The Role of EZH2 in the Induction and Maintenance of Acute Myeloid Leukaemia

Faisal Basheer

Downing College

University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy

September 2017

Declaration

The work presented in this dissertation was carried out under the supervision of Professor Brian Huntly from November 2013 to November 2016 in the Department of Haematology, University of Cambridge.

This dissertation 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. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or 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.

It does not exceed the prescribed word limit set by the Clinical Medicine and Clinical Veterinary Medicine Degree Committee.

Faisal Basheer

September 2017

Acknowledgements

First and foremost, I would like to thank my supervisor, Professor Brian Huntly, for providing me with the opportunity to work towards this thesis - for his help, encouragement and support over the last four years of my PhD. His dedicated high-level supervision, meticulous attention to detail and excellent direction throughout this time has undoubtedly helped me to maximise my potential in the lab throughout a highly interesting and rewarding project.

I would also like to thank all of my colleagues and friends in the Huntly laboratory over the past four years - working together with all of you as a team has been an unforgettable and rewarding experience. Firstly, George Giotopoulos for all the technical training he has given me regarding most aspects of this project, in particular with the planning and setting up of the numerous mouse models required, but also for his constant help, supervision and general troubleshooting outwith this over the course of the entire four years. I am grateful to Eshwar Meduri for his help and assistance with analysis and referencing of the large genomic datasets generated during this project, using his bioinformatic expertise. Haiyang Yun and Daniel Sasca have been very supportive through the latter stages of this project, assisting not only in mouse work but also with helpful detailed discussions and their observations when analysing the genomic datasets. I would specifically like to thank Haiyang for his teaching and supervision when performing all the chromatin immunoprecipitation and subsequent necessary steps for high throughput sequencing. Olivia Sheppard and Sarah Horton have been constantly helpful with large parts of the mouse work and I would like to express my gratitude toward them. I would also like to thank Hikari Osaki for her help and teaching during the early parts of the project involving retroviral protocols and tissue culture training. Paolo Gallipoli has been helpful in collating and working with primary samples and inhibitors. The Vassiliou Group at the Sanger Centre, in particular Milena Mazan and Monika Dudek were very helpful for overseeing the breeding and maintenance of the *Ezh*2 conditional mice and thank you to their group leader, Dr George Vassiliou for his collaboration. The CIMR Flow cytometry core have also been very accommodating throughout this project and I am grateful to them.

I would also like to thank the Wellcome Trust PhD Programme for Clinicians for awarding me with a junior clinical research fellowship which has allowed me to carry out this research, and all the patients with AML who have generously donated samples for this project.

Finally, I would like to thank my wife, parents and family for their patience, understanding and support throughout everything.

Work done and assistance received by others

Dr George Giotopoulos:	assistance with setting up *in vitro* and *in vivo* mouse models and analysing data assistance with setup and then overseeing functional validation mouse experiment
Dr Eshwar Meduri:	bioinformatic analysis of RNA-sequencing and ChIP-sequencing data
Dr Haiyang Yun:	supervision and assistance with ChIP-sequencing experiments
Dr Daniel Sasca:	assistance with library preparation of ChIP samples
Genomics Core, CRUK Cambridge Institute:	RNA-sequencing and high throughput sequencing of ChIP samples

Abbreviations

AML	Acute Myeloid Leukaemia
AML1-ETO	Acute Myeloid Leukaemia 1 - Eight Twenty One
BM	Bone Marrow
BSA	Bovine Serum Albumin
ChIP	Chromatin immunoprecipitation
cDNA	Complementary DNA
C_t	Threshold cycle
DLBCL	Diffuse Large B Cell Lymphoma
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic acid
EED	Embryonic Ectoderm Development
ES	Embryonic Stem
EZH1	Enhancer of Zeste Homologue 1
EZH2	Enhancer of Zeste Homologue 2
FBS	Foetal Bovine Serum
FCS	Foetal Calf Serum
FL	Follicular Lymphoma
GFP	Green Fluorescent Protein
GSK	GlaxoSmithKline
HEPES	4-(2-hydroxyethyl)-1piperazineethanesulfonic acid
HSC	Haematopoietic Stem Cell

HSPC	Haematopoietic Stem and Progenitor Cell
IC_{50}	Half maximal inhibitory concentration
IRES	Internal Ribosomal Entry Site
LB	Luria-Bertani
LT-HSC	Long Term HSC
MCS	Multiple Cloning Site
MDS	Myelodysplasia
MLL	Mixed Lineage Leukaemia
MPN	Myeloproliferative Neoplasm
MPP	Multipotent progenitor
MOZ-TIF2	Monocytic Leukaemia Zinc finger – Transcriptional Intermediary Factor 2
MTA	Materials Transfer Agreement
NPM1	Nucleophosmin
PI	Propidium Iodide
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PRC1	Polycomb Repressive Complex 1
PRC2	Polycomb Repressive Complex 2
RE	Restriction Enzyme
RNA	Ribonucleic Acid
RPMI	Roswell Park Memorial Institute
RUNX1	Runt-related transcription factor 1
SCF	Stem Cell Factor

SDS	Sodium Dodecyl Sulphate
SDS-PAGE	SDS-Poly Acrylamide Gel Electrophoresis
SEM	Standard Error of the Mean
ST-HSC	Short Term HSC
SUZ12	Suppressor of Zeste 12
TSS	Transcriptional Start Site
UV	Ultraviolet
WHO	World Health Organisation
WT	Wild Type
YFP	Yellow Fluorescent Protein

The role of EZH2 in the induction and maintenance of Acute Myeloid Leukaemia

Faisal Basheer, Downing College

Summary

Acute myeloid leukaemia (AML) is an aggressive haematological cancer that remains an unmet medical need, with over 70% of patients still succumbing to this disease. Despite this, a deeper understanding of AML biology has allowed for the identification of novel agents with efficacy against specific molecular subtypes. Recently, a better understanding of aberrant epigenetic control in AML has led to the development of promising small molecule inhibitors of epigenetic regulators. The work described in this thesis details the results of genetic and pharmacological experiments designed to perturb the histone methyltransferase EZH2, a critical regulator of gene silencing during development and cell fate decisions and a potential therapeutic target, in the context of AML. The core aims were to assess its role during the development and subsequent maintenance of AML both *in vitro* and *in vivo* using mouse models, and to subsequently determine mechanisms by which its activity may facilitate leukaemia and how inhibition may abrogate leukaemic growth.

Through utilising a conditional *Ezh2* knock-out murine model together with bone marrow retroviral transduction and transplantation assays, *Ezh2* was shown to have markedly differing roles during the induction and maintenance of three well-characterised subtypes of AML (MLL-AF9, AML1-ETO9a, MOZ-TIF2) both *in vitro* and *in vivo*, serving to highlight its complex nature. Whilst there were no significant effects *in vitro*, prior *Ezh2* loss led to a significant acceleration in the induction of MLL-AF9 and AML1-ETO9a leukaemias *in vivo*, whilst significantly decelerating induction of MOZ-TIF2 leukaemias *in vivo*. This finding of a single gene demonstrating both tumour-suppressive and oncogenic roles, clearly dependent upon subtype of AML during the same induction phase of disease was particularly surprising.

Contrary to the induction experiments, *Ezh2* loss *in vitro* completely abolished the growth of MLL-AF9, AML1-ETO9a and MOZ-TIF2 transformed murine cell lines, strongly suggesting that it was an oncogenic facilitator of the transformed state. When examined *in vivo*, *Ezh2* loss caused a significant prolongation in the latency of disease development of MLL-AF9 and AML1-ETO9a, suggesting that it was required for the maintenance of both as per the *in vitro* assays. Though a similar trend was established for secondary MOZ-TIF2 leukaemias *in vivo*, this was not statistically significant.

Thus, completely contrasting, stage-specific requirements of *Ezh2* during the evolution of AML were demonstrated. The diametrically opposite effects of *Ezh2* loss between the induction and maintenance phases of MLL-AF9 and AML1-ETO9a leukaemias *in vivo*, was a novel finding, suggesting that this gene was unexpectedly capable of fulfilling both tumour-suppressive and oncogenic roles within the same subtypes of AML, but at different phases of disease. Whilst the maintenance experiments provided evidence for targeting EZH2 therapeutically, the acceleration seen upon *Ezh2* loss during AML induction raised caution over EZH2 inhibition. This prompted a detailed mechanistic analysis to reconcile these disparate effects.

Utilising RNA-sequencing, further genomic characterisation of differential gene expression changes following *Ezh2* loss in murine HSPCs was made. This demonstrated a significant upregulation of genes - consistent with its role as a transcriptional repressor. A number of these upregulated genes were known oncogenes that were potentially facilitating leukaemia induction. When compared to differential gene expression datasets generated from RNA-sequencing of the $Ezh2^{+/+}$ and $Ezh2^{-/-}$ MLL-AF9 and AML1-ETO9a leukaemias that had shown a clear acceleration in the induction experiments, the overlaps identified *Lin28b* (a negative regulator of the let-7 microRNA tumour suppressor family and itself implicated as an oncogene in a number of other cancer models) and *Plag1* (a zinc-finger transcription factor also exhibiting oncogenic function in other leukaemia models and cancers) as being significantly upregulated. These were common to the *Ezh2*-null non-transformed, MLL-AF9 and/or AML1-ETO9a states, implying they may be functioning as downstream mediators of *Ezh2* loss and co-operating with these AML oncogenes in accelerating leukaemogenesis.

A more detailed characterisation of the epigenetic changes following *Ezh2* loss in the non-transformed state was then conducted in order to provide a mechanism at the chromatin level for the de-repression of genes seen during induction. ChIP-sequencing of the repressive histone mark, H3K27me3, in the non-transformed state following *Ezh2* loss identified that a significant number of peaks were downregulated across the genome. When compared to the above dataset, it was apparent that a clear subset of the genes upregulated following *Ezh2* loss (approximately 20%) were associated with a significant downregulation of H3K27me3, particularly at their promoter regions. Interestingly, whilst there were no significant changes in the activating H3K27Ac mark (typically associated with enhancers) near these genes, the majority were also populated with the activating H3K4me3 mark at their promoters, suggesting bivalency. This suggested loss of *Ezh2* resulted in the resolution of these two conflicting histone marks, in favour of the activating H3K4me3 mark, thus explaining their increase in expression at the transcript level. Strikingly, following *Ezh2* loss, a few select candidate genes also demonstrated a

clear increase in H3K4me3 at their promoters – this included *Lin28b* and *Plag1.* Whilst directing further scrutiny of these two candidate genes, this detailed characterisation also revealed that loss of *Ezh2* was interfering with H3K4me3 at highly specific genes, either directly or indirectly through changes in H3K27me3, suggesting a significant and active "cross-talk" between H3K27me3 and H3K4me3 in gene regulation.

Whilst overexpression of *Lin28b* in MLL-AF9 driven murine primary leukaemias *in vivo* showed a trend towards acceleration of this aggressive disease, overexpression of *Plag1* led to a clear and significant acceleration of MLL-AF9 driven primary leukaemias *in vivo,* functionally validating the genomic data and the hypothesis that this candidate gene was cooperating with MLL-AF9 following *Ezh2* loss. The synergism between *Ezh2* loss and MLL-AF9 is likely to be mediated by the concerted action of several genes and this data indicates that *Plag1* is clearly one of these downstream targets.

To explain the opposing effects of *Ezh2* loss between induction and maintenance of MLL-AF9 leukaemias, differential gene expression changes were assessed in spleen tumours from wild-type MLL-AF9 primary murine leukaemias cultured *in vitro*, following treatment with either vehicle control or GSK343 (an EZH2 inhibitor). These cultured tumours showed marked sensitivity to GSK343 *in vitro*, and following RNA-sequencing, demonstrated an almost entirely different gene set was being de-repressed following Ezh2 inhibition of these established leukaemias to that during induction, potentially reconciling the opposite effects seen between induction and maintenance following Ezh2 loss. This provided a rationale for subsequent experiments designed to assess the preclinical *in vitro* efficacy of GSK343 in AML.

Lastly, functional *in vitro* experiments demonstrated clear sensitivity of the Kasumi and KG1 AML cell lines, immortalised MLL-AF9 and MOZ-TIF2 murine HSPCs, wild-type MLL-AF9 primary leukaemia spleen tumours and importantly, human primary AML samples to GSK343. There was no evidence that Ezh2/EZH2 inhibition across these cell types accelerated their proliferative or clonogenic growth. This correlated directly with the maintenance *in vitro* and *in vivo* experiments where *Ezh2* was demonstrated to have an oncogenic role, thus providing clear evidence that EZH2 inhibition may be a viable therapeutic strategy in AML.

In summary, the work presented herein provides data demonstrating the contrasting, stage-specific roles of EZH2 during AML evolution, as well as mechanistic and functional data to potentially explain and reconcile these differences, and finally, preclinical data affording confidence that the therapeutic targeting of EZH2 in this aggressive haematological cancer is a viable opportunity.

Table of Contents

List of Tables and Figures

CHAPTER 5 Genomic analysis of transcriptional programmes subverted by Ezh2 during AML induction and maintenance

Chapter 1

Introduction

1.1 Acute Myeloid Leukaemia: incidence, aetiology and clinical features

Acute myeloid leukaemia (AML) is an aggressive cancer infiltrating bone marrow, blood and other tissues with a characteristically clonal, abnormally differentiated proliferation of early myeloid cells that gain a survival advantage. These rapidly subsume normal bone marrow function through impaired haematopoiesis leading to profound bone marrow failure. The majority of patients present with combined leucocytosis and anaemia or thrombocytopenia, with fatigue, weight loss and anorexia being common symptoms. If untreated, death usually results within months of diagnosis due to bleeding or severe infections.

AML is the most common leukaemia in adults, accounting for a quarter of all leukaemias in adults, with an incidence of 6 per 100,000 males and 4 per 100,000 females of the UK population annually. [1] The incidence of AML increases with age, from 17 - 25 per 100,000 population between males and females at 65 years old, to a peak of 35 per 100,000 in males and 23 per 100,000 in females at 75 years of age. In most cases the aetiology is unclear, arising as a *de novo* malignancy, but many risk factors and several pre-leukaemic conditions are known. Myeloproliferative neoplasms and myelodysplasia can transform into AML taking on an aggressive course that inevitably leads to death if untreated. Other predisposing conditions include Down Syndrome that has a 50-fold increase in AML incidence during the first 5 years of life. There is also a degree of genetic predisposition, with first degree relatives of patients with AML have a 2-3 fold increased risk of developing AML themselves. Inherited myeloid neoplasms associated with germline mutations in a number of genes have also been described, as have SNPs that increase the likelihood of AML development. [2,3]

AML can arise in patients as a consequence of prior therapies for other malignancies that often involve cytotoxic drugs, other chemotherapy and ionising radiation. As a result of patients surviving longer from these therapies the number of therapy-related AML cases is continuing to rise. Alkylating agents and ionising radiation are significantly associated with therapy-related AML, commonly occurring as advanced therapy-related myelodysplasia progressing to overt AML approximately 5-10 years after treatment. Topoisomerase II drugs are strongly associated with overt AML with a latency period of about 1-5 years.

1.2 AML cytogenetic, molecular heterogeneity and prognostic classification

AML is a considerably heterogeneous disease with several distinct subgroups characterised by cytogenetic, molecular, gene expression and morphological characteristics. The molecular heterogeneity of AML in particular has expanded significantly over the past 15 years, yielding better insights into prognostic stratification. Conventional cytogenetic testing and, more recently, screening for select molecular lesions has allowed for a more detailed classification of AML. The World Health Organisation classifies AML according to clinical, morphological cytogenetic and molecular factors. A recent revision of the WHO classification of AML was published in 2016 which introduces further molecular changes to the section on AML with recurrent genetic abnormalities and identifies new provisional entities with characteristic clinicopathological features. The 2016 revised WHO classification is shown below (Table 1.1):

Acute myeloid leukemia (AML) and related neoplasms
AML with recurrent genetic abnormalities
AML with t(8;21)(q22;q22.1);*RUNX1-RUNX1T1*
AML with inv(16)(p13.1q22) or t(16;16)(p13.1;q22);*CBFB-MYH11*
APL with *PML-RARA*
AML with t(9;11)(p21.3;q23.3);*MLLT3-KMT2A*
AML with t(6;9)(p23;q34.1);*DEK-NUP214*
AML with inv(3)(q21.3q26.2) or t(3;3)(q21.3q26.2); *GATA2, MECOM*
AML (megakaryoblastic) with t(1;22)(p13.3;q13.3);*RBM15-MKL1*
Provisional entity: AML with BCR-ABL1
AML with mutated *NPM1*
AML with biallelic mutations of *CABPA*
Provisional entity: AML with mutated RUNX1
AML with myelodysplasia-related changes
Therapy-related myeloid neoplasms
AML, NOS
AML with minimal differentiation
AML without maturation
AML with maturation
Acute myelomonocytic leukemia
Acute monoblastic/monocytic leukemia
Pure erythroid leukemia
Acute megakaryoblastic leukemia
Acute basophilic leukemia
Acute panmyelosis with myelofibrosis
Myeloid sarcoma
Myeloid proliferations related to Down syndrome
Transient abnormal myelopoiesis (TAM)
Myeloid leukemia associated with Down syndrome
Blastic plasmacytoid dendritic cell neoplasm

TABLE 1.1 **2016 Revised WHO Classification of Acute Myeloid Leukaemia and its subtypes**
Table adapted from Blood. 2016; 127(20):2391-2405 [4]

The major difference to previously is the inclusion of genetic information with morphology, immunophenotyped and clinical presentation. Within 'AML with recurrent genetic abnormalities' there are 11 subtypes defined according to their chromosomal translocations. The preceding 2008 classification had introduced AML with mutated NPM1 (nucleophosmin) and AML with mutated *CEBPA*. The latter has been updated to reflect the added prognostic benefit of biallelic *CEBPA* mutations. Additionally, there has been the introduction of two new provisional entities: AML with BCR-ABL1 and AML with mutated RUNX1 (de novo cases of AML with RUNX1 mutation that are not associated with cytogenetic abnormalities). The former describes rare *de novo* cases of AML that may benefit from tyrosine kinase inhibitor (TKI) therapy whilst the latter classifies *de novo* AML cases with this mutation that are not associated with myelodysplasia-related cytogenetic abnormalities. Finally, a section on familial myeloid neoplasms has been included, highlighting subtypes associated with germline mutations in at least 10 genes.

In current practice, conventional cytogenetics and results from *NPM1, FLT3* and *CEBPA* mutation screening are being used to risk stratify patients newly diagnosed with AML conferring a guide to likely prognosis. With the emergence of further molecular markers (detailed in the next section) a revised genetic risk stratification has been proposed in the 2017 European LeukemiaNet (ELN) guidelines:

Risk Category	Genetic Abnormality
Favorable	▪ t(8;21)(q22;q22.1); RUNX1-RUNX1T1 ▪ inv(16)(p13.1q22) or t(16;16)(p13.1;q22); CBFB-MYH11 ▪ Mutated NPM1 without FLT3-ITD or with FLT3-ITD^{low} ▪ Biallelic mutated CEBPA
Intermediate	▪ Mutated NPM1 and FLT3-ITD^{high} ▪ Wild type NPM1 without FLT3-ITD or with FLT3-ITD^{low} (w/o adverse risk genetic lesions) ▪ t(9;11)(p21.3;q23.3); MLLT3-KMT2A ▪ Cytogenetic abnormalities not classified as favorable or adverse
Adverse	▪ t(6;9)(p23;q34.1); DEK-NUP214 ▪ t(v;11q23.3); KMT2A rearranged ▪ t(9;22)(q34.1;q11.2); BCR-ABL1 ▪ inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2); GATA2,MECOM(EVI1) ▪ -5 or del(5q); -7; -17/abn(17p) ▪ Complex karyotype, monosomal karyotype ▪ Wild type NPM1 and FLT3-ITD^{high} ▪ Mutated RUNX1 ▪ Mutated ASXL1 ▪ Mutated TP53

TABLE 1.2 **2017 Revised European LeukaemiaNet risk stratification of AML by genetics**
Table adapted from 2017 European LeukaemiaNet guidelines

These risk categories are likely to expand further in the future to incorporate more molecular markers as more clinical data becomes available regarding their prognostic impact in AML patients.

1.3 The molecular landscape of AML

As alluded to above, with the advent of new genomic techniques such as next-generation sequencing in recent years, a significant number of new genetic mutations within AML - their variety and frequency, co-operative nature or mutual exclusivity, clonal evolution and epigenetic hallmarks - have been described. A landmark study was the analysis of the genomes of 200 patients with AML (25% via whole-genome sequencing and the remainder with whole-exome sequencing, RNA and miRNA-sequencing and DNA-methylation status) provided by The Cancer Genome Atlas Research Network. [5] Frequently mutated genes in AML were organised into multiple functional categories (Figure 1.1):

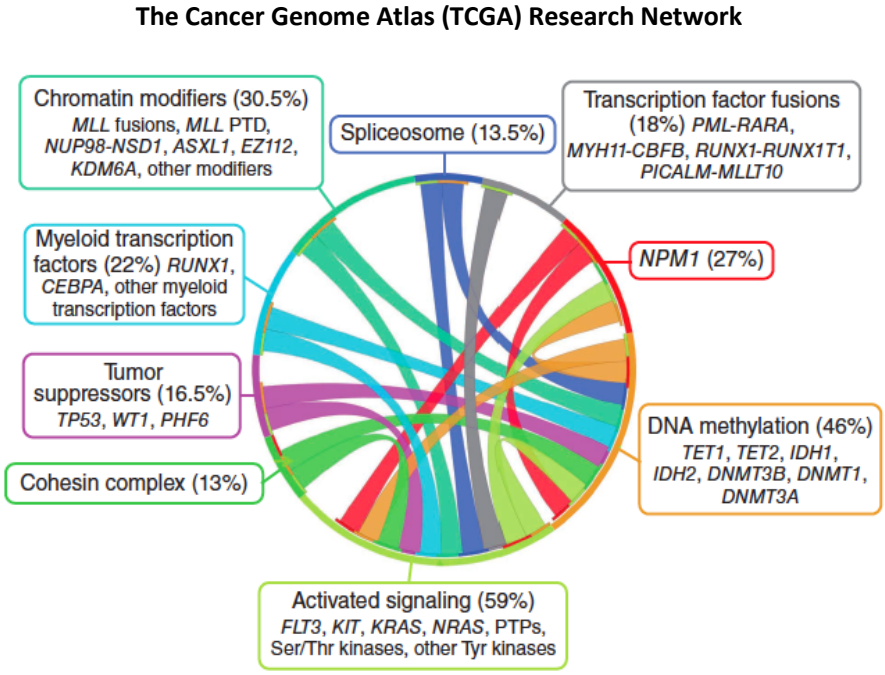


FIGURE 1.1 **Circos plot showing a panoramic view of genetic events leading to the pathogenesis of AML**
Adapted from Chen et al. *Nature Genetics* 2013 [6]

Twenty-three genes were found to be mutated at significantly higher rates than expected, including known drivers of AML such as DNMT3A, FLT3, NPM1, IDH1, IDH2, U2AF1, EZH2 and SMC1a. 99.5% of all cases had at least one important mutation. Through an analysis of variant-allele frequencies, >50%

of AML cases were made up of a founding clone and subclones, demonstrating the possibility of clonal evolution during disease progression.

Four new functional classes were identified from analysing this dataset, showing complex patterns of association. In particular, mutations in epigenetic pathways including enzymes involved in DNA methylation and chromatin modifications were shown to be present in over 70% cases.

More recently, 1540 patients intensively treated for AML across multi-centre trials were genetically profiled through cytogenetic analyses and targeted sequencing.[7] Co-mutation patterns emerged and segregated AML cases into 11 classes with distinct clinical phenotype and outcome. Three additional heterogenous classes appeared from this analysis: AML with mutations in chromatin and RNA-splicing regulators, AML with TP53 mutations and/or chromosomal aneuploidies and AML with IDH^{R172} mutations. This study categorised driver mutations and identified nonoverlapping subgroups of patients, facilitating a fully genomic classification of AML. In addition, it provided further evidence to allow mutated TP53, RUNX1 and ASXL1 to be included in the 2017 ELN recommendations adverse risk group whilst strongly putting forward the case for other molecular lesions such as DNMT3A, SRSF2, IDH2 to be considered for inclusion into risk stratification for AML patients. Through these analyses a genetic approach to classifying AML with prognostic implications was furthered, whilst obtaining similar results to the TCGA dataset. They also corroborated several epigenetic modifiers that were prominent in the TCGA analysis.

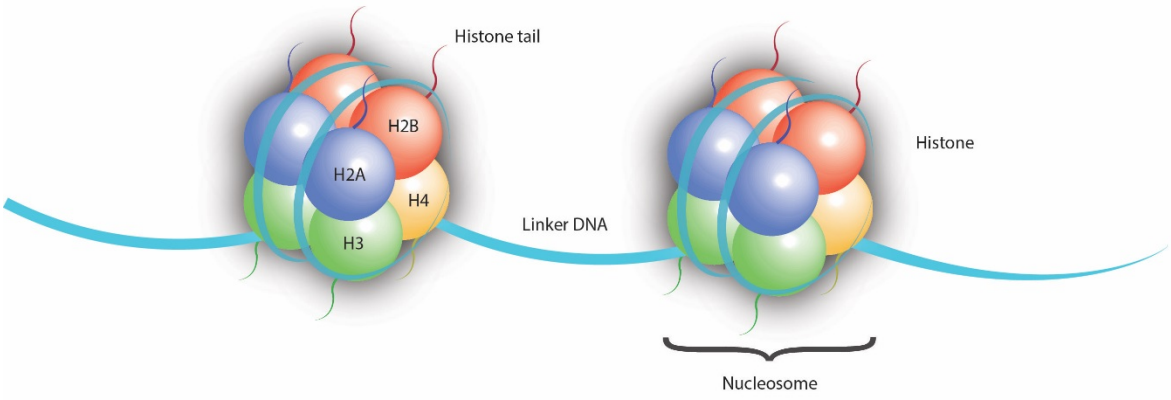
A follow up landmark study by the same authors analysed genetic data across 111 cancer genes, cytogenetic profiles and clinical information from the same 1540 AML patients. [8] Even within the 11 subgroups defined above, there was considerable variability in responses to treatment and clinical outcomes between patients. Whilst allowing a prediction for overall patterns such as survival outcomes from such data, it did not fully inform at the individual patient level. Through accumulating this molecular and clinical data as a knowledge bank, they were able to provide more accurate statements about the individual patient's likely responses through their therapy course for AML and compare likelihoods of favourable outcomes in varying treatment scenarios. This allowed for the provision of clinical advice regarding treatment that adopted a more personalised approach, particularly with respect to recommendations regarding proceeding to allogeneic transplant in first complete remission (CR1). Utilising their multistage predictive model generated from all the information provided by this detailed analysis of 1540 AML patients, and through knowledge of the complement of somatic mutations and karyotype particular to that patient, as an example, they were able to predict that an individual's risk of relapse after CR1 was considerably higher than the favourable risk score that would have been estimated by current ELN criteria. This allowed them to suggest that

allogeneic transplantation would be indicated as opposed to routine surveillance as had been suggested by the current clinical criteria. This move toward deeper characterisation at the molecular level, subsequent risk stratification and individual treatment prediction in AML suggests a future where therapeutic strategies and courses may be better informed and predicted, in order to suit the individual patient better. On a larger scale, this type of modelling also provides additional benefits for healthcare economy through being better able to predict the most appropriate therapy courses and potentially reducing costs through better resource allocation, for a condition that is also relatively expensive to treat.

The cytogenetic and molecular heterogeneity seen across AML cases has been a significant impediment to the development of effective treatments for the condition as whole. The mainstays of AML therapy (cytarabine and anthracyclines) have remained unchanged over the past 35 years. There are over 2500 deaths in the UK alone from AML each year. Whilst curable in up to 40% of patients <60 years of age, over 70% of patients diagnosed with AML eventually succumb to the condition. Outcomes in older patients remain unacceptable with median survivals of 5 to 10 months due largely to the inability to tolerate the aggressive chemotherapy regimens (that centre around the two drugs mentioned above) required for remission induction and consolidation. Through large robust studies such as those described above, the characterisation of this heterogeneous, difficult-to-treat condition is becoming more precise. The identification of critical driver lesions of AML that define subtypes followed by specific molecular mutations that may have either a prognostic impact or therapeutic implications has become increasingly important in facilitating the development of novel therapies. Novel therapeutics with lower toxicities, targeting these molecular lesions that may augment or replace standard chemotherapy would go some way to improving abysmal survival outcomes seen in older AML patients. Ongoing efforts to understand the molecular biology of several of these recently identified molecular drivers within this fascinating condition to a greater degree are most certainly required, and are paving the way for the development of important novel therapeutic strategies. As discussed above, the TCGA dataset revealed the presence of a significant number of mutations in epigenetic modifiers (>70%) in AML which has fuelled further investigation into this area in recent years.

1.4 Epigenetic regulation

Histones are a family of scaffolding proteins that associate in octamers (a tetramer of H2A and H2B plus two dimers of H3 and H4) with 147bp of DNA in the nucleus to form the nucleosome; the basic functional unit of chromatin (Figure 1.2). Chromatin facilitates compact storage of the whole 2 metres of eukaryotic cellular DNA within the nucleus and is widely recognised to exist in two states - euchromatin or heterochromatin. Euchromatin has a relaxed conformation that provides a permissive environment for active transcription, whilst heterochromatin is tightly packaged and contains predominantly inactive genes. Nucleosomes package DNA in a finely regulated manner dependent upon the cell's requirements for various dynamic DNA-templated processes. Chromatin modifying enzymes regulate post-translational modifications of histones through tightly regulated catalytic processes. Modifications to chromatin may alter function in two non-mutually exclusive ways: firstly, they may enhance or attenuate non-covalent interactions between histones and DNA, thereby governing accessibility to specific loci, and secondly the modifications that they regulate may facilitate specific recruitment of chromatin regulators to appropriate sites to alter DNA-templated processes.



Histone tail modifications

Chromatin Modification	Nomenclature	Chromatin-Reader Motif	Function
Acetylation	K-Ac	Bromodomain, PHD Fingers	Transcription, repair, replication, condensation
Methylation (lysine)	K-me1/2/3	Chromodomain, Tudor domain	Transcription and repair
Methylation (arginine)	R-me1/2s/2a	Tudor domain	Transcription
Phosphorylation (ser and thr)	S-ph, T-ph	14-3-3, BRCT	Transcription, repair and condensation
Ribosylation	E-ar	Macro, PBZ domain	Transcription and repair
SUMOylation	K-su	sumo interaction motif	Transcription and repair
Biotinylation			
Citrullination			

FIGURE 1.2 The structure of histones and a summary of common histone modifications

The term 'epigenetics' has come to encompass the general study of such chromatin based events and their regulation of DNA-based processes[9], although it was originally coined by Waddington to describe inherited changes in gene expression independent of DNA sequence changes.[10] Four different DNA modifications and at least 16 different classes of histone modification have been described.[11] These modifications are inherently malleable – they are laid down by protein complexes with catalytic function termed epigenetic "writers" and removed by other catalytic proteins known as epigenetic "erasers". The majority of histone modifications are found on the unstructured N-terminal histone tails that emanate from the globular core of the nucleosome and individual modifications include acetylation, methylation, biotinylation, citrullination, phosphorylation, ribosylation and SUMOylation.[12] Their effect is cumulative with the overall state of chromatin being dictated by combinations of these modifications. A key advance in our understanding of chromatin regulation was the recognition that most chromatin regulators contain specialised domains that confer the ability to survey the epigenetic landscape and anchor themselves to specific sites within the genome. These epigenetic "readers" can recognise and bind to post-translationally modified sites on histone tails "interpreting" the information contained in the modifications to alter locus–specific DNA functions, such as transcription, DNA repair, replication and recombination through the recruitment of specific proteins.

Aberrant regulation of these processes through the acquisition of somatic mutations in the proteins that mediate them has been increasingly recognised as a typical hallmark of leukaemia and other cancer genomes. Equally, the plasticity of the epigenome by comparison with the fixed nature of the genome, allows for its potential therapeutic manipulation in pathological conditions. Moreover, pathological dysfunction reveals mechanistic insight into the function of chromatin modifier proteins. As discussed in Section 1.3, AML is an exemplar condition in which epigenetic dysregulation has become increasingly recognised. An increasing number of epigenetic modifier/readers potentially relevant to AML pathogenesis e.g. TET2, IDH1/2, DNMT3a, ASXL1, BRD4 have been described. Recently, a better understanding of aberrant epigenetic control in AML has led to the development of promising small molecule inhibitors of epigenetic regulators. Epigenetic changes are plastic and potentially reversible, making them attractive therapeutic targets. Theoretically if a tumour-specific epigenetic state was reversible, the cancer cells within that state could re-acquire the ability to differentiate normally and surrender their proliferative advantage, restoring the balance with normal haematopoiesis.

1.5 Polycomb repressive complexes

The term *Polycomb* (Pc) referred to a Drosophila mutant with failure of normal body segmentation. [13] This mutant allowed the identification of the Polycomb Group of genes (PcG) – a group of genes that are required for normal body segmentation – their mutations result in phenotypes similar to *Polycomb*. Early embryonic lethality was observed upon deletion of genes encoding some of these proteins (*Ezh2, Eed, Suz12, Ring1b)* in mice. [14] PcG proteins are found in several families of multiproteic complexes, the most well defined of these are the Polycomb Repressive Complexes 1 and 2 (PRC1 and PRC2).

Enhancer of Zeste Homologue 2 (EZH2) functions via a catalytic "**S**uppressor of variegation 3-9, **E**nhancer of zeste and **T**ri-thorax" (SET) domain, as a histone di- and tri- methyltransferase for lysine 27 of histone 3 (H3K27me2/me3). It forms the core of PRC2 with three other obligate members (Embryonic Ectoderm Development (EED), Suppressor of Zeste 12 (SUZ12) and RbAp46/48) (Figure 1.3). Additionally, EZH1 is a non-canonical PRC2 member which complements the role of EZH2. EZH1 and EZH2 are differentially expressed with PRC2 complexes containing EZH1 instead of EZH2 displaying low methyltransferase activity relative to PRC2 containing EZH2.[15] PRC2 is well conserved throughout evolution and acts via several mechanisms to initiate polycomb-mediated gene repression, utilising direct inhibition of RNA polymerase II and chromatin compaction, which at the H3K27me3 mark, blocks access to transcription factors, and promotes transcriptional repression of genes. The H3K27me3 mark then functions as a docking site to recruit a second polycomb complex PRC1, which assists in maintaining gene repression via ubiquitination of H2AK119 (not shown). PRC1 contains 4 core components, RING2 (Ring1B in mice), BMI1, HPH, and CBX (one of various CBX proteins), and possesses E3 ubiquitin ligase activity for the mono-ubiquitylation of histone H2A at lysine 119 (H2AK119ub1). PRC1 and PRC2 occupy a common set of target sites *in vivo.* Whilst this sequential model of recruitment of PRC2 and PRC1 complexes has been well characterised [16, 17], more recently, variant PRC1 complexes containing PCGF1/3 proteins have been described - these have been shown to unexpectedly possess the ability to be recruited actively to target chromatin sites in a PRC2 independent manner. Furthermore KDM2B-mediated targeting of the variant PRC1 complexes to CpG islands was shown to be required for normal levels of PRC2 at subsets of target sites, providing an alternate mechanism of hierarchical recruitment. [18]

Early purifications of PRC2 allowed for the identification of 3 additional proteins – Aebp2, Pcl and Jarid2. AEBP2 serves to enhance enzymatic activity of the PRC2 complex and facilitates DNA binding. The PCL group (Pcl1/2/3) share similar motifs and interact with PRC2 through Ezh2 and Suz12 whilst Jarid2 belongs to the Jumonji family of proteins and facilitates PRC2 recruitment, though does not appear to enhance enzymatic function.[19]

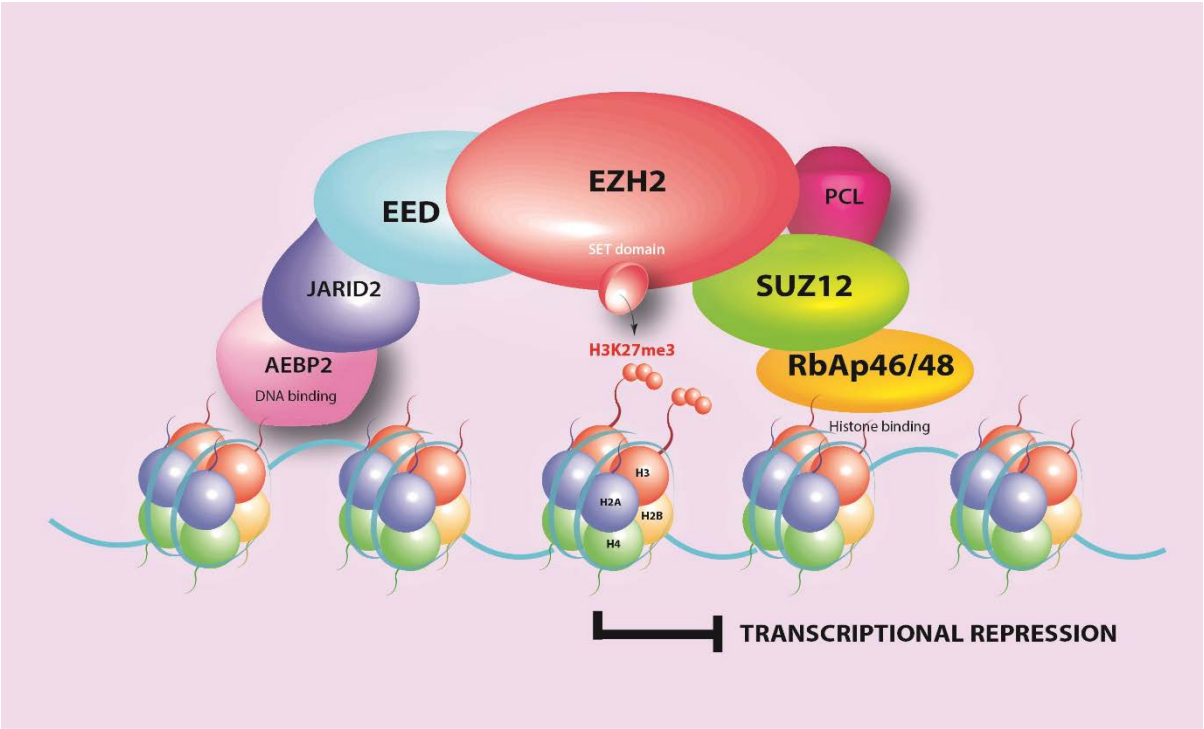


FIGURE 1.3 **The PRC2 complex**
EZH2; histone methyltransferase responsible for H3K27 di- and tri-methylation

Histone methylation occurs on the tails of arginine, lysine and histidine amino acid residues, and whilst it does not alter the overall charge of the histone, the methylation of different residues has varying effects. Methylation of H3K4, H3K36 and H3K79 is associated with gene activation in euchromatin, whilst methylation of H3K27, H3K9 and H4K20 is associated with heterochromatin formation and gene silencing. Lysine residues can be mono-, di- or tri-methylated and each methyltransferase may catalyse one or two of the transitions between these three potential states. PRC2 is the only complex that possesses enzymatic activity that results in di and tri methylation of H3K27. Enrichment of H3K27me3 correlates directly with gene silencing [20]

In mammals, binding patterns for H3K27me3 include small domains covering a few kilobases (kb) and large domains covering >100kb (e.g. those containing *Hox* loci).[21] H3K27me3 enrichment appears to be centred primarily around the transcriptional start site (TSS) at promoters with some found at intergenic regions.[22] In embryonic stem (ES) cells, chromatin organisation tends to be more open and flexible allowing for an overall higher rate of transcription. To corroborate this idea, the H3K4me3 mark associated with gene activation and transcriptional activity was found to be present at the vast majority of all PRC2 targeted genes in ES cells, demonstrating the so-called 'Bivalent Domain'. [23] Whilst initially believed to be ES-cell specific, this pattern has since been demonstrated in other differentiated

somatic cells, highlighting the importance of coexisting histone modifications in the fine tuning of gene transcription.

1.6 EZH2 in haematopoiesis and cancer

EZH2 has been demonstrated to have vital roles in stem cell development and haematopoiesis. It is essential for normal foetal haematopoiesis with Ezh2 deficient murine embryos dying of severe anaemia.[24] In the same study however, *Ezh2* deletion in adult murine bone marrow (BM) did not significantly compromise haematopoietic stem cell (HSC) function, importantly highlighting its potential as a target in haematological disease. EZH2 contributes to the regulation of cell fate decisions (as alluded to above), controlling gene expression to settle upon a balance between self-renewal and differentiation.[25] EZH2 has a predominant role in embryonic tissues and becomes downregulated in some adult differentiated tissues. In haematopoiesis, it has been shown to stabilise chromatin structure to maintain the long-term self-renewal potential of HSCs through the repression of pro-differentiation genes.[26]

Aberrant EZH2 expression and activity have been linked to tumorigenesis, through its ability to block genes necessary for differentiation, allowing retention of a cancer stem cell phenotype. This is recognised in several cancer models, including breast and prostate.[27-29] These studies demonstrated a clear correlation in both transcript and protein levels with poorer outcomes in these malignancies, also showing promoted tumour progression and metastasis.

Outwith its increased expression at both transcript and protein levels, gain-of function mutations of EZH2 have been described. Within lymphoma, a landmark study described a cohort of patients with diffuse large B-cell lymphoma (DLBCL) and follicular lymphoma (FL) with *EZH2* variants. [30] Through whole genome DNA sequencing of an index case of FL they demonstrated a mutation affecting exon 15 of the *EZH2* gene, encoding part of the catalytic SET domain. This resulted in the substitution of tyrosine to histidine at codon 641 (Y641). Upon screening further DLBCL and FL cases, the same point mutation was revealed in approximately 21% and 7% of DLBCL (exclusively in the germinal centre subtype) and FLs respectively. Data from a subsequent study demonstrated that the Y641 mutation in EZH2 conferred an aberrant functional interaction between mutant and WT gene products.[31] The malignant phenotype arising in Y641 mutated B-cell lymphoma was shown to occur due to an altered cooperation between mutated and WT EZH2 with an over-all hyper-trimethylating effect at H3K27, resulting in gene repression. In succession to these and other studies, small molecule inhibitors of EZH2 were developed to target this mutation in lymphoma. Utilising GSK126 (a novel highly selective EZH2

methyltransferase inhibitor) McCabe et al. demonstrated reduced global H3K27me3 levels in EZH2 mutant DLBCL cell lines, which resulted in inhibition of proliferation of these cells.[32] Furthermore, the compound significantly inhibited growth of EZH2 mutant DLBCL xenografts in mice. This has led to a demonstrable reduction of cancer stem cell renewal capability in lymphomas, with this and other agents (e.g. EZH2 inhibitors EPZ-6438, CPI-1205) moving into the clinical arena via ongoing early phase trials. [NCT Trial numbers: NCT02082977, NCT01897571, NCT02395601 for compounds GSK2816126, EPZ-6438 and CPI-1205 respectively]

1.7 EZH2 in myeloid malignancy

The role of EZH2 in myeloid malignancies, by comparison, is less well-defined. Although loss of function mutations have been described at low frequency in myeloid malignancies including myelodysplasia (MDS) and AML[33], evidence for overactivity of EZH2 in established myeloid leukaemia also exists. Another study demonstrated *Ezh2* loss contributed to development of a *RUNX1* mutant driven MDS in mice, however prevented its transformation to AML through PRC1-mediated repression of Hoxa9.[34]

AML associated with the rearrangement of the mixed lineage leukaemia (MLL) gene has been genetically demonstrated, in two separate studies, to rely on *Ezh2* for leukaemia maintenance.[35, 36] AML cases with translocations of the MLL gene occur in around 70% of infantile leukaemias and are characterised by aggressive leukaemias often refractory to conventional therapies. MLL is a histone methyl-transferase containing a SET (Suppressor variant 3-9, enhancer of Zeste, trithorax) domain and is involved in tri-methylation of lysine 4 on histone H3, leading to transcriptional activation. In MLL rearranged leukaemias the N-terminus of MLL is fused to one of over 70 fusion partner genes.[37] Most of these result in loss of activity at the SET domain but activate pro-leukaemogenic changes in gene expression.

In the first study[35], conditional deletion of *Ezh2* in granulocyte-macrophage progenitors already transformed by the MLL-AF9 fusion compromised growth severely *in vitro* and delayed progression of AML *in vivo*. Through genomic analysis, a significant reduction in H3K27me3 in *Ezh2*-deficient transformed cells was seen at several genes important for developmental and differentiation processes. This included the Ink4a/Arf tumor suppressor locus (also known as *CDKN2A*) which was significantly de-repressed upon *Ezh2* loss as well as *Egr1* (a positive regulator of myeloid differentiation), the latter of which was functionally validated through overexpression *in vitro* demonstrating a significant drive toward differentiation of these MLL-AF9 transformed cells. In the second study[36], conditionally *Ezh2*-deleted MLL-AF9 driven AMLs failed to accelerate upon secondary

transplantation *in vivo* complementing the findings of the first study, however H3K27me3 levels persisted at specific loci, suggesting *Ezh1* was likely partially compensating. Ablation of *Eed* however led to a complete ablation of PRC2 function and did not allow leukaemic growth. Both studies suggested that Ezh2/PRC2 function is required for the maintenance of MLL-AF9 driven AML.

More recently, and somewhat contrary to the oncogenic role of *Ezh2* in myeloid malignancies described above, early *Ezh2* loss has been shown to co-operate with *Jak2*-V617F driven myeloproliferative neoplasms, markedly advancing myelofibrosis *in vivo* and reducing survival.[38] Through further genomic analysis in this study, several highly-upregulated target genes from HSCs in the *Ezh2* null state were described, including *Lin28b* and *Hmga2.* These are well described oncogenes (discussed later in this thesis) with a clear association to other advanced human malignancies.[39] Overexpression of the latter (*Hmga2*), but not the former, partly recapitulated the accelerated synergistic myelofibrotic phenotype *in vitro* and *in vivo*. A similar study demonstrated synergistic effects between *Ezh2* loss and *Jak2*-V617F driven myelofibrosis *in vivo* with de-repression of several other targets, again including *Hmga2,* in the *Ezh2* deficient state.[40] Overexpression of these *in vitro* using *Jak2*-V617F murine bone marrow significantly increased clonogenic output, once again providing evidence for its role in mediating the synergism between *Ezh2* loss and *Jak2*-V617F.

These clearly disparate findings regarding EZH2's oncogenic or tumour-suppressive function in cancers and their subtypes, particularly within myeloid malignancies, suggest that further studies are required to delineate its role in greater detail. Being an epigenetic regulator responsible for gene silencing at a huge variety of loci, provides an added complexity. This warrants further investigation to try and reconcile the protein's apparent ability to induce both pro-oncogenic and tumour-suppressive effects. Given EZH2 inhibitors are currently in pre-clinical and early phase clinical trials, this is particularly required as inhibition of this protein could be potentially beneficial or detrimental depending on context.

1.8 H3K27me3 in AML

Work relevant to this thesis performed by Drs Mike Chapman and Emma Gudgin in the Huntly lab (unpublished data) just prior to this dissertation, has also indirectly suggested EZH2 dysfunction in AML. To characterise histone modification patterns across AML patient samples, the repressive H3K27me3 (catalysed by EZH2) and the activating mark H3K9/14 acetylation were assessed for approximately 300 genes known to be important in leukaemogenesis and haematopoiesis by Chromatin Immuno-Precipitation followed by hybridisation to microarrays ("ChIP on CHIP"). The rationale was that investigating a combination of activating and inactivating histone modifications in the chromatin landscape of AML versus control populations may provide additional information regarding patterns/mechanisms of dysregulation compared to the direct readout provided by gene expression profiling. Chromatin from 53 AML patients and 7 AML cell lines were examined, using normal CD34+ haematopoietic stem and progenitor cells (HSPC), T cells and neutrophils as controls. Following bioinformatics analysis, a variety of subpatterns were demonstrated. The most striking observation however, was that this analysis dichotomised patients into two groups, one that either demonstrated an increased H3K27 methylation and a decreased H3K 9/14 acetylation pattern – a gene "repressive" status – and vice versa (a gene "activated" status) (Figure 1.4). This suggested that overactivity of EZH2 and an H3K27me3 repressive pattern may be involved in the pathogenesis of certain subtypes of AML.

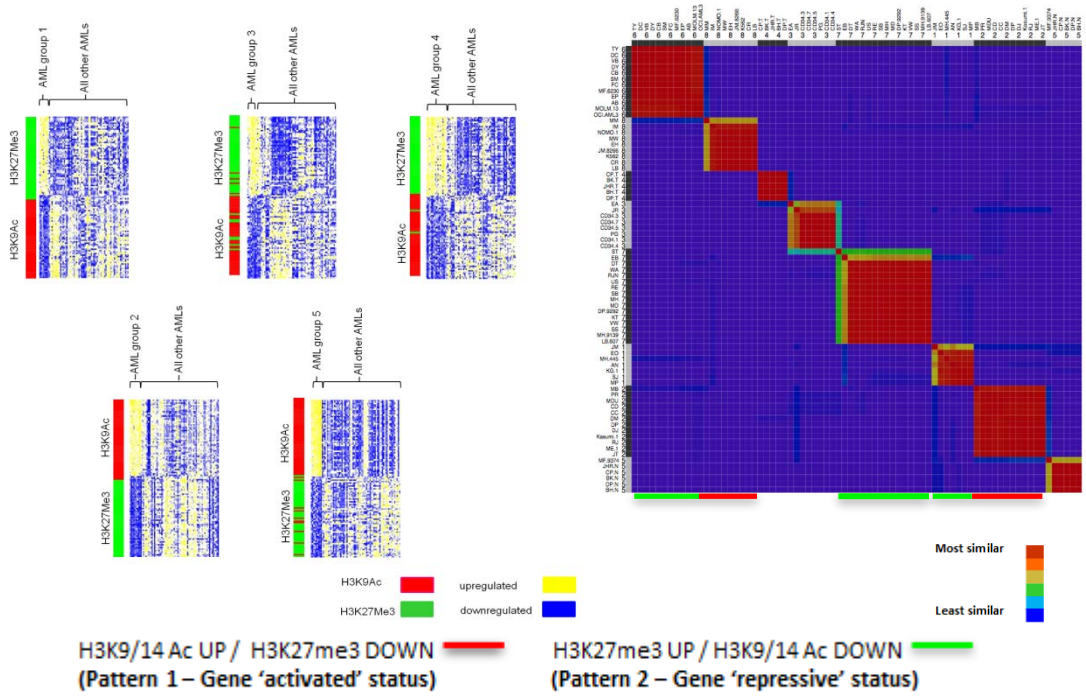


FIGURE 1.4 **Bioinformatics analysis of chromatin activation/repression states in AML patients and AML cell lines suggests patterns of EZH2 overactivity**

The prognostic significance of these AML patient epigenetic profiles was assessed (Figure 1.5). Though there was no significant overall survival difference between the two curves, there was a difference in median survival between the two states, with the "activated" group having a median survival of 21.4 months versus the "repressive" groups at 12.7 months median survival. This analysis was limited due to the small numbers involved (n=43, only patients enrolled in MRC trials were included), but again correlated with the hypothesis that EZH2 overactivity may be of pathogenetic relevance in AML.

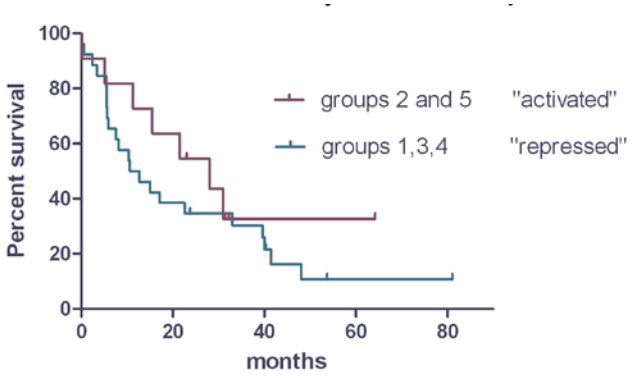


FIGURE 1.5 **A "repressed" state (upregulated H3K27me3 + downregulated H3K9/14Ac) correlates with poorer survival**

1.9 Thesis aims

The central aim of this project was to define the role of EZH2 in various stages of AML evolution. In particular, the aim was to delineate the role of EZH2 in the induction and maintenance of multiple AML subtypes through *in vitro* and *in vivo* experiments, utilising a conditional *Ezh2* knockout system. Following this, if any differences in outcomes were demonstrated, a detailed genomic analysis was proposed utilising gene expression profiling through RNA-sequencing between *Ezh2* wild-type (*Ezh2*$^{+/+}$ or wt) and *Ezh2* null (*Ezh2*$^{-/-}$) transformed or leukaemic states at various stages of AML evolution. Through Chromatin Immunoprecipitation coupled to next generation sequencing (ChIP-seq) a detailed examination of alterations in the H3K27me3 repressive mark, along with other histone activating marks and localisation of the PRC2 complex in the above conditions was also planned.

Through robustly combining these data, the aim was then to identify any targets deregulated between the two Ezh2 states that would contribute to phenotype and functionally assess these either *in vitro* or *in vivo*.

Lastly through collaboration with GlaxoSmithKline – GSK343, an active novel small molecule inhibitor of EZH2 (and its inactive homologue GSK669) were provided to the Huntly lab. The final aim was to test this inhibitor across a range of AML cell lines, murine AML tumours *in vitro* and primary AML patient samples, to assess for sensitivity/resistance, with an investigation to delineate mechanism or sensitivity through genomic analysis as discussed above.

This dissertation has thus been divided into five experimental chapters and addresses the following questions:

1. **Is *Ezh2* genetically required for the induction and maintenance of subtypes of AML *in vitro* and *in vivo*?**

2. **What underlying transcriptional programmes are subverted by *Ezh2* loss during induction or maintenance of AML to account for differences in phenotypic outcomes?**

3. **Do candidate genes (identified in Aim 2) provide a mechanistic insight to differences seen in Aim 1? Can these be functionally validated?**

4. **Is EZH2 a plausible therapeutic target (as informed by Aims 1-3) and does pharmacological inhibition with a novel EZH2 inhibitor have a role in AML?**

References

1) Cancer Research UK, Acute Myeloid Leukaemia statistics http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/leukaemia-aml

2) Godley LA. Inherited predisposition to acute myeloid leukemia. *Semin Hematol* 2014; 51: 306-21.

3) Polprasert C, Schulze I, Sekeres MA, et al. Inherited and somatic defects in DDX41 in myeloid neoplasms. *Cancer Cell* 2015; 27: 658-70.

4) Daniel A. Arber, Attilio Orazi, Robert Hasserjian, Jurgen Thiele, Michael J. Borowitz, Michelle M. Le Beau, Clara D. Bloomfield, Mario Cazzola and James W. Vardiman. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*. 2016; 127(20):2391-2405.

5) The Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 2013; 368: 2059-74.

6) Chen SJ, Shen Y, Chen Z. A panoramic view of acute myeloid leukemia. *Nat Genet.* 2013 Jun;45(6):586-7.

7) Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, Potter NE, Heuser M, Thol F, Bolli N, Gundem G, Van Loo P, Martincorena I, Ganly P, Mudie L, McLaren S, O'Meara S, Raine K, Jones DR, Teague JW, Butler AP, Greaves MF, Ganser A, Döhner K, Schlenk RF, Döhner H, Campbell PJ. Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med.* 2016 ;374(23):2209-21.

8) Gerstung M, Papaemmanuil E, Martincorena I, Bullinger L, Gaidzik VI, Paschka P, Heuser M, Thol F, Bolli N, Ganly P, Ganser A, McDermott U, Döhner K, Schlenk RF, Döhner H, Campbell PJ. Precision oncology for acute myeloid leukemia using a knowledge bank approach. *Nat Genet.* 2017 ;49(3):332-340.

9) Dawson, M.A and Kouzarides, T. Cancer epigenetics: from mechanism to therapy. *Cell.* 2012 150, 12–27.

10) Waddington, C.H. The Epigenotype. (1942) *Int J Epidemiol.* 41: 10-13.

11) Dawson, M.A., Kouzarides, T., and Huntly, B.J.P (2012) Targeting Epigenetic Readers in Cancer. *N Engl J Med.* 367: 647-657.

12) Turner, B. Reading signals on the nucleosome with a new nomenclature for modified histones. (2005) *Nat Struct Mol Biol.* 12(2), 110–112.

13) Lewis P. Pc: Polycomb. Drosophila Information Service. 1949; 21:69.

14) Schuettengruber B, Cavalli G. Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice. *Development.* 2009; 136:3531–3542.

15) Margueron R, et al. Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Molecular Cell.* 2008; 32:503–518.

16) Ku M, et al. Genome wide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet.* 2008;4.

17) Papp B., Müller J. Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. Genes Dev. 2006; 20:2041–2054.

18) Blackledge NP, et al. Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell.* 2014; 157(6):1445-59.

19) Shen X, et al. Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. *Cell*. 2009; 139:1303–1314.

20) Barski A, et al. High-resolution profiling of histone methylations in the human genome. *Cell*. 2007; 129:823–837.

21) Zhao XD, et al. Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. *Cell stem cell*. 2007; 1:286–298.

22) Bernstein BE, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*. 2006; 125:315–326.

23) Pan G, et al. Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell stem cell*. 2007; 1:299–312.

24) Mochizuki-Kashio M et al. Dependency on the polycomb gene Ezh2 distinguishes fetal from adult hematopoietic stem cells. *Blood.* 2011 ;118(25):6553-61

25) Chou RH, Yu YL, Hung MC. The roles of EZH2 in cell lineage commitment. *Am J Transl Res.* 2011; 3: 243–250

26) Kamminga LM, Bystrykh LV, de Boer A, Houwer S, Douma J, Weersing E et al. The Polycomb group gene Ezh2 prevents hematopoietic stem cell exhaustion. *Blood.* 2006; 107: 2170–2179.

27) Varambally S, Dhanasekaran SM, Zhou M, Barrette TR, Kumar-Sinha C, Sanda MG, Ghosh D, Pienta KJ, Sewalt RG, Otte AP, Rubin MA, Chinnaiyan AM. The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature*. 2002 ;419(6907):624-9

28) Kleer CG, Cao Q, Varambally S, Shen R, Ota I, Tomlins SA, Ghosh D, Sewalt RG, Otte AP, Hayes DF, Sabel MS, Livant D, Weiss SJ, Rubin MA, Chinnaiyan AM. EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc Natl Acad Sci USA* 2003 ;100(20):11606-11

29) Wagener N, Holland D, Bulkescher J, Crnković-Mertens I, Hoppe-Seyler K, Zentgraf H, Pritsch M, Buse S, Pfitzenmaier J, Haferkamp A, Hohenfellner M, Hoppe-Seyler F. The enhancer of zeste homolog 2 gene contributes to cell proliferation and apoptosis resistance in renal cell carcinoma cells. *Int J Cancer*. 2008 ;123(7):1545-50

30) Morin RD, *et al*. Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat Genetics.* 2010 ;42(2):181-5

31) Sneeringer CJ, Scott MP, Kuntz KW, Knutson SK, Pollock RM, Richon VM et al. Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. *Proc Natl Acad Sci USA* 2010; 107: 20980–20985.

32) McCabe MT, Ott HM, Ganji G, Korenchuk S, Thompson C, Van Aller GS, Liu Y, Graves AP, Della Pietra A 3rd, Diaz E, LaFrance LV, Mellinger M, Duquenne C, Tian X, Kruger RG, McHugh CF, Brandt M, Miller WH, Dhanak D, Verma SK, Tummino PJ, Creasy CL. EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature*, 2012 ;492(7427):108-12

33) Lund K, Adams PD, Copland M. EZH2 in normal and malignant hematopoiesis. *Leukemia*, 2013 ;28(1):44-9

34) Sashida G, Harada H, Matsui H, Oshima M, Yui M, Harada Y, Tanaka S, Mochizuki-Kashio M, Wang C, Saraya A, Muto T, Hayashi Y, Suzuki K, Nakajima H, Inaba T, Koseki H, Huang G, Kitamura T, Iwama A. *Ezh2* loss promotes development of myelodysplastic syndrome but attenuates its predisposition to leukaemic transformation. *Nat Commun.* 2014 Jun;5:4177

35) Tanaka S, Miyagi S, Sashida G, Chiba T, Yuan J, Mochizuki-Kashio M, Suzuki Y, Sugano S, Nakaseko C, Yokote K, Koseki H, Iwama A. *Ezh2* augments leukemogenicity by reinforcing differentiation blockage in acute myeloid leukemia. *Blood*, 2012 Aug 2;120(5):1107-17

36) Neff, T. *et al.* Polycomb repressive complex 2 is required for MLL-AF9 leukemia. *Proc Natl Acad Sci USA* 109, 5028–33 (2012).

37) Krivtsov AV, Armstrong SA. MLL translocations, histone modifications and leukaemia stem-cell development. *Nat Rev Cancer.* 2007 Nov;7(11):823-33.

38) Shimizu T, Kubovcakova L, Nienhold R, Zmajkovic J, Meyer SC, Hao-Shen H, Geier F, Dirnhofer S, Guglielmelli P, Vannucchi AM, Feenstra JD, Kralovics R, Orkin SH, Skoda RC. Loss of *Ezh2* synergizes with JAK2-V617F in initiating myeloproliferative neoplasms and promoting myelofibrosis*. J Exp Med.* 2016 Jul;213(8):1479-96.

39) Viswanathan SR, Powers JT, Einhorn W, Hoshida Y, Ng TL, Toffanin S, O'Sullivan M, Lu J, Phillips LA, Lockhart VL, Shah SP, Tanwar PS, Mermel CH, Beroukhim R, Azam M, Teixeira J, Meyerson M, Hughes TP, Llovet JM, Radich J, Mullighan CG, Golub TR, Sorensen PH, Daley GQ. Lin28 promotes transformation and is associated with advanced human malignancies. *Nat Genet.* 2009 Jul;41(7):843-8.

40) Yang Y, Akada H, Nath D, Hutchison RE, Mohi G. Loss of *Ezh2* cooperates with Jak2V617F in the development of myelofibrosis in a mouse model of myeloproliferative neoplasm. *Blood*. 2016 Jun 30;127(26):3410-23.

Chapter 2

Materials and Methods

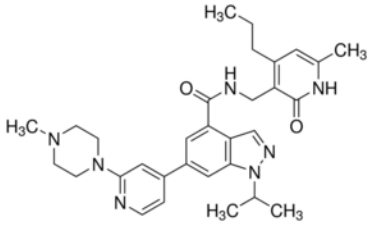
2.1 Materials

This chapter contains a description of all materials and protocols for the experimental work described in this dissertation.

2.1.1 General laboratory reagents

β-Mercaptoethanol	Sigma Aldrich	M6250
5X DNA loading buffer	Bioline	BIO-37045
Bovine serum albumin	Sigma Aldrich	A2153
Dimethyl sulfoxide (DMSO)	Sigma Aldrich	276855
Ethanol	Sigma Aldrich	
Foetal bovine serum	Sigma Aldrich	F6178
Hexadimethrine bromide (Polybrene)	Sigma Aldrich	H9268
Isopropanol	Sigma Aldrich	
L-glutamine	Sigma Aldrich	59202C
Lymphoprep™	Axis-Shield, Norway	1114547
Methanol	Fisher Chemicals	
Methocult Methylcellulose Human	Stem Cell Technologies	H4435
Methocult Methylcellulose Murine	Stem Cell Technologies	M3434
Opti-MEM reduced serum medium	ThermoFisher scientific	31985062
Penicillin/Streptomycin	Sigma Aldrich	P4333
Propidium iodide	Sigma Aldrich	P4170
Protease inhibitor	Sigma Aldrich	11 873 580 001
Puromycin	Thermo ScientifiC	A1113803
Red cell lysis buffer	5 Prime	2301310
TransIT-LT1 transfection reagent	Mirus Bio	MIR2300
Trypan Blue	Sigma Aldrich	T8154

2.1.2 GSK343: EZH2 inhibitor



GSK343: EZH2 inhibitor
60-fold selective: EZH2:EZH1
1000-fold selective: other methyltransferases

FIGURE 2.1 GSK343 structure

GSK343 was initially provided by GlaxoSmithKline through the specifications of a Materials Transfer Agreement (MTA) and later purchased through Sigma Aldrich (Sigma, Catalogue No: SML0766). Both were in powder form and this was dissolved in an appropriate volume of dimethyl sulfoxide (DMSO), (Sigma, Catalogue No: 276855) to provide a 10mM solution and stored in aliquots at -80^{o}C for use. Neat DMSO served as controls in subsequent experiments. GlaxoSmithKline also provided an inactive version of GSK343 entitled GSK669 that was again used as an additional control in some experiments. Each experiment used a freshly thawed aliquot of compound to avoid potential reduction in efficacy from repeated freeze-thawing.

2.2 Tissue Culture

All tissue culture was performed under Category 1 conditions, with cells being incubated at 37^{o}C in 5% CO_2 unless otherwise specified.

2.2.1 Cell lines

2.2.1.1 AML Cell lines

Human AML cells lines detailed below (Table 2.1) were typically maintained at a density of 0.5-2x10^{6}/ml according to optimal conditions set out by DSMZ (Deutsche Sammlung von Mikroorganismen und Zellkulturen) GmbH where they were purchased from.

Cell Line	Species	Disease	Mutation/Disease	Culture conditions
Kasumi	Human	AML	AML1-ETO	RPMI 20% FCS
KG1	Human	AML	FGFR1OP2-FGFR1	RPMI 10% FCS
K562	Human	AML	BCR-ABL	RPMI 10% FCS
NB4	Human	APML	PML-RARa	RPMI 20% FCS

TABLE 2.1 Human AML cell lines

2.2.1.2 293T cells

293T Cells were used to produce retroviral supernatant. They are a variant of the human embryonic kidney 293 cell line, having been isolated from foetal kidney cells and transformed with type 5 adenovirus. They contain a temperature sensitive SV40 large T antigen. Routine culture conditions for 293T cells were at 37 ^{o}C, 5% CO_2 in DMEM (Life Technologies, Cat No: 11965-084) + 10% heat-inactivated Fetal Bovine Serum (Sigma, Cat No: 12003C) + L-glutamine (Sigma, Cat No: G75713), being maintained at 50-70% confluency across 100mm sterile plates. To ensure no reduction in supernatant quality, passages were kept to a minimum by freezing large numbers of vials and thawing as required.

2.2.2 In vitro culture and passage of murine MLL-AF9 AML spleen tumours

All murine leukaemias generated from the *in vivo* studies described in this thesis had bone marrow and spleen tissues processed into single-cell suspensions and frozen as described in Methods 2.6.10. In order to passage these *in vitro*, they were thawed under sterile conditions as described in (Cell Storage). The splenic tumour cells were then resuspended in sterile X-Vivo 20 culture medium (Lonza, Cat No: 04-448Q) supplemented with mIL-3, IL-6 and mSCF cytokines, (Peprotech^{TM}, Catalogue Numbers: 213-03, 200-06 and 250-03, at final concentrations of 10ng/ml, 10ng/ml and 50ng/ml respectively) at approximately 0.2-0.3x10^6 cells/ml. These tumours were then passaged every 48-72 hours *in vitro* under standard tissue culture conditions in 10ml Falcon flasks and expanded. Initial GSK343 titration to assess sensitivity to GSK343, cell proliferation and viability quantifications and clonogenic methylcellulose assays were performed as described in Methods 2.3.2 to 2.3.3.

2.2.3 Primary AML sample collection/processing

Peripheral blood/bone marrow samples were collected from newly diagnosed/relapsed patients with AML by WHO Criteria[1] into cytogenetic media (RPMI supplemented with HEPES, lithium heparin and gentamicin). This was done as part of "The Causes of Clonal Blood cell disorders" study (Department of Haematology, University of Cambridge) approved by the Cambridge and Eastern Region Ethics Committee. Patients gave written informed consent and research was carried out in accordance with the Declaration of Helsinki. All samples were anonymised and were identifiable only by a tissue bank code. The mononuclear fraction was obtained following a 1:1 dilution in PBS and layering onto 1:1 volume of Lymphoprep (Axis-Shield Norway, Catalogue No: 1114544). Samples were centrifuged for 45min at 1400rpm and the interface layer between plasma and red cells containing the mononuclear fraction was carefully removed using a 1ml pipette. Finally, after washing in MACS buffer (500 ml PBS without Ca/Mg, 2.5 g BSA and 2 ml of 0.5 M EDTA, filter sterilised), cells were counted for use/storage.

2.2.4 Cell storage

2.2.4.1 Freezing

Human AML cell lines, murine cell lines and tumour cells and primary AML samples were counted and then centrifuged for 5 minutes at 1200rpm at 4^{o}C and the supernatant removed. 1x10^{7} – 1x10^{8} cells were then resuspended in 1ml of freezing medium (90% FCS + 10% DMSO) and aliquots in cryovials were made, placed into Nalgene 1^{o}C/min freezing containers (Sigma) and stored at -80^{o}C until completely frozen. These were subsequently transferred to liquid nitrogen tanks for longer term freezing at -180^{o}C.

2.2.4.2 Thawing

Frozen cells in cryovials were warmed in a 37^{o}C water bath and immediately diluted with 10ml of appropriate media. After centrifugation for 5 minutes at 1200rpm, the supernatant was removed and the cells resuspended in appropriate media, counted and cultured at 0.5-1x10^{6}/ml overnight or plated at the appropriate number into semi-solid methylcellulose for functional assays.

2.2.5 Cell counting

Counting was performed using a CASY Counter (Schärfe System GmbH), which generates a cell size profile and total viable and dead cell number (cells below average nuclear size indicated death/dying and were excluded). This method uses electric current exclusion and pulse area analysis to determine viable cell numbers. Under a low voltage field, electric current cannot pass through the intact cell membrane of a viable cell, whilst it can through the pores of a dead/dying cell, thereby allowing a means of differentiating the two.

2.3 Functional assays utilising tissue culture

2.3.1 Liquid culture growth assay

This 12 day protocol required Kasumi and KG1 cell lines maintained in liquid culture media (RPMI + 20% FCS) to be expanded in the presence of DMSO, GSK343 or GSK669 (refreshed every 72 hours) with total cell counts measured every 72 hours.

At Day 0, $0.5x10^6$/ml of Kasumi and KG1 cells in fresh media (RPMI + 20% FCS) were initially placed in a 12 well plate at 1ml/well. To each well either 1ul DMSO, GSK343 or GSK669 (at 10mM) were added. By pipetting up and down repeatedly, each well was mixed thoroughly and the plates were incubated under standard mammalian cell culture conditions (37^oC, 5% CO2).

At Day 3, each well was pipetted up and down several times to re-establish a single cell suspension and the cells were counted. Upon establishing total viable cell count, these were then transferred to Eppendorf tubes and centrifuged at 1200rpm for 7 minutes. They were then resuspended in the appropriate volume of fresh media (plus the appropriate volume of DMSO, GSK343 or GSK669), to return the concentration back to $0.5x10^6$/ml. Where cell growth had occurred, this inevitably resulted in an expanded volume after counting and resuspending, therefore the cells were subsequently stored in individual BD Falcon 5ml flasks and returned to the incubator.

This process was repeated at Days 6, 9 and 12, with increasing volumes of media (plus final concentrations of DMSO, GSK343 or GSK669 at 10μM). At each time point the total viable cell number was recorded and then plotted against time.

2.3.2 Clonogenic assays in methylcellulose with GSK343

Clonogenic potential was assessed through colony growth (colony numbers and total cell numbers) in the presence of DMSO, GSK343 or GSK669. The same method was used for human AML cell lines, immortalised murine bone marrow (BM) c-kit selected cells, murine MLL-AF9 spleen tumours cultured *in vitro*, and for primary AML samples. An aliquot of each of the AML cell lines or murine MLL-AF9 leukaemia spleen tumours was thawed (as per Methods 2.2.4.2) and then cultured in appropriate media (Table 2.1 for AML cell lines and Section 2.2.2 for murine MLL-AF9 tumours). All primary samples were obtained fresh from the patient (the same day) with written consent and the mononuclear cell fraction was resuspended in MACS buffer at ~$1x10^6$/ml (Section 2.2.3 primary samples).

The cells were counted and the volume containing the appropriate number of cells for plating (as per Table 2.2) was calculated. This was then diluted in RPMI 1640 + 20% FCS media to a final volume of 300μl and added to 3ml aliquots of freshly thawed and vigorously vortexed methylcellulose (Methocult H4435 containing human cytokines or Methocult M3434 containing murine cytokines, Stem Cell Technologies). At this stage, either 3μl DMSO and 3μl of 10mM GSK343 or GSK669 were added to each 3ml mixture of cells + methylcellulose and once again vortexed thoroughly. This achieved a final concentration of 10μM of GSK343 in the methylcellulose. The mixture was divided into 1ml aliquots in duplicate 35mm gridded tissue culture dishes using a 2ml syringe (BD Biosciences) with a 19G needle. These were then placed in a larger 150mm tissue culture dish with an additional 35mm dish containing PBS to maintain humidity. They were grown for 7-10 days under standard tissue culture conditions. At this stage, the colonies were counted using an inversion microscope, with a colony being reproducibly and consistently defined as ≥ 20 cells with a convincing core.

Cell type	Total number of cells plated per 3ml methylcellulose	Methylcellulose type
AML cell lines	$3x10^4$	Methocult H4435
Murine transformed c-kit+ bone marrow cells	$1\text{-}1.5x10^5$	Methocult M3434
Murine MLL-AF9 leukaemia spleen tumour cells cultured *in vitro*	$1.5x10^5$	Methocult M3434
Primary AML samples	$2.4x10^5$	Methocult H4435

TABLE 2.2 Numbers of cells plated for clonogenic methylcellulose assays

2.3.3 Cell proliferation and viability quantification

Using CellTiter 96® AQueous One Solution Cell Proliferation Assay (Promega G3582), the antiproliferative and cytotoxic effects of GSK343 were assayed. This product contains MTS, a tetrazolium compound [3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium]. MTS is reduced by viable, metabolically active cells into a coloured, soluble formazan product that allows for colorimetric detection. $2x10^4$ cells in 100μl media were seeded into each well of a 96-well plate. GSK343 (solubilised in DMSO) was added to the wells at 10 different concentrations reducing in a semi-logarithmic manner from 100μM to 100nM. The dilutions were created ensuring a final concentration of 1% DMSO within each condition to avoid varying levels of DMSO influencing cell proliferation/toxicity. Following 72 hours incubation, 40μl of 96® AQueous One Solution was added to each well with further incubation for 3 hours. This was then transferred to an optical plate reader and read at 490nm (excitation). Proliferation curves and IC50 values were generated using Prism® statistical software.

2.4 Flow cytometry

All flow cytometry was performed in the Cambridge Institute for Medical Research (CIMR) Flow cytometry core facility.

2.4.1 Apoptosis and cell cycle analysis by flow cytometry

A time course over 96 hours of cellular responses to DMSO, GSK343 and GSK669 (all at 10μM) was set up in liquid culture. At each time point of 24, 48, 72 and 96hrs post addition of drug to 2x10^6 cells cultured in 4ml media on 6 well plate, 500μl of cells were taken and cellular apoptosis was assessed using the Annexin V-FITC kit (Miltenyi Biotec, Catalogue No: 130-092-052), supplemented with 7-AAD (BD Bioscience, Catalogue No: 559925) as per manufacturer's instructions. Another 500μl of cells were taken for cell cycle analysis. These were washed in PBS, fixed in 70% ethanol in PBS and kept at -20°C. Following this, they were washed in PBS and resuspended in 50μg/ml propidium iodide and 0.1mg/ml RNase. Finally the samples were incubated at 37°C for 30 minutes. Flow cytometry was performed on a BD LSRFortessa 4-laser machine using FACSDiva analysis software. Data was then analysed with FlowJo software (Tree Star Inc.).

2.4.2 H3K27me3 assessment by flow cytometry

To assess for loss of H3K27me3 following Ezh2 inhibition with GSK343, *Ezh2fl/fl*; Cre wild-type MLL-AF9 AML spleen tumours (#5340 and #5342) generated from the induction *in vivo* experiments were thawed and passaged as above. $8x10^6$ cells from two MLL-AF9 spleen tumours (biological duplicates) were resuspended in pairs in X-Vivo 20 medium supplemented with cytokines (as per Methods 2.2.2) at $0.5x10^6$/ml concentrations at Day 0. At 0hr, DMSO and GSK343 (final concentration 10μM) were added to each MLL-AF9 tumour. At 24hr, 48hr and 72hrs approximately $2x10^6$ cells were taken from each of the DMSO and GSK343 treated conditions – for fixation and staining with:

a) H3K27me3 antibody conjugated to Alexa Fluor 647 fluorochrome (Abcam, Cat No: ab205729) and
b) Total Histone H3 antibody conjugated to Pacific-Blue fluorochrome (Cell Signaling technology, Cat No: 12167)

At each time point, the remaining cells were washed and returned to $0.5x10^6$/ml concentrations with fresh media and DMSO or GSK343 (10μM)

2.4.2.1 Fixation and permeabilization

$1x10^6$ cells were placed into 1.5ml Eppendorf tubes and centrifuged at 1200rpm for 5 minutes at room temperature. They were then resuspended in 100μl of BD Cytofix/Cytoperm solution (BD Biosciences, Cat No: 51-2090K2) and incubated for 20 minutes at 4°C. The cells were then washed twice with 500μl of a 1:10 dilution of 10x BD Perm/Wash cell permeabilising buffer (BD Biosciences, Cat No: 51-2091K2). After two washes, the cells were resuspended in 50μl of BD Perm/Wash buffer containing either H3K27me3-Alexa 647 conjugated antibody (1:100 dilution) or Histone H3-Pacific Blue conjugated antibody (1:50 dilution) and incubated at 4°C in the dark for 30 minutes. A further two wash steps (as already described) were undertaken to remove any excess antibody and the cells were resuspended in 200μl of BD Perm/Wash buffer mixed with 1X PBS for flow cytometry.

Flow cytometry was performed in the CIMR Flow cytometry core, on a BD Biosciences LSRFortessa 4-laser machine, as already described. Unstained cells were used to gate live cells. The 640 670/14 laser was used to detect the Alexa-647 fluorochrome conjugated to H3K27me3 antibody and the 405 450/50 laser was used for Pacific Blue conjugated to Total H3 antibody. Quantifying a reduction by flow cytometry in H3K27me3 following treatment with GSK343 was compared to both lack of reduction in H3K27me3 in the DMSO treatment arm and lack of change in both arms in total histone H3 as controls.

All flow analyses were done using FACSDiva analysis software. Data was then analysed further with FlowJo software (Tree Star Inc.) if required.

2.4.3 GFP/YFP/mCherry assessment by flow cytometry

Green and Yellow Fluorescent protein (GFP/YFP) and mCherry were fluorophores used throughout this project in several DNA constructs used across the retroviral transduction and overexpression experiments. Assessment of these markers by flow cytometry across MLL-AF9 (YFP), AML1-ETO9a (GFP) or MOZ-TIF2 (GFP) transformed murine c-kit+ bone marrow cells, mononuclear cells obtained from blood samples of mice with these leukaemias and tissues isolated from terminal leukaemic mice, allowed an estimation of percentage of transformed cells or leukaemic burden to be made. Both GFP and YFP were readily detectable on the same channel 488 530/30 using the BD LSRFortessa machine in the CIMR Flow Cytometry Core. The mCherry marker was incorporated into the MSCV-IRES-mCherry constructs used to clone candidate genes (*Lin28b* or *Plag1*) into for functional assessments. It was used to provide a surrogate measure of overexpression of these genes in murine c-kit+ bone marrow cells when assessing transduction efficiency, also to track leukaemia development in peripheral blood and to assess terminal leukaemia tissue burden. It was readily detectable on the 561 610/20 channel.

For terminal leukaemia tissues that had been subjected to potential toxicities during processing, in order to assess GFP/YFP/mCherry percentages accurately, live cells were gated for using cells negative for 7-AAD (Methods 2.4.1) detected on the 561 670/30 channel.

2.5 Protein purification, detection and analysis

2.5.1 Preparation of cell lysates for GSK343 inhibitor work

As a similar set up for a time course (Methods 2.4.1), changes in H3K27me3 in the Kasumi AML cell line in the presence of GSK343 versus DMSO control were assessed in order to identify a time point of maximal inhibition of H3K27-trimethylation. The time points chosen were 24, 48, 96 and 144hr (D1, D2, D4 and D6 respectively). Two million Kasumi cells were cultured in 4ml media with either DMSO or GSK343 at 10μM concentration. Every 24 hours the cells were counted and media + drug/DMSO were refreshed if counts were greater than $1x10^6$/ml, with cells being returned to a $0.5x10^6$/ml concentration each time. At each time point, $0.5x10^6$ cells were removed, washed and resuspended in 1ml cold PBS containing Ca^{2+}/Mg^{2+}. They were then resuspended in 20ml of 2X Dithiothreitol (DTT) with repetitive pipetting up and down to shear DNA. Using a hot block, the cells were then boiled for

5 minutes at 100°C and spun at 8000rpm for 1 minute. The resulting lysates were then stored at -80°C until required for SDS-PAGE and protein transfer.

2.5.2 Preparation of cell lysates for confirmation of overexpression of candidate genes

Either 293T HEK cells that had been successfully transfected with the MSCV-IRES-mCherry vector containing the cloned candidate gene insert (*Lin28b* or *Plag1*) OR transduced wild-type murine c-kit selected bone marrow cells with the same were used to prepare cell lysates for an assessment of overexpression of the protein for each candidate gene. Approximately 10^6 transfected or transduced cells were washed and resuspended in 40μl Laemmli buffer (made as below) supplemented with 2μl of 5% β-mercaptoethanol (Sigma; Cat No: M6250) and 0.4μl of protease inhibitor (Sigma Cat No: 11 873 580 001). These were transferred to 1.5ml TPX sonication tubes (Diagenode; Cat No: C30010009), secured into a precooled Diagenode Bioruptor Plus machine and sonicated on "high" setting for 10 cycles (30seconds on; 45 seconds off) to break down DNA interfering with the sample processing. These were transferred on ice to a heat block at boiled at 100°C for 5 minutes and spun at 8000rpm for 1 minute. The resulting lysates were then stored at -80°C until required for SDS-PAGE and protein transfer.

Laemmli buffer

3.125ml 1M Trisma pH 6.8
5ml 20% SDS
10ml 50% glycerol
5ml 0.1% bromophenol blue
26.875ml Ultra-pure water

2.5.3 SDS-PAGE and protein transfer

A 12.5% Polyacrylamide gel was prepared between 1.5mm glass spacer plates:

Lower 12.5% gel

6.25ml Acrylamide
3.75ml Lower gel buffer – containing 1.5M Tris and 0.4% SDS, pH to 8.8 (with HCl)
5ml Water
10μl TEMED
50μl APS

1ml of isobutanol was overlayed on this gel to ensure an evenly horizontal upper gel surface. After the gel was set, the isobutanol was poured away and any excess dried with Whatman 3MM paper.

The stacking gel was then prepared as follows:

Upper gel

400µl Acrylamide

800µl upper gel buffer – containing 0.5M Tris and 0.4% SDS, pH to 6.8 (with HCl)

1.9ml water

3.13µl TEMED

15.7µl 10% APS

The stacking gel was set above the lower gel and wells were formed by placement of a 1.5mm plastic comb. Once solidified, the comb was removed and the gel assembled into an electrophoresis chamber (BioRad) filled with 1x Laemmli buffer (containing Tris pH 6.8, SDS, glycerol and bromophenol blue as above). Each lysate was diluted 1:5 with 5x Laemmli buffer, denatured at 95^{o}C for 3-5 minutes and then loaded into the wells created on the gel following comb removal. Any spare wells/lanes were loaded with sample buffer to ensure even running across the gel, when run at a constant 30mA over an hour.

A Polyvinylidene Difluoride (PVDF) membrane was then cut to match the size of gel and activated by immersion into 100% methanol for a few minutes (with subsequent washing with water). A large transfer container was then filled with CAPS transfer buffer in preparation for transfer. The gel and PVDF membrane were layered sequentially into a cassette, whilst submerged in transfer buffer, between sponge and Whatman 3MM paper and bubbles were removed using a roller. The cassette was transferred to the container with an ice box to prevent overheating, and run for 2 hours at a constant 200mA. After completion of wet transfer, the membrane was labelled to highlight the side opposed to the gel during transfer.

2.5.4 Western Blotting

The membrane was then immersed in blocking buffer composed of 5% dried milk powder in 1x PBS/0.1% Tween20 for 1 hour at room temperature. Following this, it was stained overnight at 4^{o}C with appropriate antibodies (e.g. for H3K27me3, Lin28b, Plag1 and β-Actin or GAPDH as loading controls) made to the appropriate dilution as per manufacturer's instructions (in a Falcon tube placed

on a roller for constant agitation) in 5% milk powder and PBS/Tween (as above). It was then washed every 10 minutes with PBS/Tween for an hour, utilising a platform rocker. Secondary antibody conjugated to horseradish peroxidase was made up to the appropriate dilution in 0.5% milk powder and PBS/Tween and the membrane was submerged in this for 1 hour at room temperature. The blot was then washed as previously with excess buffer removal via filter paper and the membrane was immersed in Amersham ECL solution for 5 minutes.

2.5.5 Developing Western blot

Hypercassette method

After drying, the membrane was encased in cling-film and taped inside a Hypercassette (Amersham Biosciences) with a marker attached for orientation and exposed to x-ray film for 2-10 seconds (dependent upon the strength of antibody, in order to achieve a good quality exposure) and developed.

Licor Odyssey

After secondary antibody staining for 30 minutes at room temperature with Donkey Anti-Rabbit: green or Goat Anti-Mouse: Red the membrane was washed with 5% milk powder and TBST as above and then developed on a Licor Odyssey instrument, with images of the Western Blot being captured directly to computer via the supplied ImageStudio software.

2.6 Mouse work and development of induction and maintenance *in vitro / in vivo* models

2.6.1 Animal husbandry

All regulated procedures were performed in accordance with Huntly group Project license numbers 80/2454 and P86C400DB. Any procedures outlined in this report were completed under my personal Home Office licence and included irradiation, blood sampling, ear notching for identification purposes, intraperitoneal injections and humane sacrifice under Schedule 1. Direct supervision and training/assistance with these procedures was kindly provided by Dr George Giotopoulos (Huntly group) who also performed all tail vein injections required for the transplant experiments.

2.6.2 *Ezh2 fl/fl* tissues

Through collaboration with Dr George Vassiliou's laboratory, Wellcome Trust Sanger Centre, Hinxton, Cambridge UK, the Huntly laboratory had access to existing Sanger C57/Bl6 strain mice engineered to have a *loxp* site flanking exon 9 of the *Ezh2* sequence ("*Ezh2 fl/fl*" mice) in a homozygous fashion.[2] Breedings between these and mice homozygous for *Mx1-Cre* recombinase were set up courtesy of Drs Milena Mazan and Monika Dudek, Vassiliou group and subsequently litters of *Ezh2 fl/fl*; wild-type for *Mx1-Cre* (henceforth '*Ezh2 fl/fl*; wt' or 'wt') or *Ezh2 fl/fl*; heterozygous for *Mx1-Cre* (henceforth '*Ezh2 fl/fl*; Cre+' or 'Cre+'), mice were generated and selected for use following genotyping to confirm *Mx1-Cre* status. *Ezh2 fl/fl; Mx1-Cre* homozygous mice were not used due to well documented issues with Cre-mediated toxicity.

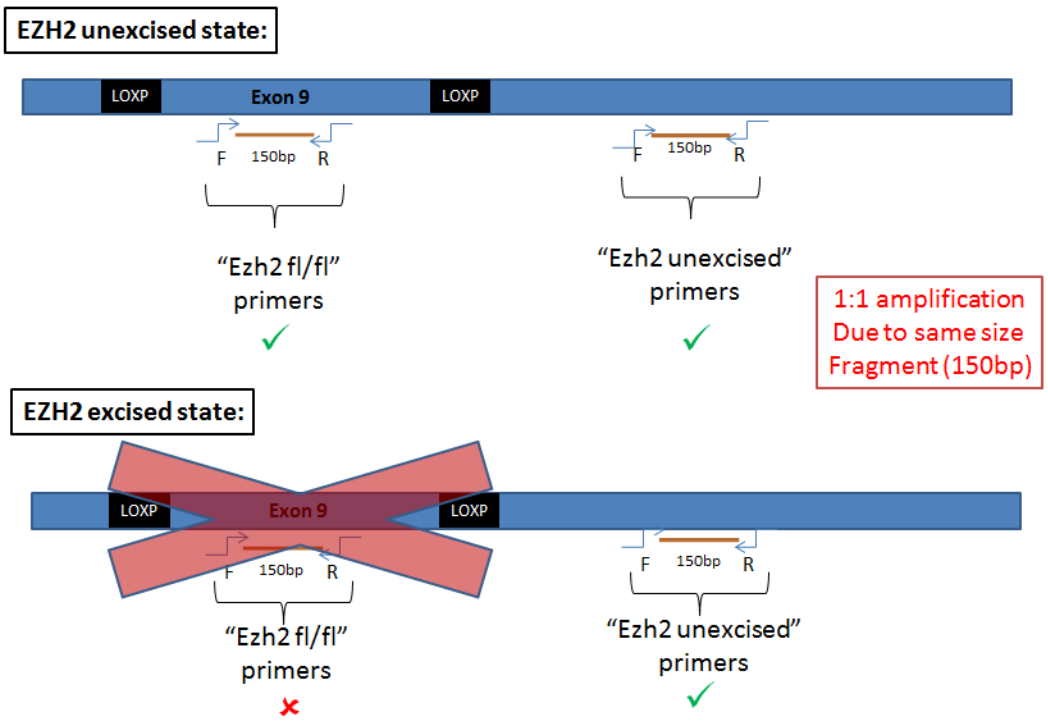
For both *in vitro* and *in vivo* induction experiments, 5-6 mice from each wt or Cre+ litter generated were used (ensuring animals were age and sex-matched) to provide bone marrow (BM).

For maintenance *in vitro* experiments, 5-6 wt mice were used, and for the maintenance *in vivo* experiments 5-6 Cre+ mice were used to generate the primary leukaemias.

2.6.3 *Ezh2* Genotyping

Due to this project being heavily dependent on accurately assessing *Ezh2* excised vs unexcised states (and their intermediaries), it was very important able to quantify these accurately at various stages during setup and analysis of these experiments. In order to do this, a qPCR strategy was developed and optimised through guidance from Dr George Giotopoulos (Huntly group) in order to better quantify/estimate levels of excision. A standard PCR and gel method, though able to provide an assessment of "excised or unexcised", was at best semi-quantitative.

Therefore, primers coding for a 150bp sequence within the *loxp* sites surrounding exon 9 of *Ezh2* were designed and a similar pair of primers (with identical melting temperatures) for an equal length 150bp sequence very closely downstream of the *loxp* sites (that would be unaffected by excision and therefore maintained during PCR amplification) were created (Figure 2.2).



***Ezh2 fl/fl* primers -**	**Forward sequence:** TATGGGGGATGAAGTTCTGG **Reverse sequence:** CCACTGTCATGGCAACAAGA

***Ezh2* unexcised primers -**	**Forward sequence:** CTTCTTCCTCCGGGTCTCTC **Reverse sequence:** GATCTTGGGGCATTGCAGTA

FIGURE 2.2 Primer design for assessing *Ezh2* excision via qPCR

In the unexcised state, both pairs of primers amplify at a 1:1 ratio by qPCR due to similar melting temperatures and identical sequence length. In the excised state, "*Ezh2 fl/fl*" primers are unable to amplify the region in relation to "*Ezh2* unexcised" primers. This difference was consistent and when compared within the sample and then to a control sample subjected to the same qPCR reaction, the ratio of the difference in C_t values was used to provide an estimation of level of excision

In the *Ezh2* unexcised state, these primers would amplify their respective sequences at an approximately 1:1 ratio given identical sequence sizes. In the *Ezh2* excised state, the "*Ezh2 fl/fl*" primers would fail to amplify whilst the "*Ezh2* unexcised" primer pair would amplify. All qPCR was done using a Stratagene MX3000P machine and data analysed through MxPro qPCR software. Each reaction used 10-20ng DNA and SYBR Green reagent (Ultrafast III, Agilent Technologies).

Trialling these two pairs of primers on control DNA (where *Ezh2* was known to be floxed but not excised i.e. ***"Ezh2 fl/fl"***) showed almost identical amplification between both primer pairs as evidenced by very

similar average C_t values (average C_t: 23.27 for "*Ezh2 fl/fl*" primers vs average C_t: 23.795 for "*Ezh2* unexcised" primers), (Figure 2.3):

Sample	fl/fl (1)	fl/fl (2)	fl/fl (3)	fl/fl Average	Un. (1)	Un. (2)	Un. (3)	Unexcised Average	ΔCt
EZH2 fl/fl	22.93	23.39	23.49	23.27	23.63	23.96	23.8	23.795	-0.525

FIGURE 2.3 **C_t values on qPCR using *Ezh2 fl/fl* primers versus *Ezh2* unexcised primers**
"*Ezh2 fl/fl*" is Control DNA obtained from *Ezh2fl/fl*; wild-type for Mx1-Cre murine HSPCs.
Mean average from triplicates

When tested in an *Ezh2* deleted scenario ("*Ezh2* -/-", Figure 2.4), as expected, amplification with the "Ezh fl/fl" primer pair was significantly decreased, therefore the average C_t value for this was typically higher when compared to the normal amplification attained by the "*Ezh2* unexcised" primer pair (average C_t: 27.685 for "*Ezh2 fl/fl*" primers vs average C_t: 23.635 for "*Ezh2* unexcised" primers):

Sample	fl/fl (1)	fl/fl (2)	fl/fl (3)	fl/fl Average	Un. (1)	Un. (2)	Un. (3)	Unexcised Average	ΔCt
Ezh2 -/-	27.8	27.57		27.685	23.74	23.53		23.635	4.05

FIGURE 2.4 **C_t values on qPCR using *Ezh2 fl/fl* primers versus *Ezh2* unexcised primers**
Ezh2 -/- is test DNA from *Ezh2* fl/fl; Mx1-Cre+ murine HSPCs that have undergone pIpC to induce *Ezh2* deletion.

In both the control (*Ezh2 fl/fl*) and test sample (*Ezh2* -/-), this change in average C_t value between both pairs of primers was calculated (ΔC_t) (Figure 2.5):

Sample	fl/fl (1)	fl/fl (2)	fl/fl (3)	fl/fl Average	Un. (1)	Un. (2)	Un. (3)	Unexcised Average	ΔCt
EZH2 fl/fl	22.93	23.39	23.49	23.27	23.625	23.96	23.8	23.795	-0.525
Ezh2 -/-	27.8	27.57		27.685	23.74	23.53		23.635	4.05

FIGURE 2.5 **Δ C_t values for control and test samples**
"*Ezh2*fl/fl": control unexcised DNA as above; *Ezh2* -/- is test DNA (where *Ezh2* deletion is expected, as above)

As seen in Figure 2.5, the Δ C_t value for the *Ezh2*-/- sample (4.05) was considerably different to the Δ C_t value attained in the control DNA (-0.525). This reflected the difference in *Ezh2* states (unexcised in the control, and excised to an unknown proportion in *Ezh2* -/-). Using the control DNA Δ C_t value as the normal, the Δ C_t value for *Ezh2*-/- was then subtracted from it, providing an estimation of the differences in amplification seen between the control and *Ezh2*-/- ($\Delta\Delta C_t$) (Figure 2.6):

Sample	fl/fl (1)	fl/fl (2)	fl/fl (3)	fl/fl Average	Un. (1)	Un. (2)	Un. (3)	Unexcised Average	ΔCt	ΔΔCt
EZH2 fl/fl	22.93	23.39	23.49	23.27	23.625	23.96	23.8	23.795	-0.525	0
Ezh2 -/-	27.8	27.57		27.685	23.74	23.53		23.635	4.05	-4.575

FIGURE 2.6 **$\Delta\Delta C_t$ values for control and test samples**
"*Ezh2fl/fl*": control unexcised DNA as above; *Ezh2-/-* is test DNA (where *Ezh2* deletion is expected, as above)

This $\Delta\Delta$ C_t value reflected the number of additional cycles required by the primers to amplify in *Ezh2* -/- compared to the control sample and when this was resolved using 2^($\Delta\Delta$ C_t), a final value was generated (termed "2^{nd} C_t"), which reflected the proportion of unexcised DNA in the sample. Knowing that no excision had taken place in the control sample, the unexcised proportion was taken to be 100%. By comparison in *Ezh2-/-*, there was only 4.2% unexcised, therefore approximately 95.8% excision (Figure 2.7):

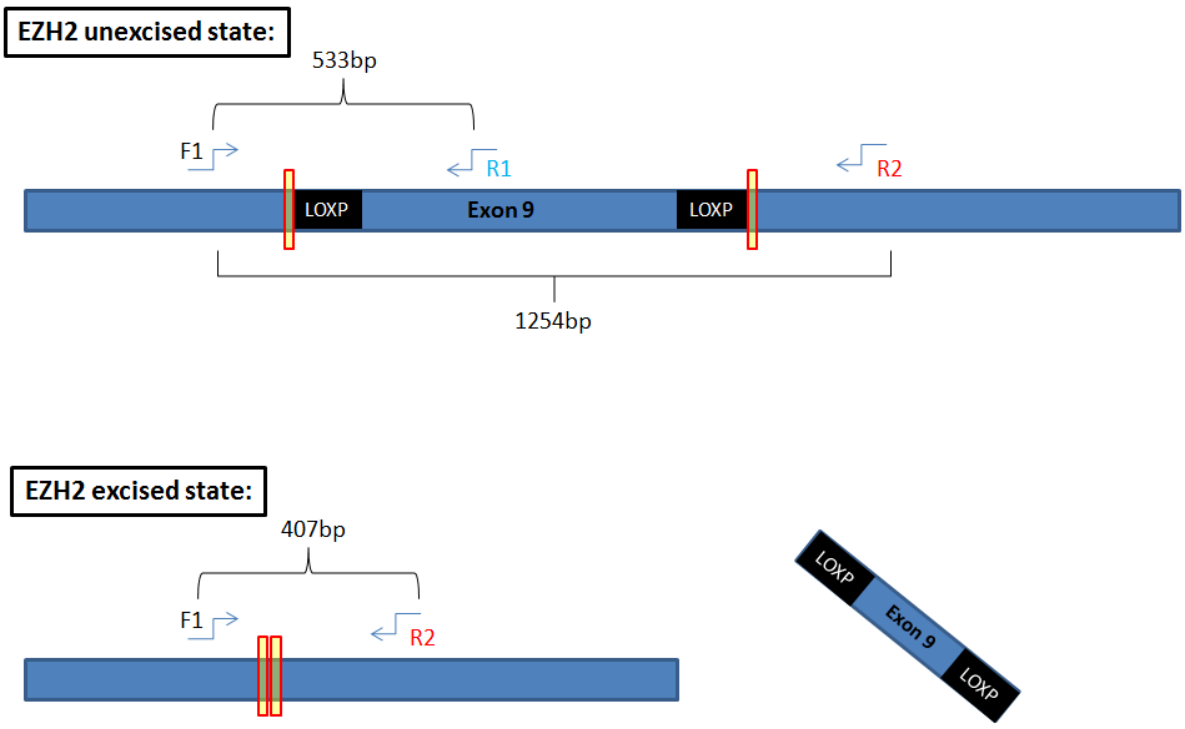
Sample	ΔCt	ΔΔCt	2ndCT	Unexcised proportion	Excision efficiency
EZH2 fl/fl	-0.525	0	1	100	0
Ezh2 -/-	4.05	-4.575	0.042	4.195539071	95.80446093

FIGURE 2.7 **Calculation of Ezh2 excision in control and test samples**
"*Ezh2fl/fl*": control unexcised DNA as above; *Ezh2-/-* is test DNA (where *Ezh2* deletion is expected, as above)

Thus, by comparing test sample Δ C_t values with control DNA Δ C_t values (where 100% of *Ezh2* was intact) each time, it was possible to generate an approximation of *Ezh2* excision, expressed as a percentage.

This assay was validated against a semi-quantitative PCR strategy developed separately, consisting of one forward primer upstream (F1) of the initial *loxp* site and two reverse primers (R1: within the *loxp* sites and R2: downstream of the second *loxp* site). The PCR products of F1+R1 and F1+R2 yield the following bands depending upon Ezh2 state (Figure 2.8):

Ezh2 fl/fl unexcised	*Ezh2* fl/fl excised	Wild-type
F1+R1 = 533bp	F1+R1 = no band	F1+R1 = 316bp
F1+R2 = 1254bp	F1+R2 = 407bp	F1+R2 = 955bp



F1: CCTGCAGCAGTTTCCTTTCTTA, R1: CCCATGTGGTAGGCAGAGAA, R2: CAGCGACACCCCAGAAAATT

FIGURE 2.8 **Primer design for assessing *Ezh2* excision by PCR and Gel electrophoresis**

When known *Ezh2* excised and unexcised DNA controls were run alongside four MOZ-TIF2 and four AML1-ETO9a DNA samples with an expected excised status, the qPCR strategy provided the following results (Table 2.3):

	Sample	fl/fl primers average	Unexcised primers average	Δ Ct	ddCt	2ndCT	Unexcised proportion	Excision efficiency
Controls	1 (Excised)	29.32	24.525	4.795	-4.533	0.043185	4.32	95.68
	2 (Excised)	29.475	24.345	5.13	-4.868	0.034236	3.42	96.58
	1 (unexcised)	24.68	24.235	0.445	-0.183	0.880666	88.07	11.93
	2 (unexcised)**	24.56	24.30	0.26	0	1	100.00	0.00
Unknowns	1 (MOZ-TIF2)	26.635	24.39	2.245	-1.983	0.252905	25.29	**74.71**
	2 (MOZ-TIF2)	26.695	24.91	1.785	-1.523	0.347881	34.79	**65.21**
	3 (MOZ-TIF2)	26.53	24.61	1.92	-1.658	0.316805	31.68	**68.32**
	4 (MOZ-TIF2)	25.605	24.255	1.35	-1.088	0.470304	47.03	**52.97**
Unknowns	1 (AML1-ETO9a)	26.18	24.595	1.585	-1.323	0.399611	39.96	**60.04**
	2 (AML1-ETO9a)	28.02	24.36	3.66	-3.398	0.094842	9.48	**90.52**
	3 (AML1-ETO9a)	26.015	24.26	1.755	-1.493	0.355191	35.52	**64.48**
	4 (AML1-ETO9a)	26.67	24.39	2.28	-2.018	0.246843	24.68	**75.32**

TABLE 2.3 **Assessment of *Ezh2* excision in unknown *Ezh2* status samples using qPCR**
Samples indicated in Yellow are known excised controls and red are known unexcised controls.
Levels of excision highlighting in bold (final column – "Excision efficiency")
(** control sample in which 1:1 amplification is expected. All other sample ΔCt values compared to this control when calculating change in ΔCt to assess excision efficiency)

When the same samples were run using the PCR and gel strategy (shown in Figure 2.9 - labels are colour matched to Table 2.3), the MOZ-TIF2 and AML1-ETO9a unknowns all demonstrated clear excision, with band intensities visually comparable to the percentages of excision calculated by the qPCR method.

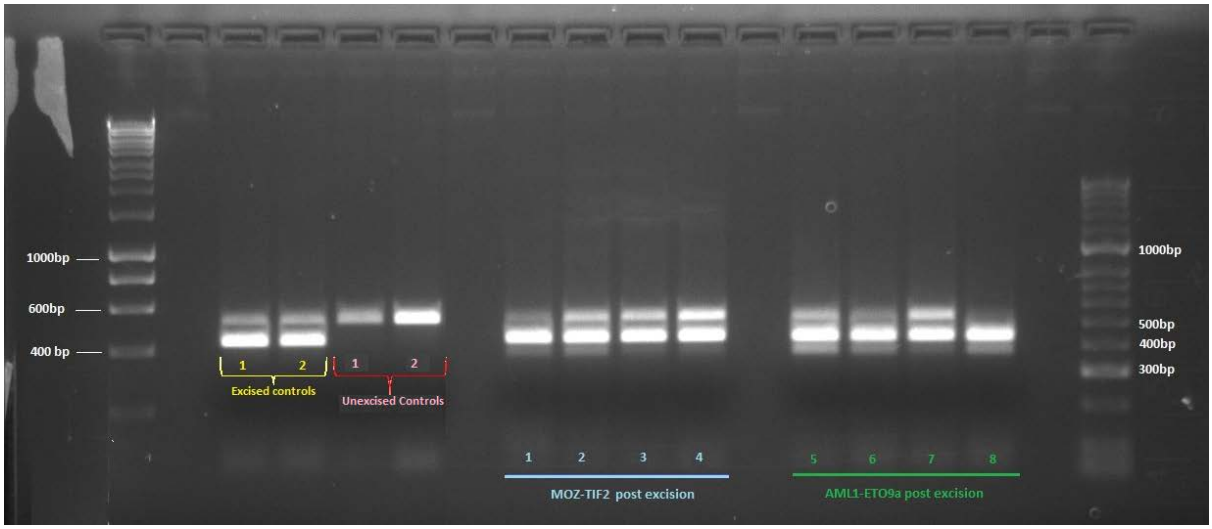


FIGURE 2.9 Assessment of *Ezh2* excision in unknown *Ezh2* status samples using PCR & Gel
F1+R1 and F1+R2 primers used for PCR

Excised controls (yellow):	show intense 407bp band expected in excised state and less intense 533bp band reflecting a small proportion in the unexcised state
Unexcised controls (red):	show a single 533bp band expected in unexcised state
Test samples (blue/green):	all demonstrate varying levels of excision (both 407/533bp bands of varying intensities) directly comparable to estimated percentage of excision by qPCR method

*additional faint band seen across test samples at 316bp position reflects residual wild-type status, as samples were taken from primary transplants into wild-type recipients, therefore reflect chimerism

With successive repeats (data not shown), the qPCR method proved to be a reliable means of assessing levels of *Ezh2* excision and was comparable to assessing bands on agarose gels after PCR using a traditional primer strategy as described above. It had the added advantage of being able to quantify levels of excision approximately in terms of percentages, thereby allowing a fairer comparison across *in vitro* and *in vivo* samples that required genotyping throughout this PhD project, therefore this method was used for genotyping and assessing *Ezh2* deletion in all experiments described.

2.6.4 *Ezh2* deletion and bone marrow harvesting

2.6.4.1 Induction experiments

At 6 weeks of age both *Ezh2 fl/fl*; wt and *Ezh2 fl/fl*; Cre+ mice were subjected to polyinosinic:polycytidylic acid (pIpC) injections to facilitate Cre-mediated excision in the Cre+ group. Four weeks following pIpC, peripheral blood samples from each animal from both groups were obtained for genotyping to ensure *Ezh2* excision had occurred in the Cre+ group and that no excision had occurred in the wt group. After this, both cohorts were culled humanely and both femurs and tibia were harvested from each mouse. Both *in vitro* and *in vivo* induction experiments utilised bone marrow harvested from the above genotypes treated with pIpC in this way.

2.6.4.2 Maintenance experiments

For *in vitro* experiments only *Ezh2 fl/fl*; wt mice at 8-12 weeks of age were required. These were genotyped for *Ezh2* status using peripheral blood samples prior to bone marrow harvesting.

For *in vivo* experiments only *Ezh2 fl/fl*; Cre+ mice at 8-12 weeks of age were required. These were again genotyped for *Ezh2* and *Mx1-Cre* status using peripheral blood samples prior to bone marrow harvesting. There was no pIpC treatment beforehand as the conditional knockout was required only at the secondary transplantation stage.

2.6.4.3 Functional validation of candidate genes (*Lin28b*, *Plag1*)

For these experiments, 5-8 wild-type C57/Bl6 mice aged 8-12 weeks were used to provide bone marrow. Harvesting of bone marrow cells was done as already described in this section. Their processing for c-kit selection, transduction and *in vitro* (Western blot, qPCR) and *in vivo* work (transplantation) is as described in the relevant sections.

Under sterile conditions, the bones from each animal were flushed with 20-30ml sterile 1xPBS through a 70μm EASYstrainer™ (Greiner) into separate 50ml Falcon tubes on ice, centrifuged at 1200rpm, 4°C and the cell pellet resuspended in 5ml of RBC lysis solution (5 Prime GmbH, Cat No: 2301310) for 5 minutes. Red cell lysis was stopped by addition of 45ml PBS and total BM cell number was counted via CASY Counter. At this stage $1x10^6$ whole BM cells from each mouse were stored for genotyping.

2.6.5 c-kit selection

All samples were centrifuged again and wt cells were pooled together separately from Cre+ cells by resuspending in FcR mouse blocking reagent (Miltenyi Biotec, Cat No: 130-092-575) as per manufacturer's instructions and MACS buffer (500 ml PBS without Ca/Mg, 2.5 g BSA and 2 ml of 0.5 M EDTA, filter sterilised). Following this, CD117 (c-kit) mouse Microbeads (Miltenyi Biotec, Cat No: 130-091-224) were added and incubated as per manufacturer's instructions then passed through a pre-separation 30μm filter and MACS LS separation column (Miltenyi Biotec, Cat No: 130-041-407 and 130-042-401 respectively) attached to a midiMACS magnetic separator (Miltenyi Biotec, Cat No: 130-042-301). C-kit selected cells were trapped in the column attached to the magnetic separator and the flow through contained c-kit negative cells. The column was then removed and flushed separately to release c-kit positive cells from both wt and Cre+ pools. Total cell numbers yielded were confirmed using a CASY counter. Typically, c-kit selection efficiency using this method was approximately 10% of total BM cells, with assessment of purity by flow cytometry following CD117-PE staining.

Following this, c-kit positive cells were resuspended at $1x10^6$ cells/ml in RPMI+20%FCS media, supplemented with mIL-3, IL-6 and mSCF cytokines, (Peprotech™; final concentrations of 10ng/ml, 10ng/ml and 100ng/ml respectively) and cultured overnight at 37°C in 12-well plates for retroviral transduction the following morning (Methods 2.6.7).

2.6.6 Production of retroviral supernatant

In accordance with Institution guidelines, all retroviral work was carried out under Category II laboratory conditions. The Huntly laboratory possesses MSCV-IRES-GFP/YFP backbone constructs harbouring MOZ-TIF2 (GFP marker) and MLL-AF9 (YFP marker) oncogenes and a MIGR1 construct that contains the AML1-ETO9a oncogene (GFP marker):

1) MSCV-IRES-MLL-AF9-YFP (referred to as 'MLL-AF9' in Methods 2.6.6.1 and 2.6.7.2)
2) MSCV-IRES-MOZ-TIF2-GFP
3) MIGR1-AML1-ETO9a-GFP (https://www.addgene.org/12433/)

These were used to transfect 293T cells as follows:

293T cells were passaged, using Trypsin-EDTA (Thermo Scientific, Cat No: 25200056) to detach from plates, under standard tissue culture conditions in DMEM + 10% FCS medium + L-glutamine, on 10cm sterile plates. $2.5x10^6$ cells were plated on Day 0 and incubated overnight. At Day 1, provided the cells

were 50-70% confluent, transfection was completed using an incubated mixture of 5μg ψ eco with 47.5μl TranSIT LT1 reagent (Mirus, Cat No: MIR2300) added dropwise to 600μl Optimem media (Life Technologies, Cat No: 31985062) with 5μg of construct DNA. The mixture was added dropwise to the plates whilst rocking to ensure even mixing then incubated at 37°C overnight. At Day 2, the media was replaced with 6ml RPMI+20%FCS media and the cells were incubated at 32°C overnight. On Day 3 (48hr post transfection) retroviral supernatant was harvested and filtered through 45μm filters on ice to remove any detached 293T cells. Supernatant was stored in aliquots at -80°C. A further 6ml media was added and again incubated overnight so that more supernatant could be harvested at Day 4 (as per Day 3).

2.6.6.1 Functional validation of candidate genes (*Lin28b, Plag1*)

For this experiment, retroviral supernatant from 293T cells either single or dual-transfected with the following constructs was collected:

1) MLL-AF9 alone
2) MLL-AF9 + MSCV-IRES-Lin28b-mCherry
3) MLL-AF9 + MSCV-IRES-Plag1-mCherry

The procedure for transfection was exactly as described above, with the only difference being in the dual-transfection, where 2.5μg of DNA of MLL-AF9 + 2.5μg DNA of candidate gene inside MSCV-IRES-mCherry vector (as listed here) was used to generate retroviral supernatant containing both constructs at a 1:1 ratio. This was stored as described above.

2.6.7 Retroviral transduction

2.6.7.1 Induction and maintenance experiments

Retroviral supernatant aliquots for each oncogene AML1-ETO9a, MOZ-TIF2, MLL-AF9 were thawed slowly on ice under sterile conditions, and the c-kit positive cells (wt or Cre+) were split into three portions depending on total viable cell number appropriate to an estimation of cells required for transplant per mouse for each oncogene. After resuspension in retroviral supernatant (1ml supernatant per $1x10^6$ cells) supplemented with cytokines (as in Methods 2.6.5) and polybrene to a final concentration of 8ng/μl, the cells were distributed evenly across a 12 or 24-well plate and centrifuged at 2500rpm for 90 minutes at 30°C. Following this the plates were incubated for 3-4 hours at 37°C at which point the media was changed (after centrifugation and resuspension) to RPMI+20%

FCS + cytokines but lacking polybrene, as above. The cells were then incubated overnight. A second round of retroviral transduction (as described) was performed on the cells the following morning, after which they were washed three times in sterile PBS.

2.6.7.2 Functional validation of candidate genes (*Lin28b*, *Plag1*)

Retroviral supernatant for MLL-AF9 alone or MLL-AF9 + either a) MSCV-IRES-*Lin28b*-mCherry or b) MSCV-IRES-*Plag1*-mCherry was thawed as above. Following c-kit selection, the cells were split into three conditions and transduced with MLL-AF9 viral supernatant alone or MLL-AF9 + a) MSCV-IRES-*Lin28b*-mCherry or b) MSCV-IRES-*Plag1*-mCherry. The transduction procedure is as described above.

2.6.8 Assessment of transduction efficiency

2.6.8.1 Induction and maintenance experiments

Following two rounds of transduction and three washes, cell numbers were attained for wt and Cre+ c-kit positive cells for each oncogene and 10^5 cells were taken to assess transduction efficiency by flow cytometry (as per Methods 2.4.3) as each construct expressed GFP (MOZ-TIF2 or AML1-ETO9a) or YFP (MLL-AF9). The same channel (488 530/30) was used to detect both. On average there was a 1-5% transduction rate for MLL-AF9 and MOZ-TIF2 (due to larger construct size) and 15-25% rate for AML1-ETO9a (smaller construct). Following this, cells were either plated into methylcellulose for assessment *in vitro*, or injected into C57/Bl6 wild-type sub-lethally irradiated mice and monitored for development of leukaemia.

2.6.8.2 Functional validation of candidate genes (*Lin28b, Plag1*) – flow cytometry

Following transduction, flow cytometry was used to assess efficiency using 488 530/30 filter for YFP (MLL-AF9) and 561 610/20 for mCherry (candidate genes). An assessment of percentage of single-transduced YFP or mCherry cells and double-transduced YFP+mCherry cells was made. After this, it was possible to calculate the appropriate number of YFP positive cells for injection into each arm of the *in vivo* study to ensure an equal dose of MLL-AF9 transduced cells was given.

2.6.8.3 Functional validation of candidate genes (*Lin28b*, *Plag1*) – quantitative RT-PCR

Following transduction, RNA was isolated from $1x10^6$ cells as per section 2.8.14.2 and cDNA was generated as per Section 2.7.1.1 to quantitate mRNA. A 20µl reaction was prepared in 96 well plates (using 10µl SYBRGreen master-mix, 1µl each of forward and reverse primers (to a final concentration

of 10μM), 1μl cDNA, ROX fluorescent reference dye 0.3μl of 1:500 dilution and 6.7μl RNase-free water). PCR amplification was performed with an initial step of 10 minutes at 95°C, followed by 40 cycles of 15 seconds at 95°C and 1 minute at 60°C. All samples were measured in duplicate and the mean C_t values were used to calculate relative expression levels. The reference housekeeping genes used were β-actin and GAPDH.

2.6.9 Irradiation and primary transplantations

Where transduced c-kit positive wt or Cre+ cells were used for *in vivo* transplantation, the recipient C57/Bl6 wild-type mice had been ear notched prior for identification purposes, and given sub-lethal or lethal irradiation (a single 5.5Gy dose or two doses of 5.5Gy on consecutive days respectively). In order to reduce skewing of results due to variability in cell numbers injected per mouse, the transduction efficiency was used to calculate an approximate number successfully transduced cells, i.e. GFP/YFP-positive (in the induction and maintenance experiments) or a ratio of YFP/mCherry to YFP positive cells (in the functional validation experiments) that were then resuspended in the correct volume of sterile PBS required for tail-vein injections. As a result, for example in the induction *in vivo* experiment, each mouse transplanted with wt cells received approximately the same number of GFP/YFP-positive cells as per the mice transplanted with Cre+ cells, ensuring a fair comparison between the two cohorts for each oncogene. The same process was applied when calculating appropriate cell numbers for transplantation in the functional validation experiments (MLL-AF9-YFP single-transduced vs MLL-AF9-YFP + MSCV-IRES-*Lin28b/Plag1*-mCherry dual-transduced).

Each mouse received cells in a 200μl volume via tail vein injection. Following injection all mice received enrofloxacin prophylaxis (Baytril™) for two weeks and were monitored closely for signs of disease, with serial peripheral bleeds and assessment of GFP/YFP expression and blood counts and clinically for signs of leukaemia. When mice showed clinical signs (piloerection, hunched posture, inactivity, inappetence, weight loss >20% compared to non-injected age-matched mice, dyspnoea) they were humanely sacrificed under Schedule 1 of the Animals Scientific Procedures Act 1986.

2.6.10 Murine leukaemia tissue processing/storage

Bone marrow and splenic tissues were processed under sterile conditions after flushing, filtering, red cell lysis and washing (as per Methods 2.6.4) and frozen in liquid nitrogen in FCS+10%DMSO aliquots.

For every animal, GFP/YFP expression in terminal peripheral blood, bone marrow and spleen was assessed by FACS and recorded. Other organs were stored in Formalin (heart, lungs, sternum, liver, kidneys, bowel) and later embedded in wax for sections to be made and stained. Cytospins of bone marrow, splenic tissue (2-4x10^5 cells/slide) were made and fixed along with a terminal blood film and stained using the Rapid Romanowsky stain (TCS Biosciences, Cat No: HS705). Histological preparations of murine tissues and staining were kindly performed by Dr George Giotopoulos (Huntly group). Images were taken using Olympus BX51 microscope with a Pixera 600ES camera and the software for image acquisition was Viewfinder Version 3.01 (Pixera). DNA was extracted from spleen or bone marrow for genotyping for each culled mouse (as per Methods 2.8.14.1). Peripheral blood counts were performed on a Woodley ABC blood counter (Woodley).

2.6.11 Induction *in vitro* assays

Following transduction of wt or Cre+ c-kit bone marrow cells from mice pre-treated with pIpC to induce *Ezh2* deletion for each oncogene (MLL-AF9, AML1-ETO9a or MOZ-TIF2), 3.3-5x10^4 transduced cells were plated for every one millilitre of M3434 methylcellulose. These were then serially replated on a weekly basis, with an assessment of total colony and cell number being made at each stage. The overall experimental plan is depicted in Figure 2.10.

2.6.12 Maintenance *in vitro* assays

In order to set up this experiment, steps from harvesting bone marrow of *Ezh2 fl/fl*;wt Cre mice (aged 8-12 weeks) through to generation of the three transformed cell lines were performed (as detailed above). GFP/YFP expression was measured by FACS and increased as expected with each successive replating. By week 4-5 all three cell lines had GFP/YFP levels consistently above 95%, and from assessment of colony appearance, number and rapid increase in total cell number per plate, appeared immortalised.

In order to then excise *Ezh2* in these established cell lines, they were taken into liquid culture (RPMI media + 20% FBS supplemented with cytokines initially and maintained at a concentration between 0.5-1x10^6/ml. After 2-3 passages on alternate days, they were resuspended into RPMI + 20% FBS with mIL-3 alone (in order to reduce cellular differentiation). Using a p-babe-Cre-puro and p-babe-puro empty vector (already available in the Huntly lab) to generate retroviral supernatant, the cell line was split into two arms and then subjected to a second round of retroviral transduction with either vector.

The transduced cells were finally seeded into M3434 methylcellulose plates (at 3.3–5x10^4 per ml) with or without 1μg/ml puromycin (Thermo Scientific, Cat No: A1113803) and assessment of colony/total cell number was made at Day 7.

Cells remaining in either conditions at Day 7 were replated into fresh M3434 at the same concentrations to allow further rounds of replating. At each weekly replating, an assessment of total colony number and cell number was made. The overall experimental plan is depicted in Figure 2.11.

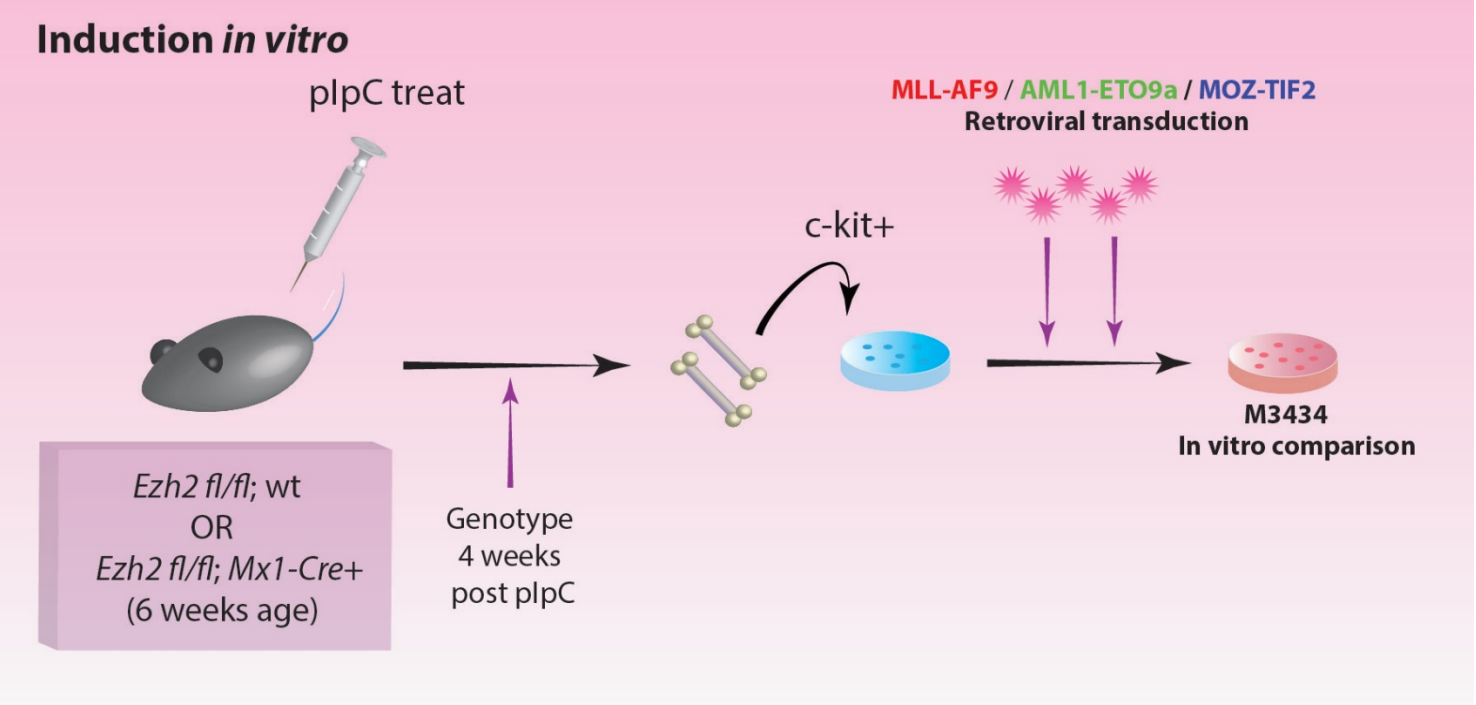


FIGURE 2.10 **Experimental plan for assessing the role of *Ezh2* in leukaemic induction *in vitro***

Ezh2 fl/fl; wt or *Ezh2 fl/fl;Mx1-Cre+* mice at 6 weeks of age were treated with pIpC to induce *Ezh2* deletion in Cre expressing mice (confirmed by genotyping from peripheral blood at 4 weeks post treatment). BM from these two cohorts was then harvested and c-kit positive HSPCs selected and pooled as either *Ezh2* non-excised or excised cells in liquid culture (with RPMI 1640+20% FCS supplemented with cytokines as detailed in main text). Following retroviral overexpression of the 3 oncogenes listed above in both groups of HSPCs, the cells were cultured in methylcellulose assays to assess clonogenic/proliferative function and compared with each other over subsequent weeks of replating as immortalisation took place.

"wt" denotes wild-type for *Mx1-Cre*, "*Mx1-Cre+*" denotes heterozygous for *Mx1-Cre*, "c-kit+" denotes c-kit/CD117 positively selected HSPC's
M3434 = Methocult GF M3434 methylcellulose medium for mouse cells with cytokines (StemCell Technologies)

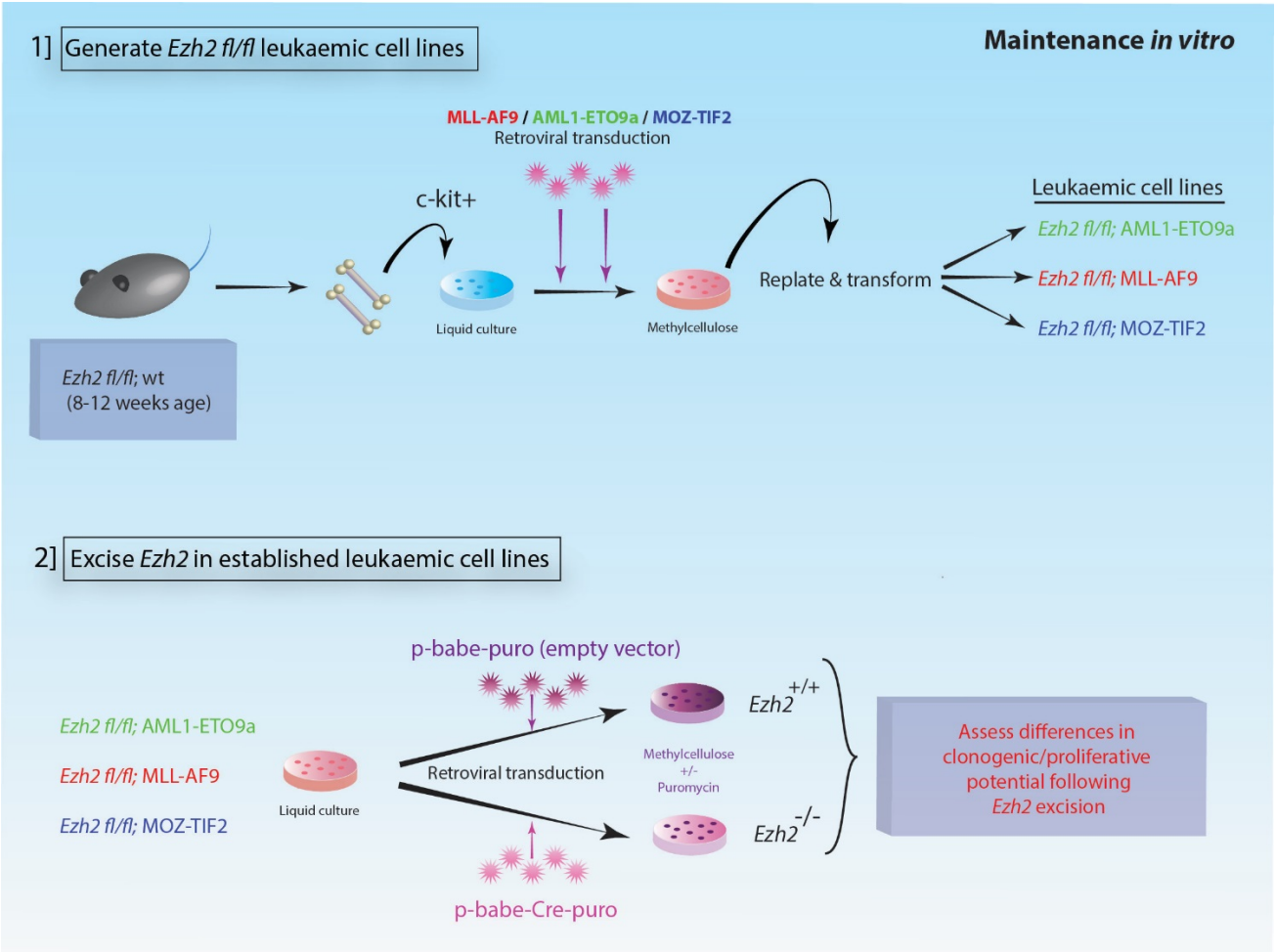


FIGURE 2.11 **Experimental plan for assessing the role of *Ezh2* in maintenance of AML *in vitro***

1) Generate *Ezh2 fl/fl* leukaemia cell lines: Bone marrow from *Ezh2 fl/fl*; wt mice at 8-12 weeks of age was harvested and c-kit positive HSPCs were selected, pooled and divided into 3 arms. Each then underwent retroviral transduction of the 3 oncogenes listed above and was allowed to transform over successive methylcellulose replatings. GFP/YFP percentage and colony number/appearance assessments at each replating provided a clear assessment of when transformation had occurred.

2) Excise Ezh2 in established leukaemic cell lines: Once established, these cells were taken into liquid culture (in RPMI 1640+20% FCS supplemented with cytokines as in main text), transduced either with a p-Babe-puro empty vector or a p-Babe-Cre-puro vector containing a Cre-recombinase (both vectors included puromycin as a selection marker) and cultured +/- puromycin in methylcellulose. Transduction with a retroviral Cre would lead to *Ezh2* deletion and the inclusion of an empty vector arm would assist with assessing any Cre-mediated toxicity

"wt" denotes wild-type for *Mx1-Cre*, "c-kit+" denotes c-kit/CD117 positively selected HSPC's

2.6.13 Induction *in vivo* model

wt (*Ezh2*$^{+/+}$) or Cre+ (*Ezh2*$^{-/-}$) cells from *Ezh2fl/fl* mice of either genotype (pre-treated with pIpC to induce *Ezh2* deletion in the Cre+ arm) were transduced as described. After assessment of transduction efficiency, an equal number of transduced cells were transplanted into lethally irradiated C57/Bl6 wild-type recipient mice, generating *Ezh2*$^{+/+}$ and *Ezh2*$^{-/-}$ leukaemia arms for each oncogene. These were bled at Day 14-21 and their full blood counts assessed to ensure adequate engraftment. They were then followed through regular assessment and blood sampling for development of leukaemia as described above. The overall experimental plan is depicted in Figure 2.12.

2.6.14 Maintenance *in vivo* model

The overall experimental plan for this is depicted in Figure 2.13.

2.6.14.1 Generation of primary leukaemias

Generation of primary murine leukaemias for each oncogene (n=5 per oncogene) was performed as described in Methods 2.6.9. The only difference for this part was that *Ezh2 fl/fl*; Cre+ mice (without any *Mx1-Cre* prior activation i.e. pIpC treatment) exclusively were harvested to provide HSPCs for retroviral transduction. Five mice per oncogene were used to generate sufficient leukaemic tissue for secondary transplants.

2.6.14.2 Maintenance *in vivo* model – secondary leukaemias and *Ezh2* deletion

Following generation of primary leukaemias, when mice were terminal, bone marrow and splenic tissues were harvested and GFP/YFP expression was assessed and recorded before freezing at -80^{o}C. Immunophenotyping was performed to characterise and confirm the leukaemia with genotyping of spleen cells to assess for any spontaneous recombination. Secondary recipient animals chosen were wild type C57/Bl6 mice at the age of 8-12 weeks. Each secondary recipient underwent lethal irradiation as per Methods 2.6.9.

Following this, for each of the three leukaemias to be tested (i.e. MLL-AF9 / AML1-ETO9a / MOZ-TIF2), 1x10^{6} spleen cells (in a volume of 200μl sterile PBS) were injected intravenously per animal (n= 20, divided into two arms), and close monitoring (via peripheral blood GFP/counts and clinical assessment)

was undertaken. The animals were maintained on prophylactic antibiotics as described above. After engraftment, at the earliest suggestion the animal was developing leukaemia (approximately Day 7 for MOZ-TIF2/MLL-AF9 and Day 15 for AML1-ETO9a from previous unpublished data in Huntly laboratory), the test arm (n= 10-12) received pIpC treatment via intraperitoneal (IP) injections (at a dose of 10μg/gram of body weight) on five alternate day doses over eleven days, with the control arm (n= 10-12) receiving IP PBS injections at equal frequency and number.

In the pIpC treated cohort, *Ezh2* deletion occurred (assessed by qPCR genotyping of DNA extracted from peripheral blood, spleen and bone marrow) with differences in survival and phenotypic outcomes being compared to the control treated (i.e. *Ezh2* unexcised) arm.

2.6.15 Functional validation of candidate oncogenes (*Lin28b/Plag1*) – *in vivo* model

MLL-AF9 or MLL-AF9+candidate oncogene (as listed above) was transduced as described above into wild-type c-kit selected bone marrow cells harvested from wild-type C57/Bl6 mice (8-12 weeks age). These cells were transplanted into wild-type lethally irradiated C57/Bl6 recipient mice (8 weeks age) via tail-vein injections in four cohorts (n= 8 per arm) and monitored to ensure engraftment had occurred adequately via full blood count measurements at 14-21 days. Development of MLL-AF9 or MLL-AF9/candidate gene driven leukaemia was monitored through flow cytometry assessment of YFP percentage (in MLL-AF9 solo control arm) and dual YFP/mCherry percentages (in MLL-AF9/candidate gene test arms) from mononuclear cells isolated from peripheral blood sampling. Monitoring and culling of mice, processing and storage of tissues was performed as described above.

The overall experimental plan for this is depicted in Figure 2.14.

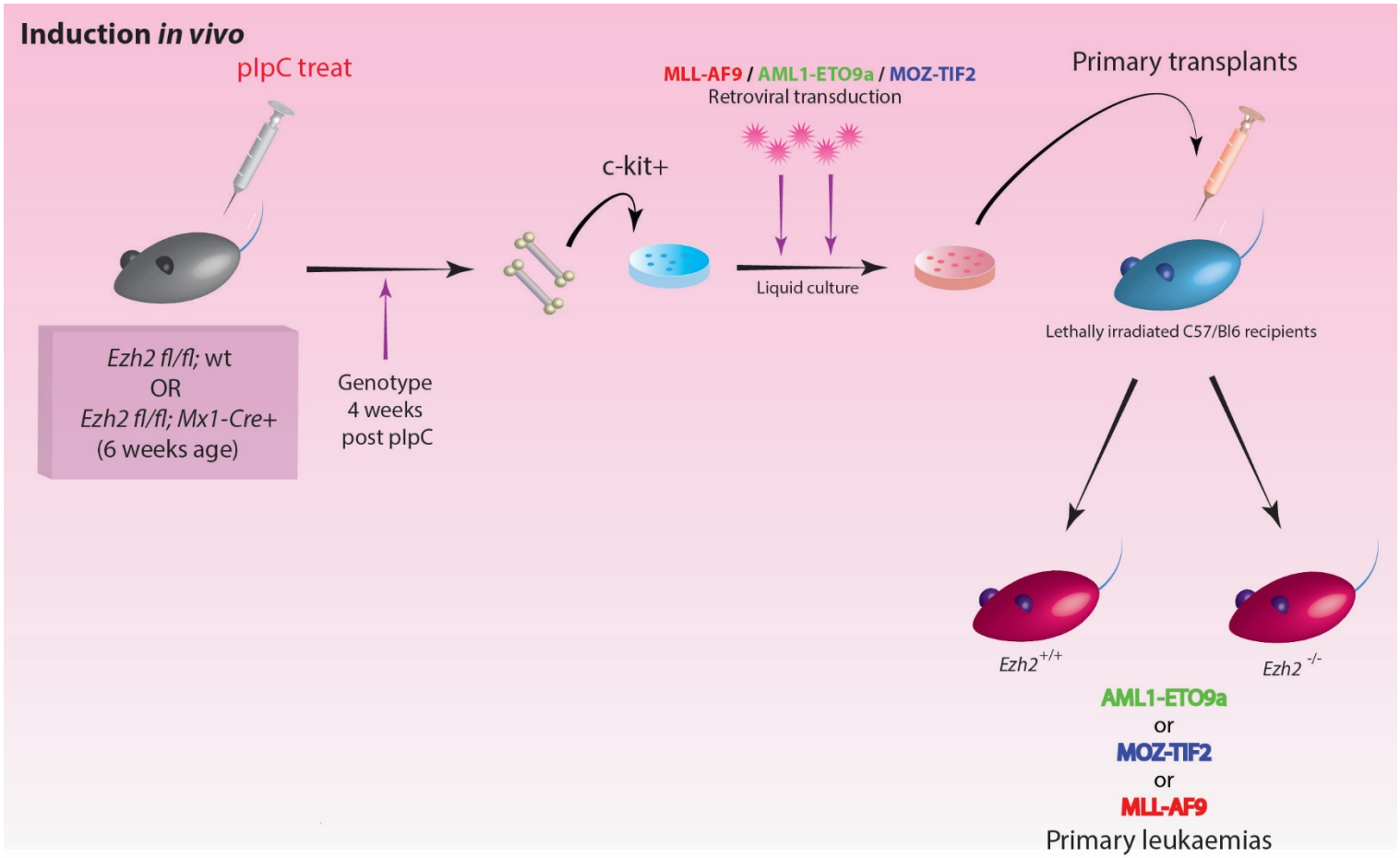


FIGURE 2.12 **Experimental plan for assessing the role of *Ezh2* in leukaemia induction *in vivo***

Ezh2 fl/fl; wt (wild-type for *Mx1-Cre*) or *Ezh2 fl/fl*; *Mx1-Cre* (heterozygous for *Mx1-Cre*) mice were treated with pIpC to induce *Ezh2* deletion in the Cre-expressing mice (confirmed by genotyping DNA extracted from peripheral blood samples 4 weeks post-treatment). BM from these two cohorts was then harvested and c-kit positive HSPCs selected and pooled in two groups (wt vs excised).
Following retroviral transduction of these HSPCs with supernatant from the 3 oncogenes listed above, these cells were transplanted into sub-lethally irradiated wild-type C57/Bl6 recipient mice. Each oncogene being tested had two arms, *Ezh2* wild type ($Ezh2^{+/+}$) versus *Ezh2* excised ($Ezh2^{-/-}$). These were closely monitored for any signs of disease and outcomes were compared with each other.

"wt" denotes wild-type for Mx1-Cre, "Mx1-Cre+" denotes heterozygous for Mx1-Cre, "c-kit+" denotes c-kit/CD117 positively selected HSPC's

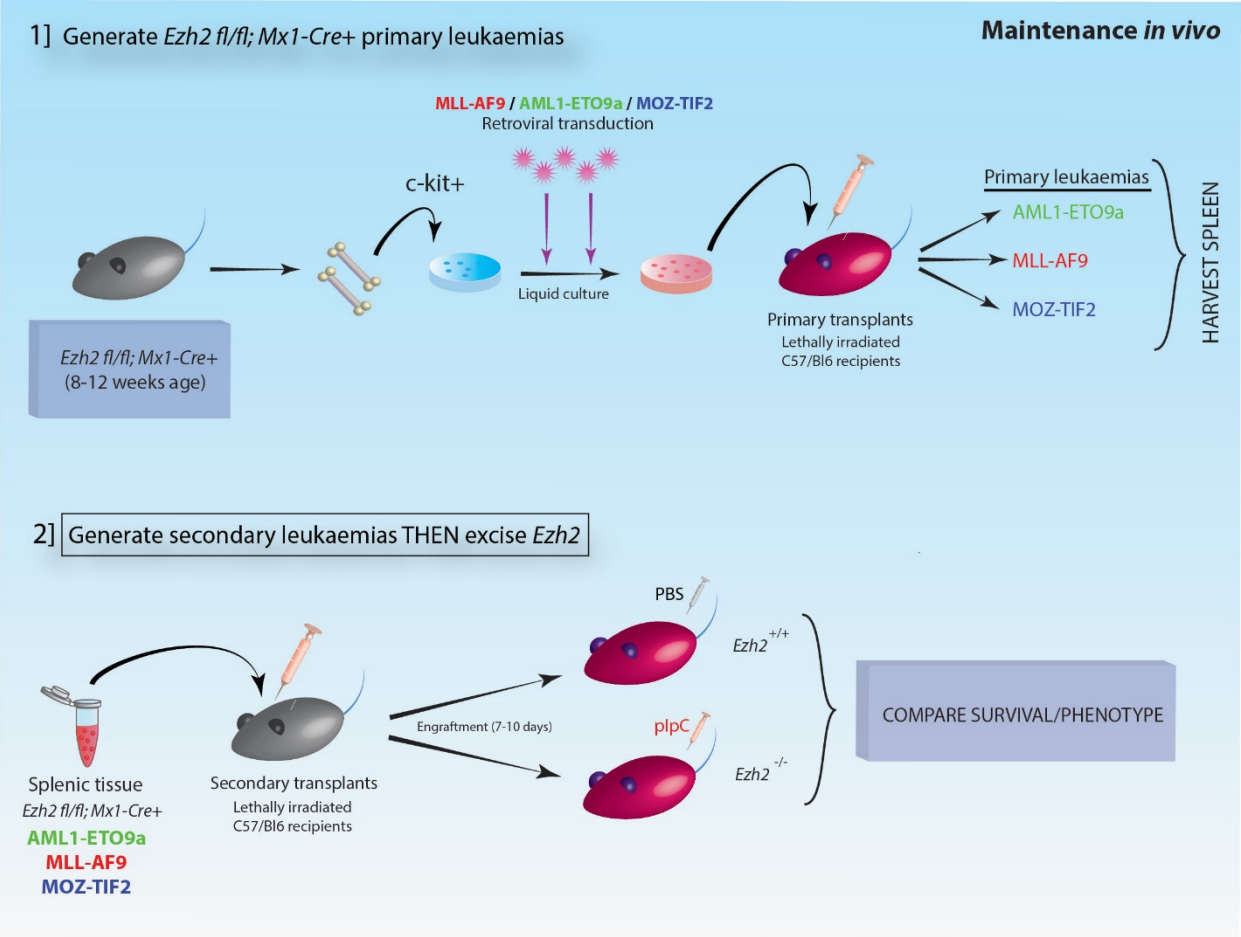


FIGURE 2.13 **Experimental plan for assessing the role of *Ezh2* in maintenance of AML *in vivo***

1) Generate *Ezh2fl/fl; Mx1-Cre+* primary leukaemias: Bone marrow from *Ezh2 fl/fl*;*Mx1-Cre+* (heterozygous for *Mx1-Cre*) mice aged 8-12 weeks was harvested and c-kit/CD117 positive HSPCs selected, pooled and then divided into 3 experimental arms. Each arm then underwent retroviral transduction using supernatant generated from the 3 oncogenes listed above. The transduced cells were immediately transplanted into lethally irradiated, wild-type C57/Bl6 recipients (n=5 per oncogene). Once leukaemia developed, these were culled and bone marrow and spleen harvested and stored. Each was genotyped to assess for any spontaneous recombination

2) Generate secondary leukaemias THEN excise *Ezh2*: Secondary transplantations into lethally irradiated wild-type C57/Bl6 recipients were then set up through tail vein injections of spleen cells from a single primary leukaemia representative of each oncogene. At the earliest sign of leukaemia, mice were given intraperitoneal pIpC injections to induce *Ezh2* deletion (vs. intraperitoneal PBS as control). These were monitored clinically for disease and differences in survival outcomes were assessed

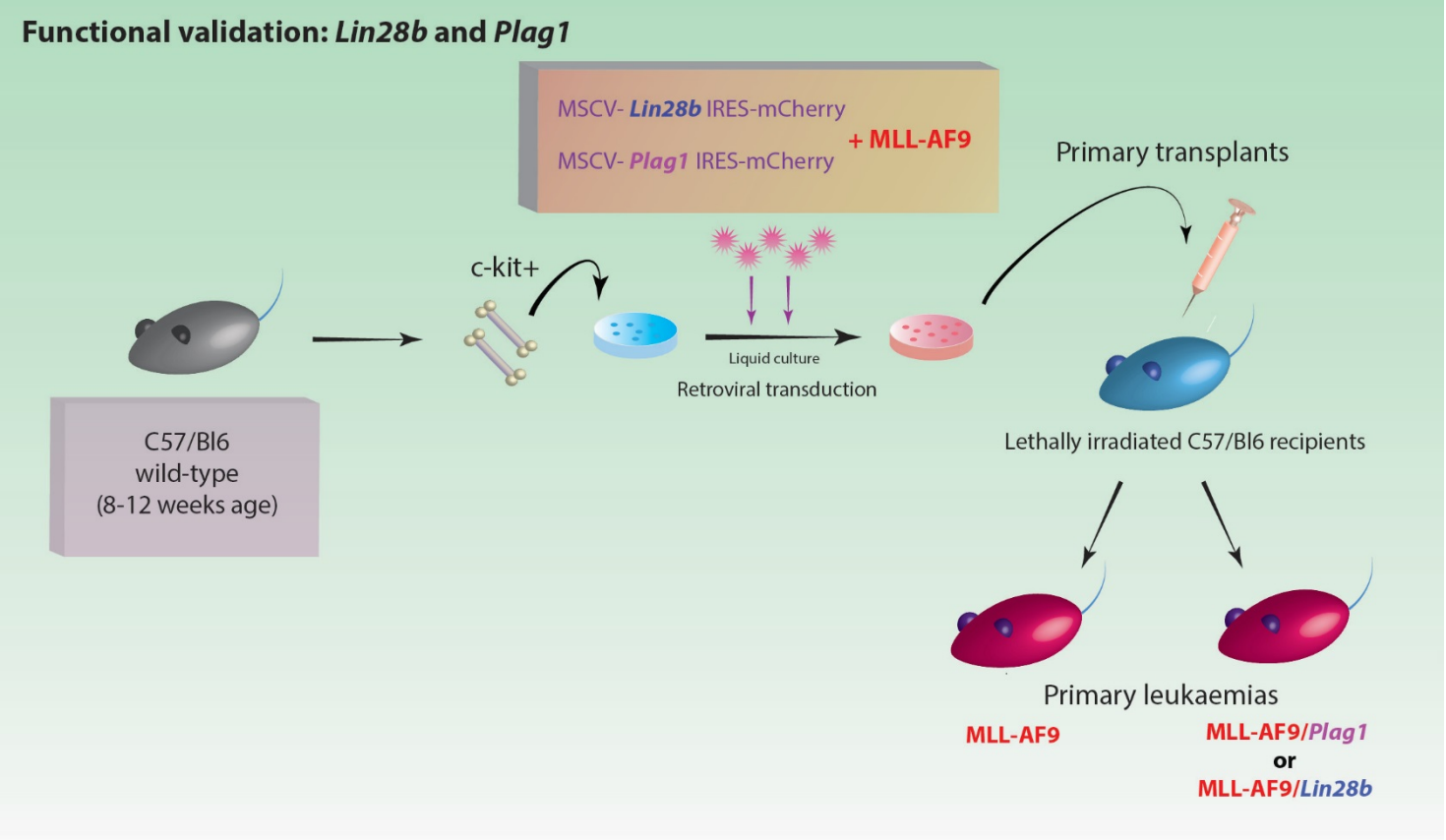


FIGURE 2.14 **Experimental plan used to functionally validate *Lin28b* or *Plag1* overexpression with MLL-AF9 *in vivo***

C57/Bl6 wild-type recipient mice aged 8-12 weeks were harvested for whole bone marrow. c-kit positive HSPCs were then selected and placed into liquid culture, divided into three arms. Each arm was then transduced with a) MLL-AF9, b) MSCV-*Lin28b*-IRES-mCherry + MLL-AF9 or c) MSCV-*Plag1*-IRES-mCherry + MLL-AF9 retroviral supernatant. After assessment of transduction efficiency by flow cytometry, cells numbers were then adjusted accordingly (as described in main text) and injected into lethally irradiated C57/Bl6 wild-type recipients. These were monitored clinically for disease and differences in survival outcomes were assessed

2.7 DNA protocols

The 3 fusion oncogenes to produce immortalise *in vitro* and produce AML *in vivo* used in this project were already available in the Huntly lab. For an analysis of the oncogenic role of two candidate genes (*Lin28b* and *Plag1*) in the induction of leukaemia, the coding sequences of *Plag1* were cloned from cDNA using the steps detailed below. For Lin28b, a GFP-tagged ORF clone containing *Mus musculus* Lin-28 homolog B (*Lin28b*) was purchased from OriGene (Cat No: MG225677). The vector that each candidate gene was cloned into was pMSCV-IRES-mCherry that was purchased from AddGene (Cat No: 52114) through an MTA.

The cloning strategies for each candidate gene were as follows:

Lin28b - EcoRI and XhoI sites introduced at 5' and 3' ends
Plag1 - EcoRI and XhoI sites introduced at 5' and 3' ends

Both were then cloned into multiple cloning site (MCS) of MSCV-IRES-mCherry pre-digested with EcoRI and XhoI

2.7.1 cDNA synthesis

To amplify the coding sequences of *Plag1*, cDNA was prepared through reverse transcription of RNA extracted from wild-type murine tissues (bone marrow or testis).
The Transcriptor One-Step RT-PCR kit (Roche, Cat No: 04655877001) was used for cDNA synthesis. Each reaction was set up as below:

PCR grade water	= 33μl
5x Reaction buffer	= 10μl
F primer (10μM dil.)	= 2μl
R primer (10μM dil.)	= 2μl
Transcriptor Enzyme	= 1μl
RNA (50ng/μl dil.)	= 2μl
TOTAL	= 50μl

The reaction was gently vortexed and placed into a BioRad thermal cycler with an initial denaturation step of 94°C of 10 seconds, followed by an annealing step of 55-62°C for 30 seconds and an elongation

step at 68°C for 1 minute per kilobase of size of PCR product. A further 25 cycles were undergone in the same format introducing an additional 5 seconds per kilobase to each elongation step. The primers used are detailed in Appendix A 2.1.

2.7.1.1 First-Strand cDNA synthesis

To synthesise first strand cDNA, the Superscript III Reverse Transcriptase Kit (Invitrogen, Cat No: 18080093) was used as per manufacturer's instructions. This method was used to isolate cDNA from transduced c-kit selected murine bone marrow cells to assess expression levels of candidate genes by qPCR. 500ng of RNA purified as per Section 2.8.14.2 was mixed with 1μl of oligo (dT) primers (50μM) and 300ng of random hexamer primers (provided in kit), 1μl dNTP (10mM) and sterile distilled water to 13μl. Samples were heated for 5 minutes at 65°C to denature the RNA secondary structure followed by a 1 minute incubation step on ice to allow the primers to anneal to RNA. After a brief centrifugation, 4μl of 5X First-Strand Buffer, 1μl 0.1M DTT, 1μl RNAse OUT (recombinant RNase inhibitor at 40 units/μl) and 1μl SuperScript III reverse transcriptase (200 units/μl) were added and gently mixed, then incubated for 5 minutes at 25°C. The reaction was then incubated at 50°C for 30-60 minutes to allow reverse transcription to take place. Finally, the reaction was inactivated at 70°C for 15 minutes.

2.7.2 Agarose Gel electrophoresis

1-2% (w/v) of agarose was mixed into 1xTAE (Tris-Acetate-EDTA buffer) in a glass jar and heated in a microwave oven to near boiling point until it had completely dissolved. It was then allowed to cool slightly and ethidium bromide was added to a final concentration of 0.1μg/ml. The gel was poured into a gel tank containing a gel comb and allowed to cool to room temperature thereby becoming set. It was then loaded into a Bio-Rad Sub-Cell compartment with the comb/top of the gel closest to the anode and submerged in 1xTAE buffer. The gel comb was then removed yielding empty wells for loading DNA samples/DNA ladder into.

10-100μl DNA for analysis were mixed with 2-20μl 5x DNA loading buffer (Bioline, Cat No: Bio-37045) and loaded into the wells in the gel, with 5μl of DNA Hyperladder II (Bioline, Cat No: Bio-33053) in separate wells to assess DNA size and quantity. Application of 80-100 volts were applied to the gel via a Bio-Rad PowerPac power supply for 30 - 90 minutes depending on the percentage of agarose in the gel and the size of DNA fragments to be visualised.

When complete, the gel was viewed under a UV light source using a Transilluminator (Life Technologies, Cat No: TFX-35M). The appropriate sized band was cut and the DNA was then extracted from the gel.

2.7.3 DNA gel extraction

Once cut from the agarose gels, each piece of gel containing DNA was weighed. Using a MiniElute Gel extraction kit (QIAGen, Cat No: 28604), the DNA was purified from the gel. Briefly, 3 volumes of QG Buffer (provided in the kit) were used per one volume of gel, then incubated at 50°C for 10 minutes (with additional vortexing during incubation to assist with dissolving). Once the agarose had been solubilised completely, one gel volume of isopropanol was added to the mixture and the sample placed in a MiniElute column for centrifugation at 13000rpm for 1 minute. The flow through was discarded and the column was washed through centrifugation with 500µl QG Buffer and 750µl of PE Buffer. Finally, DNA was eluted from the column into 30µl of provided EB Buffer after 1 minute of incubation at room temperature. This was quantified via a Nanodrop reader (Methods 2.8.14)

2.7.4 Primer design

Utilising software Primer 3 (http://primer3.ut.ee/) oligonucleotides were designed to have restriction enzyme (RE) recognition sites on both 5' and 3' ends, with 4-6 random nucleotides added 5' to the RE site of the forward primer and 3' to the reverse primer to ensure adequate enzymatic cleavage.

Oligonucleotides used to clone the coding sequences from cDNA were designed to be approximately 18-22 base pairs (bp) long, with a G-C content of 50-60% and a melting temperature of 55-65°C.
All custom oligonucleotides were ordered from Sigma-Aldrich.

2.7.5 PCR

The appropriate coding sequence from the genes of interest above was then amplified through PCR. Utilising the primers designed above the cDNA was subjected to PCR to amplify the coding sequences further. This was followed by separation using gel electrophoresis and extraction (as described above). Then a separate PCR was performed to introduce appropriate restriction sites to the

inserts using primers in order to facilitate cloning into the retroviral vector. Each PCR reaction was set up as follows:

25µl	MyTaq HS 2x Mix (Bioline, Catalogue No: Bio-25045)
0.5µl	Platinum Taq DNA Polymerase (Invitrogen, Catalogue No: 10966026)
2µl	Fwd Primer (10mM)
2µl	Rev Primer (10mM)
10ng	DNA
Rest to 50µl	PCR grade water

The PCR was set in a Bio-Rad thermal cycler as follows:

a. 96°C for 10mins (1 cycle)
b. 96°C for 30s
c. 60°C for 30s
d. 72°C for 1min (35 cycles of b-d)
e. 72°C for 10min
f. 4°C hold

Following the second PCR, the products were again run on gel electrophoresis and the appropriate sized band was cut and purified.

2.7.6 Restriction enzyme digestion

Once the PCR DNA or vector product had been purified, both were digested with the appropriate restriction enzymes (New England Biolabs). This was a double digest with a combination of two enzymes with a compatible reaction buffer. In a 50µl reaction, 2-5µg DNA was combined with 5µl of the appropriate 10X NEB restriction enzyme buffer (supplied with the enzyme), 2µl of enzyme and DEPC-treated water to 50µl. The reaction was then incubated at the recommended digestion temperature (usually 37°C) for a minimum of 2-12 hours. Digested PCR DNA products were purified using a QIAquick PCR Purification column based kit (QIAGen, Catalogue No: 28104) and eluted in water. Digested vectors or DNA were then run on an agarose gel (as described above) and the appropriate

sized band corresponding to the linearised vector was excised and purified using a MiniElute Gel extraction column based kit. Restriction enzymes used are detailed in Appendix A 2.3.

2.7.7 DNA ligation

Following restriction enzyme digestion, DNA inserts were ligated to plasmid vectors with the relative amounts calculated according to the formula:

Insert mass (ng) = Molar ratio x [Insert Size (Kb) x Vector mass (ng)]/Vector Size (Kb)

The molar ratio used was 3:1 insert:vector.
Approximately 90ng of vector DNA was used per ligation as follows:

1μl T4 DNA Ligase (New England Biolabs, Cat No: M0202S)
2μl 10X T4 DNA Ligase reaction buffer (0.5M Tris-HCl, 100mM $MgCl_2$, 100mM DTT, 10mM ATP, pH 7.5)
Vector DNA
Insert DNA (as per formula above)
to 20μl PCR grade water

The reaction was then incubated at 16°C overnight.

2.7.8 Transformation

Chemically competent *E. Coli* cells (One Shot Stbl3, ThermoFisher, Cat No: C737303) were transformed for all cloning reactions. 20μl of ligation reaction was added to 100μl of bacteria that had been slowly thawed on ice. The cells and DNA were mixed by gently flicking the tube. This was incubated on ice for 30 minutes. The cells were then heat-shocked for exactly 1 minute in a pre-heated waterbath at 42°C and immediately cooled on ice for another 30 minutes. Following this, 250μl of Luria Bertani (LB) broth media was added and the tubes were incubated for an hour in a 37°C shaking incubator. 30-100μl of transformed cells were evenly spread onto a pre-warmed (37°C) ampicillin containing agar plate and incubated overnight at 37°C

2.7.9 Miniprep plasmid DNA purification

To assess for successful ligation and integration of insert into vector, single colonies were plucked from the plates and then expanded overnight in 3ml of LB broth containing 100μg/ml ampicillin in a 37°C shaking incubator. Plasmid DNA was then purified using a QIAGen Spin Miniprep kit (Qiagen, Catalogue no: 27104). This is a buffer/centrifugation column based system that uses a modified alkaline lysis method to yield 5-20μg plasmid DNA from the bacteria. All steps were carried out at room temperature. Briefly, 2-3ml of bacteria were centrifuged at 4000rpm and resuspended in 250μl of provided Buffer P1 (containing RNAse A). 250μl of Buffer P2 was then added with gentle mixing by inversion of the tube 4-6 times. These reactions allowed for lysis of bacterial cells and denaturation of dsDNA to ssDNA. 350μl of N3 Buffer was then added with inversion of the tube 4-6 times to neutralise the lysis reaction, allowing the plasmid DNA to renature whilst longer genomic DNA would fail to reanneal properly. This was centrifuged for 10 minutes at 13000rpm and the supernatant (approximately 750μl) was decanted into the provided QIAprep spin column. The column was centrifuged at 13000rpm for 1 minute, the flow-through discarded and the column then washed with 500μl of PB Buffer to remove any trace nuclease activity. The column was then washed with 750μl PE Buffer, and a final centrifugation step at 10000rpm for 1 minute to dry was performed. The purified DNA was then eluted by transferring the column to a new 1.5ml microcentrifuge tube, adding 30μl EB Buffer. The column was then incubated for 1 minute and centrifuged at 10000rpm for 1 minute.

2.7.10 Sequence verification

Upon completion of cloning, constructs containing DNA sequences inserted into retroviral vectors as detailed above were verified through sequencing at Cambridge Biosciences (Cambridge, UK). DNA from miniprep plasmid purification was sent at a concentration of 100ng/μl and sequencing primers at 3.2pmol/μl.

2.7.11 Maxiprep plasmid DNA purification

To amplify large quantities of sequence verified constructs maxipreps were performed using a Roche Genopure Plasmid Maxi Kit (Roche, Cat No: 03143422001) according to the manufacturer's instructions. Single colony starter cultures (that had been sequence verified) were expanded as for minipreps and then diluted in 250ml LB broth containing 100μg/ml ampicillin and grown overnight in

a 37°C shaking incubator. The cultures were centrifuged at 4000rpm at 4°C and carefully resuspended in the provided Suspension Buffer (with added RNAse) and mixed gently through inversion 6-8 times. 12ml of Lysis Buffer was added with incubation at room temperature for 2-3 minutes, then 12ml pre-cooled Neutralisation Buffer was added to the suspension and mixed through inversion 6-8 times. The lysate was then cleared through centrifugation at 4000rpm for 45 minutes at 4°C. The supernatant was then passed through a provided filter paper that had been pre-folded and moistened with a few drops of Equilibration Buffer. The collected flow-through was then passed through a provided column pre-equilibrated with 6ml Equilibration Buffer. The column was then washed with 10ml Wash Buffer twice and it was inserted into a 50ml Falcon tube. The plasmid DNA was then eluted from the column with 15ml prewarmed (to 50°C) Elution Buffer. The eluted plasmid was precipitated with 11ml isopropanol and immediately centrifuged at 4000rpm for 30 minutes at 4°C. The plasmid pellet was then resuspended in 4°C 70% ethanol and centrifuged once more at 4000 rpm, 4°C for 15 minutes. The ethanol was carefully removed with a pipette tip and the DNA pellet allowed to air-dry for 10 minutes at room temperature. Finally, it was dissolved in 500μl double-distilled water and its concentration and quality measured with a Nanodrop reader.

2.8 Analysis of transcriptional programmes subverted by *Ezh2* loss

The induction and maintenance *in vitro* and *in vivo Ezh2* genetic deletion experiments gave rise to interesting and varying stage-specific roles for *Ezh2* that warranted further inquiry. Therefore, further genomic analyses of $Ezh2^{+/+}$ and $Ezh2^{-/-}$ tissues (both normal and leukaemic) were undertaken. The following sections detail the methods used therein.

2.8.1 Lineage depletion

Murine bone marrow was extracted individually from *Ezh2 fl/fl*; wild-type (wt) or *Ezh2 fl/fl*; *Mx1-Cre* (Cre+) heterozygous mice (as already described), that had been pIpC treated at 6 weeks of age to cause *Ezh2* genetic deletion in the *Mx1-Cre* heterozygous cohort. These had been previously genotyped prior to culling, using DNA extracted from peripheral blood samples taken four weeks after pIpC treatment, to ensure adequate deletion. Each mouse was processed individually throughout to ensure biological replicates were possible. In order to obtain a single-cell suspension before magnetic separation, the cells were passed through individual Pre-Separation filters (Miltenyi Biotec, Cat No: 130-041-407).

Utilising a Miltenyi Lineage Cell depletion kit (Miltenyi Biotec, Cat No: 130-090-858), following a wash step with MACS buffer, cells were resuspended in 40µl per 10^7 total cells and magnetically labelled with 10µl per 10^7 total cells of the provided Biotin-antibody cocktail containing lineage-marker antibodies for mature cells [CD5, B220, CD11b, Gr-1, Ter-119]. After an incubation period at 4°C for 10 minutes, a further 30µl MACS buffer per 10^7 total cells was added. Finally, 20µl per 10^7 total cells of Anti-Biotin Microbeads were mixed into the sample and incubated for a further 15 minutes at 4°C. The cells were then washed by adding 2ml MACS buffer per 10^7 total cells and centrifuged at 1200rpm for 10 minutes. They were then resuspended in 500µl of MACS buffer ready for magnetic separation. The cells were then loaded through into a MACS LS separation column (Miltenyi Biotec; Cat No: 130-042-401) attached to a midiMACS magnetic separator (Miltenyi Biotec; Cat No: 130-042-301). Lineage positive selected cells were trapped in the column attached to the magnetic separator and the flow through contained the lineage negative HSPCs. The column was washed three times with 3ml MACS buffer to wash through the lineage negative HSPCs. The column was then removed and flushed separately with 5ml MACS buffer to release the lineage positive HSPCs.

2.8.2 Cell counting: haemocytometer

In order to preserve cell numbers, 10µl from each lineage negative flow-through sample was taken and a total cell number was estimated using a glass 'hemacytometer' (Sigma-Aldrich, Cat No: Z359629). Following a careful 1:2 dilution with Trypan Blue dye (Thermo Scientific, Cat No: 15250061) 10µl of cell sample was carefully spread across the hemacytometer grid under a glass cover slip. The total number of viable cells (those that did not take up the Trypan blue dye) present in one corner (16 small squares of the grid) were counted. This was done across 2 to 4 such grids in each corner and then divided by the number of corners to yield an average cell number. After correcting for the initial 1:2 Trypan blue dilution, this number was multiplied by 10^4 to yield a final cell concentration of number of cells x 10^6/ml.

2.8.3 Lineage depletion: purity assessment by flow cytometry

Once lineage negative HSPC's had been isolated through the negative selection process as described above, 100µl volume of Lineage negative cells (from flow-through) and Lineage positive cells (isolated from the column) was kept aside. These were then stained for flow cytometry through the addition of 5µl of APC Lineage Positive antibody cocktail (BD Biosciences, Cat No: 558074) and incubated at 4°C for 30 minutes. Following this, the staining was inactivated by adding 1ml of 1X PBS and the cells were centrifuged for 5 minutes at 1400rpm. Finally, they were resuspended in 400µl of 1X PBS and transferred to polystyrene FACS tubes (BD Biosciences, Cat No: 352058) for flow cytometry. Flow cytometry was performed in the Cambridge Institute for Medical Research (CIMR) Flow cytometry core, on a BD Biosciences LSRFortessa 4-laser machine, using a laser (640 670/14) to detect the APC fluorochrome in the antibody cocktail. Unstained live cells were used to set initial forward area (FSC-A) and side scatter area (SSC-A) gates, to capture the cell populations. Cell doublets were excluded by gating FSC-W and FSC-H within this population. Finally, using APC cocktail stained lineage positive cells isolated from the column gates were set around the population that were negative. This was then used to calculate the percentage of Lineage negative cells isolated (i.e. negative for APC after staining) to determine the purity of the depletion. All flow analyses were done using FACSDiva analysis software. Data was then analysed further with FlowJo software (Tree Star Inc.) if required.

2.8.4 Lineage depletion: DNA extraction and genotyping

From the remaining bone marrow lineage positive fraction isolated from flushing through the LS column after magnetic selection, approximately $5x10^6$ cells were removed and DNA was extracted for genotyping. DNA was extracted using the QIAGen AllPrep DNA/RNA Mini kit as described in Methods 2.8.14. Genotyping was then performed as described in Methods 2.6.3. A comparison to the original

peripheral blood genotyping was made. As *Mx1-Cre* is predominantly expressed in haematopoietic tissues, *Ezh2* deletion, as estimated on genotyping cells isolated from bone marrow through qPCR, was consistently higher that the level of deletion estimated from genotyping peripheral blood cells.

2.8.5 Lineage depletion: Cross-linking for ChIP-sequencing

Once the lineage negative fraction isolated in the flow-through had been checked to have high purity by flow cytometry (typically >90%) the cells were resuspended in 1ml of 1X PBS in 1.5ml Eppendorf tubes to attain a concentration of 0.5-1x10^6/ml. Inside an air-flow fume hood, 27µl of 37% formaldehyde (Sigma-Aldrich, Cat No: F8775) was added resulting in a final 1% formaldehyde concentration to facilitate histone/protein crosslinking to DNA. The samples were placed on a shaker at 30rpm, room temperature for crosslinking: 5 minutes for histones and 15 minutes for proteins/transcription factors. Following this, 62.5µl of 2M Glycine (Sigma-Aldrich, Cat No: G8898; 7.5grams dissolved in 50ml ultra-sterile water) was added and mixed in to stop crosslinking. The samples were placed on a shaker at 30rpm, room temperature for 10 minutes. From this point, all samples were kept on ice or at 4°C. After crosslinking had been stopped by the addition of Glycine, the cells were centrifuged for at 1000rpm for 10 minutes at 4°C, washed in 1ml PBS and centrifuged once more at the same settings. Finally, the supernatant was aspirated, leaving the crosslinked cell pellet. This was snap frozen by submerging the Eppendorf containing the pellet into a mixture of dry ice and neat isopropanol (Sigma-Aldrich, Catalogue No: W292907) for approximately 30 seconds. All cross-linked lineage negative bone marrow samples were then stored at -80°C for later use in chromatin immunoprecipitation (ChIP).

2.8.6 Chromatin immunoprecipitation (ChIP)

Lineage depleted bone marrow HSPCs that had previously been isolated, cross-linked and stored at 1x10^6 – 5x10^6 cells at -80°C were thawed slowly on ice. For histone modifications, the 'iDeal ChIP-seq kit for Histones' (Diagenode; Cat No: C01010050) was used, and for PRC2 complex member binding the 'iDeal ChIP-seq kit for Transcription factors' (Diagenode; Cat No: C01010055) was used. 1x10^6 cells in duplicate from each condition (wt or Cre+) were used for histone modification ChIP and 5x10^6 cell pellets in duplicate (from wt or Cre+) were used for PRC2 complex member ChIP. Unless indicated otherwise, all buffers used were supplied in the Diagenode kits.

2.8.7 ChIP for histone modifications

Upon thawing, the cells were lysed sequentially using 1ml Lysis buffer iL1 for $1x10^6$ cells and then 1ml Lysis buffer iL2 with pelleting by centrifugation at 1600rpm for 5 minutes at 4°C then discarding the supernatant between each lysis step. They were then sheared using a mixture of Shearing buffer iS1 and protease inhibitor cocktail (100µl for $1x10^6$ cells; 1:200 dilution of protease inhibitor cocktail). The cell suspension was then transferred into 1.5ml TPX sonication tubes (Diagenode; Cat No: C30010009) and incubated on ice for 10 minutes. These were then secured into a precooled Diagenode Bioruptor Plus machine for sonication. Sonication on "high" setting for 10 cycles (30 seconds on; 45 seconds off) was then performed. The samples were removed and subjected centrifugation up to 2000rpm for a few seconds and transferred to a new 1.5ml TPX sonication tube. They then underwent a further 10 cycles of sonication at high setting (30 seconds on; 45 seconds off). Once shearing was complete, the samples were centrifuged at 13000rpm for 10 minutes and the supernatant containing sheared chromatin was collected for magnetic immunoprecipitation. At this point, 1/10th volume (10µl) of sheared chromatin was removed and kept aside at 4°C as "Input" for each sample, representing total chromatin that was not subjected to immunoprecipitation.

20µl of Diamag protein-A coated magnetic beads were used per immunoprecipitation (IP) reaction. These were initially washed with 4 times with 40µl of 1x ChIP buffer iC1 by resuspending them manually using a Gilson pipette inside an Eppendorf tube. The tube was then placed into a 1.5ml Diagenode Magnetic Rack (Cat No: B04000003) and after one minute the beads were completely captured by the magnet against the inner side of the Eppendorf tube. The supernatant (wash) was removed, the Eppendorf then lifted from the Magnetic Rack and the beads resuspended with 40µl of the next wash. After the fourth wash the beads were suspended in 20µl 1x ChIP buffer iC1.

The final immunoprecipitation reaction was prepared as below:

ChIP reaction mix							
No of Immunoprecipitations	5% BSA (µl)	200x PI cocktail (µl)	5x buffer iC1 (µl)	Sheared chromatin (µl)	Magnetic beads (µl)	ChIP-seq grade water (µl)	Antibody (µl)
1	6	1.5	56	100	20	116.5 to x	1

Each IP reaction was then incubated at 4°C overnight under constant rotation at 40rpm on a rotating wheel. After overnight incubation, the tubes were placed into the Magnetic Rack allowing the immunoprecipitated chromatin to be separated. This was then washed and eluted sequentially using Wash buffers iW1-4 at 4°C and Elution buffers iE1/iE2 at room temperature respectively. At this stage

the Input samples were also mixed with the same volume of Elution buffer iE2 and both Input and IP reaction were incubated at 65°C for a minimum of four hours in a thermomixer at 800rpm to facilitate de-crosslinking.

Finally, each Input and IP reaction were purified using a QIAgen PCR purification kit. Purified IP samples were then frozen at -20°C for later analysis.

2.8.8 ChIP for PRC2 complex members

These steps were performed as per the iDeal ChIP-seq kit for Transcription factors kit and largely overlap with those described above. Cell lysis, chromatin shearing and magnetic immunoprecipitation were performed in a similar way as described, using the buffers provided. The same proportion of sheared chromatin from each reaction was kept aside as Input. In this protocol the DiaMag Protein-A coated magnetic beads were pre-incubated for four hours with antibody at 4°C on a rotating wheel at 40rpm before immunoprecipitation. Following overnight immunoprecipitation, the beads from each IP and Input were washed and antibody-protein-DNA complexes eluted in a similar way. After decross-linking the samples were purified using a QIAgen PCR purification kit. Purified IP samples were then frozen at -20°C for later analysis.

2.8.9 ChIP qPCR analysis

To validate the ChIP reactions prior to sequencing, qPCR analysis was performed. As the repressive histone mark H3K27me3 and activating H3K27Ac and H3K4me3 histone marks were being assessed, qPCR to assess relative levels at control regions, including the first intron of *HoxC10* (normally repressed and at low level expression) and the promoter of *Gapdh* (normally constitutionally active and present at high levels) was performed using primers detailed in Appendix A2.1.

2.8.9.1 H3K27me3 ChIP

The expectation was that *HoxC10* would be enriched compared to Input in H3K27me3 ChIP in Cre- conditions and following *Ezh2* loss (i.e. Cre+ condition) due to loss of H3K27me3 this enrichment would disappear. *Gapdh* would be undetectable compared to Input in both Cre- and Cre+ conditions.

2.8.9.2 H3K27Ac and H3K4me3 ChIP

The expectation was that there would be no difference for *HoxC10* enrichment between Cre- and Cre+ conditions compared to Input for H3K27Ac and H3K4me3 ChIP, and *Gapdh* would remain enriched at similar (high) levels for both histone marks between Cre- and Cre+ conditions.

Each qPCR reaction was performed using Agilent Technologies 2X Brilliant III SYBR Green qPCR Master Mix (Agilent, Cat No: 600882-51) with the provided reference dye. The reactions were performed in duplicate for each condition on 96 well plates using a Stratagene Mx3000P PCR machine:

1x Reaction

SYBR Green	5µl
Primers (F+R)	0.6µl (final concentration 6uM)
Rox Ref Dye (1:500)	0.15µl
H_2O	3.75µl
DNA (1:5 dilution)	0.5µl

The PCR reaction was set at the following conditions:

1 cycle	95°C, 03:00 minutes
40 cycles	95°C, 00:10 seconds
	60°C, 00:20 seconds
1 cycle	95°C, 01:00 minute
	95°C, 00:30 seconds

The results were analysed in MxPro – Mx3000P software version 4.10, where the threshold fluorescence was set at 0.1 dRn. By including the Input for each ChIP condition as a comparator, it was possible to calculate the percentage enrichment compared to Input for each primer, in order to validate the ChIP performed before amplification prior to sequencing.

2.8.10 Amplification of ChIP products

Utilising the Illumina TruSeq ChIP sample prep kit (Illumina, Cat No: IP-202-1012) the ChIP products containing fragmented DNA post sonication were subjected to an end repair process. The ChIP samples were mixed with a resuspension buffer and End-Repair mix and incubated in a pre-heated thermal cycler for 30 minutes at 30°C. Following this, the samples were purified using DNA Clean and Concentrator Kit (Zymo Research, Catalogue No: D4014) where the end repaired DNA fragments were bound to a column through centrifugation, washed twice and eluted to recover purified DNA. The addition of A-bases to 3' end of the DNA fragments was then performed by resuspending the DNA with Resuspension Buffer and A-Tailing Mix and incubating in a pre-heated thermal cycler as follows:

37°C	30 minutes
70°C	5 minutes
4°C	Hold

Libraries were then prepared through ligation of specific indexed Adapters to each sample using a Ligation Mix and diluted Adapter Index. Multiplex indices were chosen following manufacturer's recommendation. The ligations were incubated in a preheated thermal cycler for 10 minutes at 30°C followed by the addition of Stop Ligation buffer to inactivate the ligation. Upon completion, the samples were purified once more using the DNA Clean and Concentrator Kit as described above.

The A-tailed, indexed ChIP samples were then PCR amplified using PCR Primer cocktail and master mix supplied in the Illumina TruSeq kit using a thermal cycler at the following conditions:

98°C	30 seconds	
98°C	10 seconds	x18 cycles
60°C	30 seconds	
72°C	30 seconds	
72°C	5 minutes	
4°C	- Kept on hold	

2.8.11 Size selection of Library

In order to select the appropriate sized precipitated DNA fragments (250 - 450 base pairs), the samples were then loaded into a 2% agarose low melting gel (prepared by dissolving 2g Agarose low-melting Bio-Rad in 100ml 1XTAE buffer and Invitrogen SYBR-Green as a nucleic acid stain) in an electrophoresis device specifically used for ChIP-seq samples and run at 90-120 volts for 60 minutes. The 100bp Hyperladder II from BioLine was run at either end of the gel to characterise the size of the DNA fragments. Through UV light exposure the correct size range of fragments were identified compared to the DNA ladder, and cut out with a fresh scalpel, with care being taken to avoid contamination with adapter dimers (running at 126 base pairs in size).

The cut-out gel pieces were then weighed and the DNA fragments (250 - 450 base pairs) within were extracted using a QIAgen MiniElute Gel Extraction Kit (QIAgen, Cat No: 28604). Utilising the column extraction system provided within the kit, a maximum of 400μg gel piece per column was dissolved using Buffer QG at 50°C for 10 minutes maintaining a pH of <7.5 using 3M sodium acetate and then loaded onto the column for centrifugation. DNA was adsorbed onto the silica membrane of the column during centrifugation and unwanted impurities, such as salts, unincorporated nucleotides, agarose and SYBR-Green were discarded in the flow through. Further wash and spin steps were performed before a final elution step with a 10μl of Elution Buffer containing 10mM Tris.Cl at pH 8.5 to yield the purified DNA libraries selected by gel electrophoresis.

Appropriate size selection of these samples was then confirmed through usage of an Agilent Technologies 2100 Bioanalyzer system (Agilent, Catalogue No: G2940CA) and an Agilent DNA 1000 kit (Agilent, Catalogue No: 5067-1504). Briefly, a microfluidics chip within the Agilent DNA 1000 kit was prepared by loading the provided gel-dye mix and spreading evenly using the Chip Priming station. Five microlitres of DNA marker and 1μl ladder (provided in the kit) were then loaded and then 1μl of each gel purified sample were loaded. The chip was then vortexed for 1 minute at 2400rpm and inserted into the Bioanalyzer. Each microfluidics chip was able to analyse up to 12 samples. The results were analysed in Agilent 2100 Expert software, generating both a gel electrophoresis trace for all the samples and individual electropherograms for each sample. It was then possible to ascertain the quality and average fragment length (in base pairs) of each sample after gel purification and identify any that had residual primer-dimer contamination and required a repeat size-selection.

2.8.11 Library Quantification

KAPA Library Quantification kits (KAPA, Cat No: KK4824) were used to quantify each library. Each sample was diluted to 1:10^6 using PCR-grade water. The kit also provided a set of six DNA standards of dsDNA concentrations on a logarithmic scale to quantify against. Both samples and standards were loaded in duplicates into a qPCR reaction using KAPA SYBR FAST qPCR Master mix and PCR-grade water as follows:

KAPA SYBR FAST qPCR Master mix	12µl
PCR-grade water	4µl
Diluted DNA library or DNA standard (1-6)	4µl
	20µl

The qPCR reaction was run on a Stratagene Mx3000P PCR as follows:

95°C	5 minutes	
95°C	30 seconds	x35 cycles
60°C	45 seconds	

Using MxPro software, the C_t values for each standard and sample were extracted. The mean C_t values for each standard (considering the logarithmic increases) were then plotted on a linear scale allowing a standard line of best fit curve to be constructed. The mean C_t values for each sample were then assessed against this allowing a concentration for each sample to be calculated after correction for the initial 1:10^6 dilution.

Finally, a size adjusted concentration was calculated for each sample, to take into account the difference in size between the average fragment length of the library (as identified by the Bioanalyzer) and the DNA standard (452bp):

$$\text{Average Concentration of sample (pM)} \times \frac{452}{\text{Avg. fragment length (bp)}} = \text{Size adjusted concentration (pM)}$$

2.8.12 High throughput Sequencing

Once the samples had been correctly quantified (according to size-adjusted concentrations) they were pooled for multiplexing for single-read sequencing on an Illumina HiSeq 4000 machine at the Genomics Core, CRUK Cambridge Institute. Each lane of sequencing was able to accept 4 pooled samples with a total amount of 225fmol in 15µl volume. Therefore, up to 56.25fmol of DNA from each labelled sample was used.

2.8.13 Bioinformatics analysis

All bioinformatics analysis was performed by Dr Eshwar Meduri in the Huntly lab. This included analysis of RNA-sequencing and ChIP-sequencing data.

2.8.13.1 RNA sequencing results

Paired end RNA-seq reads were quality filtered similar to Chip-Seq reads and were mapped using STAR[3] against the mouse genome (mm10). Read counts were quantified with HTSeq[4] and Differential expression analysis was carried out with these counts using Bioconductor package DESeq2. [5]

The differential gene expression list generated for all RNA-sequencing experiments was rearranged by p adjusted values and filtered to include only p <0.05 results (for significantly down- or up-regulated genes). It was then sorted by Log_2fold change and filtered to include only genes that had Log_2fold changes in expression -0.5 < (x) > 0.5. Volcano plots to illustrate differences in gene expression between *Ezh2*$^{+/+}$ and *Ezh2*$^{-/-}$ genotypes in different conditions were also provided by Dr Eshwar Meduri.

2.8.13.2 ChIP-sequencing results

Adapter sequences were trimmed for all paired end reads and mapped against mm10 reference genome using Bowtie2. [6] Uniquely mapped reads were retained and peaks were called using SICER [7] with W200 and G600 for broad peaks (histone marks) and MACS2[8] for narrow peaks (Ezh2 and Eed) with –nomodel and –nolambda parameters. Motif analysis and peaks were annotated using HOMER. [9] Peaks in intergenic regions were assigned to genes if they were within the 100kb window from the TSS. Differential binding analysis was performed using DiffBind. [10] Overlapping analysis of peaks was performed using intersectbed from bedtools. [11]

After peak calling, differentially bound histone and protein peak lists generated from *Ezh2*$^{+/+}$ and *Ezh2*$^{-/-}$ conditions were initially filtered to select all those peaks with a False Discovery Rate (FDR) of <0.01.

They were then rearranged in order of fold change and only those with Fold changes -1.5 < (x) > 1.5 were selected. When comparing against gene expression changes from RNA-seq, all intergenic peaks further than 100kb from the transcriptional start site (TSS) were removed, yielding a final list of peaks.

2.8.14 DNA/RNA extraction

All DNA and RNA samples from a maximum of $5x10^6$ cells (mouse bone marrow/spleen, *in vitro* cultured mouse bone marrow cells, *in vitro* cultured mouse spleen AML tumours, human AML cell lines and primary AML samples) were extracted through use of a QIAGen AllPrep DNA/RNA Mini kit (QIAGen Cat No: 80204). Briefly, this kit allowed for purification of genomic DNA and total RNA simultaneously from animal cells.

2.8.14.1 DNA

The cells were lysed and homogenised with 600µl RLT Lysis Buffer (QIAGen, Cat No: 79216) and 6µl 14.3M β-mercaptoethanol, then placed into the DNA mini column provided. A high-speed centrifugation kit allowed DNA to bind to the column and the total RNA to pass into the flow-through. The DNA was then washed using the included wash buffers and finally eluted through centrifugation at high speed with the provided elution buffer.

2.8.14.2 RNA

Total RNA from the original flow-through was mixed with one volume of 70% ethanol and a maximum of 700µl was transferred to the RNeasy spin column provided. A centrifugation step allowed RNA to bind to the column. This was then sequentially washed using the provided RNA wash buffers and finally eluted in RNase-free water.

All DNA and RNA samples were quantified using a Thermo Scientific NanoDrop 1000 Spectrophotometer. One µl of sample was pipetted onto a pedestal on the instrument attached to the receiving fibre optic cable, with a second pedestal containing a source fibre optic cable being placed directly over this, causing the sample to bridge the gap between both ends. Using a pulsed xenon flash lamp, a light source was passed through the source cable through the sample and measured through the spectrometer to analyse the light. The instrument then provided a reading of DNA/RNA purity and concentration via supplied PC based software (NanoDrop 1000 v 3.8).

2.8.15 RNA-sequencing

RNA was extracted from cells as described above. Purity and quantity were assessed using the Nanodrop 1000 instrument. Approximately 5 μg RNA was submitted for paired-end sequencing on an Illumina HiSeq 4000 machine at the Genomics Core, CRUK Cambridge Institute. Library preparation and processing of samples prior to sequencing was performed at the Institute as part of a Core Facility service.

2.8.16 RNA-sequencing of MLL-AF9 tumours subjected to Ezh2 pharmacological inhibition

A satisfactory time point for H3K27me3 reduction following GSK343 treatment of two MLL-AF9 spleen tumours *in vitro* was identified, from performing the time course flow cytometry experiment detailed above. RNA was then extracted from the remaining DMSO and GSK343 treated cells in culture, quantified and sent for sequencing as per Methods 2.8.16.

2.9 Statistical Methods

The results detailed in this thesis are expressed as mean ± standard error of the mean (SEM) with 95% confidence intervals in brackets (95% CI). Error bars on graphs represent SEM. A Student's two-tailed t-test was used to compare colony formation in human and murine primary cells and cell lines treated with DMSO and GSK343. The Logrank test was used to compare survival outcomes for all *in vivo* experiments Prism 7 (GraphPad software).

References

1) Daniel A. Arber, Attilio Orazi, Robert Hasserjian, Jurgen Thiele, Michael J. Borowitz, Michelle M. Le Beau, Clara D. Bloomfield, Mario Cazzola and James W. Vardiman. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*. 2016; 127(20):2391-2405.

2) Skarnes, W. C. *et al.* A conditional knockout resource for the genome-wide study of mouse gene function. *Nature.* 2011; 474, 337–42.

3) Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013 Jan 1;29(1):15-21.

4) Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015 Jan 15;31(2):166-9.

5) Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.

6) Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods.* 2012; 9, 357-9

7) Xu, S., Grullon, S., Ge, K. & Peng, W. Spatial clustering for identification of ChIP-enriched regions (SICER) to map regions of histone methylation patterns in embryonic stem cells. *Methods in molecular biology.* 2014; 1150, 97-111.

8) Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome biology.* 2008; 9, R137.

9) Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell.* 2010; 38, 576-589.

10) Ross-Innes, C. S. *et al.* Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature.* 2012; 481, 389-393.

11) Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010; 26, 841-842.

Appendix A 2.1 Oligonucleotide primers

Cloning Primers

Name	Sequence
Lin28 Cloning F_EcoRI	TATGGAATTCATGGCCGAAG
Lin28 Cloning R_XhoI	TATGCTCGAGACAGTCTTTTTCC
Plag1cloneEcoRIF	TATGGAATTCATGGCCACTGTCATTCCTG
Plag1cloneXhoIR	TATGCTCGAGCAGGGCTACACAAGCACCTC

cDNA synthesis primers

Name	Sequence
Plag1 F primer	ATGGCCACTGTCATTCCTG
Plag1 R primer	CAGGGCTACACAAGCACCTC

Genotyping primers

Name	Sequence
Ezh2 fl/fl primer F	TATGGGGGATGAAGTTCTGG
Ezh2 fl/fl primer R	CCACTGTCATGGCAACAAGA
Ezh2 unexcised primer F	CTTCTTCCTCCGGGTCTCTC
Ezh2 unexcised primer R	GATCTTGGGGCATTGCAGTA
Ezh2 F1 primer	CCTGCAGCAGTTTCCTTTCTTA
Ezh2 R1 primer	CCCATGTGGTAGGCAGAGAA
Ezh2 R2 primer	CAGCGACACCCCAGAAAATT

ChIP qPCR primers

Name	Sequence
Hoxc10 int1 F	ATATGCACTCTTTTGGCGCG
Hoxc10 int1 R	TCATCTCCACAAGCACAGCC
Gapdh F	TCATCCACCTCCCCACAGTA
Gapdh R	TTTCTTCTCCTGCAGCCTCG

RT qPCR primers

Name	Sequence
Lin28b F	ATGGCACTTCTTTGGCTGAG
Lin28b R	ATAGGTGGAGACGGCAGGAT
Plag1 F	GACAAGGCCTTTAACAGTGTTG
Plag1 R	TCAGGAGAGTGAGTAGCCATG

Appendix A 2.2 Antibodies

Name	Clone/Catalogue Number	Manufacturer
H3K27me3	07-449	Millipore
H3K4me3	C15410003	Diagenode
H3K27Ac	ab4729	Abcam
Anti EZH2	D2C9	Cell Signaling Technology
Anti EED	CS204392	Millipore
Total H3 Pacific-Blue	12167	Cell Signaling Technology
H3K27me3 Alexa Fluor 647	ab205729	Abcam
CD117 (c-kit) magnetic beads	130-091-224	Mitenyi
APC Lineage Positive Cocktail	558074	BD Biosciences
Annexin V-FITC kit	559763	Miltenyi Biotech
7-AAD	51-68981E	BD Bioscience
LIN28B	37362	Signalway Antibody
Anti PLAG1	TA335811	OriGene
Anti β-Actin	ab6276	Abcam
Anti GAPDH	ab9484	Abcam
Donkey Anti-Rabbit Green	92632213	Odyssey
Goat Anti-Mouse Red	92668070	Odyssey

Appendix A 2.3 Restriction Enzymes & Buffers

Name	Catalogue Number	Manufacturer
EcoRI	R0101S	New England Biolabs
XhoI	R0146S	New England Biolabs
CutSmart Buffer	B7204S	New England Biolabs

Chapter 3

Characterising the role of EZH2 in the induction of Acute Myeloid Leukaemia

3.1 Aims

This section details the results of experiments designed to investigate whether genetic loss of Ezh2 prior to leukaemogenesis mediated by ectopic expression of three disparate driver oncogenes, serves to influence AML induction.

3.2 Introduction

This experiment was designed to test the genetic requirement of *Ezh2* in the generation of leukaemia, using HSPC from existing mice possessing a conditional *Ezh2 fl/fl* (floxed) allele that were also heterozygous for the expression of the Mx1-Cre recombinase. By activation of *Mx1-Cre* with polyinosinic:polycytidylic acid (pIpC), recombination *in vivo* occurred resulting in *Ezh2* excision. Through deletion of Ezh2 prior to immortalisation/transformation, an assessment of whether Ezh2 is a requirement for induction of leukaemia or not was made. $Ezh2^{+/+}$ and $Ezh2^{-/-}$ HSPCs were harvested and retrovirally transduced with three specific AML-associated fusion oncogenes, representing several AML genotypes:

MLL-AF9:	potent oncogene capable of producing short-latency AML. Human disease is associated with intermediate-poor prognosis. Transforms via gene activation. Positive control for maintenance experiments[1,2]
AML1-ETO9a:	the t(8;21) is the commonest translocation seen in adult AML, associated with a repressive gene expression programme that transforms through the activation of oncogenic gene programmes. Typically associated with a good prognosis
MOZ-TIF2:	a strong oncogene, associated with poor prognosis

Experimental designs similar to the above for assessing epigenetic regulators, have been previously performed in the Huntly lab, validating this strategy as a means to assess the role of *Ezh2* in AML.[3]

The importance of *Ezh2* in the maintenance of MLL-AF9 leukaemias has been demonstrated in recent years.[1,2] The mixed lineage leukaemia (MLL) gene is involved in translocations in AML cases which invariably have a poor outcome. MLL is a histone methyltransferase, containing a SET domain and is involved in post translational trimethylation of lysine 4 on histone 3 (H3K4me3), facilitating transcription. In MLL-rearranged leukaemias the N-terminus of MLL is fused to one of over 70 fusion partner genes, with all of these rendering loss of the SET domain. These fusions activate pro-leukaemic changes in gene expression. Many fusion partners of MLL are members of the super-elongation complex, which is a key regulator of transcriptional elongation. MLL-AF9 overexpression consistently immortalises haematopoietic progenitors and generates aggressive short-latency AML *in vivo*.[4] Characteristic pathological features of the disease included enlarged pale livers and spleens, pale femurs and marked peripheral white cell counts, anaemia and thrombocytopenia.

The *AML1-ETO9a* transcript was shown to be an alternative spliced isoform of the *AML1-ETO* transcript, including an extra exon, exon 9a, of the *ETO* gene.[5] The commonly known *AML-ETO* fusion transcript contains *AML1* exons 1-5 and *ETO* exons 2-11, producing a 752-amino acid fusion protein AML1-ETO (AE). The addition of Exon 9a provides a stop codon after the last amino acid encoded by exon 8, resulting in a 575-amino acid protein AML1-ETO9a (AE9a). Both the larger form of AML1-ETO and AML1-ETO9a are expressed in patients with this leukaemia. Whilst its overexpression in murine bone marrow demonstrated increased self-renewal and a partial block in myeloid differentiation, AML1-ETO (via a MSCV-AML1-ETO-IRES-eGFP cDNA construct) was shown to be insufficient to generate leukaemia in murine overexpression and transplantation experiments on its own. [6] Expression of AML1-ETO9a (via a MIGR1 cDNA construct) in murine foetal liver cells from C57/Bl6 mice however, lead to the development of leukaemia when these cells were transplanted into irradiated recipient mice. The median survival time was 25 weeks from transplantation. AML1-ETO9a overexpression in wild-type C57/Bl6 bone marrow has been shown to produce leukaemia when transplanted similarly, though these have a longer latency (30-65 weeks). [7] Mice typically showed pale femurs, enlarged liver/spleens but no abnormalities in thymuses or lymph nodes. Terminal blood counts typically demonstrated a raised white cell count, profound anaemia and mild thrombocytopenia.

The MOZ-TIF2 fusion is associated with AML with inv(8) (p11q13). MOZ belongs to the MYST family of histone acetyltransferases and TIF2 is a member of the p160 nuclear receptor transcriptional coactivator family that associates with CREB binding protein (CBP). MOZ-TIF2 (via a MSCV-MOZ-TIF2-IRES-GFP construct) has transforming properties and reliably causes AML in murine bone marrow transplantation assays. [8] The original study described an aggressive leukaemia with a median latency of 90 days, with pathological findings of generalised lymphadenopathy, widespread infiltration into cervical nodes and salivary glands, enlarged spleens/livers and raised white cell counts.

Thus far, no assessments of the role of *Ezh2* in the induction of experimental AML have been described. As these oncogenes generate obvious leukaemic immortalisation and transformation *in vitro* and *in vivo* in normal HSPC, they were reliable and consistent tools to functionally assess the role of *Ezh2* in the generation (and maintenance) of leukaemia, using:

1) growth in methylcellulose and liquid culture *in vitro* and
2) *in vivo* transplantation

A comparison between $Ezh2^{+/+}$ and $Ezh2^{-/-}$ cells transformed with the same oncogenes was made using the above assays. A broad schema of this plan is provided in Figures 2.10 (*in vitro*) and 2.12 (*in vivo*)

3.3 Results and discussion

3.3.1 *Ezh2* does not appear to be a requirement for induction of AML *in vitro*

After retroviral transduction of wt (*Ezh2*$^{+/+}$) vs Cre+ (*Ezh2*$^{-/-}$) HSPCs with each of the three oncogenes, 50,000 cells were plated into 1ml of methylcellulose M3434 in duplicate (as in Methods 2.3.2 and 2.6.11) and their clonogenic/proliferative growth potential assessed with successive weekly replatings (for 4-5 replatings). The aim was to test whether *Ezh2* excised clones would be rapidly outgrown by *Ezh2* wt clones. Across all three subtypes of AML, transformation occurred between both study arms and there was no clear difference in clonogenic or proliferative potential in the *Ezh2*$^{-/-}$ transformed cell lines.

The only modest difference seen in the first plating, where consistently (across all three AML subtypes) the total number of colonies and total cell number was markedly lower in the *Ezh2*$^{-/-}$ arm than in the *Ezh2*$^{+/+}$ arm. However, this apparent disadvantage disappeared from Week 2 onwards and there was no noticeable difference between the two arms thereafter (Figure 3.1).

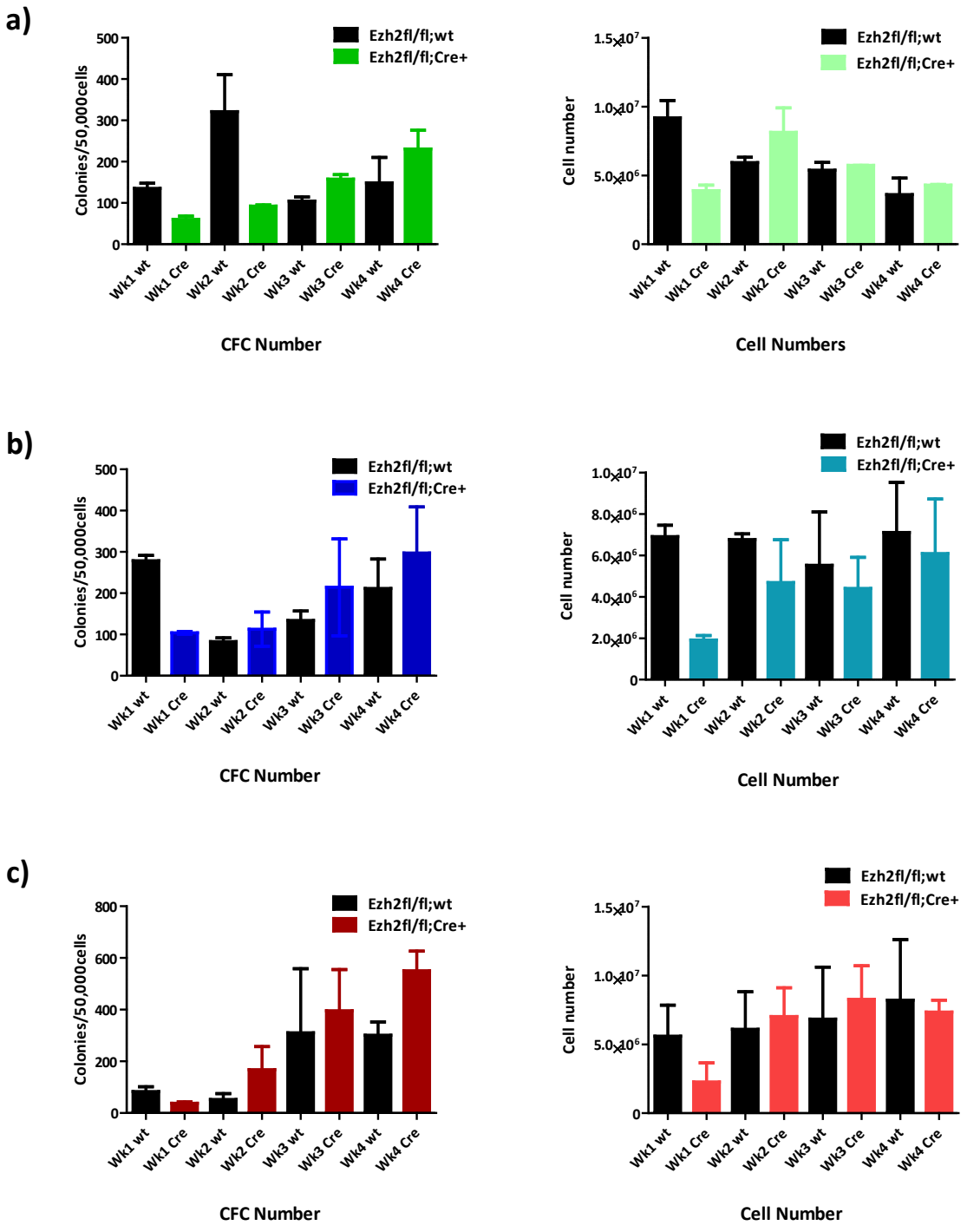


FIGURE 3.1 **Clonogenic (colony number, left) and proliferative (total cell number, right) potential of AML1-ETO9a (a), MOZ-TIF2 (b) and MLL-AF9 (c) transformed *Ezh2*$^{+/+}$ versus *Ezh2*$^{-/-}$ c-kit selected murine bone marrow cells on weekly replatings**

However, within the first week (even for the very potent MLL-AF9 fusion oncogene) transformation and immortalisation would not be expected to occur. The Week 1 colonies that formed were not dense, round and with clear margins as would be expected from an established MLL-AF9 or MOZ-TIF2 immortalised cell line. As each oncogene construct harboured a fluorescent marker readily detectable by flow cytometry (GFP for AML1-ETO9a and MOZ-TIF2 and YFP for MLL-AF9), it was possible to assess percentage of transduced cells expressing the fusion oncogene at each weekly replating and it was clear at Week 1 that only a minority were expressing the marker/fusion oncogene. With each successive replating, the proportion of cells positive for the marker (and hence fusion oncogene expression) increased and with MLL-AF9 and MOZ-TIF2, by week 3-4 90-100% cells (and by Week 5-6 for AML1-ETO9a) were positive for marker, with colonies taking on the classical immortalised appearance.

These data suggested that the *Ezh2*$^{-/-}$ c-kit selected bone marrow cells may have decreased inherent clonogenic and/or proliferative potential compared to the *Ezh2*$^{+/+}$ cells before the effects of the oncogene would take over and drive them toward immortalisation.

This prompted an assessment of the serial replating potential of untransduced *Ezh2*$^{-/-}$ vs *Ezh2*$^{+/+}$ c-kit selected bone marrow cells through weekly replatings (Figure 3.2):

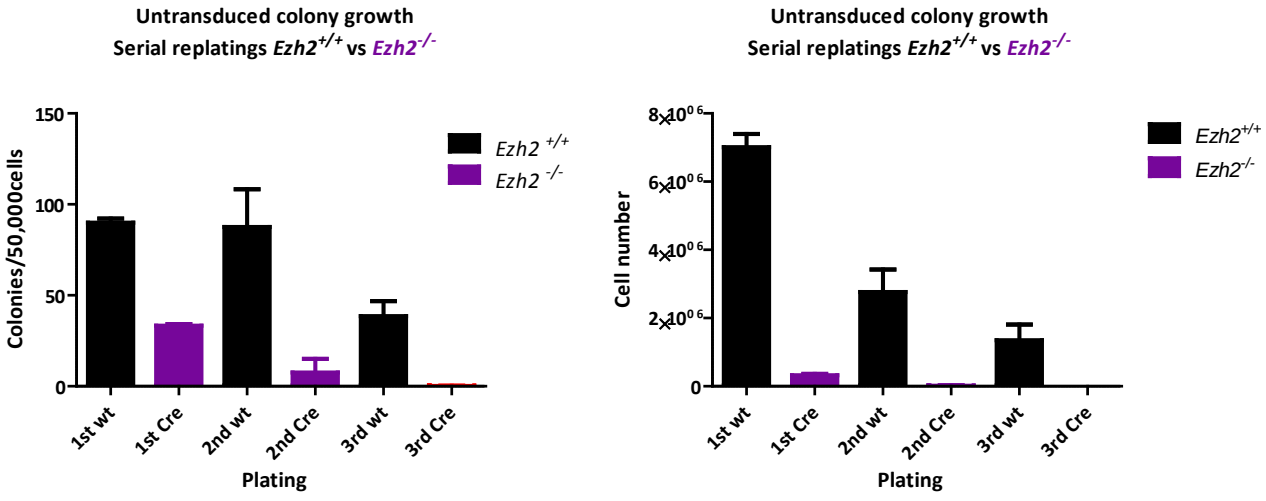


FIGURE 3.2 ***Ezh2***$^{-/-}$ **HSPCs have a decreased clonogenic and proliferative potential**
Clonogenic (colony number) and proliferative (total cell number) potential of untransduced *Ezh2*$^{+/+}$ versus *Ezh2*$^{-/-}$ c-kit BM cells on weekly replatings

This clearly demonstrated that untransduced *Ezh2*$^{-/-}$ bone marrow cells had a significant reduction in clonogenic and proliferative potential from as early as the first replating and totally exhausted by Week 2, compared to *Ezh2*$^{+/+}$ cells that replated to Week 3 and beyond. This was expanded to an assessment for 9 separate mice (all harvested at 8 weeks of age) across both genotypes, again demonstrating a clearly significantly lower clonogenic potential in the *Ezh2*$^{-/-}$ HSPCs (Figure 3.3):

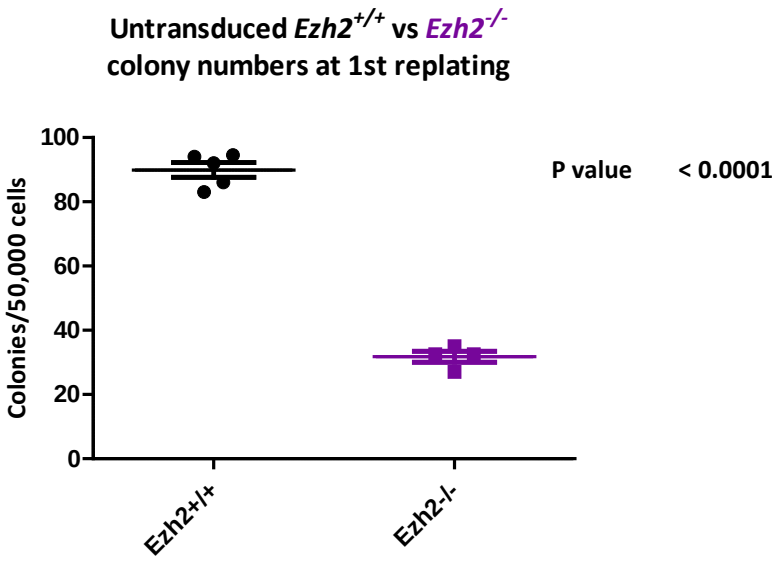


FIGURE 3.3 **Average clonogenic output at Day 7 from untransduced *Ezh2*$^{+/+}$ (n=5) versus *Ezh2*$^{-/-}$ (n=4) c-kit selected murine BM cells**

These results suggest at the early HSPC level at least, in adult murine BM cells *Ezh2* deletion significantly reduces repopulation potential *in vitro*, however does not alter outcome of the transformation or immortalisation significantly in these three AML-associated oncogenes.

To exclude differences in the size or frequency of specific progenitor populations within the HSPC compartment between untransduced *Ezh2*$^{+/+}$ and *Ezh2*$^{-/-}$ genotypes, immunophenotyping designed to characterise these populations thoroughly was kindly performed by Dr George Giotopoulos, Huntly lab (data not shown). This was performed for 12 individual mice of the same age as used in these experiments, from each *Ezh2*$^{+/+}$ (wt) or *Ezh2*$^{-/-}$ (Cre+) study arm to reduce phenotypic bias. There was no significant difference in numbers of long-term or short-term HSCs (LT-HSCs and ST-HSCs), lineage negative cells or early myeloid progenitors seen.

3.3.2 Loss of *Ezh2* prior to induction of AML *in vivo* has dramatic subtype-dependent effects

The *in vivo* experiments for each oncogene were performed next (set up as per Figure 2.12 and Methods 2.6.13). In addition, to assess the kinetics of disease induction, we performed serial peripheral blood counts to assess total white cell count (WCC) and for percentage of GFP (and therefore oncogene) expressing cells.

3.3.2.1 Loss of *Ezh2* retards MOZ-TIF2 AML induction *in vivo*

In this experiment, peripheral bleeds as early as Day 21 post-transplant demonstrated a significant difference in WCC and GFP-expression, with lower levels of both seen in the *Ezh2* excised cohort (Figure 3.4):

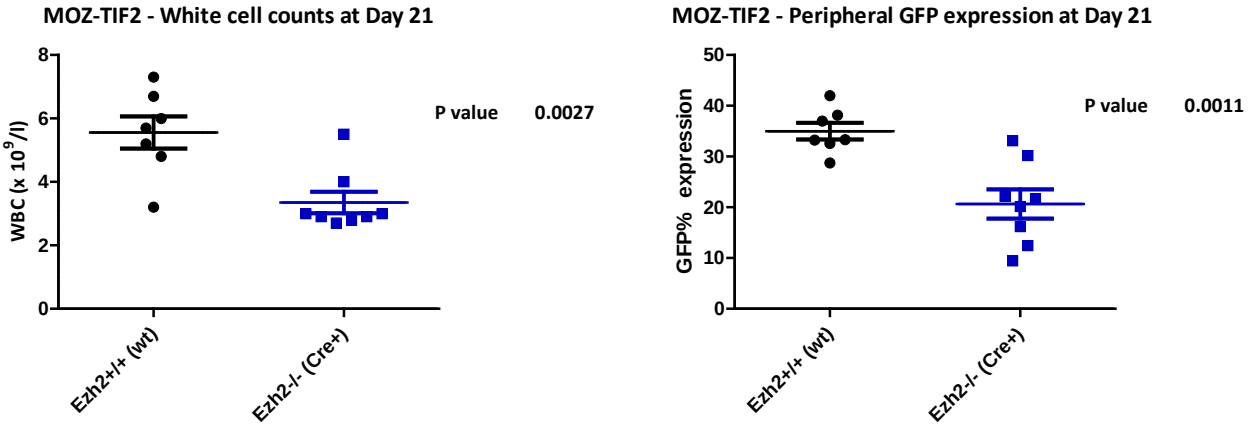


FIGURE 3.4 ***Ezh2*$^{-/-}$ MOZ-TIF2 transplanted mice demonstrate lower white cell counts and peripheral blood GFP expression at Day 21**

As the disease latency was relatively short due to the potency of MOZ-TIF2 as an oncogene, it was apparent that when assessed early i.e. just after engraftment, loss of *Ezh2* was associated with significantly less GFP positive cells in peripheral blood. The inference from this early time point was that *Ezh2* seemed to be required in leukaemic transformation or simply that *Ezh2* may be required for efficient transplantation. Mice became visibly moribund as early as Day 45 post-transplant and all were terminal between Days 49 and 90. When terminal, they exhibited significant clinical deterioration with emaciation, piloerection, hunched posture and on occasion, marked dyspnoea. At necropsy, some exhibited enlarged cervical lymph nodes consistent with the original description. [8]

In the cohort of *Ezh2*$^{+/+}$ (wt) versus *Ezh2*$^{-/-}$ (Cre+) MOZ-TIF2 transformed mice, the median survival differed significantly, from 51 (wt) to 76 (Cre+) days – a median increase of 50%, data shown in Figure 3.5:

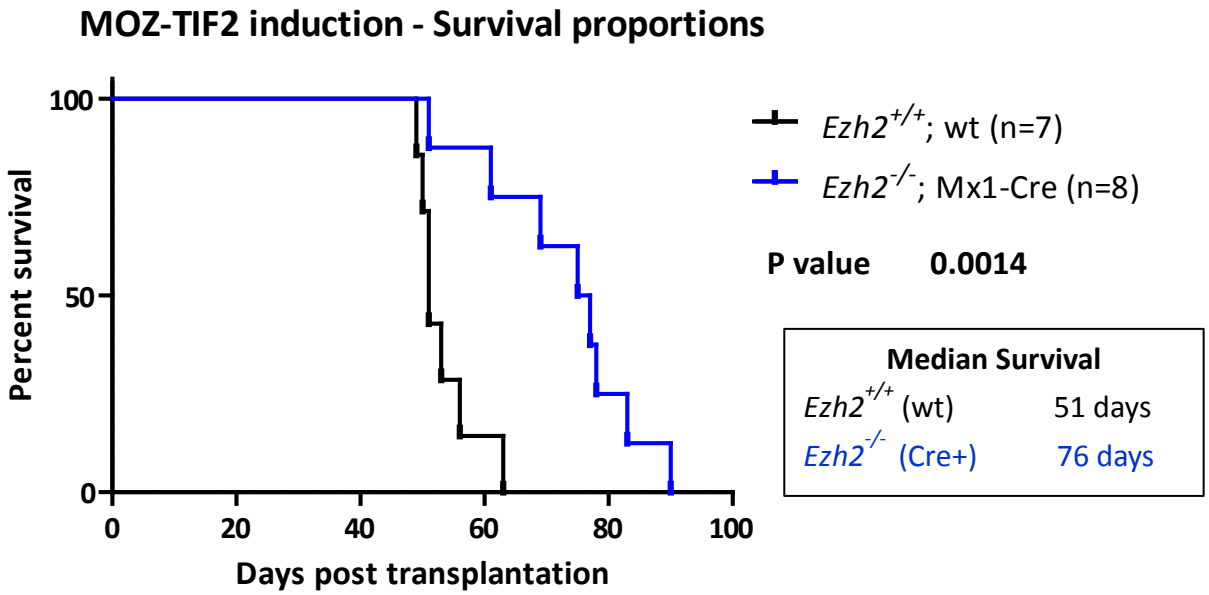


FIGURE 3.5 **Kaplan-Meier survival curve of *Ezh2*$^{+/+}$ vs *Ezh2*$^{-/-}$ MOZ-TIF2 leukaemic mice**
Ezh2$^{-/-}$ (Cre+) MOZ-TIF2 leukaemias demonstrate a significant increase in survival compared to *Ezh2*$^{+/+}$ (wt)

Notably, there were no significant differences in terminal white cell counts, spleen or liver weights (Figure 3.6, below) nor in GFP distribution between haematopoietic compartments (Figure 3.7, overleaf) between the *Ezh2*$^{+/+}$ and *Ezh2*$^{-/-}$ cohorts:

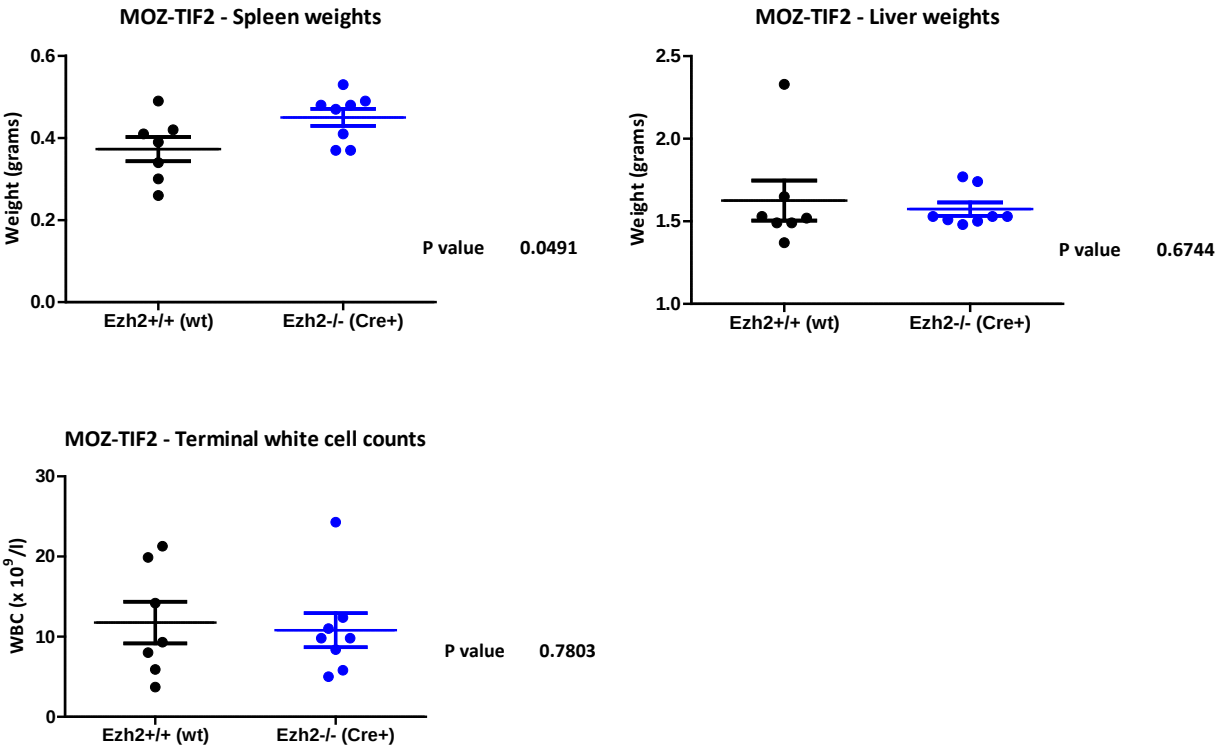


FIGURE 3.6 **Spleen/Liver weights and terminal white cell counts of *Ezh2*$^{+/+}$ vs *Ezh2*$^{-/-}$ MOZ-TIF2 leukaemic mice**

No significant differences are seen between wt and Cre+ mice

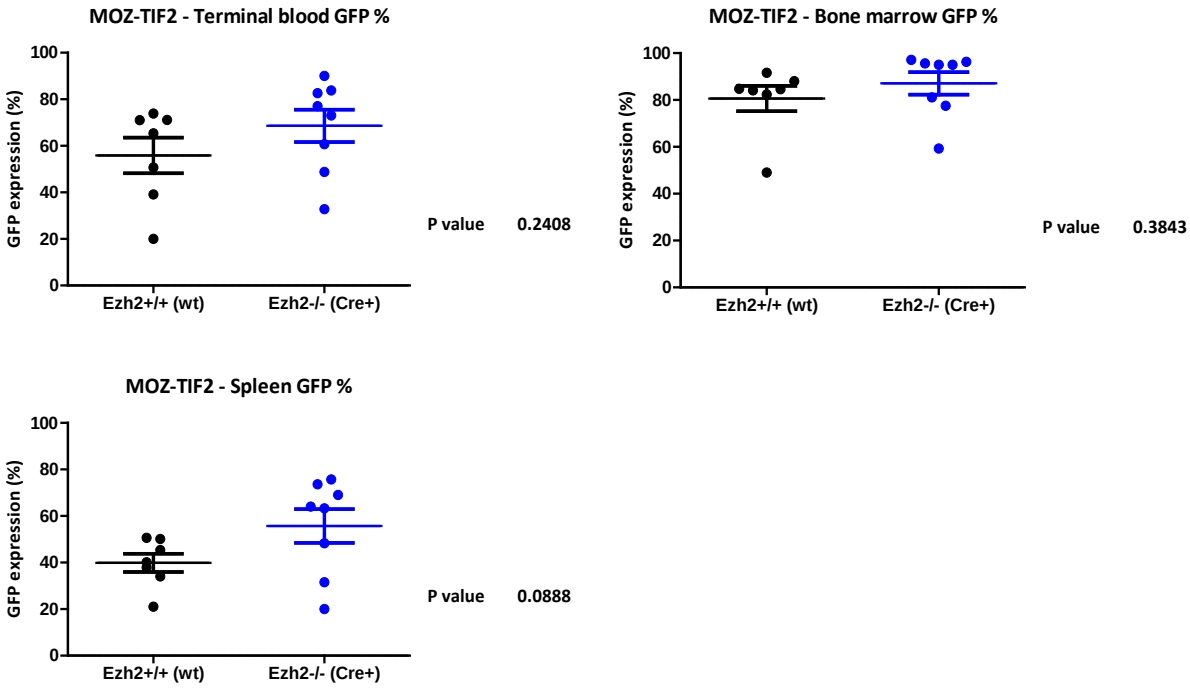


FIGURE 3.7 **GFP expression in terminal blood samples, bone marrow and spleen of *Ezh2*$^{+/+}$ vs *Ezh2*$^{-/-}$ MOZ-TIF2 leukaemic mice**

No significant differences are seen between wt and Cre+ mice

Analysis of GFP expression within tissue compartments infiltrated by AML, revealed GFP expression between 20-90% in terminal peripheral blood, 49-97% in marrow and 20-75% in spleen (Table 3.1):

						GFP%				
Mouse #	Cre status	Day culled	WCC ($x10^9$/l)	Hb (g/dl)	Plts ($x10^9$/l)	Periph Blood	Bone Marrow	Spleen	Spleen wt (g)	Liver wt (g)
4423	wt	49	5.9	136	250	71.1	82.4	40.1	0.26	1.49
4426	wt	50	19.9	126	245	73.9	84.1	37.9	0.3	1.37
4422	wt	51	9.3	154	385	50.7	84.8	34.0	0.42	1.52
4427	wt	51	14.2	158	409	20.0	49.0	21.0	0.41	1.49
4430	Cre+	51	11.0	117	134	32.8	59.3	20.0	0.37	1.48
4425	wt	53	3.7	149	326	39.1	84.6	50.6	0.34	1.53
4428	wt	56	8.0	162	292	65.4	91.6	45.4	0.39	1.65
4429	Cre+	61	9.8	103	85	48.8	77.5	31.5	0.53	1.53
4424	wt	63	21.3	148	194	71.0	88.0	50.1	0.49	2.33
4433	Cre+	69	8.4	156	151	73.0	96.3	63.3	0.37	1.53
4432	Cre+	75	24.8	115	121	82.6	95.6	73.6	0.49	1.53
4434	Cre+	77	5	100	83	60.7	95.0	64.0	0.41	1.51
4436	Cre+	78	9.8	121	30	83.8	95.0	69.0	0.48	1.74
4431	Cre+	83	12.7	71	61	90.0	97.1	75.7	0.47	1.77
4435	Cre+	90	5.8	66	122	77	81.1	48.3	0.48	1.5

TABLE 3.1 Terminal data for all MOZ-TIF2 leukaemic mice (*Ezh2*$^{+/+}$ (wt) and *Ezh2*$^{-/-}$ (Cre+) arms)

Immunophenotyping between the two cohorts (kindly performed by Dr George Giotopoulos, Huntly Lab) revealed no significant differences, with the following myeloid profile: Mac1+/Gr1+, c-kit positive leukaemia splenic tissues (also negative for CD4/B220 and Ter119). There were no differences in GFP distribution amongst haematopoietic compartments (Figure 3.7).

These data suggested no difference between phenotypes of the leukaemias seen in the *Ezh2*$^{+/+}$ and *Ezh2*$^{-/-}$ cohorts and are in keeping with described MOZ-TIF2 murine leukaemia models.[8] *Ezh2* clearly exhibited an oncogenic function in MOZ-TIF2 leukaemogenesis in this *in vivo* experiment, with its loss delaying onset of disease.

3.3.2.2 Loss of *Ezh2* accelerates AML1-ETO9a mediated leukaemogenesis *in vivo*

The MOZ-TIF2 induction experiment produced results strongly favouring a role for *Ezh2* as an oncogenic facilitator of disease. As with that experiment, an assessment of peripheral blood GFP and white cell counts was made at Day 21 for AML1-ETO9a transplanted mice. In contrast to MOZ-TIF2, there was no significant difference in GFP expression or total white cell count in the blood between AML1-ETO9a transduced; *Ezh2*$^{+/+}$ (wt) versus *Ezh2*$^{-/-}$ (Cre+) mice (Figure 3.8), suggesting that loss of *Ezh2* did not affect early AML1-ETO9a transformation. However, AML1-ETO9a transduced wild-type bone marrow cells that have been transplanted into healthy recipients have a significantly longer disease latency[7] than wild-type MOZ-TIF2 transplants, so we considered it possible that this difference between these two subtypes of AML (at this early time point) reflected the disease latencies.

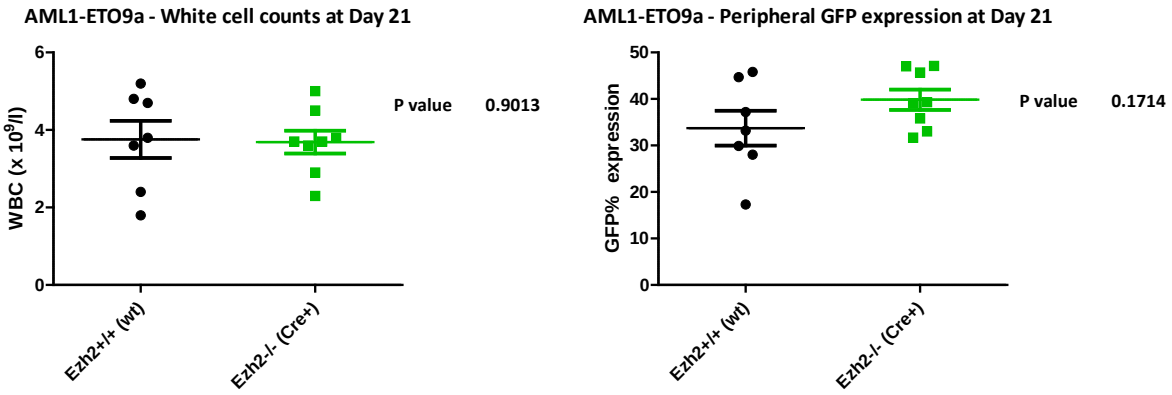


FIGURE 3.8 **White cell counts and peripheral blood GFP expression between *Ezh2*$^{+/+}$ (wt) and *Ezh2*$^{-/-}$ (Cre+) AML1-ETO9a mice at Day 21 post transplant**

Serial measurements of peripheral blood GFP percentage and WCC at regular intervals allowed us to monitor the kinetics of the disease (Figure 3.9):

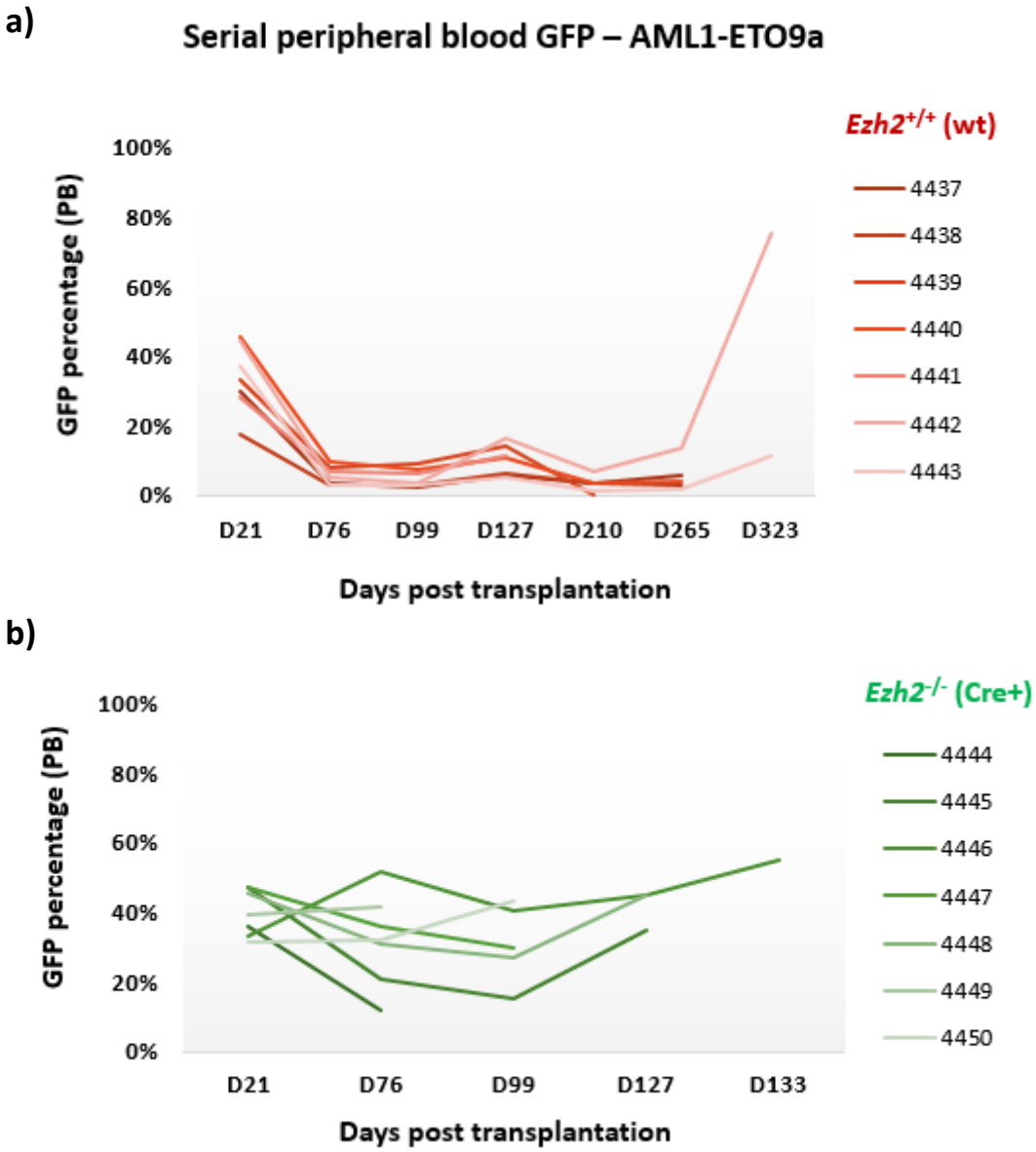


FIGURE 3.9 **Serial bleeds assessing GFP percentage in peripheral blood (PB) from (a) *Ezh2*$^{+/+}$ (wt); AML1-ETO9a and (b) *Ezh2*$^{-/-}$ (Cre+); AML1-ETO9a cohorts over time (in days)**

(a) *Ezh2*$^{+/+}$ mice (#4437 - 4443, n=7) show similar levels of GFP expression in peripheral blood at Day 21 post transplantation to *Ezh2*$^{-/-}$ mice. This GFP expression then falls and remains low level (<20%) until significantly later, increasing rapidly just before mice become terminal

(b) In contrast, *Ezh2*$^{-/-}$ (#4444 - 4450, n=7) mice display consistently higher GFP expression in peripheral blood across serial bleeds until they become terminal

Strikingly, whilst all *Ezh2*$^{+/+}$ (wt) animals had a significant fall in GFP following the initial Day 21 bleed, only retaining a low level of peripheral GFP expression until disease onset close to a year later, this was not the case with the *Ezh2*$^{-/-}$ (Cre+) animals. As shown above, all mice retained a markedly high peripheral GFP percentage until time of death which was significantly accelerated compared to the wild-type cohort. This suggests that not only is disease accelerated in the *Ezh2*$^{-/-}$ cohort (as seen below), but that its early kinetics are also altered, with all animals constantly demonstrating a higher circulating GFP expression after engraftment than in the *Ezh2*$^{+/+}$ cohort.

The *Ezh2*$^{-/-}$ (Cre+) cohort became terminal as early as Day 86 of the experiment, with clinical signs of emaciation, piloerection, hunched posturing and noticeably pale extremities. By Day 138 the entire cohort had died of leukaemia. Each animal had a pancytopenia with marked anaemia and thrombocytopenia, with higher GFP expression in bone marrow than spleen or peripheral blood. In contrast, the *Ezh2*$^{+/+}$ (wt) cohort, typically with the longer latency, began to develop GFP positive leukaemia at Day 323, at which point the experiment was terminated due to the clear significance of the result. A few of the *Ezh2*$^{+/+}$ (wt) cohort died of other causes than leukaemia (GFP negative in PB, BM and spleen) – bilateral hydronephrosis, enlarged thymi and livers on necropsy). The survival outcomes are shown below (Figure 3.10):

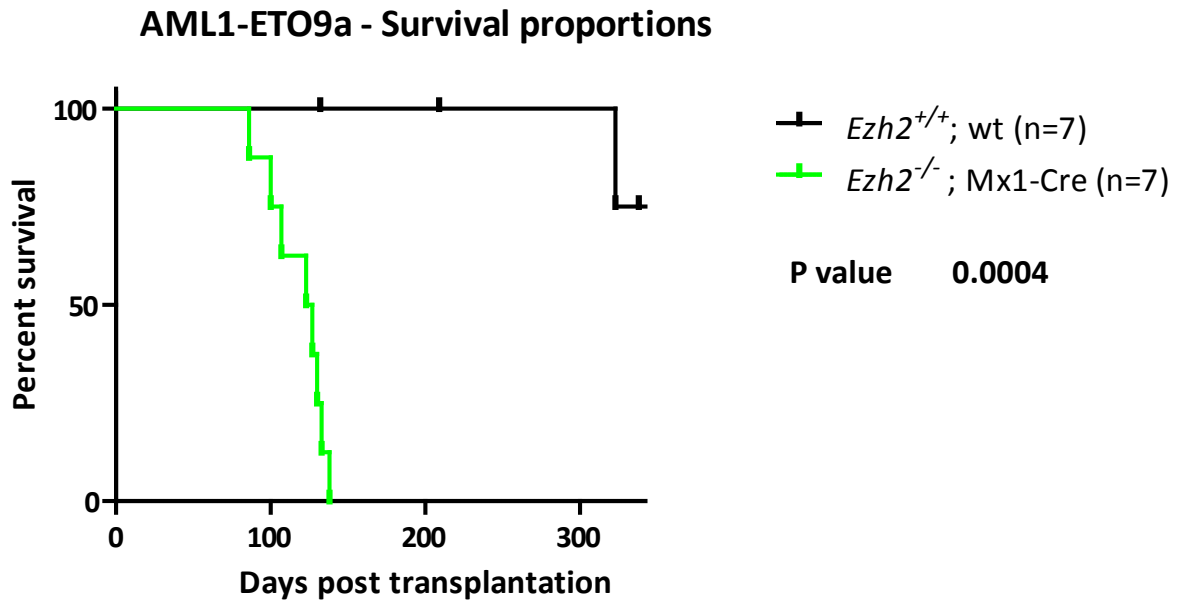


FIGURE 3.10 **Kaplan-Meier survival curve of *Ezh2*$^{+/+}$ (wt) vs *Ezh2*$^{-/-}$ (Cre+) AML1-ETO9a leukaemic mice**

Ezh2$^{-/-}$ (Cre+) AML1-ETO9a leukaemias develop significantly quicker than *Ezh2*$^{+/+}$ (wt) AML1-ETO9a leukaemias

Three notches for *Ezh2*$^{+/+}$ (wt) curve (black) indicate 3 mice culled for reasons other than leukaemia as indicated in main text, excluded from survival curve

GFP percentages in bone marrow and spleen in both cohorts were typically high, varying from 32-91% and 45-77% respectively (Table 3.2). All mice had enlarged livers and spleens, with markedly low haemoglobins at time of culling, consistent with the original description of AML1-ETO9a transplanted mice. Immunophenotyping of leukaemic splenic tissues (Dr George Giotopoulos, Huntly lab) demonstrated a solely c-kit positive population (negative for Mac1/Gr1 myeloid markers and also negative for CD4/B220 and Ter119 lymphoid/erythroid markers), again consistent with the original study. [5] Cytospins of peripheral blood and spleen prepared freshly upon culling demonstrated leukaemic blasts in Cre+ mice. Histopathology of bone marrow and spleens showed dense infiltration in both with a monomorphic population of primitive blasts (Figure 3.11).

						GFP%				
Mouse #	**Cre status**	**Day culled**	**WCC ($x10^9$/l)**	**Hb (g/dl)**	**Plts ($x10^9$/l)**	**Periph Blood**	**Bone Marrow**	**Spleen**	**Spleen wt (g)**	**Liver wt (g)**
4449	Cre+	86	5.6	36	122	29.1	85.5	56.0	0.28	1.16
4450	Cre+	100	3.4	**	40	43.2	86.0	66.0	0.32	1.23
4444	Cre+	107	2.4	**	22	36.9	43.5	45.8	0.31	1.11
4447	Cre+	123	3.1	44	15	43.0	76.9	76.7	0.40	1.10
4448	Cre+	127	3.1	30	73	46.0	76.8	77.0	0.26	1.39
4446	Cre+	130	7.4	35	37	55.0	91.0	67.5	0.39	1.33
4441	*wt*	*132*	*15.2*	*82*	*105*	*5.0*	*1.6*	*2.2*	*0.51*	*1.50*
4445	Cre+	133	4.5	28	120	34.7	61.0	61.9	0.28	1.08
4451	Cre+	138	1.2	26	31	8.9	80.9	66.4	0.29	1.29
4439	*wt*	*204*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*0.31*	*1.32*
4442	wt	323	27.4	53	273	75.6	32.1	66.5	0.59	2.88

TABLE 3.2 **Terminal data for AML1-ETO9a leukaemic mice (*Ezh2*$^{+/+}$ (wt) and *Ezh2*$^{-/-}$ (Cre+) arms)**

Rows in grey/italic indicate animals that died of non-leukaemia causes (GFP negative) – findings at necropsy included bilateral hydronephrosis, massive thymic enlargement, massive hepatomegaly

** indicate haemoglobins lower than the lower limit of detection by blood analyser

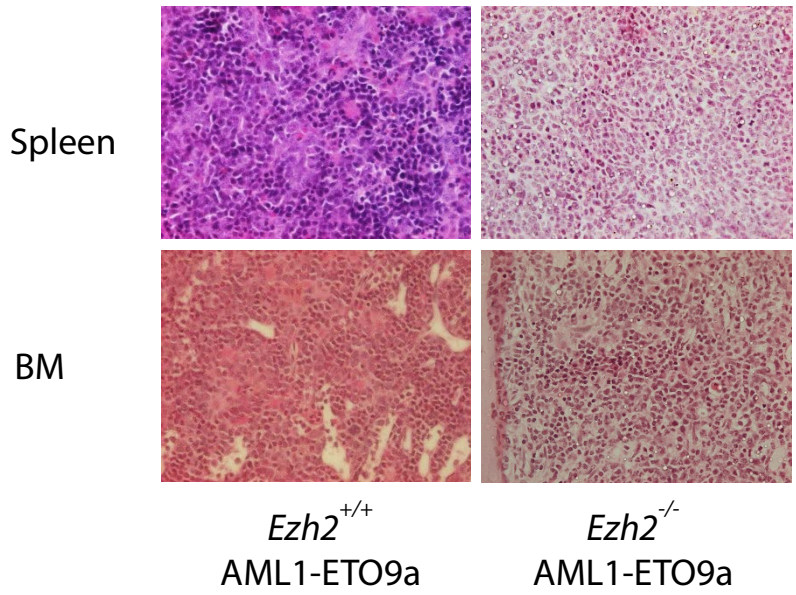


FIGURE 3.11 **Histopathology of bone marrow and spleen taken at time of death in *Ezh2*$^{+/+}$ (wt) vs *Ezh2*$^{-/-}$ (Cre+) AML1-ETO9a murine leukaemias**
Rapid Romanowsky stain. x40 magnification

Genotyping of two mice from each arm (*Ezh2fl/fl*; wt and *Ezh2 fl/fl;* Cre+ AML1-ETO9a leukaemic mice) from DNA extracted from spleen tissues, demonstrated 99-100% excision in the Cre+ group versus no excision in the wt group, confirming the leukaemias maintained the genotypes of the transplanted cells.

This experiment resulted in the highly significant and surprising finding that *Ezh2* loss prior to induction of AML1-ETO9a leukaemogenesis actually accelerated its development whilst showing no obvious phenotypic differences to the *Ezh2* replete cohort, suggesting a potent tumour suppressive function, in stark contrast to the role of *Ezh2* in transformation by MOZ-TIF2.

3.3.2.3 Loss of *Ezh2* accelerates MLL-AF9 mediated leukaemogenesis *in vivo*

The same experimental setup was then applied to assessing the role of *Ezh2* in MLL-AF9 mediated leukaemogenesis. As described above, the MLL-AF9 translocation is a particularly potent oncogenic driver of AML producing aggressive, short latency disease[9], so the expectation here was that loss of *Ezh2* prior to MLL-AF9 transformation *in vivo* was less likely to produce a significant effect on outcome, if at all any. Once again, leukaemia development was tracked by regular clinical monitoring of mice and peripheral blood sampling to assess whole blood count and GFP percentages. At Day 40, there was a significant increase in circulating GFP positive cells in the *Ezh2*$^{-/-}$ (Cre+) cohort compared to *Ezh2*$^{+/+}$ (wt) (Figure 3.12):

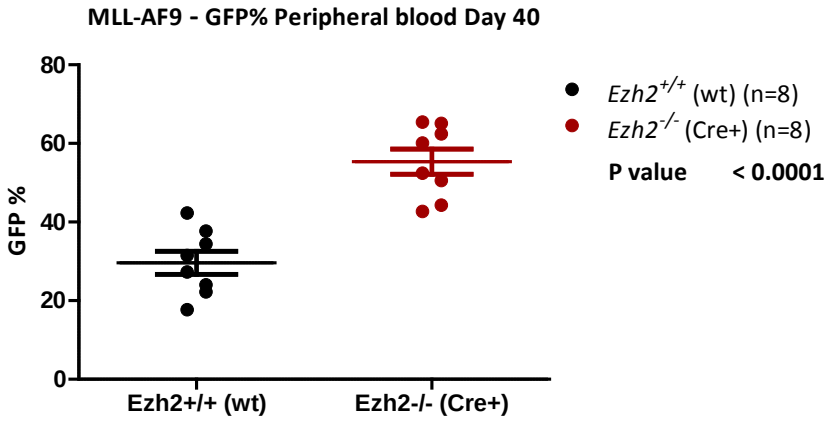


FIGURE 3.12 **GFP expression in peripheral blood between *Ezh2*$^{+/+}$ (wt) and *Ezh2*$^{-/-}$ (Cre+) MLL-AF9 mice at Day 40 post-transplant**
Ezh2$^{-/-}$ MLL-AF9 transplanted mice display a significantly higher level of GFP in peripheral blood prior to becoming terminal compared to *Ezh2*$^{+/+}$ MLL-AF9 transplanted mice

This translated into earlier development of terminal leukaemias in this cohort. As early as Day 41, animals in the *Ezh2*$^{-/-}$ (Cre+) arm became terminal with emaciation, ruffled fur, reduced movements and marked dyspnoea. All animals exhibited marked hepatosplenomegaly, elevated white cell count, anaemia and strongly expressed GFP in peripheral blood, spleen and bone marrow tissues. There was a significant acceleration of development of MLL-AF9 driven AML in the *Ezh2*$^{-/-}$ (Cre+) arm as seen in Figure 3.13:

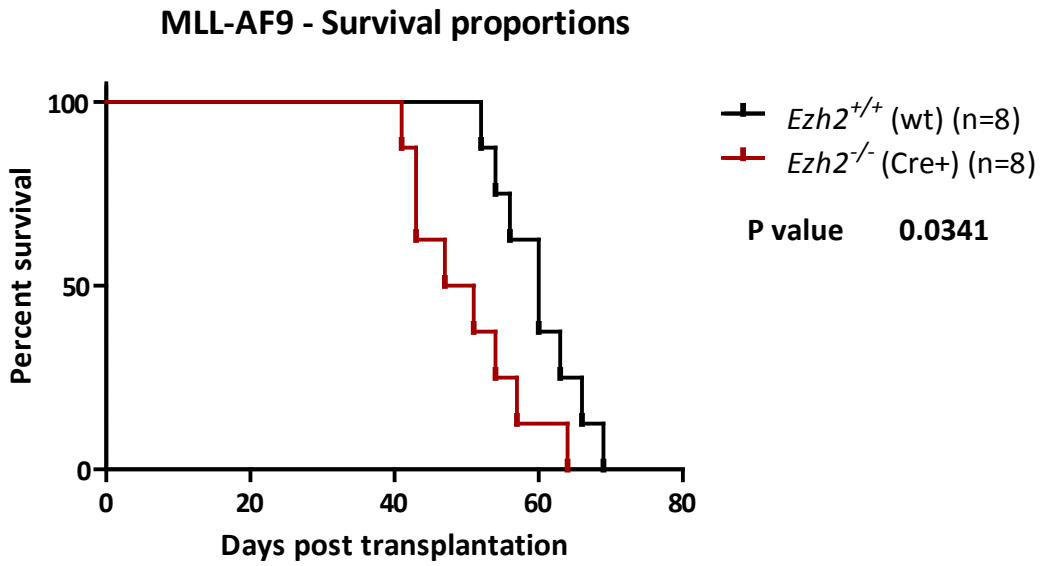


FIGURE 3.13 **Kaplan-Meier survival curve of *Ezh2*$^{+/+}$ (wt) vs *Ezh2*$^{-/-}$ (Cre+) MLL-AF9 leukaemic mice**

Ezh2$^{-/-}$ (Cre+) MLL-AF9 leukaemias develop significantly quicker than *Ezh2*$^{+/+}$ (wt) MLL-AF9 leukaemias

Mice in the *Ezh2*$^{+/+}$ (wt) arm survived significantly longer (median survival 60 days compared to 49 days in the *Ezh2*$^{-/-}$ (Cre+) group). GFP percentages in bone marrow and spleen in both cohorts were typically high, varying from 80-96% and 49-84% respectively (Table 3.3). Immunophenotyping of leukaemic splenic tissues was not different between genotypes (Dr George Giotopoulos, Huntly lab) confirming a significant myeloid population, strongly positive for Mac1 and c-kit (whilst negative for CD4/B220 and Ter119 lymphoid/erythroid markers).

Mouse #	Cre status	Day culled	WCC ($x10^9$/l)	Hb (g/dl)	Plts ($x10^9$/l)	GFP% Periph Blood	GFP% Bone Marrow	GFP% Spleen	Spleen wt (g)	Liver wt (g)
5329	Cre+	41	32.7	136	113	45.90%	84.40%	57.00%	0.531	2.287
5335	Cre+	43	67.2	70	280	50.00%	80.30%	49.20%	0.8	2.88
5336	Cre+	43	36.1	96	496	52.80%	80.20%	63.60%	0.62	2.72
5331	Cre+	47	108.4	65	110	67.90%	87.00%	71.60%	0.61	2.71
5333	Cre+	51	20.4	40	59				0.7	3
5326	wt	52								
5334	Cre+	54	142.4	69	379	78.60%	93.10%	69.60%	1.15	4.18
5342	wt	54	22.6	46	69	50.30%	90.70%	71.50%	0.89	2.74
5328	wt	56	51.9	78	157	60.40%	94.00%	60.50%	0.57	2.66
5332	Cre+	57	26.1	44	310	75.70%	93.60%	75.80%	1.01	3.89
5340	wt	60	54.7	133	209	75.50%	96.20%	65.50%	0.75	2.86
5338	wt	60	>80	160	449	72.50%	83.60%	77.80%	0.76	2.28
5327	wt	63	126	89	88	65.60%	84.20%	57.50%	0.96	4.05
5330	Cre+	64	19.2	109	332	57.50%	93.50%	84.70%	0.89	2.89
5339	wt	66							0.71	3.09
5341	wt	69	107.1	146	212	56.20%	93.10%	59.50%	0.75	2.54

TABLE 3.3 **Terminal data for all MLL-AF9 leukaemic mice (*Ezh2*$^{+/+}$ (wt) and *Ezh2*$^{-/-}$ (Cre+) arms)**
Blank cells – data not available

There were no significantly different phenotypic effects between the two cohorts. Average spleen and liver sizes and blood counts were similar as were distribution of leukaemic cells (assessed by GFP proportions) (Figures 3.14 and 3.15).

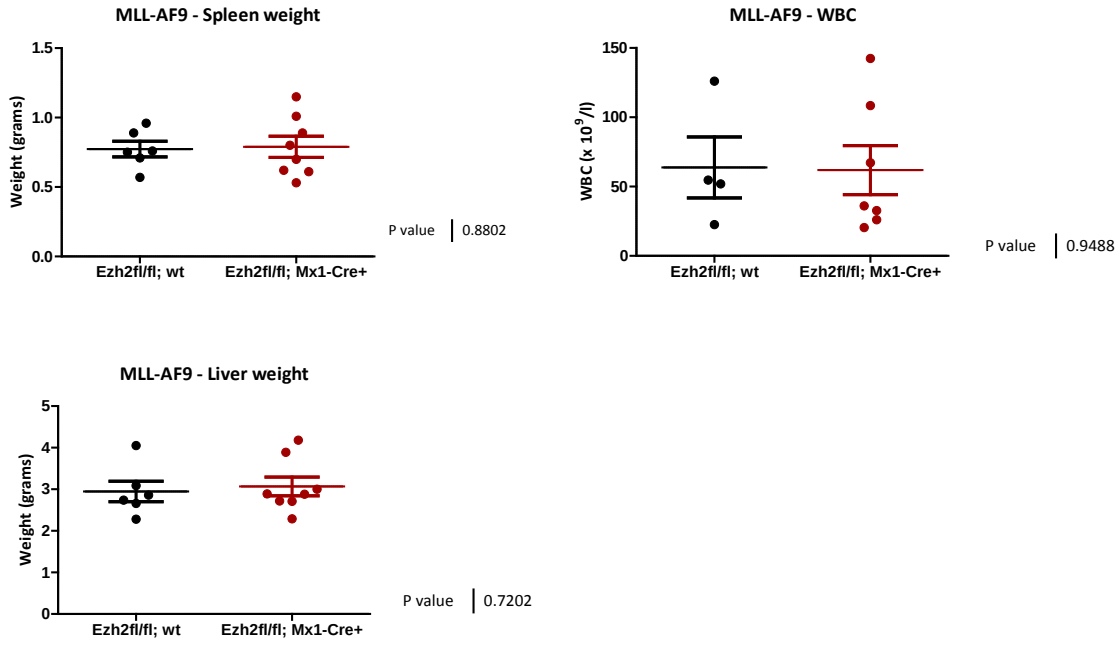


FIGURE 3.14 **Spleen/Liver weights and terminal white cell counts of *Ezh2*$^{+/+}$ vs *Ezh2*$^{-/-}$ MLL-AF9 leukaemic mice**

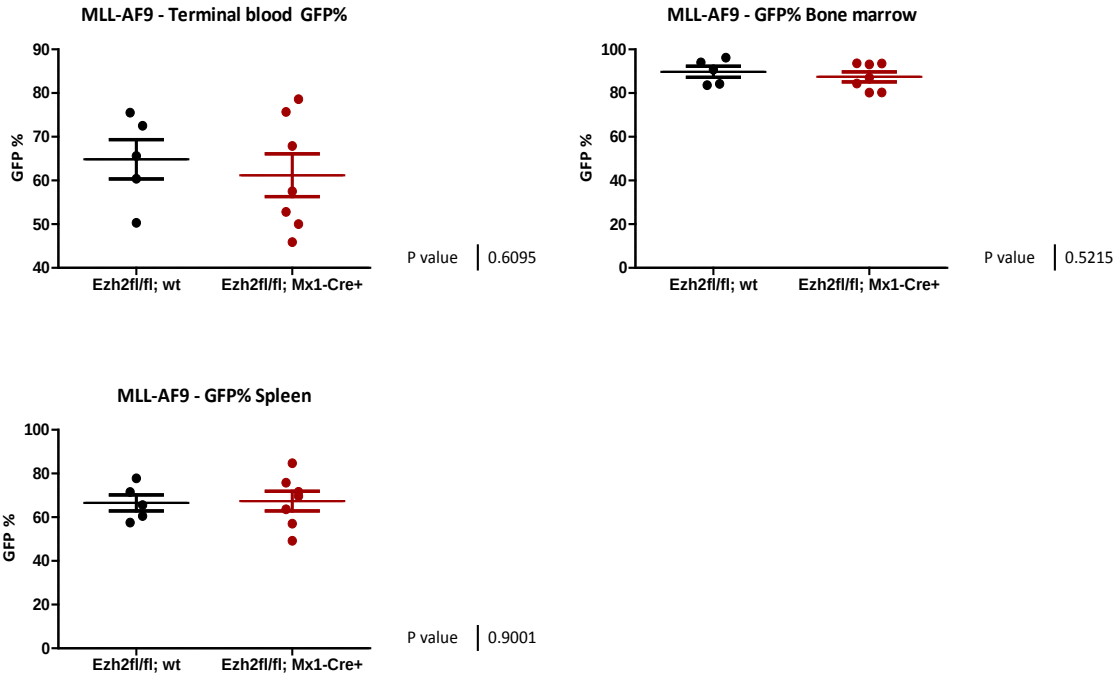


FIGURE 3.15 **GFP expression in terminal blood samples, bone marrow and spleen of *Ezh2*$^{+/+}$ vs *Ezh2*$^{-/-}$ MLL-AF9 leukaemic mice**

3.3.3 Discussion and future work

Though *Ezh2* loss *in vitro* failed to demonstrate any marked differences in immortalisation for each of the three oncogenes described above, its loss *in vivo* prior to transformation and leukaemia development resulted in profound phenotypic outcomes starkly contrasting each other, depending upon AML subtype. *In vivo*, in MOZ-TIF2 leukaemogenesis, *Ezh2* appeared to have clear oncogenic function as its absence permitted a significant increase in survival. Surprisingly, the same experiment for AML1-ETO9a and MLL-AF9 demonstrated a clear synergism between *Ezh2* loss and these oncogenes in driving AML generation, demonstrate a clear tumour-suppressive role. This was corroborated by the genotypes of *Ezh2fl/fl*; Cre+ leukaemic mice in both fusions being 99-100% *Ezh2* deficient via qPCR assessment at time of culling, suggesting that the $Ezh2^{-/-}$ transformed cells have a clear selective advantage. It is interesting that *Ezh2* loss has the same acceleratory effect upon two very different subtypes of AML. MLL-AF9 is characteristically known to promote transformation through gene activation whilst AML-ETO (and AML1-ETO9a) progress through the repression of gene transcription. That loss of Ezh2, a transcriptional repressor itself, should facilitate the generation of both these AML subtypes that alone progress by contrasting mechanisms, is highly unusual. Furthermore, similar to MLL-AF9, the MOZ-TIF2 oncogene transforms through activation of genes, yet *Ezh2* loss here led to a clearly opposite effect, with a pronounced increase in survival. The expectation would have been that any effect of *Ezh2* loss would be similar between MLL-AF9 and MOZ-TIF2 which transform in similar ways, are more aggressive and are clinically associated with poorer prognosis, whilst being different in the less aggressive AML1-ETO9a setting. However, this is not the case and as the *in vivo* effects are clearly significant across all three subtypes, this is a surprising result that reinforces the idea that *Ezh2* has varying roles during induction of leukaemia, in a subtype-dependent context.

Though different diseases, the findings in these MLL-AF9 and AML1-ETO9a *in vivo* experiments align with recently published data supporting the tumour suppressive role of *Ezh2* in myeloproliferative neoplasms,[10,11] which conclude by suggesting caution when considering EZH2 inhibitors in clinical programmes.

As this experiment assessed *Ezh2* loss during the early stages of AML evolution, in view of the *in vitro* experiments suggesting an early clonogenic and proliferative disadvantage of $Ezh2^{-/-}$ cells, an assessment of functional and phenotypic differences between wild-type $Ezh2^{+/+}$ and $Ezh2^{-/-}$ bone marrow was made. Functional assays *in vitro* demonstrated significantly diminished replating potential of *Ezh2* deficient cells and this difference was confirmed across 5 biological replicates. Furthermore, there was no significant alteration of early progenitor or HSC compartments following *Ezh2* loss that

could account for the stark differences in leukaemia outcomes seen following transformation and *in vivo* transplantation.

Given the major function of Ezh2 is to mediate histone methyltransferase activity through its catalytic SET domain, thus facilitating transcriptional repression, the next logical step was to assess differences in gene expression following *Ezh2* loss in the wild-type/non-transformed setting. In doing so, it would be possible to establish which transcriptional programmes are altered in these c-kit positive HSPCs when *Ezh2* is deleted, with the suggestion that these would either hinder or co-operate in a subtype-specific manner with the individual driver oncogene in this system once it was overexpressed. To further interrogate how *Ezh2* loss might mechanistically alter transcriptional programmes and their outputs, it was also important to assess changes in histone modifications – particularly trimethylation of lysine 27 at Histone H3 (H3K27me3), which it is responsible for. We also wanted to align this with an assessment of acetylation of lysine 27 of histone H3 (H3K27Ac) and trimethylation of lysine 4 at histone H3 (H3K4me3) – both associated with activation of transcription, to search for any overlaps between this and a loss of a repressive mark and increased gene expression. Through interlinking altered gene expression and analysis of potential alterations of the epigenetic landscape following *Ezh2* deletion in non-transformed HSPCs, we wished to identify potential candidate gene facilitators of leukaemia development in the AML1-ETO9a or MLL-AF9 setting. Given their opposing mechanisms of transformation, yet the same outcome following *Ezh2* loss, we expected to find differences in transcriptional changes between these leukaemias. By comparison of gene changes following *Ezh2* loss in the non-transformed setting with altered gene expression in these *Ezh2*-deficient leukaemias, we aimed to reveal key functional regulators that were possibly co-operating with these two oncogenes in producing the profound acceleration in disease. This is detailed further in Chapter 5.

Additionally, we wished to compare and contrast this with any functional differences seen in the maintenance experiments detailed in the next section.

References

1) Tanaka S, Miyagi S, Sashida G, Chiba T, Yuan J, Mochizuki-Kashio M, Suzuki Y, Sugano S, Nakaseko C, Yokote K, Koseki H, Iwama A. *Ezh2* augments leukemogenicity by reinforcing differentiation blockage in acute myeloid leukemia. *Blood*. 2012 ;120(5):1107-17

2) Neff T. *et al.* Polycomb repressive complex 2 is required for MLL-AF9 leukemia. *Proc Natl Acad Sci USA*. 2012;109, 5028–33.

3) Giotopoulos G, Chan WI, Horton SJ, Ruau D, Gallipoli P, Fowler A, Crawley C, Papaemmanuil E, Campbell PJ, Göttgens B, Van Deursen JM, Cole PA, Huntly BJ. The epigenetic regulators CBP and p300 facilitate leukemogenesis and represent therapeutic targets in acute myeloid leukemia. *Oncogene*. 2016;35(3):279-89.

4) Corral J, Lavenir I, Impey H & Warren A. An MLL–AF9 fusion gene made by homologous recombination causes acute leukemia in chimeric mice: a method to create fusion oncogenes. *Cell*. 1996; 85, 853–861.

5) Yan M, Kanbe E, Peterson LF, Boyapati A, Miao Y, Wang Y, Chen IM, Chen Z, Rowley JD, Willman CL, Zhang DE. A previously unidentified alternatively spliced isoform of t(8;21) transcript promotes leukemogenesis. *Nat Med*. 2006 Aug;12(8):945-9.

6) de Guzman CG, Warren AJ, Zhang Z, Gartland L, Erickson P, Drabkin H, Hiebert SW, Klug CA.Hematopoietic stem cell expansion and distinct myeloid developmental abnormalities in a murine model of the AML1-ETO translocation. *Mol Cell Biol*. 2002 Aug;22(15):5506-17.

7) Rasmussen KD, Jia G, Johansen JV, Pedersen MT, Rapin N, Bagger FO, Porse BT, Bernard OA, Christensen J, Helin K. Loss of TET2 in hematopoietic cells leads to DNA hypermethylation of active enhancers and induction of leukemogenesis. *Genes Dev*. 2015 May 1;29(9):910-22.

8) Deguchi K, Ayton PM, Carapeti M, Kutok JL, Snyder CS, Williams IR, Cross NC, Glass CK, Cleary ML, Gilliland DG. MOZ-TIF2-induced acute myeloid leukemia requires the MOZ nucleosome binding motif and TIF2-mediated recruitment of CBP. *Cancer Cell*. 2003 Mar;3(3):259-71.

9) Dawson MA, Prinjha RK, Dittmann A, Giotopoulos G, Bantscheff M, Chan WI, Robson SC, Chung CW, Hopf C, Savitski MM, Huthmacher C, Gudgin E, Lugo D, Beinke S, Chapman TD, Roberts EJ, Soden PE, Auger KR, Mirguet O, Doehner K, Delwel R, Burnett AK, Jeffrey P, Drewes G, Lee K, Huntly BJ, Kouzarides T. Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. *Nature*. 2011 Oct 2;478(7370):529-33.

10) Shimizu T, Kubovcakova L, Nienhold R, Zmajkovic J, Meyer SC, Hao-Shen H, Geier F, Dirnhofer S, Guglielmelli P, Vannucchi AM, Feenstra JD, Kralovics R, Orkin SH, Skoda RC. Loss of *Ezh2* synergizes with *JAK2*-V617F in initiating myeloproliferative neoplasms and promoting myelofibrosis. *J Exp Med*. 2016 Jul 25;213(8):1479-96.

11) Yang Y, Akada H, Nath D, Hutchison RE, Mohi G. Loss of *Ezh2* cooperates with J*ak2*V617F in the development of myelofibrosis in a mouse model of myeloproliferative neoplasm. *Blood*. 2016 Jun 30;127(26):3410-23.

Chapter 4

Characterising the role of Ezh2 in the maintenance of Acute Myeloid Leukaemia

4.1 Aims

This series of experiments was designed to examine the genetic role of *Ezh2* in established leukaemia, and whether its loss, either *in vitro* or *in vivo,* would alter immortalisation by AML oncogenes and the maintenance of these aggressive tumours.

4.2 Introduction

Retroviral transduction followed by overexpression of MLL-AF9, AML1-ETO9a and MOZ-TIF2 within murine HSPCs consistently produces immortalisation of these cells *in vitro* under the appropriate murine cytokine conditions. Transplantation of these cells into lethally irradiated wild-type C57/Bl6 recipient mice through tail-vein injection and then engraftment, eventually generates leukaemias with disease latencies reflective of potency of each oncogene. *In vitro* the immortalised cells continue to form colonies and proliferate indefinitely, whilst *in vivo* once leukaemic, mice will rapidly deteriorate clinically and show terminal signs such as weight loss, inappetence, ruffled fur and dyspnoea, with the leukaemias able to perpetuate disease through serial transplantation. Through retroviral overexpression of these oncogenes in c-kit positive HSPC's harvested from *Ezh2 fl/fl*; wild-type, it was possible to conditionally delete Ezh2 at any stage of AML evolution. After immortalisation or engraftment had occurred, the role of Ezh2 in maintaining this state was then interrogated.

To test the requirement of Ezh2 in maintenance of the immortalised state *in vitro*, *Ezh2 fl/fl* (**wild-type** for *Mx1-Cre*) c-kit selected HSPCs from mice were immortalised/transformed by retroviral transduction of the same three oncogenes. Once established in weekly methylcellulose replatings (i.e. clear colony formation with 100% GFP/YFP expression), these cell lines were transferred to liquid culture and retrovirally transduced a second time with a p-Babe-Cre-puro vector or p-Babe-puro empty vector (both containing a puromycin selection marker), allowing for excision of Ezh2 at a stage where immortalisation was already established. The readout from this assay was through methylcellulose replatings, again assessing clonogenic and proliferative potential post Ezh2-excision (Figure 2.11 in Methods).

For the maintenance *in vivo* assays, primary leukaemias were generated for the same three oncogenes from *Ezh2 fl/fl* mice (**heterozygous** for *Mx1-Cre*) and then transplanted into lethally irradiated wild-type C57/Bl6 secondary recipients. Once the cells had engrafted, PBS (control) or pIpC was then administered intraperitoneally to activate *Mx1-Cre* and delete *Ezh2*, thus testing its requirement for the maintenance of the leukaemia (Figure 2.13 in Methods).

As mentioned in Chapter 3, *Ezh2* deletion during the maintenance stage of MLL-AF9 transformed granulocyte macrophage progenitors (GMPs), severely compromised growth *in vitro* and the progression of AML *in vivo*. [1] This was performed utilising a similar *Ezh2*$^{flox/flox}$ knockout system, though employing *α Rosa: Cre-ERT*. Having seen the opposite effect on MLL-AF9 AML induction *in vivo*, we were keen to demonstrate that we could also see a delayed progression of AML *in vivo* using our conditional knockout system, therefore we carried out the maintenance experiments using MLL-AF9 (to serve as a positive control) as well as in AML1-ETO9a and MOZ-TIF2 leukaemias. Having seen varying effects dependent on subtype in AML induction, it was important to ascertain if *Ezh2* played any oncogenic or tumour-suppressive role in maintenance, especially of the latter 2 subtypes.

4.3 Results and discussion

4.3.1 *Ezh2* is an absolute requirement for the maintenance of multiple subtypes of AML *in vitro*

As described in Methods 2.6.12, methylcellulose platings with or without puromycin selection following transduction of each transformed *Ezh2 fl/fl* cell line with either empty vector or p-babe-Cre-puro were set up, with the intention of noting differences in colony formation, total cell numbers over successive weekly replatings, and subsequently assessing any differences in cellular phenotype. Colonies growing in the +puromycin plates were those that had successfully been transduced with either vector. Use of the empty vector (hereafter “EV transduced”) in tandem with p-babe-Cre-puro (hereafter “Cre-transduced”) was to provide a control and assess for Cre-mediated toxicity.

For MLL-AF9, *Ezh2* deletion following successful Cre-transduction resulted in no colonies and virtually no cells available to replate after 7 days in methylcellulose (Figure 4.1, right):

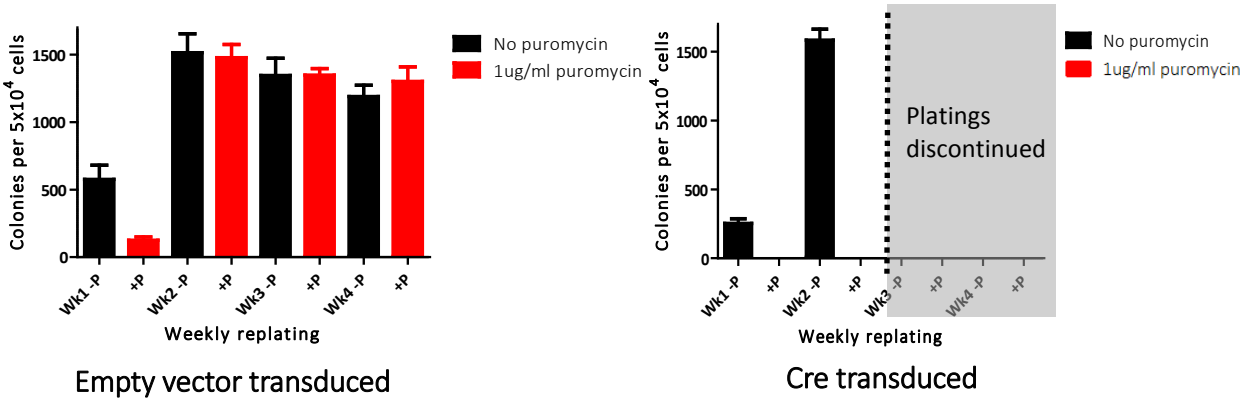


FIGURE 4.1 **Methylcellulose replatings for empty vector-transduced versus Cre-transduced *Ezh2 fl/fl*; MLL-AF9 transformed cell lines**

Ezh2 is absolutely required for maintenance of MLL-AF9 immortalised HSPCs *in vitro*
+P = with puromycin 1μg/ml, -P = without puromycin
Wk 1, 2, 3 and 4 indicate successive weekly replatings (after 7 days)

In contrast, successful empty-vector transduced *Ezh2 fl/fl*; MLL-AF9 cell lines (as evidenced by puromycin selection - indicated with red bars) produced less colonies than non-transduced at Week 1, indicating modest toxicity of the additional transduction process. However by Week 2, the clonogenic potential of both successfully and unsuccessfully transduced empty vector cells was restored to similar high levels. This was maintained through 4 weeks of replatings (Figure 4.1, left). Taken together, it therefore appeared that retroviral insertion and subsequent expression of the Cre- recombinase completely abolished growth of the already immortalised *Ezh2fl/fl*; MLL-AF9 cell line.

The effects of Cre recombinase mediated toxicity to mammalian cells have been widely reported in the literature.[2,3] In order to ensure that the profound inhibitory effect on colony growth seen on expressing Cre- from p-babe-Cre-puro transduction in the *Ezh2 fl/fl*; MLL-AF9 cell line was due to loss of *Ezh2* and not Cre-toxicity, the same experiment was performed on a separately generated wild-type; MLL-AF9 immortalised cell line (this was generated using wild-type murine c-kit positive HSPCs and subjecting them to MLL-AF9 transduction and immortalisation through successive methylcellulose replatings – performed at the same time as developing the *Ezh2fl/fl*; MLL-AF9 cell line).

If the effects were solely due to Cre-mediated toxicity, it was expected that in this wild-type; MLL-AF9 immortalised cell line where no sites were floxed, the same abolishment of colony growth at one week would be seen. However, despite being transduced with p-babe-Cre-puro (confirmed through its growth in puromycin), the wild-type; MLL-AF9 cell line was able to produce colonies at Week 1 and beyond (Figure 4.2). This established that the effects seen were due to *Ezh2* loss, suggesting its importance in maintenance of MLL-AF9 immortalised cells *in vitro*.

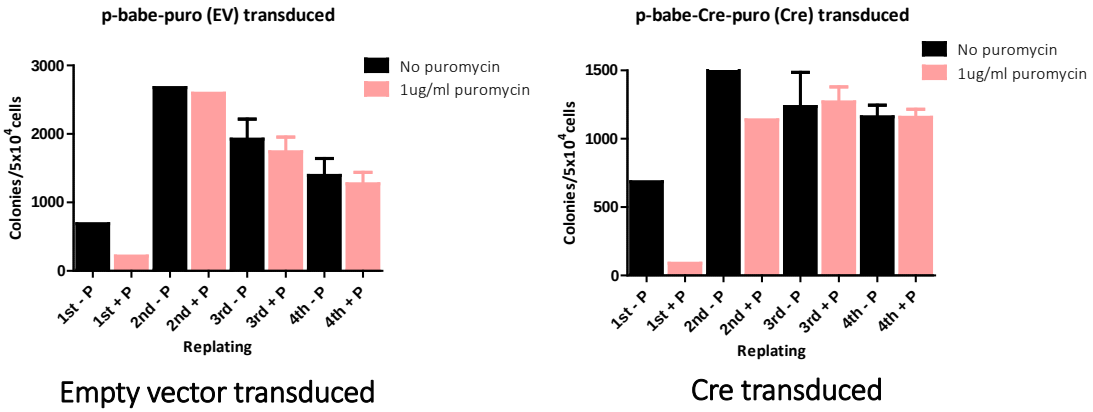


FIGURE 4.2 **Methylcellulose replatings for empty vector-transduced versus Cre-transduced wild-type; MLL-AF9 transformed cell lines**

No significant Cre-mediated toxicity is seen

+P = with puromycin 1μg/ml, -P = without puromycin

1st, 2nd, 3rd, 4th indicate successive weekly replatings (after 7 days)

When the same experiment was performed for the *Ezh2fl/fl*; AML1-ETO9a transformed cell line, it showed the same outcome – inhibited growth in the Cre-transduced +puromycin arm (compared to normal growth in both +/-puro arms for the EV-transduced cells). A single colony was present at Week 1 but when cells from this were harvested, washed and replated they did not re-establish clonogenic growth at Week2 onwards (Figure 4.3):

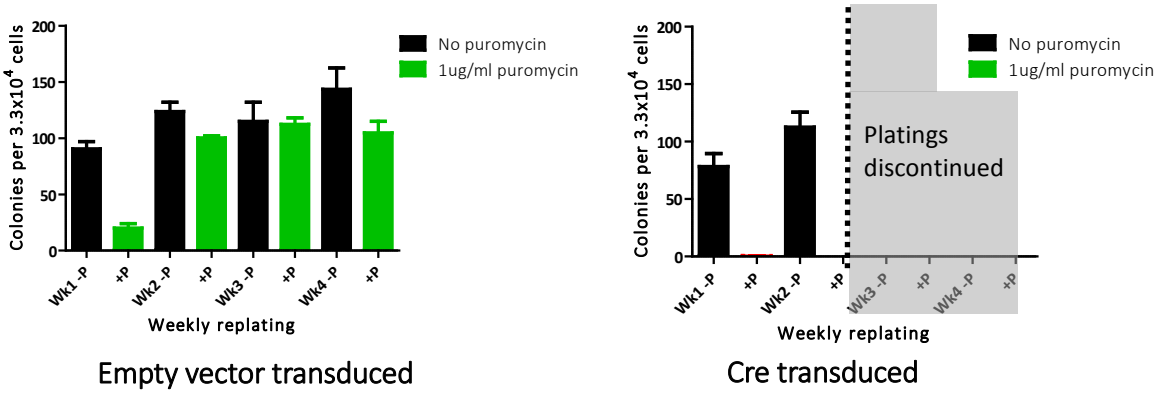


FIGURE 4.3 **Methylcellulose replatings for empty vector-transduced versus Cre-transduced *Ezh2 fl/fl*; AML1-ETO9a transformed cell lines**

Ezh2 is absolutely required for maintenance of AML1-ETO9a immortalised HSPCs *in vitro*

+P = with puromycin 1μg/ml, -P = without puromycin

Wk 1, 2, 3 and 4 indicate successive weekly replatings (after 7 days)

Similarly, an equivalent wild-type; AML1-ETO9a immortalised cell line was unaffected by p-babe-Cre-puro transduction (similar to the MLL-AF9 experiment above) and established clonogenic growth under puromycin selection in a similar fashion to Figure 4.2 (data not shown).

Finally, the same experimental outcome for the *Ezh2fl/fl*; MOZ-TIF2 transformed cell line was seen. Ezh2 loss following Cre-transduction (selected again through growth in puromycin) definitively abolished growth, whilst transduction with the empty vector did not significantly affect the cell line at all (Figure 4.4):

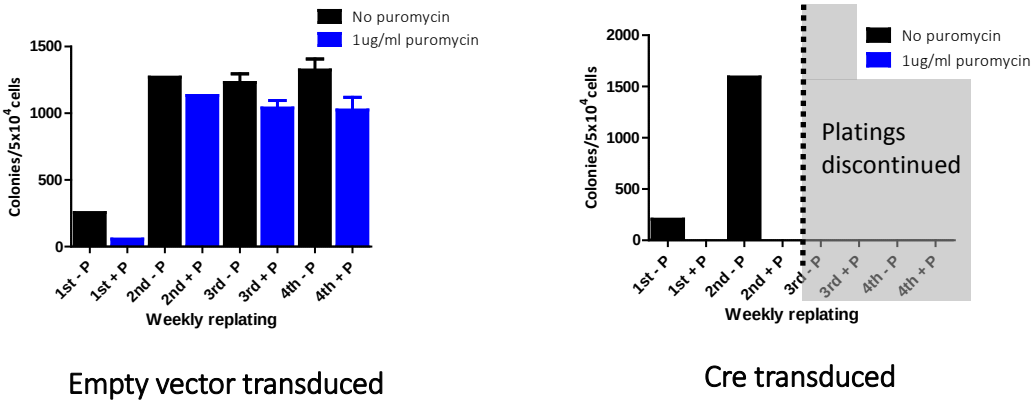


FIGURE 4.4 **Methylcellulose replatings for empty vector-transduced versus Cre-transduced *Ezh2 fl/fl*; MOZ-TIF2 transformed cell line**

Ezh2 is absolutely required for maintenance of MOZ-TIF2 immortalised HSPCs *in vitro*
+P = with puromycin 1μg/ml, -P = without puromycin
1st, 2nd, 3rd, 4th indicate successive weekly replatings (after 7 days)

The substantial effects of Ezh2 deletion in these established immortalised cell lines were so significant that due to a complete absence of cells after overexpression of Cre recombinase through a second retroviral transduction, a number of factors such as:

a) percentage of *Ezh2* excised (as per qPCR strategy)
b) functional differences e.g. changes in cell cycle progression, apoptosis and differentiation

could not be assessed. These *in vitro* data fit with the original hypothesis provided in this thesis and provide a compelling argument towards therapeutically targeting EZH2. However, it was important to assess whether this functional role was also maintained *in vivo* – this is explored in the next section.

4.3.2 Loss of *Ezh2* interferes with maintenance of secondary MLL-AF9 leukaemias *in vivo*

Primary transplants

CD117 (or c-kit) magnetic selection was performed on red-cell lysed, washed whole bone marrow taken from 2-3 *Ezh2 fl/fl*; Cre+ mice (with no prior pIpC treatment) to yield HSPCs. A typical harvest was approximately 40-50x10^6 whole bone marrow cells per mouse (aged 8-12 weeks). Magnetic labelling and selection typically yielded in the order of 10% c-kit positive cells from whole marrow (data not shown). These c-kit positive cells were then stabilised in overnight liquid culture in RPMI 20% and cytokines (as per Methods 2.6.5), and were then subjected to two rounds of retroviral transduction with retroviral supernatant derived from transfection of 293T cells with MSCV-IRES-MLL-AF9-YFP construct. Following the second round of transduction, an assessment of transduction efficiency was made using flow cytometry to assess the percentage of YFP positive (i.e. transduced) cells. For MLL-AF9 (in this and across all the other experiments detailed above) the typical transduction efficiency was 0.5-4%, an example of this is shown in Figure 4.5:

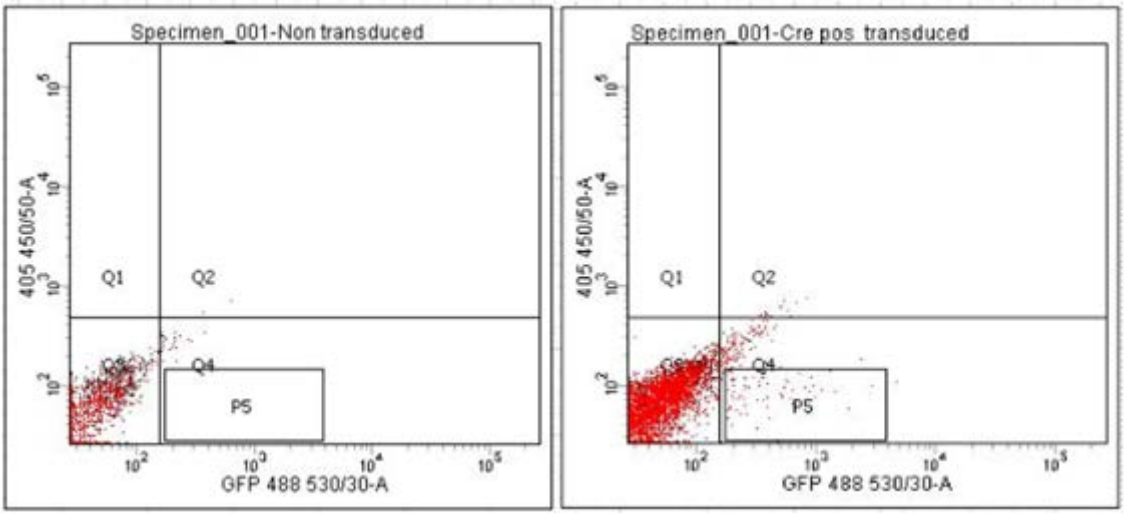


FIGURE 4.5 **Transduction efficiency of MLL-AF9 into c-kit positive HSPCs**
Cells gated initially by forward and side-scatter, doublets removed and then gated for YFP positivity
P5 indicates cells gated for YFP expression (detected on same channel as GFP, i.e. 488 530/30)

The relatively low levels of transduction efficiency were expected, given the relatively large size of the construct. Despite this, due to the potency of the oncogene, small numbers of cells YFP positive cells only were required to produce leukaemia (in the order of 10^3-10^4 YFP positive cells transplanted per mouse).

Using the percentage transduced and the total number of cells isolated, it was possible to adjust the volume injected per mouse to ensure each animal received exactly the same number of YFP-positive cells (this type of calculation was also used in all of the induction experiments for transplantation, in order to ensure each arm could be compared directly). Five wild-type C57/Bl6 mice were lethally irradiated (5.5Gy in 2 fractions) then injected with these transduced c-kit positive cells and monitored for disease, through tracking YFP percentages in peripheral blood samples across time and through clinical monitoring. By Day 40 and 41, all of these primary transplants were exhibiting clear terminal signs, with very high white cell counts and YFP-positivity (example in Figure 4.6):

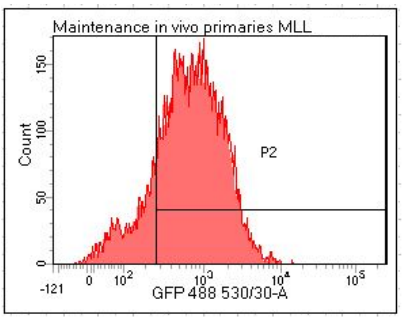


FIGURE 4.6 **FACS analysis to estimate percentage of YFP-positive cells in peripheral blood of *Ezh2 fl/fl*; Cre+ MLL-AF9 transplanted primary recipient mouse, Day 40**
P2 indicates cells positive for YFP (detectable on GFP channel 488 530/30) after red cell lysis of whole blood sample. This particular sample exhibited approximately 82% YFP positivity, with WCC of $76x10^9$/L

All animals had high white cell counts, hepatosplenomegaly and strongly YFP positive tissues at the time of sacrifice. Leukaemic splenic tissues and bone marrow were isolated, processed and stored in liquid nitrogen. In generating these transplants, the idea was to have conditional *Ezh2 fl/fl*; Cre+ MLL-AF9 leukaemic tissues that we could then transplant into lethally irradiated C57/Bl6 secondary recipients. Following engraftment there, PBS or pIpC would be administered, causing *Ezh2* deletion in the latter to occur and thus to test our hypothesis of whether *Ezh2* was important in MLL-AF9 maintenance.

Given the presence of Cre recombinase in the cells originally used to generate these primary leukaemias, even though it was not activated at that stage by design, we assessed whether any spontaneous Cre-mediated recombination of *Ezh2* had taken place by the time the animals were terminal for MLL-AF9 leukaemia, using DNA extracted from splenic tissues and the qPCR method already described. Surprisingly, all of the 5 primary MLL-AF9 leukaemias had some degree of *Ezh2*

excision (even though there was no prior pIpC treatment). This varied from 80-90% *Ezh2* excision. We opted to choose splenic tissue from the least *Ezh2* excised of these animals to act as donor material for secondary transplants. Though the level of spontaneous Cre-mediated excision was higher than we expected, we used the spleen from one mouse (with the lowest level of Ezh2 excision - approximately 80%) to generate secondary transplants and test our hypothesis. With hindsight, as the activation of *Mx1-Cre* involves an interferon mediated immune/inflammatory response, it is reasonable to expect there to be some degree of activation during the significant events of leukaemogenesis, and this has been recently demonstrated in the transplant setting.[6] This experimental design (i.e. generating primary leukaemias that are conditional for the gene under investigation followed by transplanting into secondary recipients and then conditionally deleting the gene to assess its effects) has been used in the literature before, though perhaps the exact level of excision is underestimated – as assessments of gene excision/genotype are usually performed via the gel method, which does not yield as precise results as the qPCR method.

Secondary transplants and *Ezh2* functional assessment

Splenic tissues from a primary *Ezh2 fl/fl*; Cre+ MLL-AF9 leukaemic mouse with the lowest level of excision were thawed, washed and counted. This particular animal was 76% YFP positive in splenic tissue at the time of culling. From this, approximately $1x10^6$ cells were transplanted via tail-vein injection into 24 wild-type C57/Bl6 sub-lethally irradiated recipient mice (12 per arm) in a secondary transplantation assay. Both arms were allowed 7 days for engraftment (7 days was chosen based on previous experiments in the Huntly lab, and also given the highly aggressive nature of secondary MLL-AF9 leukaemias it was better to choose an earlier time-point), then one cohort was treated with intraperitoneal PBS injections (control) and the other pIpC injections (test arm). Five doses were given on alternate days. Through pIpC injection, the *Mx1-Cre* was activated through an inflammatory response, facilitating *Ezh2* deletion. Mice in both arms were then tracked for clinical signs of deterioration, with the expectation that the latency (at least for the wild-type) arm would be much shorter than the primary transplants.

As demonstrated in Figure 4.7, loss of *Ezh2* in the pIpC arm delayed onset of leukaemia significantly:

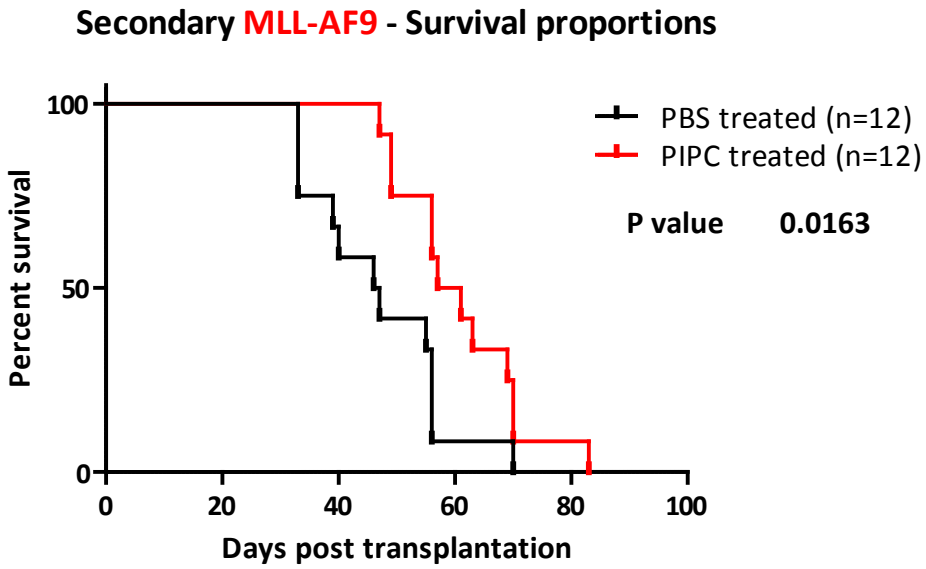


FIGURE 4.7 **Kaplan-Meier survival outcomes for *Ezh2fl/fl*; Cre+ MLL-AF9 secondary recipient mice treated with PBS or pIpC**

Mice in the pIpC treated (i.e. *Ezh2* excised) arm demonstrate a significant increase in survival compared to PBS treated

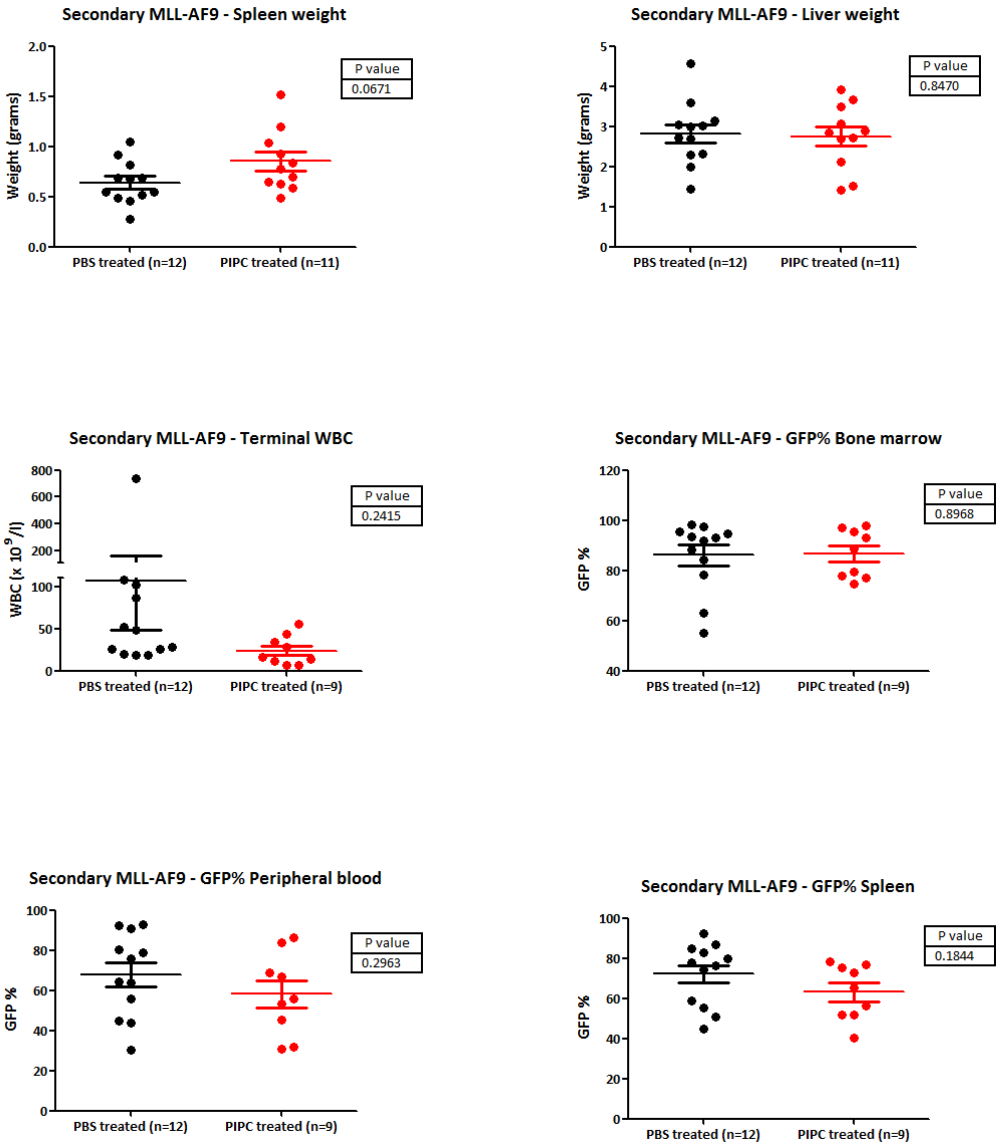


FIGURE 4.8 **Spleen and liver weights, terminal white blood cell counts and YFP percentages for blood, bone marrow and spleen for secondary MLL-AF9 recipients treated with PBS or pIpC.**
There were no significant differences seen between both treatment arms
YFP detectable on same laser/channel as GFP

There were no significant differences seen between the PBS and pIpC treated arms in terms of YFP distribution, terminal white cell counts or organ size (Figure 4.8). The leukaemias were positive for myeloid antigens Mac-1 and Gr-1 and negative for CD4/B220/Ter-119 (lymphoid and erythroid antigens). Each animal was genotyped using DNA extracted from splenic tissues. Details of all secondary leukaemias (where available) are in Table 4.1:

Mouse #	PBS/PIPC treated	Day culled	WCC (x10^9/L)	Hb (g/dl)	Plts (x10^9/L)	GFP% Periph Blood	GFP% Bone Marrow	GFP% Spleen	Spleen wt (g)	Liver wt (g)	*Ezh2* excision
4851	PBS	33	19.1	158	508	44.0	55.0	45.0	0.46	2.32	69.36%
4855	PBS	33	19.6	157	412	44.6	63.0	51.0	0.28	1.44	42.20%
4841	PBS	39	108.0	49	85	90.80	94.8	79.7	0.92	3.03	
4853	PBS	39	18.4	46	154	55.8	88.3	58.8	0.52	2.72	81.50%
4838	PBS	40	740.0	121	253	78.8	84.2	92.4	1.05	4.57	93.10%
4843	PBS	46	25.5	61	223	63.7	93.1	74.3	0.82	3.15	74.70%
4844	PIPC	47	11.3	36	197	68.6	95.6	76.7	0.78	2.71	81.40%
4850	PBS	47	28.0	79	193	75.8	93.5	76.5	0.68	2.69	88.28%
4832	PIPC	49							1.2	3.48	
4846	PIPC	49	13.4	45	193	86.3	97.9	75.2	0.84	3.06	77.70%
4842	PBS	55	52.1	92	270	64.4	91.7	87.0	0.69	3.02	72.80%
4848	PIPC	56							0.93	3.66	67.35%
4835	PIPC	56	6.4	28	199	84.0	97.2	78.3	0.65	2.11	78.39%
4839	PBS	56	26.3	114	295	80.4	95.4	77.6	0.55	3	70.17%
4840	PBS	56	102	136	169	92.8	98.2	84.6	0.55	2.3	76.84%
4854	PBS	56	87	132	229	92.5	97.5	82.8	0.69	3.6	58.82%
4833	PIPC	57							Enlarged		
4847	PIPC	61	55	111	137	67	88.7	72.7	1.04	3.92	38.87%
4836	PIPC	63	6.3	83	250	31.8	77.8	40.3	0.31	1.52	27.80%
4845	PIPC	69	16.4	47	57	55.6	93.1	65.5	0.7	2.84	3.10%
4852	PBS	70	48.3	150	374	30.5	78.2	55.5	0.49	1.98	0%
4834	PIPC	70	28.6	124	304	30.9	74.5	51.6	0.49	1.42	3.50%
4837	PIPC	70	43.7	118	164	53.4	79.4	51.6	0.63	2.68	0%
4849	PIPC	83	34.1	108	184	45.2	76.9	56.3	0.59	2.88	0%

TABLE 4.1 Terminal data for all secondary MLL-AF9 leukaemia transplanted mice (*Ezh2fl/fl*; Cre+ treated with PBS or pIpC)

Ezh2 excision was assessed in splenic tissue for all mice where it was available. This is shown in the final column of Table 4.1. Whilst there are variable levels of excision in both cohorts, particularly in the mice that succumbed quickly to the disease, for those mice with longer survival there was a very low level of recombination, even in the pIpC treated cohort, suggestive of selective growth of the *Ezh2*$^{+/+}$ leukaemia cells. These results could potentially be explained by a slight degree of variance in the natural course of this secondary leukaemia in each animal. Some animals in the PBS treated arm had more aggressive, short latency leukaemias, presumably with a significant inflammatory response that was indirectly causing Cre mediated recombination. However the longest latency leukaemias, with virtually no excision of *Ezh2* at time of death, suggested recapturing and proliferation of any initial *Ezh2* replete clones that harboured a selective advantage, and with the longer latency, competitively outgrew *Ezh2* excised leukaemic cells, establishing their dominance at the time the animal became terminal. Having caused maximal excision with pIpC at the outset, this simply delayed this process and highlighted that *Ezh2* was facilitating the transformed state.

4.3.3 *Ezh2* is required for the maintenance of secondary AML1-ETO9a leukaemias *in vivo*

Primary transplants

The same process outlined in Section 4.3.2 was used to generate primary *Ezh2* conditional AML1-ETO9a leukaemias. Retroviral transduction was performed on c-kit selected bone marrow HSPCs taken from 2-3 *Ezh2 fl/fl*; Cre+ mice using retroviral supernatant derived from HEK 293T cells transfected with the MIGR1-AML1-ETO9a-GFP construct. This reliably yielded, on average, a higher transduction efficiency of 15-25% per experiment given the construct size is considerably smaller than MLL-AF9 (Figure 4.9). Given the lower potency of this particular oncogene, this also meant that more GFP positive cells were required per mouse for each transplant (in the order of 10^5 per mouse).

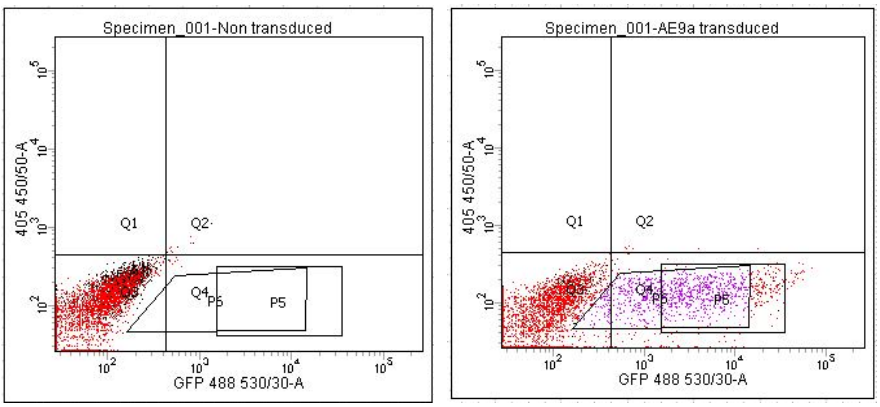


FIGURE 4.9 **Transduction efficiency of AML1-ETO9a into c-kit positive HSPCs**
Live cells gated initially by forward and side-scatter, doublets removed and then gated for GFP positivity. Population positive for GFP, reflecting AML1-ETO9a transduced cells seen in Q4, right (here ~21%)

After transduction, as above, the cell dosage was adjusted based on total cell number and percentage of GFP positivity, and an equal number of GFP positive cells were injected into five primary lethally irradiated C57/Bl6 recipients. These animals were then tracked with serial bleeds and GFP assessments, and clinically for the development of leukaemia. The expectation was that these animals would have a long latency (close to one year) as seen in previous studies. [4] However disease features became prominent at approximately 3 months, with significantly raised GFP in peripheral blood sampling and all 5 animals succumbed to leukaemia between Days 96 and 138 post transplantation. Each displayed hepatosplenomegaly, marked anaemia and thrombocytopenia as seen in the induction experiments. Splenic tissue and bone marrow was harvested and frozen in liquid nitrogen. Spleen cells were c-kit positive but negative for other myeloid antigens such as Mac-1 and Gr-1 (also negative for B220/CD4 and Ter119). Given the latency for these primary leukaemias was slightly shorter than

expected, genotyping to assess *Ezh2* recombination was performed, as with the MLL-AF9 primary leukaemias. Unsurprisingly given the presumably pro-inflammatory nature of leukaemogenesis it appeared once again that spontaneous Cre-recombinase activation had occurred, demonstrating 60-90% *Ezh2* excision had occurred using the qPCR method across the five animals (from DNA extracted from splenic tissues). This disease latency tallied with the acceleration and shorter disease latency seen in the *Ezh2fl/fl*;Cre+ pIpC pre-treated arm in the induction *in vivo* experiments, suggesting again even partial Ezh2 loss through spontaneous Cre effects was powerful enough to facilitate accelerated leukaemogenesis. Nonetheless, having seen this already happen with the MLL-AF9 primary transplants as described above, we decided to proceed with secondary transplants using splenic tissue from the mouse with the least *Ezh2* excision (approximately 60%).

Secondary transplants and *Ezh2* functional assessment

Splenic tissue from one of the primary *Ezh2fl/fl*;Cre+ AML1-ETO9a leukaemic mice with the lowest level of spontaneous *Ezh2* excision was thawed. This had an approximately 60% level of *Ezh2* excision. After washing the cells, approximately $1x10^6$ cells were injected via tail-vein into 20 sub-lethally irradiated wild-type C57/Bl6 mice (split thereafter into two arms). The cells were allowed 7 days for engraftment and the mice were then treated with intraperitoneal injections of either PBS (control) or pIpC (test arm), to induce complete *Ezh2* excision in the latter. Five doses were given on alternate days. They were then tracked for aggressive, short latency secondary leukaemias, through peripheral blood sampling, GFP assessment and clinical monitoring. *Ezh2* excision in the pIpC treated arm lead to a significant increase in survival compared to PBS treated (Figure 4.10):

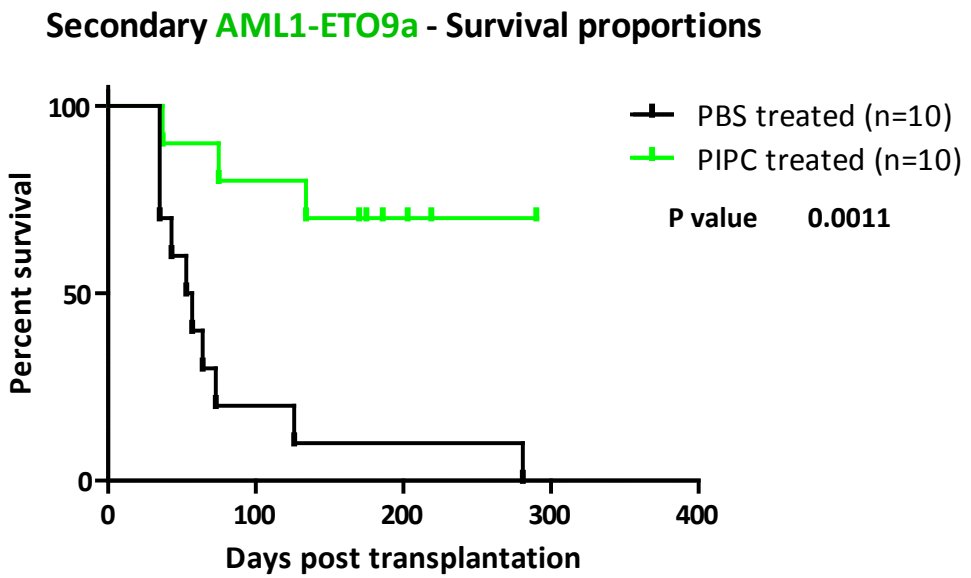


FIGURE 4.10 **Kaplan-Meier survival outcomes for *Ezh2fl/fl*; Cre+ AML1-ETO9a secondary recipient mice treated with PBS or pIpC**

Mice in the pIpC treated (i.e. *Ezh2* excised) arm demonstrate a significant increase in survival compared to PBS treated

As early as Day 35 post transplant, PBS treated animals became pre-terminal and were sacrificed. By Day 126, all PBS treated animals had been culled due to leukaemia development. In the pIpC arm only three out of ten animals actually developed secondary leukaemias, in general significantly later than the PBS treated animals. The remainder died of GFP-negative other causes (likely secondary to the effects of irradiation and aging), with several having massive thymic enlargement on necropsy (but GFP negative throughout all organs and blood counts not indicative of leukaemia) (Figure 4.11):

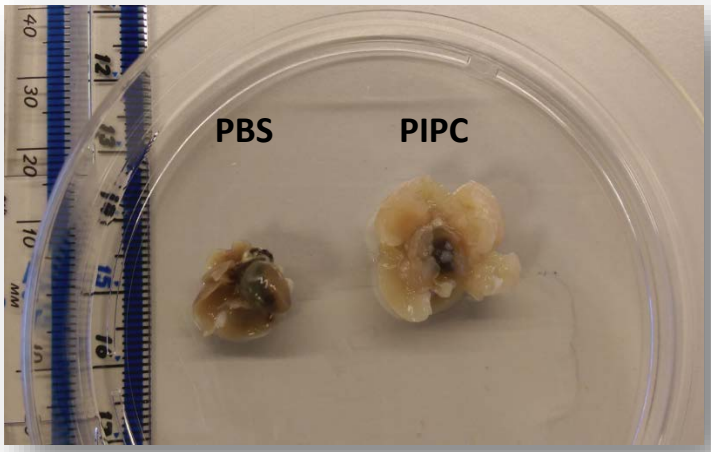


FIGURE 4.11 **Marked thymic enlargement seen in non-leukaemic mice transplanted with *Ezh2fl/fl*;Cre+ AML1-ETO9a splenic tissues and treated with pIpC (right) compared to PBS treated secondary leukaemic mouse (heart and lungs with minimal thymus)**

This suggested that maximal *Ezh2* excision in primary AML1-ETO9a transplanted leukaemic tissues, rendered a significant delay in the development of secondary leukaemia allowing other effects such as aging/post-irradiation to come into play, and/or through inhibiting leukaemia development, was influencing the development of other disease processes (outside the scope of this study). Details of all mice in this experiment (where available) are provided in Table 4.2:

						GFP%					
Mouse #	**PBS/PIPC treated**	**Day culled**	**WCC (x10^9/L)**	**Hb (g/dl)**	**Plts (x10^9/L)**	**Periph blood**	**Bone marrow**	**Spleen**	**Spleen wt (g)**	**Liver wt (g)**	**Excision**
4818	PBS	35	15.6	42	42	35.6	43.2	84.1	0.9	1.4	90.09%
4822	PBS	35	29.6	29	125	54.6	54.8	90.9	0.69	1.44	89.76%
4825	PBS	35	4.2	**	103	42.8	54.7	93.8	0.62	1.22	86.99%
4830	PIPC	37	40.5	46	55	92.0	85.0	90.0	0.69	2.03	90.95%
4821	PBS	43	74.0	54	165	94.7	30.6	82.8	1.01	2.77	84.63%
4823	PBS	56	15.6	36	134	41.8	50.1	82.5	1.02	2.45	79.17%
4826	PBS	57	176.0	40	115	93.9	58.8	84.2	0.74	3.18	83.43%
4817	PBS	64	26.1	73	95	53.6	53.4	55.5	1.02	3.16	41.49%
4820	PBS	73							0.64	1.95	75.51%
4814	PIPC	75	220	40	135	97.7	49.6	88.3	1.09	3.55	89.96%
4819	PBS	126	37.1	84	91	90.1			0.84	2.64	86.08%
4813	PIPC	134	75	56	171	39.4	77.0	66.9	1.52	3.22	90.70%
4812*	***PIPC***	*170*				*<10%*	*<10%*	*<10%*			
4815*	***PIPC***	*170*				*<5%*	*<5%*	*<5%*			
4816*	***PIPC***	*175*	*13.8*	*102*	*209*	*<2%*	*<2%*	*<2%*	*0.83*	*5.46*	
4831*	***PIPC***	*186*									
4827*	***PIPC***	*203*									
4829*	***PIPC***	*219*				*<1%*	*<1%*	*<1%*	*0.46*	*2.77*	
4824	***PBS***	*281*	*N*	*N*	*N*	*<1%*	*<1%*	*<1%*			
4828	***PIPC***	*290*	*7.7*	*158*	*172*	*<1%*		*<1%*	*1.28*	*3.24*	

TABLE 4.2 **Terminal data for all secondary AML1-ETO9a leukaemia transplanted mice (*Ezh2fl/fl*;Cre+ treated with PBS or pIpC)**

Mice identified in Bold indicate those with GFP negative disease at time of culling

*indicates mice with massive thymic enlargement at death. qPCR assessment of Ezh2 excision not performed in these latter cases as no evidence of leukaemia as determined by GFP expression

Ezh2 excision was assessed in splenic tissue for only leukaemic mice where tissue/DNA was available. This is shown in the final column of Table 4.2. There were varying levels of excision seen (between 41-90%) and did not follow a particular treatment arm. In view of the aggressive nature of these secondary leukaemias, we assumed that this level of excision was likely due to sustained Cre-activation caused by ongoing inflammation/interferon responses in the face of leukaemia progression in the PBS treated arm, though the phenotypic effects on survival clearly favour the pIpC treated arm (i.e. maximally *Ezh2* deleted after engraftment), producing a profound survival advantage. This again suggests that *Ezh2* appears to be a facilitator of the transformed state, with its complete removal at an early phase of leukaemia progression either blocking leukaemia completely (seven of ten pIpC treated mice did not develop GFP-positive leukaemia) or delaying it sufficiently to allow other disease processes to occur. Finally, there were no significant phenotypic effects on GFP distribution (though GFP percentage was marginally higher in bone marrow in the pIpC treated arm), peripheral blood counts or terminal organ size for all leukaemic mice between both arms (where available) (Figure 4.12):

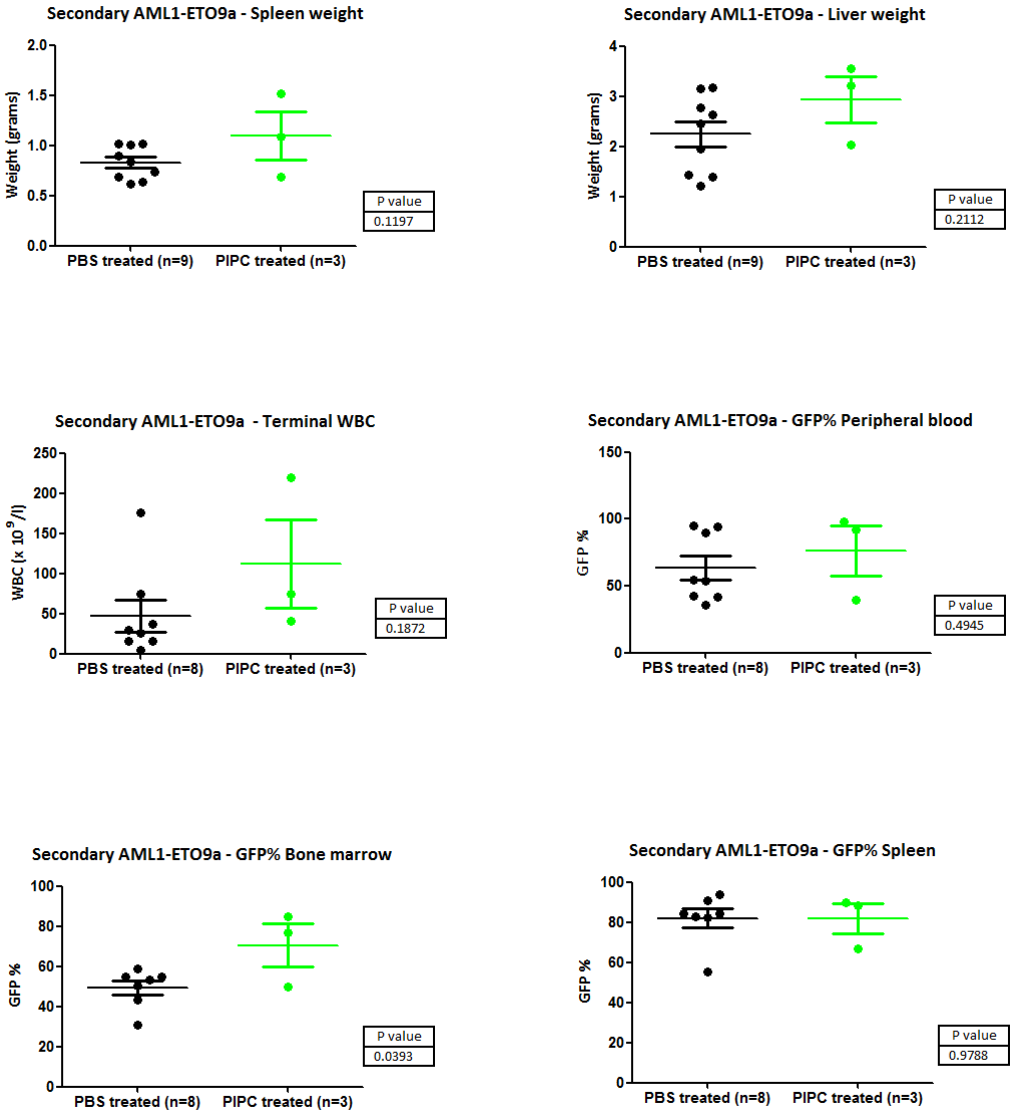


FIGURE 4.12 **Spleen and liver weights, terminal white blood cell counts and GFP percentages for blood, bone marrow and spleen for secondary AML1-ETO9a recipients treated with PBS or pIpC**

pIpC treated mice limited to those that developed leukaemia (n=3)

4.3.4 Loss of *Ezh2* during maintenance of MOZ-TIF2 leukaemias in vivo does not significantly alter outcome

Primary Transplants

As per MLL-AF9 and AML1-ETO9a, primary leukaemias were generated using c-kit positive bone marrow transduced from *Ezh2fl/fl*; Cre+ mice with retroviral supernatant produced from HEK 293T cells transfected with the MSCV-IRES-MOZ-TIF2-GFP vector. As with MLL-AF9, transduction efficiencies were typically low, ranging from 1-5%, assessed by GFP expression through flow cytometry (Figure 4.13):

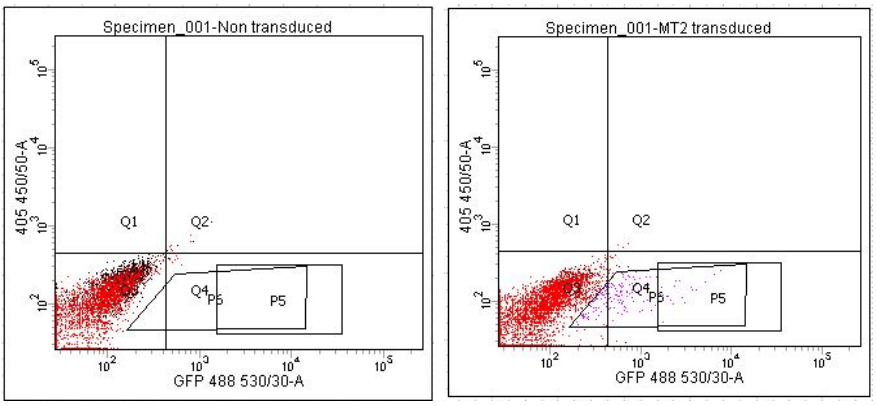


FIGURE 4.13 Transduction efficiency of MOZ-TIF2 into c-kit positive HSPCs
Live cells gated initially by forward and side-scatter, doublets removed and then gated for GFP positivity.
Population positive for GFP, reflecting MOZ-TIF2 transduced cells seen in Q4, right (here ~1%)

After transduction, as above, the cell dosage was adjusted based on total cell number and percentage of GFP positivity, and an equal number of cells were injected into five primary lethally irradiated C57/Bl6 recipients. These animals were then tracked with serial bleeds and GFP assessments, and clinically for the development of leukaemia. The expectation was that these animals would have a relatively short latency as seen in previous studies.[5] As expected, animals developed leukaemia as early as Day 47 post-transplant and by Day 83 all five had been culled with splenic tissues and bone marrow harvested. Through immunophenotyping these were confirmed to be Mac-1, Gr-1 and c-kit positive (myeloid), and again DNA from spleen was extracted to assess for any Ezh2 excision due to spontaneous Cre-mediated recombination for the reasons described above. Once again this was present, with a range from 65-74% *Ezh2* excision seen across the five mice.

Secondary transplants and Ezh2 functional assessment

Splenic tissue from one of the primary *Ezh2fl/fl*; Cre+ MOZ-TIF2 leukaemic mice with the lowest level of spontaneous *Ezh2* excision was thawed. This had an approximately 65% level of *Ezh2* excision. After washing the cells, approximately $1x10^6$ cells were injected via tail-vein into 20 sub-lethally irradiated wild-type C57/Bl6 mice (split thereafter into two arms). The cells were allowed 7 days for engraftment and then the mice were treated with intraperitoneal injections of either PBS (control) or pIpC (test arm), to induce complete *Ezh2* excision in the latter. Five doses were given on alternate days. They were then tracked for aggressive, short latency secondary leukaemias, through peripheral blood sampling, GFP assessment and clinical monitoring.

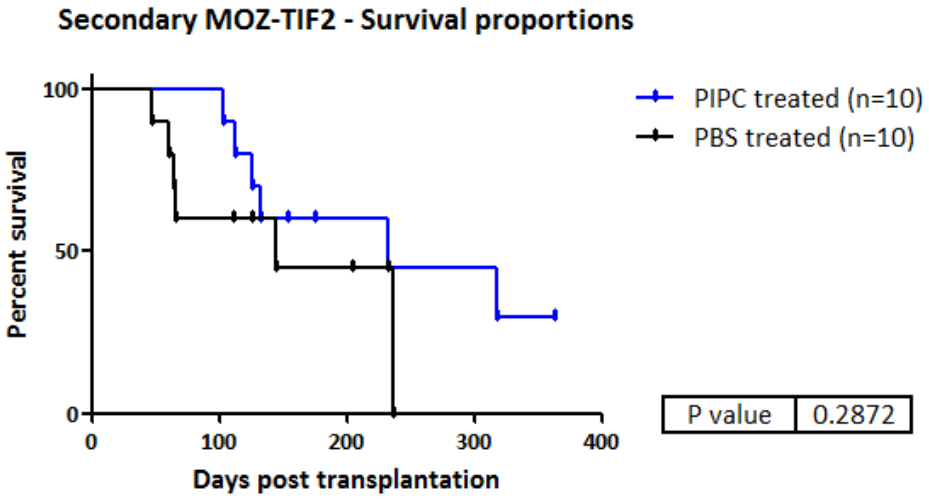


FIGURE 4.14 **Kaplan-Meier survival outcomes for *Ezh2fl/fl*; Cre+ MOZ-TIF2 secondary recipient mice treated with PBS or pIpC**
Though there was an initial trend toward survival advantage in the pIpC treated (i.e. *Ezh2* excised) arm, this was not statistically significant by the conclusion of this experiment

Though there appeared to be an initial early survival advantage of pIpC treated *Ezh2fl/fl*; Cre+ MOZ-TIF2 leukaemia secondary recipients, this was negated by around Day 140 post transplantation, with the final outcome being statistically non-significant (Figure 4.14). Furthermore, there was a considerable delay in some of the pIpC treated mice developing leukaemia, therefore the experiment was terminated at Day 320 in view of the outcome.

There was no significant difference between organ size, terminal blood counts or GFP distribution across haematopoietic compartments (where data available) (Figure 4.15):

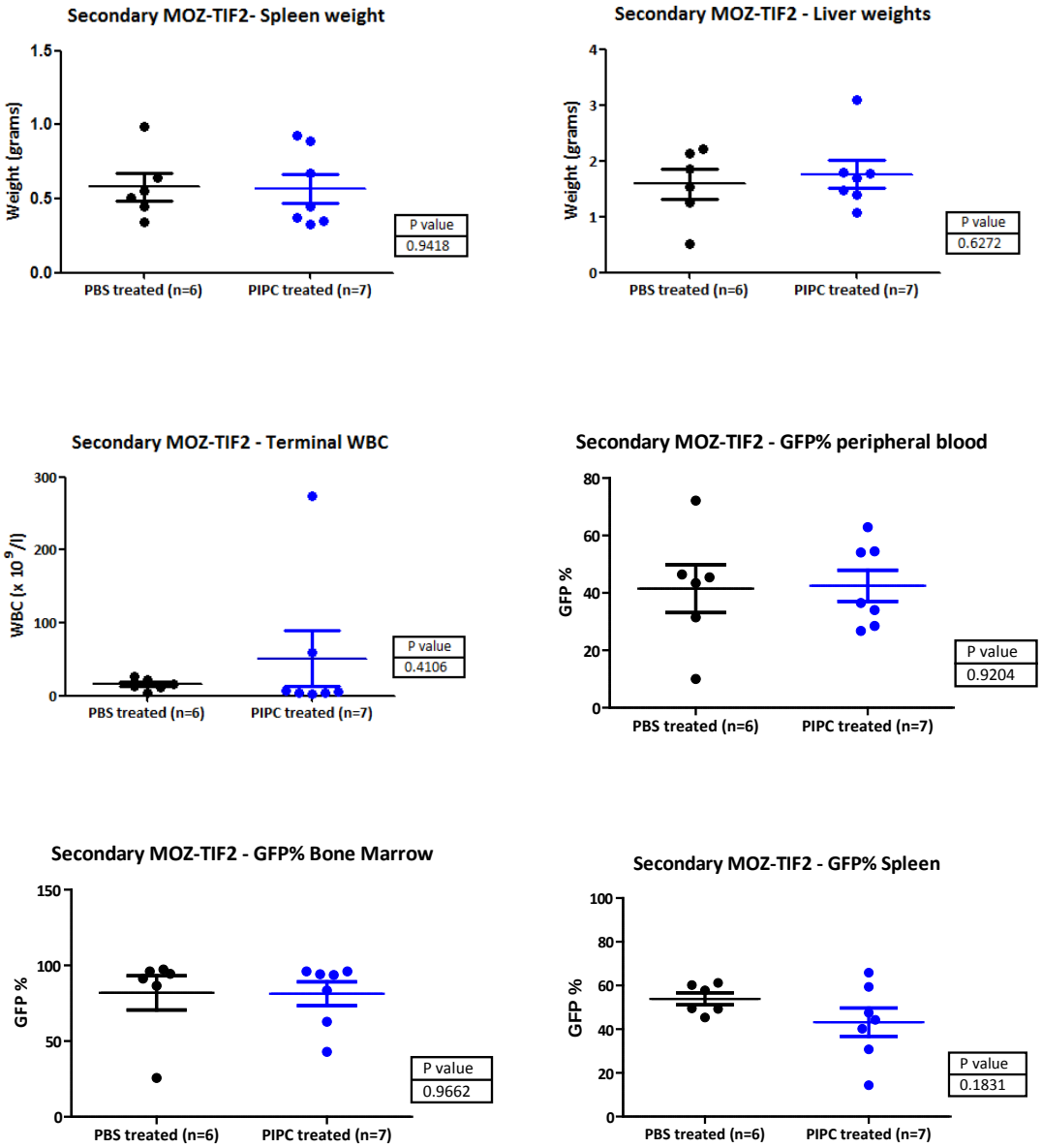


FIGURE 4.15 **Spleen and liver weights, terminal white blood cell counts and GFP percentages for blood, bone marrow and spleen for secondary MOZ-TIF2 recipients treated with PBS or pIpC**

In view of the non-significance attained between the study arms in this experiment, *Ezh2* excision status was not checked. As the experiment was prematurely terminated at Day 320, not all mice were analysed. Table 4.3 summarises the data for the secondary MOZ-TIF2 transplants:

							GFP%			
Mouse #	**PBS/PIPC treated**	**Day culled**	**WCC ($x10^9$/L)**	**Hb (g/dl)**	**Plts ($x10^9$/L)**	**Periph Blood**	**Bone Marrow**	**Spleen**	**Spleen wt (g)**	**Liver wt (g)**
4803	PBS	47	13.1	173	275	45.4	96.0	49.5	0.34	1.25
4806	PBS	61	16.4	78	176	46.4	94.4	57.8	0.45	1.53
4805	PBS	65	22.7	95	109	31.5	91.3	49.3	0.64	1.86
4798	PBS	66	26.2	123	103	43.4	97.3	61.2	0.55	2.21
4807	PIPC	103	6.9	133	223	28.5	94.2	44.3	0.45	1.48
4800*	***PBS***	*111*	*75.1*	*149*	*419*	*<10%*	*<10%*	*<10%*	*0.33*	*2.46*
4794	PIPC	113	59.1	167	372	54.1	83.4	30.8	0.37	1.8
4793	PIPC	126	3.5	142	279	36.5	96.0	65.9	0.67	1.77
4799*	***PBS***	*126*	*29.2*	*115*	*169*	*<1%*	*<1%*	*<1%*	*0.81*	*2.77*
4808	PIPC	133	274	110	101	54.5	62.8	59.4	0.89	3.1
4797	PBS	145	11.4	142	286	10.0	25.7	45.4	0.51	4.9
4796	PIPC	154	4.5	154	915	34.0	42.9	14.4	0.93	1.08
4792*	***PIPC***	175	9.5	154	915	*<5%*	*<2%*	*<2%*		
4801*	***PBS***	205	64.6	131	194	*0.0*	*0.0*	*0.0*	0.64	2.54
4804*	***PBS***	232	12.7	94	189	*7*	*4.7*	*2.2*	0.66	3.22
4809	PIPC	232	6.2	137	386	26.8	93.7	47.5	0.349	1.4
4802	PBS	237	3.8	124	265	72.2	86.6	60.2	0.99	2.14
4811	PIPC	318	2.7	144	597	62.9	96.1	40.2	0.33	1.7

TABLE 4.3 **Terminal data for all secondary MOZ-TIF2 leukaemia transplanted mice (*Ezh2fl/fl*;Cre+ treated with PBS or pIpC)**

*indicates mice with that died of causes other than leukaemia (i.e. GFP negative with no other phenotypic features)

4.4 Discussion and future work

The experiments detailed in this chapter showed striking effects of *Ezh2* loss during maintenance of AML. *In vitro*, *Ezh2* was clearly an absolute requirement across two aggressive subtypes (MLL-AF9, MOZ-TIF2) and a third, less aggressive subtype (AML1-ETO9a), with its genetic loss in already transformed cell lines completely stopping growth of these. These very prominent results translated generally into the *in vivo* studies, which demonstrated *Ezh2* loss interfered with maintenance of MLL-AF9 and AML1-ETO9a leukaemias, with its loss causing a significant improvement in survival in both cases. Furthermore, our results for the MLL-AF9 experiments were in alignment the previously published data. [1] With the inference being that *Ezh2* is a general facilitator of the transformed state, this provided strong support towards an oncogenic role for *Ezh2* in AML maintenance. However, common to both the primary and secondary leukaemias in this chapter, where characterised, was the problem of spontaneous Cre-mediated recombination occurring during leukaemia development. The experiment was deliberately designed in such a way so as to delete *Ezh2* only at the maintenance stage i.e. disease propagation in the secondary leukaemias, however it was hampered by early deletion during the development of the primary leukaemias and possible selection for the *Ezh2* deficient cells during leukaemia induction, and quite possibly further enhanced during progression of the secondary leukaemias, even in the absence of pIpC treatment. Spontaneous Mx1-Cre recombinase activation when modelling haematopoietic disease has recently been reported in the literature and highlights potential pitfalls in utilising this in conditional knock-out experiments.[6] However, in these maintenance *in vivo* experiments, whilst we accept the eventual comparison was not strictly between $Ezh2^{+/+}$ versus $Ezh2^{-/-}$, it was instead a comparison of a baseline of partially excised *Ezh2* (in the PBS arm) versus wholly excised *Ezh2* (in the pIpC treated arm) which clearly significantly influenced survival outcomes and still provides functional relevance and evidence of *Ezh2*'s oncogenic role in AML maintenance. In this respect, the prediction is that the requirement of *Ezh2* is actually underestimated.

Interestingly, the maintenance study findings were in stark contrast to the findings in the induction *in vivo* experiments detailed in the previous chapter, where early loss of *Ezh2* clearly favoured MLL-AF9 and AML1-ETO9a driven leukaemogenesis. This diametrically opposite effect of *Ezh2* loss during different stages of AML evolution (and also variance in induction, dependent upon AML subtype) serves to highlight how complex this epigenetic regulator is, playing two very different roles as tumour-suppressor and oncogenic facilitator in specific contexts. Having seen varying effects dependent upon AML subtype in the induction experiments, from a therapeutic perspective, it was reassuring to see a uniform absolute inhibitory effect of *Ezh2* loss upon maintenance of the all three subtypes *in vitro* and a similar, though less absolute effect upon secondary leukaemia progression *in vivo*. This suggests that

there may be an overlap in programmes controlled by *Ezh2* that are relevant to AML maintenance across these three subtypes and would be interesting to explore further. Furthermore, it supports an argument towards pharmacologically targeting EZH2 in AML.

As alluded to in the previous chapter, to try and explain this oncogenic effect, the next steps in these maintenance experiments (as per the original aims of this thesis) were to assess for any differences in gene expression through RNA-sequencing of AML tumours from mice that were *Ezh2* replete versus *Ezh2* deplete. In doing so, we would be able to characterise which transcriptional programmes were being subverted by loss of *Ezh2* during maintenance of disease, to account for the oncogenic effect it was clearly producing when present. An additional aim was to assess differences in H3K27 trimethylation between the representative tumours from the two maintenance arms to determine if any gene expression differences were in part due to loss of histone methyltransferase activity. As a result of the higher than expected levels of spontaneous recombination in both PBS treated MLL-AF9 and AML1-ETO9a partially *Ezh2*-deleted arms, we felt that comparing gene expression and histone marks from our experiments designed to genetically ablate *Ezh2 in vivo* would provide relatively modest differential results due to the modest differences in *Ezh2* expression. This would therefore not be a direct comparison of complete $Ezh2^{+/+}$ against a fully *Ezh2*-deficient state as it was in the induction experiments.

To bypass this issue and still assess which transcriptional programmes were being governed by Ezh2 in the maintenance of these AML murine tumours, we utilised a novel EZH2 inhibitor or its vehicle control (described in Chapter 7) against established *Ezh2fl/fl*; wt MLL-AF9 tumours cultured *in vitro* to draw out any differences in gene expression. This is detailed in Chapters 5 and 7.

Having completed the Induction and Maintenance experiments, it was possible to conclude that *Ezh2* had markedly opposite roles during these stages of AML evolution, serving to highlight the complexity of this epigenetic regulator. By moving on to analyse the genomic alterations produced by *Ezh2* loss in either stage, our hypothesis was that *Ezh2* loss de-repressed a completely different gene expression set in either cellular context, and we hoped to demonstrate this. In doing so, we would then be able to identify downstream oncogenic facilitators that were de-repressed in AML induction, and tumour-suppressive candidates that were de-repressed in AML maintenance as a result of *Ezh2* loss/Ezh2 inhibition.

References

1) Tanaka S, Miyagi S, Sashida G, Chiba T, Yuan J, Mochizuki-Kashio M, Suzuki Y, Sugano S, Nakaseko C, Yokote K, Koseki H, Iwama A. *Ezh2* augments leukemogenicity by reinforcing differentiation blockage in acute myeloid leukemia. *Blood*. 2012 Aug 2;120(5):1107-17

2) Loonstra A, Vooijs M, Beverloo HB, Allak BA, van Drunen E, et al. Growth inhibition and DNA damage induced by Cre recombinase in mammalian cells. *Proc Natl Acad Sci USA.* 2001; 98: 9209–9214.

3) Higashi AY, Ikawa T, Muramatsu M, Economides AN, Niwa A, et al. Direct hematological toxicity and illegitimate chromosomal recombination caused by the systemic activation of CreERT2. *J Immunol.* 2009 182: 5633–5640.

4) Rasmussen KD, Jia G, Johansen JV, Pedersen MT, Rapin N, Bagger FO, Porse BT, Bernard OA, Christensen J, Helin K. Loss of TET2 in hematopoietic cells leads to DNA hypermethylation of active enhancers and induction of leukemogenesis. *Genes Dev.* 2015 May 1;29(9):910-22

5) Deguchi K, Ayton PM, Carapeti M, Kutok JL, Snyder CS, Williams IR, Cross NC, Glass CK, Cleary ML, Gilliland DG. MOZ-TIF2-induced acute myeloid leukemia requires the MOZ nucleosome binding motif and TIF2-mediated recruitment of CBP. *Cancer Cell.* 2003 Mar;3(3):259-71.

6) Velasco-Hernandez T, Säwén P, Bryder D, Cammenga J. Potential Pitfalls of the *Mx1-Cre* System: Implications for Experimental Modeling of Normal and Malignant Hematopoiesis. *Stem Cell Reports.* 2016 Jul 12;7(1):11-8.

Chapter 5

Genomic analysis of transcriptional programmes subverted by Ezh2 during AML induction and maintenance

5.1 Aims

This section details the genomic analysis performed with the aim of characterising the mechanisms by which *Ezh2* has opposing effects in induction and maintenance of AML. A significant proportion of the data discussed in this section was analysed and referenced with the help of Dr Eshwar Meduri, a bioinformatician within the Huntly group.

5.2 Introduction

Having established the completely contrasting role of *Ezh2* between AML induction and maintenance, the next step was to analyse the transcriptional programmes manipulated and altered by *Ezh2* loss in either phase of disease. The stark contrast in *Ezh2* functioning as a tumour-suppressor during induction of MLL-AF9 and AML1-ETO9a leukaemia *in vivo* when compared to its oncogenic role in the maintenance of these leukaemias *in vivo*, provided a substantial challenge to explain. As *Ezh2* is an epigenetic regulator, we employed both differential gene expression analysis (through RNA-sequencing) and histone modification / PRC2 complex member binding analysis (through ChIP-sequencing) to provide answers to this question.

We initially chose to focus more on the most unexpected finding (and contrary to the original hypothesis) that *Ezh2* loss accelerated MLL-AF9 and AML1-ETO9a leukaemia induction. Both leukaemia models are well characterised in the literature and have direct clinical relevance, particularly with the human equivalent of AML1-ETO fusion – the t(8;21) (q22;q22.1) or *RUNX1-RUNX1T1* translocation being the most common fusion oncogene seen in AML. Two recent studies have extensively profiled the genomic landscape of the core-binding factor leukaemias (*RUNX1-RUNX1T1* or *CBFB-MYH11* rearrangements) through whole genome or whole-exome sequencing of paediatric and adult cases.[1,2] Both have demonstrated that not only are there an array of relatively underappreciated co-operating mutations with these translocations with remarkably different spectra between the two, but that there appeared to be a clear enrichment for certain mutations in the *RUNX1-RUNX1T1* cases, particularly the cohesin complex members and a number of important epigenetic regulators. Surprisingly, and highly

relevant to our finding with the induction experiments, was that in both studies, *EZH2* was exclusively mutated (all putative loss-of-function mutations) in a reasonable proportion of *RUNX1-RUNX1T1* cases compared to all *CBFB-MYH11* cases analysed. In the first study, there were 7 *EZH2* mutated cases in a total of 106 cases (~7%, p-value 0.006) and in the second there were 7 *EZH2* mutated cases from a total of 85 cases. The inference from this, particularly considering the cooperative nature of *Ezh2* loss and AML1-ETO9a induction in the *in vivo* experiment, might be a functional association between *EZH2* mutations and *RUNX1-RUNX1T1* driven leukaemias in humans. It would be interesting to characterise variant allele frequencies in larger human AML datasets of *EZH2* mutations in t(8;21) cases to see if this could firstly corroborate a link between *EZH2* mutations in t(8;21) leukaemias and then inform regarding order of mutation, and the tantalising possibility that *EZH2* mutations might occur before or at the time of the translocation and co-operate to drive leukaemogenesis.

In the induction experiments, *Ezh2* had been clearly deleted in the *Ezh2 fl/fl;* Mx1-Cre heterozygous arm (whilst it was impossible to excise in the *Ezh2 fl/fl*; wt arm by design due to lack of Mx1-Cre) prior to retroviral overexpression of and transformation by MLL-AF9 or AML1-ETO9a. This suggested that *Ezh2* loss may have fostered an altered epigenetic and/or transcriptional state, by virtue of its role as a transcriptional repressor, resulting in the de-repression of genes in the non-transformed state that would potentially co-operate with the driver oncogene in hastening leukaemia development. The initial focus therefore was to characterise differentially expressed genes in HSPCs following *Ezh2* loss in the non-transformed setting (i.e. $Ezh2^{+/+}$ vs $Ezh2^{-/-}$) and to then compare this with the same analysis performed in the $Ezh2^{+/+}$ and $Ezh2^{-/-}$ MLL-AF9 and AML1-ETO9a leukaemias. In triangulating these datasets, we aimed to find common genes that were de-repressed following *Ezh2* loss that would functionally prioritise those able to augment the process of transformation by the fusions.

To enhance the analysis in the non-transformed setting further, we then sought to assess whether loss of *Ezh2* had genome-wide effects upon its repressive H3K27me3 mark and also whether there were any associated changes in other linked local histone activating marks – specifically H3K27Ac and H3K4me3. We additionally pursued binding of the PRC2 complex in the $Ezh2^{+/+}$ setting, through ChIP-sequencing analysis of Ezh2 and Eed binding. We hypothesised that genes de-repressed and activated by *Ezh2* loss would feature a significant reduction in H3K27me3 at their cis-regulatory promoters and enhancers, and reduced PRC2 binding in the absence of Ezh2 (as directed by Eed localisation). At the same time, we were keen to see whether these regions also had the presence of H3K27Ac/H3K4me3 facilitating transcriptional activation and whether there was any change in these marks to further explain increased gene expression. From this comprehensive analysis, we aimed to identify any candidate drivers of disease and functionally validate them.

To address the maintenance *in vivo* experiments, we adopted a different strategy from the original aim, given the unexpected interference of spontaneous Cre-recombination activity in the primary and secondary leukaemias. Despite showing a clear effect on survival, we felt these would only allow a comparison between partially and completely *Ezh2* excised leukaemias, meaning that any differential effects at the genomic level would be relatively diluted. We were able to establish effective liquid culture growth of thawed and washed *Ezh2fl/fl*; wt MLL-AF9 spleen tumour cells from two separate mice with relative ease, allowing us to utilise chemical inhibition of Ezh2 to abrogate its function. We then used GSK343 (a novel EZH2 inhibitor available to us through collaboration with GlaxoSmithKline) or vehicle control (DMSO) against these tumours in liquid culture to impede growth (full details in Chapter 7). Furthermore, we demonstrated a clear reduction in H3K27me3 of these cells following GSK343 treatment, as early as 24 hours. Having seen the growth inhibitory effects on these MLL-AF9 tumours, we used this pharmacological 'knockdown' of Ezh2 as a system to assess which transcriptional programmes were being affected, through RNA-sequencing of the tumours +/- GSK343 inhibitor. Given the opposite effect of *Ezh2* loss (i.e. delaying leukaemia progression) in maintenance, we expected to find a different gene-set altered in maintenance when compared to induction, to account for the phenotypic effects.

5.2 Differential gene expression analysis between *Ezh2*$^{+/+}$ and *Ezh2*$^{-/-}$ states in the non-transformed setting reveals an altered transcriptional programme that may facilitate leukaemia induction

As described in Methods 2.8.1, lineage negative selection was performed on whole bone marrow from pIpC pre-treated *Ezh2 fl/fl;* wt and *Ezh2 fl/fl; Mx1-Cre*+ mice, to isolate early HSPCs of each genotype. A yield of approximately 1x10^6 lineage marker negative cells per mouse on average was attained. The purity (typically 85-95% pure) of these was confirmed by flow cytometry analysis (example in Figure 5.1):

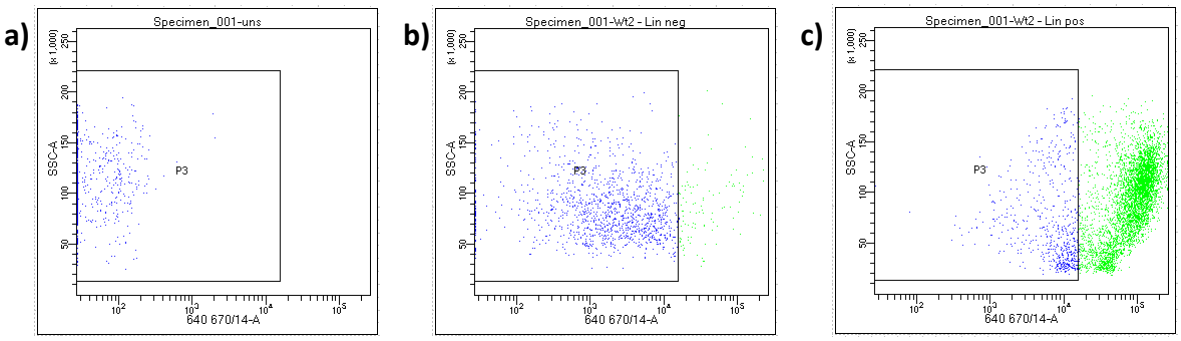


FIGURE 5.1 **FACS analysis of Lineage marker depleted murine bone marrow cells**
Live cells firstly gated by forward and side-scatter, doublets then removed. Lineage antibody cocktail stained cells gated on side-scatter and 640 670/14-A channel
a) Unstained cells (control)
b) Lineage depleted cells in flow-through (~92% of events in this example)
c) Lineage positive (mature cells) isolated from column after selection

DNA was extracted from flow-through lineage positive cells and used for genotyping by qPCR to confirm both presence of *Mx1-Cre* and adequate *Ezh2* excision in the *Ezh2 fl/fl; Mx1-Cre*+ arm (compared to lack of *Mx1-Cre* and no excision in the *Ezh2 fl/fl;* wt arm). *Ezh2* excision was typically as high as 90-98% in these bone marrow cells.

Having established the correct genotypes after pIpC treatment, lineage negative cells were pooled together from 2-3 mice within each genotype and RNA was extracted (as described in Methods 2.8.14.2). This was then quantified using a Nanodrop Spectrophotometer and approximately 5μg per condition were submitted for paired-end sequencing on an Illumina HiSeq 4000 machine at the Genomics Core, CRUK Cambridge Institute (Methods 2.8.15).

Once sequencing data returned, it was processed as described in Methods 2.8.13. Through this a filtered list of significantly differentially expressed genes was generated. Significant gene expression results were those with a p-value adjusted for multiple hypothesis testing of <0.05 with a $\log_2$ fold change of greater or less than 0.5.

Consistent with its role as a transcriptional repressor, there were considerably more significantly upregulated genes (n= 476) following *Ezh2* deletion than those downregulated (n= 48). A list of the top 25 genes de-repressed following *Ezh2* loss is given in Table 5.1 (genes were ranked in order of $\log_2$ fold change). The full list of differentially expressed genes (both up- and downregulated) is given in Appendix A5.1.

Gene ID	log_2FoldChange	Adjusted p-value
Lin28b	7.2612	1.1E-10
Scn4b	5.9047	8.96E-28
Pawr	5.7497	1.52E-29
Bmpr1a	5.6969	3.95E-59
Pcolce2	5.5429	1.66E-13
Ackr3	5.0924	0.016653374
Spats2	4.9315	3.5E-44
Cttnbp2	4.5789	0.010903684
Cma1	4.5477	2.32E-21
Cx3cl1	4.5347	0.002186949
Fjx1	4.4137	0.000000302
Adamts20	4.3240	0.0000676
Thbs4	4.3219	0.001302677
Syt14	4.2766	0.002186949
Yes1	4.2602	0.000247669
Igf2bp3	4.2270	1.05E-30
Trnp1	4.2021	0.022904117
Vpreb1	4.0931	3.86E-21
Gfra2	4.0801	0.00000825
Dsp	4.0644	0.008409518
Ddah1	4.0147	8.07E-13
Cnn1	3.9225	0.001753485
Igll1	3.8967	1.51E-18
Itga7	3.8774	0.037784092
Dscam	3.8103	0.000287645

TABLE 5.1 **Top 25 differentially upregulated genes in *Ezh2*$^{-/-}$ lineage negative HSPCs**
Arranged in order of log_2 fold change (highest first)

Notably, *Ezh2* appears in the downregulated list, aligning with its confirmed deletion by qPCR genotyping prior to RNA-sequencing. The whole list is presented in the form of a volcano plot (Figure 5.2):

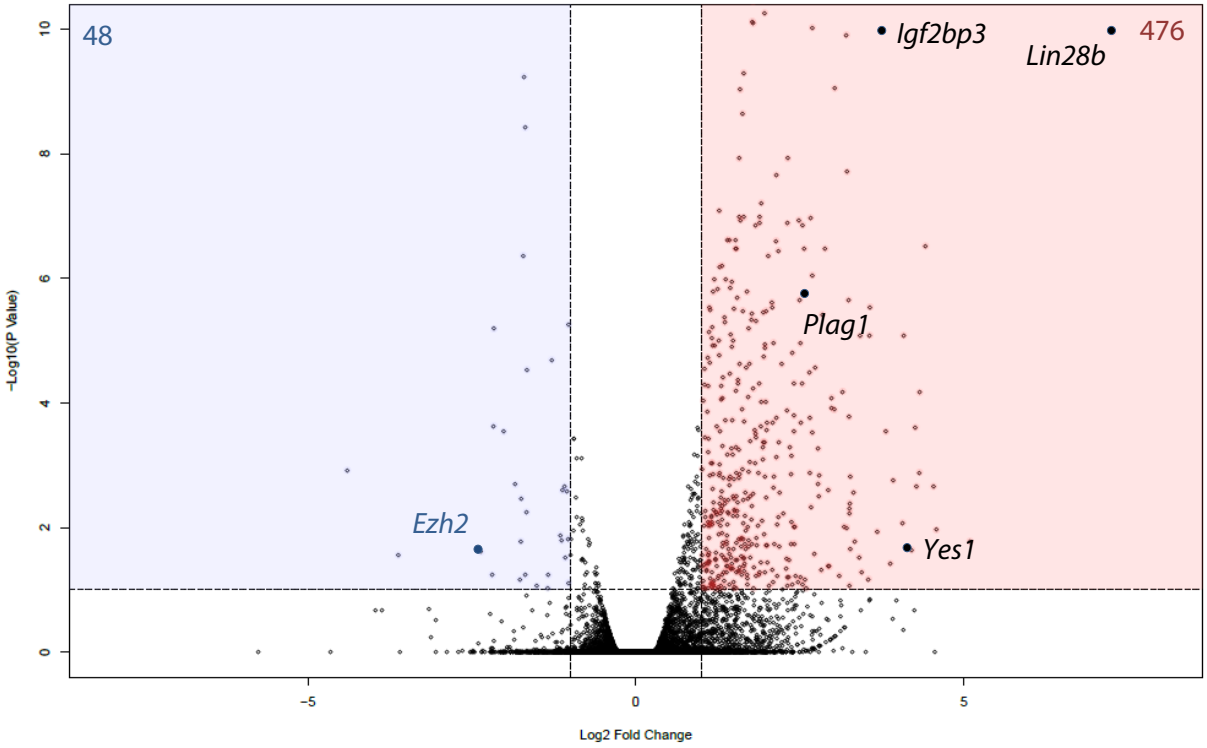
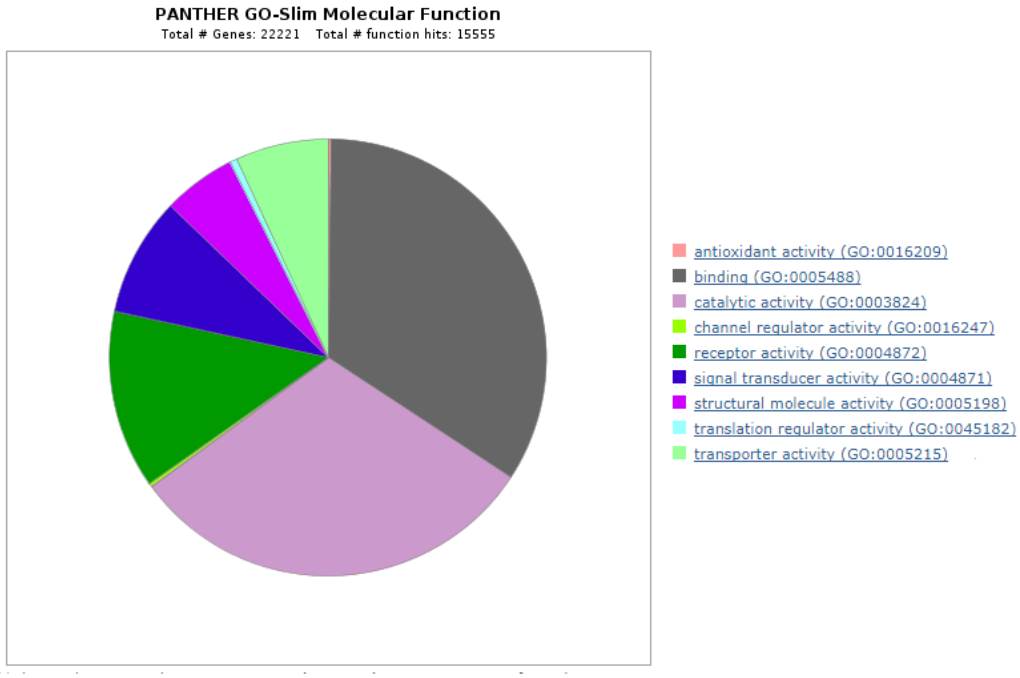


FIGURE 5.2 **Volcano plot representing differentially expressed genes determined by RNA-sequencing between *Ezh2*$^{+/+}$ and *Ezh2*$^{-/-}$ lineage negative HSPCs**
The plot shows both negative and positive log2fold changes in gene expression (x-axis) and their adjusted p-value demonstrating significance (y-axis)
Certain genes of interest, e.g. *Lin28b*, *Igf2bp3*, *Yes1* and *Plag1* are highlighted
Ezh2 appears in the downregulated (blue)

Within the top 25 were several potentially interesting candidate genes (exhibiting a significant log_2fold increase). *Lin28b*, a negative regulator of the let-7 micro-RNA tumour-suppressor family, was maximally upregulated, featuring at the top of this list. As discussed in Chapter 1, this candidate (along with *Igf2bp3* – also present in the top 25) was recently implicated as a potential downstream oncogenic effector following *Ezh2* loss in cooperation with *JAK2*-V617F in driving myelofibrosis. [3] Other candidates such as *Igf2bp3* – another oncogenic driver in solid malignancies[4], and a target of let-7 repression, were also upregulated. Additionally, the tyrosine kinase Yes1, which has recently been shown to play a facilitator role in AML1-ETO leukaemia maintenance was also upregulated. [5] A number of other candidates overlapped between this and the top 10 upregulated genes following *Ezh2* loss in the myeloproliferative disorder study, such as *Pcolce2, Ddah1, Bmpr1a, Spats2* etc., suggesting their roles in facilitating myeloid leukaemogenesis. [3]

Gene ontology analysis (via http://www.geneontology.org/) of the 476 upregulated genes following *Ezh2* loss revealed enrichment within biological processes for positive regulation of calcium ion transport, cell-cell adhesion, immune responses and negative regulation of cell proliferation. Analysis by molecular function demonstrated enrichment particularly for cytokine receptor activity and binding, catalytic activity and receptor activity suggesting Ezh2 plays a role in regulating these processes (Figure 5.3):



GO molecular function	Mus musculus - REFLIST (22221)	Fold Enrichment	p-value
cytokine receptor activity (GO:0004896)	89	7.24	5.34E-04
cytokine binding (GO:0019955)	102	5.79	1.45E-02
kinase binding (GO:0019900)	719	2.76	1.29E-04
protein kinase binding (GO:0019901)	645	2.5	1.87E-02
receptor binding (GO:0005102)	1671	2.02	2.72E-04
protein dimerization activity (GO:0046983)	1305	1.97	2.12E-02
enzyme binding (GO:0019899)	2024	1.75	1.97E-02
protein binding (GO:0005515)	8402	1.49	6.41E-11
binding (GO:0005488)	12610	1.31	4.52E-10
Unclassified (UNCLASSIFIED)	1859	0.66	0.00E+00

FIGURE 5.3 **Gene ontology analysis of genes deregulated following *Ezh2* loss**
Molecular functions shown
Panther overrepresentation test (release Apr 2017)
GO Ontology database (release Aug 2017)

5.3 Differential gene expression analysis between *Ezh2*$^{+/+}$ and *Ezh2*$^{-/-}$ in the MLL-AF9 and AML1-ETO9a induction leukaemias divulges key potential downstream mediators of AML induction

Having established a clear gene set differentially regulated by Ezh2 in the non-transformed setting, the same process was applied to the leukaemias that had accelerated upon *Ezh2* loss in the induction experiments. Two representative tumours (*Ezh2*$^{+/+}$ *and Ezh2*$^{-/-}$) from each of the MLL-AF9 and AML1-ETO9a induction *in vivo* experiments had DNA and RNA extracted from spleen cells (previously frozen) after thawing and washing. The DNA was used to confirm that levels of *Ezh2* excision were >95% in both Cre+ tumours (as expected) and that there was no excision in the wt arm (Figure 5.4):

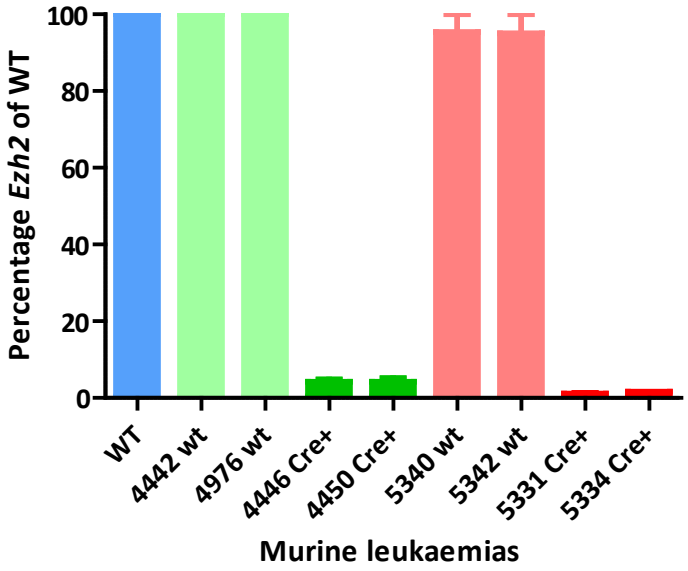


FIGURE 5.4 **Percentage of *Ezh2* across *Ezh2*$^{fl/fl}$; wt Cre (wt) and *Ezh2*$^{fl/fl}$; heterozygous Cre (Cre+) AML1-ETO9a and MLL-AF9 murine leukaemias as assessed by q-PCR**
WT (blue) = *Ezh2*$^{fl/fl}$ wild-type DNA (control)
AML1-ETO9a leukaemias in light/dark green
MLL-AF9 leukaemias in light/dark red
Numbers indicate identity of mouse. Data represent the mean and SEM of two independent experiments

Once confirmed, RNA was extracted from the same splenic tissues and after quantifying, and sent for paired-end sequencing on an Illumina HiSeq 4000 machine at the Genomics Core, CRUK Cambridge Institute. The sequencing data was again processed as per Section 5.2 and filtered to yield a list of significantly differentially expressed genes with a p-value adjusted for multiple hypothesis testing of <0.05 with a $\log_2$ fold change of greater or less than 0.5.

For the MLL-AF9 leukaemias, differential expression was more balanced than in the non-transformed setting suggesting a possible interaction with *Ezh2*'s repressive role, with 368 genes upregulated and 278 downregulated following *Ezh2* loss. This may also correlate with the aggressive nature of MLL-translocated leukaemias driven by different transcriptional programmes that are likely capable of overwhelming other programmes. For the AML1-ETO9a leukaemias, the balance was similar to the non-transformed setting with 388 genes upregulated and 55 genes downregulated following *Ezh2* loss (again consistent with AML1-ETO9a being a less aggressive leukaemia and the possibility that there is less interaction between AML1-ETO9a and Ezh2). The top 25 genes upregulated following *Ezh2* loss in both these induction leukaemias are listed in Table 5.2 (with the entire lists in Appendices A 5.2 and A 5.3 respectively). Figure 5.5 shows these in volcano plot format for a direct comparison:

MLL-AF9

Gene ID	log_2FoldChange	Adjusted p-value
Lin28b	8.3571	0.000409407
Spag6	7.8819	1.39E-20
Cacnb2	6.2302	1.31E-32
2900092N22Rik	6.1739	4.48E-27
Prg2	6.0452	0.003174392
Hist1h2al	5.8225	0.025360131
Gm37437	5.6912	8.53E-12
C330013J21Rik	5.6301	2.15E-07
Gm13269	5.3726	3.78E-11
Gm37891	5.3422	3.02E-09
Gm38273	5.3385	1.29E-10
Retnla	5.3260	2.97E-21
Gm44246	5.2468	2.83E-05
Gm37006	5.2299	5.79E-19
A1cf	5.0160	0.034048087
Gm4969	5.0027	7.34E-06
Gm38105	4.9686	4.85E-08
Col19a1	4.8757	0.000120183
Sorcs2	4.5664	3.05E-18
C330013J21Rik	4.5477	0.001277686
Slc22a3	4.4833	1.21E-14
Gm20539	4.3688	0.000573742
9430087J23Rik	4.3135	2.19E-05
Ikzf3	4.1700	0.000295085
Tmem178	4.1529	1.35E-18

AML1-ETO9a

Gene ID	log_2FoldChange	Adjusted p-value
Retnla	9.6530	1.44E-18
Fbp1	8.1861	6.59E-12
Perp	6.8145	7.72E-10
Cygb	6.7721	0.035404506
Ceacam10	6.4928	2.11E-10
BC018473	6.3207	0.011154342
Flt1	6.2741	0.026791608
Csgalnact1	6.0371	6.91E-06
Nat8l	5.9729	0.023345235
Slc35d3	5.8415	2.66E-09
Stfa2l1	5.8361	0.030335267
Gm5483	5.8153	0.032768258
Trpv3	5.7963	0.000799199
Mmp8	5.6479	4.08E-09
Glrp1	5.6026	0.021156303
Ighv1-23	5.5947	0.002173496
Crispld2	5.5663	1.30E-08
Gm2895	5.5628	0.000814866
Chil3	5.5373	6.28E-09
Plcb4	5.5205	1.68E-07
Gm16031	5.5128	1.66E-05
Mcam	5.4957	0.000409001
Hif3a	5.4893	3.67E-08
Prune2	5.4747	0.000538379
Gp1bb	5.3697	1.30E-05

TABLE 5.2 **Top 25 differentially upregulated genes in *Ezh2*$^{-/-}$ MLL-AF9 and AML1-ETO9a leukaemias**

Arranged in order of log_2 fold change (highest first)

Left (red) – MLL-AF9 leukaemias, Right (green) – AML1-ETO9a leukaemias

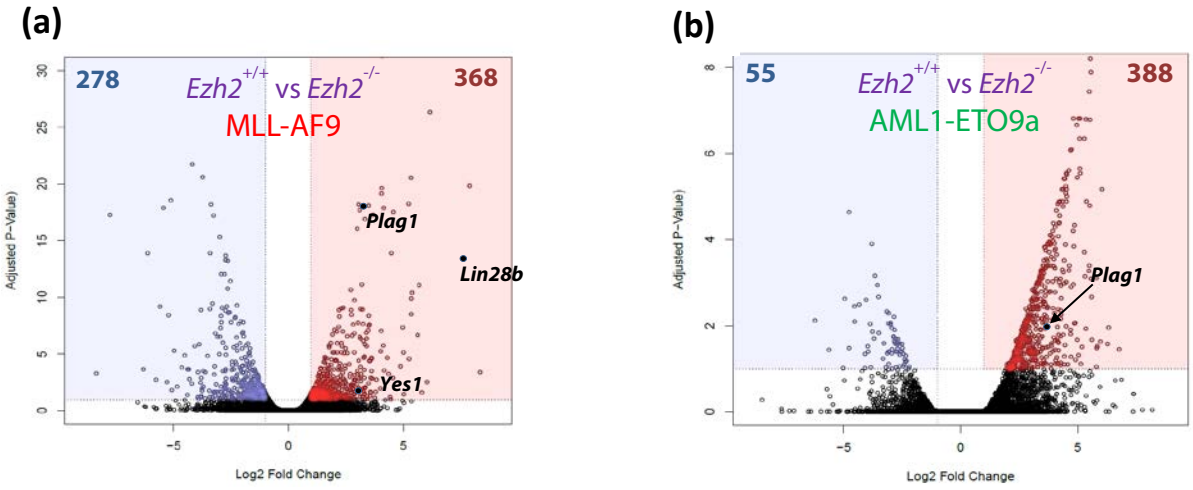


FIGURE 5.5 **Volcano plots representing differentially expressed genes determined by RNA-sequencing between *Ezh2*$^{+/+}$ and *Ezh2*$^{-/-}$ MLL-AF9 (a) and AML1-ETO9a leukaemias (b)**

Certain genes of interest, e.g. *Lin28b*, *Yes1* and *Plag1* are highlighted.
Upregulated (red), downregulated (blue)
The plot shows both negative and positive $\log_2$fold changes in gene expression (x-axis) and their adjusted p-value demonstrating significance (y-axis)

Strikingly, the gene most significantly upregulated in the *Ezh2*-null MLL-AF9 leukaemias was also *Lin28b*, as per the non-transformed setting. Other genes of interest, further down the list, were also present such as transcription factor *Plag1* and *Yes1*. The AML1-ETO9a leukaemias, whilst also having a significant number of upregulated genes, had a lesser degree of overlap with interesting candidates identified in the non-transformed differential gene expression analysis. However, *Plag1* was notably present and significantly upregulated, and was common to all three groups. A triangulation overlapping all genes differentially upregulated following *Ezh2* loss in the non-transformed, MLL-AF9 and AML1-ETO9a leukaemias was made (Figure 5.6):

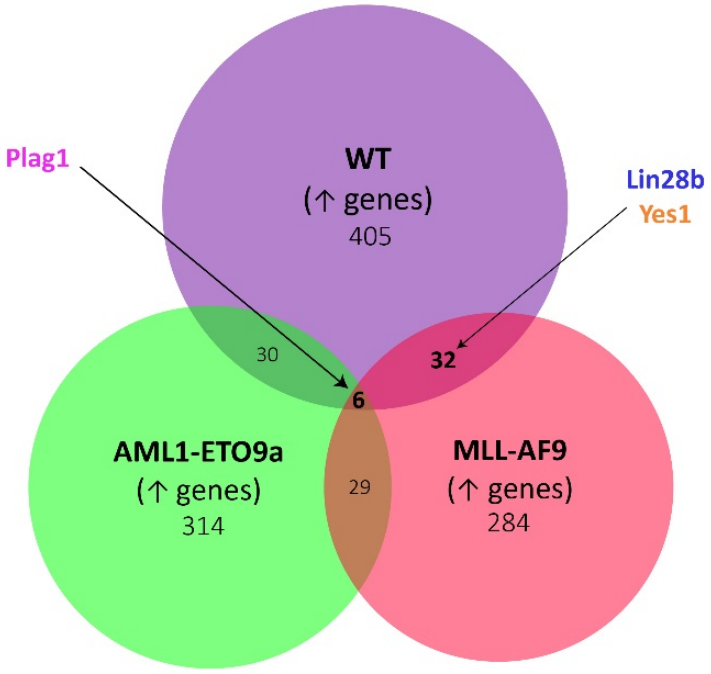


FIGURE 5.6 **Integration of upregulated genes in wild-type (non-transformed or 'WT' setting) with upregulated genes in MLL-AF9 and AML1-ETO9a leukaemias occurring after *Ezh2* deletion**

Though previously noted in the upregulated list for non-transformed, it was only after analysis of the differential gene expression data for both leukaemias that demonstrated a very small number of genes, including *Plag1,* that were common to all three conditions. Plag1 is a transcription factor, and has been previously described as an oncogene which augments *Cbfb*-MYH11 driven leukaemias.[6,7] The other 5 genes common to all three groups were *Bmpr1a, Perp, Ptk7, Cacnb2,* and *Tjp2*, which appeared to have no functional significance in the literature as potential drivers of AML. *Lin28b,* and to a lesser extent *Yes1*, were commonly upregulated between the WT and MLL-AF9 *Ezh2*-null leukaemias, suggesting these may also be important mediators of the co-operation between *Ezh2* loss and leukaemia induction.

5.4 Loss of *Ezh2* in the non-transformed state results in a significant reduction in H3K27 trimethylation at multiple sites across the genome

Having characterised the effects of *Ezh2* loss at the gene expression level in the non-transformed setting, where a significant number of genes were upregulated (some of whom were putative candidates for facilitating MLL-AF9 and AML1-ETO9a leukaemia induction), it was then imperative to determine whether this degree of upregulation was mechanistically as a direct result of loss of canonical Ezh2 function. As discussed at length in the introduction, EZH2 is the core catalytic component of the PRC2 complex, and contains the requisite SET domain necessary to methylate the lysine 27 residue of histone H3. Ezh1 containing PRC2 complexes (in the absence of Ezh2) have been shown have little methyltransferase activity, and our expectation was to see a relatively significant loss of H3K27me3 following *Ezh2* deletion.

As described in Section 5.2, lineage marker negative murine HSPCs for both $Ezh2^{+/+}$ and $Ezh2^{-/-}$ genotypes were harvested and isolated and adequate *Ezh2* excision was confirmed. For each ChIP-seq experiment two mice i.e. biological replicates were used per condition. Following this, the cells were lysed and nuclei were crosslinked with formaldehyde to allow histone modifications and proteins interacting with DNA to be preserved. Chromatin immunoprecipitation (ChIP) for H3K27me3 was then performed and samples were prepared for high-throughput sequencing (as per Methods 2.8.6 to 2.8.12). Each sample was ratified for adequate enrichment through ChIP-qPCR (as detailed in Methods 2.8.9). Once the sequencing data was available it was analysed as per Methods 2.8.13, and a SICER based peak-calling method was used to call differentially bound H3K27me3 peaks. Peaks were called using a p-value of significance at $1x10^{-10}$ and there was reasonable concordance in total numbers of peaks called between biological duplicates for each condition (wt vs Cre+). Following the generation of this list, it was filtered to select all those peaks with a False Discovery Rate (FDR) of <0.01. They were then rearranged in order of fold change and only those with Fold changes $-1.5 < (x) > 1.5$ were selected. All intergenic peaks further than 100kb from the transcriptional start site (TSS) were removed, yielding a final list of peaks which were then referenced to genes, when comparing with RNA-seq changes.

Following *Ezh2* loss, there was a large number of H3K27me3 peaks whose differential binding was downregulated (n= 3821). In comparison, only a handful of peaks (n=4) were identified whose differential binding was significantly increased. The downregulated peaks were distributed primarily across intergenic and intron sites, with some also present at promoter and transcriptional start sites (TSS) (Figure 5.7):

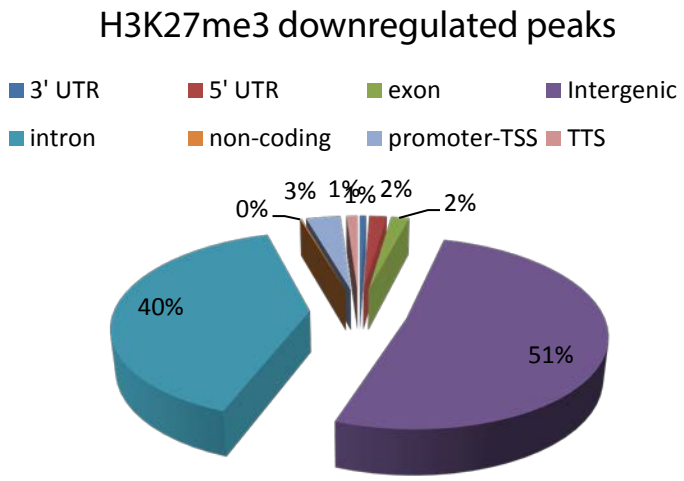


FIGURE 5.7 **Distribution of significantly downregulated H3K27me3 peaks following *Ezh2* deletion in lineage negative murine HSPCs (n=3821)**

5.5 Reduction of H3K27me3 occurs strictly at promoters/TSS of a narrow subset of all genes that are upregulated after *Ezh2* loss

Once aligned to the reference genome ('mm10' - the *Mus musculus* assembly (Genome Reference Consortium Mouse Build 38 [GCA_000001635.2])), the regions where H3K27me3 binding was significantly reduced following *Ezh2* deletion could be associated with 2706 genes. This list of genes was then overlapped with the list of 478 upregulated genes following *Ezh2* deletion (Appendix A5.1) to determine which were being directly upregulated as a result of H3K27me3 loss. Approximately 20% of all genes (91 of 478) upregulated on *Ezh2* loss were associated with a significant reduction in H3K27me3 (Figure 5.8):

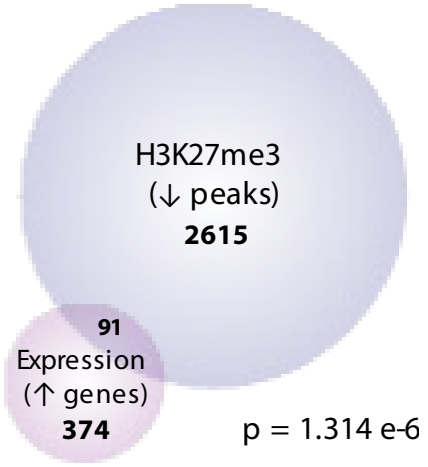


FIGURE 5.8 Overlap of all upregulated genes and genes at which H3K27me3 peaks were significantly downregulated following *Ezh2* loss

When the genomic location of these 91 overlapping regions was determined, unlike the global pattern, there was a striking enrichment for promoter/TSS regions (75 genes out of 91, i.e. 82%).

This overlap identified a narrow subset of 91 upregulated genes that were associated with a reduction in H3K27me3, suggesting that loss of Ezh2 function at their promoters/TSS regions directly contributed to their increased expression. This list of 91 genes is provided in Table 5.3:

2700081O15Rik	Cd3g	Dscam	Hdgfrp3	Myom3	Ppp4r4	Sntg2
6330403A02Rik	Cd5l	Dsp	Homer2	Ncam1	Prune2	Spats2
Adamts20	Cdkn2a	E330013P04Rik	**Igf2bp3**	Ndn	Rasal2	Stc2
Adcyap1r1	Cecr2	Elfn1	Jam3	Nkain1	Raver2	Syt13
Afp	Chic1	Enpep	Jazf1	Otoa	Reep1	Syt14
Ank2	Chst1	Eomes	Jph1	Pamr1	Ric3	Trpc1
Arhgap29	Cnnm1	Fam149a	Kank1	Pbx3	Rora	Unc13b
Atp8b5	Cplx2	Fjx1	**Lin28b**	Pcolce2	Rtn1	**Yes1**
Bend5	Cttnbp2	Fmn2	Lpl	Peg12	Slc16a2	Zdhhc2
Bex2	D030025P21Rik	Gca	Lrig3	Penk	Slc1a3	Zfp105
Bmpr1a	Dcbld2	Glis3	Mboat1	**Plag1**	Slc26a3	Zfp462
Cacna1e	Ddah1	Gzma	Mlph	Plekhh2	Slc2a1	Zfp9
Cd109	Dock3	H2afy2	Myo1e	Plxna2	Snhg11	Zswim5

TABLE 5.3 **Genes whose expression is significantly upregulated following *Ezh2* loss that are also associated with a significant reduction in H3K27me3 binding**

Within this subset of genes are some of the key candidates (highlighted in red) also present in the leukaemia gene expression datasets. It is interesting to note that around 80% of genes upregulated following *Ezh2* loss did not have an associated reduction in H3K27me3, suggesting alternative and possibly indirect mechanisms for upregulation. Additionally, the smaller subset of genes downregulated after *Ezh2* loss (n=48) had no overlap with changes in H3K27me3.

Gene ontology analysis of these 91 genes demonstrated an enrichment in metabolic, developmental and regulatory processes. Molecular functions predominantly included binding, catalytic and receptor activity (Figure 5.9). These data imply that in particular, Ezh2 enzymatic function is directly related to these processes and functions.

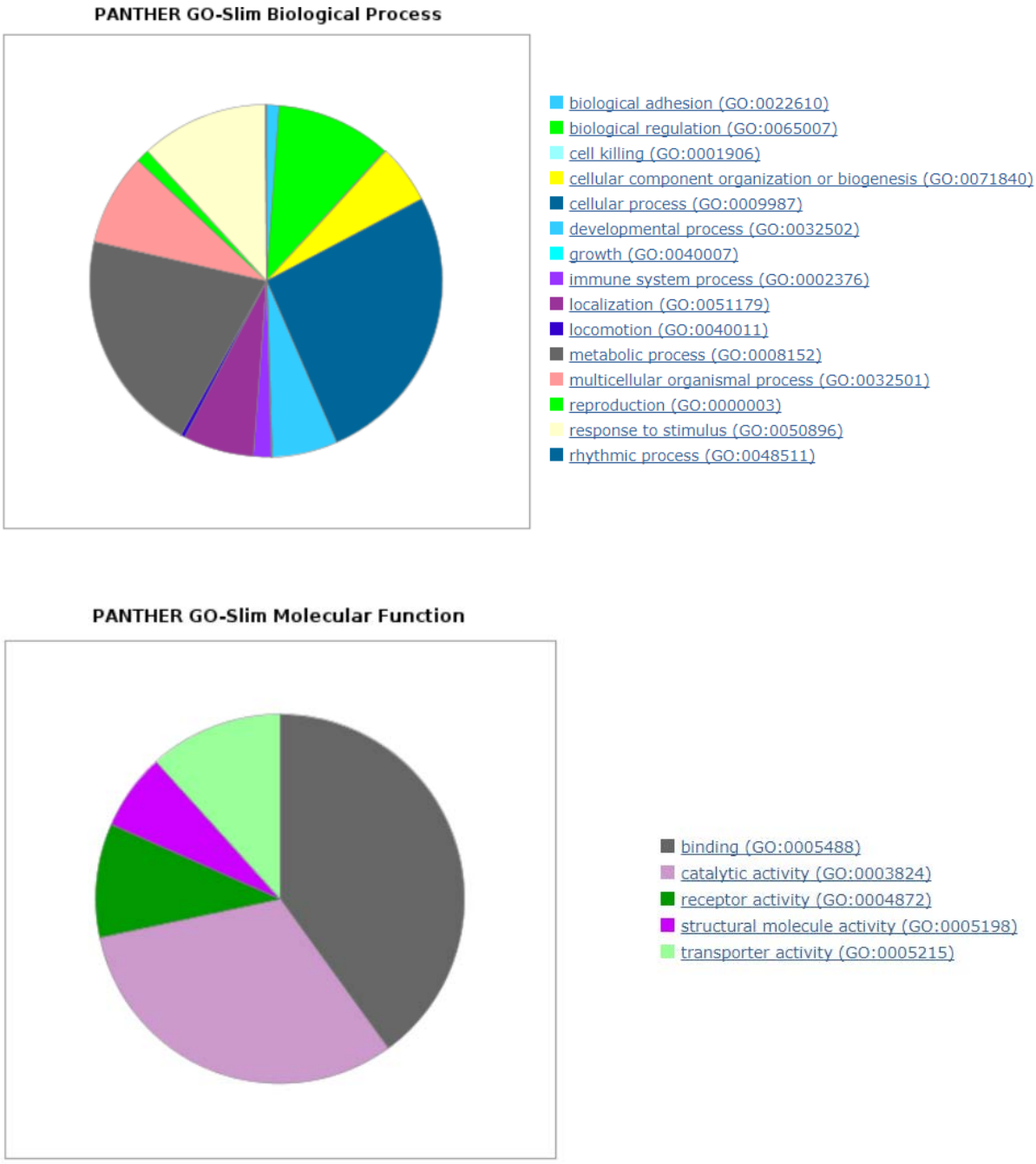


FIGURE 5.9 **Gene ontology analysis of genes upregulated via direct loss of H3K27me3 following *Ezh2* loss**
Enriched biological processes and molecular functions shown
Panther overrepresentation test (release Apr 2017)
GO Ontology database (release Aug 2017)

Visualisation of the H3K27me3 changes in a Genome Browser (UCSC Bioinformatics, University of California Santa Cruz) following *Ezh2* loss at these selected genes showed a clear loss of H3K27me3 peaks at the promoter regions (example in Figure 5.10):

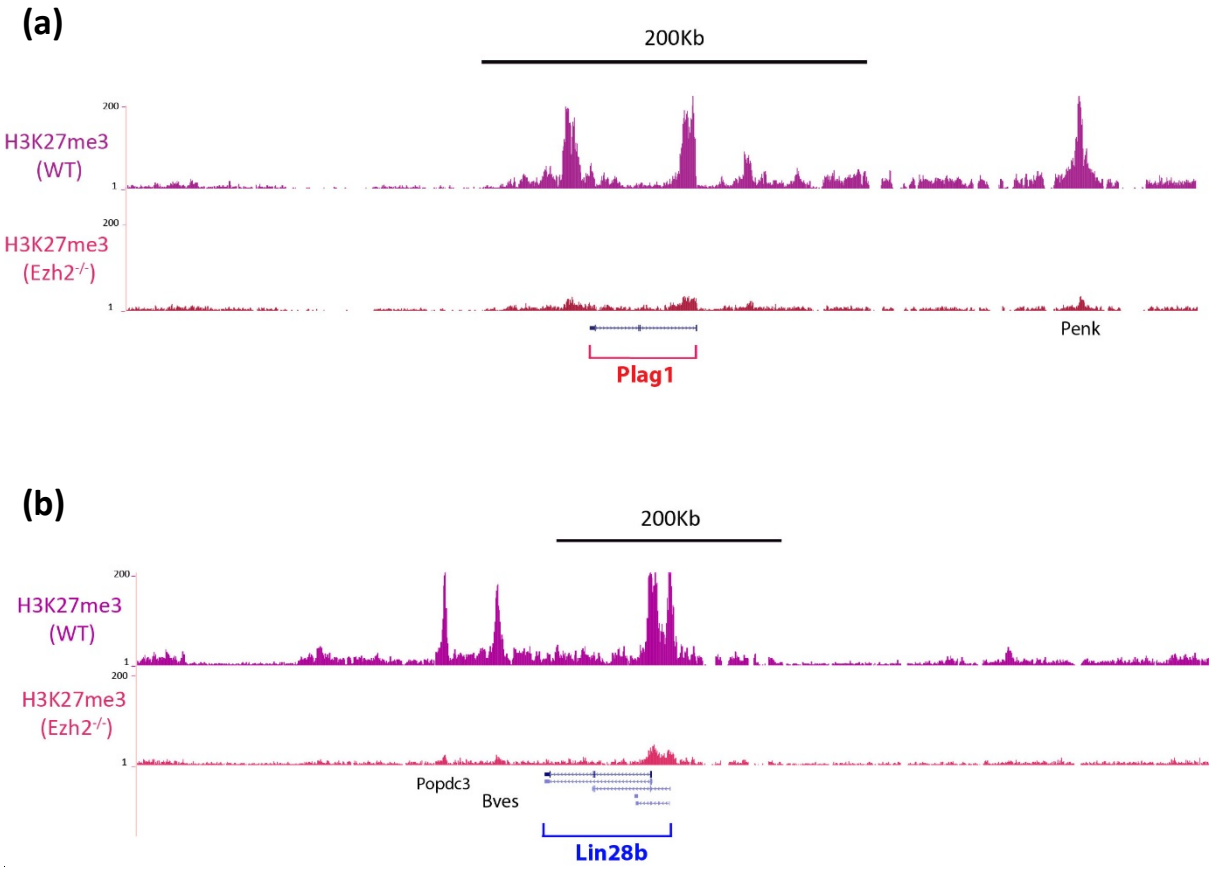


FIGURE 5.10 **Change in H3K27me3 peaks at (a) *Plag1* and (b) *Lin28b* genes before and after *Ezh2* loss**
Data presented in UCSC Genome Browser on Mouse Dec. 2011 (GRCm38/mm10) Assembly

For *Lin28b* the reduction in the H3K27me3 peak was 5.06 fold, whilst it was 1.95 fold for *Plag1*, both with adjusted p-values of $< 1x10^{-6}$. There were similar changes in H3K27me3 across all 91 genes selected in this overlap.

Having established that a number of these genes are upregulated in association with a loss of the H3K27me3 repressive mark, we rapidly moved to assess the presence of, and any alterations in potentially interrelated histone activating marks – H3K4me3 and H3K27Ac.

5.6 Analysis of H3K27Ac changes does not reveal any additional mechanistic insight into associated gene expression changes following *Ezh2* loss

ChIP-sequencing of the activating histone modification H3K27Ac was also performed on chromatin derived from lineage negative HSPCs isolated from murine bone marrow of pIpC pre-treated *Ezh2*$^{+/+}$; wt and *Ezh2*$^{-/-}$; Cre+ mice (in biological duplicates). The data was analysed as detailed above and a filtered list of differentially bound H3K27Ac peaks was generated using the same FDR and fold change criteria detailed above. Following the application of these filters, 846 genes were revealed to be differentially H3K27 acetylated. Twenty-four of these were associated with significant upregulation of H3K27Ac mainly at intergenic sites (Table 5.4), consistent with the fact that H3K27 acetylation primarily occurs at enhancer regions at some distance away from promoters or TSS. However, the majority - 822 genes had a significant downregulation.

Gene Name	Fold	p-value
Gm5150	2.44	1.3E-10
Fcgr3	2.41	1.7E-45
Xist	2.36	1.40E-53
Eps8l1	2.12	8.6E-15
Ccr1	2.09	3E-114
Lrrk2	1.85	4.5E-29
Atxn7l3b	1.83	4.4E-11
Ikbkg	1.75	1.07E-05
Zfp316	1.71	2.92E-06
Klhdc4	1.7	1.06E-08
Kctd16	1.69	8.91E-13
Eea1	1.69	0.00026
Slc4a11	1.68	0.00042
Mpp7	1.66	0.00225
Klhdc4	1.65	2E-05
Foxd4	1.63	2.71E-07
Pitx2	1.62	0.00054
9530068E07Rik	1.62	2.81E-06
F630028O10Rik	1.61	1.91E-25
Papss1	1.6	1E-30
Cybb	1.59	3.40E-10
Ldoc1	1.58	0.00115
1700017J07Rik	1.53	4.2E-08
Cd53	1.52	1.08E-28

TABLE 5.4 **Genes at which H3K27Ac is significantly upregulated following Ezh2 loss**
'Fold' represents log2fold change

This latter finding was somewhat surprising and unexpected. Overlaps between either up or downregulated H3K27Ac genes and genes increased in expression following *Ezh2* loss failed to show any specific correlation (for example there was no overlap in the genes identified in Table 5.4 and all genes upregulated after *Ezh2* loss). When a more targeted approach was adopted in scrutinising specific genes, such as some of the candidates discussed above, there was no differential change in H3K27Ac for either *Plag1* or *Yes1*. Counterintuitively, there appeared to be a reduction in H3K27Ac for *Lin28b* and *Igf2bp3*. As this particular dataset did not reveal a significant influence of changes in this histone modification on gene expression following *Ezh2* loss, no further analysis was undertaken.

5.7 Analysis of H3K4me3 localisation reveals bivalent domains that are subject to alteration following *Ezh2* loss, favouring transcriptional activation

As discussed earlier, the H3K4me3 mark is a histone modification associated with transcriptional activation. Its colocalization with H3K27me3 at promoters has been recognised in embryonic stem cells and this combination is widely considered to poise the expression of developmental genes, with its resolution in favour of H3K4me3 allowing rapid induction of differentiation programmes.[8] This then allows for timely activation, whilst maintaining repression in the absence of differentiation signals. In mammals, the SET1A, SET1B and mixed lineage leukaemia (MLL) proteins 1-4 are the enzymes responsible for laying the H3K4me3 mark, using additional subunits to form complexes for their activity.

Using an orthogonal dataset of H3K4me3 peaks across the genome in wild-type C57/Bl6 murine derived lineage negative HSPCs in duplicate, available in the Huntly lab (courtesy of Dr Haiyang Yun), we were able to gain a cursory glance at the presence or absence of H3K4me3 peaks at the previously identified H3K27me3 laden genes described in Section 5.4. Interestingly, of the 91 genes upregulated AND having a significant downregulation in H3K27me3 after *Ezh2* loss, 79 (~87%) also had a clear H3K4me3 peak from this dataset, at the same promoter. An example is shown in Figure 5.11:

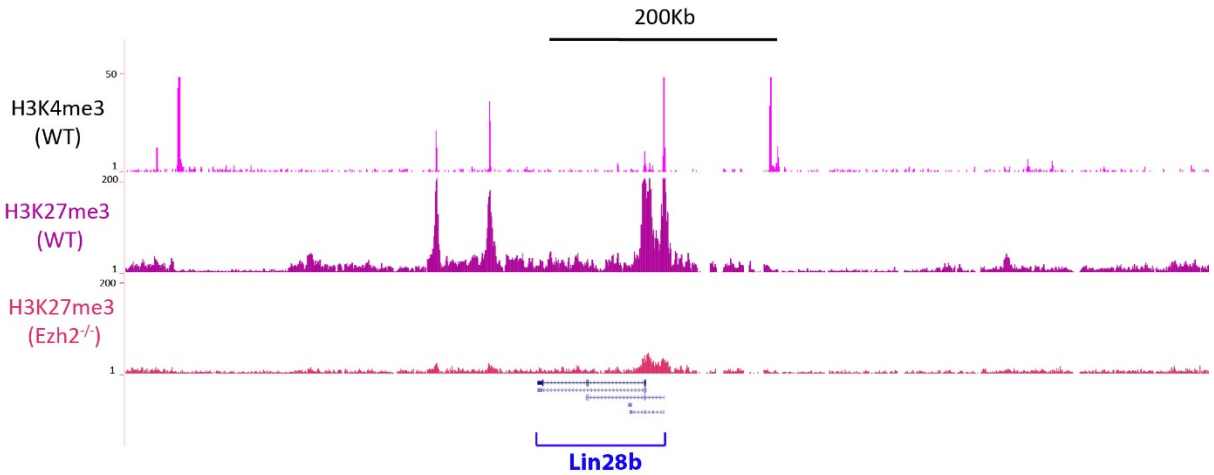


FIGURE 5.11 **H3K4me3 is present at the *Lin28b* promoter, also marked by H3K27me3**
Data presented in UCSC Genome Browser on Mouse Dec. 2011 (GRCm38/mm10) Assembly

This suggested bivalency (i.e. the presence of both an activating and repressive mark) at these genes, and the tantalising possibility that in the wild-type scenario at the HSPC level, these genes were being held in a poised transcriptional state by the presence of the conflicting histone marks. With the loss of *Ezh2*, causing a significant reduction of H3K27me3 at these gene promoters, we hypothesised that the balance would resolve towards transcriptional activation due to the presence of H3K4me3, facilitating increased gene expression.

Additionally, overlapping the remainder of the genes (n=2561) at which H3K27me3 was downregulated without any changes in gene expression with this H3K4me3 dataset showed that only 24% (n=562 genes), i.e. the minority had a co-existing H3K4me3 peak. This strongly suggested the possibility that the selected subset of 91 genes were being regulated by both H3K27me3 and H3K4me3 and loss of *Ezh2* was resulting in the resolution of a conflict between the opposing marks in favour of facilitating transcription.

In order to test the hypothesis made upon this observation, we proceeded to perform ChIP-sequencing of H3K4me3 (in biological duplicates) of both genotypes, using lineage negative HSPCs derived from pIpC pre-treated $Ezh2^{+/+}$ (wt) and $Ezh2^{-/-}$ (Cre+) mice (as per H3K27me3). The sequencing data was analysed in the same way as detailed above, generating a list of differentially bound H3K4me3 peaks. Peaks with a False Discovery Rate (FDR) of <0.01 and fold changes $-1.5 < (x) > 1.5$ were selected. In total, there were 75 genes at which significantly differential H3K4me3 peaks were seen between $Ezh2^{+/+}$ and $Ezh2^{-/-}$ states – 45 upregulated and 31 downregulated following *Ezh2* loss. Importantly, this gain of H3K4me3 at these genes included the promoter regions of *Lin28b* and *Plag1* (example shown in Figure 5.12):

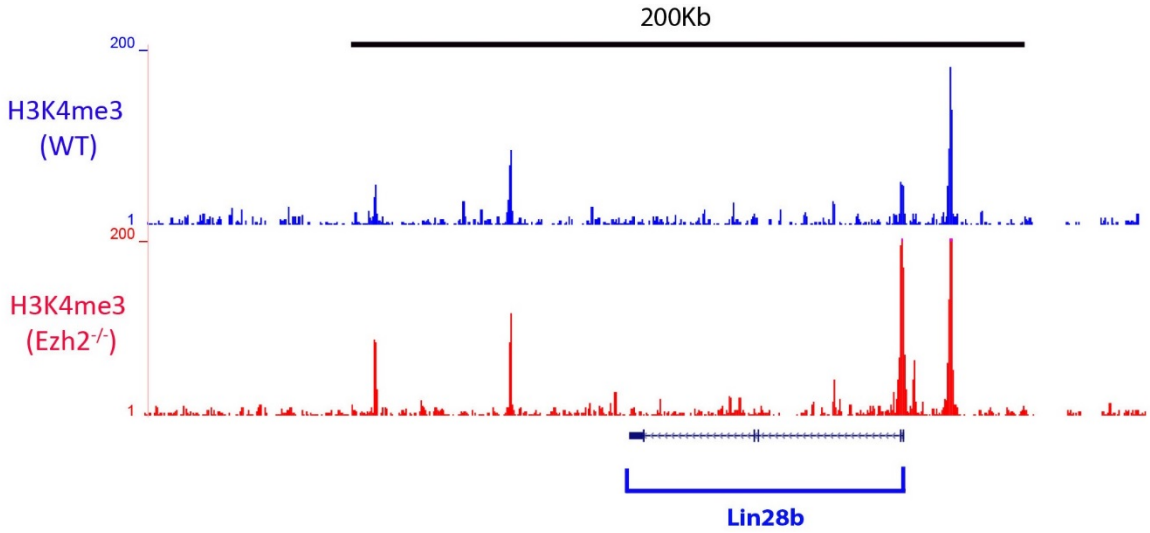


FIGURE 5.12 **H3K4me3 is upregulated at the *Lin28b* promoter following *Ezh2* loss**
Data presented in UCSC Genome Browser on Mouse Dec. 2011 (GRCm38/mm10) Assembly

The differentially upregulated H3K4me3 peaks corresponding to the 45 genes are listed in Table 5.5:

Gene Name	Fold	p-value
Cdk5rap1	7.77	1.19E-20
Rfx6	3.16	1.39E-08
5730412P04Rik	3.12	2.66E-06
Lin28b	2.96	2.06E-08
Plekha7	2.86	1.44E-07
A530006G24Rik	2.81	3.03E-05
Amph	2.62	0.000222
Dmrt1	2.45	6.24E-06
Nox4	2.4	4.23E-05
Plb1	2.29	0.000327
Susd5	2.26	6.60E-10
C2cd4b	2.2	1.93E-07
Ifi27l2a	2.2	5.42E-06
Pld5	2.13	2.38E-05
Krt1	2.08	6.90E-09
Hmgcll1	2.04	6.79E-07
Cdh15	1.99	6.88E-05
9430019J16Rik	1.98	0.000122
Dzip1l	1.94	7.21E-07
Mos	1.93	6.99E-09
Cnnm1	1.91	4.49E-30
Ttyh1	1.89	8.43E-05
Zfp641	1.88	1.52E-06
Bend3	1.87	3.49E-08
Flnb	1.85	1.02E-06
Tfap2c	1.85	0.000166
Peak1	1.84	0.000109
Plag1	1.84	0.000299
Rgs3	1.83	0.000155
Rgs22	1.82	0.000191
Wdr17	1.81	1.57E-06
Lamb1	1.8	0.000124
Wnt16	1.8	0.000149
Syt14	1.76	1.01E-18
Nrn1	1.75	0.000123
Sp7	1.74	1.97E-05
Shank1	1.72	1.60E-05
Echdc2	1.68	1.50E-08
Got2	1.67	0.000188
Sbspon	1.66	9.48E-06
Pde11a	1.62	3.33E-07
Lrig2	1.56	4.79E-05
Uggt2	1.54	9.73E-08

TABLE 5.5 **Genes at which H3K4me3 is significantly upregulated following *Ezh2* loss**
'Fold' is $\log_2$fold change, n = 45 genes

The overlap between differentially upregulated H3K4me3 genes and H3K27me3 downregulated genes that exhibited increased gene expression (n=91) was relatively small – only 4 genes were common to all three criteria – and these included *Lin28b* and *Plag1*.

However, a comparison between H3K4me3 peaks present at genes in the *Ezh2*$^{+/+}$ (wt) state and this selected subset of genes, revealed the majority (80 out of 91, i.e. 88%) at least had a recognisable H3K4me3 peak at the same promoter region as they had appeared to in the initial observation from the orthogonal dataset.

Combining these datasets suggested that *Ezh2* loss could in fact alter both H3K27me3 and H3K4me3 at a highly-selected set of upregulated genes leading to their increased expression. These included two candidate oncogenes, *Lin28b* and *Plag1,* which had both been identified as being significantly increased in expression in the non-transformed *Ezh2*-null state AND in *Ezh2*-null MLL-AF9 and AML1-ETO9a leukaemias. In addition, most of the upregulated genes that were associated with downregulated H3K27me3 peaks were also associated with the presence of a H3K4me3 mark at their promoters.

Taken together, these integrated data suggested that *Ezh2* loss would result in a resolution of conflict between these two competing histone modifications at the bivalent domains, favouring (or in the case of some important candidates, e.g. *Lin28b* and *Plag1* increasing) H3K4me3, providing an additional mechanistic explanation for why these genes were upregulated.

5.8 Assessment of PRC complex localisation in the absence of Ezh2

As discussed above, one of the other aims of this genomic characterisation was to delineate where Ezh2 itself was binding in the non-transformed state (in the *Ezh2*$^{+/+}$; wt condition) through ChIP-sequencing. In doing so, we posited seeing a substantial overlap where H3K27me3 was reduced following *Ezh2* loss, and an association with upregulated genes. Additionally, we sought to characterise PRC2 complex differential binding in the non-transformed state in the absence of Ezh2, through ChIP-seq profiling the binding of another core PRC2 complex member (Eed) in both *Ezh2*$^{+/+}$ and *Ezh2*$^{-/-}$ states.

Ezh2 binding

The same methods were used to obtain chromatin from lineage negative HSPCs isolated from murine bone marrow of pIpC pre-treated *Ezh2*$^{+/+}$; wt mice (in biological duplicates). In view of them being wild-type i.e. lacking *Mx1-Cre* expression, there was no excision expected and this was confirmed through qPCR genotyping before continuing. A total of $5x10^6$ cells were used per replicate with a longer 15-minute cross-linking time than for histones, given the relatively lower abundance of proteins when compared to the above histone modifications. Following this ChIP-sequencing of Ezh2 was performed and data analysed as already described. Peaks were called at a p-value of $1x10^{-5}$ (considerably less stringent than that used to call the histone modification peaks detailed above) due to the poorer quality of antibody binding and non-specific binding. The initial results showed an immediate problem between biological replicates themselves, with a marked difference in the total number of significant peaks called between identical duplicates (2900 peaks in replicate 1 and 4300 peaks in replicate 2). There was a limited overlap of 1280 genes between these two duplicates (<50%) and therefore it was felt that due to poor concordance between these two datasets (most likely for technical reasons such as poor specificity of the antibody) further analysis of this dataset was unlikely to reveal meaningful results.

Eed binding

As per Ezh2, $5x10^6$ lineage negative HSPCs per replicate were used for ChIP-sequencing to track Eed binding and PRC2 localisation both in the presence and absence of Ezh2. The cells were obtained as described above, from pIpC pre-treated *Ezh2*$^{+/+}$; wt and *Ezh2*$^{-/-}$; Cre+ mice (whose excision in the Cre+ arm had been confirmed prior to use). ChIP-qPCR for control positive and negative regions did appear to show enrichment as expected between the two conditions (data not shown). Following sequencing,

data was analysed as already described. Peaks were called at a p-value of $1x10^{-5}$. Unfortunately, (likely for the same technical reasons as described above), there was again a significant variance in numbers of peaks between biological replicates in both wt and Cre+ conditions (e.g. 1801 peaks for replicate 1 and 10191 peaks for replicate 2 in the Cre+ setting and 1741 peaks for replicate 1 and 7841 peaks for replicate 2 in the wt setting). Additionally, when visualised on a UCSC genome browser session, these peaks were of poor quality and not readily distinguishable from the baseline (example shown in Figure 5.13), with poor concordance between replicates. It was therefore decided not to attempt further analysis of this dataset.

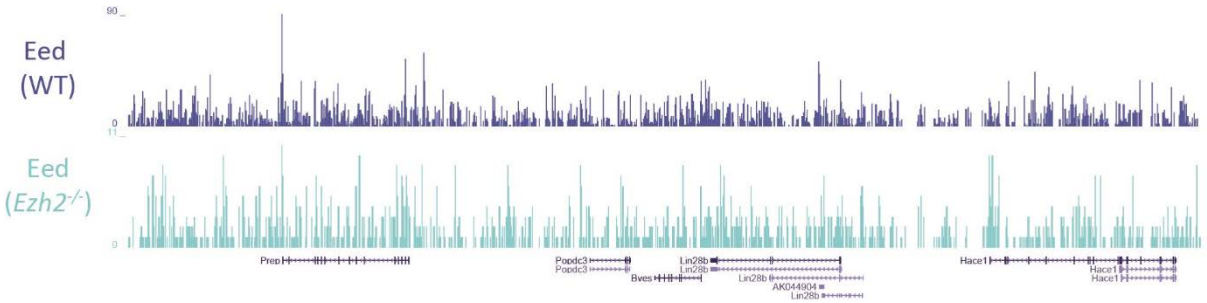


FIGURE 5.13 **Eed binding between *Ezh2*$^{+/+}$ and *Ezh2*$^{-/-}$ genotypes**
Data presented in UCSC Genome Browser on Mouse Dec. 2011 (GRCm38/mm10) Assembly
Lin28b is shown (centre-right)

Both these ChIP-sequencing datasets highlighted technical difficulties in obtaining adequate results when using antibodies against a protein marker as opposed to those against a histone modification. There was a reasonable degree of heterogeneity between biological replicates and even when the stringency of peak calling was relaxed (relative to that used for the histone modifications datasets above), it was still not possible to have a reliable filtered set of results for either Ezh2 or Eed binding that we could use to compared to the gene expression and histone modification datasets.

5.9 Analysis of differential gene expression resulting from pharmacological Ezh2 inhibition in MLL-AF9 murine leukaemias *in vitro* reveals an entirely different gene expression set involved in their maintenance

As discussed at the beginning of this chapter, the profoundly different functional outcomes of *Ezh2* loss in MLL-AF9 and AML1-ETO9a leukaemia induction versus their maintenance led us to hypothesise that different transcriptional programmes were being altered by *Ezh2* loss, or that, at the very least, there would be minimal overlap between these two programmes. Having compared the transcriptional and epigenetic differences between a 100% *Ezh2* replete (wt) and 95%+ *Ezh2* deleted (Cre+) scenarios through genetic excision in the induction *in vitro* experiments, we felt it would be an unfair comparison to any differential gene expression analysis to compare the partially *Ezh2* excised 'wild-type' leukaemias and their completely *Ezh2* deleted counterparts in the maintenance experiments. Differences at the genomic level would likely be diluted by the relative lack of genotype extremes (as the induction experiments had cleanly provided) in the maintenance *in vivo* experiments, due to spontaneous Cre-mediated *Ezh2* excision.

We therefore sought a different approach to characterise the altered transcriptional programmes that were possibly leading to abrogation of the leukaemias following complete *Ezh2* excision. Instead of further genetic manipulation, we employed pharmacological inhibition of Ezh2 in established leukaemia tumours to produce a more pronounced Ezh2 deficiency that we could compare to vehicle control treatment alone. As MLL-AF9 tumours tend to proliferate more aggressively in liquid culture than AML1-ETO9a tumours do, we decided to conduct this experiment utilising the former. After thawing and washing spleen cells from two separate *Ezh2fl/fl*; wt MLL-AF9 primary leukaemia mice whose tissues had been processed and stored in liquid nitrogen during the induction *in vivo* experiment, we were able to culture and expand these reliably and consistently *in vitro* using sterile X-Vivo 20 culture medium, as they were already established leukaemias (as described in Methods 2.2.2). Full details of the following are provided in Chapter 7, but briefly, both spleen tumours showed an almost equal sensitivity to GSK343 (dissolved in DMSO as vehicle) in liquid culture *in vitro*, with an inhibitory effect on liquid culture growth as early as 24 hours and IC_{50} inhibition as low as concentrations of 10μM. Having established sensitivity, both tumours were treated with DMSO (Control) or GSK343 in liquid culture and cells harvested at every 24 hours for up to 72 hours to assess reduction in H3K27me3 by flow cytometry (Methods 2.4.2). Biological and technical replicates demonstrated that a near total loss of H3K27me3 was seen in each tumour as early as 24 hours of

treatment with GSK343 (Figure 5.14). Therefore, this time point was chosen and cells were harvested from both tumours in liquid culture at 24 hours from both DMSO and GSK343 treated conditions.

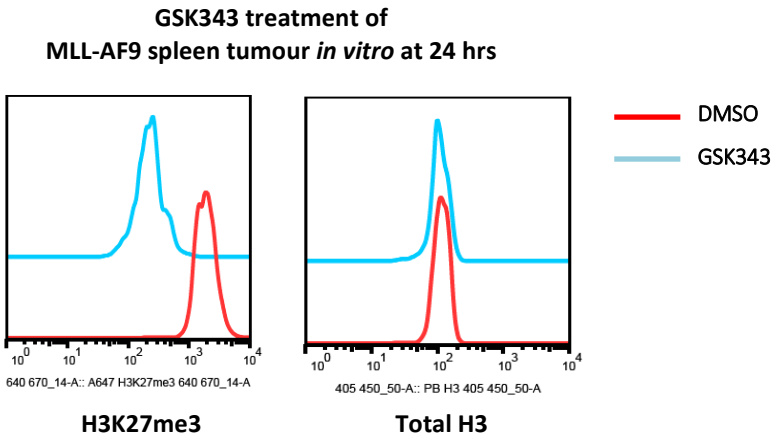


FIGURE 5.14 **Loss of H3K27me3 at 24hours following Ezh2 inhibition with GSK343 in MLL-AF9 spleen tumour #5340 *in vitro***
Comparisons made to DMSO (vehicle control treated) cells and Total Histone H3 of the same cells

RNA was extracted from these cells and after quantification, 5µg from each sample was sent for library preparation and sequencing (Methods 2.8.15). Through this, a filtered list of differentially expressed genes of significance between DMSO and GSK343 treated tumours was generated. Significant gene expression results were those with a p-value adjusted for multiple hypothesis testing of <0.05 with a $\log_2$ fold change of greater or less than 0.5. With the application of these criteria, a total 91 genes were differentially regulated – considerably less than in the MLL-AF9 induction programme. Of these, likely as a result of Ezh2 inhibition, 79 genes were significantly upregulated. These are listed in Table 5.6:

GeneName	log_2FoldChange	padj			
Gm15907	4.8379	0.030197352	continued		
Cth	4.2341	1.73E-09	Stc2	1.2234	0.00018
Fstl1	3.6056	3.41E-06	Mib2	1.2071	0.00041
Extl1	3.3829	2.13E-16	Cd33	1.2008	0.03199
Aldh1l2	3.3746	4.94E-26	Zbtb4	1.1954	0.01159
Trib3	2.6706	2.88E-05	Mfhas1	1.1908	0.01377
Zfhx3	2.5334	0.000105543	Npc1	1.1858	0.00027
Ccl3	2.4599	0.000105452	Grn	1.1751	0.00192
Chac1	2.3940	0.000278152	Gpt2	1.1718	0.00104
Tnks1bp1	2.2944	0.005801715	Klhdc1	1.1710	0.03292
Slc7a7	2.2733	0.019959395	Mcfd2	1.1694	0.03292
Cd28	2.2643	3.68E-12	Ppp3cc	1.1469	0.01996
Gas6	2.2169	0.019959395	Parp12	1.1419	0.02567
Wfdc17	2.2051	2.71E-13	Dgkg	1.1390	0.00501
Pdpn	2.1102	0.021463925	Hipk2	1.1187	0.00926
Fam129a	2.0788	4.13E-08	Scpep1	1.1171	0.01996
Atf5	1.9756	4.13E-08	Grina	1.1108	0.00144
Oas3	1.8819	3.41E-06	Trpm2	1.0988	0.00636
Slc26a11	1.7986	0.007793617	Bcar3	1.0974	0.03032
Itgb3	1.7881	1.27E-06	Hmox1	1.0962	0.00598
Acot2	1.7761	0.005844631	Plekho1	1.0886	0.02104
Nupr1	1.7188	1.16E-07	Guca1a	1.0826	0.02496
Arrdc3	1.6459	0.004928	St3gal1	1.0788	0.00584
Clic5	1.6414	0.005712155	Kdelr3	1.0777	0.02567
Pycr1	1.6190	3.41E-06	Ttc12	1.0662	0.01907
Pck2	1.5974	1.88E-08	Sesn2	1.0656	0.00356
Gadd45a	1.5746	7.09E-05	Mcfd2	1.0540	0.02226
Xrcc5	1.5383	0.003561327	Camk1	1.0433	0.00571
Tbc1d16	1.4587	0.005837158	Clic4	1.0220	0.00507
Cdr2l	1.4492	1.85E-05	Mt1	1.0097	0.00508
Thbs1	1.4216	0.006314509	Osm	1.0086	0.03292
Ptms	1.3354	0.022609809	Timp2	1.0033	0.02104
Fcgr3	1.3201	0.000278152	Malat1	1.0029	0.00584
Hacd4	1.2904	0.000512542	Crebrf	1.0008	0.01996
Ankrd63	1.2799	0.000278152	Ddx58	0.9434	0.02904
Grcc10	1.2792	0.001039787	Tcp11l1	0.9411	0.03292
Ctsh	1.2525	0.045022393	Gns	0.9315	0.01996
Cd200r1	1.2504	0.000593059	Scarb2	0.9225	0.03199
Arl4c	1.2290	0.017273765	Eif4ebp1	0.8655	0.03842
Mir22hg	1.2281	0.009120184	B2m	0.8552	0.04589

TABLE 5.6 **Genes upregulated following Ezh2 inhibition with GSK343 in established MLL-AF9 murine leukaemia**

'padj' is adjusted p-value for multiple hypothesis testing

Albeit that there are significant differences between genetic loss and reversible inhibition of Ezh2, this set of de-repressed genes involved in the maintenance of these MLL-AF9 leukaemias is almost mutually exclusive with the de-repressed gene set seen in the both the non-transformed and MLL-AF9 leukaemia induction *in vivo* setting. Overlapping these datasets revealed very few common candidates, none of which had immediate obvious relevance to either phase of AML evolution (Figure 5.15):

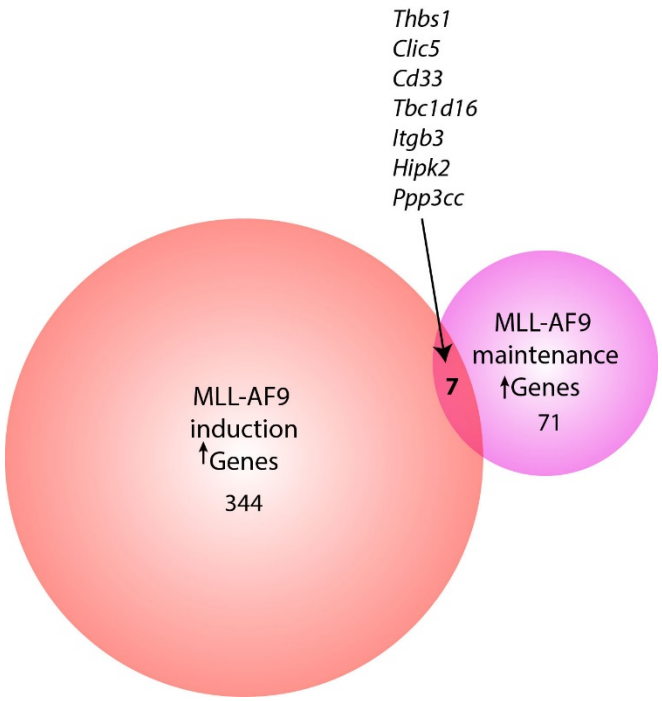


FIGURE 5.15 **Overlap of genes de-repressed following *Ezh2* loss in MLL-AF9 leukaemia induction and genes de-repressed following Ezh2 inhibition in MLL-AF9 leukaemia maintenance**

This analysis suggested that the significantly opposing functional differences seen in the *in vivo* induction and maintenance experiments subsequent to *Ezh2* loss could be explained by the de-repression of a completely different gene expression set in either cellular context. At the time of submission, further work such as gene set enrichment analysis on these datasets is being performed to see if there is any alignment with specific cellular programmes.

5.10 Discussion and future work

The genomic analysis described in this chapter served to provide a better mechanistic understanding of the different outcomes seen in leukaemia induction and maintenance, particularly in the previously described *in vivo* experiments. It also allowed for the identification of certain genes that were de-repressed following *Ezh2* loss that could potentially co-operate with MLL-AF9 and AML1-ETO in accelerating leukaemia induction.

Through comparison of the non-transformed state in both the presence and absence of *Ezh2*, a detailed characterisation of differential gene expression and relevant histone modifications was made. When these datasets were compared, it was apparent that a number of genes that were de-repressed following *Ezh2* loss were associated with a significant reduction in H3K27me3 (the direct output of Ezh2 enzymatic function). Interestingly, a number of these genes were also populated with the activating H3K4me3 mark at their promoters (79 out of 91), suggesting bivalency and a possible shift in balance towards transcriptional activation due to the resolution of the conflict between these two histone marks in favour of H3K4me3 upon *Ezh2* loss. Strikingly, a few select candidate genes (*Lin28b* and *Plag1* amongst others) also demonstrated a significant increase in H3K4me3 following *Ezh2* loss, potentially explaining why they were so significantly upregulated. It is highly interesting that *Ezh2* loss does not only downregulate H3K27me3, but also in some way is modulating H3K4me3 upregulation, particularly at a limited set of genes where the level of upregulation is incredibly marked. A recent study characterising the role of the histone demethylase UTX (responsible for demethylation of the H3K27me3 mark – effectively the opposite role of EZH2) in the epigenetic regulation of invariant natural killer T cells, identified that UTX knock-out led to the downregulation of a number of genes. Genes that were most downregulated in the UTX knock-out scenario were revealed to have an accumulation of H3K27me3 at their promoters (consistent with repression), but also *a marked decrease* in H3K4me3 when both histone marks were assessed by ChIP-sequencing.[9] The deficiency in UTX appeared to affect the chromatin state around a particular subset of downregulated genes. Whilst a completely separate enzyme with canonically the opposite function to EZH2, UTX loss in this study appears to be demonstrating a clear effect on H3K4me3. The common link between fluxes in EZH2 and UTX status is centred around changes in H3K27me3. Alterations in H3K27me3 resulting from changes in these two, particularly at the promoters of certain subsets of genes (both in this study and in our data above), appear to directly influence significant changes in their H3K4me3 state (in the opposite direction), providing a valid mechanism for the changes seen in transcriptional output. However, the mechanisms that underlie this are not entirely clear and would be very interesting to further

characterise. Taken together and particularly supported by the opposing effects of EZH2 and UTX and their diametrically opposite effects on the H3K4me3 mark at highly specific genes, this suggests a significant and active 'cross-talk' between H3K27me3 and H3K4me3 in gene regulation.

Having compared the gene expression changes resulting from *Ezh2* loss in the non-transformed setting to those seen in MLL-AF9 and AML1-ETO9a during leukaemia induction, further evidence was provided for some of these candidate genes facilitating a functional role in accelerating leukaemogenesis. We hypothesised that early *Ezh2* loss in the non-transformed state led to the de-repression of a number of genes, of which *Lin28b* and *Plag1* were plausibly at the forefront in co-operating with MLL-AF9 and/or AML1-ETO9a to accelerate leukaemia induction. *Lin28b* was maximally upregulated in both the non-transformed and MLL-AF9 gene expression sets (appearing at the top of both lists in order of $\log_2$fold change), and with its general role in tumour induction and recent implication as a mediator of *Ezh2* loss synergising with *JAK2*-V617F myeloproliferative neoplasms[3] warranted further inquiry into its functional role in AML induction.

The evidence implicating *Plag1* as a facilitator of *Ezh2* loss in AML induction was potentially even stronger. Despite it not appearing in the top 25 upregulated genes for each of the three conditions, it was the only known oncogene common to all three groups, and has also been implicated in the accelerated development of *Cbfb-Myh11* murine AMLs.[4,5] Both candidates had a reduction in H3K27me3 and increase in H3K4me3 to explain the mechanism behind their upregulation. Having established these as potential candidates, the next step was to attempt to functionally validate these through overexpression of either of these genes in tandem with MLL-AF9 or AML1-ETO9a, ideally *in vivo*, with the aim of phenocopying Ezh2 loss to accelerate leukaemia induction.

Finally, we identified a largely different transcriptional programme altered by *Ezh2* loss to potentially explain the opposite and negative effects seen in the maintenance *in vivo* MLL-AF9 leukaemia experiments, albeit that we utilised reversible inhibition of Ezh2 rather than genetic ablation. It is also arguable that the lack of overlap seen between the gene sets is in part due to experimental differences (genetic versus pharmacological inhibition) and that the two are not directly comparable. However, we maintain that this is a better comparison than that between partially *Ezh2*-excised versus completely *Ezh2*-excised states.

The induction experiments had brought up valid concerns that use of EZH2 inhibitors mandated caution (also exhibited in the myeloproliferative studies), considering that *Ezh2* loss in the murine experiments was clearly accelerating leukaemia development. However, this latter characterisation of genes altered in maintenance has allayed these fears largely because they are almost mutually

exclusive gene sets. For the maintenance phase, these findings reassured regarding the use of EZH2 inhibitors, especially as they are currently in clinical trials, and the next step then was to test GSK343 further in murine and human leukaemia cell lines and in primary AML samples.

References

1) Duployez N, Marceau-Renaut A, Boissel N, Petit A, Bucci M, Geffroy S, Lapillonne H, Renneville A, Ragu C, Figeac M, Celli-Lebras K, Lacombe C, Micol JB, Abdel-Wahab O, Cornillet P1, Ifrah N, Dombret H, Leverger G, Jourdan E, Preudhomme C. Comprehensive mutational profiling of core binding factor acute myeloid leukemia. *Blood*. 2016 May 19;127(20):2451-9

2) Faber ZJ, Chen X, Gedman AL, Boggs K, Cheng J, Ma J, Radtke I, Chao JR, Walsh MP, Song G, Andersson AK, Dang J, Dong L, Liu Y, Huether R, Cai Z, Mulder H, Wu G, Edmonson M, Rusch M, Qu C, Li Y, Vadodaria B, Wang J, Hedlund E, Cao X, Yergeau D, Nakitandwe J, Pounds SB, Shurtleff S, Fulton RS, Fulton LL, Easton J, Parganas E, Pui CH, Rubnitz JE, Ding L, Mardis ER, Wilson RK, Gruber TA, Mullighan CG, Schlenk RF, Paschka P, Döhner K, Döhner H, Bullinger L, Zhang J, Klco JM, Downing JR. The genomic landscape of core-binding factor acute myeloid leukemias. *Nat Genet.* 2016 Dec;48(12):1551-1556

3) Shimizu T, Kubovcakova L, Nienhold R, Zmajkovic J, Meyer SC, Hao-Shen H, Geier F, Dirnhofer S, Guglielmelli P, Vannucchi AM, Feenstra JD, Kralovics R, Orkin SH, Skoda RC. Loss of *Ezh2* synergizes with JAK2-V617F in initiating myeloproliferative neoplasms and promoting myelofibrosis. *J Exp Med.* 2016 Jul 25;213(8):1479-96.

4) Lederer M, Bley N, Schleifer C, Hüttelmaier S. The role of the oncofetal IGF2 mRNA-binding protein 3 (IGF2BP3) in cancer. *Semin Cancer Biol.* 2014 Dec; 29:3-12.

5) Vegi NM, Klappacher J, Oswald F, Mulaw MA, Mandoli A, Thiel VN, Bamezai S, Feder K, Martens JH, Rawat VP, Mandal T, Quintanilla-Martinez L, Spiekermann K, Hiddemann W, Döhner K, Döhner H, Stunnenberg HG, Feuring-Buske M, Buske C. MEIS2 Is an Oncogenic Partner in AML1-ETO-Positive AML. *Cell Reports.* 2016 Jul 12;16(2):498-507

6) Landrette SF, Kuo YH, Hensen K, Barjesteh van Waalwijk van Doorn-Khosrovani S, Perrat PN, Van de Ven WJ, Delwel R, Castilla LH. *Plag1* and *Plagl2* are oncogenes that induce acute myeloid leukemia in cooperation with Cbfb-MYH11. *Blood*. 2005 Apr 1;105(7):2900-7

7) Castilla LH, Perrat P, Martinez NJ, et al. Identification of genes that synergize with *Cbfb*-MYH11 in the pathogenesis of acute myeloid leukemia. *Proc Natl Acad Sci USA*. 2004 ;101:4924-4929.

8) Bernstein BE, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*. 2006; 125:315–326.

9) Beyaz S, Kim JH, Pinello L, Xifaras ME, Hu Y, Huang J, Kerenyi MA, Das PP, Barnitz RA, Herault A, Dogum R, Haining WN, Yilmaz ÖH, Passegue E, Yuan GC, Orkin SH, Winau F. The histone demethylase UTX regulates the lineage-specific epigenetic program of invariant natural killer T cells. *Nat Immunol.* 2017 Feb;18(2):184-195.

Chapter 6

Functional analysis of candidate genes identified by genomic analysis

6.1 Aims

This chapter addresses the following question:

1. Does overexpression of *Lin28b* or *Plag1* phenocopy loss of *Ezh2* in accelerating MLL-AF9 leukaemia induction *in vivo*?

6.2 Introduction

LIN28

Lin28 was initially characterised in the *Caenorhabditis elegans* nematode as being an important regulator of developmental timing[1], and is highly conserved across evolution. Its mammalian homologues *Lin28a* and *Lin28b* are RNA-binding proteins that bind to the terminal loops of the let-7 microRNA tumour-suppressor family and prevent their post-translational processing into functional mature miRNAs.[2] The let-7 microRNAs in turn negatively regulate *LIN28B* expression as well as many other genes such as *HMGA2*, *c-Myc* and *K-Ras*, normally promoting cell differentiation. [3,4] LIN28 has been described as an important reprogramming factor in concert with KLF4, SOX2 and NANOG during the induction of pluripotency in adult human fibroblast cells. [5] In murine haematopoiesis, the Lin28b protein, let-7 and its target gene *Hmga2* have been identified as important determinants of enhanced self-renewal characteristics that foetal haematopoietic stem cells (HSCs) display. [6] Reintroduction of *Lin28b* into purified adult HSCs resulted in a nearly 10-fold greater numerical expansion and increased lineage reconstitution following transplantation into recipient mice, suggesting the reacquisition of foetal-like self-renewal capacity. The overexpression of *Lin28b* in adult HSCs reactivated *Hmga2*, a key let-7 miRNA target. Hmga2 has been demonstrated to regulate growth and metabolism pathways through important regulators such as insulin-like growth factor (IGF) and the oncogenic Myc pathway and is well described as a downstream mediator of oncogenic function associated with the Lin28b-let-7 pathway.[7] LIN28 and LIN28b have been shown to be overexpressed in a number of primary human tumours and human cancer cell lines and this overexpression was linked to de-repression of let-7 targets. [8] Elevated LIN28/LIN28B expression across many cancer types (including advanced breast, lung and liver) correlate with poorer prognosis and clinical outcomes. [9]

Recently, Ezh2 was characterised as a regulator of the *Lin28b-let-7* pathway in adult murine HSCs through H3K27me3, with its loss resulting in the re-activation of foetal-like gene signatures, through ectopic expression of *Lin28b*. [10] The suggestion here was that Ezh2 co-operated with let-7 microRNAs to silence foetal-like gene signatures in adult bone marrow HSCs, potentially restricting their propensity to transformation.

Ectopic expression of Lin28b in a transgenic mouse model on a C57/Bl6 strain (via a HS 21/45-vav vector for targeted expression in haematopoietic tissues) resulted in aberrant T-cell development and invariably led to the development of aggressive peripheral T-cell lymphoma with associated systemic inflammatory changes, with a latency as early as 10 months. [11] In the recent study describing *Ezh2* loss cooperating with *JAK2*-V617F to drive myelofibrosis, the *Lin28b-let-7-Hmga2* axis was suggested as a key candidate pathway mediating this process. [12]

PLAG1

The PLAG family comprises 3 members (PLAG1, PLAGL1 and PLAGL2), all of which function as transcription factors. They are highly conserved in structure and function and the PLAG1 gene (pleomorphic adenoma gene) has been associated with benign tumours harbouring 8q12 translocations including pleomorphic adenomas of the salivary glands, lipoblastomas and hepatoblastomas. [13,14] *Plag1* was identified (through a retroviral insertional mutagenesis screen) as a potential co-operative gene with *Cbfb-MYH11* leukaemias *in vivo*. [15] In a successive study by the same group, co-expression of *Plag1* and *Cbfb-MYH11* in a bone marrow transduction and transplantation assay caused the development of AML in 100% of recipient mice, with a latency of 3-12 weeks, whilst expression of *Plag1* or *Cbfb-MYH11* alone was insufficient to cause leukaemia. [16] *In vitro, Plag1* overexpression also caused expansion of murine haematopoietic progenitors. In the same study, 5% of primary AML samples had significantly elevated levels of the *PLAG1* transcript across all cytogenetic groups, suggesting PLAG1's promiscuity in co-operating with other mutations associated with AML.

The gene expression and epigenetic analysis of the induction *in vivo* leukaemia experiments documented thus far revealed several plausible downstream effectors of *Ezh2* loss whose de-repression was potentially capable of augmenting MLL-AF9 or AML1ETO9a leukaemogenesis. After detailed genomic analysis the most striking candidates were *Lin28b* and *Plag1* (for the reasons discussed in Chapter 5), and coupled with clearly supportive data from other studies in the literature

as discussed above, we felt these two candidates specifically warranted further functional investigation.

6.3 Functional assessment of a retroviral overexpression system for *Lin28b* and *Plag1*

The retroviral vectors MSCV-*Lin28b*-IRES-mCherry and MSCV-*Plag1*-IRES-mCherry were generated for this experiment through cloning each insert into the MSCV-IRES-mCherry vector as described in Methods 2.7. Having established these, the aim was to utilise them to overexpress each of the putative co-operating oncogenes with MLL-AF9 in c-kit selected HSPCs harvested from the bone marrow of wild-type C57/Bl6 mice. These transduced cells would then be immediately transplanted into irradiated wild-type C57/Bl6 recipients to determine whether either would co-operate to cause acceleration of MLL-AF9 induction *in vivo*, thus phenocopying the loss of *Ezh2*. This would be compared against a single transduced MLL-AF9 control arm. The MLL-AF9 leukaemia model was chosen to assess these two candidate genes over AML1-ETO9a as its latency on the bone marrow transplantation experiments described in Chapter 3 described was significantly shorter (6-10 weeks), to enable a faster readout. This was balanced against the potential pitfall that the MLL-AF9 leukaemia programme is so aggressive that it risked overwhelming or diluting any additional co-operative function we were aiming to demonstrate through the overexpression of either of these single candidate genes.

Prior to commencing the functional *in vivo* transplantation experiment, it was imperative to assess the function of each newly generated construct. Firstly, retroviral supernatant was generated from dual transfection of 293T cells with equal amounts of MSCV-MLL-AF9-IRES-YFP and either a) MSCV-*Plag1*-IRES-mCherry or b) MSCV-*Lin28b*-IRES-mCherry (Methods 2.6.6.1). To confirm co-expression of YFP and mCherry within 293T cells (indicating successful dual transfection), these were assessed by flow cytometry and compared to single MSCV-MLL-AF9-IRES-YFP transfected 293T cells (Figure 6.1)

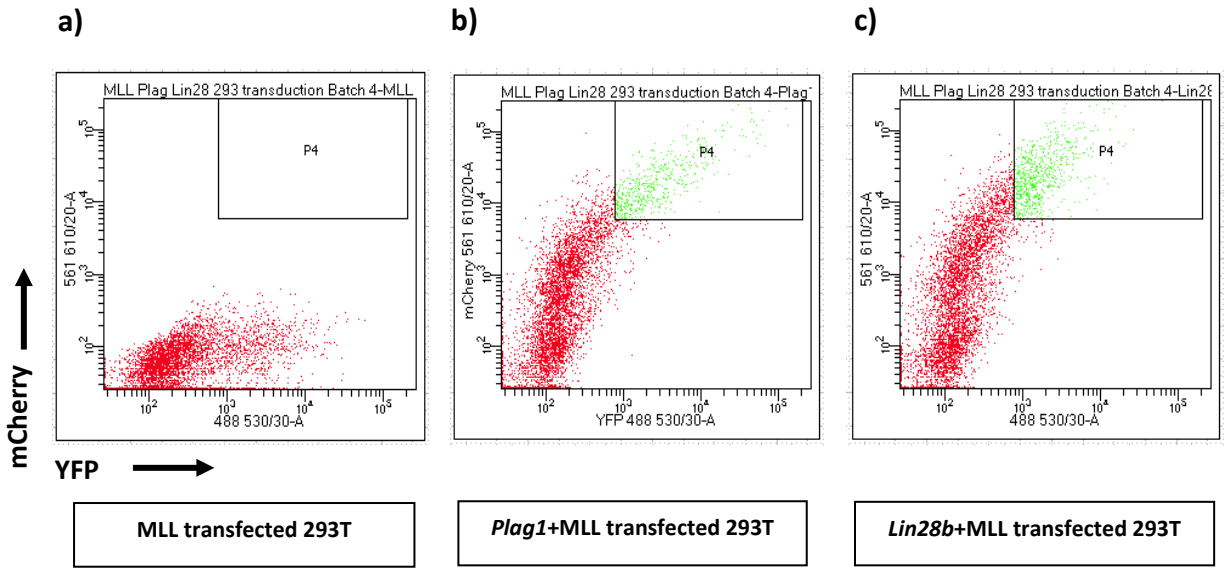


FIGURE 6.1 **Flow cytometry of (a) single (MLL-AF9), and (b) or (c) dual (*Plag1*+MLL or *Lin28b*+MLL) transfected 293T cells**

Live cells firstly gated by forward and side-scatter, doublets then removed (not shown). YFP detected on 488 530/30A channel (x-axis), mCherry detected on 561 610/20A channel (y-axis)
a) 293T cells transfected with MSCV-MLL-AF9-IRES-YFP alone
b) 293T cells transfected with MSCV-MLL-AF9-IRES-YFP AND MSCV-*Plag1*-IRES-mCherry
c) 293T cells transfected with MSCV-MLL-AF9-IRES-YFP AND MSCV-*Lin28b*-IRES-mCherry

Having demonstrated the presence of dual transfected 293T cells in similar proportions for both MLL-AF9+ *Plag1*- or *Lin28b*- conditions, and importantly that it did not deviate in favour of singly transfected MLL-AF9 cells, the next step was to determine whether the retroviral supernatant produced from each would provide appropriate transduction. Retroviral transduction was performed as detailed in Methods 2.6.7.2. The expectation was that transduction of c-kit positive bone marrow cells with supernatant from conditions (b) and (c) in Figure 6.1 above, would result in four populations: (1) non-transduced cells, (2) single MLL-AF9 (YFP+) transduced cells, (3) single *Plag1*- or *Lin28b*- (mCherry+) transduced cells and (4) double transduced MLL-AF9 + *Plag1* OR *Lin28b* (dual YFP/mCherry positive) cells.

Ideally, the double transduced population (4) would form the majority of the sum of (2), (3) and (4) and would allow for a fair functional assessment of overexpression of either candidate gene in the presence of MLL-AF9.

Given the potency of the MLL-AF9 oncogene, a major concern was whether there would be the presence of a significant sized population (2), i.e. single MLL-AF9 (YFP+) transduced cells. If present, alone these could potentially override any additive functional effect either candidate gene may

demonstrate. As the experimental design was to transplant bulk transduced cells and not flow sort population (4) for each condition (to avoid further stress to the cells being transplanted), this concern required addressing before any *in vivo* transplants utilising these could be set up.

To test the retroviral supernatants, mock transductions were performed with these in duplicate (Methods 2.6.7.2) utilising c-kit selected HSPCs isolated from the bone marrow of 2 wild-type C57/Bl6 recipient mice. A representative example of this is shown in Figure 6.2:

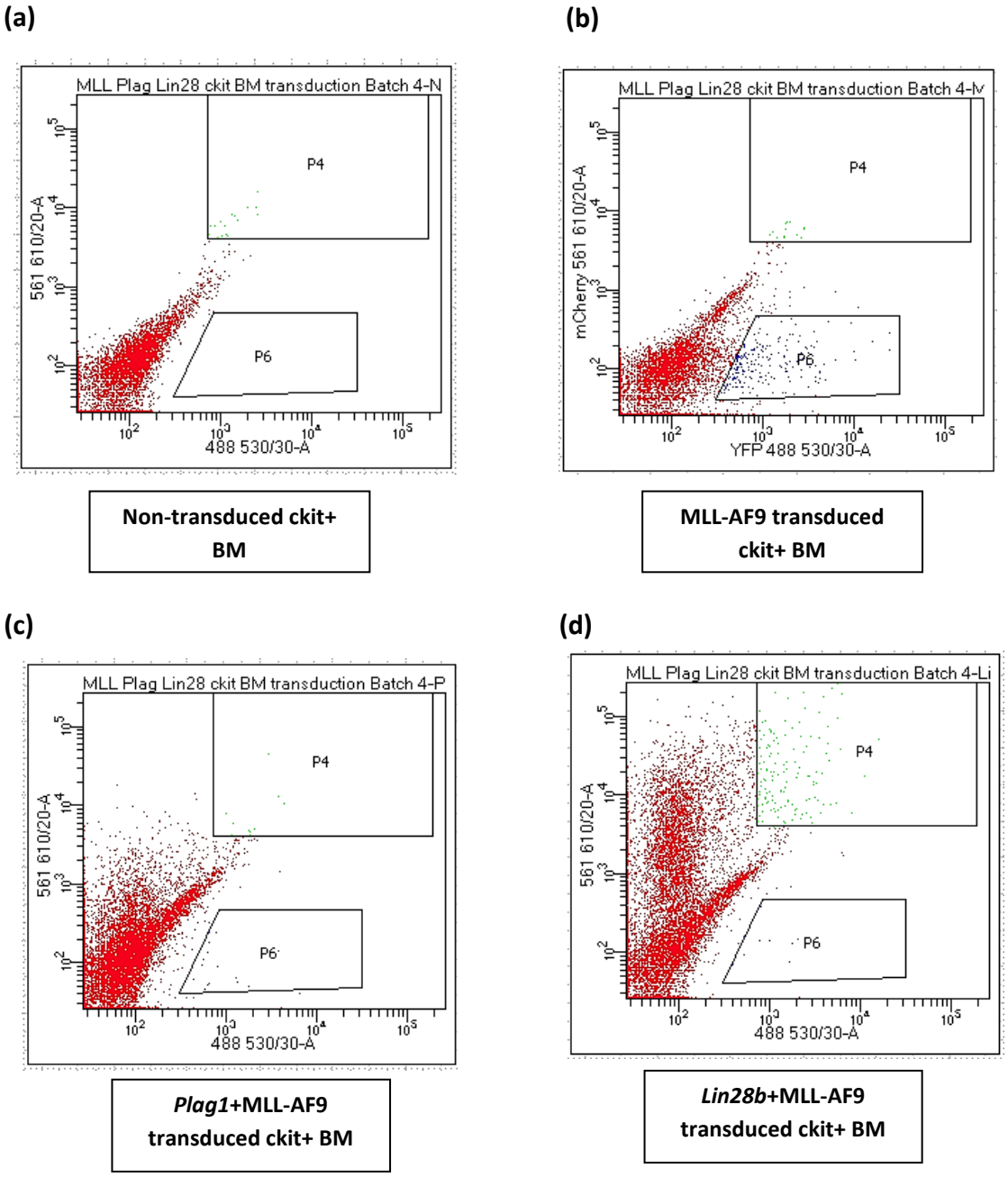


FIGURE 6.2 **Flow cytometry analysis of single (MLL-AF9) and dual (*Plag1*+MLL-AF9 or *Lin28b*+MLL-AF9) transduced c-kit selected BM HSPCs shows dual transduction <u>without</u> a significant single MLL-AF9 transduced population**

Live cells firstly gated by forward and side-scatter, doublets then removed (not shown).
YFP detected on 488 530/30A channel (x-axis), mCherry detected on 561 610/20A channel (y-axis)

a) Non-transduced bone marrow (Control)
b) MLL-AF9 transduced bone marrow (P6 approximately 4-5% in this case)
c) *Plag1* and MLL-AF9 transduced – dual transduced cells in P4 (~1%), P6 (~0.5%)
d) *Lin28b* and MLL-AF9 transduced – dual transduced cells in P4 (8-10%), P6 (~0.5%)

For *Plag1* + MLL-AF9 transduced cells (Figure 6.2c), dual transduction was lower than anticipated (P4), but perhaps reflected the larger insert size of *Plag1* (1500bp). There were, however, a reasonable proportion of single *Plag1* transduced cells (as indicated by single mCherry expression), but virtually no single MLL-AF9 transduced cells (as indicated in P6). *Plag1* overexpression alone did not induce disease in the *CBFB-MYH11* study with 100% of single *Plag1* transplanted C57/Bl6 mice remaining alive and well through the entire experiment, so this was not a particular concern.[16]

For *Lin28b* + MLL-AF9 transduced cells (Figure 6.2d), dual transduction was reasonably high (P4), and higher than with *Plag1* + MLL-AF9, perhaps owing to the smaller insert size of *Lin28b* (800bp). There was a significantly higher proportion of single-transduced *Lin28b* cells (as indicated by single mCherry expression), but only a small number of single MLL-AF9 transduced cells. Whilst there have been no reports of the effects of *Lin28b* retroviral overexpression in the haematopoietic system in the literature, as mentioned before, a transgenic mouse model employing *Lin28b* enforced expression led to the development of peripheral T cell lymphoma, with a disease latency considerably longer than that of MLL-AF9 transplanted mice, so we were not unduly concerned that these single-transduced *Lin28b* cells would influence the development of AML in this experiment or affect outcome.

These mock transductions allayed concerns regarding any interference of additional single MLL-AF9 transduced cells and each batch of retroviral supernatant was tested in this way prior to use in the final *in vivo* assessment.

6.4 Assessment of *Lin28b* and *Plag1* overexpression in c-kit+ transduced HSPCs

The presence of mCherry following transduction was a surrogate marker for adequate expression of the retroviral construct. However, to ensure the constructs were leading to an overexpression of both candidate genes appropriately at the transcript and protein level, it was imperative to assess this appropriately.

RNA was extracted from the c-kit positive HSPCs following their transduction with *Lin28b* or *Plag1* in the assessments described above. This was then used to quantitatively assess expression using qPCR (as described in Methods 2.6.8.3). The housekeeping gene used as a reference was *Gapdh* and for each condition being assessed, the other was used as a negative control (Figure 6.3):

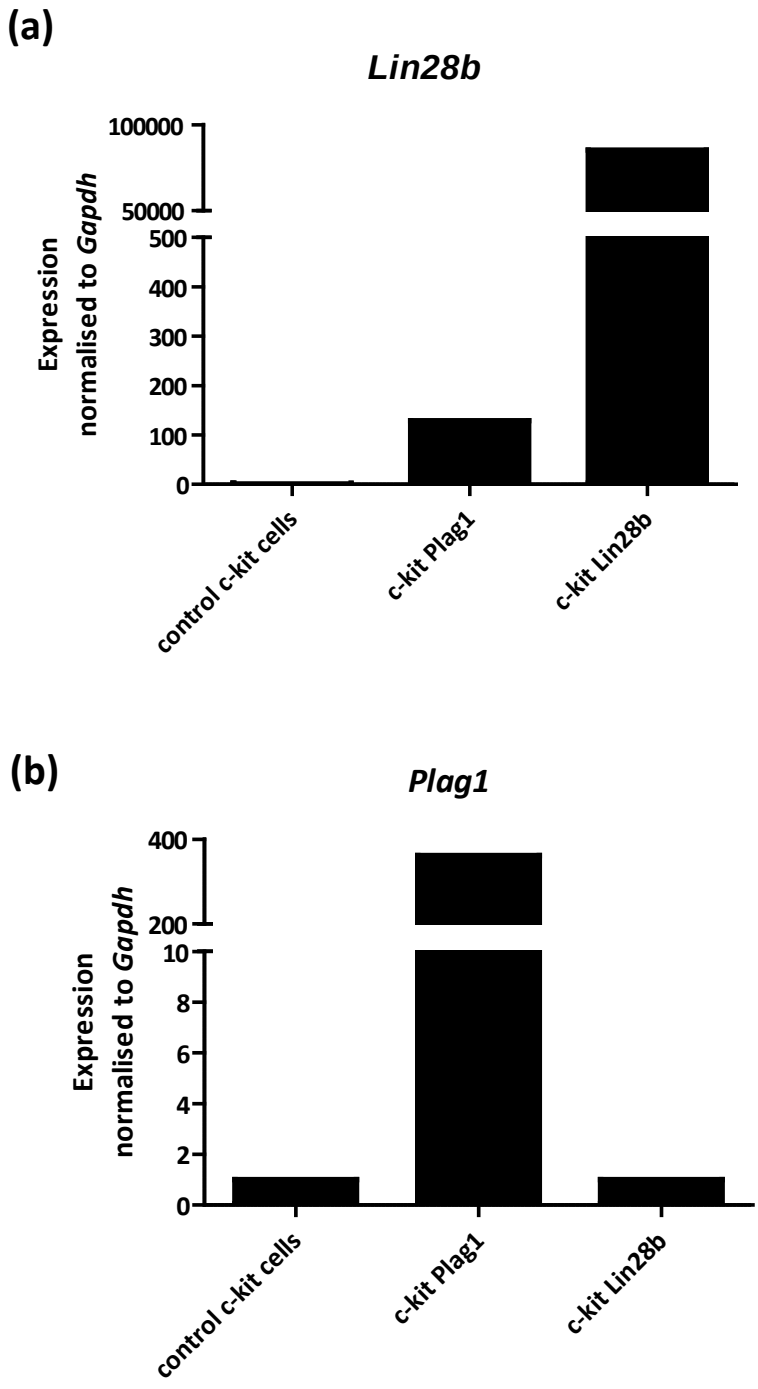


FIGURE 6.3 **RT qPCR demonstrating overexpression of (a) *Lin28b* following transduction with MSCV-*Lin28b*-IRES-mCherry construct and (b) *Plag1* following transduction with MSCV-*Plag1*-IRES-mCherry construct in c-kit selected wild-type HSPCs normalised to *Gapdh* control**

a) *Lin28b* is significantly overexpressed following single transduction of c-kit selected HSPCs with its construct. Note *Plag1* expression is also modestly increased when normalised to *Gapdh* control (discussed further in main text)
b) *Plag1* is significantly overexpressed following single transduction of c-kit selected HSPCs with its construct compared to *Gapdh* control and *Lin28b*

Both qPCR assays demonstrated that each construct was able to produce an appropriately significant increase in expression of the gene at the transcript level within the c-kit selected HSPCs when compared to control. Interestingly, *Lin28b* overexpression appeared to be unexpectedly causing a modest degree of *Plag1* overexpression (Figure 6.3a), suggesting the possibility that it may exert some control over *Plag1* expression independent of Ezh2 status. A cursory glance at a publicly available micro RNA database (www.mirdb.org) reveals the *Plag1* gene in *Mus Musculus* is a predicted target of 180 miRNA's in order of interaction, amongst which three of the let-7 family appear in the top ten. Though speculative, this suggests *Lin28b* may exert a direct positive regulatory effect on *Plag1* through releasing let-7 mediated suppression and would warrant further investigation.

Cell lysates were prepared from *Lin28b* or *Plag1* transduced c-kit selected HSPCs for protein extraction (as described in Methods 2.5.2). Western blotting was then performed to assess for overexpression of each candidate at the protein level within these cells (Figure 6.4):

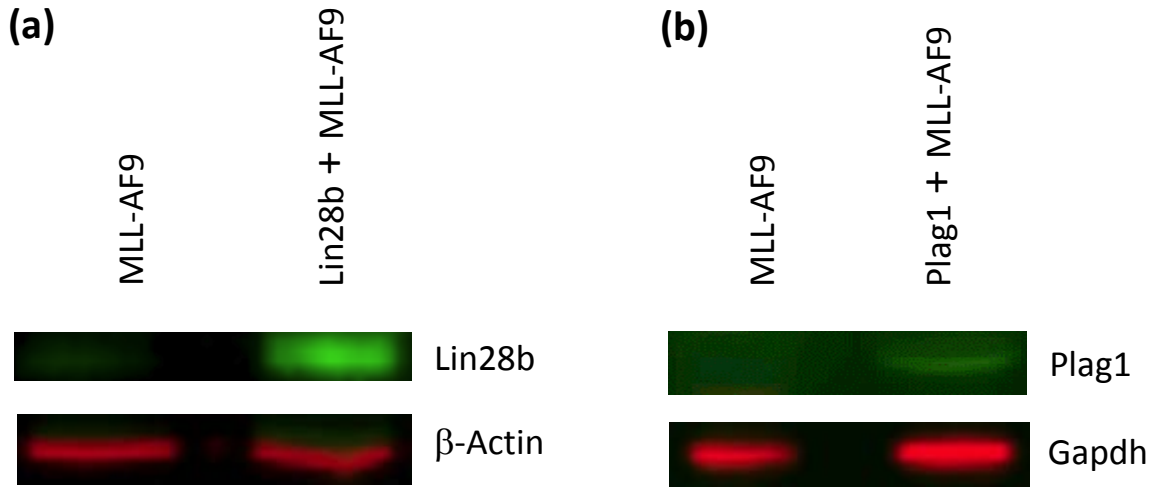


FIGURE 6.4 **Western blot demonstrating overexpression of (a) Lin28b protein after transduction with *Lin28b* and MLL-AF9 constructs, (b) Plag1 protein after transduction with *Plag1* and MLL-AF9 constructs, compared to single-transduced MLL-AF9 control**

a) Lin28b protein is significantly overexpressed following transduction with viral supernatant from dual- *Lin28b* and MLL-AF9 construct transduced cells. β-actin used as loading control
b) Plag1 protein is overexpressed following transduction with viral supernatant from dual-*Plag1* and MLL-AF9 construct transduced cells. Gapdh used as loading control

Both Lin28b and Plag1 were overexpressed at the protein level compared to minimal expression in single MLL-AF9 transduced c-kit positive HSPCs (Plag1 to a slightly lesser extent than Lin28b as evidence by blot).

Having demonstrated that both constructs were capable of causing adequate overexpression of *Lin28b* and *Plag1* at both the transcript and protein level (following transduction within the same c-kit HSPC population we intended to use for the transplantation experiments to validate either candidate gene), we proceeded to the functional *in vivo* experiment.

6.5 Overexpression of *Plag1* phenocopies *Ezh2* loss in synergizing with MLL-AF9 to accelerate leukaemia induction

Four wild-type C57/Bl6 mice aged 8-12 weeks were culled and whole bone marrow was harvested as previously described. From this, c-kit+ HSPCs were selected (as described in Methods 2.6.5) and cultured overnight in appropriate culture medium. The cells were then divided into three conditions and transduced with retroviral supernatant:

a) MSCV-MLL-AF9-IRES-YFP
b) MSCV-MLL-AF9-IRES-YFP + MSCV-IRES-*Lin28b*-mCherry
c) MSCV-MLL-AF9-IRES-YFP + MSCV-IRES-*Plag1*-mCherry

As described in Section 6.3, an estimation of transduction efficiency for each arm was made through flow cytometry so that the total number of cells injected per lethally irradiated C57/Bl6 wild-type recipient mouse could be adjusted to ensure that equal numbers were being injected per animal and that the ratio of double transduced m-Cherry/YFP+ cells to single YFP+ cells was noted. For the *Lin28b* test arm, the ratio injected per mouse was approximately 3:1 and for the *Plag1* test arm the ratio injected per mouse was 1:1 (this was expected as the number of dual-transduced cells were always lower in the *Plag1* mock transductions than the *Lin28b* mock transductions as described above). The control arm (i.e. MLL-AF9 alone) also had the same number of YFP+ cells injected as per the two test arms. An overview of the experiment is depicted in Figure 6.5:

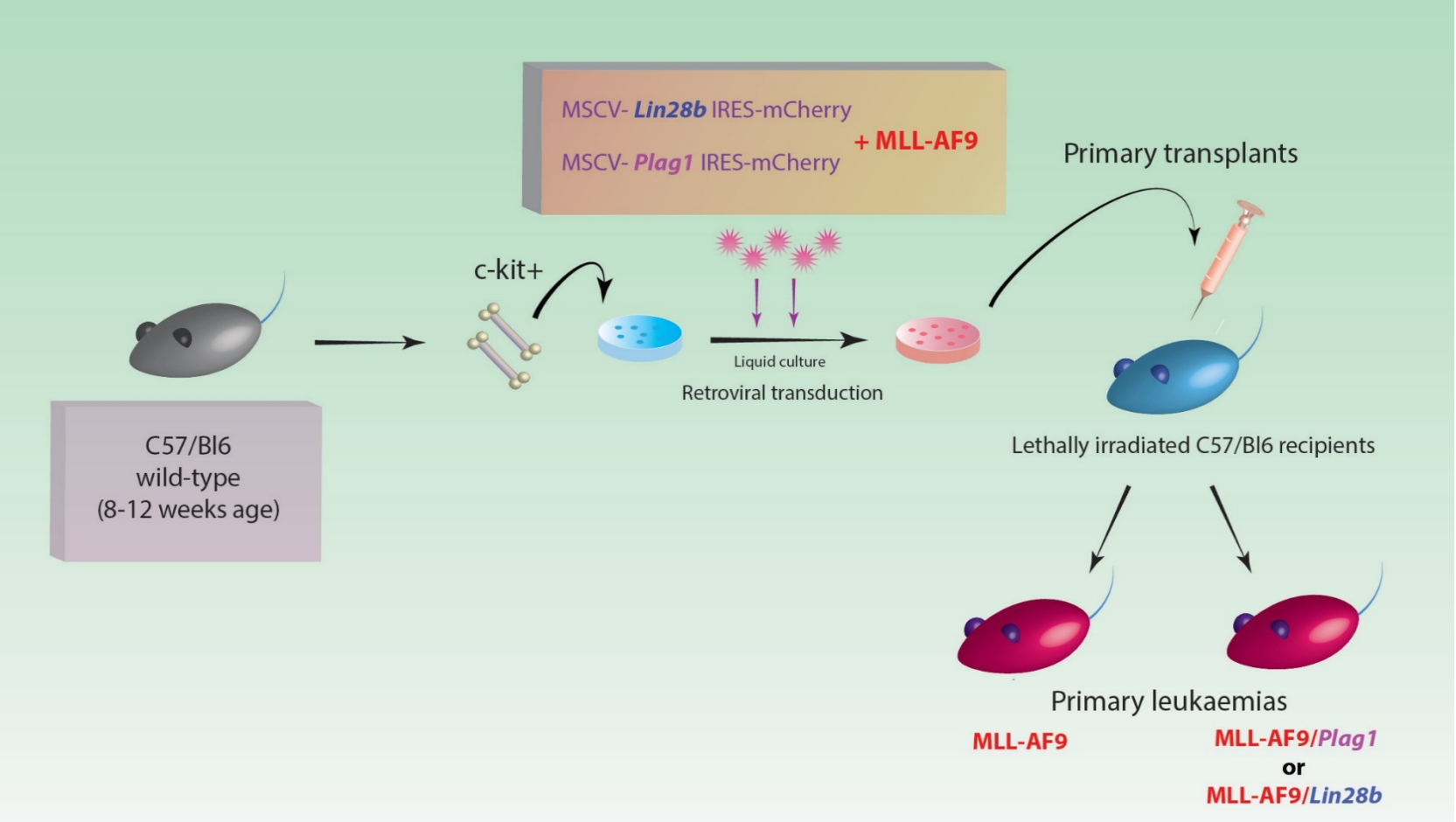


FIGURE 6.5 **Experimental design used to functionally validate *Plag1* or *Lin28b* overexpression with MLL-AF9 *in vivo***

BONE MARROW TRANSDUCTION AND TRANSPLANTATION ASSAY:
C57/Bl6 wild-type recipient mice aged 8-12 weeks are harvested for whole bone marrow. c-kit positive HSPCs are then selected and placed into liquid culture divided into three arms. Each arm is then transduced with a) MLL-AF9, b) MSCV-*Lin28b*-IRES-mCherry + MLL-AF9 or c) MSCV-*Plag1*-IRES-mCherry + MLL-AF9 retroviral supernatant. After assessment of transduction efficiency by flow cytometry, cells numbers are then adjusted accordingly (as described in main text) and injected into lethally irradiated C57/Bl6 wild-type recipients. These are monitored clinically for disease and differences in survival outcomes are assessed

Mice transplanted with *Plag1* + MLL-AF9 transduced c-kit+ HSPCs became moribund as early as Day 35 post transplantation and overall developed leukaemia significantly quicker than those transplanted with MLL-AF9 transduced c-kit+ HSPCs (Figure 6.6):

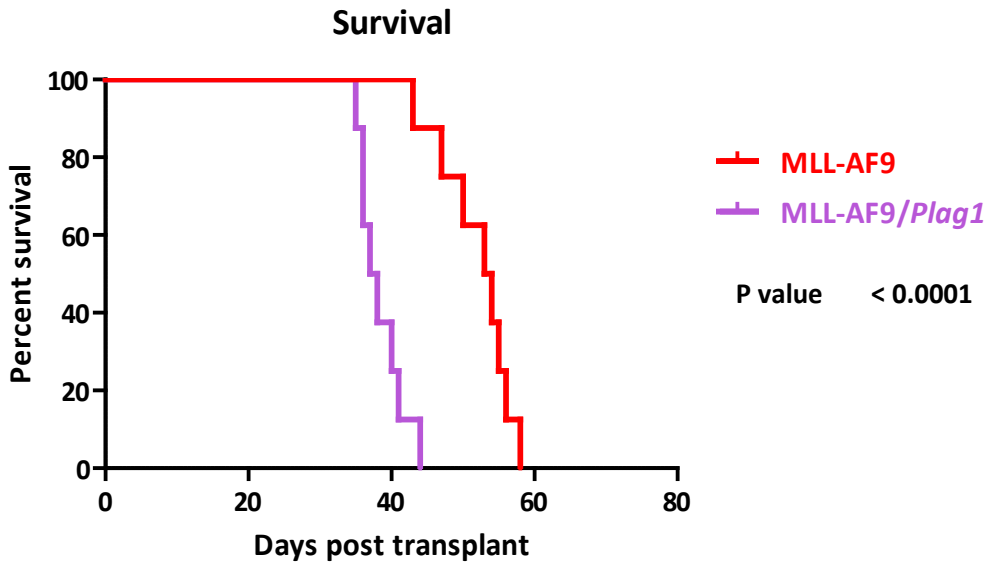


FIGURE 6.6 ***Plag1* co-operates with MLL-AF9 to accelerate the induction of AML in mice**
Kaplan-Meier survival curve of mice that were transplanted *Plag1* + MLL-AF9 transduced c-kit+ HSPCs or MLL-AF9 transduced c-kit HSPCs alone

Mice exhibited elevated white cell counts and anaemia as they became terminal. The median survival differed significantly from 37.5 days in the *Plag1* + MLL-AF9 group to 53.5 days in the MLL-AF9 control group. Analysis of both spleen and bone marrow from *Plag1* + MLL-AF9 transplanted mice demonstrated an increase in the ratio of double transduced m-Cherry/YFP+ cells to single YFP+ cells from approximately 1:1 at the start of transplant to as high as 5:1 in BM and 3.1:1 in spleen (averaged across 5 animals), clearly indicating that dual overexpressing *Plag1* and MLL-AF9 cells had a selective advantage over single MLL-AF9 transduced cells alone, and were accelerating disease. At the time of writing these animals are being further characterised, but there are no apparent significant phenotypic differences.

Mice transplanted with *Lin28b* + MLL-AF9 transduced c-kit HSPCs failed to show any significant acceleration of AML induction compared to those transplanted with MLL-AF9 transduced c-kit+ HSPCs alone (Figure 6.7):

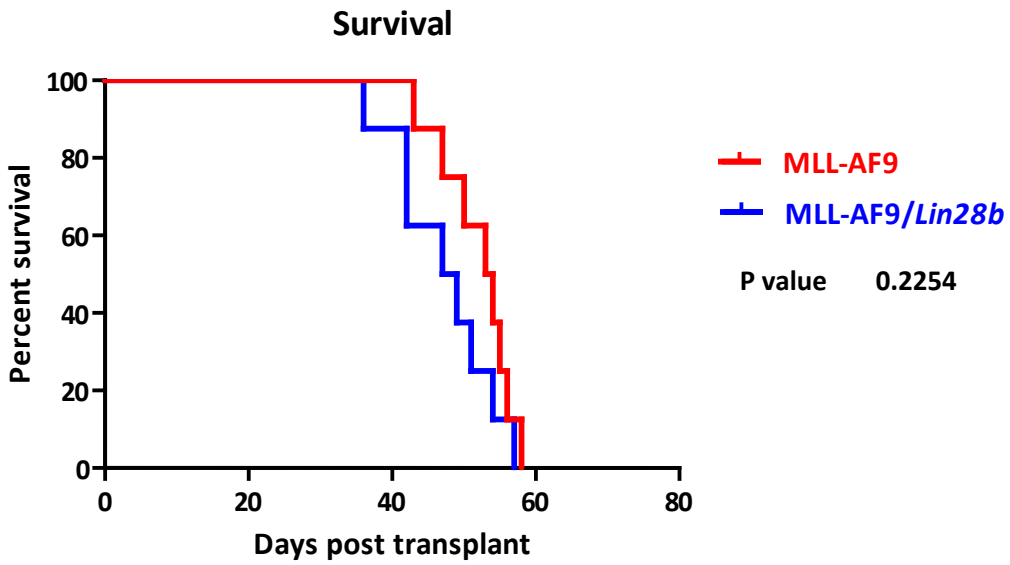


FIGURE 6.7 ***Lin28b* overexpression with MLL-AF9 does not significantly accelerate the induction of AML in mice**

Kaplan-Meier survival curve of mice that were transplanted with *Lin28b* + MLL-AF9 transduced c-kit+ HSPCs or MLL-AF9 transduced c-kit HSPCs alone

Though animals transplanted with *Lin28b* + MLL-AF9 transduced c-kit HSPCs did develop disease slightly earlier, becoming moribund as early as Day 36, the overall survival trend did not demonstrate a significant alteration in outcome. Median survival was 48 days in the *Lin28b* + MLL-AF9 transplanted group compared to 53.5 days in the MLL-AF9 control group. There was also no significant expansion in the ratio of double transduced m-Cherry/YFP+ cells to single YFP+ cells in haematopoietic tissues of terminal mice suggesting no functional advantage for dual overexpressing *Lin28b* and MLL-AF9 cells. Again at the time of writing, the leukaemias are being further characterised but there are no obvious phenotypic differences.

6.6 Discussion and future work

The loss of *Ezh2* prior to leukaemia induction with MLL-AF9 and AML1-ETO9a dramatically accelerated leukaemogenesis. Through detailed characterisation of the genomic changes resulting in *Ezh2* loss we were able to discover that the overexpression of a few select genes resulted from a loss of repression by H3K27me3 and an upregulation of the activating histone mark H3K4me£ at their promoters. From this select group both *Lin28b* and *Plag1* were identified as possible oncogenic mediators of *Ezh2* loss in augmenting disease development.

Having established and validated a retroviral overexpression system for either candidate gene, we demonstrated that overexpression of *Plag1* co-operated with MLL-AF9 in clearly accelerating the development of AML (compared to MLL-AF9 driven AML) without any notable changes in phenotype. Haematopoietic tissues were infiltrated by an increased ratio of dual *Plag1/*MLL-AF9 expressing leukaemia cells, on cursory analysis by flow cytometry. Overexpression of *Plag1* was able to phenocopy loss of *Ezh2* in accelerating the induction of MLL-AF9 leukaemia. This significant result was further corroborated by the fact that the *Plag1*/MLL-AF9 combination resulted in a marked expansion of the cells over single MLL-AF9 expressing cells. The ability of a single gene to be able to augment and drive as aggressive a leukaemia programme as MLL-AF9 is remarkable. In its role as a transcription factor, it is possible that overexpression of *Plag1* may be driving another transcriptional programme that synergizes with MLL-AF9 and facilitates transformation. PLAG1 regulates gene expression through DNA binding at promoters. The insulin like growth factor gene (*IGF2*) is known to contain multiple PLAG binding sites within its promoter. [17] The *Cbfb-MYH11* and *Plag1* study proposed the possibility that *Plag1* may expand haematopoietic precursors through driving the *Igf2* pathway and found that *Igf2* transcript levels in bone marrow cells overexpressing *Plag1* were increased. Furthermore, loss of imprinting at the *IGF2* locus resulting in overexpression of IGF2 has been described in AML cases. [18]

Though the genomic analysis provided evidence of its role as a downstream mediator of *Ezh2* loss, the overexpression of *Lin28b* did not significantly accelerate MLL-AF9 driven disease. Whilst also being repeatedly identified as a potential co-operator with *JAK2*-V617F following *Ezh2* loss in the MPN study, when functionally assayed *in vivo, Lin28b* did not produce a significantly synergistic effect - rather this was borne out by *Hmga2,* which *Lin28b* potentially exerts control over (through suppression of let-7 microRNAs). [12]

Though speculative at present, it is not unreasonable to propose that there are multiple pathways through which *Ezh2* loss in haematopoietic progenitors may facilitate transformation and drive the

MLL-AF9 leukaemia programme. Whilst the overexpression of *Lin28b* described above did not demonstrate a significant acceleration in MLL-AF9 disease induction, following *Ezh2* loss, *Lin28b* was clearly the most upregulated gene in both the non-transformed setting and MLL-AF9 leukaemias from the gene expression datasets, suggesting it may still contribute. As discussed above, the Lin28b protein blocks normal maturation of inhibitory let-7 microRNAs. One of the downstream targets of the let-7 miRNA suppressor family is the insulin-like growth factor binding protein *Igf2bp3,* and following *Ezh2* loss, this gene was also noted to be significantly upregulated in both the non-transformed and MLL-AF9 induction leukaemias as described in Chapter 5. IGF2BP3 is known to facilitate translation of an IGF2 transcript variant through direct binding to its 5'UTR sequence. [19]

The synergism between *Ezh2* loss and MLL-AF9 is likely to be mediated by the concerted action of several genes and this data indicates that *Plag1* is clearly one of these downstream targets. A possible sequence of events following loss of Ezh2 in haematopoietic progenitors might be that, firstly through direct chromatin effects, the upregulation of a number of genes including *Plag1* and *Lin28b* occurs. *Plag1* may then facilitate transformation potentially through an expansion of progenitors by driving the Igf2 pathway (as also proposed in the *Cbfb-MYH11* study). The concurrent upregulation of *Lin28b*, whilst not driving MLL-AF9 leukaemia forwards directly on its own, may also facilitate the proposed Plag1-Igf2 pathway in a number of ways. Firstly, through directly increasing the expression of *Plag1* as surprisingly noted in the qPCR assessments in Figure 6.3, and secondly through releasing let-7 mediated suppression of *Igf2bp3.* (Interestingly, *Ezh2* loss itself may also fuel the upregulation of *Igf2bp3* as observed in our gene expression datasets, where there was an associated reduction in H3K27me3 at its promoter). As described, *IGF2BP3* is known to increase the translation of *IGF2* mRNA transcripts. For these reasons, a plausible common pathway between *Plag1* and to a lesser extent *Lin28b*, following *Ezh2* loss might be through the Igf2 pathway, promoting an environment conducive to MLL-AF9 mediated leukaemic transformation. A summary of these pathways is shown in Figure 6.8.

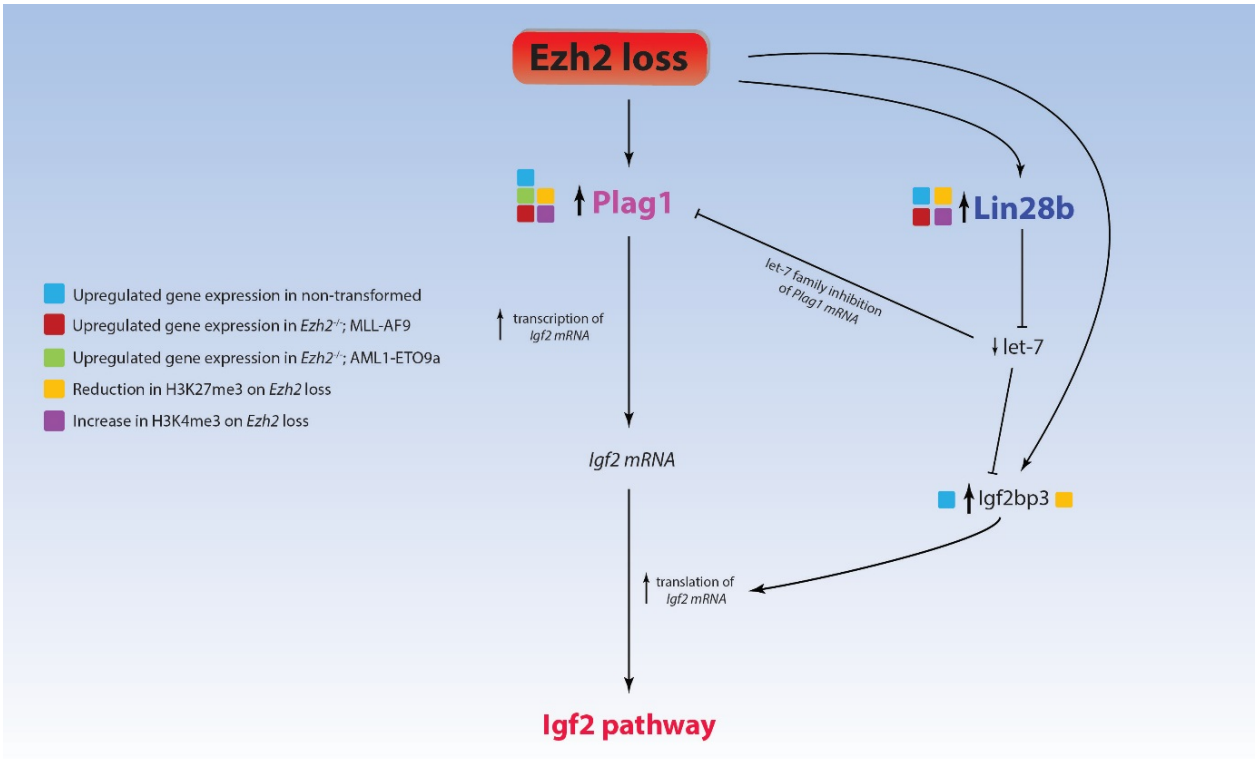


FIGURE 6.8 **Proposed molecular model of *Ezh2* loss and downstream effects**
Ezh2 loss leads to direct upregulation of *Plag1* and *Lin28b* through loss of H3K27me3 and an increase in H3K4me3 at their promoters. Lin28b overexpression negatively regulates maturation of the let-7 miRNA family, releasing let-7 inhibition on *Igf2bp3* and on *Plag1*. Plag1 and Igf2bp3 co-operate to potentially drive the Igf2 axis through increased transcription and translation of Igf2 mRNA respectively

Though all of this is highly speculative, it would be interesting to examine some of these hypotheses further, potentially starting by assessing any functional changes induced in murine HSPCs following *Plag1* overexpression and whether these were related to components of the Igf2 pathway. It would be interesting to assess and compare any differences in expression of *Igf2* between MLL-AF9 and *Plag1* + MLL-AF9 leukaemic tissues.

Finally, whilst demonstrating that *Plag1* is able to copy *Ezh2* loss in synergizing with MLL-AF9 in accelerating leukaemogenesis, it would be important to assess (ideally *in vivo*) if it has the same capabilities in driving AML1-ETO9a leukaemogenesis, which canonically drives AML in a different fashion to MLL-AF9 through repression of multiple pathways. At the time of writing, an *in vitro* characterisation of this is being performed through retroviral transduction and overexpression of *Plag1* together with AML1-ETO9a versus AML1-ETO9a alone in c-kit+ HSPCs.

References

1) Moss EG, Lee RC, Ambros V.The cold shock domain protein LIN-28 controls developmental timing in C. elegans and is regulated by the lin-4 RNA. *Cell*. 1997 Mar 7;88(5):637-46.

2) Viswanathan SR, Daley GQ, Gregory RI. Selective blockade of microRNA processing by Lin28. *Science.* 2008 Apr 4;320(5872):97-100.

3) Kumar MS, Lu J, Mercer KL, Golub TR, Jacks T. Impaired microRNA processing enhances cellular transformation and tumorigenesis. *Nat Genet.* 2007 May;39(5):673-7

4) Kumar MS, Erkeland SJ, Pester RE, Chen CY, Ebert MS, Sharp PA, Jacks T. Suppression of non-small cell lung tumor development by the let-7 microRNA family. *Proc Natl Acad Sci USA.* 2008 Mar 11;105(10):3903-8

5) Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R, Slukvin II, Thomson JA. Induced pluripotent stem cell lines derived from human somatic cells. *Science*. 2007 Dec 21;318(5858):1917-20.

6) Copley MR, Babovic S, Benz C, Knapp DJ, Beer PA, Kent DG, Wohrer S, Treloar DQ, Day C, Rowe K, Mader H, Kuchenbauer F, Humphries RK, Eaves CJ. The Lin28b-let-7-Hmga2 axis determines the higher self-renewal potential of fetal haematopoietic stem cells. *Nat Cell Biol.* 2013 Aug;15(8):916-25.

7) Shyh-Chang N1, Daley GQ. Lin28: primal regulator of growth and metabolism in stem cells. *Cell Stem Cell.* 2013 Apr 4;12(4):395-406.

8) Viswanathan SR, Powers JT, Einhorn W, Hoshida Y, Ng TL, Toffanin S, O'Sullivan M, Lu J, Phillips LA, Lockhart VL, Shah SP, Tanwar PS, Mermel CH, Beroukhim R, Azam M, Teixeira J, Meyerson M, Hughes TP, Llovet JM, Radich J, Mullighan CG, Golub TR, Sorensen PH, Daley GQ. Lin28 promotes transformation and is associated with advanced human malignancies. *Nat Genet.* 2009 Jul;41(7):843-8.

9) Zhou J, Ng SB, Chng WJ. LIN28/LIN28B: an emerging oncogenic driver in cancer stem cells. *Int J Biochem Cell Biol.* 2013 May;45(5):973-8

10) Oshima M, Hasegawa N, Mochizuki-Kashio M, Muto T, Miyagi S, Koide S, Yabata S, Wendt GR, Saraya A, Wang C, Shimoda K, Suzuki Y, Iwama A. *Ezh2* regulates the Lin28/let-7 pathway to restrict activation of fetal gene signature in adult hematopoietic stem cells*. Exp Hematol*. 2016 Apr;44(4):282-96

11) Beachy SH, Onozawa M, Chung YJ, Slape C, Bilke S, Francis P, Pineda M, Walker RL, Meltzer P, Aplan PD. Enforced expression of Lin28b leads to impaired T-cell development, release of inflammatory cytokines, and peripheral T-cell lymphoma. *Blood*. 2012 Aug 2;120(5):1048-59

12) Shimizu T, Kubovcakova L, Nienhold R, Zmajkovic J, Meyer SC, Hao-Shen H, Geier F, Dirnhofer S, Guglielmelli P, Vannucchi AM, Feenstra JD, Kralovics R, Orkin SH, Skoda RC. Loss of *Ezh2* synergizes with JAK2-V617F in initiating myeloproliferative neoplasms and promoting myelofibrosis*. J Exp Med.* 2016 Jul 25;213(8):1479-96

13) Hibbard MK, Kozakewich HP, Dal Cin P, et al. *PLAG1* fusion oncogenes in lipoblastoma. *Cancer Res.* 2000;60:4869-4872

14) Zatkova A, Rouillard JM, Hartmann W, Lamb BJ, Kuick R, Eckart M, von Schweinitz D, Koch A, Fonatsch C, Pietsch T, Hanash SM, Wimmer K. Amplification and overexpression of the *IGF2* regulator *PLAG1* in hepatoblastoma. *Genes Chromosomes Cancer*. 2004 Feb;39(2):126-37.

15) Castilla LH, Perrat P, Martinez NJ, et al. Identification of genes that synergize with *Cbfb-MYH11* in the pathogenesis of acute myeloid leukemia. *Proc Natl Acad Sci U S A.* 2004;101:4924-4929.

16) Landrette SF, Kuo YH, Hensen K, Barjesteh van Waalwijk van Doorn-Khosrovani S, Perrat PN, Van de Ven WJ, Delwel R, Castilla LH. *Plag1* and *Plagl2* are oncogenes that induce acute myeloid leukemia in cooperation with *Cbfb-MYH11*. *Blood*. 2005 Apr 1;105(7):2900-7

17) Hensen K, Van Valckenborgh IC, Kas K, Van de Ven WJ, Voz ML. The tumorigenic diversity of the three PLAG family members is associated with different DNA binding capacities. *Cancer Res.* 2002;62:1510-1517.

18) Wu HK, Weksberg R, Minden MD, Squire JA. Loss of imprinting of human insulin-like growth factor II gene, *IGF2*, in acute myeloid leukemia. *Biochem Biophys Res Commun.* 1997;231:466-472.

19) Liao B, Hu Y, Herrick DJ, Brewer G. The RNA-binding protein IMP-3 is atranslational activator of insulin-like growth factor II leader-3 mRNA during proliferation of human K562 leukemia cells. *J Biol Chem* 2005;280:18517–24.

Chapter 7

Pharmacological inhibition of EZH2 in the context of AML

7.1 Aims

This chapter details the preclinical *in vitro* activity of GSK343, an EZH2 inhibitor, and aims to answer the following:

1. Is EZH2 a plausible therapeutic target (as informed by work in previous chapters) and does pharmacological inhibition with an EZH2 inhibitor have a role in AML?

7.2 Introduction

Abnormal regulation of chromatin modifiers is a common and recurrent theme in the pathogenesis of AML. As discussed above, recent studies have consolidated the significant number of epigenetic regulators that are mutated and potentially implicated as drivers of AML. [1,2] The plasticity and potential reversibility of chromatin modifications makes them tantalising targets for therapeutic intervention. EZH2 has histone methyltransferase activity through its catalytic SET domain, which is potentially druggable for therapeutic benefit. However, based on its contrasting roles in various malignancies (as detailed in Chapter 1) and the phase and subtype of disease as demonstrated in this thesis thus far, the context in which this is proposed is clearly of significant importance. In mice, *Ezh2* is essential for normal foetal haematopoiesis with *Ezh2* deficient murine embryos dying of severe anaemia.[3] In the same study, *Ezh2* deletion in adult murine bone marrow did not significantly compromise haematopoietic stem cell function, importantly highlighting its potential as a target in haematological disease without predicting disastrous toxicities for normal haematopoiesis. Following extensive preclinical work in the lymphoma arena, EZH2 has emerged as a potential therapeutic target in DLBCL and FL with gain-of-function EZH2 mutations[4] and EZH2 inhibitors GSK126, EPZ-6438, and CPI-1205 have been in Phase I Clinical trials for the past 2-3 years [NCT Trial numbers: NCT02082977, NCT01897571, NCT02395601 for compounds GSK2816126, EPZ-6438 and CPI-1205 respectively].

In functional assessments of the role of Ezh2 in AML, we have identified markedly disparate outcomes following genetic loss of *Ezh2* at various stages of leukaemia evolution. The finding and subsequent molecular characterisation of *Ezh2* loss accelerating MLL-AF9 driven leukaemogenesis *in vivo* (and similar observations in AML1-ETO9a leukaemogenesis), raised concerns that EZH2 inhibition in AML may be an issue. However, upon assessment of *Ezh2*'s role in maintaining established AML *in vivo* (for

both the same models) we found the stark opposite – that genetic loss delayed leukaemia progression considerably. We were able to characterise this further mechanistically and demonstrate that a different transcriptional programme was being subverted in AML maintenance to that during induction, providing some reassurance that targeting EZH2 in established AML might be a viable option. Taking this forward, we aimed to assess sensitivity of murine and human AML cell lines as well as human AML primary samples to another EZH2 inhibitor, GSK343 to see if it was possible to further consolidate the hypothesis that EZH2 has an oncogenic role in the maintenance of AML.

Through the specifications of an MTA between the Huntly laboratory and GlaxoSmithKline, the active small molecule inhibitor of EZH2, compound GSK343, and its inactive homolog GSK669 were provided.

GSK343 is a specific inhibitor of the histone H3-lysine 27 (H3K27) methyltransferase EZH2. It inhibits EZH2 enzymatic activity with an IC_{50} of 4nM. [5] A schematic of its structure is shown in Figure 2.1 (Methods 2.1.2). The compound displays 60 fold selectivity for EZH2 over EZH1, and 1000 fold or greater selectivity against other histone methyltransferases (Figure 7.1). EZH1 is highly homologous to EZH2 with 76% sequence identity overall and 96% sequence identity within the catalytic SET domain. The structural factors responsible for the selectivity over EZH1, and for the different levels of selectivity for EZH2 over EZH1, are presently unclear. GSK343 is available as a chemical probe as part of the Structural Genomics Consortium epigenetics initiative and commercially available through Sigma-Aldrich.

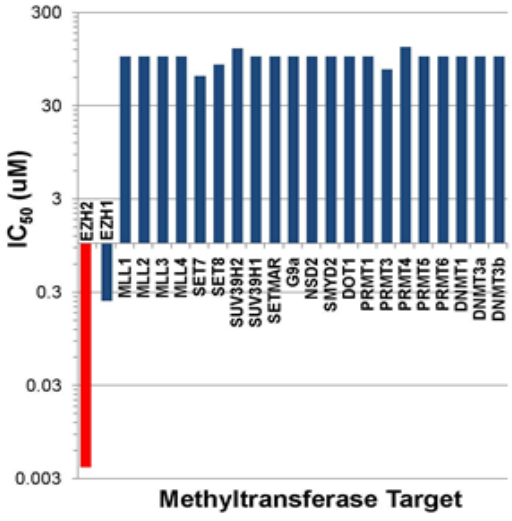


FIGURE 7.1 GSK343 displays marked selectivity for EZH2 over other histone methyltransferases (Adapted from GSK/SGC)

Its inactive homolog GSK669 was also provided to the Huntly laboratory and served as a control, demonstrating no efficacy or significant cellular toxicity.

7.3 GSK343 demonstrates efficacy against human AML cell lines

GSK343 was initially tested against a panel of human AML cell lines available in the Huntly lab (Table 2.1, Methods 2.2.2.1). Of these, the Kasumi and KG1 cell lines exhibited sensitivity above the others. Unresponsive cell lines included K562, MV4-11 and U937 (which harbour the BCR-ABL, MLL-AF4 and CALM-AF10 translocations respectively). A titration of different concentrations was used (assessed concurrently with GSK669 as a control) to roughly determine what concentration could be used for further assessments (Figure 7.2):

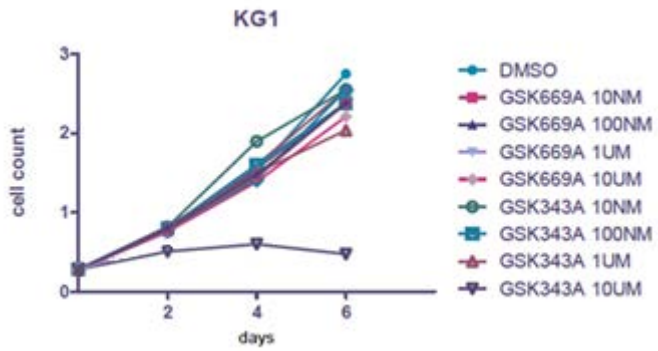


FIGURE 7.2 **The KG1 cell line is most sensitive to GSK343 at 10μM following 6 days in liquid culture growth**
y-axis: Cell count is $x10^6$

This established the 10μM concentration as an appropriate starting point for assessing sensitivity to GSK343. Kasumi and KG1 cells were then expanded in liquid culture over a 12-day time course (as described in Methods 2.3.1) in the presence of DMSO (vehicle control), GSK343 at 10μM and GSK669 at 10μM (Figure 7.3):

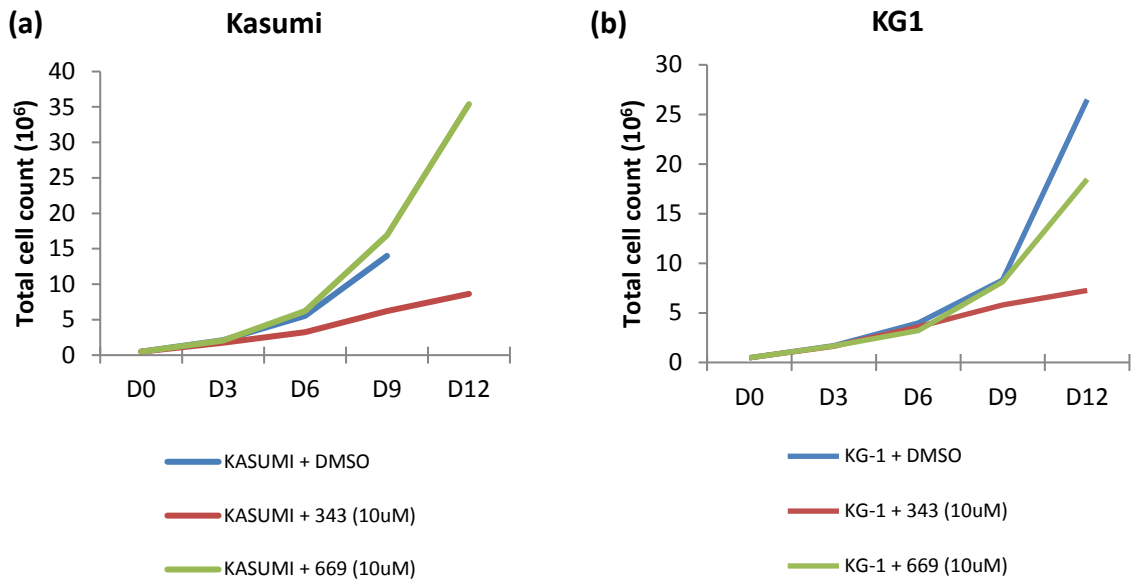


FIGURE 7.3 **GSK343 at 10μM concentration markedly impedes liquid culture growth of AML cell lines (a) Kasumi, (b) KG1**
Total cell count measured across a 12-day period in the presence of DMSO (blue), GSK343 (red) and GSK669 (green)

This was highly interesting given the Kasumi cell line harbours the t(8;21) RUNX1-RUNX1T1 translocation, and corroborated with the clear disruption we had seen following *Ezh2* loss in the maintenance experiments (detailed in Chapter 4), where *in vitro Ezh2* loss completely abolished established AML1-ETO9a driven immortalisation, and *in vivo Ezh2* loss clearly led to the disruption of AML1-ETO9a leukaemia progression. This suggested that pharmacological inhibition of the AML-ETO programme was similar to genetic disruption of *Ezh2* in the murine experiments.

A colony formation assay utilising methylcellulose (as described in Methods 2.3.2) to assess clonogenic and proliferative potential of Kasumi and KG1 cells was then performed. Both cell lines exhibited sensitivity and reduced colony formation and cell proliferation in the presence of GSK343 at 10μM when total colony and cell number were assessed at weekly replatings (Figure 7.4, data shown for KG1 as an example):

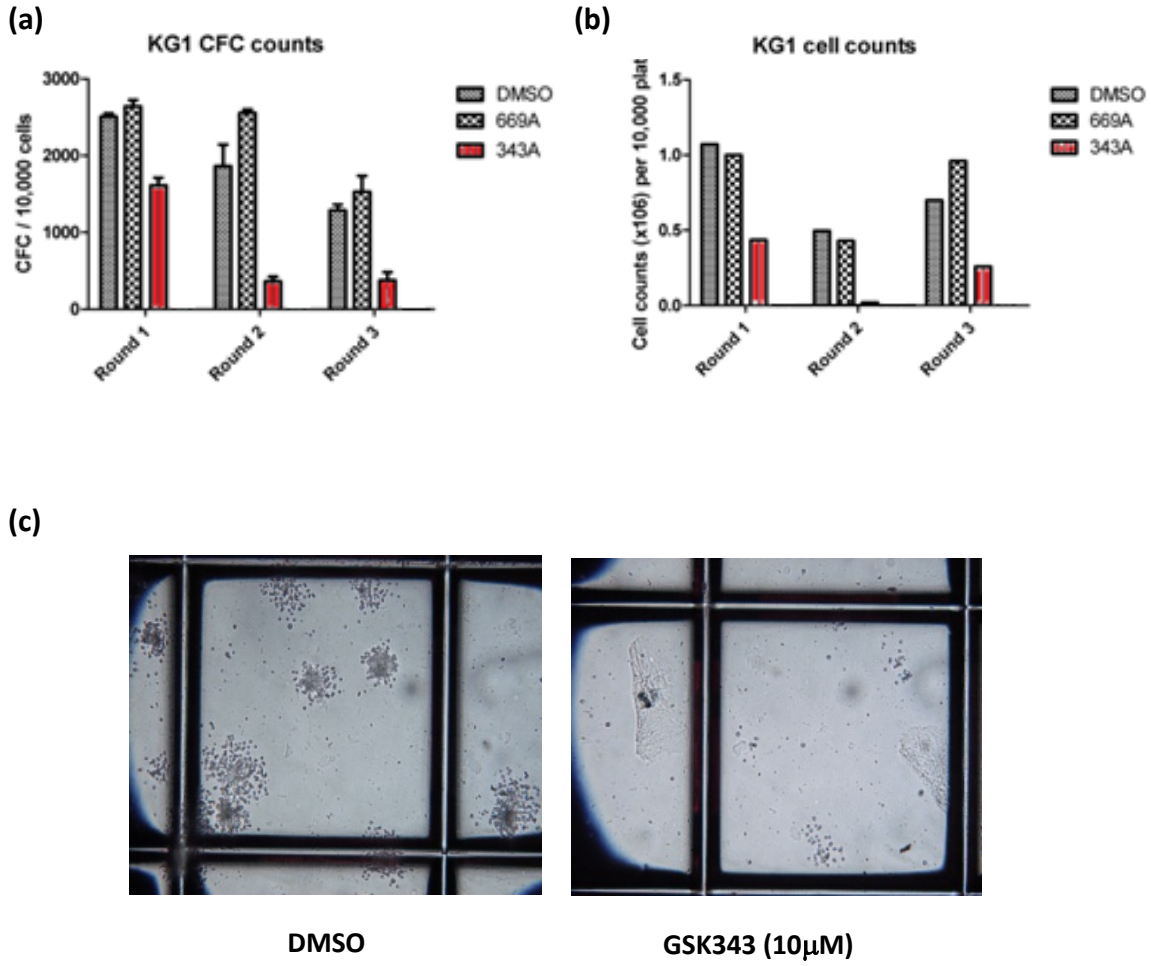


FIGURE 7.4 **GSK343 significantly reduces the clonogenic and proliferative potential of the KG1 cell line in a colony formation assay**

(a) Total colony numbers at weekly replatings for DMSO, GSK669 and GSK343 (at 10μM) treatment. A significant reduction in colony numbers following GSK343 treatment is seen (drugs added at the first round only). 10^4 cells plated each week

(b) Total cell numbers at weekly replatings for DMSO, GSK669 and GSK343 (at 10μM) treatment. A significant reduction in colony numbers following GSK343 treatment is seen (drugs added at the first round only). 10^4 cells plated each week

(c) Morphology of KG1 colonies treated with DMSO (control) left, and GSK343 (10μM) right (x10 magnification). Colonies are dense and well formed in DMSO, but barely present in GSK343

The results for the Kasumi cell line were similar. Both the liquid culture and colony formation assay experiments suggested that GSK343 has efficacy against the Kasumi and KG1 cell lines and the next step was to perform some functional assays to verify this.

7.4 GSK343 treatment is accompanied by modest late apoptosis and G_1 cell cycle arrest

Upon staining, apoptotic cells are Annexin V positive and 7AAD negative. When the Kasumi cell line was treated with GSK343 in liquid culture, though modest, the maximal level of apoptosis (as assessed according to Methods 2.4.1) was consistently seen at 96 hours (with some initial degree noted as early as 72 hours, data not shown) as shown in Figure 7.5. The response was similar for KG1 (data not shown).

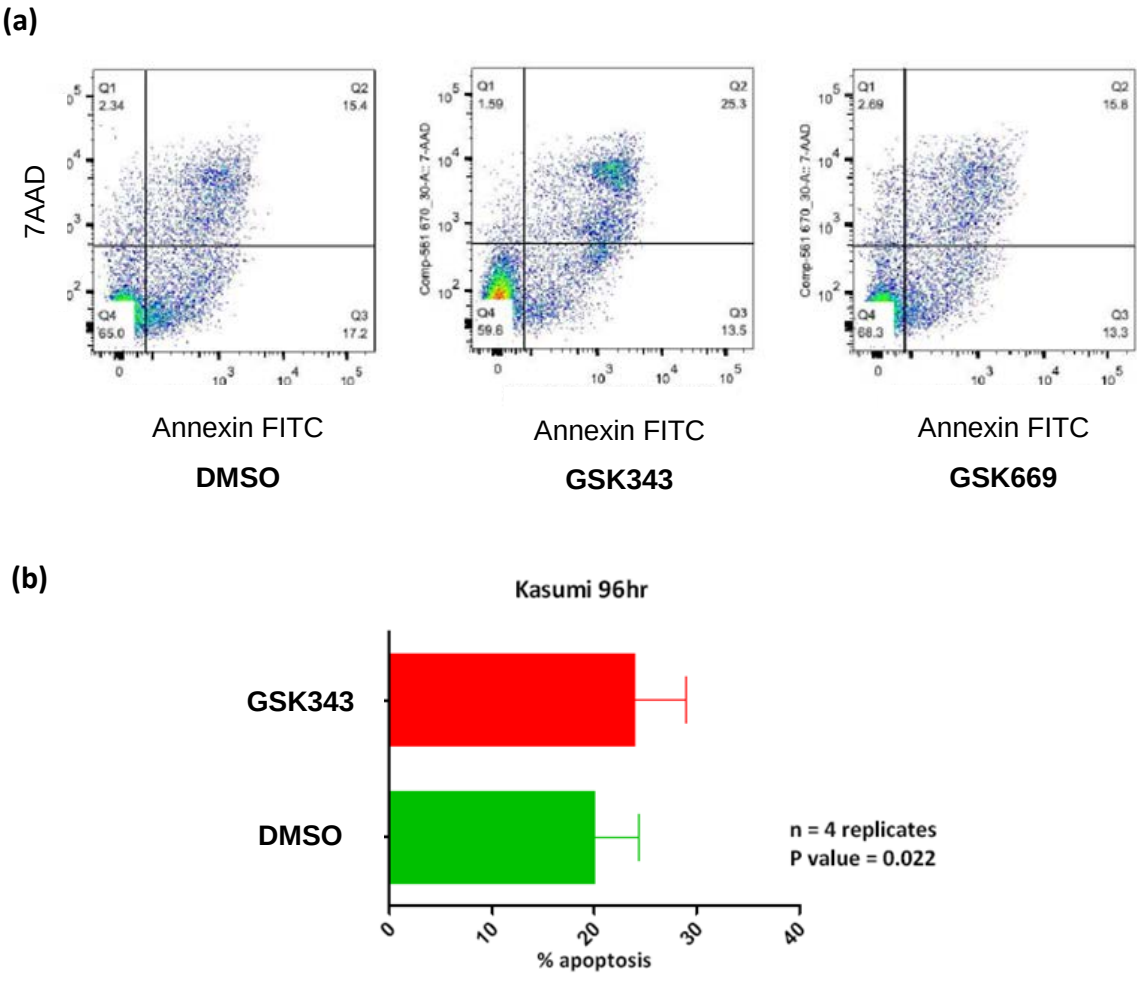


FIGURE 7.5 **GSK343 treatment of the Kasumi cell line in liquid culture causes induction of apoptosis at 96 hours (over 4 replicates)**

(a) Flow cytometry analysis of DMSO, GSK343 and GSK669 treated cells at 96hrs post treatment. Live cells gated by forward and side-scatter, doublets removed and then gated for 7-AAD (y-axis) and Annexin FITC (x-axis)

(b) Percentage of Kasumi cells undergoing apoptosis at 96hrs (DMSO vs GSK343). Average of 4 replicates

Cell cycle analysis was also performed and this demonstrated G_1 cell cycle arrest for both cell lines, most prominent at 96hrs for the KG1 cell line (Figure 7.6). Similar G_1 cell cycle arrest was seen at 96hrs with Kasumi cells (data not shown). The data suggested an antiproliferative effect on Kasumi/KG1 at 72-96 hours post treatment, suggesting that any methylation related changes were occurring before this time point.

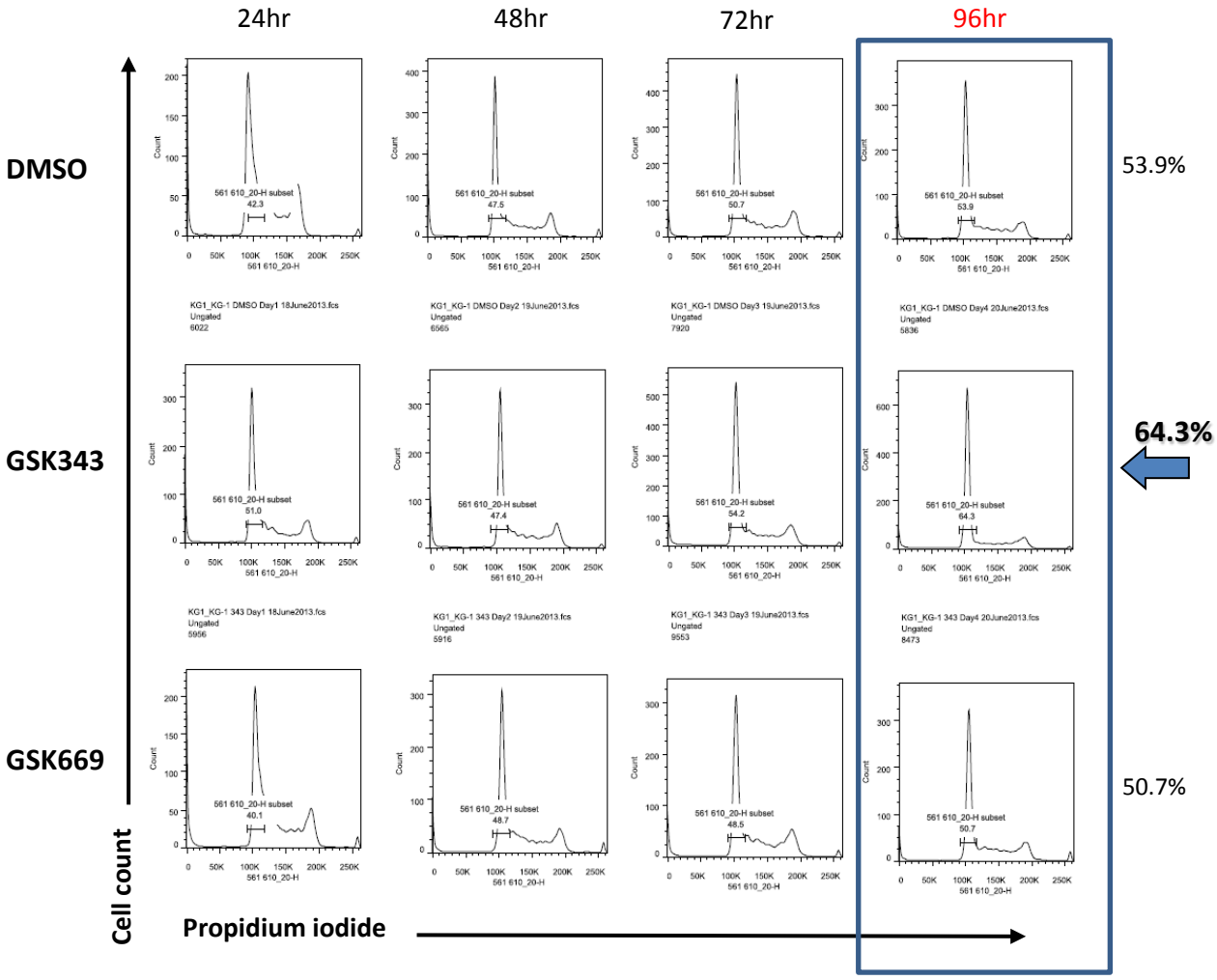


FIGURE 7.6 **GSK343 treatment of the KG1 cell line in liquid culture causes modest G1 cell cycle arrest and is most pronounced at 96 hours**

Flow cytometry analysis of DMSO, GSK343 and GSK669 treated cells at 96hrs post treatment. Live cells gated by forward and side-scatter, doublets removed and then assessed for total cell counts propidium iodide stain.

Though modest, an increase consistent with G_1 cell cycle arrest is seen at 96 hours following GSK343 treatment (highlighted). GSK343 decreases progression from G_0/G_1 to S and G_2M phases. In GSK343 treated cells at 96 hours, 64.3% of cells are in G_0/G_1 phase compared to 50.7-53.9% in GSK669 and DMSO treated conditions respectively.

7.5 GSK343 demonstrates efficacy at micromolar concentrations

Through employing a cell proliferation assay (as described in Methods 2.3.3), the efficacy of GSK343 against cell proliferation of Kasumi and KG1 was tested to determine the optimal concentration of GSK343 required to reduce proliferation by 50% (IC_{50}). The results shown in Figure 7.7 suggest an IC_{50} of 10-20μM for these two cell lines, corroborating the concentrations used previously:

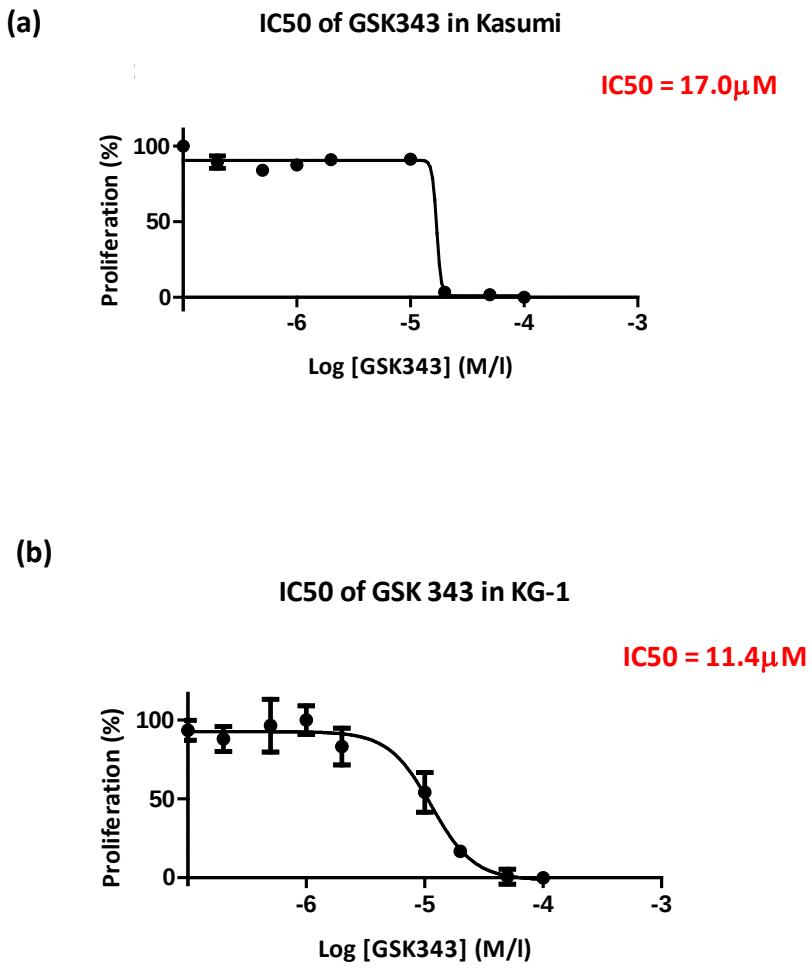


FIGURE 7.7 **GSK343 inhibits cell proliferation of (a) Kasumi and (b) KG1 cell lines at IC_{50}'s of 17μM and 11.4μM respectively**

The [3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium] MTS compound was used to assay cell proliferation at 72 hours across a range of different GSK343 concentrations (from 100nM to 100μM, represented on a logarithmic scale, x-axis). The proliferation reduced in a sigmoid manner as the concentration of drug was increased. Each assay was repeated in triplicate.

7.6 H3K27 trimethylation is inhibited as early as 48 hours by GSK343

Lysates of Kasumi cells treated with DMSO/GSK343 were prepared and processed for Western blotting as described in Methods 2.5. Figure 7.8 shows a time course of progressive H3K27me3 changes with EZH2 inhibition, with a noticeable reduction in H3K27me3 as early as 48hrs.

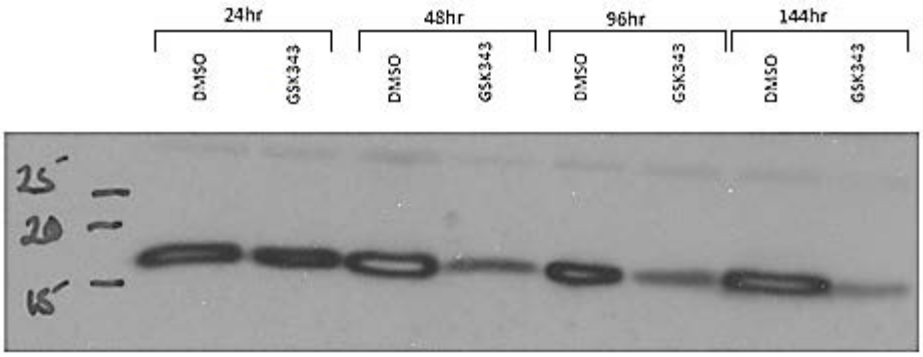


FIGURE 7.8 **GSK343 treatment of the Kasumi cell line results in a reduction of H3K27me3 as early as 48 hours**
Western blot showing decrease in H3K27me3 with GSK343 treatment compared to DMSO at 24hr, 48hr, 96hr and 144hr

By Day 6 (144hrs) both Kasumi and KG-1 (and NB4 to a lesser degree) show a considerable reduction in H3K27me3 as demonstrated in Figure 7.9:

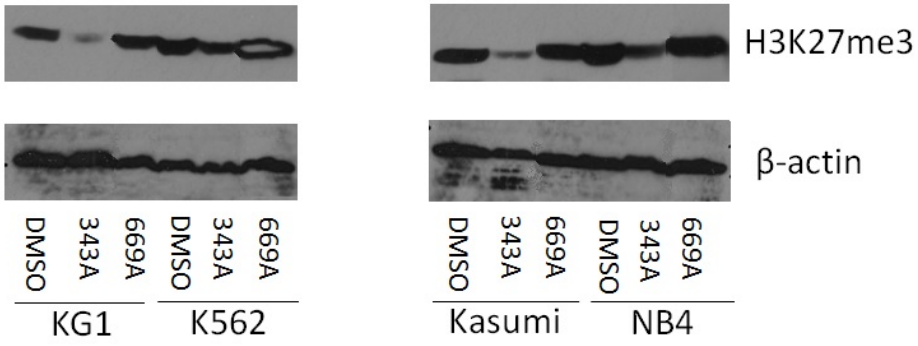


FIGURE 7.9 **At 144hrs GSK343 treatment results in a significant reduction of H3K27me3 in Kasumi, KG1 and NB4 human AML cell lines**
DMSO/GSK343/GSK669 against sensitive KG1, Kasumi, partially sensitive NB4 and resistant K562 AML cell lines, at 144 hours post treatment

As epigenetic modifications and their downstream effects tend to occur over several days, these findings of EZH2 inhibition leading to a decrease in H3K27me3 from 48 hours onwards in the sensitive AML cell lines were expected, and were compatible with the modest functional effects of apoptosis and cell cycle arrest that were demonstrated to occur at 96 hours and beyond.

7.7 Transformed *Ezh2fl/fl*;wt murine cell lines exhibit sensitivity to GSK343

The immortalised cell lines generated through transduction of *Ezh2fl/fl*;wt c-kit selected HSPCs with oncogenes MLL-AF9, AML1-ETO9a and MOZ-TIF2 (detailed in Chapter 4) were maintained in methycellulose replatings on a weekly basis. After a second transduction with a separate retroviral Cre-recombinase, the resulting *Ezh2* deletion resulted in complete abolishment of clonogenic growth.

To gauge whether pharmacological Ezh2 inhibition had any similar effect to genetic *Ezh2* deletion in these cell lines, a clonogenic assay was performed using established *Ezh2fl/fl*;wt MLL-AF9 and *Ezh2fl/fl*;wt MOZ-TIF2 cells treated with either DMSO (control) or GSK343 (at 10μM). The assay was set up as described in Methods 2.3.2 seeding an initial $5x10^4$ cells per millilitre of methylcellulose. GSK343 treatment resulted in a significant decrease in clonogenic growth assessed by total colony number and a marked decrease in colony size and density (Figure 7.10) in both murine cell lines. The effects were more marked in the MOZ-TIF2 cell line, with a dramatic reduction in colony size as demonstrated. Though not absolute, as was genetic deletion of *Ezh2*, this assay provided further evidence to suggest that pharmacological inhibition of Ezh2 was capable of exerting a negative effect on established leukaemia.

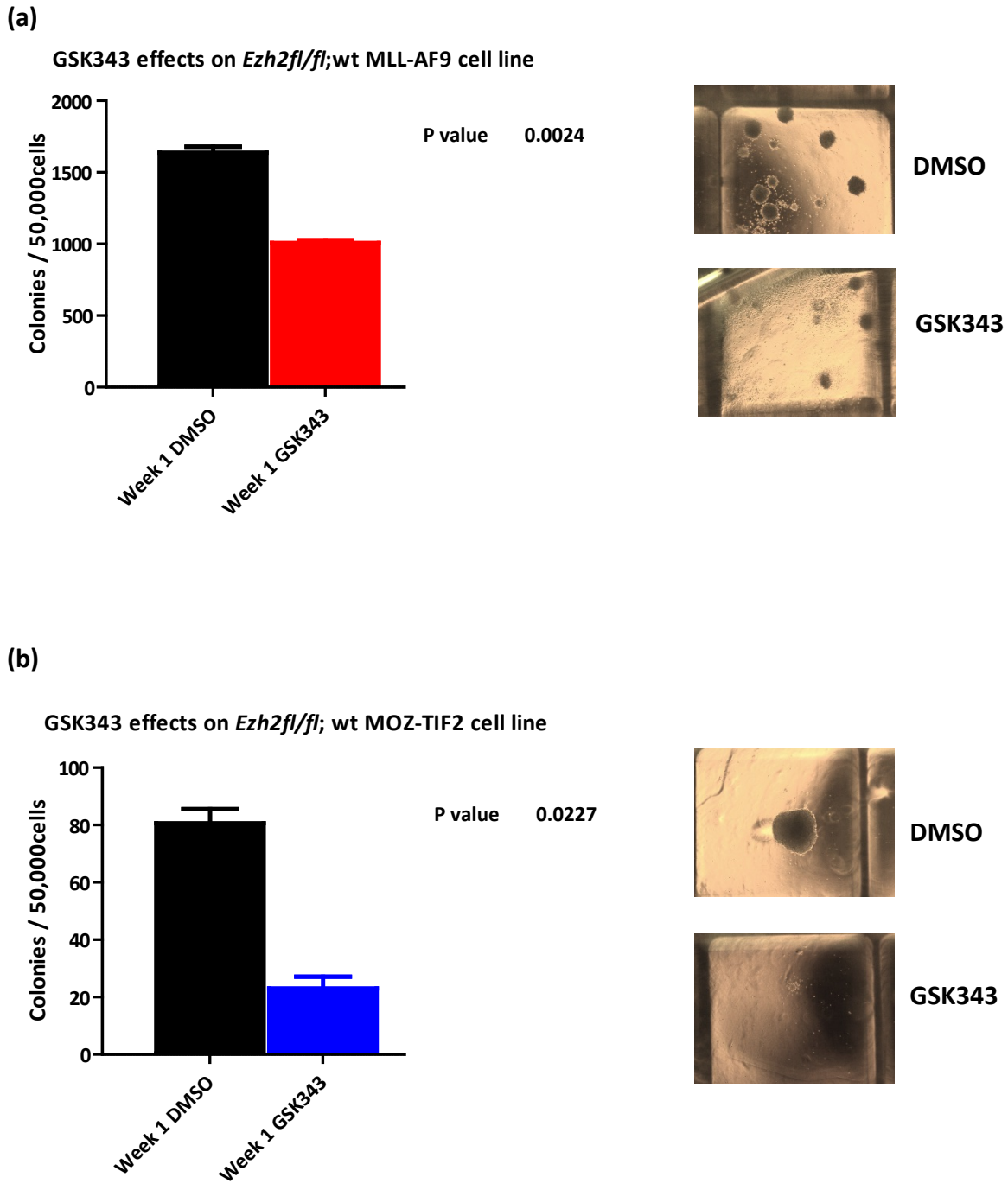


FIGURE 7.10 **GSK343 significantly inhibits clonogenic growth of transformed murine (a) MLL-AF9 and (b) MOZ-TIF2 cell lines *in vitro***

(a) LEFT: Immortalised *Ezh2 fl/fl*;wt MLL-AF9 cells exhibit reduced colony formation at 1 week in the presence of GSK343 (10μM) compared to DMSO control. (n=3) RIGHT: Reduction in colony size and density on GSK343 treatment (x20 magnification)

(b) LEFT: Immortalised *Ezh2 fl/fl*;wt MOZ-TIF2 cells exhibit reduced colony formation at 1 week in the presence of GSK343 (10μM) compared to DMSO control. (n=3) RIGHT: Reduction in colony size and density on GSK343 treatment (x20 magnification)

7.8 Murine *Ezh2fl/fl*;wt MLLAF9 spleen tumours exhibit sensitivity to GSK343 when cultured *in vitro*

Having established that the t(8;21) *RUNX1-RUNX1T1* translocated Kasumi AML cell line (and KG1) and murine transformed MLL-AF9 and MOZ-TIF2 cell lines were both exhibiting a degree of sensitivity to EZH2/Ezh2 inhibition with GSK343 at the same concentration, it was important to assess whether murine leukaemic tissues generated *in vivo* were also affected by Ezh2 inhibition. Genetic deletion of *Ezh2* in established *Ezh2* conditional MLL-AF9 and AML1-ETO9a leukaemias *in vivo* had demonstrated a disruption in leukaemia progression in the transplantation assay described in the maintenance experiments (Chapter 4). We speculated that pharmacological inhibition of Ezh2 through GSK343 treatment of these tumours (in the absence of any *Ezh2* deletion) would also cause impaired growth and proliferation.

To assay this, the splenic tumour tissues from two *Ezh2fl/fl*; wt MLL-AF9 leukaemia (also referred to as $Ezh2^{+/+}$ MLL-AF9 tumours) were placed into liquid culture in sterile conditions and expanded as described in Methods 2.2.2. Two tumours of the same genotype were used to provide biological replicates. Both behaved aggressively in liquid culture, expanding rapidly and requiring passage frequently, showing no signs of exhaustion or morphological differentiation. To establish whether these tumours were sensitive to GSK343, an initial titration of varying concentrations (from 5uM to 250uM) was carried out. Equal numbers of cells from both tumours were cultured in the presence of DMSO or GSK343 at concentrations of 5μM, 10μM, 50μM, 100μM and 250μM for 24 hours and total cell numbers were established at the end of this period. Both tumours (henceforth #5340 and #5342) were inhibited considerably from concentrations of 10μM and upward (data not shown).

Having established this as a starting point, a more precise assessment utilising the cell proliferation [3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium] MTS assay was performed for each tumour (Figure 7.11):

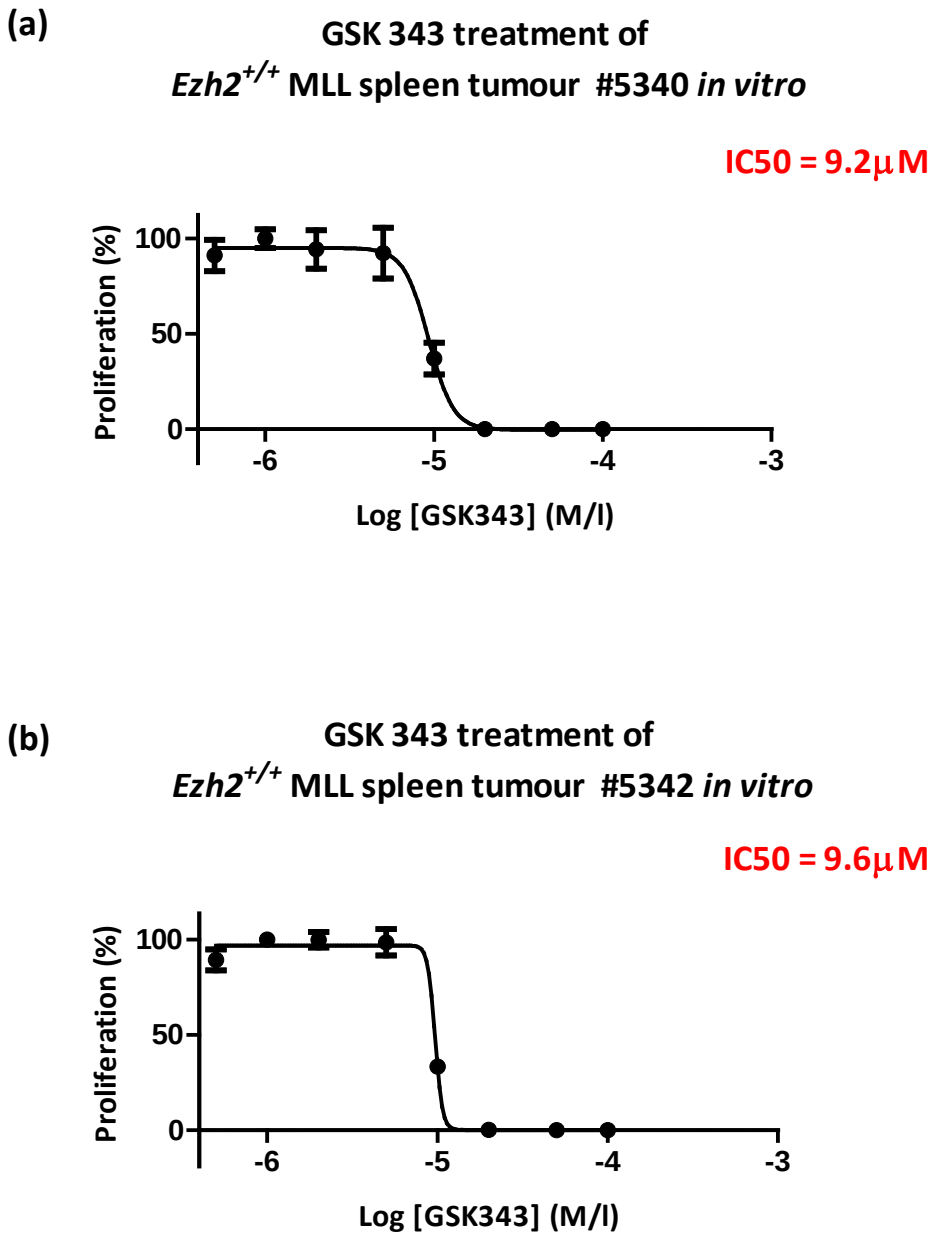


FIGURE 7.11 **IC_{50} assay demonstrating that GSK343 inhibits proliferation of two murine *Ezh2*$^{+/+}$ MLL-AF9 tumours (a) #5340 and (b) #5342 cultured *in vitro* with IC50's of approximately 10μM**

The [3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium] MTS compound was used to assay cell proliferation of *Ezh2*$^{+/+}$ MLLAF9 spleen leukaemia cells cultured *in vitro* at 72 hours across a range of different GSK343 concentrations (from 5μM to 500μM, represented on a logarithmic scale, x-axis). The proliferation reduced in a sigmoid manner as the concentration of drug was increased. Each assay was repeated in triplicate

Strikingly, both tumours #5340 and #5342 were inhibited at 9.2μM and 9.6μM respectively – approximate to the 10μM concentration that the AML cell lines and transformed murine cell lines exhibited sensitivity at.

To test this further, colony forming assays in methylcellulose were performed (as per Methods 2.3.2). Cells from both tumours #5340 and #5342 were taken from liquid culture into methylcellulose treated with either DMSO (control) or GSK343 at 10μM. At Day 7, the total number of colonies produced in each condition for each tumour was assessed, along with a record of colony morphology. Both tumours exhibited a significant reduction in colony formation and size with GSK343 treatment compared to DMSO (Figure 7.12, representative data shown for tumour #5340):

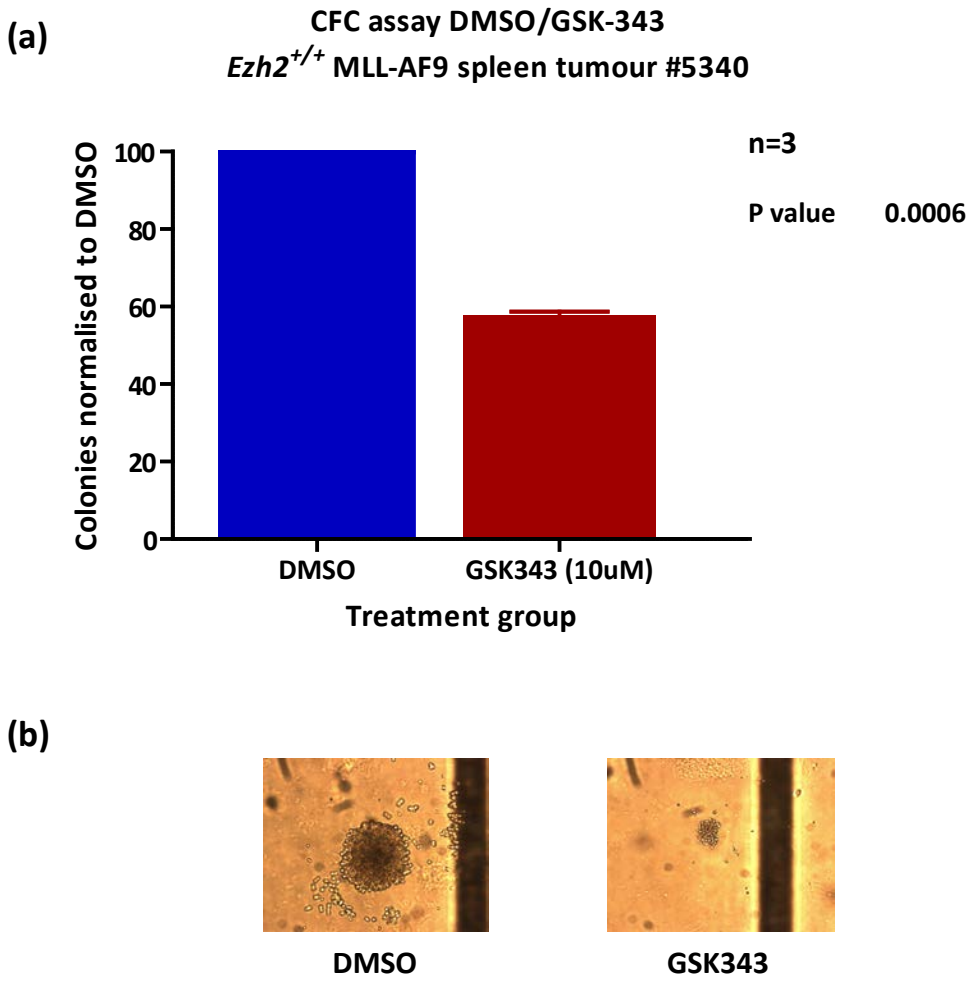


FIGURE 7.12 **GSK343 significantly inhibits clonogenic growth of murine *Ezh2*$^{+/+}$ MLL-AF9 spleen tumour *in vitro***

(a) The total number of colonies formed following treatment with GSK343 at 10μM is significantly reduced compared to DMSO treatment at Day 7 in methylcellulose colony forming assay (data is normalised to total colonies produced in DMSO), performed in triplicate

(b) Reduction in colony size and density on GSK343 treatment (x20 magnification)

7.9 Murine *Ezh2fl/fl*;wt MLLAF9 spleen tumours demonstrate loss of H3K27me3 as early as 24 hours post GSK343 treatment

In order to further characterise the sensitivity of these murine *Ezh2*$^{+/+}$ MLL-AF9 tumours (#5340 and #5342) to GSK343 and to assess whether it was due direct inhibition of Ezh2 enzymatic function, an assessment of changes in H3K27me3 following GSK343 treatment of each tumour in liquid culture was made using flow cytometry. Briefly, both tumours were maintained in liquid culture in sterile conditions in the presence of either DMSO or GSK343 at 10μM. At 24, 48 and 72 hours, cells were harvested from each condition for fixation and permeabilization. Following permeabilization, each condition (DMSO or GSK343 from each tumour #5340 and #5342) was divided into two further conditions and stained for H3K27me3 or total histone H3 (control) using fluorochrome conjugated antibodies (as described in Methods 2.4.2 and 2.4.2.1). As early as 24 hours, there was a near total loss of H3K27me3 with no changes in total histone H3 (a representative example is shown in Figure 7.13):

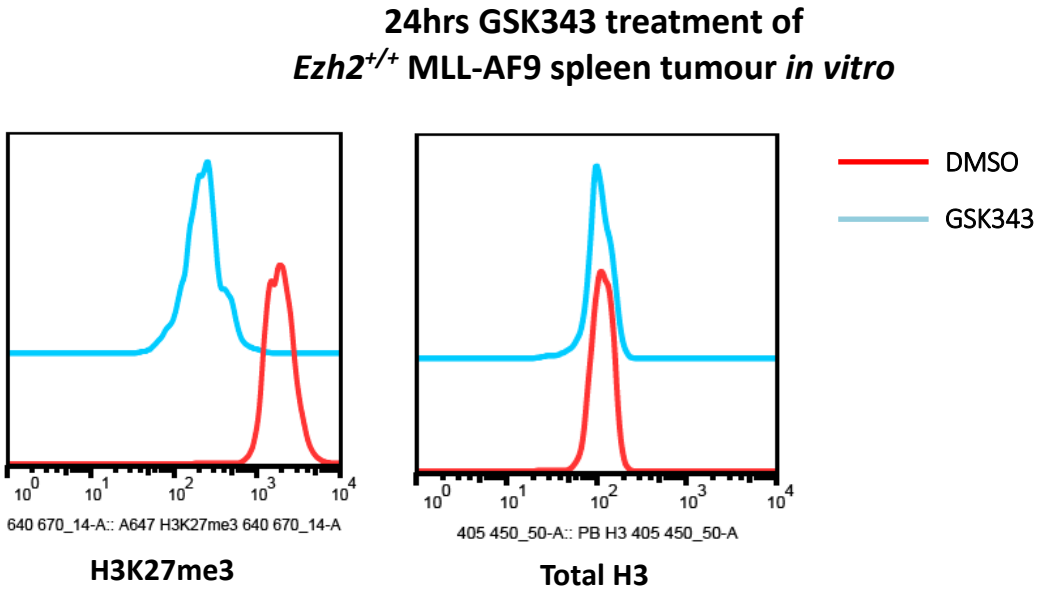


FIGURE 7.13 **GSK343 significantly reduces H3K27me3 in *Ezh2*$^{+/+}$ MLL-AF9 spleen tumour in vitro as early as 24 hours (tumour #5340)**
Histogram representing changes in H3K27me3 with GSK343 treatment
LEFT: Total H3K27me3 (detected on 640 670/14 channel) – Following GSK343 treatment (blue) a near total loss is seen compared to DMSO treatment (red)
RIGHT: Total histone H3 (control, detected on 405 450/50 channel) – no changes seen between DMSO (red) and GSK343 (blue) treatment

This reduction in H3K27me3 demonstrated GSK343 cellular inhibition correlated with and appeared to be functioning through direct inhibition of Ezh2 catalytic function. Both MLL-AF9 tumours exhibited marked sensitivity to the inhibitor, having a significant reduction in total cell number within 24 hours when treated with GSK343 as already described. Therefore, it was not surprising to see that the reduction in H3K27me3 was taking place as early as 24 hours post treatment, though it was earlier than that seen in the Kasumi cell line where H3K27me3 reduction occurred from 48 hours and beyond. This demonstrated that the murine leukaemias were even more sensitive to GSK343 than the human AML cell lines or that the flow cytometric assay was more sensitive to levels of H3K27me3.

As described in Chapter 5, having established that this sensitivity at 24 hours was possibly secondary to an active reduction in H3K27me3, it was felt that any resultant gene expression changes would become apparent as early as this time point. Therefore, RNA was harvested from both tumours treated with DMSO and GSK343 at 24 hours and sent for RNA-sequencing to analyse which transcriptional programmes or tumour suppressive genes were potentially being de-repressed through this pharmacological Ezh2 inhibition and H3K27me3 loss, to account for this marked sensitivity.

It would be useful to adopt the same technical approach and test *Ezh2*$^{+/+}$ AML1-ETO9a murine tumour cells in the same way to corroborate the same level of sensitivity and loss of H3K27me3 and this work is being planned.

7.10 Clonogenic growth of AML patient samples is inhibited by GSK343 treatment

GSK343 was tested against primary AML samples collected by the Huntly lab (Methods 2.2.3). 80,000 mononuclear cells were plated in duplicate in methylcellulose containing appropriate human cytokines plus either DMSO or GSK343 (10μM) and incubated for 7-10 days, following which colony numbers were counted (Methods 2.3.2). Colony numbers in GSK343 treated samples were normalised to paired DMSO samples. Despite being a relatively small sample size (n=14), there was a significant reduction in average colony numbers when treated with GSK343 compared to DMSO (Figure 7.14):

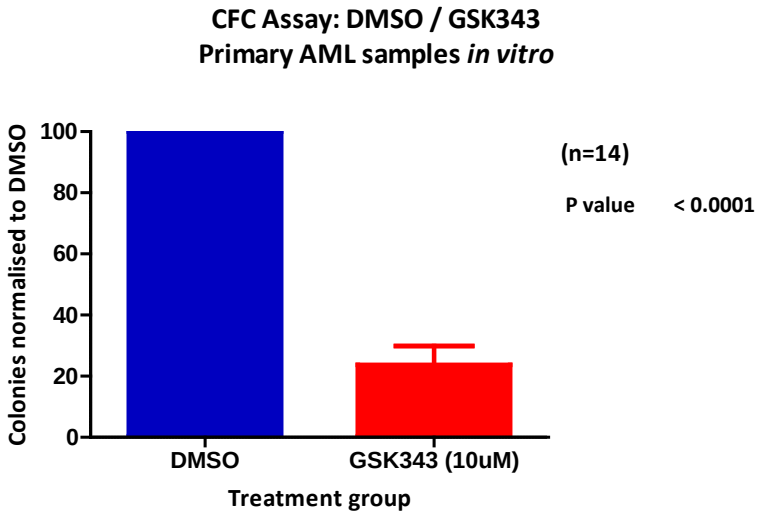


FIGURE 7.14 **GSK343 significantly inhibits clonogenic growth of primary human AML samples *in vitro***

Colonies counted D7-10 after seeding 80,000 cells/ml in methylcellulose.
The total number of colonies formed following treatment with GSK343 at 10μM is significantly reduced compared to DMSO treatment at Day 7 in methylcellulose colony forming assay (data is normalised to total colonies produced in DMSO), performed in duplicate for each sample (n=14).
An average reduction in colony production of ~75% is seen in GSK343 treated samples

Due to the brittle nature of colonies (even when supported in methylcellulose specifically optimised to support human HSPCs) formed from mononuclear cells isolated from whole blood or bone marrow of newly diagnosed AML patients, it was only possible to assess this effect at 7-10 days post seeding – multiple weekly replatings were not possible. The 14 cases that this assay was performed against were unselected, in order to contain a range of ages and variety of karyotypes across both sexes (Table 7.1):

Patient	Sex	Age	Karyotype/Cytogenetic abnormality
1	M	83	N/A
2	F	45	Trisomy 15
3	F	59	N/A
4	M	74	46, XY
5	M	21	Monosomy 1p and trisomy 8q
6	M	65	46, XY
7	F	70	46, XX
8	F	70	Complex
9	M	66	Trisomy 8 in 2 of 22 cells (47, XY, +8[2]/46, XY[20])
10	F	70	46, XX
11	F	46	46, XX
12	F	56	Complex
13	M	63	APML - PML/RAR
14	M	47	MLL-AF9

TABLE 7.1 **Primary AML samples treated with GSK343 - patient characteristics**
Sex, age and karyotype or cytogenetic abnormality associated with each case is detailed. 'N/A' indicates samples where karyotyping failed.
Clonogenic assay results for these patient samples is shown in Figure 7.12

Interestingly, Patient 14 had the MLL-AF9 translocation detected on cytogenetic testing and its colony formation assay showed an approximately four-fold reduction in colony numbers when treated with GSK343 compared to DMSO. This was in keeping with the sensitivity we had seen of the murine MLL-AF9 tumours to GSK343. Examples of reduction in colony size seen in primary AML samples following GSK343 are shown in Figure 7.15:

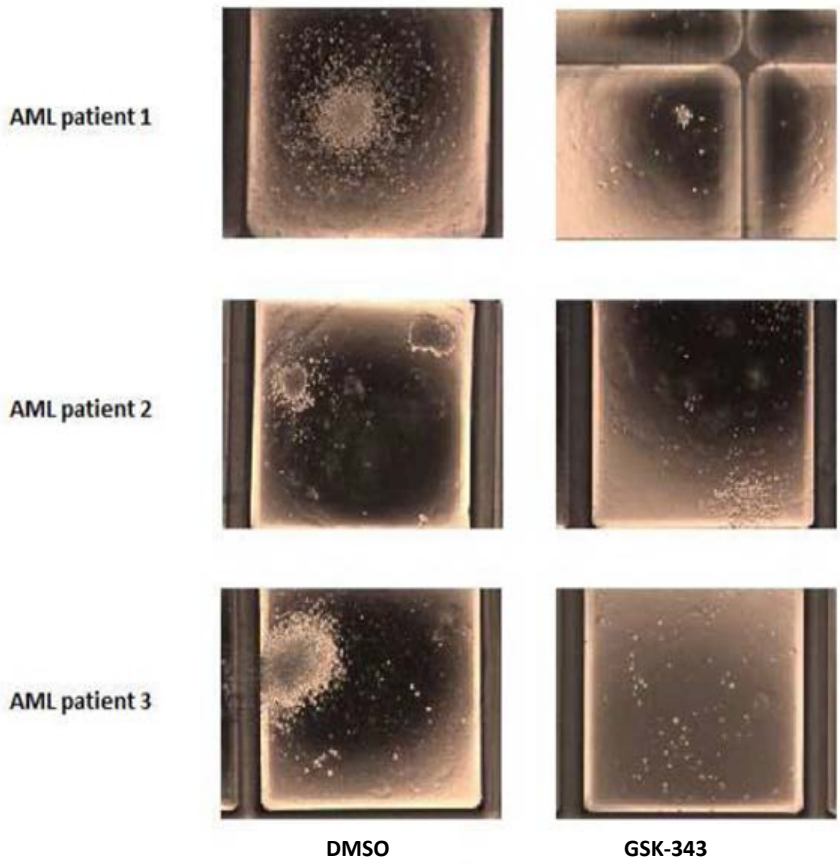


FIGURE 7.15 **Marked reduction in colony size and density seen in the when unselected primary human AML samples are treated with GSK343 in colony formation assay (compared to DMSO)**
x20 magnification. 3 patient samples shown

7.11 GSK343 does not significantly eliminate colony formation by normal CD34+ control cells

Although other EZH2 inhibitors are now in Phase I clinical trials, we wanted to assess whether the effect of EZH2 inhibition by GSK343 was leukaemia specific as opposed to secondary to general toxicity to primary haematopoietic cells *in vitro*. Primary human CD34+ cells were used as a control. These were harvested through GCSF-stimulated apheresis from patients with non-Hodgkin's lymphoma/myeloma in clinical remission prior to autologous stem cell transplantation (with consent as per Methods 2.2.3). Following harvesting these samples were subjected to magnetic bead separation to further purify and obtain in the order of >90% pure CD34+ cells. These purified CD34+ samples were already present in the Huntly lab and were used, amongst other things, to assess haematopoietic toxicity to novel inhibitor compounds.

An assessment of toxicity secondary to GSK343 was made at 10μM and 20μM. Though a very small sample size (n=2), the results suggest that at 10μM no significant toxicity is present, though considerable toxicity is seen at 20μM (Figure 7.16):

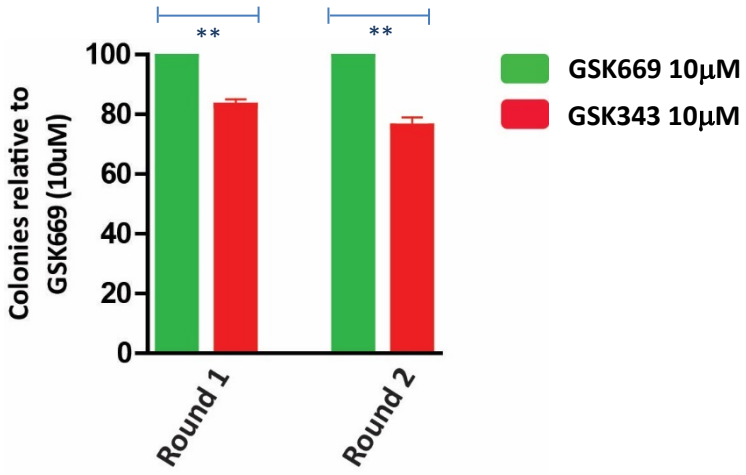


FIGURE 7.16 **GSK343 at 10μM does not exhibit a significant effect on normal CD34+ cells**
GSK343 versus normal CD34+ HSPCs, results normalised to GSK669 colony numbers. Colonies counted D7-10 after seeding 50,000 cells/plate in methylcellulose
At 10μM, GSK343 treatment does not significantly reduced colony growth in the first and second round of replating.
** = p-value indicates no significance between GSK669 and GSK343 at 10μM
Error bars represent the SEM. (n=2 samples)

7.11 Discussion and future work

These data demonstrate that GSK343 appears to have a wide range of efficacy against some human AML cell lines, but a larger number of murine cell lines immortalised with AML driver oncogenes, murine MLL-AF9 leukaemias cultured *in vitro* and, most importantly, unselected primary AML samples. Having demonstrated that loss of *Ezh2* was capable of accelerating AML1-ETO9a and MLL-AF9 leukaemogenesis *in vivo* we felt there were valid concerns about the use of EZH2 inhibitors in AML. However we were greatly reassured by our genetic data that *Ezh2* loss antagonised the maintenance of the same leukaemias *in vivo.* We used this as a surrogate to postulate that EZH2/Ezh2 pharmacological inhibition with GSK343 would be efficacious and inhibitory toward leukaemic cell line/tumour growth and expansion *in vitro*.

We were able to formally demonstrate that GSK343 caused effective inhibition of expansion of the Kasumi and KG1 cell lines, with the effect on the former being very interesting given it is driven by the AML-ETO *RUNX1-RUNX1T1* translocation. This corroborated our demonstration that genetic *Ezh2* loss was inhibitory to AML1-ETO9a transformed cells *in vitro* and AML1-ETO9a driven secondary leukaemias *in vivo*. At no point did we see an acceleration or augmentation of cellular growth and expansion in the Kasumi cell line when treated with GSK343. We were able to demonstrate that GSK343 was causing a clear reduction in total H3K27me3 in the Kasumi (and KG1/NB4) cell line. This again demonstrated the oncogenic effect that EZH2 was likely exerting in the maintenance of this AML-ETO driven cell line, further strengthening our argument that intrinsic EZH2 expression (through the maintenance of the repressive H3K27me3 mark) likely has negative influence over a different and presumably tumour-suppressive transcriptional programme in the maintenance of AML.

Importantly, *Ezh2fl/fl*;wt MLL-AF9 (and MOZ-TIF2) transformed murine cell lines exhibited marked sensitivity to GSK343, exhibiting a clear reduction in clonogenic potential, and two *Ezh2*$^{+/+}$; MLL-AF9 murine primary leukaemias whose spleen tumour tissues were cultured *in vitro* demonstrated a distinct sensitivity to GSK343 at a consistent 10μM concentration. We were able to demonstrate the reduction in clonogenic potential, liquid culture expansion as early as 24 hours and that this appeared to be secondary to a near total loss of H3K27me3 as early as 24 hours following GSK343 treatment. This suggested that GSK343 was exerting its effects through direct epigenetic changes, making a formal assessment of differentially regulated genes in the MLL-AF9 leukaemias following Ezh2 inhibition possible (Chapter 5). As well as doing this, it provided insight and corroborated our demonstration that genetic *Ezh2* loss in established MLL-AF9 leukaemias delayed their progression *in vivo*. Once again there was no indication that Ezh2 inhibition had any positive effect on the expansion of these MLL-AF9 transformed murine cell lines and tumours *in vitro*.

Pharmacological inhibition of Ezh2 therefore appears pre-clinically effective in AML-ETO and MLL-AF9 driven leukaemias, albeit that we have not demonstrated this *in vivo*. It is remarkable to see that EZH2/Ezh2 appears to be exerting an oncogenic function in the maintenance of two very different subtypes of AML. MLL-AF9 is a very aggressive leukaemia model *in vitro* and *in vivo* that transforms through the activation of genetic programmes and in humans the translocation is associated with an intermediate-poor prognosis. In contrast the AML1-ETO9a is a less aggressive leukaemia model both *in vitro* and *in vivo* that is described to function through the repression of genetic programmes. Additionally, its translocation in human AML is associated with a good prognosis. Why two such disparate drivers of AML, acting through differing programmes, behave in the same way following genetic or pharmacological targeting of Ezh2/EZH2 is of considerable interest and warrants further characterisation. A starting point would be the comparison of differential gene changes following pharmacological Ezh2 inhibition of the MLL-AF9 murine leukaemias *in vitro* (already performed) and that of AML1-ETO9a murine leukaemias similarly cultured and treated *in vitro* (not yet performed), which we suggest would provide further mechanistic insights into this sensitivity. Additionally, in the genetic and pharmacological experiments of *in vitro* MOZ-TIF2 maintenance and *in vitro* MOZ-TIF2 immortalised cells maintenance respectively, *Ezh2* deletion / Ezh2 inhibition clearly antagonised the maintenance of the leukaemia. Though clinically less common, the MOZ-TIF2 leukaemia programme is also an aggressive one, associated with poor prognosis and acts, as MLL-AF9, through the activation of genes.

Taken together, these results demonstrate that, in the maintenance phase of disease for a wide range of AML, targeting Ezh2/EZH2 is a viable therapeutic option.

When this assessment was extrapolated to primary human AML samples *in vitro*, we saw a pronounced inhibitory effect on clonogenic potential across a wide variety of AML samples that appeared independent of sex, age and cytogenetic abnormality. Once again, following EZH2 inhibition, at no point did we see any augmentation of clonogenic or proliferative ability across any of these primary samples. Though this cohort did not include any t(8;21) *RUNX1-RUNX1T1* or inv(8)(p11q13) *MOZ-TIF2* translocated samples, it did include a primary sample from a patient with *MLL-AF9* translocation, which exhibited marked sensitivity to GSK343. It would be useful to expand this further using more AML samples especially ones from patients with the aforementioned translocations (currently ongoing at time of writing). Additionally, we saw only low levels of toxicity from GSK343 to normal haematopoietic cells, though this was in a very limited number and ideally requires further work in more normal CD34+ samples to consolidate this.

Taken together all this data, as discussed above, provides preclinical evidence that EZH2 inhibition in established human AML (i.e. the maintenance phase of the disease) is a plausible therapeutic option and should be free of concerns that disease may be accelerated as the latter occurs in a different and independent phase (induction) of disease.

References

1) The Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 2013; 368: 2059-74

2) Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, Potter NE, Heuser M, Thol F, Bolli N, Gundem G, Van Loo P, Martincorena I, Ganly P, Mudie L, McLaren S, O'Meara S, Raine K, Jones DR, Teague JW, Butler AP, Greaves MF, Ganser A, Döhner K, Schlenk RF, Döhner H, Campbell PJ. Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med.* 2016 ;374(23):2209-21

3) Mochizuki-Kashio M et al. Dependency on the polycomb gene *Ezh2* distinguishes fetal from adult hematopoietic stem cells. *Blood* **(**2011)

4) McCabe MT, Ott HM, Ganji G, Korenchuk S, Thompson C, Van Aller GS, Liu Y, Graves AP, Della Pietra A 3rd, Diaz E, LaFrance LV, Mellinger M, Duquenne C, Tian X, Kruger RG, McHugh CF, Brandt M, Miller WH, Dhanak D, Verma SK, Tummino PJ, Creasy CL. EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature*, 2012 ;492(7427):108-12

5) Sharad K. Verma, Xinrong Tian, Louis V. LaFrance, Celine Duquenne, Dominic P. Suarez, Kenneth A. Newlander, Stuart P. Romeril, Joelle L. Burgess, Seth W. Grant, James A. Brackley, Alan P. Graves, Daryl A. Scherzer, Art Shu, Christine Thompson, Heidi M. Ott, Glenn S. Van Aller, Carl A. Machutta, Elsie Diaz, Yong Jiang, Neil W. Johnson, Steven D. Knight, Ryan G. Kruger, Michael T. McCabe, Dashyant Dhanak, Peter J. Tummino, Caretha L. Creasy, and William H. Miller. Identification of Potent, Selective, Cell-Active Inhibitors of the Histone Lysine Methyltransferase EZH2. *ACS Medicinal Chemistry Letters*. 2012; 3, 1091-6.

Chapter 8

Summary and conclusions

The unifying theme of the work detailed in this thesis is the multifaceted, heavily context-dependent role that EZH2, a key epigenetic regulator, plays in acute myeloid leukaemia.

The dysregulation of epigenetic mechanisms across AML is a widely-recognised feature and as more detailed functional and genomic analyses are being performed to better characterise these, the greater the potential for further classifying disease and developing novel therapies to treat this aggressive cancer will become. The molecular processes that govern haematopoiesis are controlled by lineage-specific transcription factors which are, in turn, regulated by and contribute to a higher level of epigenetic control. Through direct surveillance and modification of chromatin, epigenetic readers, writers and erasers are able to control changes in chromatin structure and gene transcription. Mutations in histone modifiers are increasingly implicated in haematological malignancies, especially across AML, and examining these closely at the molecular level allows for a better understanding of how epigenetic dysregulation may be targeted for therapeutic benefit.

In this thesis, data addressing the role of EZH2 in various phases of AML evolution with perturbations performed both genetically and pharmacologically, have been presented. Through the use of three markedly different AML driver oncogenes, via retroviral transduction/overexpression and bone marrow transplantation assays on the backdrop of a conditional *Ezh2* knock-out murine model, it was possible to clearly demonstrate that *Ezh2* has entirely opposite roles dependent upon subtype and phase of disease. The purposeful use of three separate models of AML, clinically associated with a range from good and intermediate to poor prognosis, each with markedly different mechanisms of leukaemic transformation, allowed for a better experimental representation of the considerable heterogeneity seen at the cytogenetic and molecular level within human AML. Moreover, the characterisation of *Ezh2*'s role in the early stages of leukaemogenesis – a critical period of disease development - provided a better understanding of specific transcriptional programmes that were being de-repressed in its absence and contributing to the significant alterations in disease outcome. By thoroughly examining this and then contrasting to its function in AML maintenance – the therapeutically significant phase of disease – an additional layer of understanding into the complexities of how this transcriptional repressor exerts a variety of effects was provided.

In Chapter 3 it was demonstrated that whilst not providing any strong influence *in vitro*, *Ezh2* loss led to unexpected and varying outcomes in the induction of leukaemia *in vivo* across three subtypes of

AML. Whilst attaining a significant 50% increase in median survival in the MOZ-TIF2 experiment, *Ezh2* loss served to considerably accelerate the functionally repressive and less aggressive AML1-ETO9a leukaemia programme *and* accelerate the considerably more aggressive, short-latency MLL-AF9 leukaemia programme *in vivo.* How the manipulation of one gene is able to induce such disparate outcomes between three heterogenous subtypes of AML *in vivo,* serves to portray how remarkably complex this epigenetic modifier is. Whilst not examined further in this thesis, due to focusing on the concern that *Ezh2* loss clearly accelerated the induction of the more clinically relevant MLL-AF9 and AML1-ETO9a leukaemias, the observation that following *Ezh2* loss dramatic slowing of MOZ-TIF2 leukaemogenesis (whose mechanisms of transformation are felt to be similar in some ways to MLL-AF9) was also important, suggesting it has an oncogenic role within this subtype, but a tumour-suppressive role within the other two, during the same phase of disease. This clearly warrants further investigation, perhaps with the starting point of assessing changes in gene expression in these induction MOZ-TIF2 leukaemias and comparing with the non-transformed, MLL-AF9 and AML1-ETO9a gene changes. The hypothesis would be (as per the maintenance experiments) that the continued presence of *Ezh2* is required for efficient transformation by MOZ-TIF2.

The recent recognition of mutations of epigenetic regulators, including chromatin modifiers such as *EZH2*, as being associated more exclusively with the t(8;21) *RUNX1-RUNX1T1* translocated subset of core-binding factor leukaemias, suggests a plausible co-operative and mechanistic link between the two, as it did in the AML1-ETO9a induction experiment. It would be important to examine larger AML datasets for the co-occurrence of *EZH2* mutations in t(8;21) mutated cases, and ideally, an assessment of variant allelic frequencies to inform regarding order of mutation. Additionally, a comparison of differences in survival outcomes of *EZH2* mutated versus *EZH2* wild-type cases and the same assessment specifically within AML1-ETO rearranged cases (to determine if *EZH2* status alters outcomes) would be highly interesting. The suggestion from the *in vivo* work was clearly that *Ezh2* loss was a co-operative event in AML1-ETO9a murine leukaemogenesis and adds weight towards speculation that *EZH2* mutation in t(8;21) translocated human leukaemias may occur preceding or at the time of transformation. The tumour suppressive effect of *Ezh2* in the induction of *JAK2*-V617F driven MPN has been well characterised recently (with some of its downstream mediators), and its loss led to a synergism with *JAK2*-V617F in accelerating disease development, supporting the data in this thesis that *Ezh2* has a tumour suppressive role in myeloid malignancy, perhaps more so in its development than its maintenance or progression.

Ezh2 loss in established, immortalised murine haematopoietic cell lines for all three driver oncogenes assessed in the induction experiments, led to a complete abolishment of growth *in vitro*. The possibility

that this impressive outcome was secondary to Cre-recombinase toxicity was subsequently excluded. *In vivo* experiments assessing loss of *Ezh2* during the maintenance of AML, whilst showing no significant effect on secondary MOZ-TIF2 leukaemia progression, demonstrated a clear increase in survival for both AML1-ETO9a and MLL-AF9 secondary leukaemias. This suggested *Ezh2* was exhibiting an oncogenic role in AML maintenance, facilitating the transformed state, through potentially repressing tumour-suppressive elements via its role as a transcriptional repressor. This was backed up by preceding studies that had demonstrated similar findings in the maintenance of MLL-AF9 driven AML in mice. Additionally, in demonstrating *Ezh2* had oncogenic function in AML maintenance, further evidence for pursuing EZH2 as a therapeutic target in AML was provided, leading to functional experiments with an EZH2 inhibitor to assess efficacy against murine and human AML cell lines, murine leukaemias and primary AML samples. Importantly, the outcomes in the maintenance experiments were in stark contrast to its apparent tumour-suppressive function in the same two AML subtypes during leukaemogenesis. The realisation that one gene was able to have such pronounced and opposing effects at two different phases of the same two subtypes of disease was an incredibly interesting one and prompted a detailed genomic analysis to explain this 'Jekyll and Hyde' behaviour.

Initially focusing on the tumour suppressive nature of *Ezh2*, with its loss facilitating progression from the normal state to overt AML1-ETO9a or MLL-AF9 leukaemia, a characterisation of differential gene expression changes before and following *Ezh2* loss in the normal state was made. Consistent with its role as a transcriptional repressor, deletion of *Ezh2* in normal HSPCs led to the upregulation of significantly more genes than those downregulated (in the order of 9:1). When gene expression changes between both $Ezh2^{+/+}$ and $Ezh2^{-/-}$ AML1-ETO9a and MLL-AF9 leukaemias were characterised in the same way, a similar though less pronounced shift in balance towards de-repressed genes was revealed, suggesting in the transformed state, the effects of *Ezh2* loss were being diluted and overridden by the leukaemia programmes themselves. This change was least marked in the differentially expressed genes for the MLL-AF9 leukaemias, where the balance between upregulated and downregulated genes was close to 1.5:1, perhaps reflecting the more aggressive nature of this disease or the effects of an activating oncogenic programme. Given these changes appear to vary depending upon the oncogene, it would be interesting to examine changes in binding patterns of each of these in the context of *Ezh2* loss or reduced H3K27me3. Nonetheless, a number of potential oncogenic candidate drivers of disease presented themselves when the upregulated genes for all three cohorts were overlapped. *LIN28B*, a well-documented candidate known to repress the let-7 microRNA tumour-suppressor family, with a clear correlation to poor prognosis and outcomes across a range of malignancies was highlighted in these gene expression datasets. It was the gene with the highest ranked $\log_2$fold change between the normal and MLL-AF9 leukaemia states. The recent documentation

of *Lin28b* and its potential downstream regulators, including *Hmga2, Igf2bp3* as oncogenic facilitators of *Ezh2* loss in collaborating with *JAK2*-V617F mediated MPN further served to fuel interest in exploring this candidate. Whilst not quite as highly upregulated as *Lin28b*, this overlapping analysis also revealed *Plag1* as another plausible oncogenic mediator of *Ezh2* loss. Plag1 is a zinc-finger transcription factor that has been implicated as a collaborator in the pathogenesis of *Cbfb-MYH11* translocated murine AML *in vivo* – the equivalent of the inv(16) *CBFβ-MYH11* core-binding factor leukaemias in human AML. *PLAG1* has been documented as being associated with a number of other benign tumours and has been demonstrated to have increased levels of expression across AML subtypes. In these analyses, *Plag1* stood out (from a limited group of six genes) as being the most likely potential oncogene mediator de-repressed upon *Ezh2* loss and was commonly upregulated in all three groups – non-transformed, AML1-ETO9a and MLL-AF9.

Mechanistically, further genomic analyses informed the potential regulation of these genes by *Ezh2*. Through characterising global changes in H3K27me3 following *Ezh2* deletion utilising ChIP-sequencing in the normal non-transformed state, it was possible to identify genes that were being directly upregulated as a result of loss of *Ezh2* enzymatic activity. Approximately 20% of all genes upregulated were associated with a reduction in H3K27me3, suggesting alternative unexplored and potentially indirect mechanisms by which the remaining 80% were being upregulated. However, within this smaller subset again were *Lin28b, Plag1* and *Igf2bp3* amongst others. Characterisation of changes in the H3K27Ac activating mark, generally present at enhancer regions, did not reveal any significant insights into the upregulation of these genes. However, genome wide analysis of the H3K4me3 activating mark, associated primarily with promoter regions, revealed a significant proportion of the subset of genes upregulated after *Ezh2* loss AND which demonstrated a loss of H3K27me3 at their promoter, also demonstrated the presence of H3K4me3 at their promoter – suggesting these were bivalent domains – held in a poised transcriptional state by two conflicting histone marks, and that *Ezh2* loss was shifting the balance in favour of gene activation at least through loss of H3K27me3. When probed further however, it became apparent that *Ezh2* loss was also causing a significant upregulation of H3K4me3 at these bivalent domains for some, but not all of this subset of genes – including the promoters of *Lin28b* and *Plag1*. This surprising finding, that changes in H3K27me3 resulting directly from alterations in the enzyme responsible for its deposition were also capable of either directly or indirectly modulating H3K4me3 was highly interesting, and warrants further work to determine what is unique about these particular genes and how *Ezh2* loss might mechanistically lead to an upregulation of H3K4me3 at their promoters. A recent study assessing the UTX protein, which functions as a histone demethylase for H3K27me3 (the opposite function to EZH2), has shown that its loss, although in the different context of variant natural killer cells, results in increased H3K27me3 and

the downregulation of H3K4me3 again at particular subsets of genes, corresponding to the downregulation of those genes. This suggests that changes in H3K4me3 status are being directly regulated in some way by fluxes in H3K27me3. The complex interplay or 'cross-talk' between these conflicting histone marks and the enzymes controlling their deposition and removal is highly interesting, even in the normal state, and unravelling this might further inform how *Ezh2* loss fosters an epigenetic state that is co-operative for AML transformation. Analysing changes in H3K27me3 and H3K4me3 in the $Ezh2^{+/+}$ and $Ezh2^{-/-}$ AML1-ETO9a and MLL-AF9 leukaemias and comparing with that in the normal state would be a starting point.

Having identified plausible downstream effectors of *Ezh2* loss and building a strong case that either or both of these were potentially facilitating leukaemogenesis, it was then imperative to functionally validate them *in vivo*. Through cloning, both *Lin28b* and *Plag1* were inserted into the murine stem cell virus (MSCV) retroviral construct under the control of the IRES promoter, and therefore could be overexpressed upon retroviral transduction. Whilst potentially contributing to both AML1-ETO9a and MLL-AF9 leukaemogenesis, it was decided that they would be functionally validated in the MLL-AF9 setting, given the shorter latency of wild-type MLL-AF9 disease *in vivo*. Through co-expression of *Lin28b* or *Plag1* with MLL-AF9 using the same retroviral transduction / overexpression and bone marrow transplantation assays used in the preceding *Ezh2* induction and maintenance experiments, it was possible to address the question of whether either of these oncogenes would accelerate MLL-AF9 leukaemogenesis *in vivo.* Whilst ectopic expression of *Lin28b* failed to accelerate MLL-AF9 compared to the wild-type arm (even though there was an early trend that it was doing so), strikingly, overexpression of *Plag1* in concert with MLL-AF9 led to a dramatic increase in the rate of leukaemia development, with an almost 30% reduction in median survival, without obviously altering phenotype (currently being analysed at time of writing). The fact that overexpression of a single gene, amongst the potent milieu of transcriptional changes driven by such an aggressive leukaemia programme, was able to exert its effect enough to phenocopy *Ezh2* loss and significantly alter survival outcome, was extremely striking. Speculation as to how these two might collaborate to induce AML faster potentially implicates the IGF2 pathway – a known target of *PLAG1*, though requires further studies to confirm this hypothesis. Though *Lin28b* overexpression alone was unable to phenocopy *Ezh2* loss, given its profound upregulation in both normal and MLL-AF9 settings after *Ezh2* deletion and the fact that it in turn influences the upregulation of a number of potential oncogenic candidates (surprisingly including *Plag1* itself, from the qPCR analysis following *Lin28b* transduction), it remains likely to play a role. The synergism between *Ezh2* loss and MLL-AF9 or AML1-ETO9a likely results from the concerted action of several genes that are altered in their expression, with this data indicating, for MLL-AF9 at least, that *Plag1* is one of these downstream targets. Further work is required to assess if overexpression of *Plag1*

is capable of augmenting AML1-ETO9a leukaemogenesis as well as MLL-AF9, ideally *in vivo*, though given the long latency of wild-type disease, it would be more feasible to initially attempt this *in vitro* - and therefore this is being done at the time of writing. If borne out across both leukaemia subtypes, given their significant differences in mode of leukaemia induction, this *Ezh2-Plag1* axis has the potential to be examined in other AML models.

To explain the disparity in roles between induction and maintenance of MLL-AF9 leukaemias, a comparison of differential gene expression, following Ezh2 pharmacological inhibition in established murine MLL-AF9 spleen tumours cultured *in vitro,* was made to the induction differential gene expression dataset. MLL-AF9 spleen tumours were markedly sensitive to the EZH2 inhibitor *in vitro* and this was associated with a significant downregulation of H3K27me3 (as seen in the genetic experiments). Given the opposite outcomes seen in either phase of disease, the prediction was that Ezh2 was exerting control over different, and likely mutually exclusive, transcriptional programmes. Analysis of RNA-sequencing data in this scenario demonstrated differential regulation of a narrower set of genes than in induction, and that again more genes were upregulated than downregulated following Ezh2 inhibition. Despite this, comparing this dataset to genes upregulated following *Ezh2* loss in induction of MLL-AF9 leukaemias revealed virtually no overlap, strongly suggesting that the opposing effects of *Ezh2* loss or inhibition in either scenario could be explained by the fact that *Ezh2* loss was resulting in the de-repression of a completely different gene expression set in either context. This demonstration provided reassurances that inhibition of EZH2 in the context of established AML – which is therapeutically more relevant – was unlikely to be associated with adverse effects on leukaemia progression. It should be noted however, that these two experimental conditions cannot be fully compared, in that they represent Ezh2 perturbation in the context of different cellular states with one utilising full genetic ablation (induction), whilst the other utilises reversible chemical inhibiton (maintenance).

To address the question of whether EZH2 might be a valid target in AML, experiments designed to preclinically test the EZH2 inhibitor GSK343 further were performed. Both the Kasumi and KG1 human AML cell lines exhibited sensitivity to GSK343 in liquid culture growth and demonstrated reduced clonogenic and proliferative potential in colony formation assays in the presence of drug. Further analysis revealed that in both cell lines, GSK343 treatment resulted in a modest increase in apoptosis, as late as 96 hours post treatment, and that there was a mild induction of G_1 cell cycle arrest at around the same time point. The sensitivity displayed by the Kasumi cell line to GSK343 was of particular interest, given it harbours the t(8;21) *RUNX1-RUNX1T1* translocation – equivalent to the AML1-ETO9a translocation oncogene employed in the murine experiments that demonstrated a clear sensitivity to

Ezh2 loss both *in vitro* and *in vivo*. Having established a link between the murine equivalent and human t(8;21) sensitivities to *Ezh2* loss/EZH2 inhibition respectively, this fortified the idea that, whilst detrimental in the leukaemogenesis/induction pathway, EZH2 inhibition may play a role in established t(8;21) translocated leukaemias.

Ezh2 inhibition with GSK343 appeared to have similar inhibitory effects on established immortalised MLL-AF9 and MOZ-TIF2 murine cell lines *in vitro*, and when taken further by assessing this efficacy on *in vivo* generated MLL-AF9 leukaemias whose spleen tumours were cultured *in vitro*, once again a clear sensitivity to Ezh2 inhibition was demonstrated. GSK343 reduced expansion of these tumours in liquid culture, decreased clonogenic and proliferative potential, and was associated with a clear loss of H3K27me3 at a relatively early time point after treatment. As with AML1-ETO9a, these findings were consistent with the maintenance *in vitro* and *in vivo* experiments demonstrating MLL-AF9 sensitivity to *Ezh2* loss.

Finally, extending this work further, the effects of EZH2 inhibition on human AML were modelled through assessing the sensitivity of a number of primary human AML samples to GSK343 in colony formation assays. Across a range of unselected AML cases, including one case harbouring the MLL-AF9 translocation, it was possible to demonstrate a considerable degree of sensitivity through reduced clonogenic and proliferative potential. Interestingly the sensitive AML cell lines, immortalised murine cell lines, MLL-AF9 murine tumours had IC_{50} measurements in the 10-20μM range and the sensitivities demonstrated across these range of samples (including the primary samples) were all at the 10μM concentration. At no point did GSK343 appear to exert a positive effect on growth or expansion across these murine and human samples. No significant toxicity was seen when GSK343 was tested against a limited number of normal human CD34+ cells in the same assay. Further testing against more normal and AML samples (ideally with the latter containing t(8;21) or *MOZ-TIF2* translocated cases) are warranted and will be conducted in the near future. Furthermore, it would be interesting to consider a drug trial in mice either with MLL-AF9 or AML1-ETO9a driven leukaemia to see if the effects of pharmacological inhibition could mimic that seen with genetic deletion in the maintenance experiments.

The work set out in this thesis answers its initial aim through demonstrating that *Ezh2* is differentially required for the induction and maintenance of AML *in vitro* and *in vivo*. This novel finding appears to be the first description of a single gene possessing significantly disparate behavioural requirements (i.e. tumour suppressor versus oncogenic roles) between different phases of the evolution of subtypes of a single disease. Given its role as an epigenetic regulator, controlling transcription at a potentially genome-wide level, this dramatic difference in role is inherently plausible. The work herein provides

mechanistic explanations for how loss of this gene is capable of accelerating AML induction, through detailed analysis of alterations in transcription and at the chromatin level, providing some novel observations regarding changes in histone modifications that result in the de-repression of genes. It then furthers this by functionally assessing plausible candidates and validating a single target, establishing a novel axis to explain the tumour suppressive behaviour seen in AML induction. To explain the reversal in role regarding the phenotypic outcomes seen in the maintenance of AML *in vitro* and *in vivo,* it then provides an assessment of transcriptional changes in this setting, demonstrating that they are clearly different from that seen in the induction programmes, allaying concerns regarding the therapeutic targeting of EZH2. Finally, through preclinical studies, this work provides uniform evidence that suggests there is scope in therapeutically targeting EZH2 in established AML. It will be interesting to see if further preclinical data to support these observations appears, and whether future Phase I/II clinical trials of EZH2 inhibitors would subsequently expand their scope from solid tumours and B-cell lymphomas to include AML cases.

The data presented in this thesis has been a detailed characterisation of the role of EZH2 in AML and whilst fulfilling the aims set out in the Introduction chapter, has also created new questions that require further exploration:

Is it possible to reconcile the increase in survival seen following *Ezh2* loss in the induction of MOZ-TIF2 leukaemias with the opposite reduction in survival seen in MLL-AF9 and AML1-ETO9a leukaemia during the same phase of disease evolution? Is there any overlap with changes in transcriptional programmes in this setting for MOZ-TIF2 induction to that seen in the maintenance experiments?

Does the *Ezh2-Plag1* axis also play a similar role in AML1-ETO9a leukaemogenesis? Through which pathways does *Plag1* mechanistically act to accelerate MLL-AF9 leukaemia development? In larger patient datasets is there a direct link between *EZH2* mutations and t(8;21) or MLL-AF9 translocated human leukaemias with any influence on survival outcomes? Do *PLAG1 or LIN28B* expression levels have any association with this?

Through what mechanisms does *Ezh2* loss in normal HSPCs lead to an upregulation of H3K4me3 at a limited set of gene promoters, and why are these restricted? Is this directly dependent upon changes in H3K27me3? Do changes in these facilitate a state that is more prone to transformation in AML? What are the associated changes in H3K27me3 and H3K4me3 in the AML-ETO9a and MLL-AF9 induction leukaemias in the absence of *Ezh2,* and do they overlap those seen in the non-transformed state?

Answering these questions (with some experiments to address these being performed at the time of writing) would further this work and provide important insights into EZH2's direct link to human AML, better defining its enigmatic and apparent dual nature that is so context-specific. Therefore, in summary, this work highlights the importance and inherent complexities of epigenetic dysregulation in AML, through defining the role of the epigenetic regulator EZH2 in different phases of AML evolution, and concluding that despite its dual nature, it remains a viable therapeutic target. Through further preclinical and clinical characterisation, EZH2 inhibition will hopefully be added to a growing repository of therapies in acute myeloid leukaemia, with the intent of impacting upon this aggressive disease whose outcomes continue to remain unacceptable in this era.

Published work relating to this thesis

Basheer F, Huntly BJ. BET bromodomain inhibitors in leukemia. *Exp Hematol.* 2015 Aug;43(8):718-31

Horton SJ, Giotopoulos G, Yun H, Vohra S, Sheppard O, Bashford-Rogers R, Rashid M, Clipson A, Chan WI, Sasca D, Yiangou L, Osaki H, Basheer F, Gallipoli P, Burrows N, Erdem A, Sybirna A, Foerster S, Zhao W, Sustic T, Petrunkina Harrison A, Laurenti E, Okosun J, Hodson D, Wright P, Smith KG, Maxwell P, Fitzgibbon J, Du MQ, Adams DJ, Huntly BJP. Early Loss of CREBBP Confers Malignant Stem Cell Properties on Lymphoid Progenitors. *Nature Cell Biology.* 2017 Sep;19(9):1093-1104

Manuscripts in preparation

Contrasting stage-specific requirements for the histone methyltransferase EZH2 during evolution of Acute Myeloid Leukaemia

Appendix A 5.1

Genes differentially expressed between *Ezh2*$^{+/+}$ and *Ezh2*$^{-/-}$ lineage negative murine HSPCs in the non-transformed setting. Significant gene expression results were those with a p-value adjusted for multiple hypothesis testing of <0.05 with a $\log_2$ fold change of greater or less than 0.5.

Gene ID	$\log_2$ Fold Change	Adjusted p-value	Gene ID	$\log_2$ Fold Change	Adjusted p-value
Lin28b	7.2612	1.10E-10	*Klra4*	3.2902	2.01E-11
Scn4b	5.9047	8.96E-28	*Fbxo2*	3.2685	0.001512777
Pawr	5.7497	1.52E-29	*Cdkn2a*	3.2671	0.022365402
Bmpr1a	5.6969	3.95E-59	*1700001O22Rik*	3.2592	0.004042759
Pcolce2	5.5429	1.66E-13	*Lta*	3.2571	0.005009866
Ackr3	5.0924	0.016653374	*Fbln1*	3.2571	0.005967392
Spats2	4.9315	3.50E-44	*Fgf13*	3.2471	0.000167925
Cttnbp2	4.5789	0.010903684	*Ddx43*	3.2413	2.23E-06
Cma1	4.5477	2.32E-21	*Klra8*	3.2194	1.94E-08
Cx3cl1	4.5347	0.002186949	*A930004D18Rik*	3.2137	0.010015884
Fjx1	4.4137	3.02E-07	*Zswim5*	3.2083	1.23E-10
Adamts20	4.3240	6.76E-05	*Mcpt4*	3.1720	0.009652117
Thbs4	4.3219	0.001302677	*D030025P21Rik*	3.1520	6.54E-05
Syt14	4.2766	0.002186949	*Mpped2*	3.0310	8.89E-10
Yes1	4.2602	0.000247669	*Cd109*	3.0233	0.000128851
Igf2bp3	4.2270	1.05E-30	*6330403A02Rik*	2.9847	8.23E-05
Trnp1	4.2021	0.022904117	*Igkv13-84*	2.9838	0.000119993
Vpreb1	4.0931	3.86E-21	*Wtip*	2.9444	0.042280005
Gfra2	4.0801	8.25E-06	*Homer2*	2.9347	0.002498901
Dsp	4.0644	0.008409518	*Fbln1*	2.9300	0.041344594
Ddah1	4.0147	8.07E-13	*Gm30948*	2.8837	3.34E-07
Cnn1	3.9225	0.001753485	*Jph1*	2.8431	3.82E-06
Igll1	3.8967	1.51E-18	*Usp44*	2.8349	1.87E-12
Itga7	3.8774	0.037784092	*Myom3*	2.8263	0.048648156
Dscam	3.8103	0.000287645	*Kank1*	2.7940	0.001439296
Dock3	3.7397	1.85E-19	*Hspa12a*	2.7916	0.003136277
Bex2	3.7122	1.35E-11	*Gfra4*	2.7791	0.001970539
Elfn1	3.6790	0.011629161	*Plppr4*	2.7776	0.000542065
Penk	3.5996	8.18E-16	*Pacsin1*	2.7347	2.71E-05
Ttc9	3.5696	2.88E-06	*Nkain1*	2.7187	0.026991604
Adcyap1r1	3.5625	8.23E-06	*Zfp462*	2.7166	2.64E-14
Snhg11	3.4513	2.16E-13	*Arxes2*	2.6878	9.46E-11
Sntg2	3.4200	8.39E-06	*Aif1l*	2.6876	8.93E-07
Rab15	3.4024	0.029712934	*Slc1a3*	2.6861	0.000298674
Lhfpl4	3.3341	0.016653374	*2600014E21Rik*	2.6685	0.004625159
Tmeff1	3.3174	0.002742996	*Ptprm*	2.6614	1.07E-07

Gene ID	Log_2 Fold Change	Adjusted p-value	Gene ID	Log_2 Fold Change	Adjusted p-value
Ank2	2.6589	2.86E-11	*Otoa*	2.1525	0.008985433
Hmgcs2	2.6524	0.000173767	*Ric3*	2.1421	0.000173075
Fam149a	2.6502	0.001314767	*Mkx*	2.1415	0.000746497
Rtn1	2.6471	3.24E-05	*Ptchd2*	2.1392	2.20E-08
Alpl	2.6256	3.07E-11	*Spred3*	2.1383	2.50E-07
Syt13	2.5665	2.24E-12	*Laptm4b*	2.1180	1.84E-13
Cyyr1	2.5662	3.34E-07	*Grb7*	2.1110	0.018499526
Susd5	2.5409	1.40E-07	*Plag1*	2.1060	5.04E-16
Rhbdf1	2.5352	4.97E-05	*Ar*	2.1046	0.001491956
9630013D21Rik	2.5337	7.25E-15	*Raver2*	2.0966	0.0022645
Gzma	2.5272	4.03E-16	*Perp*	2.0956	1.08E-05
Cd5l	2.5127	0.000207577	*Ncam1*	2.0890	1.58E-18
Bfsp2	2.5081	1.11E-05	*Klrc2*	2.0870	0.000204548
Plekhh2	2.4946	2.23E-06	*Slc35f2*	2.0810	1.48E-11
Ppm1e	2.4920	5.53E-27	*Nid2*	2.0792	2.88E-06
Cadm3	2.4834	1.16E-07	*E330013P04Rik*	2.0773	0.047818044
Efna5	2.4223	0.009652117	*Dusp9*	2.0733	2.43E-06
Hap1	2.4164	0.018211409	*Fzd3*	2.0190	4.31E-07
Olfml2a	2.4120	0.009652117	*Slamf8*	2.0028	0.036496196
Unc13b	2.4116	0.00016128	*Slc30a4*	2.0007	0.00021033
Enpep	2.4114	4.77E-05	*Zfp503*	1.9989	0.019232312
Zfp105	2.4107	2.86E-11	*Gfpt2*	1.9927	0.000267527
Bex1	2.4094	0.000287645	*Clip4*	1.9876	0.003359541
Ighv3-8	2.3872	3.22E-13	*Tln2*	1.9855	0.019433058
Stc2	2.3864	1.58E-05	*Lonrf2*	1.9850	3.29E-06
Gm43194	2.3740	0.008088538	*Cd3g*	1.9811	9.55E-05
Pydc4	2.3709	0.000708895	*Pamr1*	1.9700	1.30E-05
Uchl1	2.3671	0.001386665	*Msrb3*	1.9696	1.17E-05
Egfr	2.3653	0.018211409	*Mar-09*	1.9661	0.001491956
Cplx2	2.3512	8.36E-22	*Uaca*	1.9630	5.50E-11
Ndn	2.3471	0.000512003	*Eomes*	1.9598	0.000424129
Ppp4r4	2.3368	0.033494793	*Nlgn2*	1.9564	0.013049867
Zfp827	2.3143	1.19E-08	*C1qb*	1.9509	1.84E-05
Neo1	2.3087	1.31E-07	*Tmem178*	1.9498	3.07E-11
Hba-x	2.3057	0.000131138	*Ptprn*	1.9457	2.50E-12
5730559C18Rik	2.2968	0.031561786	*Hs3st3b1*	1.9416	3.55E-06
Tnfsf11	2.2924	0.045938796	*H2-Q7*	1.9410	0.000419471
Klri2	2.2703	3.12E-11	*Lrch2*	1.9383	0.000791065
Plcd3	2.2493	0.006047599	*Ptprf*	1.9336	0.00052296
Ebf1	2.2219	2.40E-05	*Ngfr*	1.9210	0.007828262
Klk1	2.1897	0.002768636	*H2afy2*	1.9125	6.14E-08
Lingo3	2.1850	0.020288823	*Bik*	1.9121	9.55E-05
Adamts14	2.1795	0.000424625	*Camk2n1*	1.9076	0.036456706
Kdelr3	2.1732	0.006669568	*Adam33*	1.9026	0.000234465
Lpl	2.1714	2.95E-18	*Stk39*	1.8891	1.05E-07
Cnnm1	2.1708	3.57E-07	*Trdc*	1.8834	1.29E-07

Gene ID	Log_2 Fold Change	Adjusted p-value	Gene ID	Log_2 Fold Change	Adjusted p-value
Acsbg1	1.8796	4.97E-05	*Vpreb3*	1.6490	0.049110172
Gzmb	1.8751	5.15E-12	*Ccl5*	1.6466	1.05E-07
Gpc3	1.8682	0.001331854	*Dcbld2*	1.6465	0.032053258
Mapk11	1.8586	0.009306334	*Gbp4*	1.6457	0.041343784
Rag2	1.8451	0.018982672	*BC023719*	1.6443	0.029669006
Dmc1	1.8358	0.014469499	*Sdc3*	1.6418	5.04E-10
Zfp9	1.8331	4.88E-06	*Ptprk*	1.6405	0.003660368
Klrb1c	1.8326	0.000356514	*Plekhd1*	1.6381	0.008619841
Lysmd2	1.8260	1.43E-07	*Sdc4*	1.6357	0.015288254
Prf1	1.8241	0.00030616	*Tcf7*	1.6293	3.30E-06
Samd3	1.8207	0.000915544	*Arhgap5*	1.6263	2.26E-09
Nrbp2	1.8133	0.000275601	*Sh2d1b1*	1.6222	0.001142403
Zfp618	1.7978	0.04306599	*Gjb2*	1.6198	0.000124505
Hes1	1.7888	5.97E-05	*Pcsk9*	1.6129	0.014568168
Ppp3cc	1.7876	8.05E-11	*Slamf6*	1.6088	0.020520932
Rras2	1.7851	3.62E-11	*Dok4*	1.6069	0.029304498
Hist3h2ba	1.7844	0.002857106	*Hmox1*	1.5985	0.015148933
Pde8a	1.7735	7.50E-11	*Ctsf*	1.5956	1.16E-07
Olfml2b	1.7695	0.018277837	*H2-Q6*	1.5908	9.37E-10
Prune2	1.7666	3.73E-06	*Rundc3b*	1.5836	0.049958762
Id3	1.7662	0.00447536	*Ptk7*	1.5816	0.001418362
Fzd4	1.7645	4.64E-06	*Plcl1*	1.5811	0.043948923
Cbr3	1.7634	0.00735573	*Lifr*	1.5796	2.40E-05
Hoxa10	1.7450	0.029669006	*Dmxl2*	1.5780	1.05E-07
Hpgds	1.7450	0.003095991	*Chst1*	1.5750	1.19E-08
Blnk	1.7448	0.000405842	*Hotairm1*	1.5722	0.015629399
Slfn5	1.7350	0.0009039	*C1qc*	1.5722	0.001620482
Btn1a1	1.7324	2.32E-05	*Trbv13-2*	1.5661	0.000560947
Bmi1	1.7298	3.49E-13	*Klre1*	1.5595	4.82E-05
Arl4d	1.7197	0.001275814	*Nbea*	1.5568	4.25E-05
Fmn2	1.7194	0.015763733	*Il22ra1*	1.5538	0.049497332
Nkx1-2	1.7184	0.01979553	*Pbx3*	1.5492	0.003984213
4933440M02Rik	1.7184	6.43E-06	*Oasl2*	1.5473	2.07E-06
Gkap1	1.7166	1.72E-12	*Cacnb2*	1.5473	0.006592219
Fcna	1.7100	8.06E-05	*Kcnc1*	1.5472	0.005380876
Gcat	1.7075	0.003382588	*Notch3*	1.5471	0.033494793
C1qa	1.7066	0.009768013	*Adra2a*	1.5324	0.017969399
Sorcs2	1.6944	1.67E-06	*C4b*	1.5314	0.001313899
Zfp580	1.6801	2.71E-05	*Dip2c*	1.5309	3.32E-07
Abcg3	1.6763	0.010944972	*Xcl1*	1.5291	0.003136277
Mlph	1.6756	0.022207452	*Rapgef5*	1.5255	0.000507638
Slc16a2	1.6672	0.002710722	*Chd7*	1.5225	6.25E-05
Prkg1	1.6633	0.006355029	*Zbtb10*	1.5185	2.44E-07
Kcnj10	1.6614	0.018794747	*Bend5*	1.5180	0.032730211
Spic	1.6566	0.038273745	*Ccr5*	1.5171	3.34E-07
Gimap7	1.6513	0.000195244	*Mreg*	1.5068	0.005319185

Gene ID	Log_2 Fold Change	Adjusted p-value	Gene ID	Log_2 Fold Change	Adjusted p-value
Jam3	1.5028	0.005167454	*Stx1a*	1.3400	0.029712934
Prrg4	1.4973	3.08E-06	*Cxcr3*	1.3350	0.001626747
Gimap3	1.4945	0.001696653	*Zdhhc2*	1.3291	3.94E-05
Tbc1d30	1.4929	0.003468624	*H2-Aa*	1.3270	0.017831065
Cd247	1.4850	0.022402666	*Gimap8*	1.3229	0.019886567
Pmaip1	1.4818	0.003123615	*Klrk1*	1.3227	0.000497758
Ubtd2	1.4796	9.92E-06	*Sep-10*	1.3194	8.34E-05
Tmem151b	1.4794	0.006355029	*Gm37795*	1.3149	0.012298401
Pkia	1.4739	0.044988956	*Il7r*	1.3126	0.005396486
Klrc1	1.4739	0.000675351	*Thy1*	1.3123	6.27E-07
Lix1l	1.4716	0.000187107	*Tgtp1*	1.3116	0.00735573
Fam184a	1.4662	1.25E-05	*Cpne7*	1.3112	0.004594572
Gmfg	1.4651	0.038345424	*Atp8b5*	1.3107	0.002606342
Il2rb	1.4617	1.14E-06	*Prkcdbp*	1.3054	0.027902112
Eml5	1.4555	0.000544739	*Arhgef5*	1.3021	5.30E-05
Afp	1.4537	0.029712934	*Gimap3*	1.2992	8.70E-05
Dtna	1.4529	0.045496728	*Tox*	1.2966	0.005009866
Fam212b	1.4517	0.045309023	*Ciita*	1.2929	0.003272852
Mllt4	1.4469	8.00E-06	*Cd4*	1.2920	0.009404166
H1f0	1.4388	1.42E-06	*A430105I19Rik*	1.2881	0.005801168
Abca5	1.4370	0.001504638	*Lrrc16b*	1.2805	0.000279419
Tdrkh	1.4352	3.42E-05	*Tmc3*	1.2780	0.004037481
Selm	1.4287	2.44E-07	*Parp12*	1.2776	6.50E-07
Reep1	1.4254	0.00204637	*Cdkl5*	1.2765	0.000961543
Rora	1.4239	0.026093987	*Trbc1*	1.2738	0.001312081
Tbx21	1.4221	0.001642171	*Fkbp11*	1.2708	1.70E-05
Chic1	1.4133	0.005955875	*Gprin3*	1.2703	0.001439296
B3galnt1	1.4127	0.022365402	*Tjp2*	1.2694	8.23E-08
Shisa2	1.4097	0.003660368	*Prss16*	1.2667	9.90E-06
Wdr54	1.4031	0.041832986	*Maf*	1.2640	0.012634679
Kif19a	1.4003	0.005657275	*Cdkn1a*	1.2534	0.00832091
E130012A19Rik	1.3970	2.41E-07	*H2-Q5*	1.2462	1.49E-06
Mpzl3	1.3954	0.008831026	*Trbc2*	1.2355	0.006304986
Lrig3	1.3951	0.000665237	*Irf2bpl*	1.2329	0.000242105
Tgfa	1.3869	0.033494793	*Pou6f1*	1.2301	0.042280005
Myh7b	1.3855	0.013585246	*Jazf1*	1.2242	0.005611335
Mageh1	1.3843	0.042542979	*Trpc1*	1.2206	0.025239268
Gbp9	1.3813	0.016530871	*Sapcd1*	1.2156	0.033494793
Pbx3	1.3797	0.018277837	*Col27a1*	1.2146	0.0315174
Rasal2	1.3678	1.05E-06	*Ptpru*	1.2133	0.005628872
Rtp4	1.3643	0.000191433	*Ptpdc1*	1.2021	1.18E-05
Ccdc85c	1.3618	0.033494793	*Lamc1*	1.1934	1.05E-06
Tcrg-C4	1.3576	0.00408088	*Mn1*	1.1841	0.00509308
Arhgap29	1.3545	5.10E-06	*Ppp2r3a*	1.1827	0.003984213
H2-Ab1	1.3507	0.003660368	*Fkbp9*	1.1800	1.67E-06
Ctsw	1.3491	4.23E-06	*Aldh5a1*	1.1781	0.001384499

Gene ID	Log_2 Fold Change	Adjusted p-value	Gene ID	Log_2 Fold Change	Adjusted p-value
Kbtbd11	1.1704	1.19E-05	*Il2ra*	1.0067	0.0276757
Cd96	1.1699	0.002354766	*Dusp23*	1.0062	0.041427634
Plxna2	1.1660	0.008831026	*Rbp1*	0.9910	0.032475731
Ckb	1.1654	0.000915544	*Myo1e*	0.9887	0.005861915
Fam174b	1.1623	6.19E-06	*Hpgd*	0.9886	0.049275679
Ccnd1	1.1607	9.03E-06	*Ldhb*	0.9850	0.020288823
Sema7a	1.1597	0.008198811	*B4galnt4*	0.9822	0.016747744
Socs3	1.1503	0.006725463	*Tnfrsf10b*	0.9784	0.032652245
Socs1	1.1500	0.006592219	*Parm1*	0.9524	0.003290689
Cacna1e	1.1488	0.000915544	*Camk1d*	0.9515	0.000278543
Degs2	1.1386	0.008619841	*Cers4*	0.9514	0.04306599
Pik3c2a	1.1329	3.21E-06	*Neto2*	0.9499	0.041902127
Abtb2	1.1274	0.001446088	*Tmem98*	0.9499	0.003473129
Camk4	1.1256	0.009472597	*Sall2*	0.9497	0.013068533
Adgra3	1.1247	2.25E-05	*Cmtm4*	0.9458	0.001983469
Setbp1	1.1210	0.034485601	*Nlrc5*	0.9435	0.000250821
Pde3b	1.1199	2.91E-06	*Pkd2*	0.9430	0.000708073
Itm2a	1.1183	0.008583215	*Slc22a21*	0.9408	0.035918809
Cysltr1	1.1141	0.014568168	*Slc9a5*	0.9360	0.004467257
Myo10	1.1139	0.025581518	*Car13*	0.9312	0.001512777
2700081O15Rik	1.1129	7.33E-06	*Rxra*	0.9310	0.005367173
Slc26a3	1.1106	0.034485601	*Pvrl1*	0.9290	0.016530871
Ncs1	1.1093	0.000371582	*Adora2a*	0.9186	0.013230002
Fas	1.1068	0.036902743	*Mybl1*	0.9098	0.022207452
Zap70	1.1053	0.027902112	*Pycr1*	0.9079	0.046847181
Cecr2	1.1042	5.37E-05	*Pik3c2b*	0.9023	0.001826211
Tcrg-C1	1.1030	0.027902112	*Basp1*	0.9017	0.032475731
Zbp1	1.1029	0.008052666	*Rhobtb3*	0.8997	0.002891349
Ifnlr1	1.1023	0.009306334	*Plekha5*	0.8919	0.002887594
Glis3	1.1014	0.038273745	*Tspan3*	0.8881	0.000675351
Slc39a14	1.1013	0.000608477	*Ptk2*	0.8881	0.015109867
Id2	1.0968	1.85E-05	*Rbm44*	0.8828	0.03164704
Ppp1r9a	1.0881	0.000138188	*Sv2a*	0.8793	0.047558474
Plekhh1	1.0774	0.046681042	*Tulp3*	0.8788	0.006685678
Syde1	1.0761	0.048842368	*Armcx4*	0.8786	0.02468981
Coro2a	1.0689	0.022365402	*Alox12e*	0.8726	0.018677664
Perm1	1.0673	0.038477979	*Drc7*	0.8684	0.047818044
Peg12	1.0604	0.00693804	*Phlpp1*	0.8647	0.002810651
Epb41l5	1.0571	0.005427571	*1700025G04Rik*	0.8603	0.011839211
Tanc2	1.0516	0.000356514	*Adgrl1*	0.8516	0.003227465
Pdia5	1.0487	5.21E-05	*Slc7a6*	0.8491	0.005861915
Grap2	1.0449	2.84E-05	*Afap1l1*	0.8423	0.002367104
Zfhx3	1.0390	0.009152793	*Nusap1*	0.8390	0.016653374
Gimap6	1.0386	0.039978028	*Homer1*	0.8383	0.021022149
Khdrbs3	1.0299	9.21E-05	*Zbtb16*	0.8365	0.008985433

Gene ID	Log_2 Fold Change	Adjusted p-value	Gene ID	Log_2 Fold Change	Adjusted p-value
Mfsd6	0.8314	0.009306334	*Emc9*	-0.8110	0.011212359
Prkch	0.8284	0.029712934	*Sphk1*	-0.8124	0.03164704
Faah	0.8255	0.014568168	*Spon2*	-0.8170	0.007828262
Cdc42bpb	0.8171	0.005409736	*Pklr*	-0.8197	0.007234536
Slc2a1	0.8135	0.009306334	*Serinc3*	-0.8238	0.000770619
Foxd2os	0.8107	0.016654582	*Igkv4-57*	-0.8277	0.033494793
Ada	0.8100	0.008271908	*Ifi30*	-0.8495	0.008746739
Kif1b	0.8000	0.018677664	*Fads3*	-0.8984	0.006669568
H2-Q4	0.8000	0.002158617	*1700006J14Rik*	-0.9028	0.000785826
Plcg1	0.7984	0.013230002	*Ntn4*	-0.9166	0.003229724
Crim1	0.7983	0.018651241	*Atp1b2*	-0.9430	0.000374676
Celsr2	0.7979	0.036496196	*Cela1*	-0.9481	0.00038049
Hdgfrp3	0.7965	0.018677664	*Cpt1c*	-0.9503	0.011008918
Ms4a4b	0.7944	0.021695541	*Hpn*	-0.9906	0.015425655
Lama5	0.7917	0.012965366	*Aqp11*	-0.9939	0.027248364
Ptrf	0.7837	0.040389149	*Ighm*	-1.0285	5.58E-06
Gmds	0.7805	0.016530871	*C1qtnf6*	-1.0345	0.015109867
Ighv2-2	0.7473	0.036456706	*Ighv1-19*	-1.0504	0.002584592
Hspa2	0.7442	0.030045938	*Enpp3*	-1.0761	0.030714021
St8sia6	0.7442	0.031677663	*Ccdc74a*	-1.0851	0.002186949
Hipk2	0.7416	0.02048884	*Ighv7-1*	-1.1167	0.00255965
Lax1	0.7405	0.036653999	*Ighv2-6*	-1.1303	0.016184476
Serpine2	0.7384	0.008957346	*Gm15915*	-1.1501	0.013230002
Foxo1	0.7383	0.025460628	*Slc26a1*	-1.2826	2.04E-05
Bmp1	0.7327	0.021025807	*Ighv7-3*	-1.5705	3.07E-11
Plekha8	0.7303	0.032475731	*Ighv1-78*	-1.6593	2.93E-05
Lgalsl	0.7253	0.013310215	*Igkv14-100*	-1.6679	0.005628872
Sep-06	0.7246	0.014195588	*Ighv1-81*	-1.6871	3.81E-09
Npl	0.7202	0.033765948	*Hddc3*	-1.7055	5.74E-10
Lacc1	0.7191	0.048648156	*Ighg1*	-1.7190	4.31E-07
Gca	0.7095	0.029712934	*Igkv4-61*	-1.7474	0.003406732
Klf7	0.6965	0.036496196	*Nuf2*	-1.7540	0.016654582
Mboat1	0.6888	0.04479626	*Otogl*	-1.8386	0.002034566
Bbx	0.6750	0.03567193	*Nuf2*	-2.0160	0.000287645
Ptprcap	0.6514	0.032052391	*Ighv8-12*	-2.1654	6.45E-06
Prkar2a	0.6428	0.049685333	*Ighv1-18*	-2.1684	0.000234465
Casp3	-0.6045	0.043459246	*Ezh2*	-2.3631	0.024614917
Shkbp1	-0.6261	0.051042526	*Igfn1*	-3.6222	0.0276757
Nxpe4	-0.7026	0.017969399	*Igkv9-123*	-4.3084	8.36E-22
Setd6	-0.7153	0.018848593	*Eif3j2*	-4.3162	2.40E-46
Fzd7	-0.7251	0.015052321	*D430041D05Rik*	-4.3998	0.001202676
Spire2	-0.7732	0.036902743	*Pcna-ps2*	-4.7606	6.86E-77
Ctse	-0.8034	0.0325769			

Appendix A 5.2

Genes differentially expressed between *Ezh2*$^{+/+}$ and *Ezh2*$^{-/-}$ lineage negative murine HSPCs in the MLL-AF9 leukaemias. Significant gene expression results were those with a p-value adjusted for multiple hypothesis testing of <0.05 with a $\log_2$ fold change of greater or less than 0.5.

Gene ID	$\log_2$ Fold Change	Adjusted p-value	Gene ID	$\log_2$ Fold Change	Adjusted p-value
Lin28b	8.3571	0.000409407	*Asprv1*	3.6325	2.75E-05
Spag6	7.8819	1.39E-20	*Gm14327*	3.6317	0.000331242
Cacnb2	6.2302	1.31E-32	*Gm42632*	3.6274	2.15E-05
2900092N22Rik	6.1739	4.48E-27	*Gm37240*	3.5931	0.024579692
Prg2	6.0452	0.003174392	*Gm20407*	3.5920	0.001311807
Hist1h2al	5.8225	0.025360131	*Nxnl2*	3.5829	0.022219115
Gm37437	5.6912	8.53E-12	*Zcchc11*	3.5457	4.10E-05
C330013J21Rik	5.6301	2.15E-07	*Kbtbd12*	3.5427	3.88E-05
Gm13269	5.3726	3.78E-11	*Thbs1*	3.5343	2.55E-08
Gm37891	5.3422	3.02E-09	*Yes1*	3.5098	4.13E-09
Gm38273	5.3385	1.29E-10	*Tcf24*	3.5073	3.37E-08
Retnla	5.3260	2.97E-21	*P2ry14*	3.5025	0.004047245
Gm44246	5.2468	2.83E-05	*Gm26603*	3.5001	8.10E-19
Gm37006	5.2299	5.79E-19	*Bmi1*	3.4840	8.31E-08
A1cf	5.0160	0.034048087	*Pax5*	3.4226	2.03E-06
Gm4969	5.0027	7.34E-06	*Epha7*	3.4150	4.51E-07
Gm38105	4.9686	4.85E-08	*P2rx3*	3.4039	1.36E-07
Col19a1	4.8757	0.000120183	*P2ry14*	3.3568	0.008757237
Sorcs2	4.5664	3.05E-18	*C79798*	3.3476	1.26E-17
C330013J21Rik	4.5477	0.001277686	*6720464F23Rik*	3.3472	1.47E-06
Slc22a3	4.4833	1.21E-14	*Ptpn13*	3.3420	0.009865244
Gm20539	4.3688	0.000573742	*Gm37524*	3.3274	0.030624084
9430087J23Rik	4.3135	2.19E-05	*Bcl2a1b*	3.3259	1.68E-05
Ikzf3	4.1700	0.000295085	*P2ry14*	3.3200	0.000259914
Tmem178	4.1529	1.35E-18	*Oaf*	3.3196	0.038098799
Gm17110	4.1371	1.84E-06	*Tns1*	3.3179	0.001284206
Gm8730	4.0936	4.97E-32	*Epha7*	3.2927	2.16E-05
Bmpr1a	4.0751	2.42E-20	*Tiam1*	3.2920	0.01747164
Gm11915	4.0618	6.77E-20	*Krt18*	3.2745	6.90E-07
Retnlg	4.0438	8.95E-08	*4933406I18Rik*	3.2605	2.35E-08
Ppnr	4.0289	5.49E-06	*Zbtb16*	3.2512	0.011131775
Adgre4	4.0087	0.01081153	*Gm25436*	3.2466	0.01199519
Gm37215	3.8372	0.00192646	*Rspo1*	3.2356	1.21E-09
Aff3	3.8082	0.022634083	*Col19a1*	3.2129	7.74E-12
Aspa	3.6819	5.38E-18	*Lrrc1*	3.1815	0.000559847
Hoxc6	3.6605	0.001521857	*A430093F15Rik*	3.1722	0.001060425

Gene ID	Log $_2$ Fold Change	Adjusted p-value	Gene ID	Log $_2$ Fold Change	Adjusted p-value
Unc5cl	3.1637	0.002625895	*Trpv3*	2.5561	2.59E-07
Atp8a2	3.1576	1.02E-09	*Cep126*	2.5499	0.011624178
E330040D14Rik	3.1553	0.000143898	*Hcar2*	2.5424	0.012134889
Emid1	3.1533	4.16E-07	*Clcf1*	2.5405	0.037779742
Gm25831	3.1159	3.14E-07	*Gm29438*	2.5005	0.004019465
Plag1	3.1094	2.15E-18	*Rras2*	2.4875	3.16E-06
9530036M11Rik	3.0905	0.001298601	*Ankrd22*	2.4740	0.007639463
Dach1	3.0738	0.014048442	*Nrarp*	2.4721	0.034682741
Stfa2	3.0646	6.24E-19	*Csnk1e*	2.4682	0.040427544
4933406I18Rik	3.0631	3.88E-05	*Pbx3*	2.4675	0.000191321
Gm14399	3.0493	0.008757237	*Grb7*	2.4660	0.003016624
Cdkn2a	3.0402	0.024021077	*Skida1*	2.4574	0.0001055
Fam135a	3.0352	1.77E-08	*Fam135a*	2.4431	0.002189662
Gm37212	3.0069	0.000612244	*Gm25465*	2.4419	0.036858362
Bmi1	3.0000	8.61E-17	*Clic5*	2.4286	8.74E-06
Sema7a	2.9938	0.0003334	*Pik3c2g*	2.4281	0.044364482
Slc14a2	2.9749	0.005013363	*Dclk2*	2.3902	0.001484416
Gm14420	2.9102	0.00183369	*Dapk2*	2.3864	8.30E-10
Gm38126	2.8776	0.001802243	*Gm37310*	2.3668	0.007304068
Ankrd13b	2.8758	0.000717807	*Adgrl4*	2.3653	1.04E-07
Heg1	2.8680	1.87E-10	*Mir7655*	2.3651	0.000449367
Gm38125	2.8652	0.037256199	*Ppp1r42*	2.3307	2.05E-06
Ly6g	2.8590	0.001827815	*Gm38302*	2.3232	0.014923603
Faah	2.8567	1.48E-07	*Eml5*	2.3069	1.06E-06
Akt3	2.8559	7.05E-06	*Erc2*	2.3013	7.34E-06
Perp	2.8501	0.003190598	*Tbc1d10c*	2.3007	9.97E-10
Slc25a23	2.8392	0.033823473	*Ica1*	2.2749	0.037009516
Gm11655	2.8253	0.012578366	*Ddx43*	2.2649	0.004734003
Gm37199	2.8115	0.005997471	*Ighd*	2.2645	0.002012049
A930004D18Rik	2.8101	2.54E-08	*Gm38235*	2.2599	0.000859248
Tet1	2.8078	0.014341187	*Mmp9*	2.2553	0.000194374
Ptk7	2.7687	2.29E-07	*Gm13371*	2.2450	7.33E-05
Stau2	2.7541	0.006885773	*AI854703*	2.2383	0.025360131
Crtac1	2.7456	0.047914637	*Pard6b*	2.2218	0.002026092
Rfx3	2.7415	3.46E-09	*Calcrl*	2.2192	5.11E-09
Gm26756	2.7404	0.005997471	*Mir1927*	2.2149	0.040219402
Pbx3	2.7246	1.42E-09	*Dennd3*	2.2017	1.42E-09
Ldlrad4	2.7115	2.25E-11	*Gm16158*	2.1990	0.030778582
Gramd2	2.7111	0.016320984	*Gm5483*	2.1982	0.005257418
Ccdc8	2.7106	0.005239955	*Add3*	2.1975	0.000133328
Cmya5	2.6839	0.000581586	*Lrrc1*	2.1963	0.011999887
Gm25180	2.6672	0.00100639	*Elovl7*	2.1936	0.000245282
Sh3bp5	2.6443	0.016480515	*Galnt3*	2.1880	4.22E-08
Tacstd2	2.6391	8.21E-05	*Ssbp2*	2.1848	0.032830254
Slain1	2.6053	0.021404084	*Med12l*	2.1739	0.003654687
Coro2a	2.5600	8.17E-06	*Gm15818*	2.1539	0.001269854

Gene ID	Log $_2$ Fold Change	Adjusted p-value	Gene ID	Log $_2$ Fold Change	Adjusted p-value
BC100530	2.1350	0.003285881	*Flt3*	1.7897	0.032343751
Sh3bp4	2.1250	7.05E-06	*6330409D20Rik*	1.7851	0.000204552
Fam169b	2.1240	1.15E-06	*Cd33*	1.7804	1.84E-06
Tln2	2.1113	1.11E-07	*Ppm1e*	1.7771	2.51E-06
Glis2	2.1012	3.88E-05	*Fgr*	1.7740	0.004402726
Stfa2l1	2.0952	0.036015993	*Zfp831*	1.7355	9.14E-05
Ipcef1	2.0781	0.0353509	*Sgms1*	1.7254	0.029047921
Klhl41	2.0721	0.033869367	*Mapk4*	1.7241	0.000478668
Prss30	2.0639	0.007035403	*Stfa3*	1.7187	0.009973649
Fam73a	2.0614	0.015684058	*Ear2*	1.7128	0.047376298
Sesn1	2.0610	0.041278046	*Smim5*	1.7125	0.026427153
Agpat4	2.0585	3.56E-05	*Gm26522*	1.7080	0.027652291
Gm22883	2.0534	0.037633378	*Gp1bb*	1.7064	0.002701185
Gm5416	2.0348	3.69E-05	*Gm42571*	1.7055	0.000925419
Phlpp1	2.0094	7.32E-08	*Gm17207*	1.7047	0.007503325
Tnik	2.0053	0.016766204	*Mfsd6*	1.7032	0.000191321
Add3	1.9945	7.34E-06	*Cxadr*	1.7002	0.00054897
Gm15356	1.9934	0.002479963	*Gm16059*	1.6974	0.002625895
Slc1a3	1.9914	0.005941906	*Spns3*	1.6966	0.039333199
Smpd3	1.9841	0.000229904	*Gm36975*	1.6955	0.001069895
Six5	1.9606	0.003151284	*Tspan13*	1.6903	0.023961069
Bcas1os2	1.9556	0.000977746	*Gm28373*	1.6800	0.001060242
Gm37802	1.9361	0.002974898	*Hip1r*	1.6740	2.44E-05
Slpi	1.9345	0.047535854	*Klf5*	1.6737	0.018984682
Gphn	1.9320	7.05E-06	*Alcam*	1.6600	0.00088187
Fmnl2	1.9246	0.003175948	*Adcy4*	1.6571	0.004385005
Pde3b	1.9225	2.01E-07	*Klrb1f*	1.6560	0.038098799
Epb41l4b	1.9199	0.01083626	*Gm15880*	1.6469	0.029769008
Prkca	1.9092	0.014256054	*Sh2d3c*	1.6203	0.004683723
Ntng2	1.9085	0.002029901	*Tbc1d16*	1.6148	0.040219402
Gm37334	1.9076	0.013686056	*Gm43430*	1.6011	0.002446865
Sox13	1.9031	0.014836458	*Tjp2*	1.5998	0.004786272
C030034L19Rik	1.8795	0.021996195	*Nrip1*	1.5959	0.00088187
Gm37465	1.8786	0.034964156	*Mar-01*	1.5947	0.015152184
Gkap1	1.8758	2.92E-05	*Map3k9*	1.5942	1.79E-05
Il1b	1.8703	0.020617799	*Spns2*	1.5925	5.97E-05
Hoxaas2	1.8685	0.020379532	*Pik3cb*	1.5913	0.002607396
Zfp599	1.8656	0.001542566	*Ddx3y*	1.5889	2.37E-05
Col4a1	1.8623	8.61E-05	*Csnk1e*	1.5875	0.004265588
Gcnt2	1.8561	0.002385479	*Slc7a8*	1.5837	0.000190855
Col5a1	1.8465	1.06E-06	*Kdm5b*	1.5833	0.017430509
Thbd	1.8309	0.004636264	*Parp3*	1.5823	0.030624084
Syne1	1.8223	0.000459336	*Itgb3*	1.5820	0.025360131
Trib1	1.8096	0.008227053	*Caskin2*	1.5815	0.022896818
Gm43999	1.8003	0.009445025	*Ralgps2*	1.5788	0.028031315
Ralgps2	1.7964	0.019679195	*Hsh2d*	1.5728	0.022219115

Gene ID	Log $_2$ Fold Change	Adjusted p-value	Gene ID	Log $_2$ Fold Change	Adjusted p-value
Gm5751	1.5723	0.012596437	*A330023F24Rik*	1.3141	0.024183645
Chrnb1	1.5613	0.023699414	*Scmh1*	1.3051	0.034048087
Rltpr	1.5593	0.00331841	*Pde7a*	1.2992	0.016840344
Sycp2	1.5589	0.031830631	*Syne1*	1.2984	0.000973551
Tmem216	1.5535	0.008368056	*Gm44292*	1.2890	0.016768979
Smim5	1.5520	0.000527932	*Ssh1*	1.2862	0.021996195
Hipk2	1.5480	5.12E-05	*A430046D13Rik*	1.2832	0.002673331
Cebpe	1.5288	0.032343751	*Runx2os2*	1.2809	0.008368056
Vipr1	1.5254	0.023983074	*Gm42463*	1.2778	0.045349683
2700081O15Rik	1.5132	0.019960492	*Adpgk*	1.2767	0.024892656
Csnk1e	1.5119	0.000516684	*Inpp5j*	1.2690	0.003151284
Lurap1	1.5082	0.004603258	*Gm10612*	1.2684	0.009471129
9830107B12Rik	1.5058	0.034816645	*Acox3*	1.2664	0.024431725
Ppp3cc	1.5030	0.016651481	*Acss1*	1.2572	0.003175948
Anxa11os	1.4972	0.011855976	*Abcb9*	1.2545	0.011667527
Chil5	1.4870	0.02308857	*Srgap3*	1.2433	0.040452683
Myadml2	1.4826	0.008117382	*Btbd11*	1.2250	0.030209302
Neo1	1.4685	0.048810554	*6030400A10Rik*	1.2236	0.034045645
Ccdc85c	1.4667	0.021196459	*Carns1*	1.2111	0.025551012
Lrrk2	1.4615	0.022543245	*Dsp*	1.2065	0.005359344
Ranbp17	1.4594	0.021692373	*Ap1s3*	1.2040	0.021485406
B630019K06Rik	1.4526	0.018016455	*Gm17552*	1.1988	0.037633378
Pdlim1	1.4512	0.000911926	*Ccdc102a*	1.1947	0.037633378
Mbp	1.4371	0.031074368	*Ugcg*	1.1922	0.025159991
Fndc9	1.4304	0.001181623	*9530085L11Rik*	1.1899	0.027232914
Gm44001	1.4167	0.04778651	*Gm43274*	1.1805	0.014883992
Map3k5	1.4159	0.036496618	*Gm42572*	1.1753	0.015684058
Igf1r	1.4151	0.000177284	*Tmco3*	1.1718	0.01083626
Kif21b	1.4130	0.00020904	*Itgb5*	1.1712	0.019753071
2310001H17Rik	1.3992	0.007353507	*Ptprcap*	1.1529	0.023459094
Samd9l	1.3962	0.004393371	*Cd84*	1.1462	0.015975805
Gm20721	1.3946	0.016177947	*Socs6*	1.1421	0.022075368
Olfm4	1.3944	0.011257381	*A130048G24Rik*	1.1363	0.018723227
Rere	1.3792	0.002259632	*Ccl6*	1.1329	0.010968802
9530077C05Rik	1.3661	0.008614204	*Gm10097*	1.1326	0.029510447
Olr1	1.3642	0.000432923	*Lilr4b*	1.1257	0.022889331
Cd5	1.3552	0.021692373	*Fgd6*	1.1204	0.015482829
Plxnc1	1.3496	0.003471371	*Ell2*	1.1201	0.02699617
Hmgn1	1.3438	0.0007259	*Ltb4r1*	1.1167	0.010410422
Sep-01	1.3424	0.015852596	*Raph1*	1.1148	0.014632041
Prex1	1.3329	0.001698042	*Zfp318*	1.0950	0.017430509
Gpr18	1.3317	0.006211376	*A630072L19Rik*	1.0948	0.047573494
Tmem71	1.3314	0.039757716	*Arhgap27*	1.0907	0.044291239
Gnas	1.3267	0.00705825	*Phtf2*	1.0880	0.033070808
Cyp4f18	1.3202	0.048812272	*Gm37159*	1.0865	0.035350682
Chn2	1.3151	0.021544955	*Gm20632*	1.0843	0.029047921

Gene ID	Log $_2$ Fold Change	Adjusted p-value	Gene ID	Log $_2$ Fold Change	Adjusted p-value
Gm43982	1.0790	0.029313718	*Il3ra*	-1.1983	0.005342288
9630013D21Rik	1.0746	0.022219115	*Btbd19*	-1.2289	0.029103733
Kdm6b	1.0623	0.01897076	*Galc*	-1.2434	0.004317886
Ric1	1.0600	0.019960492	*Cyp27a1*	-1.2511	0.005212461
Pik3ap1	1.0515	0.029999353	*Tlr6*	-1.2526	0.005013363
Trp53bp2	1.0282	0.040219402	*H2-Ab1*	-1.2558	0.015391194
Clec2i	1.0239	0.032319532	*Gm16104*	-1.2660	0.022488475
Prrc2b	1.0201	0.039567406	*Lpcat1*	-1.2807	0.002860596
Wipi1	1.0177	0.048810554	*Phyh*	-1.2850	0.0015981
Dennd1c	1.0141	0.028031315	*Pyhin1*	-1.2924	0.001213877
Crtc3	1.0130	0.028282334	*Dhrs9*	-1.3028	0.008113507
Zbtb10	1.0073	0.044314642	*Mllt4*	-1.3070	0.001829311
Rabgap1l	1.0056	0.045556763	*C3*	-1.3085	0.000752295
Mylk	0.9860	0.02872854	*Hs6st1*	-1.3200	0.00365675
Cd2ap	0.9854	0.039775258	*Oas1c*	-1.3242	0.013546745
Xylt1	0.9838	0.039775258	*Atp10a*	-1.3273	0.001150452
Strbp	0.9592	0.047090629	*Tfec*	-1.3304	0.003938258
Rabgap1l	0.9559	0.044088806	*Ptger2*	-1.3447	0.018016455
Gfod1	0.9553	0.044769427	*Adam3*	-1.3455	0.008757237
Arsb	0.9421	0.044291239	*Ldhb*	-1.3455	0.004079882
Stom	-0.9416	0.043399503	*Chpt1*	-1.3504	0.004603258
Calr	-0.9648	0.0338262	*Pyroxd2*	-1.3543	0.001542566
Mt1	-0.9727	0.039449055	*Rexo2*	-1.3625	0.000753803
Tax1bp1	-0.9866	0.028524483	*Zfp217*	-1.3859	0.016768979
Limd1	-0.9944	0.030624084	*Tmem38b*	-1.3922	0.000512489
Tmx4	-0.9996	0.027583885	*Chpt1*	-1.4004	0.023680887
Tlr4	-1.0113	0.038663847	*Gpr65*	-1.4194	0.0010031
Top1mt	-1.0126	0.048670386	*Plpp5*	-1.4238	0.049400656
Vps13c	-1.0325	0.018632515	*1810011H11Rik*	-1.4371	0.000331242
B2m	-1.0403	0.015781075	*Sema6b*	-1.4467	0.015545321
Ppp1r21	-1.0406	0.019571318	*Rps2*	-1.4526	0.003260338
4930503L19Rik	-1.0455	0.047028515	*Egln3*	-1.4545	0.000225606
Deptor	-1.0500	0.037009516	*Praf2*	-1.4570	0.018209427
Grn	-1.0970	0.008827397	*Dnajc12*	-1.4595	0.008757237
Tmem150b	-1.1092	0.023699414	*Glce*	-1.4695	0.035183671
Serp1	-1.1253	0.006210509	*Layn*	-1.4700	0.034816645
Dpp7	-1.1359	0.015391194	*Smad1*	-1.4806	0.011201557
Tspan31	-1.1364	0.010130058	*Asns*	-1.4923	0.016122878
Ogfrl1	-1.1393	0.023699414	*Fam114a1*	-1.5018	0.000889241
App	-1.1537	0.028524483	*Gm13881*	-1.5039	0.000911926
Hspa13	-1.1614	0.015100438	*Ptch2*	-1.5070	0.037256199
Pkd2	-1.1724	0.024813193	*Lonrf3*	-1.5116	0.00032763
Dhx58	-1.1747	0.011841109	*Zfp948*	-1.5124	0.005331053
Ifih1	-1.1763	0.019002737	*Trem3*	-1.5247	4.93E-05
D5Ertd605e	-1.1816	0.019429285	*Arhgap31*	-1.5464	0.018718378
Oas1a	-1.1832	0.020892644	*Cmtm3*	-1.5674	0.000255641

Gene ID	Log $_2$ Fold Change	Adjusted p-value	Gene ID	Log $_2$ Fold Change	Adjusted p-value
Prdx4	-1.5715	7.81E-05	*C1qb*	-1.8355	0.001031584
Gpr179	-1.5768	0.004378766	*Gbp2*	-1.8362	0.004019465
Mrc1	-1.5800	0.002489643	*Mfap3l*	-1.8525	0.006140216
Slc6a13	-1.5801	0.007798183	*9330159M07Rik*	-1.8596	0.035681842
Ppm1f	-1.5866	0.034380078	*Adamts2*	-1.8716	1.81E-05
1810008I18Rik	-1.5868	0.003471371	*Slc28a2*	-1.8974	0.002320235
C1qc	-1.5886	0.015391194	*Gm29395*	-1.9138	0.024545915
Gm38244	-1.5950	0.001838628	*Cndp2*	-1.9256	0.005307977
Gm15232	-1.6052	0.005076312	*Afap1*	-1.9265	0.022543245
Gm11685	-1.6056	0.037256199	*Gm11505*	-1.9278	2.59E-07
Tlr1	-1.6099	0.029139505	*AI429214*	-1.9322	0.001838628
Celsr1	-1.6122	0.039449055	*Lmo1*	-1.9396	0.000401478
Akap17b	-1.6289	0.003031951	*Ces2g*	-1.9460	0.040219402
Copz2	-1.6307	0.009511636	*Prtn3*	-1.9503	1.54E-08
Gstm7	-1.6323	0.044291239	*Gstm1*	-1.9600	0.008687739
Pafah2	-1.6565	6.17E-05	*Chst1*	-1.9622	0.036015993
Gm13431	-1.6591	0.039775258	*Ldlrad3*	-1.9625	1.23E-07
Cfp	-1.6607	3.36E-06	*Bhlha15*	-1.9644	0.000442066
Mmp19	-1.6628	1.50E-05	*Mcf2l*	-1.9654	0.000111053
Gm37233	-1.6673	0.00322379	*9430069I07Rik*	-1.9697	0.006289198
Adhfe1	-1.6688	0.009700028	*0610038B21Rik*	-1.9731	0.003446389
Srxn1	-1.6707	0.000169352	*Endod1*	-1.9864	1.44E-07
Man1a	-1.6792	7.54E-06	*Otogl*	-1.9920	3.40E-06
Mrgpra2a	-1.6887	0.007144454	*Lgals3bp*	-1.9947	0.002496489
Mgst2	-1.7024	0.004986334	*Gm16619*	-1.9963	0.00331841
F10	-1.7101	9.30E-06	*Cyp39a1*	-1.9967	0.000118406
Ptgir	-1.7179	0.026297264	*Slc36a3os*	-2.0011	0.008780287
Gm37261	-1.7218	0.000596947	*Gm9898*	-2.0069	0.048331067
Ptprm	-1.7291	9.68E-05	*Vat1*	-2.0095	2.38E-05
Zdhhc2	-1.7350	0.015836166	*9330159M07Rik*	-2.0584	1.25E-08
Mgam	-1.7470	0.02872854	*Ms4a6d*	-2.0606	0.0003334
6330562C20Rik	-1.7479	0.014836458	*Nid2*	-2.0689	0.000226246
Cebpa	-1.7509	1.21E-06	*Fam20c*	-2.0711	0.007258212
Slc46a1	-1.7598	0.000250224	*Mpo*	-2.0980	1.16E-06
F10	-1.7624	0.0010031	*Sntb1*	-2.1010	0.03386502
Il1rl2	-1.7700	0.035350682	*Pla2r1*	-2.1128	0.00331841
C3ar1	-1.7775	0.005907562	*Dab2*	-2.1275	0.000166664
Jag1	-1.7854	0.0093633	*Rassf10*	-2.1580	2.35E-08
Scamp5	-1.7860	0.011131775	*Ifi27*	-2.1725	0.029739338
Rhoj	-1.7938	0.028770254	*D430019H16Rik*	-2.1856	0.001822599
Il1f9	-1.8023	3.78E-05	*Ms4a3*	-2.1922	0.003446389
Baiap3	-1.8036	0.046418775	*A930033H14Rik*	-2.2336	0.01177775
Hal	-1.8119	0.008227053	*Guca1a*	-2.2373	0.0330242
Car2	-1.8240	0.000100645	*Esr1*	-2.2384	4.92E-06
Spry2	-1.8313	1.86E-06	*Gm15658*	-2.2390	0.000629931
Tubb6	-1.8342	0.025879577	*Sec24d*	-2.2452	5.09E-10

Gene ID	Log_2 Fold Change	Adjusted p-value	Gene ID	Log_2 Fold Change	Adjusted p-value
Gm16193	-2.2618	1.86E-06	*S1pr3*	-2.7882	9.35E-13
Mertk	-2.2630	3.62E-07	*Gm38399*	-2.7906	0.042090255
Six1	-2.2668	7.54E-05	*Tnfsf13os*	-2.7952	0.009629654
Mpo	-2.2677	0.000957854	*1810006J02Rik*	-2.8030	0.008033807
Fstl1	-2.3089	5.05E-05	*Slfn4*	-2.8454	0.000326946
Tnr	-2.3131	0.001802243	*Pm20d1*	-2.8523	9.46E-07
Mfsd7a	-2.3145	5.31E-09	*Prps1l1*	-2.8847	8.60E-10
Soga1	-2.3385	6.90E-09	*Il22ra2*	-2.9168	1.69E-08
Elane	-2.3526	0.009940874	*Amot*	-2.9212	0.023354938
Rassf8	-2.3789	2.57E-06	*Slfn4*	-2.9270	9.35E-13
Kcnip3	-2.3846	0.000227241	*Gm15401*	-2.9298	0.008143725
Slamf9	-2.3976	0.002356368	*Ube2l6*	-2.9299	4.56E-09
Gstm1	-2.4287	0.002078787	*Col6a6*	-2.9463	5.16E-08
Six4	-2.4393	0.024892656	*2210010C04Rik*	-2.9526	0.022543245
4933406C10Rik	-2.4437	0.026493429	*Muc13*	-2.9813	2.89E-07
Six4	-2.4491	9.64E-05	*Tg*	-2.9830	4.68E-16
Slc13a3	-2.4560	0.001525379	*Mok*	-3.0091	1.07E-05
F13a1	-2.4562	5.04E-05	*2210417A02Rik*	-3.0690	0.022543245
2510009E07Rik	-2.4740	0.000260618	*5930430L01Rik*	-3.0829	3.00E-07
Spocd1	-2.4744	0.038983543	*Tslp*	-3.1442	0.004930889
Atp6v0d2	-2.4961	1.69E-08	*Gm12498*	-3.1614	0.021826097
Pou4f1	-2.5218	0.02169934	*Slc4a8*	-3.1792	1.42E-07
Clec4b2	-2.5305	3.70E-12	*Mamdc2*	-3.2085	0.000129356
Adgrb1	-2.5647	2.08E-09	*Muc13*	-3.2241	0.023680887
Atp6v0c	-2.5697	0.015482829	*Gm44107*	-3.2265	0.000644778
Gpnmb	-2.6210	2.34E-09	*Fmn1*	-3.2362	0.021198775
1810006J02Rik	-2.6440	6.02E-14	*C4b*	-3.2448	0.024545915
Hfe	-2.6495	1.69E-11	*Mgl2*	-3.2533	6.20E-18
Arhgef38	-2.6828	6.83E-10	*Cd4*	-3.3012	3.68E-10
Mcpt8	-2.6893	0.04298596	*Mgl2*	-3.3611	6.24E-19
Mxi1	-2.7072	0.012569999	*Trf*	-3.4025	1.23E-14
Col6a3	-2.7131	0.004786272	*Pde7b*	-3.4124	1.09E-09
Tnfsf13b	-2.7235	0.00147448	*Gm13391*	-3.4381	0.016768979
Gm33370	-2.7252	1.13E-05	*A730036I17Rik*	-3.5060	1.58E-05
Rasgef1a	-2.7258	0.000227407	*Oas1h*	-3.6116	0.000400577
Aox1	-2.7266	2.89E-06	*Tspan10*	-3.6921	0.008045858
Tuba8	-2.7337	2.11E-14	*Ctsg*	-3.6947	0.000335796
Igfbp4	-2.7356	5.12E-14	*Serpinf1*	-3.7255	2.55E-21
Ccno	-2.7372	1.60E-09	*Gja5*	-3.7456	0.017866928
4930403O15Rik	-2.7463	0.021928008	*Trim29*	-3.7524	0.011269432
F3	-2.7495	0.003557965	*Sdc1*	-3.7871	0.001326144
Rnf186	-2.7496	0.006912851	*Oas1g*	-3.7985	1.29E-09
Gstm2	-2.7521	0.00350006	*Tmem37*	-3.8149	0.00350006
Far2	-2.7551	0.000327591	*Pvrl1*	-3.8692	1.46E-06
Slc26a9	-2.7610	0.01194795	*Mtus1*	-3.9265	0.000401478
4933440M02Rik	-2.7619	0.0121059	*4933415J04Rik*	-4.0628	2.04E-06

Gene ID	Log_2 Fold Change	Adjusted p-value	Gene ID	Log_2 Fold Change	Adjusted p-value
Slc32a1	-4.1723	1.88E-22	*Dsel*	-5.1182	2.65E-19
Enpep	-4.1915	0.004273077	*Gm14204*	-5.2199	3.91E-09
Nr1i2	-4.2885	0.001624286	*Clec11a*	-5.2832	0.003016624
5830416I19Rik	-4.3164	0.000227407	*Grid1*	-5.4359	1.23E-18
Spink2	-4.3694	0.000228138	*Cfap57*	-5.5802	6.83E-10
8430423G03Rik	-4.3783	0.041665883	*Wdr66*	-5.7590	0.001997299
Ces1d	-4.4945	1.29E-05	*Vmn2r96*	-6.1210	1.30E-14
Ccl9	-4.6210	0.000149592	*F7*	-6.2165	1.95E-57
Pcdhb5	-4.8523	0.008227053	*BC018473*	-6.3030	0.000217409
Gm26569	-4.9763	5.34E-06	*BC018473*	-7.7651	5.38E-18
Colec12	-5.0789	0.015741546	*St6galnac1*	-8.3590	0.000527932

Appendix A 5.3

Genes differentially expressed between *Ezh2*$^{+/+}$ and *Ezh2*$^{-/-}$ lineage negative murine HSPCs in the AML1-ETO9a leukaemias. Significant gene expression results were those with a p-value adjusted for multiple hypothesis testing of <0.05 with a $\log_2$ fold change of greater or less than 0.5.

Gene ID	Log 2 Fold Change	Adjusted p-value	Gene ID	Log 2 Fold Change	Adjusted p-value
Retnla	9.6530	1.44E-18	*Retnlg*	5.0326	1.55E-07
Fbp1	8.1861	6.59E-12	*Strc*	4.9783	3.69E-06
Perp	6.8145	7.72E-10	*Bmpr1a*	4.9762	0.021598398
Cygb	6.7721	0.035404506	*Pglyrp1*	4.9567	2.21E-07
Ceacam10	6.4928	2.11E-10	*Sycp2*	4.8865	0.008703859
BC018473	6.3207	0.011154342	*Gm13152*	4.8595	0.008801495
Flt1	6.2741	0.026791608	*Spats2*	4.8577	4.71E-05
Csgalnact1	6.0371	6.91E-06	*Slco3a1*	4.8352	0.000141954
Nat8l	5.9729	0.023345235	*Dmxl2*	4.8183	0.00110084
Slc35d3	5.8415	2.66E-09	*C4b*	4.8152	1.55E-07
Stfa2l1	5.8361	0.030335267	*Ccnjl*	4.8054	6.27E-05
Gm5483	5.8153	0.032768258	*Lcn2*	4.7325	8.01E-07
Trpv3	5.7963	0.000799199	*Ptk7*	4.7171	0.001259817
Mmp8	5.6479	4.08E-09	*Lrg1*	4.6977	8.53E-07
Glrp1	5.6026	0.021156303	*4933431E20Rik*	4.6962	0.015731238
Ighv1-23	5.5947	0.002173496	*P2ry2*	4.6951	0.012546359
Crispld2	5.5663	1.30E-08	*Tmtc2*	4.6434	0.000409001
Gm2895	5.5628	0.000814866	*Lbp*	4.6241	2.47E-06
Chil3	5.5373	6.28E-09	*Gpr84*	4.6224	0.000523592
Plcb4	5.5205	1.68E-07	*Peli2*	4.5327	1.01E-05
Gm16031	5.5128	1.66E-05	*Cxcr2*	4.5267	3.08E-06
Mcam	5.4957	0.000409001	*Serpine2*	4.5194	6.91E-06
Hif3a	5.4893	3.67E-08	*Prkg1*	4.5144	1.30E-05
Prune2	5.4747	0.000538379	*Cd33*	4.5132	0.000966345
Gp1bb	5.3697	1.30E-05	*Ctsg*	4.5075	2.91E-06
Mdga1	5.3479	1.62E-07	*Bmx*	4.5033	0.010339075
Serpinb10	5.3467	0.011815172	*Slco3a1*	4.4927	8.40E-06
AI838599	5.3064	0.000713284	*4930438A08Rik*	4.4904	5.77E-06
P3h2	5.2869	4.51E-07	*Vegfc*	4.4748	0.000131329
Flt1	5.2794	0.000197457	*Amica1*	4.4534	0.027996315
Plekhg1	5.1845	0.031026582	*Cd177*	4.4457	3.91E-06
Ms4a3	5.1059	2.93E-06	*Gm37297*	4.4202	0.000311294
Dgat2	5.0958	1.55E-07	*Fam101b*	4.4197	3.91E-06
Scrg1	5.0916	2.29E-06	*Gng11*	4.4166	0.001029411
Mgam	5.0799	4.51E-07	*Pi16*	4.3868	6.91E-06
Gm11827	5.0371	0.048643768	*Mapk13*	4.3865	0.000579374

Gene ID	Log 2 Fold Change	Adjusted p-value	Gene ID	Log 2 Fold Change	Adjusted p-value
Slco4c1	4.3808	0.016352668	*Plekhf1*	3.9438	0.003977234
Slc44a1	4.3706	1.70E-05	*Osgin1*	3.9418	0.038397449
Mpo	4.3532	1.30E-05	*Lmo1*	3.9323	0.000965573
Gk5	4.3498	0.000505877	*Wfdc21*	3.9320	6.32E-05
Il1r2	4.3235	6.84E-06	*Clec5a*	3.9280	8.84E-05
Ctla2b	4.2909	0.005502957	*Gp6*	3.8775	0.026099589
G0s2	4.2880	0.002281246	*Gm26512*	3.8741	0.029023903
Amer2	4.2779	0.001200592	*Hk3*	3.8695	0.000252546
Ctbp2	4.2729	0.006939604	*C3*	3.8365	8.04E-05
Thbs1	4.2698	0.004389357	*Krba1*	3.8338	0.001627021
Orm1	4.2692	0.017656814	*Cebpe*	3.8303	0.000129641
Cacnb2	4.2628	0.000804811	*Elane*	3.8286	8.04E-05
Fpr2	4.2410	1.30E-05	*Mpo*	3.8233	0.044634913
Olr1	4.2299	2.84E-05	*Tarm1*	3.8203	0.000409001
Ppp1r3d	4.2186	4.14E-05	*Mcemp1*	3.8119	0.000125006
Vnn3	4.2110	0.000995779	*Foxd4*	3.8089	0.007485866
Itih5	4.2042	1.37E-05	*Prtn3*	3.7885	0.000110839
Dio2	4.1940	4.25E-05	*Kcnj2*	3.7789	0.00019597
Akt3	4.1816	0.025193588	*Arsb*	3.7746	0.000114626
Zfp800	4.1805	0.01591885	*Smoc2*	3.7731	0.001696001
Clec4a2	4.1781	0.037880126	*Ngp*	3.7428	0.000129641
Itgb2l	4.1718	8.46E-05	*Slfn1*	3.7413	0.000409001
Gcsam	4.1645	0.004210085	*Aatk*	3.7380	0.031161065
Fpr1	4.1571	1.52E-05	*S100a9*	3.7360	0.00012773
Crtac1	4.1443	0.023230866	*Atp11a*	3.7226	0.000945457
Cldn1	4.1359	0.010687937	*Dnph1*	3.7104	0.003681518
Gm38001	4.1284	0.021598398	*S100a8*	3.7100	0.000129005
Prrt4	4.0996	0.000221127	*Wwc2*	3.7066	0.000314124
Gpc1	4.0906	2.10E-05	*Ctbp2*	3.7028	0.003977234
Scmh1	4.0810	0.000804811	*Capns2*	3.6632	0.006264829
Mtus1	4.0538	0.000162994	*Olfml2b*	3.6607	0.000286963
Ccpg1	4.0379	0.02214357	*Arhgef28*	3.6530	0.03186035
Mylk	4.0376	5.15E-05	*Adgrl4*	3.6467	0.000422333
Adgrb1	4.0340	0.036545179	*Mogat2*	3.6423	0.001573232
F5	4.0323	0.001696001	*Tgfbr3*	3.6333	0.000177671
Cdc42bpa	4.0219	0.000456102	*Pdia5*	3.6205	0.003277539
Rnd1	4.0181	0.000176966	*Egln3*	3.6179	0.000456102
Prss16	4.0034	0.000123335	*Gm38215*	3.6100	0.002008633
Phactr2	4.0027	0.005144036	*Ceacam18*	3.6083	0.011720032
Tacstd2	3.9811	0.01896859	*Afap1l1*	3.5993	0.000153426
Cd300lf	3.9797	0.006939604	*Dock5*	3.5991	0.008505266
Msrb2	3.9771	0.000286963	*Clec12a*	3.5975	0.006811352
A530064D06Rik	3.9693	0.000416172	*Cers6*	3.5902	0.001256239
Slc28a3	3.9662	0.007423496	*B430306N03Rik*	3.5868	0.002761282
Bst1	3.9622	4.22E-05	*Pls1*	3.5864	0.005519174
C1rl	3.9501	0.000109172	*Siglece*	3.5842	0.000149575

Gene ID	Log 2 Fold Change	Adjusted p-value	Gene ID	Log 2 Fold Change	Adjusted p-value
Nucb2	3.5603	0.003691292	*Anxa1*	3.2801	0.000981415
Abcd2	3.5586	0.000305445	*Rab32*	3.2728	0.001054663
Fam217b	3.5447	0.000605649	*Gm13375*	3.2718	0.025889705
Ccr1	3.5342	0.000409001	*Card9*	3.2711	0.003622805
Sgms2	3.5338	0.00049831	*Aph1b*	3.2640	0.00182194
Rhobtb3	3.5324	0.002833777	*Ckap4*	3.2593	0.002368454
Ccdc122	3.5296	0.002708803	*Arl11*	3.2544	0.004229527
B4galt6	3.5270	0.000505877	*Rgcc*	3.2483	0.003480985
Ltf	3.5173	0.000409001	*Kctd1*	3.2451	0.003640933
Zfp449	3.5079	0.032352251	*Cped1*	3.2347	0.005800464
Fam92a	3.5028	0.008541462	*Sep-05*	3.2344	0.002362986
Ric3	3.4940	0.034241133	*Gatm*	3.2330	0.001128954
Pira2	3.4848	0.000682284	*Gapt*	3.2320	0.002637033
Zfp820	3.4836	0.006861602	*Galc*	3.2294	0.001744978
Pram1	3.4828	0.000799199	*Gm29125*	3.2278	0.020108289
6330403L08Rik	3.4825	0.017491763	*Lpcat2*	3.2228	0.001256239
Atp8b4	3.4781	0.000678839	*Plppr3*	3.2221	0.002433827
Afap1	3.4746	0.000523592	*Gm26510*	3.2143	0.012067299
Bicd1	3.4591	0.006962375	*Actn1*	3.2135	0.009854763
Fgd4	3.4529	0.00110084	*Il1b*	3.2089	0.047454967
Agap1	3.4451	0.00161913	*Serpina3i*	3.1939	0.036725086
Tnfrsf1a	3.4364	0.000563048	*Mctp1*	3.1803	0.001706699
Acpp	3.4323	0.000830293	*Bcl6*	3.1759	0.039790742
Il13ra1	3.4308	0.000523592	*F730035M05Rik*	3.1745	0.043548421
Panct2	3.4237	0.012454234	*Ldhb*	3.1685	0.008382869
Them7	3.4216	0.019695097	*Hgf*	3.1627	0.03997003
Hp	3.4177	0.000606731	*Sorl1*	3.1509	0.001449066
Neurl1b	3.4160	0.002362986	*Tarm1*	3.1444	0.017656814
Tns1	3.4098	0.040503919	*Gpr65*	3.1329	0.001611033
Clec5a	3.4076	0.012067299	*Laptm4b*	3.1315	0.005902668
Hsd11b1	3.3973	0.003274721	*Tlr2*	3.1301	0.002637764
Megf9	3.3916	0.000765291	*Cass4*	3.1262	0.044634913
Igsf6	3.3704	0.000713284	*Gk5*	3.1169	0.033287325
Cyp4b1	3.3672	0.030082072	*Ptprc*	3.1129	0.008541462
Anxa3	3.3590	0.000942723	*Fgr*	3.0981	0.002250867
Ndrg1	3.3565	0.000814866	*App*	3.0920	0.00244345
Ano10	3.3379	0.000966345	*Cebpb*	3.0892	0.00411519
Fcnb	3.3356	0.00110084	*Slc35f2*	3.0887	0.030410259
Rhou	3.3320	0.003640933	*Neto2*	3.0865	0.007134391
Enah	3.3281	0.016578309	*Vwf*	3.0835	0.025833008
Hspa12a	3.3249	0.024307674	*Abcc2*	3.0807	0.04544032
Lyz2	3.3249	0.000960005	*Hvcn1*	3.0792	0.041780018
Chd7	3.3194	0.00417792	*Pla2g7*	3.0710	0.001482345
Hook1	3.3080	0.042300222	*F13a1*	3.0702	0.004366123
Ptges	3.3059	0.006962375	*Ltbr*	3.0653	0.003274721
Slc2a9	3.2986	0.040141191	*Ccdc8*	3.0477	0.002761282

Gene ID	Log 2 Fold Change	Adjusted p-value	Gene ID	Log 2 Fold Change	Adjusted p-value
Ldlrad3	3.0461	0.003622805	*Sla*	2.8281	0.008580213
Zik1	3.0454	0.033287325	*Sell*	2.8233	0.006962375
Synm	3.0407	0.040815551	*Fkbp11*	2.8171	0.019725533
Selp	3.0397	0.015894177	*Clec4e*	2.8132	0.008433658
Itm2a	3.0326	0.006962375	*E130012A19Rik*	2.8124	0.026121203
Slc25a24	3.0283	0.039189302	*Plac8*	2.8073	0.017875682
Nfil3	3.0192	0.005533167	*Ctla2a*	2.7896	0.004056424
Lpgat1	3.0189	0.002498172	*Eif4e3*	2.7831	0.009764698
Ighv1-5	3.0100	0.018185192	*Rxra*	2.7761	0.01223428
Fut4	3.0079	0.005902668	*Anxa2*	2.7692	0.008540913
Znrf3	3.0066	0.037157487	*Slc23a2*	2.7654	0.00754827
Gm14548	3.0046	0.004229527	*Rgs18*	2.7382	0.012676074
Slc36a4	3.0034	0.033533276	*Vsir*	2.7382	0.012873361
Gramd2	3.0007	0.046062404	*Tlr4*	2.7325	0.009243087
Gbe1	2.9704	0.021728494	*Trim30b*	2.7322	0.024028463
Reck	2.9646	0.01358965	*Asrgl1*	2.7313	0.008433658
Gca	2.9640	0.006124391	*Cybb*	2.7311	0.008721279
Ralb	2.9621	0.00411519	*Ccl6*	2.7310	0.042102606
Rasgrf2	2.9474	0.020983308	*Apobr*	2.7278	0.014997163
Glrx	2.9454	0.003561523	*Lrp5*	2.7211	0.011888345
Gm38146	2.9452	0.045892593	*Il6ra*	2.7187	0.019725533
Gm9949	2.9412	0.042102606	*Nedd4l*	2.7173	0.045086469
Antxr2	2.9362	0.007637767	*Reep3*	2.7143	0.009764698
Slc26a9	2.9357	0.022610022	*B3gnt7*	2.7073	0.01731534
Dscam	2.9354	0.010751171	*4930539E08Rik*	2.6978	0.030697474
Scel	2.9338	0.025989497	*Atp10a*	2.6957	0.011575935
Prrg4	2.9249	0.042709665	*Pstpip1*	2.6936	0.013946069
Slc22a21	2.9206	0.043972032	*Gstm1*	2.6908	0.018430476
Dhrs7	2.9135	0.006124391	*Nid2*	2.6827	0.020983308
Trim30b	2.9134	0.035160413	*Ebi3*	2.6619	0.020108289
Mctp2	2.9077	0.018185192	*Nadk*	2.6611	0.030335267
Sort1	2.9021	0.040436889	*Rap2a*	2.6608	0.018185192
Pak1	2.9006	0.025193588	*Scarb2*	2.6604	0.010322684
Adgra3	2.9004	0.037157487	*Nfam1*	2.6573	0.020134486
Ltf	2.8941	0.033287325	*Emb*	2.6489	0.01591885
1700020L24Rik	2.8810	0.039189302	*Clec4d*	2.6409	0.01896859
Svip	2.8804	0.010222274	*Slc31a2*	2.6371	0.018797177
Pygl	2.8642	0.007558363	*Slc15a3*	2.6326	0.011998272
Treml1	2.8642	0.004542927	*1700025G04Rik*	2.6208	0.028370788
Fam134b	2.8599	0.005636077	*Gm9733*	2.6024	0.033287325
Sox6	2.8563	0.024601538	*Msrb1*	2.6002	0.024124491
Mmp9	2.8538	0.007357828	*Arhgef6*	2.5995	0.014680924
Iqgap2	2.8478	0.007619683	*Slpi*	2.5976	0.015894177
Chst12	2.8466	0.007651711	*Ptafr*	2.5976	0.024524194
Myo5a	2.8440	0.025193588	*Fgl2*	2.5916	0.012537033
Tjp2	2.8347	0.00914445	*Cd9*	2.5899	0.018748591

Gene ID	Log 2 Fold Change	Adjusted p-value	Gene ID	Log 2 Fold Change	Adjusted p-value
Alcam	2.5867	0.029385461	*Sphk1*	-2.6058	0.047454967
As3mt	2.5856	0.031384956	*Fgf2*	-2.6134	0.025833008
Ltb4r1	2.5759	0.018430476	*Gm16565*	-2.6807	0.020987633
Dhrs9	2.5528	0.037157487	*Padi3*	-2.6814	0.025833008
Ifngr1	2.5507	0.017656814	*Hdac11*	-2.6895	0.033967631
Arhgap18	2.5456	0.020509598	*Gm43429*	-2.7102	0.040141191
Tst	2.5376	0.044634913	*G530011O06Rik*	-2.7236	0.027506243
Plcg2	2.5360	0.020983308	*Slc25a21*	-2.7651	0.020983308
Trem3	2.5186	0.021336137	*Mast4*	-2.7709	0.03997003
Gng12	2.5121	0.025368605	*F830212C03Rik*	-2.8080	0.025989497
Plag1	2.4945	0.028756058	*Car1*	-2.8263	0.012552935
Fam114a1	2.4808	0.046228659	*Gdpd1*	-2.8345	0.022610022
Trem1	2.4798	0.023856314	*Atp2b4*	-2.8448	0.006264829
Fam84b	2.4793	0.044634913	*Cnnm1*	-2.8679	0.017873672
Dram1	2.4790	0.029627535	*Trbv13-2*	-2.8715	0.018455669
Tyrobp	2.4663	0.028511836	*Mid1*	-2.9279	0.035181044
Pik3r5	2.4567	0.026791608	*Hkdc1*	-2.9379	0.011575935
Timp2	2.4536	0.039832362	*Abcb4*	-2.9419	0.043621415
Arrdc4	2.4493	0.040436889	*Gm15726*	-2.9923	0.008541462
Itga5	2.4454	0.040436889	*Gm37368*	-3.0007	0.009481421
Tlr13	2.4446	0.025889705	*4933401H06Rik*	-3.0634	0.040436889
Abtb2	2.4395	0.030335267	*Rbm20*	-3.0824	0.005260718
Ncf4	2.4380	0.036876071	*Gm21817*	-3.0884	0.040794683
Gda	2.4358	0.020983308	*Sowaha*	-3.0888	0.008967509
Tlr6	2.4317	0.038891244	*Spin4*	-3.1353	0.043666066
Msra	2.4312	0.033287325	*Nmnat2*	-3.1534	0.004702454
Tmem55a	2.4105	0.040819144	*Crnde*	-3.2790	0.025540735
Plxnd1	2.4083	0.039832362	*Acp5*	-3.4981	0.014635565
S100a11	2.4079	0.039790742	*1700063H04Rik*	-3.5109	0.002121898
Hip1	2.3967	0.031557115	*Pmfbp1*	-3.5572	0.001147118
Cd52	2.3820	0.024524194	*Maob*	-3.6657	0.000701274
Itga2b	2.3646	0.022039622	*Gm38059*	-3.7000	0.020346609
Entpd1	2.3555	0.044634913	*Tspan8*	-3.7644	0.013784324
Qser1	2.3487	0.043365654	*Gata3*	-3.7965	0.000126706
Napsa	2.3401	0.036510638	*P4htm*	-3.8678	0.009278296
Ncoa1	2.3347	0.040815551	*Gm10825*	-3.9527	0.040436889
Aldh3b1	2.3343	0.047454967	*Slc30a2*	-3.9719	0.030697474
Pik3ip1	2.3021	0.044781897	*C1qtnf4*	-4.0384	0.00248504
Ipcef1	2.2913	0.039823027	*Sphk1*	-4.0883	0.045913063
Ddx60	2.2735	0.049341605	*Gm42784*	-4.3317	0.003248461
Scd1	-2.3459	0.049341605	*Zfp185*	-4.5258	0.008106877
Naaladl1	-2.4453	0.045987471	*Arhgef25*	-4.5347	0.003569835
Samd11	-2.4876	0.033287325	*Fgf2os*	-4.7687	0.033287325
Dtx3	-2.5085	0.026717686	*Nags*	-4.7690	2.36E-05
Fgf10	-2.5171	0.039790742	*Stard13*	-4.9495	0.00239471
Il1rl2	-2.5854	0.041139826	*Magi1*	-5.6302	0.036840441

Gene ID	Log 2 Fold Change	Adjusted p-value	Gene ID	Log 2 Fold Change	Adjusted p-value
P4ha3	-6.2278	0.007676119	*Dbx2*	-6.5337	3.06E-10
			Lonrf2	-8.1378	7.72E-10