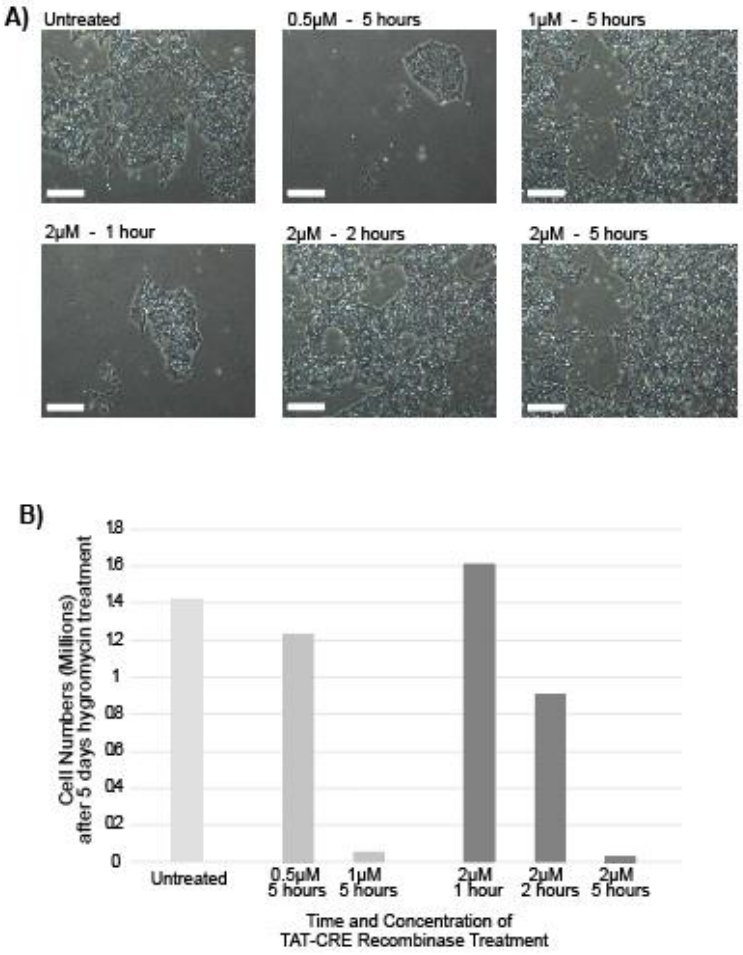


Figure 5.13. Loss of hygromycin resistance after TAT-CRE Recombinase treatment. Cell colonies were treated with TAT-CRE Recombinase for one of the six treatment conditions as indicated, to induce LoxP excision of the hygromycin resistance gene. These six cell cultures colonies were grown for 48 hrs and then hygromycin selection was undertaken using 10 μg/μl for 5 days. (A) Cells were then imaged after 5 days antibiotic selection. Representative phase-contrast images of each TAT-CRE Recombinase treated culture, with 100 μm scale bars (B) Cell counts of each of the six TAT-CRE Recombinase treated cell cultures, n=1 replicate.

To isolate clones that had complete removal of the hygromycin resistance, and could therefore be utilised for retargeting, colonies were picked from the two successfully treated populations and screened for antibiotic excision (Figure 5.14). Three possible alleles could be present in these cell types – a wild-type allele should be present in all cells as the lines used were heterozygous. The other two possibilities arise from the different alleles formed with

successful and unsuccessful excision from the TAT-CRE Recombinase treatment (Figure 5.14a). To test which alleles were present in the clones, three different PCRs were performed, all of which use the same left primer located in the endogenous *EED* gene outside of the homology arm region. The wild-type PCR is to confirm the presence of the wild-type allele as expected. The mAID PCR will target within the mAID sequence, to confirm that the line that is carried forward is successfully tagged, and this reaction should be positive regardless of the TAT-CRE Recombinase success. The final PCR targets within the PGK promoter, which is between the LoxP sites and therefore will only be present if the antibiotic was not excised (Figure 5.14a); an untreated HDF_TIR1 EED_mAID_Hygromycin sample was used as a positive control for this reaction. From the genotyping results, all clones had a wild-type band present (Figure 5.14b). Also, all clones had the expected mAID PCR product, which was present in the parental HDF_TIR1 EED_mAID_Hygromycin but not the HDF wild type. Only 1 out of 10 clones displayed a positive PCR using the PGK promoter primer, producing a PCR product that size matched the PCR product observed in the untreated sample. This suggesting that the TAT-CRE Recombinase was approximately 90% successful at removing the antibiotic resistance gene between the LoxP sites. Surprisingly, a second wild-type band was not seen in the excised clones, as would have been expected (Figure 5.14a). The removal of the antibiotic gene would change the product from the wild-type PCR on the edited allele from 4287 bp to 2030 bp, and therefore a second band was expected above the 1438 bp wild type allele product. No band was detected, but with no positive control available it is difficult to determine if this was due to a technical problem in the PCR or an unexpected occurrence during LoxP excision. Regardless of this unexpected result, two clones were taken forward for retargeting of the second allele, using the previously successful EED-mAID-Hygromycin template targeting. At the same time, the EED-mAID-Blasticidin template was also used in parallel experiments as a successful introduction of one transgene but not the other would clarify if the original targeting issue arose due to the use of the blasticidin.

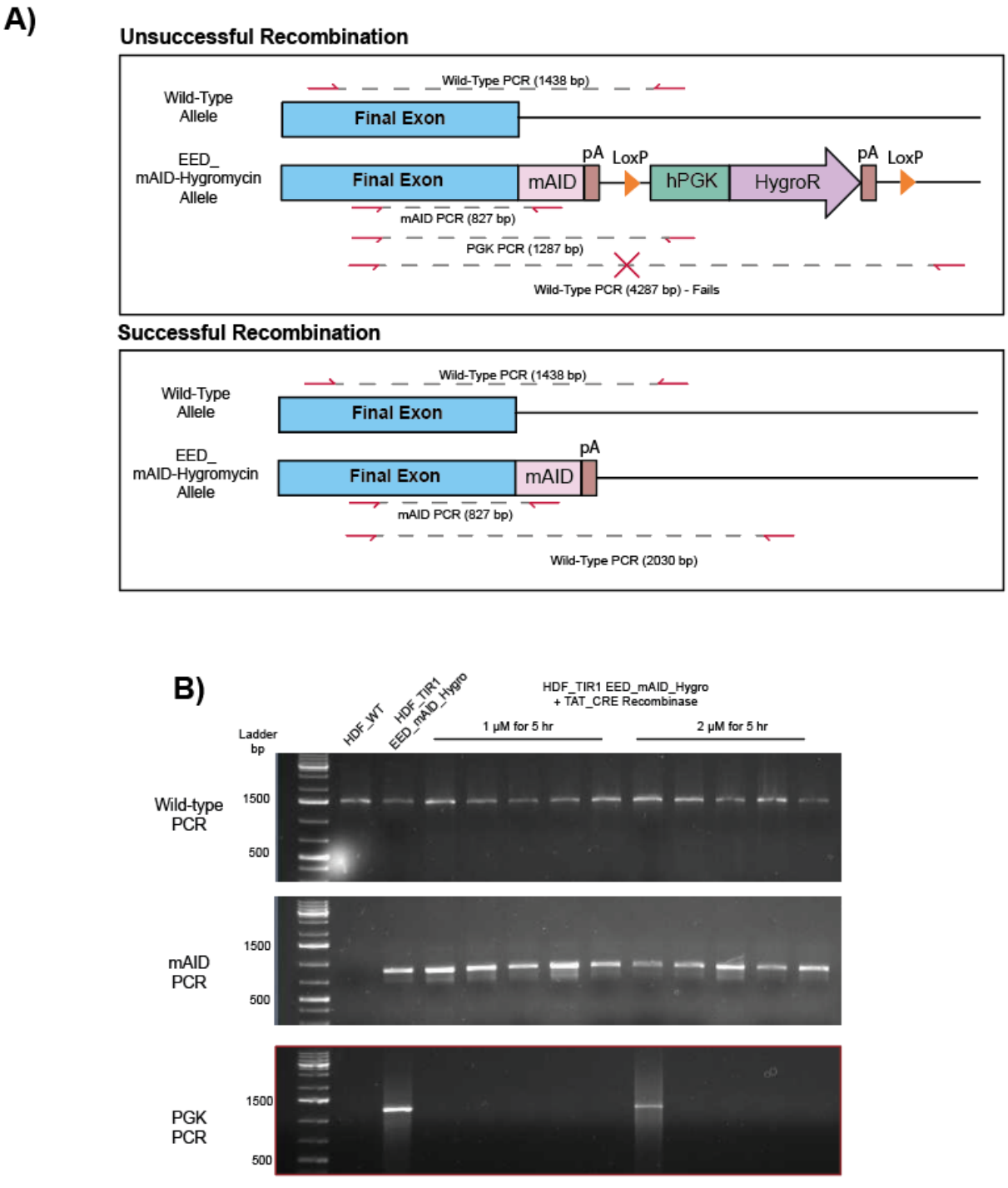


Figure 5.14. Genotyping of TAT-CRE Recombinase treated clones. (A) Schematic demonstrating the different alleles resulting from successful and unsuccessful LoxP recombination following TAT-CRE Recombinase. The three PCRs used to determine the presence of the possible alleles present are shown. (B) Genotyping of individual clones from the two treatment conditions as indicated.

Unfortunately, the re-transfection of the TAT-CRE Recombinase lines with either the EED_mAID_Hygromycin or EED_mAID_Blasticidin plasmids did not produce biallelic tagged EED cells. Previous transfection using the EED_mAID_Hygromycin plasmid had allowed the selection of successfully edited clones. The inability to then isolate resistant colonies suggests there is a problem in targeting the second allele. It is not possible to deduce from the targeting experiment why homozygously tagged colonies could not be isolated but potential limitations of the AID that may contribute to this are discussed in 5.4.1. Further possible strategies to overcome this problem and potential alternative protein degradation systems are also discussed.

5.3.2 Design and production of an inducible CRISPR interference system targeting *EED*

In developing an inducible system to target EED for depletion in hPSCs, two diverse methods for targeting this protein were selected. As an alternative to the AID system, the CRISPRi knockdown system was also selected for introduction into the cells. The CRISPRi system will target earlier than the AID, acting at the gene levels to inhibit transcription, and thereby block protein production. Utilising cells that had previously been edited to express a dCas9-KRAB complex (Mandegar et al, 2016), the only element of the system that was still required was the gRNA to direct the repressive fusion protein to the gene of interest. To select the gRNAs to test for inducing silencing of *EED*, sequences were taken from Gilbert et al, 2014 who developed rules that could predict efficient targeting gRNA sequences for all genes. Of the ten gRNAs suggested for *EED*, four were selected, gRNA_1, gRNA_6, gRNA_7 and gRNA_9, for testing due to their location near to the *EED* transcriptional start site and not overlapping with other shortlisted gRNAs (Figure 5.15).

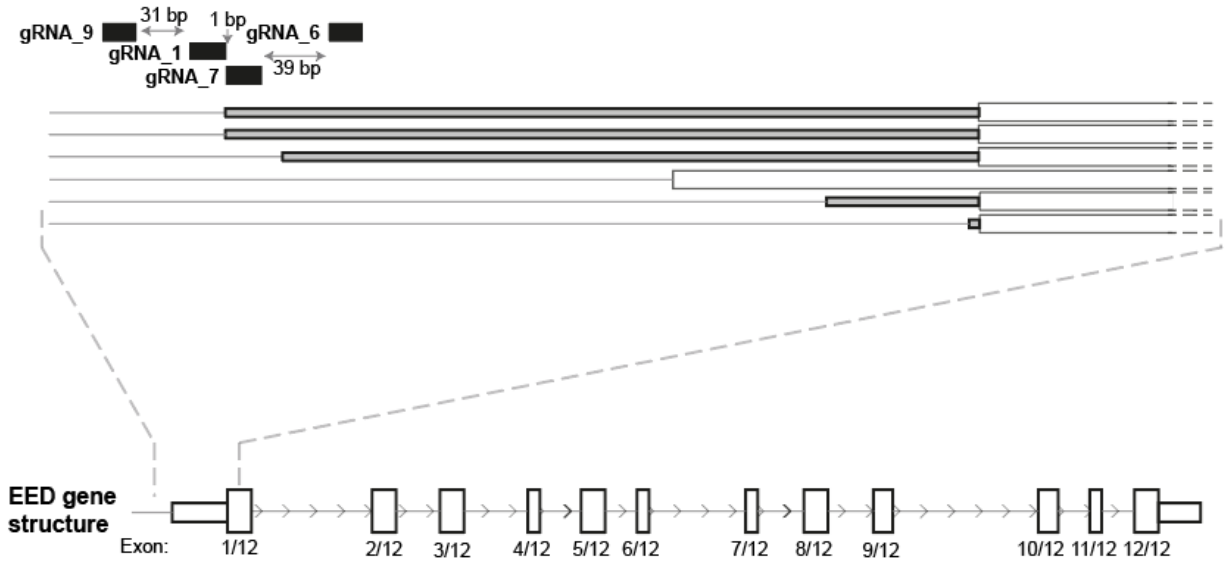


Figure 5.15. Selecting guide RNAs to target *EED* silencing. The location of the selected gRNAs relative to the transcriptional sites of multiple *EED* isoform, from the UCSC Genome Browser on Human Dec. 2013. The 5' untranslated region is indicated in grey and the start of the coding gene in each isoform can be seen, as represented by the white box region. The distance between each gRNA in terms of base pairs (bp) is also shown.

To determine the efficiency of each of the gRNAs designed, optimisation of the dCAs9-KRAB expression was first required. This fusion protein is not expressed constitutively but rather is induced by doxycycline. Upon addition of doxycycline the TREG3 promoter will become active. This induces transcription of mCherry, and the dCas9-KRAB protein, therefore mCherry can act as an indicator for the presence of dCas9-KRAB (see Figure 5.2). The expression of mCherry can therefore be used as a readout of approximate dCAS9-KRAB expression in a cell population (Figure 5.16). To optimise the doxycycline concentration required and to ensure all cells are responsive, doxycycline was added to cells for 48 hrs at five concentrations ranging from 0.1 µM to 3 µM. The addition of doxycycline did not have a negative effect on the cells at any of the concentrations used, and moreover, greater than 99% of the cells demonstrated mCherry expression at doxycycline concentrations of 0.2 µM and above. This level of mCherry

positivity occurred at maximum levels by 0.5 µM and the expression was robust at all timepoints with little variation over multiple replicates (Figure 5.16).

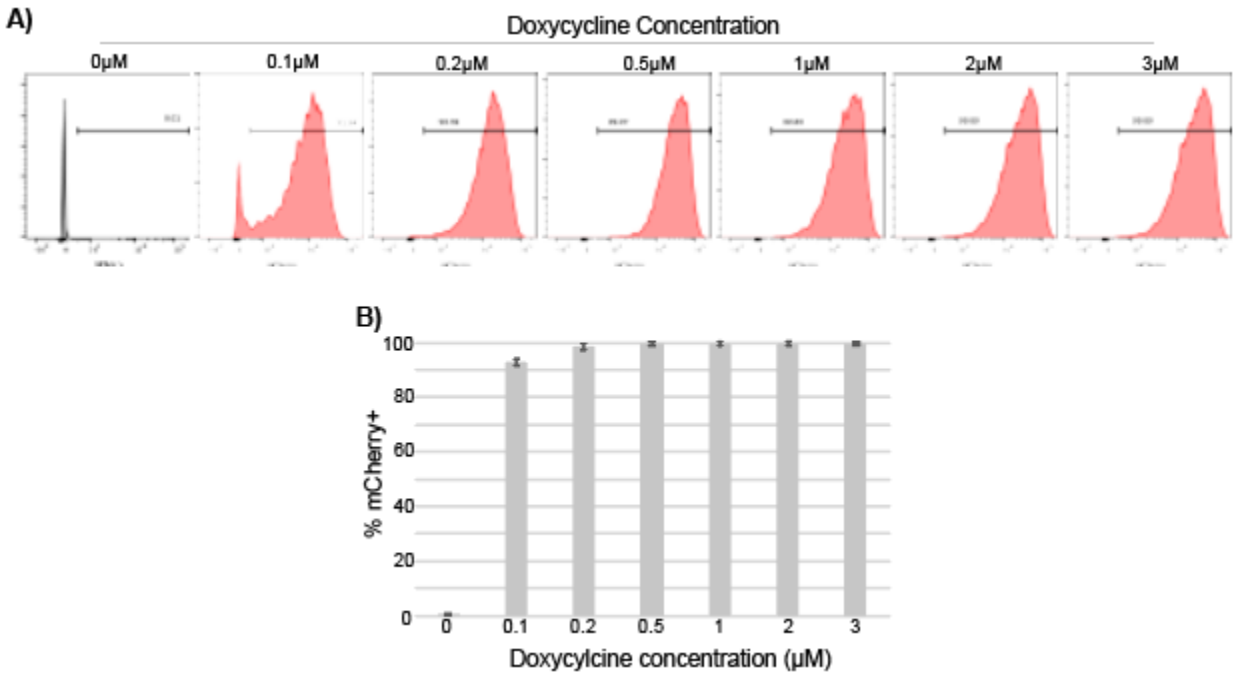


Figure 5.16. Optimising doxycycline treatment. A) Flow cytometry histograms showing mCherry expression induced by doxycycline after 48 hrs of treatment with different concentrations. B) Quantification of the percentage of cell positive for mCherry expression after 48 hrs of treatment with the indicated concentrations of doxycycline, n=3, error bars represent the standard deviation.

To determine which gRNAs led to the best knockdown efficiency in the cells, a number of different transfection methods and selection protocols were tested and the overall silencing of these different conditions on *EED* levels was analysed by RT-qPCR. Firstly, all four gRNAs were nucleofected into the cells individually, and a multi-gRNA nucleofection was also carried out with all four gRNAs introduced into a single sample. These five bulk populations were then treated with blasticidin at 10 µg/ml for a minimum of 10 days and a further 2-3 passages to isolate a resistant population that should stably express the introduced gRNAs. After selection, cells were either maintained as a bulk population, or samples were treated to create single cells, followed by colony outgrowth and individual picking of the colonies to take forward. If successful

integration of the gRNA has occurred in an open chromatin region, the construct will express both the mKate2 fluorescent protein as well as blasticidin resistance. Therefore, after antibiotic selection, all cultures to be tested for *EED* knockdown were analysed under a fluorescent microscope to ensure the expression of mKate2. Once it had been confirmed that cultures are both blasticidin resistant and had mKate2 expression, they were treated with 2 µM of doxycycline for approximately 96 hrs, after which RNA was harvested for *EED* expression analysis by RT-qPCR (Figure 5.17). These results indicated that a number of conditions will cause a large reduction in the levels of *EED* expression after exposure to doxycycline, although the effect is highly variable between the different treatments. The gRNA 7 and gRNA 9 had better silencing efficiency than either gRNA 1 or gRNA 6, particularly in the bulk cell populations. Interestingly, the clonal lines from gRNA 1 and gRNA 6 did have some silencing effects, although this was not to the same levels as the silencing seen in gRNA 7 or gRNA 9 clonal lines (Figure 5.17). The multi-gRNA transfected lines, both bulk and clonal, did have some silencing, but this was less than the silencing effect of gRNA 7 or gRNA 9. The inclusion of gRNA 1 and gRNA 6, which are less efficient gRNAs, within these samples could account for the reduced levels of silencing. Due to variability in the silencing efficiency, and the potential of off-target effects from the gRNAs, three different cell lines were taken forward to study the ability of CRISPR interference to reduce *EED* levels in hPSCs - the samples selected were the bulk gRNA 7 line, the clonal gRNA 9.A line and the clonal gRNA 1.C line. In all three lines, levels of *EED* expression were reduced to less than 50% compared to the non-doxycycline treated samples (Figure 5.17).

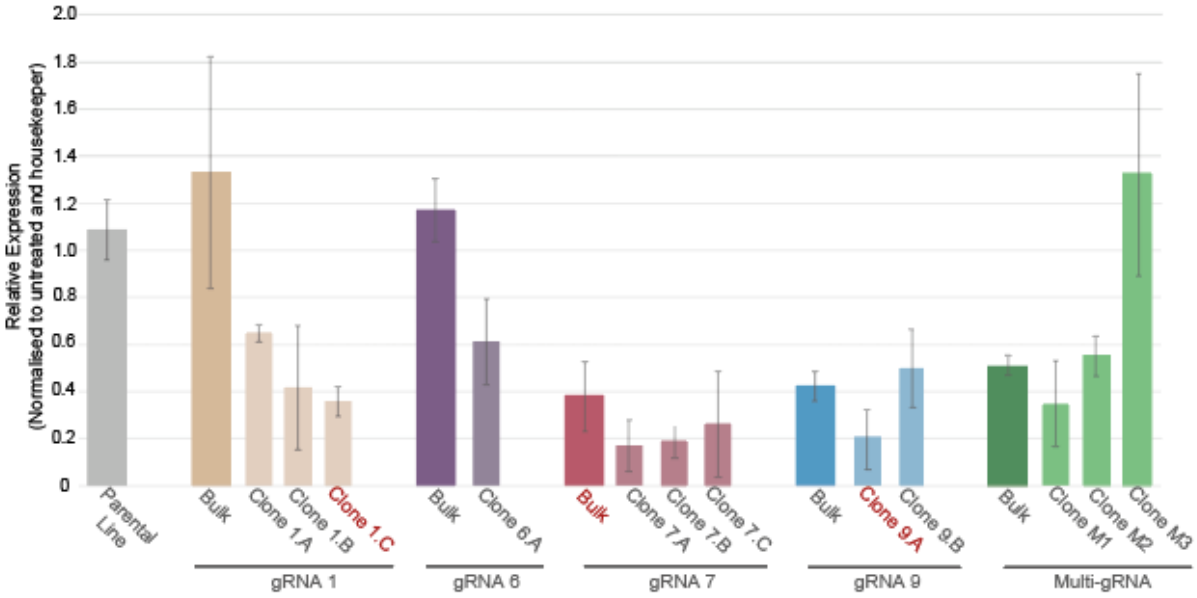


Figure 5.17. The ability of gRNAs in CRISPR interference to decrease EED gene expression. A Cas9_KRAB hPSC line was transfected with four different guide RNAs (gRNA 1, 6, 7 or 9) either individually or in combination (multi-gRNA). After selection for gRNA introduction, cells were maintained as a bulk population or individual clones were isolated. Each population was then treated with 2 µM of doxycycline for 96 hrs and the relative levels of EED compared to untreated cells. Average of n=3 ± S.E.M.

The cell lines gRNA 1.C (gRNA_1), gRNA 7 Bulk (gRNA_7) and gRNA 9.A (gRNA_9) were taken forward and the dynamics of *EED* gene silencing in these lines was analysed to determine how quickly an efficient knockdown was achieved. To determine the dynamics of knockdown, doxycycline was added to cells at 2 µM and kept in the media throughout the experiment. Samples were collected for RT-qPCR expression analysis every 24 hrs, for the first 10 days, and then every 48 hrs until day 20 (Figure 5.18). Using RT-qPCR analysis, the levels of *EED* mRNA in cells were compared to day 0, which had not been exposed to doxycycline. A parental line containing the Cas9-KRAB without gRNAs was also treated with the same concentration of doxycycline to confirm that this treatment does not have any effect on the cells. For all three gRNA lines, *EED* was not decreased within 24 hrs of doxycycline addition, although reductions in *EED* gene expression was observed after 48 hrs of doxycycline addition. In the case of gRNA_1, *EED* levels are reduced to approximately 50% by 48 hrs addition, with levels of *EED* staying similar throughout the remaining points of time course (Figure 5.18). For the gRNA_7 cell line, the levels of *EED* gene expression fell sharply between 24 hrs and 48 hrs of

doxycycline, the levels of *EED* were then maintained at similar levels for the remaining timepoints. The levels of *EED* expression in gRNA_9 cell line was also below 50% and this silencing occurred between 24 hrs and day 4 of doxycycline treatment. The CRISPRi system can therefore cause a strong reduction of *EED* gene expression to less than 50% of the levels normally present, within 48 to 96 hrs of doxycycline addition (Figure 5.18).

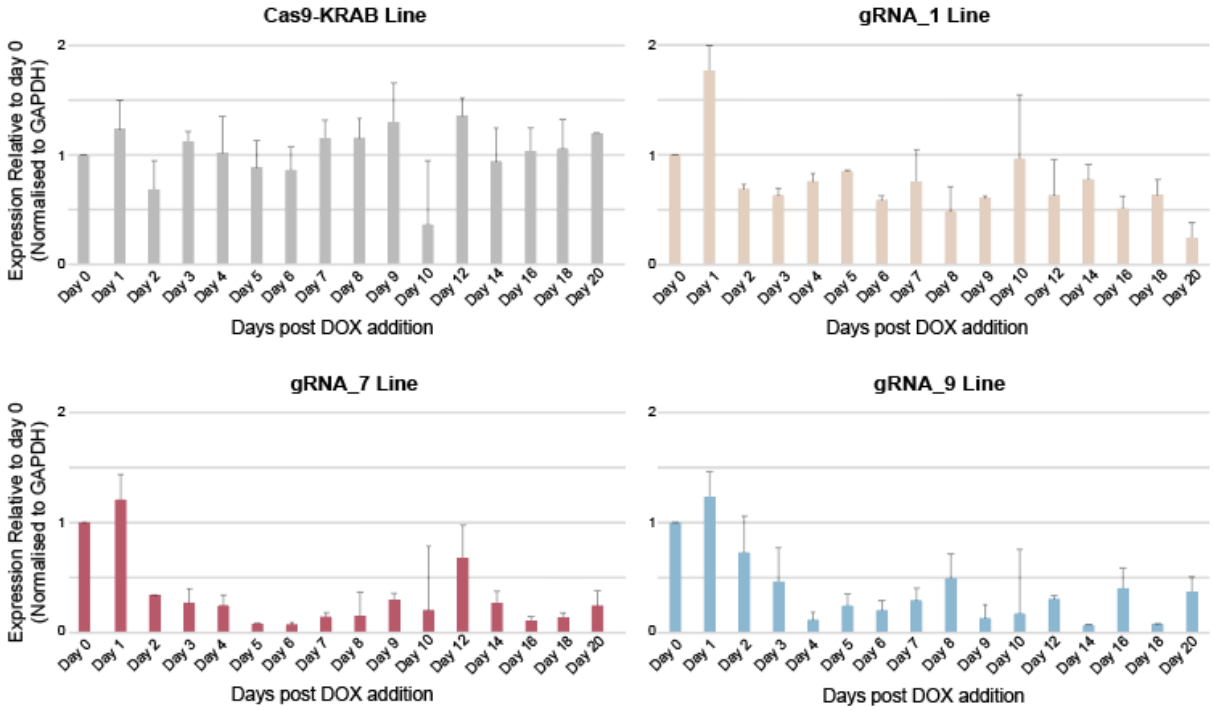


Figure 5.18. Silencing of *EED* gene expression through the CRISPR interference system. A Cas9-KRAB parental line, and three *EED* gRNA expressing lines were analysed for EED expression. On day 0, 2 µM of doxycycline was added to each population and samples were isolated every 24 hrs, for 10 days, followed by every 48 hrs for a further 10 days. Relative *EED* expression was calculated compared to day 0 expression. Average of n=3 ± S.E.M.

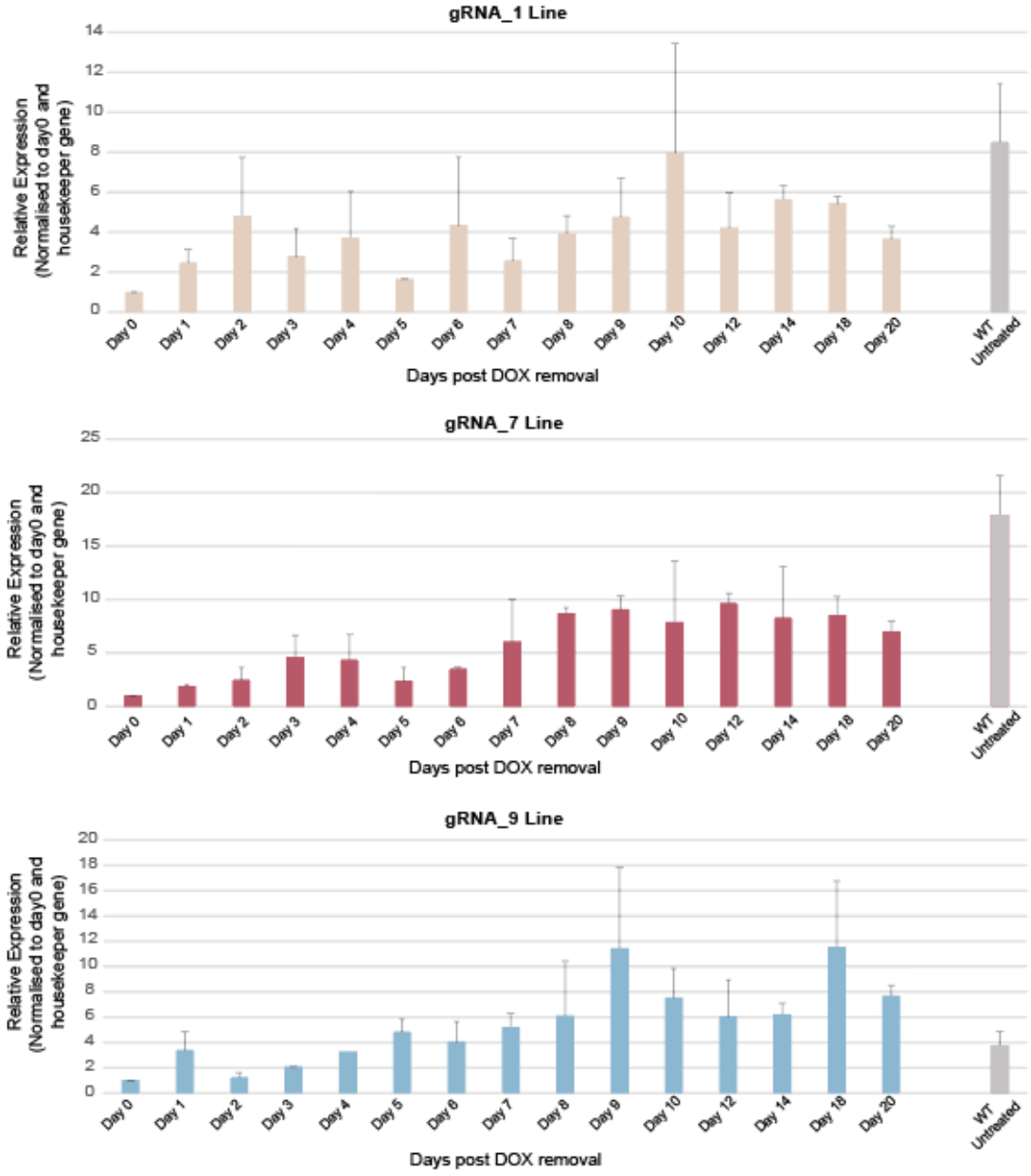


Figure 5.19. Recovery of EED gene expression after removal of Cas9-KRAB silencing. Three Cas9-KRAB EED gRNA expressing lines were analysed for EED expression after doxycycline removal. Cells were treated with 2 µM for a minimum of 20 days before removal. Doxycycline was removed at day 0, and samples were isolated every 24 hrs for 10 days, followed by every 48 hrs for a further 10 days. Relative EED expression was calculated compared to day 0 expression. The relative EED expression of Cas9-KRAB untreated parental line compared to day 0 was analysed. Average of n=2 ± S.E.M.

The CRISPRi system was selected because it offers the potential for a highly inducible, controlled and potentially reversible loss of function approach. After establishing the silencing ability of the system, the next step was to confirm the ability to reverse the *EED* depletion. To examine this, cell lines were treated with doxycycline for a minimum of 20 days, before removal of doxycycline and then analysis of *EED* levels. Expression of *EED* in all time points was compared to the time point of doxycycline removal (day 0). Samples were collected every day for 10 days post-removal and then collected every second day until 20 days doxycycline removal (Figure 5.19). Analysis of all three cell lines demonstrated a variable but clear recovery in *EED* expression, occurring rapidly after doxycycline removal and with an average increase in all lines only 24 hrs post-removal. Although the variability, as would be predicted, was observed in the first few days following post-removal, by day 6-7 post-removal all 3 lines have a minimum 2-fold increase in *EED* levels. In the gRNA_7 and gRNA_9 cell lines, the levels increased approximately 7-fold by 20 days post-removal. The gRNA_1 cell line increased by approximately 4-fold in *EED* expression by day 20, however the *EED* silencing in this line was less than that seen in gRNA_7 and gRNA_9 (Figure 5.19).

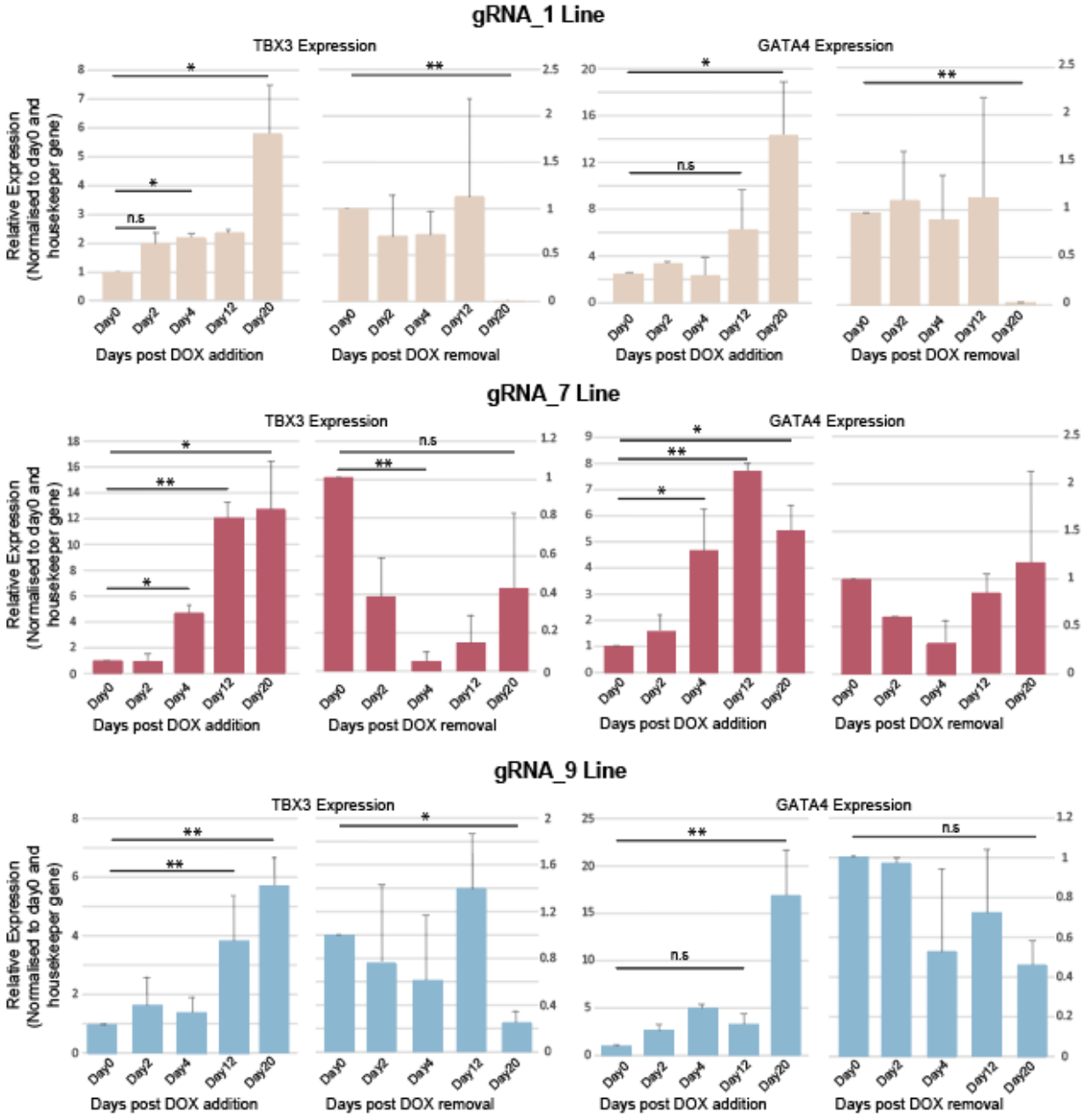


Figure 5.20. The ability of CRISPRi to control Polycomb-target gene expression by EED targeting. Three Cas9-KRAB EED gRNA expressing lines were analysed for TBX3 and GATA4 expression after treatment with 2 µg/ml doxycycline, and after doxycycline removal. Relative gene expression was calculated compared to day 0 of either the addition of doxycycline or the removal of doxycycline, as indicated. Average of n=2 or n=3 ± S.E.M, data was compared using a one-way ANOVA followed by Bonferroni's multiple comparison test ns = $p>0.05$, * $p\leq0.05$, ** $p\leq0.01$.

To determine if the CRISPRi system was able to have a functional and reversible effect within the hPSCs, the expression of genes normally repressed by the PRC2 complex in these cell types was analysed. The *TBX3* and *GATA4* gene promoters in hPSCs are normally marked by H3K27me3, and it was previously demonstrated (Figure 4.2) that the addition of small molecule inhibitors targeting PRC2 complex can result in increased expression of these genes. To determine if the CRISPRi system can also induce a loss of gene repression, the expression of these two genes following doxycycline addition was analysed at several timepoints from 48 hrs to 20 days post treatment. The time points at which gene de-repression was observed is variable between both the gRNA cell line and the gene analysed, but overall by day 20 of doxycycline treatment for all cell lines there was a clear increase in *GATA4* and *TBX3* expression (Figure 5.20). Furthermore, a significant gene de-repression was detected as early as day 4 of doxycycline addition for *TBX3* and *GATA4* in the gRNA_7 cell line. For *GATA4* expression, gRNA_1 and gRNA_9 both resulted in a >10-fold increase at day 20 of doxycycline addition, and gRNA_7 had over a 5-fold addition by day 20 (Figure 5.20). After the removal of doxycycline from the gRNA_1 cell line, almost all TBX3 and *GATA4* expression is lost by day 20, which may indicate the recovery of EED function in the cells. For gRNA_7 and gRNA_9 cell lines, a decrease in expression was observed 20 days post doxycycline removal, although this effect was variable and was not a consistent or significant decrease for *TBX3* and *GATA4* (Figure 5.20). The *EED* levels are known to recover in these lines, therefore the inability to reintroduce efficient gene repression could be caused by changes that might have occurred in the cells, which cannot be rescued by re-expressing the PRC2 complex. Although further experiments should be undertaken to confirm the depletion of EED at a protein level, which was not possible in this study due to problems with antibody efficiency, initial data suggest that the CRISPRi system is able to cause depletion of EED levels in a cell at a functional level, which can be at least partially reversible.

5.4 Discussion

Loss of function studies are important in determining the function of a protein within different contexts. These studies can be targeted at the gene, RNA or protein level, all with the intention of creating a knockout to study the effect this will cause in specific settings. The decision of which system to select is determined by a number of factors and are chosen based on both the target of the system and the question being investigated by the loss of function study. In this study, two different approaches were selected for use in hPSCs, with the aim of targeting the EED component of the PRC2 complex. First, the AID system, which despite considerable efforts was not fully integrated into the cells, and second, the CRISPRi system, which appears to successfully allow the targeted and reversible depletion of EED.

5.4.1 The auxin inducible degron system was not successfully introduced.

Within this study it was not possible to produce hPSCs that contained all of the components required to target the endogenous EED protein using the AID degron system. Cells were generated that stably expressed the OsTIR1 protein, which is necessary for introducing the tagged protein of interest to the ubiquitin system, thereby targeting it for degradation. The next essential stage of this system is to tag both alleles of the protein of interest with an AID degradation peptide, which would allow the protein to be recognised by the OsTIR1 protein in the presence of the Auxin/IAA plant hormone (Natsume et al., 2016). In these experiments, cell lines with only a single tagged EED allele were successfully produced.

The reasons that the second allele of these cell lines could not be successfully edited is unclear from the data collected. The inability to insert an mAID tag to the second EED allele could have been caused by multiple factors, although it was not possible to determine the cause in this study. A potential issue in targeting a second allele that was investigated in the study was the use of the blasticidin antibiotic in allowing selection of resistant lines to indicate successful transfection. When carrying out two editing experiments in parallel with the EED_mAID_Hygro targeting plasmid and the EED_mAID_Bsr plasmid, no blasticidin resistant colonies were isolated but hygromycin resistant and successfully edited lines were formed. This may suggest that the problems arose during molecular cloning in forming the EED_mAID_Bsr plasmid, resulting in plasmid which will integrate a non-functional resistance gene, and therefore could not be used as a successful positive selection marker. The use of TAT_CRE and the unsuccessful retargeting by the previously used EED_mAID_Hygro plasmid, would, however,

suggest another problem with the second allele targeting that is unrelated to the inability to isolate blasticidin resistant clones. The inability to target the second allele may instead be the result of issues within the AID targeting strategy adopted. The insertion of a tag to the site of interest relies on homology arms that allows the insertion of the mAID tag to a specific region, in this case just prior to the *EED* stop codon. As the first round of gene targeting allowed the successful integration of a mAID tag to the *EED* gene this shows that the homology arms are successful in directing this edit to the desired region of the genome. However, after the first successive round of targeting, there is the possibility that the target site may have been changed, which could affect the ability to insert the tag on the second allele. This may be caused by disturbing the gRNA recognition site thereby removing the ability of the Cas9 to cleave the DNA, or it may disrupt the ability of the homology arm template to recognise its complementary region in the genomic DNA and can therefore not introduce the new sequence.

Another reason that could contribute to the inability of the second allele tagging is due to the high levels of basal degradation in the protein of interest that can occur with the AID system. This basal degradation has clearly been demonstrated in other studies, such as that by Li et al, 2019, which targeted the endogenous human seipin gene, a transmembrane protein, for AID mediated degradation. However, it was observed that the addition of the mAID tag to the protein, as used in this study, caused a severe reduction of the protein levels before IAA addition, resulting in cells which exhibited a seipin knockout phenotype (Li et al., 2019). Similarly, a study targeting HJURP, a centromeric histone chaperone protein, demonstrated that tagging with mAID resulted in depletion of >90% of HJURP protein, even in the absence of auxin in human cell lines (Zasadzińska *et al.*, 2018). Moreover, targeting of some proteins in previous studies resulted in an inability to introduce a biallelic mAID tag in OsTIR1 expressing cells (Natsume *et al.*, 2016), similar to the problems seen in this study.

It has been demonstrated previously that hESCs with knockouts of core PRC2 components, including EED, will undergo spontaneous differentiation with a severely reduced self-renewal ability, resulting in difficulties to maintain the PRC2-disrupted hESC lines (Collinson *et al.*, 2016; Shan *et al.*, 2017). If the degron system in this study, targeting the EED protein, had similar levels of basal degradation observed in previous studies, obtaining a homozygous mAID targeted cell line could result in cells that have a severely reduced growth ability. Combined with the low efficacy of homology directed repair, estimated to be well below <10% (Lombardo *et al.*, 2007; Steyer *et al.*, 2018), it may be extremely difficult to isolate homozygous targeted cells due to their possible slow growth and out-competing by other, non-targeted, cells.

5.4.2 Methods and improvements to systems which regulate protein stability in cells.

As the auxin-inducible degron system is a promising molecular tool, but suffers from the problem of high basal degradation, a number of novel adaptations have been created to overcome this issue. The first way to overcome the basal degradation is by limiting the expression of the OsTIR1 protein, for instance by controlling the expression of OsTIR1 with an inducible promoter (Natsume *et al.*, 2016). Substituting an inducible TET promoter in place of the constitutive CAG CMV promoter that was used in this study resulted in the tight control of OsTIR1 expression (Natsume *et al.*, 2016). This is a simple but effective approach that could be used to overcome basal degradation, but still maintains a number of caveats, including any leaky expression from the TET promoter that could again result in reduced levels of the target protein. Also, importantly, the TET promoter reduces some of the dynamic control that is associated with the degron system, as the system is no longer as directly responsive to the addition of the IAA molecule as it would be in the initial designed AID system.

Other techniques to control basal degradation can be done through modifications to the AID system components. The first of these adaptations made use of the auxin responsive transcription factors (ARF), a component that is present in the natural plant auxin degradation system. The PB1 (Phox and Bem1) domain of ARF proteins interact with the Aux/IAA family of proteins in the absence of auxin in plant cells (Sathyan *et al.*, 2019). A study that introduced the ARF-PB1 domain for expression, under the same promoter as the OsTIR1 within the AAVS1 genomic safe harbour, determined that the expression of this protein resulted in two benefits in the system. The first, is a reduced auxin-independent degradation of target proteins, and secondly the ARF expression appears to promote more rapid induced degradation of tagged proteins (Sathyan *et al.*, 2019). Interestingly, the ARF-PB1 and OsTIR1 do not compete for the same binding region, and it is therefore unclear how the ARF-PB1 can cause decreased basal degradation by the OsTIR1 protein, although Sathyan et al, 2019, hypothesise a conformational change brought about by the ARF-PB1 binding may be responsible. Although this is a potentially exciting adaptation in the development of the AID system, as the mechanism through which the ARF-PB1 function is not clear, it is difficult to determine whether the inclusion of this protein within the system will have the same effect with all targeted proteins. Another study that adapted the components of the system in an attempt to overcome the high basal degradation was reported by Li et al, 2019, in which they trialled a number of variations of both the mAID tag

and the SCF interacting OsTIR1 protein. A comparison of the OsTIR1 protein with five other SCF interacting proteins, identified the *Arabidopsis thaliana* AFB2 (AtAFB2) as better candidate for use in the AID system. Protein degradation after 16 hrs was comparable to that seen with the OsTIR1 protein, but the basal degradation was five-fold higher in the OsTIR1 version of the system (Li *et al.*, 2019). However, this decrease in the basal degradation also resulted in inefficient auxin-induced depletion at 1 hour with the AtAFB2. To overcome this, the mAID tag was compared with 14 alternative degron sequences, in combination with either OsTIR1 or AtAFB2. This identified a 67 amino acid degron from the *Arabidopsis thaliana* IAA7 (termed miniIAA7) which, in combination with AtAFB2, improved auxin-inducible depletion over three-fold compared to miniAID at 1 hour of treatment, and with reduced basal degradation (Li *et al.*, 2019). This new improved version of the AID system was efficient for a variety of target proteins in both N and C-terminal tagged orientations and may be one of the best methods to use in studies going forward. However, this system requires all of the components within the current system to be replaced, which will require new targeting vectors and also retargeting of the components the original parental lines, which is highly labour and time intensive when the original degron system component have been partially or fully introduced into cells, such as in this study.

As a method to overcome the problem with pre-existing AID systems that have already been partially or fully introduced to cells, a potent auxin antagonist that binds to OsTIR1 was developed (Hayashi *et al.*, 2012). This molecule was identified through an initial screen, followed by further development, eventually producing a molecule termed auxinole. This molecule binds directly to the TIR1/AFB E3 ligase binding proteins and will block the formation of the TIR1–IAA-mAID complex and will therefore inhibit auxin-induced degradation of target proteins. This was observed in a study by Yesbolatova et al, 2019, which demonstrated reduced basal degradation of DHC1 in the presence of OsTIR1, but without the addition of IAA. The addition of auxinole was also seen to increase the recovery of proteins when added after the removal of an IAA treatment, which allows a tighter control of the dynamics of the AID system. Importantly, addition of the auxinole molecule to cell media did not appear to have any negative effect on the cells (Yesbolatova et al., 2019). The addition of auxinole to cells that are proving difficult in which to generate biallelic tagging of mAID can potentially help by restricting the degradation of the proteins, thereby allowing the maintenance and outgrowth of the targeted cell lines that are normally negatively affected by the basal degradation.

Although the AID system offers a very useful tool for eliminating proteins of interest from a cell, it is not the only protein degradation method available for use in hPSCs. Similar to the AID technique, a number of protein degradation techniques have been developed that utilise the endogenous ubiquitin-proteasomal degradation system to deplete target protein levels from a cell, although through differing targeting methods. One of these methods, which has been adapted to both basic research and is offering potential for clinical applications are PROTAC (PROteolysis Targeting Chimeric) molecules. PROTACs are heterobifunctional molecules that comprise an E3 ligand linked to a substrate protein ligand, which promotes the target substrate for proteasomal degradation (Verma et al., 2020). One of the first examples of this molecular tool was developed to target the methionine aminopeptidase-2 (MetAP-2) enzyme (Sakamoto *et al.*, 2001; Verma, Mohl and Deshaies, 2020). To target this protein for degradation each aspect of the bifunctional molecule had to be designed to specifically link the MetAP-2 enzyme to the proteasomal degradation pathway. Firstly, this required a known binder of the MetAP-2 enzyme, an example of which is the anti-angiogenic biomolecule Ovalicin, which inhibits the MetAP-2 by binding directly to the enzyme molecule. Secondly, a link into the ubiquitin-proteasome system was required. This identified the IκBα protein, a known cellular target of the Skp1-Cullin-F box E3 ligase, through the β-TRCP F box protein. The recruitment of IκBα is mediated by a 10-aa peptide within the protein. Therefore, to link the MetAP-2 to the SCF-β-TRCP and target it for proteasomal degradation, a bifunctional molecule was produced consisting of an Ovalicin molecule, attached to the IκBα peptide sequence through a small linker. The addition of this molecule, termed Protac-1, was then able to successfully target the MetAP-2 for degradation (Sakamoto *et al.*, 2001). Since this initial molecule was developed a large number of PROTACs have been developed, targeting different classes of proteins for successful degradation, including epigenetic targets, protein kinases and nuclear receptors through various ligase linking sequences (Bondeson *et al.*, 2015; Winter *et al.*, 2015; Gadd *et al.*, 2017; Salami *et al.*, 2018). This includes the recent, successful targeting to the EED proteins, which brought about the loss of PRC2 and reduction of H3K27me3 in the targeted cells (Hsu *et al.*, 2020; Potjewyd *et al.*, 2020) It estimated that PROTACs could be used to target almost any protein within the cell, provided that a ligand can be obtained for the target protein. Although, these molecules do offer great prospects, particularly within the clinical setting, the development of these bifunctional molecules can be difficult to produce in the average research laboratory setting, due to the specialist knowledge and skills required.

A technique that utilises a similar approach to induce the rapid degradation of endogenous proteins is the Trim-away technique. This technique also links a target protein to the proteasomal degradation but in place of a bifunctional molecule, the Fc receptor TRIM21 is used to target the protein of interest for degradation without prior modification required (Clift et al, 2012). To induce specific degradation, the protein of interest must be targeted with an efficient and specific antibody, and therefore this technique can be used for almost any protein that has an adequate antibody available. After the addition of the antibody to the cells, which will target the protein of interest, the introduction of either endogenous or exogenously expressed TRIM21 protein will target the antibody-antigen complex for degradation via the ubiquitin-proteasomal pathway. This degradation is reliant on the presence of both the specific antibody and TRIM21 overexpression, as the introduction of a non-specific control IgG antibody does not cause specific protein degradation, similarly protein destruction was not observed in cells which were not overexpressing the TRIM21 protein (Clift *et al.*, 2017). This system was effective in a variety of cell types, targeting a range of proteins and, importantly, it does not require extensive and time-consuming genome modification. Instead the Trim-away can be undertaken using off the shelf reagents for TRIM21 and the protein targeting antibody, although this is also the biggest limitation in this technique due to the requirement of an efficient antibody that recognises the target. If such an antibody is not readily available, or an antigen is inaccessible, through masking within the native protein or localized in a particular cellular compartment, such as the nucleus, it cannot be targeted for degradation (Clift et al., 2017; Clift et al., 2018). Regardless of these limitations the trim-away system is still a highly useful system for controlling the levels of a protein of interest within cells, and has successfully been used to study the function of proteins in zebrafish, human cell types and mouse zygotes (Clift *et al.*, 2017; Chen *et al.*, 2019; Zhou *et al.*, 2019). A variation of the Trim-away technique that removes the need for efficient and specific antibodies has developed an alternative bifunctional fusion protein. This technique, termed Predator, utilised a moiety recognition to target the protein for destruction, similar to the molecules used in PROTACs, and used the TRIM21 protein to induce ubiquitination and proteasomal degradation of the protein (Liu *et al.*, 2020). The Predator technique, with high levels of overlap between the Trim-away and PROTAC methods, have similar limitations, with difficulty in determining highly specific moieties and also developing these molecules in a non-specialised laboratory setting.

All of the above-mentioned systems will result in the inducible degradation of target proteins by linking the protein to an E3 ubiquitin ligase, thereby introducing it into the

endogenous proteasomal degradation. These techniques all induce degradation by the addition of an exogenous factor be it a small molecule, such as IAA, an engineered bifunctional molecule or an antibody, resulting in the depletion of the specific protein. In contrast, techniques exist that will protect a targeted protein from degradation, through the addition of small molecules, and only upon withdrawal of these protective molecules will the target be depleted. Two variations of this similar approach exist: the FKBP-Shield1 system and the DHFR-TMP system (Banaszynski et al., 2006; Iwamoto et al., 2010; Nabet et al., 2018). The FKBP-Shield1 system, with optimised systems commonly referred to as dTAG, was founded through the ability of a 107 residue protein – FK506- and rapamycin-binding protein (FKBP12) – when fused to the target protein to act as a destabilising domain. This destabilising domain targets the fused protein to E3 ligase and proteasomal degradation, therefore depleting levels of the target protein. However, to stabilise the protein, a small molecule ligand termed Shield1 binds to the domain and inhibits the E3 ligase interaction, thereby stabilising the protein levels. The addition and removal of Shield1 will result in a highly responsive and reversible system (Banaszynski *et al.*, 2006; Nabet *et al.*, 2018). A similar variation of this technique, also reliant on a destabilising domain and a stabilising ligand, is the DHFR-TMP system, which fuses the 159 residue E. coli dihydrofolate reductase (eDHFR) to the target protein. This causes depletion of proteins when not in the presence of the stabilising ligand, trimethoprim (TMP). The resulting cell lines had low basal degradation in the presence of TMP, a robust and predictable dose-response behaviour, and rapid degradation upon ligand removal (Iwamoto *et al.*, 2010). Both of these destabilising techniques have been employed within mammalian cells to study a variety of protein types (Banaszynski., 2008; Trentini et al., 2020; van den Berg et al., 2018). When fusion proteins are created, however, these are relatively large tags, and may not be suitable to use with some proteins due to a negative effect on the protein function and will also require genome editing to target endogenous protein. However, unlike the PROTACs and Trim-Away techniques that require the production of complex molecules or the availability of high-quality antibodies, any laboratory with the ability to undertake genomic editing can produce these destabilising protein degradation systems for any target of interest.

Therefore, although it was not possible to produce an AID system targeting EED in hPSCs, the trial of new adaptations to the AID system may still allow the development and application of this system. These systems have previously been utilised in PSCs to uncover the functional action of proteins (including the use of AID system in hPSCs and differentiation (Weintraub *et al.*, 2017; Sybirna *et al.*, 2020), demonstrating the efficiency of the system if

successfully integrated into the cells. It is therefore worthwhile to attempt alternative methods of introducing the system to hPSCs. However, if these problems persist with this approach, then a multitude of other protein degradation systems could be trialled in place of the AID system.

5.4.3 CRISPR interference to control reversible and dynamic protein levels.

Although an inducible loss of protein function system was not successfully introduced to target EED, depletion of this factor through suppressing *EED* transcription was achieved. The CRISPRi method was recently developed to create knockdowns at the gene expression level, through adaptions to the Cas9 / gRNA system. The CRISPRi system acts earlier than the protein degradation system and will act to inhibit the transcription and thereby stopping the initial formation of the targeted protein. The system can be introduced to cells to act constitutively but due to the problems encountered with PRC2 loss of function hPSCs (Collinson *et al.*, 2016; Shan *et al.*, 2017) the system can also be adapted to act in an inducible manner.

Within the CRISPRi system, a modified Cas9 protein is used, the first of these modifications is the creation of a catalytic dead protein (dCas9). As the protein would normally cleave the DNA at a PAM site (Mali et al, 2012; Cho et al, 2013; Jinek et al, 2013), this must be prevented in the case of the CRISPRi system where the cleavage of the genome is not desired. This catalytic mutant can be achieved by modification of the RuvC1 and HNH nuclease domains of the protein, resulting in a Cas9 that can still bind to the gRNA and at the site of recognition but will not cut the DNA (Qi et al, 2013). The second modification of the protein, which further increases the efficiency of silencing, is the fusion of a KRAB (Krüppel associated box) domain of Kox1 to the Cas9 protein (Gilbert et al, 2013). The KRAB domain is a transcriptional repressive domain typically consisting of approximately 75 amino acid residues, with a minimal repression module of approximately 45 amino acid residues (Margolin et al, 1994), which are present in a large number of zinc finger protein-based transcription factors (Huntley et al, 2006). Fusing a KRAB domain to dCas9 allows the targeting of this protein to specific genomic sites and when guided to a TSS can inhibit the expression of the underlying gene, beyond the levels of repression achieved with dCas9 alone (Gilbert et al, 2013). The exact method for how this silencing occurs through the Cas9-gRNA complex is still unclear. It is known that the dCas9 has some silencing effect before being fused with repressive domains, which is hypothesised to function through steric hindrance by blocking the transcriptional machinery to target and initiate transcription form the TSS (Qi et al, 2013; Gilbert et al, 2013). Introduction of the Cas9-KRAB is then shown to further increase the silencing of target genes, this is observed in a study by

Gilbert et al, 2013, in which a targeted GFP gene was decreased by 2-fold with dCas9 alone, and 5-fold with dCas9-KRAB. The KRAB domains are hypothesised to induce silencing by heterochromatin spreading (Groner *et al.*, 2010), and it was shown that the dCas9-KRAB fusion complex resulted in both increased H3K9me3 levels and reduced chromatin accessibility at targeted gene promoters (Thakore *et al.*, 2015; Adli, 2018).

The silencing machinery can only have an effect if it is properly directed to the region of interest, therefore the design and selection of effective gRNAs to induce gene repression can be a major limitation in this technique. To target the gene for silencing it is recommended that a site close to the TSS is selected, which would likely induce the strong levels of knockdown, but these gRNAs must also have a very low probability of off-target effects. In this study, four different gRNA sequences were selected, all targeting very close regions of the genome sequence. However, as can be seen (Figure 5.15), EED has a number of isoforms that are initiated from different TSSs. This further compound the problem of determining an efficient and specific silencing gRNAs, as the gRNAs that are efficient at one site, may not be close enough to direct the silencing machinery to other TSSs of the same gene. To overcome this, gRNAs were selected that targeted to the longest isoform in the expectation that this may also silence the isoforms of shorter length. These shorter isoforms however may escape silencing and could be responsible for the partial knockdown observed in almost all of the gRNA induced clones selected.

Another problem brought about by the targeting gRNA is the inability to control the expression of the gRNAs within the cell. In this study, all of the gRNAs were introduced into the Cas9-KRAB cells through a randomly integrating vector. As the regions that these were targeted into were not pre-selected, this could lead to a number of downstream issues. The first of these is lack of gRNA expression in the cells, this was overcome through the use of both an antibiotic selection marker and a mKate2 fluorescence marker (Mandegar et al, 2016). After the initial introduction of the gRNA to the cells, those that have both successful integration and stability from this site can be selected by the treatment of blasticidin. The success of this selection can then be confirmed by ensuring that all of the selected cells also have mKate2 expression. However, this resulting population will now contain cells which are likely to have a vast range of integration sites. This could result in a population in which cells express varying levels of gRNA, and therefore may have different silencing abilities. In some cases, this gRNA variation appears to have a negative effect with the bulk population unable to silence the gene to the same levels as a single clonal population. By taking forward a single clonal population, after antibiotic

selection, all of the resulting cells should have the same gRNA integration site and should therefore all express very similar levels of gRNA. Therefore, if the gRNA is efficient at repressing gene expression, this should occur at similar levels in all cells of the population after Cas9-KRAB expression is induced. Although this option is beneficial for resulting in similar levels through a population, this may lead to problems when utilising the cells for differentiation studies, as the region in which the gRNA is contained may become silenced, thereby inhibiting the expression of the gRNA and removing the ability of the CRISPRi system to silence the target gene. To overcome this potential issue a number of different clonal lines will have to be selected and used to carry out silencing studies. However, in conjugation with control cell lines, and varying treatment types, this can quickly increase the number of experimental conditions required, which may become unmanageable, particularly when studying multiple targets.

The repression of gene transcription alone will not create an instant knockdown phenotype as it is the dynamics of the protein that determines the ability to deplete the levels. The CRISPRi system relies on the endogenous proteasomal machinery to turnover and degrade the existing EED protein after inducing Cas9-KRAB expression. Unlike in the protein degradation methods, where the induction of the system should induce immediate (<1 hour) and near complete protein loss (Natsume *et al.*, 2016; Nora *et al.*, 2017). The gene repression methods will always be slower, as it first inhibits the transcription, but any mRNA still present can be converted to protein, and similarly any protein already present can continue to function until it is depleted by endogenous proteasomal degradation. Within the CRISPRi system it was observed that efficient gene silencing was not achieved until after 24/48 hrs of initiation. Due to the inability to measure protein levels with an efficient antibody it is not possible to determine the rate of protein depletion in relation to this gene silencing. However, with CRISPRi the rate of protein removal is dependent wholly on the natural turnover of the protein in the cell, if this is normally slow as the levels of degradation for this protein are low, meaning the EED may be present for multiple days after addition of DOX. For these reasons, it is thought that the CRISPRi system may not be as efficient or dynamic as inducible-protein degradation, however as inducible protein degradation was not implemented in the cells, a direct comparison cannot be undertaken between the systems.

It was attempted to produce systems within hPSCs that could deplete EED levels in a highly specific, rapid and reversible manner. The first of these systems, utilised the AID technique, which acts at a protein level to deplete the protein of interest. Due to the inability of implementing this system into hPSCs, likely caused by weakness in targeting proteins essential

in cell survival and proliferation, this system could not be trialled for EED depletion. However, a number of variations in the system would likely allow for these limitations to be overcome, or with a plethora of similar but distinct protein targeting techniques now available these could instead be adapted for the EED target in hPSCs. The second of these systems, the CRISPRi system, which will silence a gene at the transcriptional level did appear to be successful when introduced into the hPSCs. Although the silencing did not appear to be completely efficient, with some transcript remaining, it was efficient in reducing gene repression of PRC2 target genes, and therefore can likely be used to study the effect of EED / PRC2 in PSCs and in differentiation, that is currently not achievable with a full EED knockout line.

Chapter 6
Conclusions and Future Directions

6.1 Conclusions

In the years since human iPSCs were first formed, a number of interesting concepts have arisen in terms of the role that epigenetics play in these cells and the consequences for the applications of the cells, particularly in regard to epigenetic memory and epigenetic age. The term epigenetic memory is attributed to the observation that depending on the cell type used for reprogramming, a propensity towards certain lineage differentiation is observed in the resulting iPSCs. This phenomenon is seen throughout various cell and tissue types, including pancreatic β-cells (Bar-Nur et al., 2011) and is believed to be caused by the endurance of epigenetic marks from the originator cell (Kim *et al.*, 2010; Lister *et al.*, 2011). The second of these concepts is epigenetic age, in which whole organisms and individual tissues can be estimated close to biological age, based solely on DNA methylation data, analysed through model-based algorithms (Bollati *et al.*, 2009; Christensen *et al.*, 2009; Horvath, 2013). Analysis of the epigenetic age in the iPSCs demonstrated a resetting of the original somatic cells age to an age equivalent to 0 years similar to that observed in ESCs (Horvath, 2013). A very interesting study demonstrated that upon reprogramming of mesenchymal stromal cells to iPSCs, and subsequent redifferentiation to a mesenchymal stromal cell type, cells retained a rejuvenated DNA methylation profile, and were only partially functional in their immunomodulatory function (Frobel *et al.*, 2014). These concepts therefore demonstrate the potential importance of the iPSC epigenetics in forming mature and functional differentiated cell types.

This raises the question: does the presence of epigenetic memory/rejuvenated epigenetic age prevent the production of terminally differentiated cells that are representative of adult cells obtained *ex vivo*? This present study therefore attempted to investigate the hypothesis that modulation of PRC2 during *in vitro* pancreatic differentiation could improve the production of a mature endocrine cell population. To investigate this hypothesis the transcriptional and histone modifications during *in vitro* differentiation was investigated. This analysis identified genes that are differentially expressed between the *in vitro* derived endocrine cells and those isolated from

adult islets, and also indicated the persistence of H3K27me3 throughout differentiation. Comparison to an adult islet population demonstrated the existence of large numbers of H3K27me3-modified genes in the in vitro-derived cells, and these regions were enriched for differentially expressed genes. This findings extends an earlier study which identified that *in vitro* derived endocrine cells did not fully remove Polycomb modifications from endocrine-specific genes, which was believed to contribute to the partial functionality of the cells (Xie *et al.*, 2014). Previous research had also demonstrated that inhibition of the EZH1/2 during specific timepoints in differentiation could increase *NGN3* levels and increase the percentage of C-peptide positive cells in the end cell population (Xu *et al.*, 2014). To further investigate the prediction that PRC2 inhibition may be beneficial during *in vitro* differentiation, this strategy was applied in this current study at multiple points during *in vitro* differentiation and the resulting cell populations were studied in depth. Due to limitations with both glucose stimulated insulin secretion and flow cytometry analysis, the functionality of the endocrine cell before and after inhibition were difficult to measure, but hormone transcription and basal C-peptide release do not indicate increased functionality. At a transcriptional level, inhibition at an earlier stage of differentiation appeared to a have a negative effect on pancreatic endocrine gene expression, but later-stage inhibition appeared to have a positive effect, specifically on genes that are differentially expressed between *in vitro*-derived cells and adult islet cells. Future investigation to examine a prolonged protocol, optimisation of functional assays and potentially targeting of other epigenetic modifiers, could further determine if PRC2 perturbation can have a positive effect on endocrine pancreatic cells.

When analysing the global difference in transcription between cells after inhibition, small or specific changes in transcription could be masked by the larger amount of data in which no changes occur in the majority of data. However, one of the best examples in the study showing the effect of H3K27me3 in limiting pancreatic β-cell differentiation *in vitro*, is the key transcription factor *MAFA*. This gene is upregulated specifically in β-cells, and knockout of *MAFA* has severe detrimental effects on the function of pancreatic islets in glucose response (Zhang *et al.*, 2005; Wang *et al.*, 2007; Dai *et al.*, 2012). This transcription factor, in combination with other factors, has the ability to push cells towards the β-cell fate (Zhou *et al.*, 2008; Cim *et al.*, 2012). In the *in vitro*-derived cells after the standard differentiation protocol, *MAFA* is not expressed at day 27 of differentiation or in the foetal samples analysed, but as expected this key pancreatic gene is expressed in the isolated adult islet cells. Analysis of the histone modification in the day 27 population demonstrated the presence of both H3K4me3 and H3K27me3 at the *MAFA*

promoter, likely keeping the level of gene expression dampened in these cells. In the adult islets, H3K4me3 is present, however the repressive mark H3K27me3 is not detected at this site, therefore the difference in histone modifications could at least partially explain the difference in gene expression between the analysed cell populations. Interestingly, in the day 27 *in vitro*-derived cells after late-stage PRC2 inhibition, *MAFA* is seen to increase from originally not expressed in DMSO matched control to expressed in the UNC1999-treated cells. The PRC2 inhibition at a late-stage of differentiation was seen to reduce H3K27me3 levels at the *MAFA* promoter, with the levels of H3K4me3 being maintained. Although it is difficult to determine correlation and causation between the transcription and histone modifications, focusing on *MAFA* demonstrates the potential positive effect that PRC2 inhibition can have on improving pancreatic cell differentiation.

6.2 Key Points

The individual discussion sections in Chapters 3, 4 and 5 explored the results from this thesis in terms of how the new findings fit within and add to the current knowledge in the field. The overall broader conclusions are discussed above and, from this summary, the main advances uncovered in this project are as follows:

- Differences exist in the histone modification distribution between *in vitro*-derived pancreatic cells and adult pancreatic cells, particularly in the persistence of promoter H3K27me3 within the *in vitro*-derived endocrine cells. This could partially explain of limitations that exist within current *in vitro* derived pancreatic cells.

- Genes that possess aberrant histone modifications within *in vitro*-derived pancreatic cells when compared with adult pancreatic cells are enriched for differentially expressed genes, between the two cell types. The presence of repressive modifications suggests the current differentiation protocol does not sufficiently remove H3K27me3 at the necessary locations during differentiation to form fully functional pancreatic islet cells.

- Treating cells with PRC2 inhibitors at around the time of endocrine progenitor formation can decrease the formation of a pancreatic transcriptome. Demonstrating the removal of

H3K27me3 and continued differentiation was not sufficient to select for pancreatic transcriptome upregulation but was highly dependent on initial cell context.

- There was a small increase in the pancreatic transcriptional signature in terms of producing a mature pancreatic cell type when PRC2 inhibitors are added to cells at a late stage of *in vitro* pancreatic differentiation. This suggests that the removal of H3K27me3 at the endocrine progenitor cells could upregulate the pancreatic transcriptome, thereby potentially improving the functionality of *in vitro* derived cells.

- The loss of H3K27me3 at promoters in differentiating cells following PRC2 perturbation caused a large increase in H3K4me3 at these previously modified regions. This did not correlate with a specific increase in the transcription of those affected genes, suggesting interplay between the methylation modifications which were further reinforced by other histone modifications.

- Targeting of *EED* with the auxin-inducible degron system was not possible using the approach initially used, but CRISPR interference could cause efficient and reversible decrease in *EED* expression in human pluripotent cells. The CRISPRi system could therefore be used to study the effect of epigenetic modifiers on the differentiating pancreatic *in vitro* cells.

6.2 Future Directions

6.2.1 Single Cell Applications.

The continuing expansion of single-cell analysis techniques may further extend our knowledge of pancreatic cells, with the transcriptional sequencing of the human islets at a single cell level previously reported (Baron *et al.*, 2016; Segerstolpe *et al.*, 2016; Vanheer *et al.*, 2020). These types of analysis can expand our knowledge of each of the individual cell types that form the islets, allowing an insight into the mature transcriptional signature associated with each endocrine cell type, which is currently difficult due to the mixed population and difficulty in isolating all of the specific cell types. The ability to sequence single cells/very low cell numbers could also allow further analysis of the developing pancreas, which is currently challenging to

undertake due to the very limited material available for these analyses. Uncovering more about the developing human pancreas could potentially be used to further optimise the differentiation protocols, which have previously relied on information gained from rodent development to extrapolate to what occurs in human development. Application of these techniques to the *in vitro*-derived pancreatic cells can also provide information about the cell which are present in the day 27 population, and therefore highlight area in which these cells continue to remain transcriptionally distinct from mature pancreatic islet cells, and therefore potentially how to improve them.

Combining the single cell transcriptional analysis cell types with the recently developed single cell compatible ChIP-like analysis (Hainer *et al.*, 2019; Kaya-Okur *et al.*, 2019) would add a wealth of information to this field. Similar to the benefits in terms of transcriptional analysis, utilising these techniques in foetal and early developing pancreas could overcome the issue of limited tissue availability, and could determine more of the epigenome within these cell types. The ability to probe a single cell will also address the issue of heterogeneity in this cell analysis, this assay may be useful for determining the presence of possible bivalency in a mixed population such as pancreatic islets. It may also uncover if the differences in histone modifications exist between the *in vitro* and *in vivo* cells, as indicated by this and previous studies (Xie et al, 2014), are truly caused by differences within similar cell type as opposed to comparing the averages of mixed populations, thereby uncover new techniques for improving pancreatic *in vitro* differentiation.

6.2.2 Inhibition Combined with Culture Optimisation.

An exciting development in the culture of human cells, in both pancreatic and other cell types, is the ability to form more efficient cells through 3D suspension culture *in vitro* differentiation systems. It is known that the pancreas will form in a highly organised structure in the body, and a number of studies have utilised this system to promote the development of the pancreatic cell types (Rezania et al, 2012; Russ et al, 2015). The data from Chapter 4 suggests that the addition of PRC2 inhibitors induces the upregulation of mature endocrine markers. Therefore, combining these approaches may have benefits beyond those seen in the culture system used in this study, which occurs solely in a 2D culture system. If the addition of PRC2 inhibitors facilitates the upregulation of numerous genes, potentially both specific and non-specific to the pancreatic cell type, then there is the potential that the cells within the culture can direct the cell differentiation by cell-to-cell communication orchestrated through the structure of

the organoid. In this circumstance, the small molecule may act in a way to allow the maturation of the cell by removing this barrier of the inhibitory H3K27me3 modifications. This could also be combined with elongating the protocol to see what effect the release of pancreatic factors, as observed following the inhibition of PRC2 at a late stage of differentiation, would have on maturing the cells in culture.

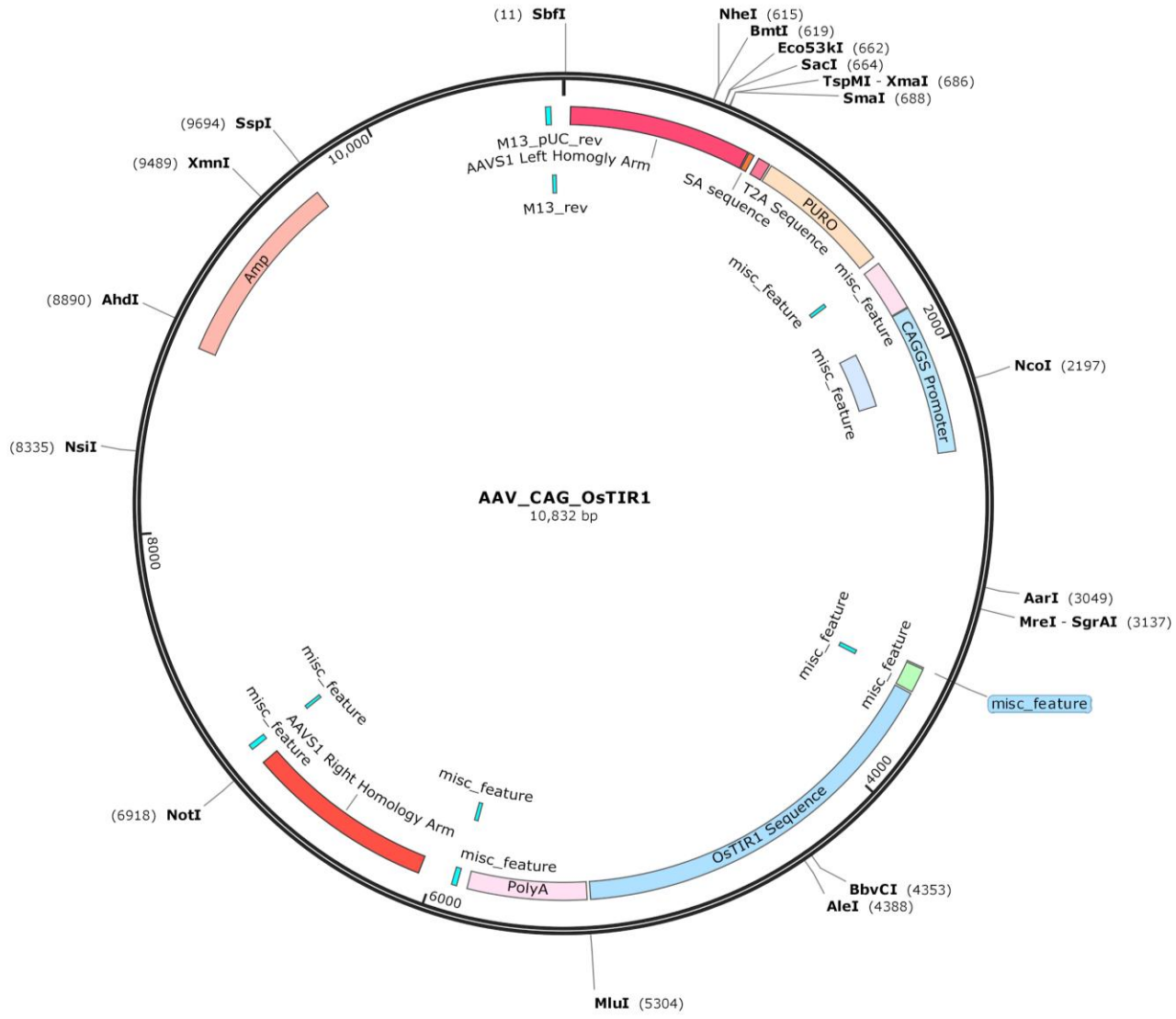
6.2.3 CRISPR interference targeting epigenetic modifiers.

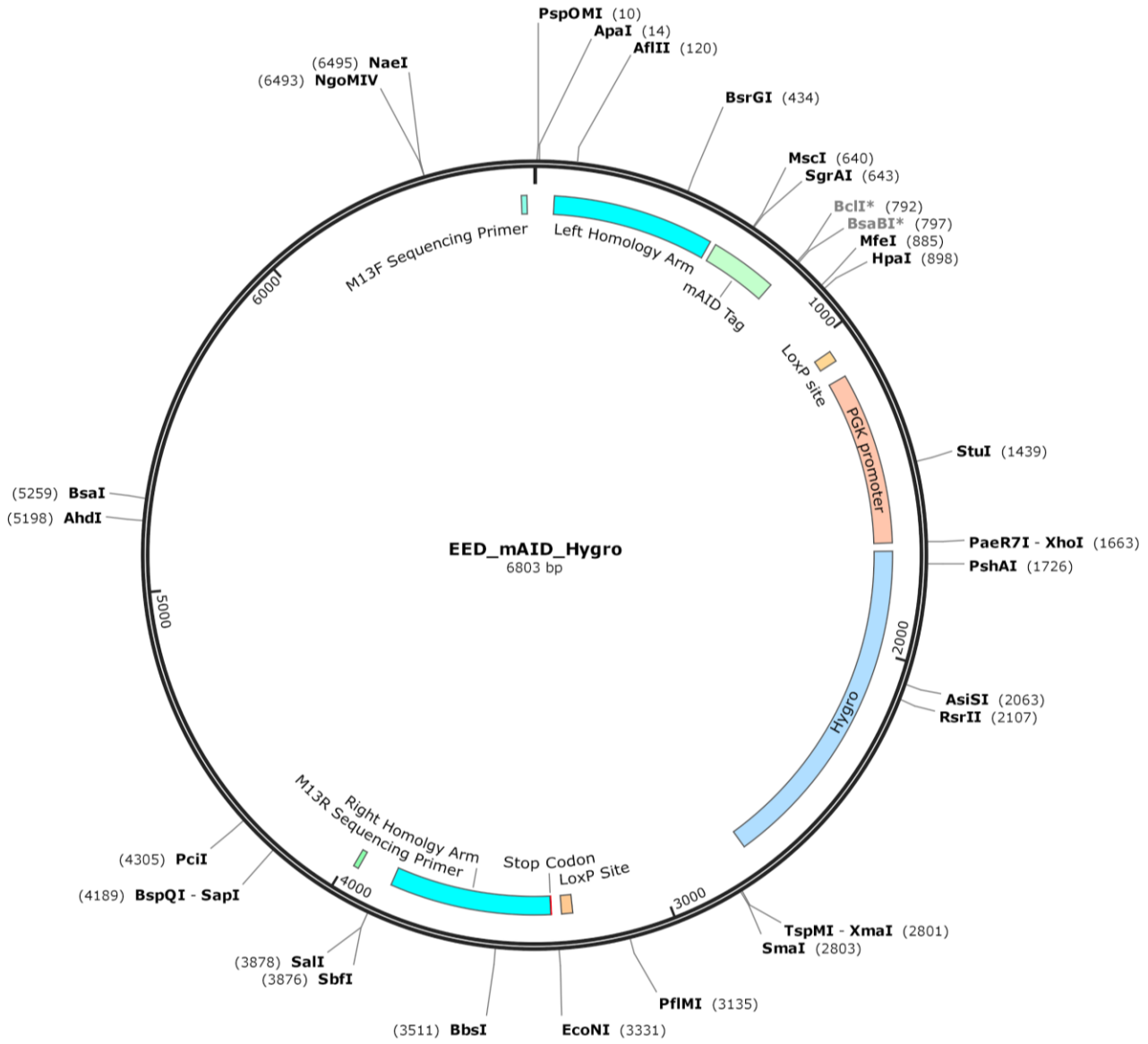
Although this study focused on PRC2 in pancreatic differentiation, the role of epigenetic modifiers in cell differentiation is a much broader field, and the work included within this thesis could serve as a proof of principle model. This is possible due to the ability to control PRC2 through the EZH1/2 enzymatic activity, and therefore molecular tools to investigate other proteins could be necessary to examine the roles of other histone modifying complexes. The adaption of CRISPR interference may benefit this, as this approach appears to be efficient at removing the majority of *EED* from the hPSC lines tested. However, further experiments are required to ensure this system can reduce H3K27me3 and allow their recovery after withdrawal of the doxycycline that induces depletion. This study demonstrated the depletion of *EED* mRNA to much lower levels, with a resulting increase in repressed gene expression but whether this translates to a similar loss of protein and prolonged PRC2 activity loss is unclear. Protein analysis of EED levels and H3K27me3 following the induction of CRISPRi will be sufficient to demonstrate the efficiency of the CRISPRi system in hPSCs. Once the efficiency of this is shown, the CRISPRi system can be used to target other modifiers during the differentiation protocol, such as those responsible for H3K4 methylation, histone acetylation and DNA methylation, to determine if it is possible to positively affect cell differentiation through control of these epigenetic modifications, and also can be used to target multiple modifying complexes in a single experiment. It could also be used to target epigenetic modifiers that are responsible for the removal of these modifications, to determine if these will have the opposing effects than observed when targeting enzymes responsible for deposition of epigenetics.

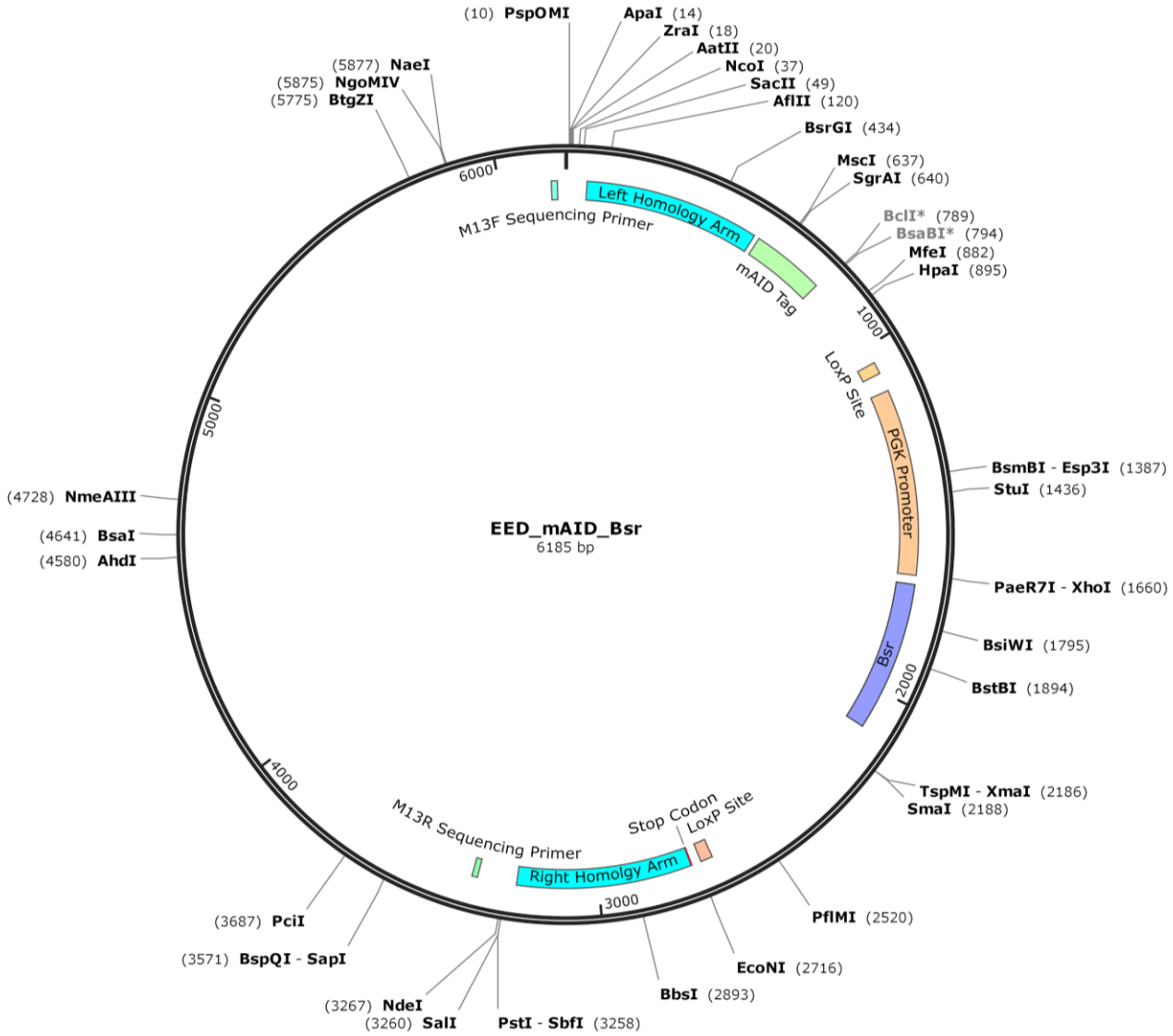
Chapter 7

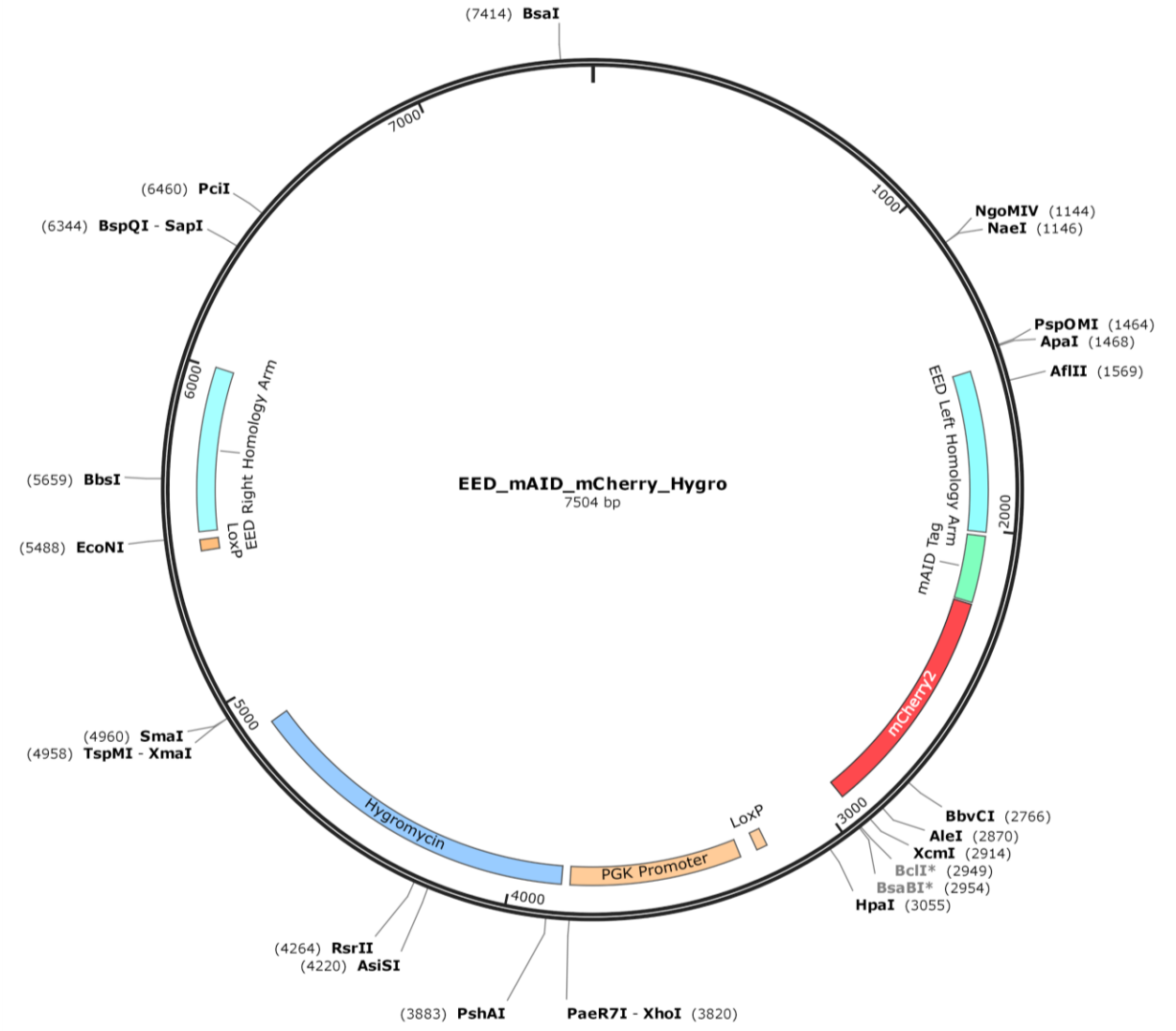
Appendix

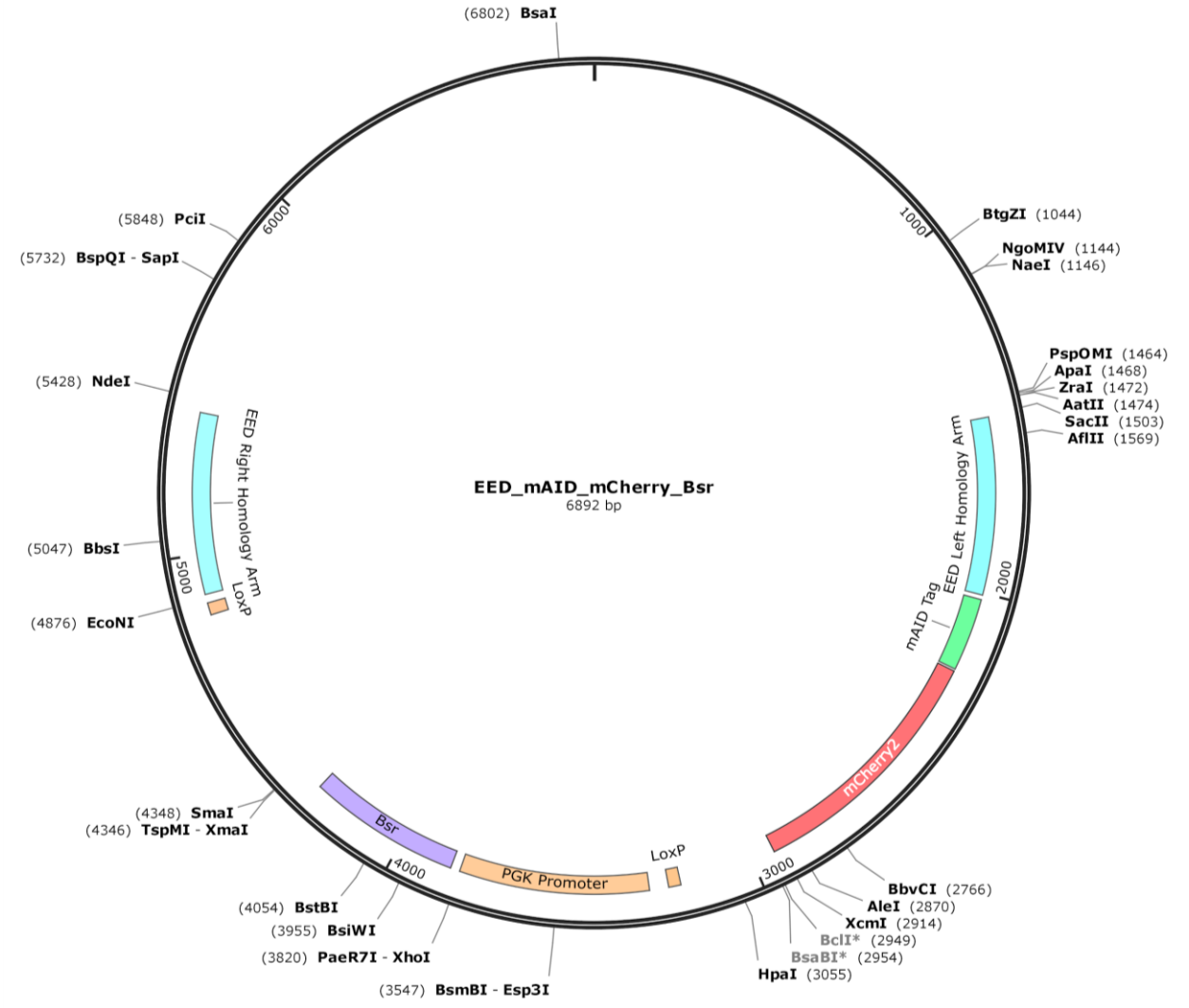
Appendix I - Plasmid Maps for Auxin Inducible Degron System




PspOMI (10)
ApaI (14)
AflII (120)
BsrGI (434)
MscI (640)
SgrAI (643)
BclI* (792)
BsaBI* (797)
MfeI (885)
HpaI (898)
StuI (1439)
PaeR7I - XhoI (1663)
PshAI (1726)
AsiSI (2063)
RsrII (2107)
TspMI - XmaI (2801)
SmaI (2803)
PflMI (3135)
EcoNI (3331)
BbsI (3511)
SalI
SbfI (3876)
(3878)
(4189) BspQI - SapI
PciI (4305)
(5198) AhdI
BsaI (5259)
NaeI (6495)
NgoMIV (6493)
M13F Sequencing Primer
Left Homology Arm
mAID Tag
LoxP site
PGK promoter
Hygro
Right Homolgy Arm
M13R Sequencing Primer
Stop Codon
LoxP Site
1000
2000
3000
4000
5000
6000
EED_mAID_Hygro
6803 bp


EED_mAID_Bsr
6185 bp
(10) PspOMI
ApaI (14)
ZraI (18)
AatII (20)
NcoI (37)
SacII (49)
AflII (120)
BsrGI (434)
MscI (637)
SgrAI (640)
BclI* (789)
BsaBI* (794)
MfeI (882)
HpaI (895)
BsmBI - Esp3I (1387)
StuI (1436)
PaeR7I - XhoI (1660)
BsiWI (1795)
BstBI (1894)
TspMI - XmaI (2186)
SmaI (2188)
PflMI (2520)
EcoNI (2716)
BbsI (2893)
PstI - SbfI (3258)
(3260) SalI
(3267) NdeI
(3571) BspQI - SapI
(3687) PciI
(4580) AhdI
(4641) BsaI
(4728) NmeAIII
(5775) BtgZI
(5875) NgoMIV
(5877) NaeI
1000
2000
3000
4000
5000
6000
M13F Sequencing Primer
Left Homology Arm
mAID Tag
LoxP Site
PGK Promoter
Bsr
Stop Codon
LoxP Site
M13R Sequencing Primer
Right Homolgy Arm


(7414) BsaI
7000
1000
NgoMIV (1144)
NaeI (1146)
PspOMI (1464)
ApaI (1468)
AflII (1569)
EED Left Homology Arm
2000
mAID Tag
mCherry2
BbvCI (2766)
AleI (2870)
XcmI (2914)
BclI* (2949)
BsaBI* (2954)
HpaI (3055)
3000
LoxP
PGK Promoter
Hygromycin
4000
(4264) RsrII
(4220) AsiSI
(3883) PshAI
PaeR7I - XhoI (3820)
5000
(4960) SmaI
(4958) TspMI - XmaI
(5659) BbsI
(5488) EcoNI
LoxP
EED Right Homology Arm
6000
(6460) PciI
(6344) BspQI - SapI
EED_mAID_mCherry_Hygro
7504 bp


(6802) BsaI
BtgZI (1044)
NgoMIV (1144)
NaeI (1146)
PspOMI (1464)
ApaI (1468)
ZraI (1472)
AatII (1474)
SacII (1503)
AflII (1569)
EED_mAID_mCherry_Bsr
6892 bp
EED Left Homology Arm
mAID Tag
mCherry2
LoxP
PGK Promoter
Bsr
BbvCI (2766)
AleI (2870)
XcmI (2914)
BclI* (2949)
BsaBI* (2954)
HpaI (3055)
(3547) BsmBI - Esp3I
(3820) PaeR7I - XhoI
(3955) BsiWI
(4054) BstBI
(4346) TspMI - XmaI
(4348) SmaI
(4876) EcoNI
(5047) BbsI
(5428) NdeI
(5732) BspQI - SapI
(5848) PciI
EED Right Homology Arm
LoxP
1000
2000
3000
4000
5000
6000

Appendix II - Sequencing Read Counts for Pancreatic *in vitro* Differentiation

Sample	Differentiation Replicate	RNA-Sequencing Readcount	H3K4me3 ChIP-Sequencing	H3K27me3 ChIP-Sequencing
Day 0	Replicate 1	15470027	33263317	19512185
	Replicate 2	13499437	27611517	27611517
	Replicate 3	6643983	12365113	16884682
Day 4	Replicate 1	15613313	26348718	9343266
	Replicate 2	10485281	23731361	24144367
	Replicate 3	14368597	17778230	22703728
Day 9	Replicate 1	34872963	29854213	46612672
	Replicate 2	14404119	13282611	10107524
	Replicate 3	18138157	86732522	15895394
Day 17	Replicate 1	6939540	25192331	35485282
	Replicate 2	17923886	45515981	35127735
	Replicate 3	24077190	32150707	55185766
Day 27	Replicate 1	14443185	14422074	27912538
	Replicate 2	5893701	22427260	33143712
	Replicate 3	9304760	14896467	18993543

Appendix III - Sequencing Read Counts from Pancreatic *in vitro* Differentiation with Inhibition

ChIP-Seq Differentiation Sample	Human Reads	Drosophila Reads	% Human reads/ Total reads
DMSO_Timepoint1_ Rep1_H3K27me3	21051272	2258243	89.4
DMSO_Timepoint1_ Rep2_H3K27me3	8160360	8209274	49.2
DMSO_Timepoint1_ Rep3_H3K27me3	15918902	19380645	44.5
DMSO_Timepoint2_ Rep1_H3K27me3	8962591	934076	89.7
DMSO_Timepoint2_ Rep2_H3K27me3	15724595	11892518	47.7
DMSO_Timepoint2_ Rep3_H3K27me3	11663134	10998402	48.1
DMSO_Timepoint3_ Rep2_H3K27me3	12819962	10660831	53.4
DMSO_Timepoint3_ Rep3_H3K27me3	8618591	5661859	58.5
DMSO_Timepoint4_ Rep1_H3K27me3	12392648	9172638	57.5
DMSO_Timepoint4_ Rep2_H3K27me3	5287818	7553912	40.6
DMSO_Timepoint4_ Rep3_H3K27me3	14574764	6011418	42.5
GSK343_Timepoint1_ Rep1_H3K27me3	21502676	2801371	87.5
GSK343_Timepoint1_ Rep2_H3K27me3	6108888	9156147	39.4
GSK343_Timepoint1_ Rep3_H3K27me3	6100296	9799499	37.8
GSK343_Timepoint2_	37556993	4204751	88.7

Rep1_H3K27me3			
GSK343_Timepoint2_ Rep2_H3K27me3	753478	1183355	38.4
GSK343_Timepoint3_ Rep1_H3K27me3	21345021	6374315	75.5
GSK343_Timepoint3_ Rep2_H3K27me3	15929843	10674463	58.9
GSK343_Timepoint3_ Rep3_H3K27me3	15891874	11560336	56.6
GSK343_Timepoint4_ Rep1_H3K27me3	8273402	88991	98.9
GSK343_Timepoint4_ Rep2_H3K27me3	4046225	7209329	35.4
GSK343_Timepoint4_ Rep3_H3K27me3	9455336	14431545	39.0
UNC1999_Timepoint1_ Rep1_H3K27me3	14383773	6171741	68.5
UNC1999_Timepoint1_ Rep2_H3K27me3	5820846	8944447	38.9
UNC1999_Timepoint1_ Rep3_H3K27me3	4345975	6968064	37.8
UNC1999_Timepoint2_ Rep2_H3K27me3	11918916	18096307	39.1
UNC1999_Timepoint2_ Rep3_H3K27me3	40918687	29541629	57.2
UNC1999_Timepoint3_ Rep1_H3K27me3	11116780	17566184	38.1
UNC1999_Timepoint3_ Rep2_H3K27me3	13315867	13665361	48.1
UNC1999_Timepoint3_ Rep3_H3K27me3	14571593	6518603	67.3
UNC1999_Timepoint4_ Rep1_H3K27me3	22475226	3876231	83.8
UNC1999_Timepoint4_ Rep2_H3K27me3	6678557	8479649	43.4

DMSO_Timepoint1_ Rep1_H3K4me3	20548928	2247512	88.8
DMSO_Timepoint1_ Rep2_H3K4me3	11347524	6425439	62.5
DMSO_Timepoint1_ Rep3_H3K4me3	33979582	9290552	77.4
DMSO_Timepoint2_ Rep1_H3K4me3	51326830	5061026	90.1
DMSO_Timepoint2_ Rep2_H3K4me3	16179542	11999317	56.4
DMSO_Timepoint2_ Rep3_H3K4me3	44063091	5913309	87.1
DMSO_Timepoint3_ Rep1_H3K4me3	29144401	2991627	89.8
DMSO_Timepoint3_ Rep2_H3K4me3	17537175	13780965	55.1
DMSO_Timepoint3_ Rep3_H3K4me3	12459576	1193628	90.1
DMSO_Timepoint4_ Rep1_H3K4me3	1354568	173561	87.6
DMSO_Timepoint4_ Rep2_H3K4me3	14969322	12662642	53.2
DMSO_Timepoint4_ Rep3_H3K4me3	13217640	10335671	55.1
GSK343_Timepoint1_ Rep1_H3K4me3	9979331	4778254	66.1
GSK343_Timepoint1_ Rep2_H3K4me3	12178869	3432939	76.2
GSK343_Timepoint1_ Rep3_H3K4me3	6943898	18163209	27.0
GSK343_Timepoint2_ Rep1_H3K4me3	81099656	11401896	86.3
GSK343_Timepoint2_ Rep2_H3K4me3	14769388	14004768	50.4
GSK343_Timepoint2_	19589347	6054782	75.3

Rep3_H3K4me3			
GSK343_Timepoint3_ Rep1_H3K4me3	22013776	3277532	86.1
GSK343_Timepoint3_ Rep2_H3K4me3	39704615	23920385	61.3
GSK343_Timepoint3_ Rep3_H3K4me3	21822104	5911521	77.0
GSK343_Timepoint4_ Rep1_H3K4me3	15641439	2490578	84.7
GSK343_Timepoint4_ Rep2_H3K4me3	9658976	16262332	36.5
GSK343_Timepoint4_ Rep3_H3K4me3	9121700	10571500	45.4
UNC1999_Timepoint1_ Rep2_H3K4me3	17699263	18072194	49.5
UNC1999_Timepoint1_ Rep3_H3K4me3	30964103	78279858	28.3
UNC1999_Timepoint2_ Rep1_H3K4me3	43132260	5577711	88.5
UNC1999_Timepoint2_ Rep3_H3K4me3	13330641	11500726	53.7
UNC1999_Timepoint3_ Rep1_H3K4me3	30847931	2947380	91.3
UNC1999_Timepoint3_ Rep2_H3K4me3	31410374	4140124	88.4
UNC1999_Timepoint3_ Rep3_H3K4me3	6310442	1832837	77.5
UNC1999_Timepoint4_ Rep1_H3K4me3	15402322	1821250	89.4
UNC1999_Timepoint4_ Rep2_H3K4me3	12115890	14099756	46.2
UNC1999_Timepoint4_ Rep2_H3K4me3	48432406	52828915	47.8
DMSO_Timepoint1_ Rep1_Input	15019653	1828870	89.1

DMSO_Timepoint1_ Rep2_Input	5172135	347417	93.7
DMSO_Timepoint1_ Rep3_Input	7328010	637100	92.0
DMSO_Timepoint2_ Rep1_Input	7360819	801726	90.2
DMSO_Timepoint2_ Rep2_Input	5569297	589542	90.4
DMSO_Timepoint2_ Rep3_Input	7360482	710428	91.2
DMSO_Timepoint3_ Rep1_Input	5977284	236803	96.2
DMSO_Timepoint3_ Rep2_Input	7360482	710428	91.2
DMSO_Timepoint3_ Rep3_Input	4923460	430421	92.0
DMSO_Timepoint4_ Rep1_Input	9254735	514720	94.7
DMSO_Timepoint4_ Rep2_Input	10093207	1008927	90.9
DMSO_Timepoint4_ Rep3_Input	2906219	708753	80.4
GSK343_Timepoint1_ Rep1_Input	7790276	595691	92.9
GSK343_Timepoint1_ Rep2_Input	6907142	843240	80.0
GSK343_Timepoint1_ Rep3_Input	4981446	1580729	75.9
GSK343_Timepoint2_ Rep1_Input	5470823	760068	87.8
GSK343_Timepoint2_ Rep2_Input	8209691	1106977	88.1
GSK343_Timepoint2_ Rep3_Input	1583564	527422	75.0
GSK343_Timepoint3_	4230285	219938	95.1

Rep1_Input			
GSK343_Timepoint3_ Rep2_Input	5165076	366726	93.4
GSK343_Timepoint3_ Rep3_Input	8926393	858307	91.2
GSK343_Timepoint4_ Rep1_Input	8842624	370706	96.0
GSK343_Timepoint4_ Rep2_Input	8760646	750126	92.1
GSK343_Timepoint4_ Rep3_Input	6694164	1823134	78.6
UNC1999_Timepoint1_ Rep1_Input	5857943	427371	92.1
UNC1999_Timepoint1_ Rep2_Input	6487104	602286	91.5
UNC1999_Timepoint1_ Rep3_Input	1817008	620909	74.5
UNC1999_Timepoint2_ Rep1_Input	85845	13260	86.6
UNC1999_Timepoint2_ Rep2_Input	12030484	1431596	89.4
UNC1999_Timepoint2_ Rep3_Input	1828267	217221	89.4
UNC1999_Timepoint3_ Rep1_Input	11799957	432948	96.5
UNC1999_Timepoint3_ Rep2_Input	15272331	865194	94.6
UNC1999_Timepoint3_ Rep3_Input	356117	133933	72.7
UNC1999_Timepoint4_ Rep1_Input	6431277	309312	95.4
UNC1999_Timepoint4_ Rep2_Input	11276994	1120964	91.0
UNC1999_Timepoint4_	4601921	1052440	81.4

Rep3_Input			

Sample Timepoint	**Differentiation Replicate**	**DMSO RNA-Seq Reads**	**GSK343 RNA-Seq Reads**	**UNC1999 RNA-Seq Reads**
Timepoint 1	Rep 1	20950275	39900574	37220127
	Rep 2	37220127	30671853	20548571
	Rep 3	20548571	41842005	37765484
Timepoint 2	Rep 1	22018723	21945135	13217588
	Rep 2	21928396	12060914	11415166
	Rep 3	11415166	38967833	9209596
Timepoint 3	Rep 1	21777709	25886624	10646463
	Rep 2	28481994	34606450	7462064
	Rep 3	7462064	17173739	16761814
Timepoint 4	Rep 1	30519411	29194594	37768216
	Rep 2	43910331	45438354	43057996
	Rep 3	34930598	33558755	15233582

Chapter 8

References

Adli, M. (2018) ‘The CRISPR tool kit for genome editing and beyond’, *Nature Communications*. Nature Publishing Group. doi: 10.1038/s41467-018-04252-2.

Agger, K. *et al.* (2007) ‘UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development’, *Nature*. Nature Publishing Group, 449(7163), pp. 731–734. doi: 10.1038/nature06145.

Allis, C. D. and Jenuwein, T. (2016) ‘The molecular hallmarks of epigenetic control’, *Nature Reviews Genetics*. Nature Publishing Group, pp. 487–500. doi: 10.1038/nrg.2016.59.

Almeida, M. *et al.* (2017) ‘PCGF3/5-PRC1 initiates Polycomb recruitment in X chromosome inactivation’, *Science*. American Association for the Advancement of Science, 356(6342), pp. 1081–1084. doi: 10.1126/science.aal2512.

Van Arensbergen, J. *et al.* (2010) ‘Derepression of polycomb targets during pancreatic organogenesis allows insulin-producing beta-cells to adopt a neural gene activity program’, *Genome Research*. Cold Spring Harbor Laboratory Press, 20(6), pp. 722–732. doi: 10.1101/gr.101709.109.

Van Arensbergen, J. *et al.* (2017) ‘A distal intergenic region controls pancreatic endocrine differentiation by acting as a transcriptional enhancer and as a polycomb response element’. doi: 10.1371/journal.pone.0171508.

Ashcroft, F. M., Harrison, D. E. and Ashcroft, S. J. H. (1984) ‘Glucose induces closure of single potassium channels in isolated rat pancreatic β-cells’, *Nature*. Nature, 312(5993), pp. 446–448. doi: 10.1038/312446a0.

Association, A. D. (2010) ‘Diagnosis and classification of diabetes mellitus’, *Diabetes Care*. American Diabetes Association, pp. S62–S69. doi: 10.2337/dc10-S062.

Atouf, F., Czernichow, P. and Scharfmann, R. (1997) ‘Expression of neuronal traits in pancreatic beta cells: Implication of neuron-restrictive silencing factor/repressor element silencing transcription factor, a neuron-restrictive silencer’, *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology, 272(3), pp. 1929–1934. doi: 10.1074/jbc.272.3.1929.

Avior, Y., Sagi, I. and Benvenisty, N. (2016) ‘Pluripotent stem cells in disease modelling and drug discovery’, *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, pp. 170–182. doi: 10.1038/nrm.2015.27.

Azuara, V. *et al.* (2006) ‘Chromatin signatures of pluripotent cell lines’, *Nature Cell Biology*. Nature Publishing Group, 8(5), pp. 532–538. doi: 10.1038/ncb1403.

Ballaré, C. *et al.* (2012) ‘Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity’, *Nature Structural and Molecular Biology*. NIH Public Access, 19(12), pp. 1257–1265. doi: 10.1038/nsmb.2434.

Ballinger, W. F. and Lacy, P. E. (1972) ‘Transplantation of intact pancreatic islets in rats’, *Surgery*. Elsevier, 72(2), pp. 175–186. doi: 10.5555/uri:pii:0039606072903315.

Banaszynski, L. A. *et al.* (2006) ‘A Rapid, Reversible, and Tunable Method to Regulate Protein Function in Living Cells Using Synthetic Small Molecules’, *Cell*. Cell Press, 126(5), pp. 995–1004. doi: 10.1016/j.cell.2006.07.025.

Banaszynski, L. A. *et al.* (2008) ‘Chemical control of protein stability and function in living mice’, *Nature Medicine*. Nature Publishing Group, 14(10), pp. 1123–1127. doi: 10.1038/nm.1754.

Banga, A. *et al.* (2012) ‘In vivo reprogramming of Sox9+ cells in the liver to insulin-secreting ducts’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 109(38), pp. 15336–15341. doi: 10.1073/pnas.1201701109.

Bannister, A. J. *et al.* (2005) ‘Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes’, *Journal of Biological Chemistry*. J Biol Chem, 280(18), pp. 17732–17736. doi: 10.1074/jbc.M500796200.

Bannister, A. J. and Kouzarides, T. (2011) ‘Regulation of chromatin by histone modifications’, *Cell Research*. Nature Publishing Group, pp. 381–395. doi: 10.1038/cr.2011.22.

Banting, F. G. *et al.* (no date) *THE CANADIAN MEDICAL ASSOCIATION JOURNAL PANCREATIC EXTRACTS IN THE TREATMENT OF DIABETES MELLITUS PRELIMINARY REPORT*.

Bar-Nur, O. *et al.* (2011) 'Epigenetic memory and preferential lineage-specific differentiation in induced pluripotent stem cells derived from human pancreatic islet beta cells', *Cell Stem Cell*. Cell Press, 9(1), pp. 17–23. doi: 10.1016/j.stem.2011.06.007.

Baron, M. *et al.* (2016) 'A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure', *Cell Systems*. Cell Press, 3(4), pp. 346-360.e4. doi: 10.1016/j.cels.2016.08.011.

Barski, A. *et al.* (2007) 'High-Resolution Profiling of Histone Methylations in the Human Genome', *Cell*. Elsevier, 129(4), pp. 823–837. doi: 10.1016/j.cell.2007.05.009.

Basford, C. L. *et al.* (2012) 'The functional and molecular characterisation of human embryonic stem cell-derived insulin-positive cells compared with adult pancreatic beta cells', *Diabetologia*. Springer, 55(2), pp. 358–371. doi: 10.1007/s00125-011-2335-x.

Becker, A. J., McCulloch, E. A. and Till, J. E. (1963) 'Cytological Demonstration of the Clonal Nature of Spleen Colonies Derived from Transplanted Mouse Marrow Cells', *Nature*. Nature Publishing Group, 197, p. 452. Available at: http://dx.doi.org/10.1038/197452a0.

Beltran, M. *et al.* (2016) 'The interaction of PRC2 with RNA or chromatin is mutually antagonistic'. doi: 10.1101/gr.197632.115.

van den Berg, J. *et al.* (2018) 'A limited number of double-strand DNA breaks is sufficient to delay cell cycle progression', *Nucleic Acids Research*. Oxford University Press, 46(19), pp. 10132–10144. doi: 10.1093/nar/gky786.

Beringer, M. *et al.* (2016) 'EPOP Functionally Links Elongin and Polycomb in Pluripotent Stem Cells', *Molecular Cell*. Cell Press, 64(4), pp. 645–658. doi: 10.1016/j.molcel.2016.10.018.

Bernstein, B. E. *et al.* (2002) 'Methylation of histone H3 Lys 4 in coding regions of active genes', *Proceedings of the National Academy of Sciences of the United States of America*. 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.

Bertero, A. *et al.* (2016) 'Optimized inducible shRNA and CRISPR/Cas9 platforms for in vitro studies of human development using hPSCs.', *Development*, 143(23), pp. 4405–4418. doi: 10.1242/dev.138081.

Bhandare, R. *et al.* (2010) 'Genome-wide analysis of histone modifications in human pancreatic islets', *Genome Research*. Cold Spring Harbor Laboratory Press, 20(4), pp. 428–433. doi: 10.1101/gr.102038.109.

Bilodeau, S. *et al.* (2009) 'SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state', *Genes and Development*, 23(21), pp. 2484–2489. doi: 10.1101/gad.1837309.

Blackledge, N. P. *et al.* (2014) 'Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation', *Cell*. Cell Press, 157(6), pp. 1445–1459. doi: 10.1016/j.cell.2014.05.004.

Blackledge, N. P. *et al.* (2020) 'PRC1 Catalytic Activity Is Central to Polycomb System Function', *Molecular Cell*. Cell Press, 77(4), pp. 857-874.e9. doi: 10.1016/j.molcel.2019.12.001.

Blackledge, N. P., Rose, N. R. and Klose, R. J. (2015) 'Targeting Polycomb systems to regulate gene expression: Modifications to a complex story', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 16(11), pp. 643–649. doi: 10.1038/nrm4067.

Bollati, V. *et al.* (2009) 'Decline in genomic DNA methylation through aging in a cohort of elderly subjects', *Mechanisms of Ageing and Development*. Elsevier, 130(4), pp. 234–239. doi: 10.1016/j.mad.2008.12.003.

Bondeson, D. P. *et al.* (2015) 'Catalytic in vivo protein knockdown by small-molecule PROTACs', *Nature Chemical Biology*. Nature Publishing Group, 11(8), pp. 611–617. doi: 10.1038/nchembio.1858.

Bonn, S. *et al.* (2012) 'Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development', *Nature Genetics*. Nat Genet, 44(2), pp. 148–156. doi: 10.1038/ng.1064.

Bonner-Weir, S. and Weir, G. C. (2005) 'New sources of pancreatic β-cells', *Nature*

Biotechnology. Nature Publishing Group, 23(7), pp. 857–861. doi: 10.1038/nbt1115.

Boyer, L. A. *et al.* (2006) 'Polycomb complexes repress developmental regulators in murine embryonic stem cells', *Nature*. Nature, 441(7091), pp. 349–353. doi: 10.1038/nature04733.

Bracken, A. P. *et al.* (2006) 'Genome-wide mapping of polycomb target genes unravels their roles in cell fate transitions', *Genes and Development*. Cold Spring Harbor Laboratory Press, 20(9), pp. 1123–1136. doi: 10.1101/gad.381706.

Bradley, J. A., Bolton, E. M. and Pedersen, R. A. (2002) 'Stem cell medicine encounters the immune system', *Nature Reviews Immunology*. Nature Publishing Group, pp. 859–871. doi: 10.1038/nri934.

Brien, G. L. *et al.* (2012) 'Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation', *Nature Structural and Molecular Biology*, 19(12), pp. 1273–1281. doi: 10.1038/nsmb.2449.

Brissova, M. *et al.* (2005) 'Assessment of human pancreatic islet architecture and composition by laser scanning confocal microscopy.', *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society*. SAGE PublicationsSage CA: Los Angeles, CA, 53(9), pp. 1087–97. doi: 10.1369/jhc.5C6684.2005.

Bruin, J. E. *et al.* (2013) 'Maturation and function of human embryonic stem cell-derived pancreatic progenitors in macroencapsulation devices following transplant into mice', *Diabetologia*. Springer, 56(9), pp. 1987–1998. doi: 10.1007/s00125-013-2955-4.

Bruin, J. E. *et al.* (2014) 'Characterization of polyhormonal insulin-producing cells derived in vitro from human embryonic stem cells', *Stem Cell Research*. Elsevier, 12(1), pp. 194–208. doi: 10.1016/j.scr.2013.10.003.

Bulger, M. and Groudine, M. (2011) 'Functional and mechanistic diversity of distal transcription enhancers', *Cell*. Cell Press, pp. 327–339. doi: 10.1016/j.cell.2011.01.024.

Bunker, C. A. and Kingston, R. E. (1994) 'Transcriptional repression by Drosophila and mammalian Polycomb group proteins in transfected mammalian cells.', *Molecular and Cellular Biology*. American Society for Microbiology, 14(3), pp. 1721–1732. doi: 10.1128/mcb.14.3.1721.

Burridge, P. W. *et al.* (2012) 'Production of de novo cardiomyocytes: Human pluripotent stem

cell differentiation and direct reprogramming', *Cell Stem Cell*. Elsevier, pp. 16–28. doi: 10.1016/j.stem.2011.12.013.

Cabrera, O. *et al.* (2006) 'The unique cytoarchitecture of human pancreatic islets has implications for islet cell function', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 103(7), pp. 2334–2339. doi: 10.1073/pnas.0510790103.

Cano-Rodriguez, D. *et al.* (2016) 'Writing of H3K4Me3 overcomes epigenetic silencing in a sustained but context-dependent manner', *Nature Communications*. Nature Publishing Group, 7. doi: 10.1038/ncomms12284.

Cao, R. *et al.* (2002a) 'Role of histone H3 lysine 27 methylation in polycomb-group silencing', *Science*. American Association for the Advancement of Science, 298(5595), pp. 1039–1043. doi: 10.1126/science.1076997.

Cao, R. *et al.* (2002b) 'Role of histone H3 lysine 27 methylation in polycomb-group silencing', *Science*. American Association for the Advancement of Science, 298(5595), pp. 1039–1043. doi: 10.1126/science.1076997.

Cao, R. and Zhang, Y. (2004) 'SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex', *Molecular Cell*. Cell Press, 15(1), pp. 57–67. doi: 10.1016/j.molcel.2004.06.020.

Cebola, I. *et al.* (2015) 'TEAD and YAP regulate the enhancer network of human embryonic pancreatic progenitors', *Nature Cell Biology*. Nature Publishing Group, 17(5), pp. 615–626. doi: 10.1038/ncb3160.

Chakrabarti, S. K. and Mirmira, R. G. (2003) 'Transcription factors direct the development and function of pancreatic β cells', *Trends in Endocrinology and Metabolism*. Elsevier Current Trends, pp. 78–84. doi: 10.1016/S1043-2760(02)00039-5.

Chamberlain, S. J., Yee, D. and Magnuson, T. (2008) 'Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency', *Stem Cells*, 26(6), pp. 1496–1505. doi: 10.1634/stemcells.2008-0102.

Chen, E. Y. *et al.* (2013) 'Enrichr: Interactive and collaborative HTML5 gene list enrichment

analysis tool', *BMC Bioinformatics*. BMC Bioinformatics, 14. doi: 10.1186/1471-2105-14-128.

Chen, G. *et al.* (2011) 'Chemically defined conditions for human iPSC derivation and culture', *Nature Methods*. Nat Methods, 8(5), pp. 424–429. doi: 10.1038/nmeth.1593.

Chen, T. and Dent, S. Y. R. (2013) 'Chromatin modifiers and remodellers: regulators of cellular differentiation', *Nature Reviews Genetics*. Nature Publishing Group, 15(2), pp. 93–106. doi: 10.1038/nrg3607.

Chen, X. *et al.* (2019) 'Degradation of endogenous proteins and generation of a null-like phenotype in zebrafish using Trim-Away technology', *Genome Biology*. BioMed Central Ltd., 20(1), p. 19. doi: 10.1186/s13059-019-1624-4.

Cheng, X. *et al.* (2012) 'Self-renewing endodermal progenitor lines generated from human pluripotent stem cells', *Cell Stem Cell*. NIH Public Access, 10(4), pp. 371–384. doi: 10.1016/j.stem.2012.02.024.

Cho, C. H. H. *et al.* (2012) 'Inhibition of activin/nodal signalling is necessary for pancreatic differentiation of human pluripotent stem cells', *Diabetologia*, 55(12), pp. 3284–3295. doi: 10.1007/s00125-012-2687-x.

Christensen, B. C. *et al.* (2009) 'Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CPG island context', *PLoS Genetics*. PLoS Genet, 5(8). doi: 10.1371/journal.pgen.1000602.

Christensen, J. *et al.* (2007) 'RBP2 Belongs to a Family of Demethylases, Specific for Tri-and Dimethylated Lysine 4 on Histone 3', *Cell*. Cell Press, 128(6), pp. 1063–1076. doi: 10.1016/j.cell.2007.02.003.

Cim, A. *et al.* (2012) 'In vivo studies on non-viral transdifferentiation of liver cells towards pancreatic β cells', *Journal of Endocrinology*. J Endocrinol, 214(3), pp. 277–288. doi: 10.1530/JOE-12-0033.

Clift, D. *et al.* (2017) 'A Method for the Acute and Rapid Degradation of Endogenous Proteins', *Cell*. Cell Press, 171(7), pp. 1692-1706.e18. doi: 10.1016/j.cell.2017.10.033.

Clift, D. *et al.* (2018) 'Acute and rapid degradation of endogenous proteins by Trim-Away', *Nature Protocols*. Nature Publishing Group, 13(10), pp. 2149–2175. doi: 10.1038/s41596-018-

0028-3.

Clouaire, T. *et al.* (2012) 'Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells', *Genes and Development*. Cold Spring Harbor Laboratory Press, 26(15), pp. 1714–1728. doi: 10.1101/gad.194209.112.

Coleman, R. T. and Struhl, G. (2017) 'Causal role for inheritance of H3K27me3 in maintaining the off state of a Drosophila HOX gene', *Science*. American Association for the Advancement of Science, 356(6333). doi: 10.1126/science.aai8236.

Collinson, A., Collier, Amanda J, *et al.* (2016) 'Deletion of the Polycomb-Group Protein EZH2 Leads to Compromised Self-Renewal and Differentiation Defects in Human Embryonic Stem Cells', *Cell Reports*. ElsevierCompany., 17(6), pp. 2700–2714. doi: 10.1016/j.celrep.2016.11.032.

Collinson, A., Collier, Amanda J., *et al.* (2016) 'Deletion of the Polycomb-Group Protein EZH2 Leads to Compromised Self-Renewal and Differentiation Defects in Human Embryonic Stem Cells', *Cell Reports*. Elsevier B.V., 17(10), pp. 2700–2714. doi: 10.1016/j.celrep.2016.11.032.

Collombat, P. *et al.* (2006) 'Specifying pancreatic endocrine cell fates', *Mechanisms of Development*. Elsevier, pp. 501–512. doi: 10.1016/j.mod.2006.05.006.

Conrad, E., Stein, R. and Hunter, C. S. (2014) 'Revealing transcription factors during human pancreatic β cell development', *Trends in Endocrinology and Metabolism*. Elsevier Inc., pp. 407–414. doi: 10.1016/j.tem.2014.03.013.

Conway, E. *et al.* (2018) 'A Family of Vertebrate-Specific Polycombs Encoded by the LCOR/LCORL Genes Balance PRC2 Subtype Activities', *Molecular Cell*. Cell Press, 70(3), pp. 408-421.e8. doi: 10.1016/j.molcel.2018.03.005.

Cook, D. L. and Hales, N. (1984) 'Intracellular ATP directly blocks K+ channels in pancreatic B-cells', *Nature*. Nature, 311(5983), pp. 271–273. doi: 10.1038/311271a0.

Cooper, S. *et al.* (2014) 'Targeting Polycomb to Pericentric Heterochromatin in Embryonic Stem Cells Reveals a Role for H2AK119u1 in PRC2 Recruitment', *Cell Reports*. Elsevier, 7(5), pp. 1456–1470. doi: 10.1016/j.celrep.2014.04.012.

Cooper, S. *et al.* (2016) 'Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk

between Polycomb complexes PRC1 and PRC2', *Nature Communications*. Nature Publishing Group, 7, p. 13661. doi: 10.1038/ncomms13661.

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*, 107(50), pp. 21931–21936. doi: 10.1073/pnas.1016071107.

Croce, L. Di and Helin, K. (2013) 'Transcriptional regulation by Polycomb group proteins', *Nature Publishing Group*, 20. doi: 10.1038/nsmb.2669.

Cui, K. *et al.* (2008) 'Resource Chromatin Signatures in Multipotent Human Hematopoietic Stem Cells Indicate the Fate of Bivalent Genes during Differentiation', *Stem Cell*. Elsevier Inc., 4(1), pp. 80–93. doi: 10.1016/j.stem.2008.11.011.

Czermin, B. *et al.* (2002) 'Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites', *Cell*. Cell Press, 111(2), pp. 185–196. doi: 10.1016/S0092-8674(02)00975-3.

D'Amour, K. A. *et al.* (2005) 'Efficient differentiation of human embryonic stem cells to definitive endoderm', *Nature Biotechnology*. Nature Publishing Group, 23(12), pp. 1534–1541. doi: 10.1038/nbt1163.

D'Amour, K. a *et al.* (2006) 'Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells.', *Nature biotechnology*, 24(11), pp. 1392–401. doi: 10.1038/nbt1259.

Dai, C. *et al.* (2012) 'Islet-enriched gene expression and glucose-induced insulin secretion in human and mouse islets', *Diabetologia*. Diabetologia, 55(3), pp. 707–718. doi: 10.1007/s00125-011-2369-0.

Dambacher, S., De Almeida, G. P. and Schotta, G. (2013) 'Dynamic changes of the epigenetic landscape during cellular differentiation', *Epigenomics*. Epigenomics, pp. 701–713. doi: 10.2217/epi.13.67.

Davidovich, C. and Cech, T. R. (2015) 'The recruitment of chromatin modifiers by long noncoding RNAs: Lessons from PRC2', *RNA*. Cold Spring Harbor Laboratory Press, pp. 2007–2022. doi: 10.1261/rna.053918.115.

Dean, P. M. and Matthews, E. K. (1968) ‘Electrical activity in pancreatic islet cells’, *Nature*. Nature, pp. 389–390. doi: 10.1038/219389a0.

Dehé, P. M. *et al.* (2006) ‘Protein interactions within the Set1 complex and their roles in the regulation of histone 3 lysine 4 methylation’, *Journal of Biological Chemistry*. J Biol Chem, 281(46), pp. 35404–35412. doi: 10.1074/jbc.M603099200.

Denissov, S. *et al.* (2014) ‘Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant’, *Development (Cambridge)*. Oxford University Press for The Company of Biologists Limited, 141(3), pp. 526–537. doi: 10.1242/dev.102681.

Doss, M. X. and Sachinidis, A. (2019) ‘Current Challenges of iPSC-Based Disease Modeling and Therapeutic Implications’, *Cells*. MDPI AG, 8(5), p. 403. doi: 10.3390/cells8050403.

Dou, Y. *et al.* (2006) ‘Regulation of MLL1 H3K4 methyltransferase activity by its core components’, *Nature Structural and Molecular Biology*. Nat Struct Mol Biol, 13(8), pp. 713–719. doi: 10.1038/nsmb1128.

Douillet, D. *et al.* (2020) ‘Uncoupling histone H3K4 trimethylation from developmental gene expression via an equilibrium of COMPASS, Polycomb and DNA methylation’, *Nature Genetics*. Nature Research, 52(6), pp. 615–625. doi: 10.1038/s41588-020-0618-1.

Du, A. *et al.* (2009) ‘Islet-1 is required for the maturation, proliferation, and survival of the endocrine pancreas’, *Diabetes*. American Diabetes Association, 58(9), pp. 2059–2069. doi: 10.2337/db08-0987.

Dumasia, N. P. and Pethe, P. S. (2020) ‘Pancreas development and the Polycomb group protein complexes’, *Mechanisms of Development*. Elsevier Ireland Ltd, p. 103647. doi: 10.1016/j.mod.2020.103647.

Dunham, I. *et al.* (2012) ‘An integrated encyclopedia of DNA elements in the human genome’, *Nature*. Nature Publishing Group, 489(7414), pp. 57–74. doi: 10.1038/nature11247.

Eskeland, R. *et al.* (2010) ‘Ring1B Compacts Chromatin Structure and Represses Gene Expression Independent of Histone Ubiquitination’, *Molecular Cell*. NIH Public Access, 38(3), pp. 452–464. doi: 10.1016/j.molcel.2010.02.032.

Evans, M. J. and Kaufman, M. H. (1981) ‘Establishment in culture of pluripotential cells from

mouse embryos', *Nature*, 292(5819), pp. 154–156. doi: 10.1038/292154a0.

Fadista, J. *et al.* (2014) 'Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 111(38), pp. 13924–13929. doi: 10.1073/pnas.1402665111.

Fan, J. *et al.* (2015) 'Metabolic regulation of histone post-translational modifications', *ACS Chemical Biology*. American Chemical Society, pp. 95–108. doi: 10.1021/cb500846u.

Faust, C. *et al.* (1998) 'The Polycomb-group gene eed is required for normal morphogenetic movements during gastrulation in the mouse embryo', *Development*, 125(22), pp. 4495–4506. Available at: http://europepmc.org/article/MED/9778508 (Accessed: 11 November 2020).

Ferrai, C. *et al.* (2017) ' RNA polymerase II primes Polycomb-repressed developmental genes throughout terminal neuronal differentiation ', *Molecular Systems Biology*. EMBO, 13(10), p. 946. doi: 10.15252/msb.20177754.

Ferrari, K. J. *et al.* (2014) 'Polycomb-Dependent H3K27me1 and H3K27me2 Regulate Active Transcription and Enhancer Fidelity', *Molecular Cell*. Elsevier, 53(1), pp. 49–62. doi: 10.1016/j.molcel.2013.10.030.

Fischle, W. *et al.* (2003) 'Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by polycomb and HP1 chromodomains', *Genes and Development*. Cold Spring Harbor Laboratory Press, 17(15), pp. 1870–1881. doi: 10.1101/gad.1110503.

Fiskus, W. *et al.* (2009) 'Combined epigenetic therapy with the histone methyltransferase EZH2 inhibitor 3-deazaneplanocin A and the histone deacetylase inhibitor panobinostat against human AML cells', *Blood*. Blood, 114(13), pp. 2733–2743. doi: 10.1182/blood-2009-03-213496.

Frobel, J. *et al.* (2014) 'Epigenetic rejuvenation of mesenchymal stromal cells derived from induced pluripotent stem cells', *Stem Cell Reports*. Cell Press, 3(3), pp. 414–422. doi: 10.1016/j.stemcr.2014.07.003.

Fuks, F. *et al.* (2003) 'The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase', *Nucleic Acids Research*. Oxford University Press, 31(9), pp. 2305–2312. doi: 10.1093/nar/gkg332.

Fursova, N. A. *et al.* (2019) 'Synergy between Variant PRC1 Complexes Defines Polycomb-Mediated Gene Repression', *Molecular Cell*. Cell Press, 74(5), pp. 1020-1036.e8. doi: 10.1016/j.molcel.2019.03.024.

Gadd, M. S. *et al.* (2017) 'Structural basis of PROTAC cooperative recognition for selective protein degradation', *Nature Chemical Biology*. Nature Publishing Group, 13(5), pp. 514–521. doi: 10.1038/nchembio.2329.

Gao, T. *et al.* (2014) 'Pdx1 maintains β cell identity and function by repressing an α cell program', *Cell Metabolism*. NIH Public Access, 19(2), pp. 259–271. doi: 10.1016/j.cmet.2013.12.002.

Gao, Z. *et al.* (2012) 'PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes', *Molecular Cell*. Elsevier, 45(3), pp. 344–356. doi: 10.1016/j.molcel.2012.01.002.

Ge, S. X., Son, E. W. and Yao, R. (2018) 'iDEP: An integrated web application for differential expression and pathway analysis of RNA-Seq data', *BMC Bioinformatics*. BioMed Central Ltd., 19(1), pp. 1–24. doi: 10.1186/s12859-018-2486-6.

Gilbert, L. A. *et al.* (2013) 'XCRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes', *Cell*. Cell Press, 154(2), p. 442. doi: 10.1016/j.cell.2013.06.044.

De Gobbi, M. *et al.* (2011) 'Generation of bivalent chromatin domains during cell fate decisions', *Epigenetics and Chromatin*. BioMed Central, p. 9. doi: 10.1186/1756-8935-4-9.

Gouon-Evans, V. *et al.* (2006) 'BMP-4 is required for hepatic specification of mouse embryonic stem cell-derived definitive endoderm', *Nature Biotechnology*. Nature Publishing Group, 24(11), pp. 1402–1411. doi: 10.1038/nbt1258.

Gradwohl, G. *et al.* (2000) 'neurogenin3 is required for the development of the four endocrine cell lineages of the pancreas', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 97(4), pp. 1607–1611. doi: 10.1073/pnas.97.4.1607.

Groner, A. C. *et al.* (2010) 'KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading', *PLoS Genetics*. PLoS Genet,

6(3). doi: 10.1371/journal.pgen.1000869.

Grzybowski, A. . *et al.* (2019) 'Native internally calibrated chromatin immunoprecipitation for quantitative studies of histone post-translational modifications', *Nature protocols*. Nat Protoc, 14(12), pp. 3275–3302. doi: 10.1038/S41596-019-0218-7.

Gu, C. *et al.* (2010) 'Pancreatic β Cells Require NeuroD to Achieve and Maintain Functional Maturity', *Cell Metabolism*. Cell Metab, 11(4), pp. 298–310. doi: 10.1016/j.cmet.2010.03.006.

Guenther, M. G. *et al.* (2007) 'A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells', *Cell*. Cell, 130(1), pp. 77–88. doi: 10.1016/j.cell.2007.05.042.

Guo, T. and Hebrok, M. (2009) 'Stem cells to pancreatic β-cells: New sources for diabetes cell therapy', *Endocrine Reviews*. The Endocrine Society, pp. 214–227. doi: 10.1210/er.2009-0004.

Gurdon, J. B. (1962) 'The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles.', *Journal of embryology and experimental morphology*, 10(4), pp. 622–40. doi: 10.1002/stem.684.

Hainer, S. J. *et al.* (2019) 'Profiling of Pluripotency Factors in Single Cells and Early Embryos', *Cell*. Cell Press, 177(5), pp. 1319-1329.e11. doi: 10.1016/j.cell.2019.03.014.

Hang, Y. and Stein, R. (2011) 'MafA and MafB activity in pancreatic β cells', *Trends in Endocrinology and Metabolism*. NIH Public Access, pp. 364–373. doi: 10.1016/j.tem.2011.05.003.

Hansen, K. H. *et al.* (2008) 'A model for transmission of the H3K27me3 epigenetic mark', *Nature Cell Biology*. Nature Publishing Group, 10(11), pp. 1291–1300. doi: 10.1038/ncb1787.

Harikumar, A. and Meshorer, E. (2015) 'Chromatin remodeling and bivalent histone modifications in embryonic stem cells', *EMBO reports*. EMBO, 16(12), pp. 1609–1619. doi: 10.15252/embr.201541011.

Harrison, K. A. *et al.* (1999) 'Pancreas dorsal lobe agenesis and abnormal islets of Langerhans in Hlxb9-deficient mice', *Nature Genetics*, 23(1), pp. 71–75. doi: 10.1038/12674.

Harshman, S. W. *et al.* (2013) 'H1 histones: Current perspectives and challenges', *Nucleic Acids Research*. Nucleic Acids Res, pp. 9593–9609. doi: 10.1093/nar/gkt700.

Hauri, S. *et al.* (2016) 'A High-Density Map for Navigating the Human Polycomb Complexome', *Cell Reports*. Elsevier B.V., 17(2), pp. 583–595. doi: 10.1016/j.celrep.2016.08.096.

Hawkins, R. D. *et al.* (2010) 'Distinct epigenomic landscapes of pluripotent and lineage-committed human cells', *Cell Stem Cell*. NIH Public Access, 6(5), pp. 479–491. doi: 10.1016/j.stem.2010.03.018.

Hayashi, K. I. *et al.* (2012) 'Rational design of an auxin antagonist of the SCF TIR1 auxin receptor complex', *ACS Chemical Biology*. ACS Chem Biol, 7(3), pp. 590–598. doi: 10.1021/cb200404c.

Hebrok, M., Kim, S. K. and Melton, D. A. (1998) 'Notochord repression of endodermal sonic hedgehog permits pancreas development', *Genes and Development*. Cold Spring Harbor Laboratory Press, 12(11), pp. 1705–1713. doi: 10.1101/gad.12.11.1705.

Heintzman, N. D. *et al.* (2007) 'Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome', *Nature Genetics*. Nat Genet, 39(3), pp. 311–318. doi: 10.1038/ng1966.

Henseleit, K. D. *et al.* (2005) 'NKX6 transcription factor activity is required for α- and β-cell development in the pancreas', *Development*, 132(13), pp. 3139–3149. doi: 10.1242/dev.01875.

Hödl, M. and Basler, K. (2012) 'Transcription in the absence of histone H3.2 and H3K4 methylation', *Current Biology*. Cell Press, 22(23), pp. 2253–2257. doi: 10.1016/j.cub.2012.10.008.

Højfeldt, J. W. *et al.* (2018) 'Accurate H3K27 methylation can be established de novo by SUZ12-directed PRC2', *Nature Structural and Molecular Biology*. Nature Publishing Group, 25(3), pp. 225–232. doi: 10.1038/s41594-018-0036-6.

Hong, S. H. *et al.* (2007) 'Identification of JmjC domain-containing UTX and JMJD3 as histone H3 lysine 27 demethylases', *Proceedings of the National Academy of Sciences of the United States of America*. Proc Natl Acad Sci U S A, 104(47), pp. 18439–18444. doi: 10.1073/pnas.0707292104.

Horvath, S. (2013) 'DNA methylation age of human tissues and cell types', *Genome Biology*. BioMed Central, 14(10), p. R115. doi: 10.1186/gb-2013-14-10-r115.

Howe, F. S. *et al.* (2017) 'Is H3K4me3 instructive for transcription activation?', *BioEssays*. John Wiley and Sons Inc., 39(1), pp. 1–12. doi: 10.1002/bies.201600095.

Hrvatin, Sinisa *et al.* (2014) 'Differentiated human stem cells resemble fetal, not adult, β cells.', *Proceedings of the National Academy of Sciences of the United States of America*, 111(8), pp. 3038–43. doi: 10.1073/pnas.1400709111.

Hrvatin, Siniša *et al.* (2014) 'Differentiated human stem cells resemble fetal, not adult, β cells', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 111(8), pp. 3038–3043. doi: 10.1073/pnas.1400709111.

Hsu, J. H. R. *et al.* (2020) 'EED-Targeted PROTACs Degrade EED, EZH2, and SUZ12 in the PRC2 Complex', *Cell Chemical Biology*. Elsevier Ltd, 27(1), pp. 41-46.e17. doi: 10.1016/j.chembiol.2019.11.004.

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

Hu, D. *et al.* (2013) 'The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells', *Nature Structural and Molecular Biology*. NIH Public Access, 20(9), pp. 1093–1097. doi: 10.1038/nsmb.2653.

Hunkapiller, J. *et al.* (2012) 'Polycomb-like 3 promotes polycomb repressive complex 2 binding to CpG islands and embryonic stem cell self-renewal', *PLoS Genetics*. Public Library of Science, 8(3), p. 1002576. doi: 10.1371/journal.pgen.1002576.

Hyun, K. *et al.* (2017) 'Writing, erasing and reading histone lysine methylations', *Experimental and Molecular Medicine*. Nature Publishing Group, p. 324. doi: 10.1038/emm.2017.11.

Inagaki, N., Gonoi, T. and Seino, S. (1997) 'Subunit stoichiometry of the pancreatic β-cell ATP-sensitive K+ channel', *FEBS Letters*. FEBS Lett, 409(2), pp. 232–236. doi: 10.1016/S0014-5793(97)00488-2.

van Ingen, H. *et al.* (2008) 'Structural Insight into the Recognition of the H3K4me3 Mark by the TFIID Subunit TAF3', *Structure*. Elsevier, 16(8), pp. 1245–1256. doi: 10.1016/j.str.2008.04.015.

Iwamoto, M. *et al.* (2010) 'A general chemical method to regulate protein stability in the mammalian central nervous system', *Chemistry and Biology*. Cell Press, 17(9), pp. 981–988.

doi: 10.1016/j.chembiol.2010.07.009.

Iwase, S. *et al.* (2007) 'The X-Linked Mental Retardation Gene SMCX/JARID1C Defines a Family of Histone H3 Lysine 4 Demethylases', *Cell*. Cell Press, 128(6), pp. 1077–1088. doi: 10.1016/j.cell.2007.02.017.

Jamieson, K. *et al.* (2013) 'Regional control of histone H3 lysine 27 methylation in Neurospora', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 110(15), pp. 6027–6032. doi: 10.1073/pnas.1303750110.

Jennings, R. E. *et al.* (2013) 'Development of the human pancreas from foregut to endocrine commitment', *Diabetes*. American Diabetes Association, 62(10), pp. 3514–3522. doi: 10.2337/db12-1479.

Jennings, R. E. *et al.* (2017) 'Laser Capture and Deep Sequencing Reveals the Transcriptomic Programmes Regulating the Onset of Pancreas and Liver Differentiation in Human Embryos', *Stem Cell Reports*. ElsevierCompany., 9(5), pp. 1387–1394. doi: 10.1016/j.stemcr.2017.09.018.

Jeon, J. *et al.* (2009) 'Endocrine cell clustering during human pancreas development', *Journal of Histochemistry and Cytochemistry*. J Histochem Cytochem, 57(9), pp. 811–824. doi: 10.1369/jhc.2009.953307.

Jermann, P. *et al.* (2014) 'Short sequences can efficiently recruit histone H3 lysine 27 trimethylation in the absence of enhancer activity and DNA methylation', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 111(33), p. E3415. doi: 10.1073/pnas.1400672111.

Jiang, J. *et al.* (2007) 'Generation of Insulin-Producing Islet-Like Clusters from Human Embryonic Stem Cells', *STEM CELLS*. Wiley, 25(8), pp. 1940–1953. doi: 10.1634/stemcells.2006-0761.

Johansson, K. A. *et al.* (2007) 'Temporal Control of Neurogenin3 Activity in Pancreas Progenitors Reveals Competence Windows for the Generation of Different Endocrine Cell Types', *Developmental Cell*, 12(3), pp. 457–465. doi: 10.1016/j.devcel.2007.02.010.

Jonsson, J. *et al.* (1994) 'Insulin-promoter-factor 1 is required for pancreas development in mice', *Nature*, 371(6498), pp. 606–609. doi: 10.1038/371606a0.

Justin, N. *et al.* (2016) 'Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2', *Nature communications*. Nature Publishing Group, 7, p. 11316. doi: 10.1038/ncomms11316.

Kalb, R. *et al.* (2014) 'Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression', *Nature Structural and Molecular Biology*. Nature Publishing Group, 21(6), pp. 569–571. doi: 10.1038/nsmb.2833.

Kanhere, A. *et al.* (2010) 'Short RNAs Are Transcribed from Repressed Polycomb Target Genes and Interact with Polycomb Repressive Complex-2', *Molecular Cell*. NIH Public Access, 38(5), pp. 675–688. doi: 10.1016/j.molcel.2010.03.019.

Kasinath, V. *et al.* (2018) 'Structures of human PRC2 with its cofactors AEBP2 and JARID2', *Science*. American Association for the Advancement of Science, 359(6378), pp. 940–944. doi: 10.1126/science.aar5700.

Kaya-Okur, H. S. *et al.* (2019) 'CUT&Tag for efficient epigenomic profiling of small samples and single cells', *Nature Communications*. Nature Publishing Group, 10(1). doi: 10.1038/s41467-019-09982-5.

Kim, H., Kang, K. and Kim, J. (2009) 'AEBP2 as a potential targeting protein for Polycomb Repression Complex PRC2', *Nucleic Acids Research*. Oxford University Press, 37(9), p. 2940. doi: 10.1093/NAR/GKP149.

Kim, K. *et al.* (2010) 'Epigenetic memory in induced pluripotent stem cells', *Nature*. Nature Publishing Group, 467(7313), pp. 285–290. doi: 10.1038/nature09342.

Kim, K. *et al.* (2011) 'Donor cell type can influence the epigenome and differentiation potential of human induced pluripotent stem cells', *Nature Biotechnology*. Nat Biotechnol, 29(12), pp. 1117–1119. doi: 10.1038/nbt.2052.

Kim, K. H. and Roberts, C. W. M. (2016) 'Targeting EZH2 in cancer', *Nature Medicine*. Nature Publishing Group, pp. 128–134. doi: 10.1038/nm.4036.

Kinkley, S. *et al.* (2016) 'ReChIP-seq reveals widespread bivalency of H3K4me3 and H3K27me3 in CD4+ memory T cells', *Nature Communications*. Nature Publishing Group, 7. doi: 10.1038/ncomms12514.

Kloet, S. L. *et al.* (2016) 'The dynamic interactome and genomic targets of Polycomb complexes during stem-cell differentiation', *Nature Structural & Molecular Biology*. Nature Publishing Group, 23(7), pp. 682–690. doi: 10.1038/nsmb.3248.

Konze, K. D. *et al.* (2013) 'An orally bioavailable chemical probe of the lysine methyltransferases EZH2 and EZH1', *ACS Chemical Biology*, 8(6), pp. 1324–1334. doi: 10.1021/cb400133j.

de Krijger, R. R. *et al.* (1992) 'The midgestational human fetal pancreas contains cells coexpressing islet hormones', *Developmental Biology*. Dev Biol, 153(2), pp. 368–375. doi: 10.1016/0012-1606(92)90121-V.

Krogan, N. J. *et al.* (2002) 'COMPASS, a histone H3 (lysine 4) methyltransferase required for telomeric silencing of gene expression', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology, 277(13), pp. 10753–10755. doi: 10.1074/jbc.C200023200.

Kroon, E., Martinson, Laura A, *et al.* (2008) 'Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells in vivo.', *Nat Biotechnol*, 26(4), pp. 443–452. doi: 10.1038/nbt1393.

Kroon, E., Martinson, Laura A., *et al.* (2008) 'Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells in vivo', *Nature Biotechnology*. Nat Biotechnol, 26(4), pp. 443–452. doi: 10.1038/nbt1393.

Ku, M. *et al.* (2008) 'Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains', *PLoS Genetics*. Public Library of Science, 4(10). doi: 10.1371/journal.pgen.1000242.

Kuleshov, M. V. *et al.* (2016) 'Enrichr: a comprehensive gene set enrichment analysis web server 2016 update', *Nucleic acids research*. Nucleic Acids Res, 44(W1), pp. W90–W97. doi: 10.1093/nar/gkw377.

Kuzmichev, A. *et al.* (2002) 'Histone methyltransferase activity associated with a human multiprotein complex containing the enhancer of zeste protein', *Genes and Development*. Cold Spring Harbor Laboratory Press, 16(22), pp. 2893–2905. doi: 10.1101/gad.1035902.

Lan, F. *et al.* (2007) 'A histone H3 lysine 27 demethylase regulates animal posterior

development', *Nature*. Nature Publishing Group, 449(7163), pp. 689–694. doi: 10.1038/nature06192.

Laprell, F., Finkl, K. and Müller, J. (2017) 'Propagation of Polycomb-repressed chromatin requires sequence-specific recruitment to DNA', *Science*. American Association for the Advancement of Science, 356(6333), pp. 85–88. doi: 10.1126/science.aai8266.

Lauberth, S. M. *et al.* (2013) 'H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation', *Cell*. Elsevier B.V., 152(5), pp. 1021–1036. doi: 10.1016/j.cell.2013.01.052.

Laugesen, A. and Helin, K. (2014) 'Chromatin repressive complexes in stem cells, development, and cancer', *Cell Stem Cell*. Cell Press, pp. 735–751. doi: 10.1016/j.stem.2014.05.006.

Lavarone, E., Barbieri, C. M. and Pasini, D. (2019) 'Dissecting the role of H3K27 acetylation and methylation in PRC2 mediated control of cellular identity', *Nature Communications*. Nature Publishing Group, 10(1). doi: 10.1038/s41467-019-09624-w.

Lawrence, M., Daujat, S. and Schneider, R. (2016) 'Lateral Thinking: How Histone Modifications Regulate Gene Expression', *Trends in Genetics*. Elsevier Ltd, pp. 42–56. doi: 10.1016/j.tig.2015.10.007.

Lee, C. H. *et al.* (2018) 'Distinct Stimulatory Mechanisms Regulate the Catalytic Activity of Polycomb Repressive Complex 2', *Molecular Cell*. Cell Press, 70(3), pp. 435-448.e5. doi: 10.1016/j.molcel.2018.03.019.

Lee, H. G. *et al.* (2015) 'Genome-wide activities of Polycomb complexes control pervasive transcription', *Genome Research*. Cold Spring Harbor Laboratory Press, 25(8), pp. 1170–1181. doi: 10.1101/gr.188920.114.

Lee, J. H. *et al.* (2007) 'Identification and characterization of the human Set1B histone H3-Lys 4 methyltransferase complex', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology, 282(18), pp. 13419–13428. doi: 10.1074/jbc.M609809200.

Lee, J. H. and Skalnik, D. G. (2005) 'CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex', *Journal of Biological Chemistry*. American Society for

Biochemistry and Molecular Biology, 280(50), pp. 41725–41731. doi: 10.1074/jbc.M508312200.

Lee, M. G. *et al.* (2005) 'An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation', *Nature*. Nature Publishing Group, 437(7057), pp. 432–435. doi: 10.1038/nature04021.

Lee, T. I. *et al.* (2006) 'Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells', *Cell*. Cell, 125(2), pp. 301–313. doi: 10.1016/j.cell.2006.02.043.

Leeb, M. *et al.* (2010) 'Polycomb complexes act redundantly to repress genomic repeats and genes', *Genes and Development*. Cold Spring Harbor Laboratory Press, 24(3), pp. 265–276. doi: 10.1101/gad.544410.

Lehnertz, B. *et al.* (2003) 'Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin', *Current Biology*. Cell Press, 13(14), pp. 1192–1200. doi: 10.1016/S0960-9822(03)00432-9.

Leibiger, I. B., Leibiger, B. and Berggren, P.-O. (2008) 'Insulin Signaling in the Pancreatic β-Cell', *Annual Review of Nutrition*. Annual Reviews, 28(1), pp. 233–251. doi: 10.1146/annurev.nutr.28.061807.155530.

Leonard, J. *et al.* (1993) 'Characterization of somatostatin transactivating factor-1, a novel homeobox factor that stimulates somatostatin expression in pancreatic islet cells.', *Molecular Endocrinology*. Oxford Academic, 7(10), pp. 1275–1283. doi: 10.1210/mend.7.10.7505393.

Lewis, E. B. (1947) 'New mutants report', *Drosophila Information Service*, 21. Available at: https://scholar.google.com/scholar_lookup?journal=Drosoph.+Inf.+Serv.&title=New+mutants:+Reports+of+P.+lewis&author=E.B.+Lewis&volume=21&publication_year=1947&pages=69& (Accessed: 22 December 2020).

Lewis, E. B. (1978) 'A gene complex controlling segmentation in Drosophila', *Nature*. Nature, pp. 565–570. doi: 10.1038/276565a0.

Li, H. *et al.* (2017) 'Polycomb-like proteins link the PRC2 complex to CpG islands', *Nature*. Nature Publishing Group, 549(7671), pp. 287–291. doi: 10.1038/nature23881.

Li, S. *et al.* (2019) 'An efficient auxin-inducible degron system with low basal degradation in human cells', *Nature Methods*. Nature Publishing Group, 16(9), pp. 866–869. doi:

10.1038/s41592-019-0512-x.

Lima, M. J. *et al.* (2012) 'Pancreatic transcription factors containing protein transduction domains drive mouse embryonic stem cells towards endocrine pancreas', *PLoS ONE*. PLoS One, 7(5). doi: 10.1371/journal.pone.0036481.

Lister, R. *et al.* (2011) 'Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells', *Nature*. Nature, 471(7336), pp. 68–73. doi: 10.1038/nature09798.

Liu, C. *et al.* (2020) 'Predator: A novel method for targeted protein degradation', *BioRxiv*. Cold Spring Harbor Laboratory, p. 2020.07.31.231787. doi: 10.1101/2020.07.31.231787.

Liu, X. *et al.* (2017) 'The immunogenicity and immune tolerance of pluripotent stem cell derivatives', *Frontiers in Immunology*. Frontiers Media S.A., p. 645. doi: 10.3389/fimmu.2017.00645.

Lombardo, A. *et al.* (2007) 'Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery', *Nature Biotechnology*, 25(11), pp. 1298–1306. doi: 10.1038/nbt1353.

Long, Y. *et al.* (2020) 'RNA is essential for PRC2 chromatin occupancy and function in human pluripotent stem cells', *Nature Genetics*. Nature Research, 52(9), pp. 931–938. doi: 10.1038/s41588-020-0662-x.

Lu, T. T. H. *et al.* (2018) 'The Polycomb-Dependent Epigenome Controls β Cell Dysfunction, Dedifferentiation, and Diabetes', *Cell Metabolism*. Cell Press, 27(6), pp. 1294-1308.e7. doi: 10.1016/j.cmet.2018.04.013.

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

Lynch, M. D. *et al.* (2012) 'An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment', *EMBO Journal*. European Molecular Biology Organization, 31(2), pp. 317–329. doi: 10.1038/emboj.2011.399.

Lyttle, B. M. *et al.* (2008) 'Transcription factor expression in the developing human fetal endocrine pancreas', *Diabetologia*. Springer, 51(7), pp. 1169–1180. doi: 10.1007/s00125-008-1006-z.

Malecki, M. T. *et al.* (1999) 'Mutations in NEUROD1 are associated with the development of type 2 diabetes mellitus', *Nature Genetics*. Nature Publishing Group, 23(3), pp. 323–328. doi: 10.1038/15500.

Mandegar, M. A. *et al.* (2016) 'CRISPR Interference Efficiently Induces Specific and Reversible Gene Silencing in Human iPSCs', *Cell Stem Cell*. Cell Press, 18(4), pp. 541–553. doi: 10.1016/j.stem.2016.01.022.

Margueron, R. *et al.* (2008) 'Ezh1 and Ezh2 Maintain Repressive Chromatin through Different Mechanisms', *Molecular Cell*. Cell Press, 32(4), pp. 503–518. doi: 10.1016/j.molcel.2008.11.004.

Margueron, R. *et al.* (2009) 'Role of the polycomb protein EED in the propagation of repressive histone marks', *Nature*. Nature Publishing Group, 461(7265), pp. 762–767. doi: 10.1038/nature08398.

Margueron, R. and Reinberg, D. (2010) 'Chromatin structure and the inheritance of epigenetic information', *Nature Reviews Genetics*. Nature Publishing Group, 11(4), pp. 285–296. doi: 10.1038/nrg2752.

Margueron, R. and Reinberg, D. (2011) 'The Polycomb complex PRC2 and its mark in life', *Nature*. NIH Public Access, pp. 343–349. doi: 10.1038/nature09784.

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

Mas, G. *et al.* (2018) 'Promoter bivalency favors an open chromatin architecture in embryonic stem cells', *Nature Genetics*. Nature Publishing Group, 50(10), pp. 1452–1462. doi: 10.1038/s41588-018-0218-5.

Matsuoka, T. *et al.* (2003) 'Members of the Large Maf Transcription Family Regulate Insulin Gene Transcription in Islet β Cells', *Molecular and Cellular Biology*. American Society for Microbiology, 23(17), pp. 6049–6062. doi: 10.1128/mcb.23.17.6049-6062.2003.

Maury, Y. *et al.* (2012) 'Human pluripotent stem cells for disease modelling and drug screening',

BioEssays. John Wiley & Sons, Ltd, 34(1), pp. 61–71. doi: 10.1002/bies.201100071.

McCabe, M. T. *et al.* (2012) 'EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations', *Nature*. Nature, pp. 108–112. doi: 10.1038/nature11606.

MCCULLOCH, E. A. and TILL, J. E. (1960) 'The radiation sensitivity of normal mouse bone marrow cells, determined by quantitative marrow transplantation into irradiated mice.', *Radiation research*, 13, pp. 115–125. doi: 10.2307/3570877.

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

Mendenhall, E. M. *et al.* (2010) 'GC-rich sequence elements recruit PRC2 in mammalian ES cells', *PLoS Genetics*. Public Library of Science, 6(12), pp. 1–10. doi: 10.1371/journal.pgen.1001244.

van Mierlo, Guido *et al.* (2019) 'Integrative Proteomic Profiling Reveals PRC2-Dependent Epigenetic Crosstalk Maintains Ground-State Pluripotency', *Cell Stem Cell*. Cell Press, 24(1), pp. 123-137.e8. doi: 10.1016/j.stem.2018.10.017.

van Mierlo, G. *et al.* (2019) 'The Complexity of PRC2 Subcomplexes', *Trends in Cell Biology*. Elsevier Ltd, pp. 660–671. doi: 10.1016/j.tcb.2019.05.004.

Miguel-Escalada, I. *et al.* (2019) 'Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes', *Nature Genetics*. Nature Research, 51(7), pp. 1137–1148. doi: 10.1038/s41588-019-0457-0.

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

Miller, T. *et al.* (2001) 'COMPASS: A complex of proteins associated with a trithorax-related SET domain protein', *Proceedings of the National Academy of Sciences of the United States of America*. First Published, 98(23), pp. 12902–12907. doi: 10.1073/pnas.231473398.

Milne, T. A. *et al.* (2002) 'MLL targets SET domain methyltransferase activity to Hox gene promoters', *Molecular Cell*. Cell Press, 10(5), pp. 1107–1117. doi: 10.1016/S1097-2765(02)00741-4.

Min, G. L. *et al.* (2007) 'Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination', *Science*. American Association for the Advancement of Science, 318(5849), pp. 447–450. doi: 10.1126/science.1149042.

Min, J., Zhang, Y. and Xu, R. M. (2003) 'Structural basis for specific binding of polycomb chromodomain to histone H3 methylated at Lys 27', *Genes and Development*. Genes Dev, 17(15), pp. 1823–1828. doi: 10.1101/gad.269603.

Miyaoka, Y. *et al.* (2014) 'Isolation of single-base genome-edited human iPS cells without antibiotic selection', *Nature Methods*. Nature Publishing Group, 11(3), pp. 291–293. doi: 10.1038/nmeth.2840.

Mobasseri, M. *et al.* (2020) 'Prevalence and incidence of type 1 diabetes in the world: A systematic review and meta-analysis', *Health Promotion Perspectives*. Tabriz University of Medical Sciences, pp. 98–115. doi: 10.34172/hpp.2020.18.

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

Mohn, F. and Schübeler, D. (2009) 'Genetics and epigenetics: stability and plasticity during cellular differentiation', *Trends in Genetics*. Elsevier Current Trends, pp. 129–136. doi: 10.1016/j.tig.2008.12.005.

Moody, J. D. *et al.* (2017) 'First critical repressive H3K27me3 marks in embryonic stem cells identified using designed protein inhibitor', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 114(38), pp. 10125–10130. doi: 10.1073/pnas.1706907114.

Müller, J. (1995) 'Transcriptional silencing by the polycomb protein in Drosophila embryos', *EMBO Journal*. Wiley-VCH Verlag, 14(6), pp. 1209–1220. doi: 10.1002/j.1460-2075.1995.tb07104.x.

Müller, J. *et al.* (2002) 'Histone methyltransferase activity of a Drosophila Polycomb group repressor complex', *Cell*. Cell Press, 111(2), pp. 197–208. doi: 10.1016/S0092-8674(02)00976-5.

Murtaugh, L. C. and Melton, D. A. (2003) 'Genes, Signals, and Lineages in Pancreas Development', *Annual Review of Cell and Developmental Biology*. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA , 19(1), pp. 71–89. doi: 10.1146/annurev.cellbio.19.111301.144752.

Murzina, N. V. *et al.* (2008) 'Structural Basis for the Recognition of Histone H4 by the Histone-Chaperone RbAp46', *Structure*. Cell Press, 16(7), pp. 1077–1085. doi: 10.1016/j.str.2008.05.006.

Nabet, B. *et al.* (2018) 'The dTAG system for immediate and target-specific protein degradation', *Nature Chemical Biology*. Nature Publishing Group, 14(5), pp. 431–441. doi: 10.1038/s41589-018-0021-8.

Nair, G. G. *et al.* (2019) 'Recapitulating endocrine cell clustering in culture promotes maturation of human stem-cell-derived β cells', *Nature Cell Biology*. Nature Publishing Group, 21(2), pp. 263–274. doi: 10.1038/s41556-018-0271-4.

Natsume, T. *et al.* (2016) 'Rapid Protein Depletion in Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors', *Cell Reports*. The Authors, 15(1), pp. 210–218. doi: 10.1016/j.celrep.2016.03.001.

Naya, F. J. *et al.* (1997) 'Diabetes, defective pancreatic morphogenesis, and abnormal enteroendocrine differentiation in BETA2/NeuroD-deficient mice', *Genes and Development*. Cold Spring Harbor Laboratory Press, 11(18), pp. 2323–2334. doi: 10.1101/gad.11.18.2323.

Nekrasov, M., Wild, B. and Müller, J. (2005) 'Nucleosome binding and histone methyltransferase activity of Drosophila PRC2', *EMBO Reports*. European Molecular Biology Organization, 6(4), pp. 348–353. doi: 10.1038/sj.embor.7400376.

Ng, J. *et al.* (2000) 'A Drosophila ESC-E(Z) Protein Complex Is Distinct from Other Polycomb Group Complexes and Contains Covalently Modified ESC', *Molecular and Cellular Biology*. American Society for Microbiology, 20(9), pp. 3069–3078. doi: 10.1128/mcb.20.9.3069-3078.2000.

Nikolaou, N. *et al.* (2016) 'Optimizing human hepatocyte models for metabolic phenotype and function: Effects of treatment with dimethyl sulfoxide (DMSO)', *Physiological Reports*. American Physiological Society, 4(21). doi: 10.14814/phy2.12944.

Nishimura, K. *et al.* (2009) 'An auxin-based degron system for the rapid depletion of proteins in nonplant cells', *Nature Methods*. Nature Publishing Group, 6(12), pp. 917–922. doi: 10.1038/nmeth.1401.

Nolden, L. *et al.* (2006) 'Site-specific recombination in human embryonic stem cells induced by cell-permeant Cre recombinase', *Nature Methods*. Nat Methods, 3(6), pp. 461–467. doi: 10.1038/nmeth884.

Nolden, L. *et al.* (2007) 'Stem cell engineering using transducible Cre recombinase.', *Methods in molecular medicine*. Methods Mol Med, 140, pp. 17–32. doi: 10.1007/978-1-59745-443-8_2.

Nora, E. P. *et al.* (2017) 'Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization', *Cell*, 169(5), pp. 930-944.e22. doi: 10.1016/j.cell.2017.05.004.

O'Carroll, D. *et al.* (2001) 'The polycomb-group gene Ezh2 is required for early mouse development.', *Molecular and Cellular Biology*, 21(13), pp. 4330–4336. doi: 10.1128/MCB.21.13.4330.

O'Geen, H. *et al.* (2017) 'DCas9-based epigenome editing suggests acquisition of histone methylation is not sufficient for target gene repression', *Nucleic Acids Research*. Oxford University Press, 45(17), pp. 9901–9916. doi: 10.1093/nar/gkx578.

Offield, M. *et al.* (1996) *PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum*, *dev.biologists.org*. Available at: https://dev.biologists.org/content/122/3/983.short (Accessed: 26 October 2020).

Oksuz, O. *et al.* (2018) 'Capturing the Onset of PRC2-Mediated Repressive Domain Formation', *Molecular Cell*. Cell Press, 70(6), pp. 1149-1162.e5. doi: 10.1016/j.molcel.2018.05.023.

Oliviero, G. *et al.* (2016) 'Dynamic Protein Interactions of the Polycomb Repressive Complex 2 during Differentiation of Pluripotent Cells.', *Molecular & cellular proteomics : MCP*, 15(11), pp. 3450–3460. doi: 10.1074/mcp.M116.062240.

Otonkoski, T. *et al.* (1988) 'Maturation of insulin response to glucose during human fetal and neonatal development. Studies with perifusion of pancreatic isletlike cell clusters', *Diabetes*. American Diabetes Association, 37(3), pp. 286–291. doi: 10.2337/diab.37.3.286.

Pagliuca, F. W. *et al.* (2014) 'Generation of functional human pancreatic β cells in vitro', *Cell*. Cell Press, 159(2), pp. 428–439. doi: 10.1016/j.cell.2014.09.040.

Palazzani, L. (2011) 'Embryo research in Italy: The bioethical and biojuridical debate', *Human Reproduction and Genetic Ethics*. Routledge, pp. 28–39. doi: 10.1558/hrge.v17i1.28.

Pan, F. C. and Wright, C. (2011) 'Pancreas organogenesis: From bud to plexus to gland', *Developmental Dynamics*, 240(3), pp. 530–565. doi: 10.1002/dvdy.22584.

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

Papapetrou, E. P. and Schambach, A. (2016) 'Gene insertion into genomic safe harbors for human gene therapy', *Molecular Therapy*. Nature Publishing Group, pp. 678–684. doi: 10.1038/mt.2016.38.

Pasini, D. *et al.* (2004) 'Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity', *The EMBO Journal*, 23(20), pp. 4061–4071. doi: 10.1038/sj.emboj.7600402.

Pasini, D. *et al.* (2007) 'The Polycomb Group Protein Suz12 Is Required for Embryonic Stem Cell Differentiation', *Molecular and Cellular Biology*. American Society for Microbiology, 27(10), pp. 3769–3779. doi: 10.1128/mcb.01432-06.

Pasini, D. *et al.* (2010) 'JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells', *Nature 2010 464:7286*. Nature Publishing Group, 464(7286), pp. 306–310. doi: 10.1038/nature08788.

Pauler, F. M. *et al.* (2009) 'H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome', *Genome Research*. Cold Spring Harbor Laboratory Press, 19(2), pp. 221–233. doi: 10.1101/gr.080861.108.

Peng, J. C. *et al.* (2009) 'Jarid2/Jumonji Coordinates Control of PRC2 Enzymatic Activity and Target Gene Occupancy in Pluripotent Cells', *Cell*. Cell Press, 139(7), pp. 1290–1302. doi: 10.1016/j.cell.2009.12.002.

Pengelly, A. R. *et al.* (2013) 'A histone mutant reproduces the phenotype caused by loss of histone-modifying factor polycomb', *Science*. American Association for the Advancement of Science, 339(6120), pp. 698–699. doi: 10.1126/science.1231382.

Pengelly, A. R. *et al.* (2015) 'Transcriptional repression by PRC1 in the absence of H2A monoubiquitylation', *Genes and Development*. Cold Spring Harbor Laboratory Press, 29(14), pp. 1487–1492. doi: 10.1101/gad.265439.115.

Perino, M. *et al.* (2020) 'Stem Cell Reports Article Two Functional Axes of Feedback-Enforced PRC2 Recruitment in Mouse Embryonic Stem Cells'. doi: 10.1016/j.stemcr.2020.07.007.

Peters, A. H. F. M. *et al.* (2003) 'Partitioning and Plasticity of Repressive Histone Methylation States in Mammalian Chromatin', *Molecular Cell*. Cell Press, 12(6), pp. 1577–1589. doi: 10.1016/S1097-2765(03)00477-5.

Petersen, M. B. K. *et al.* (2018) 'Recapitulating and Deciphering Human Pancreas Development From Human Pluripotent Stem Cells in a Dish', in *Current Topics in Developmental Biology*. Academic Press Inc., pp. 143–190. doi: 10.1016/bs.ctdb.2018.02.009.

Pethe, P., Nagvenkar, P. and Bhartiya, D. (2014) 'Polycomb group protein expression during differentiation of human embryonic stem cells into pancreatic lineage in vitro.', *BMC cell biology*, 15(1), p. 18. doi: 10.1186/1471-2121-15-18.

Petruk, S., Cai, J., *et al.* (2017) 'Delayed Accumulation of H3K27me3 on Nascent DNA Is Essential for Recruitment of Transcription Factors at Early Stages of Stem Cell Differentiation', *Molecular Cell*. Cell Press, 66(2), pp. 247-257.e5. doi: 10.1016/j.molcel.2017.03.006.

Petruk, S., Mariani, S. A., *et al.* (2017) 'Structure of Nascent Chromatin Is Essential for Hematopoietic Lineage Specification', *Cell Reports*. Elsevier B.V., 19(2), pp. 295–306. doi: 10.1016/j.celrep.2017.03.035.

Piper, K. *et al.* (2004) 'Beta cell differentiation during early human pancreas development', *Journal of Endocrinology*. J Endocrinol, 181(1), pp. 11–23. doi: 10.1677/joe.0.1810011.

Plath, K. *et al.* (2003) 'Role of histone H3 lysine 27 methylation in X inactivation', *Science*. Science, 300(5616), pp. 131–135. doi: 10.1126/science.1084274.

Poepsel, S., Kasinath, V. and Nogales, E. (2018) 'Cryo-EM structures of PRC2 simultaneously

engaged with two functionally distinct nucleosomes', *Nature Structural and Molecular Biology*. Nature Publishing Group, 25(2), pp. 154–162. doi: 10.1038/s41594-018-0023-y.

Polak, M. *et al.* (2000) *Early Pattern of Differentiation in the Human Pancreas*, *DIABETES*.

Polo, J. M. *et al.* (2010) 'Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells', *Nature Biotechnology*. NIH Public Access, 28(8), pp. 848–855. doi: 10.1038/nbt.1667.

Portoso, M. *et al.* (2017) ' PRC 2 is dispensable for HOTAIR -mediated transcriptional repression ', *The EMBO Journal*. EMBO, 36(8), pp. 981–994. doi: 10.15252/embj.201695335.

Potjewyd, F. *et al.* (2020) 'Degradation of Polycomb Repressive Complex 2 with an EED-Targeted Bivalent Chemical Degrader', *Cell Chemical Biology*. Elsevier Ltd, 27(1), pp. 47-56.e15. doi: 10.1016/j.chembiol.2019.11.006.

Del Prete, S. *et al.* (2015) 'One, two, three: Polycomb proteins hit all dimensions of gene regulation', *Genes*. MDPI AG, pp. 520–542. doi: 10.3390/genes6030520.

Qi, L. S. *et al.* (2013) 'Repurposing CRISPR as an RNA-γuided platform for sequence-specific control of gene expression', *Cell*. NIH Public Access, 152(5), pp. 1173–1183. doi: 10.1016/j.cell.2013.02.022.

Rada-Iglesias, A. *et al.* (2011) 'A unique chromatin signature uncovers early developmental enhancers in humans.', *Nature*. Nature Publishing Group, 470(7333), pp. 279–83. doi: 10.1038/nature09692.

Rao, S. S. P. *et al.* (2017) 'Cohesin Loss Eliminates All Loop Domains', *Cell*. Cell Press, 171(2), pp. 305-320.e24. doi: 10.1016/j.cell.2017.09.026.

Rastelli, L., Chan, C. S. and Pirrotta, V. (1993) 'Related chromosome binding sites for zeste, suppressors of zeste and Polycomb group proteins in Drosophila and their dependence on enhancer of zeste function', *EMBO Journal*, 12(4), pp. 1513–1522. doi: 10.1002/j.1460-2075.1993.tb05795.x.

Reijnen, M. J. *et al.* (1995) 'Polycomb and bmi-1 homologs are expressed in overlapping patterns in Xenopus embryos and are able to interact with each other', *Mechanisms of Development*. Mech Dev, 53(1), pp. 35–46. doi: 10.1016/0925-4773(95)00422-X.

Reverón-Gómez, N. *et al.* (2018) 'Accurate Recycling of Parental Histones Reproduces the Histone Modification Landscape during DNA Replication', *Molecular Cell*. Cell Press, 72(2), pp. 239-249.e5. doi: 10.1016/j.molcel.2018.08.010.

Reynolds, N. *et al.* (2012) 'NuRD-mediated deacetylation of H3K27 facilitates recruitment of Polycomb Repressive Complex 2 to direct gene repression', *EMBO Journal*. EMBO J, 31(3), pp. 593–605. doi: 10.1038/emboj.2011.431.

Rezania, A. *et al.* (2012) 'Maturation of human embryonic stem cell-derived pancreatic progenitors into functional islets capable of treating pre-existing diabetes in mice', *Diabetes*. Diabetes, 61(8), pp. 2016–2029. doi: 10.2337/db11-1711.

Rezania, A. *et al.* (2014) 'Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells', *Nature Biotechnology*. Nature Publishing Group, 32(11), pp. 1121–1133. doi: 10.1038/nbt.3033.

Rhodes, J. D. P. *et al.* (2020) 'Cohesin Disrupts Polycomb-Dependent Chromosome Interactions in Embryonic Stem Cells', *Cell Reports*. Elsevier B.V., 30(3), pp. 820-835.e10. doi: 10.1016/j.celrep.2019.12.057.

Riedel, M. J. *et al.* (2012) 'Immunohistochemical characterisation of cells co-producing insulin and glucagon in the developing human pancreas', *Diabetologia*. Springer, 55(2), pp. 372–381. doi: 10.1007/s00125-011-2344-9.

Riising, E. M. *et al.* (2014) 'Gene silencing triggers polycomb repressive complex 2 recruitment to CpG Islands genome wide', *Molecular Cell*. Cell Press, 55(3), pp. 347–360. doi: 10.1016/j.molcel.2014.06.005.

Roh, T. Y. *et al.* (2006) 'The genomic landscape of histone modifications in human T cells', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 103(43), pp. 15782–15787. doi: 10.1073/pnas.0607617103.

Roost, M. S. *et al.* (2015) 'KeyGenes, a Tool to Probe Tissue Differentiation Using a Human Fetal Transcriptional Atlas', *Stem Cell Reports*. Cell Press, 4(6), pp. 1112–1124. doi: 10.1016/j.stemcr.2015.05.002.

Rorsman, P. *et al.* (1989) 'Failure of glucose to elicit a normal secretory response in fetal

pancreatic beta cells results from glucose intensitivity of the ATP-regulated K+ channels', *Proceedings of the National Academy of Sciences of the United States of America*. Karolinska Institute, Karolinska Hospital, 86(12), pp. 4505–4509. doi: 10.1073/pnas.86.12.4505.

Rubio-Cabezas, O. *et al.* (2011) 'Permanent neonatal diabetes and enteric anendocrinosis associated with biallelic mutations in NEUROG3', *Diabetes*. American Diabetes Association, 60(4), pp. 1349–1353. doi: 10.2337/db10-1008.

Russ, H. A. *et al.* (2015) ' Controlled induction of human pancreatic progenitors produces functional beta-like cells in vitro ', *The EMBO Journal*. EMBO, 34(13), pp. 1759–1772. doi: 10.15252/embj.201591058.

Ryan, E. A., Bigam, D. and Shapiro, A. M. J. (2006) 'Current indications for pancreas or islet transplant', *Diabetes, Obesity and Metabolism*. Blackwell Publishing Ltd, 8(1), pp. 1–7. doi: 10.1111/j.1463-1326.2004.00460.x.

Sakamoto, K. M. *et al.* (2001) 'Protacs: Chimeric molecules that target proteins to the Skp1-Cullin-F box complex for ubiquitination and degradation', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 98(15), pp. 8554–8559. doi: 10.1073/pnas.141230798.

Saksouk, N. *et al.* (2014) 'Redundant Mechanisms to Form Silent Chromatin at Pericentromeric Regions Rely on BEND3 and DNA Methylation', *Molecular Cell*. Cell Press, 56(4), pp. 580–594. doi: 10.1016/j.molcel.2014.10.001.

Salami, J. *et al.* (2018) 'Androgen receptor degradation by the proteolysis-targeting chimera ARCC-4 outperforms enzalutamide in cellular models of prostate cancer drug resistance', *Communications Biology*. Nature Research, 1(1). doi: 10.1038/s42003-018-0105-8.

De Santa, F. *et al.* (2007) 'The Histone H3 Lysine-27 Demethylase Jmjd3 Links Inflammation to Inhibition of Polycomb-Mediated Gene Silencing', *Cell*. Elsevier, 130(6), pp. 1083–1094. doi: 10.1016/j.cell.2007.08.019.

Santos-Rosa, H. *et al.* (2002) 'Active genes are tri-methylated at K4 of histone H3', *Nature*. Nature Publishing Group, 419(6905), pp. 407–411. doi: 10.1038/nature01080.

Sarma, K. *et al.* (2008) 'Ezh2 Requires PHF1 To Efficiently Catalyze H3 Lysine 27

Trimethylation In Vivo', *Molecular and Cellular Biology*. American Society for Microbiology, 28(8), pp. 2718–2731. doi: 10.1128/mcb.02017-07.

Sathyan, K. M. *et al.* (2019) 'An improved auxin-inducible degron system preserves native protein levels and enables rapid and specific protein depletion', *Genes and Development*. Cold Spring Harbor Laboratory Press, 33(19–20), pp. 1441–1455. doi: 10.1101/gad.328237.119.

Satijn, D. P. *et al.* (1997) 'RING1 is associated with the polycomb group protein complex and acts as a transcriptional repressor.', *Molecular and Cellular Biology*. American Society for Microbiology, 17(7), pp. 4105–4113. doi: 10.1128/mcb.17.7.4105.

Schaffer, A. E. *et al.* (2010) 'Nkx6 transcription factors and Ptf1a function as antagonistic lineage determinants in multipotent pancreatic progenitors', *Developmental Cell*. Cell Press, 18(6), pp. 1022–1029. doi: 10.1016/j.devcel.2010.05.015.

Schaffer, A. E. *et al.* (2013) 'Nkx6.1 Controls a Gene Regulatory Network Required for Establishing and Maintaining Pancreatic Beta Cell Identity', *PLoS Genetics*, 9(1). doi: 10.1371/journal.pgen.1003274.

Scharp, D. W. *et al.* (1990) 'Insulin independence after islet transplantation into type I diabetic patient', *Diabetes*. American Diabetes Association Inc., 39(4), pp. 515–518. doi: 10.2337/diab.39.4.515.

Schmitges, F. W. *et al.* (2011) 'Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks', *Molecular Cell*. Cell Press, 42(3), pp. 330–341. doi: 10.1016/j.molcel.2011.03.025.

Schneider, J. *et al.* (2005) 'Molecular regulation of histone H3 trimethylation by COMPASS and the regulation of gene expression', *Molecular Cell*. Mol Cell, 19(6), pp. 849–856. doi: 10.1016/j.molcel.2005.07.024.

Schuettengruber, B. *et al.* (2017) 'Genome Regulation by Polycomb and Trithorax: 70 Years and Counting', *Cell*. Cell Press, pp. 34–57. doi: 10.1016/j.cell.2017.08.002.

Schulz, T. C. *et al.* (2012) 'A scalable system for production of functional pancreatic progenitors from human embryonic stem cells', *PLoS ONE*. Public Library of Science, 7(5). doi: 10.1371/journal.pone.0037004.

Schwanke, K. *et al.* (2014) 'Fast and efficient multitransgenic modification of human pluripotent

stem cells', *Human Gene Therapy Methods*. Mary Ann Liebert Inc., 25(2), pp. 136–153. doi: 10.1089/hgtb.2012.248.

Schwitzgebel, V. M. *et al.* (2003) 'Agenesis of Human Pancreas due to Decreased Half-Life of Insulin Promoter Factor 1', *The Journal of Clinical Endocrinology & Metabolism*. Oxford Academic, 88(9), pp. 4398–4406. doi: 10.1210/jc.2003-030046.

Segerstolpe, Å. *et al.* (2016) 'Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes', *Cell Metabolism*. Cell Press, 24(4), pp. 593–607. doi: 10.1016/j.cmet.2016.08.020.

Sen, S. Q. *et al.* (2019) 'Neuroblast-specific open chromatin allows the temporal transcription factor, Hunchback, to bind neuroblast-specific loci', *eLife*. NLM (Medline), 8. doi: 10.7554/eLife.44036.

Senior, P. A. and Pettus, J. H. (2019) 'Stem cell therapies for Type 1 diabetes: current status and proposed road map to guide successful clinical trials', *Diabetic Medicine*. Blackwell Publishing Ltd, 36(3), pp. 297–307. doi: 10.1111/dme.13846.

Sewalt, R. G. A. B. *et al.* (1998) 'Characterization of Interactions between the Mammalian Polycomb-Group Proteins Enx1/EZH2 and EED Suggests the Existence of Different Mammalian Polycomb-Group Protein Complexes', *Molecular and Cellular Biology*. American Society for Microbiology, 18(6), pp. 3586–3595. doi: 10.1128/mcb.18.6.3586.

Shahjalal, H. M. *et al.* (2018) 'Generation of pancreatic β cells for treatment of diabetes: Advances and challenges', *Stem Cell Research and Therapy*. BioMed Central Ltd., 9(1), pp. 1–19. doi: 10.1186/s13287-018-1099-3.

Shan, Y. *et al.* (2017) 'PRC2 specifies ectoderm lineages and maintains pluripotency in primed but not naïve ESCs', *Nature Communications*. Nature Publishing Group, 8(1). doi: 10.1038/s41467-017-00668-4.

Shao, Z. *et al.* (1999) 'Stabilization of chromatin structure by PRC1, a polycomb complex', *Cell*. Cell Press, 98(1), pp. 37–46. doi: 10.1016/S0092-8674(00)80604-2.

Shapiro, A. M. J. *et al.* (2000) 'Islet Transplantation in Seven Patients with Type 1 Diabetes Mellitus Using a Glucocorticoid-Free Immunosuppressive Regimen', *New England Journal of*

Medicine. Massachusetts Medical Society , 343(4), pp. 230–238. doi: 10.1056/NEJM200007273430401.

Shema, E. *et al.* (2016) 'Single-molecule decoding of combinatorially modified nucleosomes', *Science*. American Association for the Advancement of Science, 352(6286), pp. 717–721. doi: 10.1126/science.aad7701.

Shen, X. *et al.* (2008) 'EZH1 Mediates Methylation on Histone H3 Lysine 27 and Complements EZH2 in Maintaining Stem Cell Identity and Executing Pluripotency', *Molecular Cell*. Elsevier Inc., 32(4), pp. 491–502. doi: 10.1016/j.molcel.2008.10.016.

Shen, X. *et al.* (2009) 'Jumonji modulates Polycomb activity and self-renewal versus differentiation of stem cells', *Cell*. NIH Public Access, 139(7), p. 1303. doi: 10.1016/J.CELL.2009.12.003.

Shim, J. H. *et al.* (2007) 'Directed differentiation of human embryonic stem cells towards a pancreatic cell fate', *Diabetologia*. Springer, 50(6), pp. 1228–1238. doi: 10.1007/s00125-007-0634-z.

Shimomura, H. *et al.* (2000) *Brief Genetics Report Nonsense Mutation of Islet-1 Gene (Q310X) Found in a Type 2 Diabetic Patient With a Strong Family History*, *DIABETES*.

Silva, J. *et al.* (2003) 'Establishment of histone H3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes', *Developmental Cell*. Cell Press, 4(4), pp. 481–495. doi: 10.1016/S1534-5807(03)00068-6.

Singh, V. K. *et al.* (2015) 'Induced pluripotent stem cells: Applications in regenerative medicine, disease modeling, and drug discovery', *Frontiers in Cell and Developmental Biology*. Frontiers Media S.A. doi: 10.3389/fcell.2015.00002.

Smits, A. H. *et al.* (2013) 'Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics', *Nucleic Acids Research*. Nucleic Acids Res, 41(1). doi: 10.1093/nar/gks941.

Sneddon, J. B. *et al.* (2018) 'Stem Cell Therapies for Treating Diabetes: Progress and Remaining Challenges', *Cell Stem Cell*. Cell Press, pp. 810–823. doi: 10.1016/j.stem.2018.05.016.

Soares, L. M. *et al.* (2017) 'Determinants of Histone H3K4 Methylation Patterns', *Molecular Cell*. Cell Press, 68(4), pp. 773-785.e6. doi: 10.1016/j.molcel.2017.10.013.

Son, J. *et al.* (2013) 'Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin', *Genes and Development*. Cold Spring Harbor Laboratory Press, 27(24), pp. 2663–2677. doi: 10.1101/gad.225888.113.

Sosa-Pineda, B. *et al.* (1997) 'The Pax4 gene is essential for differentiation of insulin-producing β cells in the mammalian pancreas', *Nature*. Nature, 386(6623), pp. 399–402. doi: 10.1038/386399a0.

Steyer, B. *et al.* (2016) 'High content analysis platform for optimization of lipid mediated CRISPR-Cas9 delivery strategies in human cells', *Acta Biomaterialia*. Elsevier Ltd, 34, pp. 143–158. doi: 10.1016/j.actbio.2015.12.036.

Steyer, B. *et al.* (2018) 'Scarless Genome Editing of Human Pluripotent Stem Cells via Transient Puromycin Selection', *Stem Cell Reports*. Cell Press, 10(2), pp. 642–654. doi: 10.1016/j.stemcr.2017.12.004.

Stoffers, D. A. *et al.* (1997) 'Pancreatic agenesis attributable to a single nucleotide deletion in the human IPF1 gene coding sequence', *Nature Genetics*. Nat Genet, 15(1), pp. 106–110. doi: 10.1038/ng0197-106.

Stolovich-Rain, M. *et al.* (2015) 'Weaning Triggers a Maturation Step of Pancreatic β Cells', *Developmental Cell*. Cell Press, 32(5), pp. 535–545. doi: 10.1016/j.devcel.2015.01.002.

Sybirna, A. *et al.* (2020) 'A critical role of PRDM14 in human primordial germ cell fate revealed by inducible degrons', *Nature Communications*. Nature Research, 11(1). doi: 10.1038/s41467-020-15042-0.

Sze, C. C. *et al.* (2020) 'Coordinated regulation of cellular identity–associated H3K4me3 breadth by the COMPASS family', *Science Advances*. American Association for the Advancement of Science, 6(26), p. eaaz4764. doi: 10.1126/sciadv.aaz4764.

Tabar, V. and Studer, L. (2014) 'Pluripotent stem cells in regenerative medicine: Challenges and recent progress', *Nature Reviews Genetics*. Nat Rev Genet, pp. 82–92. doi: 10.1038/nrg3563.

Takahashi, K. *et al.* (2007) 'Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by

Defined Factors', *Cell*, 131(5), pp. 861–872. doi: http://dx.doi.org/10.1016/j.cell.2007.11.019.

Takahashi, K. and Yamanaka, S. (2006) 'Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors', *Cell*, 126(4), pp. 663–676. doi: 10.1016/j.cell.2006.07.024.

Tamburri, S. *et al.* (2020) 'Histone H2AK119 Mono-Ubiquitination Is Essential for Polycomb-Mediated Transcriptional Repression', *Molecular Cell*. Cell Press, 77(4), pp. 840-856.e5. doi: 10.1016/j.molcel.2019.11.021.

Teo, A. K. K. *et al.* (2011) 'Pluripotency factors regulate definitive endoderm specification through eomesodermin', *Genes and Development*. Cold Spring Harbor Laboratory Press, 25(3), pp. 238–250. doi: 10.1101/gad.607311.

Thakore, P. I. *et al.* (2015) 'Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements', *Nature Methods*. Nature Publishing Group, 12(12), pp. 1143–1149. doi: 10.1038/nmeth.3630.

Thalheim, T. *et al.* (2017) 'The regulatory capacity of bivalent genes—A theoretical approach', *International Journal of Molecular Sciences*. MDPI AG, 18(5). doi: 10.3390/ijms18051069.

Thomson, J. A. (1998) 'Embryonic stem cell lines derived from human blastocysts', *Science*. American Association for the Advancement of Science, 282(5391), pp. 1145–1147. doi: 10.1126/science.282.5391.1145.

Thomson, J. A. *et al.* (1998) 'Embryonic Stem Cell Lines Derived from Human Blastocysts', 282(November), pp. 1145–1148.

Trentini, D. B. *et al.* (2020) 'Role for ribosome-associated quality control in sampling proteins for MHC class I-mediated antigen presentation', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 117(8), pp. 4099–4108. doi: 10.1073/pnas.1914401117.

Uhlen, M. *et al.* (2015) 'Tissue-based map of the human proteome', *Science*. American Association for the Advancement of Science, 347(6220), pp. 1260419–1260419. doi: 10.1126/science.1260419.

Vakoc, C. R. *et al.* (2006) 'Profile of Histone Lysine Methylation across Transcribed Mammalian

Chromatin', *Molecular and Cellular Biology*. American Society for Microbiology, 26(24), pp. 9185–9195. doi: 10.1128/mcb.01529-06.

Vallier, L. *et al.* (2009) 'Signaling pathways controlling pluripotency and early cell fate decisions of human induced pluripotent stem cells', *Stem Cells*. John Wiley & Sons, Ltd, 27(11), pp. 2655–2666. doi: 10.1002/stem.199.

Vanheer, L. *et al.* (2020) 'Revealing the Key Regulators of Cell Identity in the Human Adult Pancreas', *bioRxiv*. Cold Spring Harbor Laboratory, p. 2020.09.23.310094. doi: 10.1101/2020.09.23.310094.

Vaskova, E. A. *et al.* (2013) '"Epigenetic memory" phenomenon in induced pluripotent stem cells', *Acta Naturae*. Russian Federation Agency for Science and Innovation, pp. 15–21. doi: 10.32607/20758251-2013-5-4-15-21.

Velazco-Cruz, L. *et al.* (2019) 'Acquisition of Dynamic Function in Human Stem Cell-Derived β Cells', *Stem Cell Reports*. Cell Press, 12(2), pp. 351–365. doi: 10.1016/j.stemcr.2018.12.012.

Veres, A. *et al.* (2019) 'Charting cellular identity during human in vitro β-cell differentiation', *Nature*. Nature Publishing Group, 569(7756), pp. 368–373. doi: 10.1038/s41586-019-1168-5.

Verheijen, M. *et al.* (2019) 'DMSO induces drastic changes in human cellular processes and epigenetic landscape in vitro', *Scientific Reports*. Nature Publishing Group, 9(1), pp. 1–12. doi: 10.1038/s41598-019-40660-0.

Verma, R., Mohl, D. and Deshaies, R. J. (2020) 'Harnessing the Power of Proteolysis for Targeted Protein Inactivation', *Molecular Cell*. Cell Press, pp. 446–460. doi: 10.1016/j.molcel.2020.01.010.

Verma, S. K. *et al.* (2012) 'Identification of potent, selective, cell-Active inhibitors of the histone lysine methyltransferase EZH2', *ACS Medicinal Chemistry Letters*, 3(12), pp. 1091–1096. doi: 10.1021/ml3003346.

Vermeulen, M. *et al.* (2007) 'Selective Anchoring of TFIID to Nucleosomes by Trimethylation of Histone H3 Lysine 4', *Cell*. Cell Press, 131(1), pp. 58–69. doi: 10.1016/j.cell.2007.08.016.

Vermeulen, M. *et al.* (2010) 'Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers', *Cell*. Cell, 142(6), pp. 967–980. doi:

10.1016/J.CELL.2010.08.020.

Vizán, P. *et al.* (2015) 'Role of PRC2-associated factors in stem cells and disease', *FEBS Journal*. Blackwell Publishing Ltd, 282(9), pp. 1723–1735. doi: 10.1111/febs.13083.

Voigt, P. *et al.* (2012) 'Asymmetrically modified nucleosomes', *Cell*. NIH Public Access, 151(1), pp. 181–193. doi: 10.1016/j.cell.2012.09.002.

Voigt, P. and Reinberg, D. (2011) 'Histone Tails: Ideal Motifs for Probing Epigenetics through Chemical Biology Approaches', *ChemBioChem*. Howard Hughes Medical Institute, pp. 236–252. doi: 10.1002/cbic.201000493.

Wachter, E. *et al.* (2014) 'Synthetic CpG islands reveal DNA sequence determinants of chromatin structure', *eLife*, 3, p. e03397. doi: 10.7554/eLife.03397.

Wagner, E. J. and Carpenter, P. B. (2012) 'Understanding the language of Lys36 methylation at histone H3', *Nature Reviews Molecular Cell Biology*. Nat Rev Mol Cell Biol, pp. 115–126. doi: 10.1038/nrm3274.

Wang, A. *et al.* (2015) 'Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates', *Cell Stem Cell*. Elsevier Inc., 16(4), pp. 386–399. doi: 10.1016/j.stem.2015.02.013.

Wang, H. *et al.* (2004) 'Role of histone H2A ubiquitination in Polycomb silencing', *Nature*. Nature Publishing Group, 431(7010), pp. 873–878. doi: 10.1038/nature02985.

Wang, H. *et al.* (2007) 'MAFA controls genes implicated in insulin biosynthesis and secretion.', *Diabetologia*. NIH Public Access, 50(2), pp. 348–358. doi: 10.1007/s00125-006-0490-2.

Wang, L. *et al.* (2017) 'Activin/Smad2-induced histone H3 Lys-27 trimethylation (H3K27me3) reduction is crucial to initiate mesendoderm differentiation of human embryonic stem cells', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology Inc., 292(4), pp. 1339–1350. doi: 10.1074/jbc.M116.766949.

Wang, P. *et al.* (2009) 'Global Analysis of H3K4 Methylation Defines MLL Family Member Targets and Points to a Role for MLL1-Mediated H3K4 Methylation in the Regulation of Transcriptional Initiation by RNA Polymerase II', *Molecular and Cellular Biology*. American Society for Microbiology, 29(22), pp. 6074–6085. doi: 10.1128/mcb.00924-09.

Wang, X., Paucek, R. D., *et al.* (2017) 'Molecular analysis of PRC2 recruitment to DNA in chromatin and its inhibition by RNA', *Nature Structural and Molecular Biology*. Nature Publishing Group, 24(12), pp. 1028–1038. doi: 10.1038/nsmb.3487.

Wang, X., Goodrich, K. J., *et al.* (2017) 'Targeting of Polycomb Repressive Complex 2 to RNA by Short Repeats of Consecutive Guanines', *Molecular Cell*. Cell Press, 65(6), pp. 1056-1067.e5. doi: 10.1016/j.molcel.2017.02.003.

Wang, X. *et al.* (2019) 'Point mutations in the PDX1 transactivation domain impair human β-cell development and function', *Molecular Metabolism*. Elsevier GmbH, 24, pp. 80–97. doi: 10.1016/j.molmet.2019.03.006.

Wang, Z. *et al.* (2009) 'Genome-wide Mapping of HATs and HDACs Reveals Distinct Functions in Active and Inactive Genes', *Cell*. Cell Press, 138(5), pp. 1019–1031. doi: 10.1016/j.cell.2009.06.049.

Weiner, A. *et al.* (2016) 'Co-ChIP enables genome-wide mapping of histone mark co-occurrence at single-molecule resolution', *Nature Biotechnology*. Nature Publishing Group, 34(9), pp. 953–961. doi: 10.1038/nbt.3652.

Weintraub, A. S. *et al.* (2017) 'YY1 Is a Structural Regulator of Enhancer-Promoter Loops', *Cell*. Cell Press, 171(7), pp. 1573-1588.e28. doi: 10.1016/j.cell.2017.11.008.

Wen, B. *et al.* (2009) 'Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells', *Nature Genetics*, 41(2), pp. 246–250. doi: 10.1038/ng.297.

Wiles, E. T. and Selker, E. U. (2017) 'H3K27 methylation: a promiscuous repressive chromatin mark', *Current Opinion in Genetics and Development*. Elsevier Ltd, pp. 31–37. doi: 10.1016/j.gde.2016.11.001.

Wingett, S. W. and Andrews, S. (2018) 'FastQ Screen: A tool for multi-genome mapping and quality control', *F1000Research*. F1000 Research Ltd, 7, p. 1338. doi: 10.12688/f1000research.15931.2.

Winter, G. E. *et al.* (2015) 'Phthalimide conjugation as a strategy for in vivo target protein degradation', *Science*. American Association for the Advancement of Science, 348(6241), pp.

1376–1381. doi: 10.1126/science.aab1433.

Wu, M. *et al.* (2008) 'Molecular Regulation of H3K4 Trimethylation by Wdr82, a Component of Human Set1/COMPASS', *Molecular and Cellular Biology*. American Society for Microbiology, 28(24), pp. 7337–7344. doi: 10.1128/mcb.00976-08.

Wysocka, J. *et al.* (2006) 'A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling', *Nature*. Nature Publishing Group, 442(7098), pp. 86–90. doi: 10.1038/nature04815.

Xie, R. *et al.* (2014) 'NIH Public Access', 12(2), pp. 224–237. doi: 10.1016/j.stem.2012.11.023.Dynamic.

Xu, C.-R. *et al.* (2014) 'Dynamics of genomic H3K27me3 domains and role of EZH2 during pancreatic endocrine specification.', *The EMBO journal*, 33(19), pp. 1–14. doi: 10.15252/embj.201488671.

Xu, C. R. *et al.* (2011) 'Chromatin "prepattern" and histone modifiers in a fate choice for liver and pancreas', *Science*. American Association for the Advancement of Science, 332(6032), pp. 963–966. doi: 10.1126/science.1202845.

Yamamoto, K. *et al.* (2004) 'Polycomb Group Suppressor of Zeste 12 Links Heterochromatin Protein 1α and Enhancer of Zeste 2', *Journal of Biological Chemistry*, 279(1), pp. 401–406. doi: 10.1074/jbc.M307344200.

Yan, I. *et al.* (2016) 'Epigenomic landscape of human fetal brain, heart, and liver', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology Inc., 291(9), pp. 4386–4398. doi: 10.1074/jbc.M115.672931.

Yearim, A. *et al.* (2015) 'HP1 Is Involved in Regulating the Global Impact of DNA Methylation on Alternative Splicing'. doi: 10.1016/j.celrep.2015.01.038.

Yesbolatova, A. *et al.* (2019) 'Generation of conditional auxin-inducible degron (AID) cells and tight control of degron-fused proteins using the degradation inhibitor auxinole', *Methods*. Academic Press Inc., 164–165, pp. 73–80. doi: 10.1016/j.ymeth.2019.04.010.

Yu, J. *et al.* (2007) 'Induced pluripotent stem cell lines derived from human somatic cells', *Science*, 318(5858), pp. 1917–1920. doi: 10.1126/science.1151526.

Yu, J. R. *et al.* (2019) 'PRC2 is high maintenance', *Genes and Development*. Cold Spring Harbor Laboratory Press, pp. 903–935. doi: 10.1101/gad.325050.119.

Yu, X. X. *et al.* (2018) 'Dynamics of chromatin marks and the role of JMJD3 during pancreatic endocrine cell fate commitment', *Development (Cambridge)*. Company of Biologists Ltd, 145(6). doi: 10.1242/dev.163162.

Yuan, W. *et al.* (2011) 'H3K36 methylation antagonizes PRC2-mediated H3K27 methylation', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology, 286(10), pp. 7983–7989. doi: 10.1074/jbc.M110.194027.

Yun, M. *et al.* (2011) 'Readers of histone modifications', *Cell Research*. Nature Publishing Group, pp. 564–578. doi: 10.1038/cr.2011.42.

Zasadzińska, E. *et al.* (2018) 'Inheritance of CENP-A Nucleosomes during DNA Replication Requires HJURP', *Developmental Cell*. Cell Press, 47(3), pp. 348-362.e7. doi: 10.1016/j.devcel.2018.09.003.

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

Zhang, C. *et al.* (2005) 'MafA Is a Key Regulator of Glucose-Stimulated Insulin Secretion', *Molecular and Cellular Biology*. American Society for Microbiology, 25(12), pp. 4969–4976. doi: 10.1128/mcb.25.12.4969-4976.2005.

Zhang, Q. *et al.* (2019) 'RNA exploits an exposed regulatory site to inhibit the enzymatic activity of PRC2', *Nature Structural and Molecular Biology*. Nature Publishing Group, 26(3), pp. 237–247. doi: 10.1038/s41594-019-0197-y.

Zhang, T., Cooper, S. and Brockdorff, N. (2015) 'The interplay of histone modifications – writers that read', *EMBO reports*. EMBO, 16(11), pp. 1467–1481. doi: 10.15252/embr.201540945.

Zhang, W. *et al.* (2006) 'Zfp206 regulates ES cell gene expression and differentiation', *Nucleic Acids Research*. Nucleic Acids Res, 34(17), pp. 4780–4790. doi: 10.1093/nar/gkl631.

Zhang, W. *et al.* (2013) 'Structural plasticity of histones H3-H4 facilitates their allosteric exchange between RbAp48 and ASF1', *Nature Structural and Molecular Biology*. Nat Struct Mol

Biol, 20(1), pp. 29–35. doi: 10.1038/nsmb.2446.

Zhao, J. *et al.* (2008) 'Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome', *Science*. American Association for the Advancement of Science, 322(5902), pp. 750–756. doi: 10.1126/science.1163045.

Zhao, X. D. *et al.* (2007) 'Whole-Genome Mapping of Histone H3 Lys4 and 27 Trimethylations Reveals Distinct Genomic Compartments in Human Embryonic Stem Cells', *Cell Stem Cell*, 1(3), pp. 286–298. doi: 10.1016/j.stem.2007.08.004.

Zhou, C. J. *et al.* (2019) 'Loss of CENPF leads to developmental failure in mouse embryos', *Cell Cycle*. Taylor and Francis Inc., 18(20), pp. 2784–2799. doi: 10.1080/15384101.2019.1661173.

Zhou, Q. *et al.* (2008) 'In vivo reprogramming of adult pancreatic exocrine cells to β-cells', *Nature*. Nature Publishing Group, 455(7213), pp. 627–632. doi: 10.1038/nature07314.

Zhu, Y. *et al.* (2017) 'PDX1, Neurogenin-3, and MAFA: Critical transcription regulators for beta cell development and regeneration', *Stem Cell Research and Therapy*. BioMed Central Ltd., p. 240. doi: 10.1186/s13287-017-0694-z.