Dissecting the role of β-catenin in AML by global epigenetic sequencing

Lynn, Claire

Awarding institution:
King's College London

Dissecting the Role of β-catenin in AML by Global Epigenetic Sequencing

Claire Lynn

A thesis presented for the degree of

Doctor of Philosophy

Department of Haematological and Molecular Medicine

King's College London

2019

Declaration

I hereby declare that this thesis is my own work and has not been submitted to any institution for any award. Where other sources of information have been used, they have been acknowledged. The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

Claire Lynn

2019

Abstract

Wnt–β-catenin signalling regulates diverse developmental and homeostasis processes. Aberrant regulation of this pathway is observed in numerous cancers. β-catenin, whilst largely dispensable in adult haematopoiesis, has a critical role in the development, disease progression and drug resistance of Acute Myeloid Leukaemia (AML). One of the most frequent chromosomal translocations in AML involves KMT2A, also known as Mixed Lineage Leukaemia protein (MLL). MLL-fusions initiate transformation of haematopoietic stem cells (HSCs) and progenitors such as the Granulocyte-Macrophage Progenitors (GMPs) into Leukaemia Stem Cells (LSCs). Recently we demonstrated that β-catenin is required for the development of MLL-ENL transformed LSCs originating from GMPs but not from HSC enriched populations. This study aims to uncover key targets of β-catenin and dissect mechanisms by which β-catenin controls transcription in LSCs arising from distinct cellular haematopoietic populations. To investigate β-catenin targets in AML, haematopoietic progenitors of specific origin were taken from a β-catenin conditional mouse model and transformed by retroviral MLL-ENL expression.

To this end, we performed global transcriptional profiling by RNA sequencing (RNA-Seq) on transformed cells in the presence or absence of β-catenin. β-catenin has been shown to recruit histone modifying enzymes and other epigenetic regulators; to elucidate mechanisms by which β-catenin modulates its targets in AML, ChIP sequencing (ChIP-Seq) was produced for promoter and enhancer defining histone modifications. β-catenin targets in LSCs were found to be enriched in cytokine signalling and immune pathways and contrast with those in other cell types suggesting that β-catenin may perform unique roles in AML. Interestingly, one of the few genes differentially up regulated in GMP derived LSCs upon β-catenin deletion included *Gadd45a*, a DNA damage inducible growth arrest mediator. Whilst showing little effect in vitro, *Gadd45a* overexpression eliminated leukaemia in vivo. Combined global ChIP-Seq and RNA-Seq analyses revealed β-catenin dependant changes at promoter regions suggesting that β-catenin could be actively involved in histone modification. However, further study with improved experimental design is required to ensure that the observed effects are specific to β-catenin activity.

Acknowledgements

This work would not have been possible, had it not been for the support I received from many great people. I would like to thank Prof Eric So and Prof Boris Lenhard for their advice, patience and securing the funding for this opportunity. I am extremely grateful to have had a fantastic academic panel to look up to and look out for me over the years, thank you Prof. Linda Barber, Prof. Rebecca Oakey and Prof. Francesca Ciccarelli.

I would like to thank my colleagues, past and present from both the So and Lenhard groups for their support. In particular, I would like to thank the late Amanda Wilson for her dedication, crucial support and kindness. Bernd Zeisig for answering all of my questions. Rick (Kar Lok Kong) for teaching me how to perform ChIP and retain some sanity, with fantastic patience. Jackie (Jie Chen) and Priscilla Lau for sharing debate and comic relief, especially during our sequencing library preparation madness. Fung Kan, for your help with Fluorescent-Activated Cell Separation (FACS) analysis and wisdom. Haoli Li, for bringing back my passion for bioinformatics and stimulating conversation after my accident. Magdalena Zarowiecki, for your phenomenal kindness, encouragement and advice while writing this thesis. You gave me the confidence to just keep writing, I could never express my gratitude enough!

My fellow PhD students, Abdullah, Henry, Nelson, Ray, Siyi and Mickey, your positivity, determination and resilience has been a great source of inspiration to me and I am honoured to have shared the PhD experience with you. Special thanks to Ray for his extraordinary kindness in teaching me how to do MTT and immunofluorescence assays among other things.

I would also like to take this opportunity to thank my parents, siblings (James, Louise and Amanda), and friends for their love and support. Last and certainly not least, I would like to thank my dear companion Daniel for loving and supporting me through sickness and health on this journey.

Contents

Declaration 2

Abstract 3

Acknowledgements 4

Contents 9

List of Figures 12

List of Tables 14

List of Acronyms 14

1 Introduction **18**

1.1 The Haematological System 19

1.2 Cell Identity and Transcriptional Regulation 22

1.2.1 DNA Methylation 22

1.2.2 Chromatin Organisation 23

1.2.3 The Histone Modification Code 24

1.2.4 Transcriptional read-out of histone modification 25

1.3 The Epigenetic Landscape of Haematopoietic Cells 27

1.3.1 DNA Methylation 27

1.3.2 Chromatin Remodelling 28

1.3.3 Histone Acetylation 28

1.3.4 Bivalent Promoters 29

1.3.5 Mixed lineage leukaemia protein 30

1.3.6 Transcription Factors 31

1.3.7 Characterisation of the Epigenome in Haematopoiesis 33

1.4 Leukaemia and its Subtypes . . . 33
1.5 Gene Fusions . . . 37
1.6 The Origin of Leukaemia Stem Cells . . . 40
1.6.1 Retroviral Cell-of-Origin Models . . . 41
1.6.2 Knock-in Mouse Cell-of-Origin Models . . . 42
1.7 Wnt Signalling in Development and Disease . . . 43
1.7.1 The Role of β-catenin in Leukaemia Stem Cells . . . 47
1.8 Project Aims . . . 50

2 Methods **51**

2.1 Production of β-catenin conditional pre-LSCs . . . 51
2.1.1 β-catenin conditional mouse lines . . . 51
2.1.2 Genotyping PCRs . . . 53
2.1.3 Haematopoietic stem cell and progenitor sorting . . . 54
2.1.4 Retroviral transduction and transformation assay . . . 55
2.1.5 In vivo Leukaemogenesis assay . . . 56
2.2 Cell Culture and *Ctnnb1* Knock-out . . . 57
2.3 Quantitative Reverse Transcriptase Polymerase Chain Reaction (RT-qPCR) . . . 58
2.4 Western Blot . . . 58
2.5 RNA-Sequencing (RNA-Seq) and analysis . . . 59
2.5.1 A note on experimental design . . . 59
2.5.2 RNA-Seq Libraries . . . 60
2.5.3 RNA-Seq Processing . . . 61
2.6 *Gadd45a* overexpression . . . 62
2.7 *Gadd45a* knockdown . . . 63
2.7.1 MTT assay . . . 65
2.7.2 Cytospin . . . 65
2.7.3 Annexin Staining . . . 66
2.8 Chromatin Immunoprecipitation coupled to high-throughput sequencing (ChIP-Seq) Generation . . . 66
2.8.1 ChIP q-PCR . . . 69
2.8.2 ChIP-Seq library preparation . . . 70

2.9 ChIP-Seq processing and quality assessment . . . 71

2.10 Immunofluorescence . . . 72

3 Transcriptional profiling of β-catenin conditional cell lines: Combined Batch Analysis 73

3.1 Introduction . . . 73

3.2 Methods . . . 75

3.2.1 RNA-Seq Generation . . . 75

3.2.2 Confirming β-catenin depletion . . . 76

3.2.3 RNA-Seq Analysis . . . 76

3.3 Results . . . 79

3.3.1 Validation of β-catenin knock-out . . . 79

3.3.2 Principal Component Analysis reveals batch effects and bias . . . 81

3.3.3 Batch Correction for the effect of cell-of-origin . . . 82

3.3.4 Biological themes related to the combined cell-of-origin signature . . . 85

3.3.5 Batch Correction for the effect of β-catenin deletion in GMP^{ME} . . . 86

3.3.6 Batch Correction for the effect of β-catenin deletion in LSK^{ME} . . . 86

3.3.7 Putative β-catenin target genes in MLL-ENL pre-LSCs are largely independent of cell-of-origin . . . 89

3.3.8 Biological themes related to the combined β-catenin knock-out signature 89

3.4 Discussion . . . 92

4 Transcriptional profiling of β-catenin conditional cell lines: Separate Batch Analysis 94

4.1 Introduction . . . 94

4.2 Methods . . . 95

4.2.1 Differential expression analysis . . . 95

4.2.2 Comparison of targets with published β-catenin knock-out expression data 96

4.2.3 Comparison of targets with other studies utilising Rosa-Cre-ER . . . 97

4.3 Results . . . 97

4.3.1 The observable cell-of-origin transcriptomic signature is reduced in 2015 samples . . . 97

4.3.2 β-catenin signature genes significantly overlap with those in the published dataset . . . 100

4.3.3 Putative β-catenin targets are largely unique to MLL-ENL transformed GMP and LSK . . . 103

4.3.4 Many of our β-catenin targets are also seen in experiments using the Cre-ER system in our lab . . . 106

4.4 Discussion . . . 108

5 The role of *Gadd45a* in GMP derived MLL-ENL LSCs 112

5.1 Introduction . . . 112

5.2 Methods . . . 113

5.3 Results . . . 114

5.3.1 *Gadd45a* overexpression shows little phenotypic effect in vitro . . . 114

5.3.2 *Gadd45a* overexpression in GMP^{ME} eliminates leukaemia in vivo . . . 115

5.3.3 *Gadd45a* knock-down causes an increase in proliferation and self-renewal . 117

5.3.4 β-catenin deletion results in apoptosis and differentiation in vitro . . . 118

5.3.5 β-catenin deletion resistance emerges in GMP^{ME} in vivo . . . 120

5.3.6 *Gadd45a* overexpression in LSK^{ME} eliminates leukaemia in vivo . . . 123

5.4 Discussion . . . 125

6 Integrative ChIP-Seq analysis of cell-of-origin specific β-catenin conditional cell lines 128

6.1 Introduction . . . 128

6.2 Methods . . . 129

6.2.1 ChIP-Seq Generation . . . 129

6.2.2 Correlating ChIP-Seq with RNA-Seq . . . 131

6.2.3 Regulatory state segmentation . . . 131

6.2.4 Gene-centric quantitative differential ChIP-Seq analysis . . . 132

6.2.5 Enhancer Analysis . . . 133

6.2.6 Motif Enrichment . . . 134

6.2.7 Comparisons with published β-catenin ChIP-Seq . . . 134

6.2.8 β-catenin ChIP in constitutively active β-catenin cell lines . . . 135

6.3 Results . . . 136

6.3.1 Successfully enriched ChIP was obtained . . . 136

6.3.2 Successfully enriched ChIP-Seq data were obtained . . . 138

6.3.3 Histone modification data demonstrates expected patterns at the TSS . . 140

6.3.4 Histone modification signal at gene loci correlates with corresponding gene expression data 141

6.3.5 Variation and Correlation between ChIP-Seq samples 143

6.3.6 Histone modification data segment the genome into biologically relevant chromatin states 145

6.3.7 Variation of histone modification observed at the HoxA locus 147

6.3.8 Gene-centric quantitative analysis of histone modification between samples with different cells-of-origin 150

6.3.9 Pathway analysis of differentially modified genes between different cells-of-origin 155

6.3.10 Cell-of-Origin associated Enhancer analysis 156

6.3.11 β-catenin related changes in histone modifications 157

6.3.12 β-catenin related changes in histone modifications: pathways 163

6.3.13 β-catenin related changes in histone modifications motifs 166

6.3.14 Comparison to publicly available β-catenin ChIP-Seq 167

6.3.15 β-catenin exon three cell lines could be used for further work 167

6.4 Discussion 170

7 Conclusion **174**

7.1 Suggestions for Future Work 177

References **201**

List of Figures

1.1 Current models of lineage fate determination in the haematopoietic system . . . 21

1.2 The MLL Complex. 31

1.3 Epigenetic control of myeloid differentiation 32

1.4 Mechanisms of transcriptional regulation by MLL-fusions 39

1.5 Cell-of-origin and clonal evolution model of LSCs 40

1.6 Canonical Wnt signalling 45

2.1 β-catenin conditional alleles 52

2.2 Schematic of experimental set-up 61

2.3 ChIP-Sequencing Schematic 67

3.1 Validation of RNA-Seq samples 80

3.2 Experimental variables and batch effects separate samples in principal component analysis (PCA) 82

3.3 Batch correction to extract the cell-of-origin effect 84

3.4 Biological themes of cell-of-origin signature 85

3.5 Batch correction for loss of β-catenin in GMP^{ME} 87

3.6 Batch correction for loss of β-catenin in LSK^{ME} 88

3.7 Most of the transcriptional effects of β-catenin deletion are independent of cell-of-origin 90

3.8 GO enrichment for genes deregulated after β-catenin knock-out. 91

4.1 Detectable cell-of-origin signature is markedly reduced in 2015 samples 99

4.2 Comparison of transcriptomic effects of β-catenin KO across experiments 101

4.3 *Gadd45a* upregulation is a unique effect of β-catenin deletion in GMP^{ME} 102

4.4 Comparison of β-catenin knock-out signatures with known targets and across tissues . . . 103

4.5 GSEA enrichment for genes deregulated after β-catenin knock-out. . . . 104

4.6 Specificity of β-catenin target genes compared to others using 4-OHT . . . 106

5.1 Schematic illustrating the experimental set up for this chapter. . . . 113

5.2 In vitro Overexpression of *Gadd45a* . . . 115

5.3 *Gadd45a* overexpression eliminates GMP derived MLL-ENL leukaemia . . . 116

5.4 In vitro knockdown of *Gadd45a* . . . 117

5.5 Effect of β-catenin KO with 20 nM 4-OHT. . . . 119

5.6 β-catenin KO cells showed no significant difference in two transplant experiments. 121

5.7 Survival curve for transplant of cells used for sequencing in this study . . . 122

5.8 *Gadd45a* overexpression eliminates LSK derived MLL-ENL leukaemia . . . 124

6.1 ChIP-qPCR for positive and negative control regions. . . . 137

6.2 Heatmaps and average profiles for all histone marks from the TSS to 10 kb into the gene body . . . 141

6.3 Correlation between RNA-Seq and ChIP-Seq . . . 142

6.4 Relationships between ChIP-Seq samples . . . 143

6.5 Global correlation between ChIP-Seq samples . . . 144

6.6 Global chromatin state assignment and transitions between experimental conditions 146

6.7 Chromatin state changes between conditions at the TSS . . . 147

6.8 Transcriptomic and epigenetic variation at the HoxA locus in MLL-ENL transformed cells . . . 148

6.9 Epigenetic variation at the HoxA locus in normal cells . . . 149

6.10 Summary of significantly differentially modified regions between LSK^{ME} and GMP^{ME} . . . 152

6.11 H3K27ac, H3K4me3 and gene expression at differentially modified and/or expressed genes between LSK^{ME} and GMP^{ME} . . . 153

6.12 Signal tracks of loci exhibiting cell-of-origin dependent changes in histone modification . . . 154

6.13 GO and KEGG analysis for genes with differentially bound promoters between LSK^{ME} and GMP^{ME} . . . 155

6.14 Motif analysis of enhancers exhibiting cell-of-origin specific activity . . . 156

6.15 Summary of significantly differentially modified regions following β-catenin knock-out 158

6.16 H3K4me3 and H3K27ac signal and gene expression at differentially modified or expressed genes following β-catenin knock-out 160

6.17 Histone modification and RNA-Seq tracks at the *Gadd45a* locus 161

6.18 ChIP signal at selected significantly differentially modified regions following β-catenin knock-out. 162

6.19 Gene ontology and pathways associated with increases in H3K4me3 following β-catenin deletion 163

6.20 Gene ontology and pathways associated with decreases in H3K27ac following β-catenin deletion 164

6.21 Validation of conditionally activated β-catenin MLL-ENL transformed cell lines . 169

List of Tables

1.1 Recognised ENCODE histone modifications and variants 26

1.2 FAB classification of AML 35

1.3 WHO classification of AML 35

1.4 The 6 most common MLL-fusion partners 38

1.5 TFs proposed to recruit β-catenin 46

1.6 Wnt pathway components expressed in the bone marrow environment 48

2.1 LOF *Ctnnb1*$^{\text{Kem/flox}}$ genotyping primers 53

2.2 GOF *Ctnnb1*$^{\text{Mmt/flox}}$ genotyping primers 53

2.3 RT-qPCR primers 58

2.4 Primers to amplify *Gadd45a* ORF and add restriction sites 62

2.5 *Gadd45a* shRNA Sequences 64

2.6 ChIP Lysis Buffers 68

2.7 ChIP Wash Buffers 68

2.8 ChIP-qPCR primers 70

3.1 All RNA-Seq samples produced in our lab and analysed in this chapter 77

3.2 Quality Control metrics for 2015 RNA-Seq samples 77

4.1 Published β-catenin deletion transcriptomic data 96

4.2 In-house 4-OHT treatment transcriptomic data 97

4.3 Significance of overlaps of differentially expressed genes following β-catenin deletion in LSK$^{\text{ME}}$ & GMP$^{\text{ME}}$ 102

4.4 GSEA analysis of Wnt/β-catenin related pathways 105

4.5 Significance of overlaps of differentially expressed genes following 4-Hydroxytamoxifen (4-OHT) treatment in other studies 107

6.1 ChIP-Seq samples generated for this study 130

6.2 Available published β-catenin ChIP-Seq data 135

6.3 β-catenin ChIP-qPCR primers 136

6.4 Quality control results for ChIP-Seq generated for this study 139

6.5 Enriched motifs found within promoter regions of differentially H3K27ac marked genes 166

6.6 Published β-catenin ChIP QC 168

List of Acronyms

4-OHT 4-Hydroxytamoxifen

5mC 5-methyl of Cytosine

ABL Biphenotypic Leukaemia

AGM Aorta-Gonad Mesonephros

ALL Acute Lymphoblastic Leukaemia

AML Acute Myeloid Leukaemia

B-ALL B-cell Acute Lymphoblastic Leukaemia/Lymphoma

BAF BRG1- or BRM-Associated Factors

BM Bone Marrow

CGI CpG Island

ChIP Chromatin Immunoprecipitation

ChIP-Seq Chromatin Immunoprecipitation coupled to high-throughput Sequencing

ChIP-qPCR ChIP quantitative PCR

CLL Chronic Lymphocytic Leukaemia

CLP Common Lymphoid Progenitor

CML Chronic Myeloid Leukaemia

CMP Common Myeloid Progenitor

CpG Cytosine followed by a Guanine

CPM Count Per Million

CRCs Chromatin Remodelling Complexes

CRISPR Clustered Regularly Interspaced Short Palindromic Repeat

CSC Cancer Stem Cell

DNMT DNA Methyl-Transferase

eRNA enhancer RNA

ES Embryonic Stage

ESC Embryonic Stem Cell

ETOH Ethanol

ETP Early T-cell Precursor

FAB French-American-British

FACS Fluorescent-Activated Cell Separation

FDR False Discovery Rate

GLM Generalised Linear Model

GMP Granulocyte-Monocyte Progenitor

GMP$^{\mathbf{ME}}$ MLL-ENL transformed GMP

GO Gene Ontology

GOF Gain Of Function

GSEA Gene Set Enrichment Analysis

GTF General Transcription Factor

HAT Histone Acetyl-Transferase

HDAC Histone De-ACetylase

HSC Haematopoietic Stem Cell

KAT Histone Lysine Acetyl-Transferase

KDM Histone Lysine Demethylase

KEGG Kyoto Encyclopedia of Genes and Genomes

KMT histone Lysine Methyl Transferase

KO Knock-Out

LATF Leukaemia Transcription Factor

LIC Leukaemia Initiating Cell

LMPP Lymphoid-primed Multi-Potent Progenitor

LncRNA Long non-coding RNA

LOF Loss Of Function

LRT Log Ratio Test

LSC Leukaemia Stem Cell

LSK Lineage-specific negative, Sca-1 and c-Kit positive

LSK$^{\mathbf{ME}}$ MLL-ENL transformed LSK

LT-HSC Long Term reconstituting HSC

MACS Magnetic-Activated Cell Sorting

MBD Methyl-CpG Binding Domain

MDS Myelo-Dysplastic Syndrome

MEP Megakaryocyte-Erythroid Progenitor

miRNA micro RNA

MLL Lysine Methyltransferase 2A (KMT2A), also known as Mixed Lineage Leukaemia protein

MPP Multipotent Progenitor

NITC Nonsense Induced Transcriptional Compensation

NK Natural Killer Cell

NuRD Nucleosome Remodelling and Deacetylase

ORF Open Reading Frame

PCA Principal Component Analysis

PcG Polycomb Group protein

PCR Polymerase Chain Reaction

pol II RNA polymerase II

PRC1 Polycomb Repressive Complex 1

PRC2 Polycomb Repressive Complex 2

Pre-LSC Pre-Leukaemic Stem Cell

PRMT Protein arginine N-methyltransferase

proB B-cell Precursor

RBC Red Blood Cell

RNA-Seq RNA Sequencing

RPKM Reads Per Kilobase of transcript per Million mapped reads

RT-qPCR Reverse Transcriptase quantitative PCR

RTTA Retroviral Transduction and Transformation Assay

SAM Sequence Alignment Map

shRNA Short Hairpin RNA

SM Sorting Media

SNP Single Nucleotide Polymorphism

SRA SET and RING finger-Associated

ST-HSC Short Term reconstituting HSC

T-ALL T-cell Acute Lymphoblastic Leukaemia/Lymphoma

TF Transcription Factor

TFBS TF Binding Site

TSS Transcription Start Site

UTR Untranslated Region

WHO World Health Organisation

WRE Wnt Responsive Element

WT Wild-type

ZF Zinc Finger

Chapter 1

Introduction

Acute Myeloid Leukaemia (AML) is a cancer which originates from haematopoietic stem or progenitor cells which in normal circumstances would fully differentiate and give rise to mature blood cells. Malignant transformation prevents progenitor cells from differentiating into mature blood and dramatically increases proliferation, leading to a pathogenic accumulation of immature cells. 2,856 cases of AML, 0.8% of all new cancer cases, were diagnosed in the UK in 2012. For those diagnosed with AML, aged 14 or younger, 66% survive 5 years or more, at 15-24 this drops to 60%, at 25-64 it is 40% and at 65 or older just 5% (Cancer Research UK 2015).

Improvements in the treatment of AML over the past few decades can be mostly attributed to supportive care and better diagnosis (Thein et al. 2013). Therefore therapies that effectively target the biological mechanisms underpinning the disease need to be developed. To unravel the causes of AML and therefore novel targets for therapy, we must first understand what blood is composed of and the processes involved in blood formation and transformation.

1.1 The Haematological System

Haematopoiesis is a cascade of cellular and molecular events resulting in the formation of all blood cellular constituents. Although highly complex, haematopoiesis is one of the most comprehensively studied developmental systems. Thus, the haematological system has become somewhat of a paradigm for stem cell maintenance and renewal. Classical haematopoiesis is organised into a pyramidal hierarchical structure, with Haematopoietic Stem Cells (HSCs) at the apex. Rare, long-lived, multipotent HSCs give rise to larger populations of increasingly restricted, committed intermediates, or progenitors, ultimately producing terminally differentiated, functional blood cells (Till and McCulloch 1980). HSCs are classed into two groups, Long Term reconstituting HSCs (LT-HSCs) which replenish the stem cell pool and Short Term reconstituting HSCs (ST-HSCs) which go on to form all blood lineages (Weissman et al. 2001).

Before distinct lineages are specified, ST-HSCs generate Multipotent Progenitors (MPPs), MPPs are uncommitted yet display weaker self-renewal potential. From MPPs, myelopoiesis and lymphopoiesis are segregated by the formation of their common oligopotent progenitors, Common Myeloid Progenitors (CMPs) and Common Lymphoid Progenitors (CLPs) (Weissman et al. 2001). From CMPs arise Granulocyte-Monocyte Progenitors (GMPs) and Megakaryocyte-Erythroid Progenitors (MEPs), progressing to dendritic cells, monocytes, granulocytes, erythrocytes, megakaryocytes and platelets. From CLPs arise Early T-cell Precursors (ETPs) and B-cell Precursors (proBs), leading to Natural Killer Cells (NKs), T and B cells. The current models of haematopoiesis in mouse and human including the surface marker expression of each individual population, is illustrated in Figure 1.1.

The process of haematopoiesis is understood in great detail due to pioneering work in mice, leading to the discovery and functional assessment of haematopoietic populations (Weissman et al. 2001). In the mouse embryo, primitive haematopoiesis begins in the yolk sac blood islands at Embryonic Stage (ES) 7.5 (day 7.5 after fertilisation), producing firstly Red Blood Cell (RBC) precursors, followed by lymphoid precursors (Haar and Ackerman 1971). Other haematopoietic sites emerge once the circulatory system begins to develop. HSCs emerge at the Aorta-Gonad Mesonephros (AGM) region at ES10.5 and then in the placenta. Fully fledged haematopoiesis occurs in the foetal liver (ES13.5), thymus, spleen and then finally in the Bone Marrow (BM) after birth (Orkin and Zon 2008; Costa et al. 2012).

The concept of the HSC existed over 100 years ago, when Russian biologist Maximow (1909) explained observations of morphological diversity of bone marrow by suggesting that haematopoiesis arises from a common precursor. The atomic era brought irrefutable evidence for a HSC population when radiation exposure was found to cause lethal BM failure. It was discovered that exposed animals could be rescued after injection of healthy spleen or BM cells (Lorenz et al. 1951). Models of haematopoiesis have been built upon cell purification, transplantation and in vitro clonal assays. Firstly cell separation was done by density, sensitivity to antimitotic agents and then surface glycoprotein. Cell surface markers or differentiation "antigens" allowed for the identification, isolation and observation of stem cells and progenitors using monoclonal antibodies. A rare population of cells displaying HSC characteristics was identified in mice (Spangrude et al. 1988; Okada et al. 1992). This population is lineage-specific negative, Sca-1 and c-Kit positive, therefore commonly known as Lineage-specific negative, Sca-1 and c-Kit positive (LSK) cells.

Cell surface markers have been better refined over the years (Figure 1.1) and are now commonly utilised in FACS or Magnetic-Activated Cell Sorting (MACS) to isolate the specific haematopoietic populations with very good purity (Yeung and So 2009). Functional assays continue to reform the standard model, for example, the discovery of Lymphoid-primed Multi-Potent Progenitor (LMPP) in mouse Adolfsson et al. (2005) revealed that the first lineage branch after HSC determines megakaryocyte/erythroid potential, rather than just myeloid/lymphoid. Fascinatingly, by monitoring HSC divisions in mouse, it was demonstrated that MEPs can arise directly from HSCs (Yamamoto et al. 2013).

Human haematopoiesis has been considered largely consistent with mouse haematopoiesis (Doulatov et al. 2012). Recently, the widely accepted model of haematopoiesis has been challenged in humans. (Notta et al. 2015) hypothesised a developmental shift to a "two tier" hierarchy in adult human haematopoiesis. Using a improved sorting scheme to better resolve the human CD34+ cell compartment in the different tissues, they revealed that whilst large, distinct populations of oligopotent progenitors were found in foetal liver, only multipotent and unipotent progenitor classes prevailed in adult bone marrow. This suggests that oligopotent progenitors with the same immunophenotype, thus belonging to the same purification fraction, display different levels of differentiation potential. It is therefore vital that we take care in applying studies of murine haematopoiesis to human haematopoiesis.

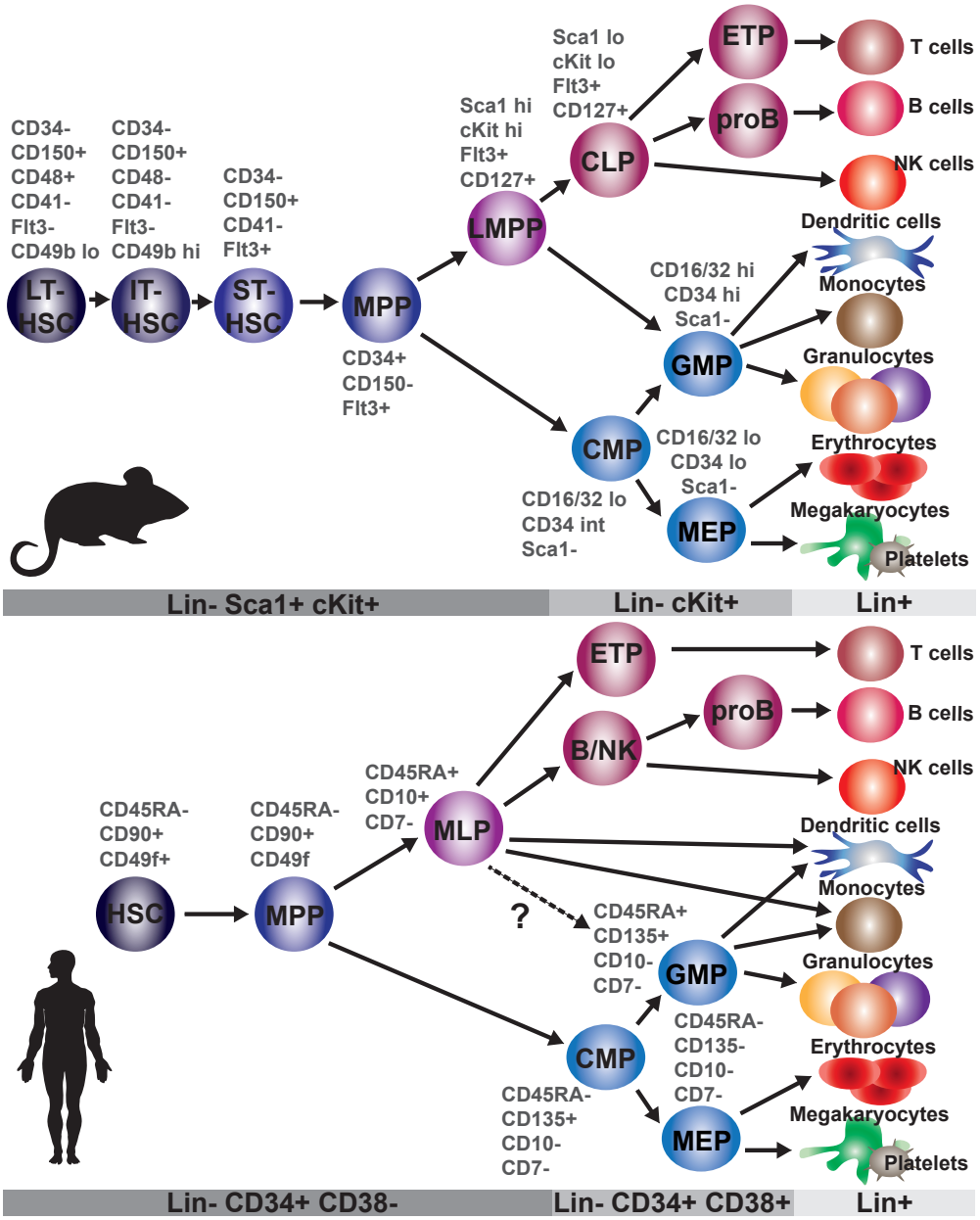

Figure 1.1: Current models of lineage fate determination in the haematopoietic system Adapted from Doulatov et al. (2012).

Haematopoiesis is governed by precise control of gene expression. Expression of genes promoting multipotency and self-renewal are shut down whilst expression of genes essential for specific cell functions are tuned up along branches of differentiation. To further understand the process of haematopoiesis on a molecular level, we must recognise the mechanisms of gene expression control.

1.2 Cell Identity and Transcriptional Regulation

Due to precise spatial-temporal regulation of gene expression, cells that are genetically identical become distinctive cell populations with defined roles in discrete tissues. The study of gene expression regulation is called epigenetics. "Epi-" comes from Greek, meaning "outside of" or "on top of", thus the literal meaning of epigenetics is the external modification of DNA. The epigenetic landscape concept was famously illustrated by Conrad Waddington in (Waddington 1957). Marbles or cells, roll down a hill, competing for the grooves or lineages, then finally come to rest at different locations at the bottom of the hill or their terminal cell fates. When the field first emerged, epigenetics was defined as "The study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence" by Russo et al. (1996). Modern usage of the term "epigenetics" commonly describes any perturbations of chromatin state (Bird 2007).

Approximately two meters of DNA are packed into the nucleus of a human cell. To protect the genome from damage and arrange it so that functional loci can be accessed, DNA is packaged with proteins such as histones into a structure known as chromatin. For a gene to be transcribed, the basic transcriptional machinery, consisting of RNA polymerase II (pol II) and General Transcription Factors (GTFs), must bind to the core promoter. The core promoter, usually containing a TATA box, named by its literal nucleotide sequence, is immediately upstream of the Transcription Start Site (TSS) (Butler and Kadonaga 2002). Transcription is enhanced by interaction of further regulatory DNA elements; the proximal promoter and enhancer regions contain recognition sites for Transcription Factors (TFs) and other transcriptional activators, recruiting coactivators to the core promoter. Regulation of gene expression is mediated by a combination of many factors: DNA methylation, chromatin remodelling, histone variants, histone modifications, histone modification readers, transcription factors, co-activators or repressors and various non-coding RNAs.

1.2.1 DNA Methylation

DNA methylation is the quintessential epigenetic mechanism as much is known about the mechanisms of its heritability and its established effects on gene expression. DNA Methyl-Transferases (DNMTs) methylate the 5-methyl of Cytosine (5mC) in DNA. This often occurs at a Cytosine followed by a Guanine (CpG) but can occur at non-CpGs (Lister et al. 2009; Xie

et al. 2012). The binding of proteins can be inhibited by methylation. 5mC readers are found in; the Methyl-CpG Binding Domain (MBD) family, the Zinc Finger (ZF) family and the SET and RING finger-Associated (SRA) family (Klose and Bird 2006). Cytosine to Tyrosine mutations are promoted by methylation as 5mC is susceptible to deamination. CpG Islands (CGIs) are regions enriched for CpGs and are broadly unmethylated. 50% of CGIs are associated with promoters of predominantly housekeeping genes (Illingworth and Bird 2009), whilst some are developmentally regulated (Deaton et al. 2011). While CGI promoter methylation implies stable repression, methylated gene bodies do not prevent elongation but a skew in methylation in exons is thought to affect splicing regulation (Jones 2012).

1.2.2 Chromatin Organisation

Primarily, the level of chromatin compaction influences DNA accessibility and thus its activity, several distinct forms have been characterised; heterochromatin is compact, often contains repetitive elements and is generally transcriptionally silenced whilst euchromatin is open permitting active transcriptional regulation. Chromatin is firstly organised into nucleosomes, 146 bp of DNA wraps around an octamer arrangement of histone proteins, inter-dispersed by linker DNA, often visually described as "beads on a string" (Luger et al. 1997). Histone octamers contain two copies of the canonical histones: H2A, H2B, H3 and H4. Further compaction of chromatin is achieved in nucleosome arrays and then chromatin fibres.

Chromatin arrangements are not only a scaffold for DNA, but a precise enzyme controlled method of transcriptional regulation. ATP hydrolysis is used by Chromatin Remodelling Complexes (CRCs) to remove nucleosomes, slide them along DNA or exchange histone proteins (Hargreaves and Crabtree 2011). Nucleosomes are shifted to allow transcriptional, replication and DNA repair machinery to make contact. Nucleosomes are also shifted or removed to deny or allow access to protein recognition sites (Narlikar et al. 2013). Canonical histones are dependent on replication for their incorporation into chromatin, however, replication independent histone variants can be incorporated into chromatin via CRCs.

Histone variants have unique properties over their canonical counterparts, some have specific functions. For example, H2A.Z locates at the $5'$ end of genes and may control transcriptional regulation (Zilberman et al. 2008; Catherine 2013). Neighbouring nucleosomes interact via their unstructured N-terminal and C-terminal tails. More often than their globular domains, unstructured N-terminal histone tails are modified by acetylation, methylation,

phosphorylation, sumoylation and ubiquitination, facilitating structural regulation. Other modifications, most with unknown function have been found (Tan et al. 2011).

1.2.3 The Histone Modification Code

Patterns and sequences of histone modifications or "marks" along histone tails are interpreted somewhat as a code, each combination resulting in a specific regulatory effect (Strahl and Allis 2000). However, only a small proportion of the potential combinations have been observed (Rando 2012), but functional classes of genes show some preference with specific combinations (Kurdistani et al. 2004). Already characterised in 1964 (Allfrey and Mirsky 1964), two of the best studied histone marks are acetylation and methylation of lysine residues. Histone tail modifications are abbreviated by the histone protein name, followed by the one letter amino acid code of the residue, the amino acid position in the tail and finally its chemical group. For example, H3K4me3 can be understood as a tri-methyl group on a lysine residue at the fourth amino acid position of Histone 3.

Histone acetylation of lysine residues is performed by Histone Lysine Acetyl-Transferases (KATs) and reversed by Histone De-ACetylases (HDACs). The enzymes responsible for the regulation of lysine acetylation were discovered only relatively recently in 1995 (Brownell and Allis 1995; Kleff et al. 1995; Taunton et al. 1996). Acetylation levels are balanced by dynamic, antagonistic action and even interaction of the enzymes. Found at active promoters and open chromatin, acetylation is usually an activating modification (Bannister and Kouzarides 2011). Lysine is positively charged, therefore addition of a negatively charged acetyl group neutralises this charge and reduces its interaction with negatively charged DNA increasing its availability and activity (Zentner and Henikoff 2013). Chromatin modulators with bromodomains can also bind and regulate chromatin structure at acetylated lysines (Bannister and Kouzarides 2011).

Methylation marks can be found at arginine and lysine residues on histone tails (Bannister and Kouzarides 2011). Varying regulation is applied by mono-, di- or trimethylation occurring at lysine residues. Long after methylation modifications were discovered, the first histone Lysine Methyl Transferase (KMT), KMT1A was found in 2000 (Rea et al. 2000). KMTs are specific for both methylation level and peptide residues (Bannister and Kouzarides 2011). Their antagonists, Histone Lysine Demethylases (KDMs) have also been discovered only recently, the first, KDM1A, was found in 2004 (Shi et al. 2004).

Methylation marks can have an activating or repressive effect depending on the residues position and methylation level (Bannister and Kouzarides 2011), this is because unlike acetylation, the charge of a histone is not altered by methylation. For example, at the promoter, H3K4me3 promotes gene expression (Barski et al. 2007). Conversely, H3K27me3 is a repressive mark generated and read by Polycomb Group proteins (PcGs). The Polycomb Repressive Complex 2 (PRC2) adds this mark and Polycomb Repressive Complex 1 (PRC1) recognises or "reads" H3K27me3 and causes chromatin to condense (Simon and Kingston 2013).

1.2.4 Transcriptional read-out of histone modification

Cis-regulatory elements, such as enhancers and promoters have distinctive histone modification patterns as shown in Table 1.1. Enhancers are genetic elements between 200 and 500 bp which boost transcription at a promoter up to thousands of kilobases away (Calo and Wysocka 2013). Enhancers regulate gene expression by physically bringing the promoter in to contact with transcription factors bound to the enhancer (Calo and Wysocka 2013; Smallwood and Ren 2013).

In recent years, the use of next generation sequencing to investigate epigenetics genomewide, such as Chromatin Immunoprecipitation coupled to high-throughput Sequencing (ChIP-Seq) has greatly accelerated. ChIP-Seq begins by fixing DNA using formaldehyde or a combination of other fixing agents. This chemically crosslinks protein to DNA, the DNA is sheared via enzymes or physical sonication. Antibodies specific to a histone modification or other proteins can be used to immunoprecipitate the DNA linked to the protein of interest. This DNA is purified, decrosslinked, sequencing adapters are ligated and the DNA is then subjected to high-throughput sequencing.

The ENCODE project identified 399,124 regions with enhancer-like features over 1640 datasets in 147 human cell types (ENCODE Project Consortium and others 2012). Enhancers commonly contain DNAse hypersensitive sites, are bound by p300 and SMARCA4 and are enriched for H3K4me1 but lack H3K4me3. Active enhancers are marked by H3K27ac and H3K4me1, poised enhancers are marked with H3K4me1, H3K4me2 and H3K27me3 (Andersson et al. 2014). Functional experiments show that the epigenetic characterisation of enhancers is in fact highly predictive of their activity (Calo and Wysocka 2013).

Table 1.1: Recognised ENCODE histone modifications and variants Adapted from ENCODE Project Consortium and others (2012). It is important to note that additional histone variants and modifications have been characterised however, epigenomic data for example of arginine methylation is currently limited.

Histone modification or variant	**Characteristics of Signal**	**Putative Functions**
H2A.Z	Peak	Histone protein variant (H2A.Z) associated with regulatory elements with dynamic chromatin
H3K4me1	Peak/region	Mark of regulatory elements associated with enhancers and other distal elements, but also enriched downstream of transcription starts
H3K4me2	Peak	Mark of regulatory elements associated with promoters and enhancers
H3K4me3	Peak	Mark of regulatory elements primarily associated with promoters/transcription starts
H3K9ac	Peak	Mark of active regulatory elements with preference for promoters
H3K9me1	Region	Preference for the 5′ end of genes
H3K9me3	Peak/Region	Repressive mark associated with constitutive heterochromatin and repetitive elements
H3K27ac	Peak	Mark of active regulatory elements; may distinguish active enhancers and promoters from their inactive counterparts
H3K27me3	Region	Repressive mark established by polycomb complex activity associated with repressive domains and silent developmental genes
H3K36me3	Region	Elongation mark associated with transcribed portions of genes, with preference for 3′ regions after intron 1
H3K79me2	Region	Transcription-associated mark, with preference for 5′ end of genes
H4K20me1	Region	Preference for 5′ end of genes

Given our understanding of the effects of epigenetics on gene transcription, next, I will describe studies that have led to the discovery of specific roles of epigenetic modulators in normal and malignant haematopoiesis.

1.3 The Epigenetic Landscape of Haematopoietic Cells

The plastic chromatin state of HSCs and early progenitors permits access to genes required for numerous lineages and low level transcription of these genes allows cells to adapt and differentiate accordingly (Hu et al. 1997; Miyamoto et al. 2002). The general theme follows, throughout haematopoietic development and differentiation genes conferring 'stemness' are down-regulated whilst differentiation genes are upregulated.

1.3.1 DNA Methylation

Following embryonic development, methylation of DNA was traditionally thought of as a stable epigenetic modification, however plasticity is retained into adulthood (Meissner et al. 2008). The DNMTs, DNMT3A and DNMT1 are critical for haematopoiesis (Bröske et al. 2009; Challen et al. 2012; Trowbridge et al. 2009). DNMT3A has been shown to be needed for HSC differentiation, however it is non-essential for lineage fate decisions. Transplanting DNMT3A Knock-Out (KO) HSCs into irradiated mice resulted over-proliferation of HSCs, however transplant of KO progenitors did not drastically affect differentiation (Challen et al. 2012). Therefore, DNMT3A must be important for silencing stem cell maintenance genes and de-repressing differentiation genes in the HSC. DNMT1 is required for both differentiation and lineage determination.

Indeed, Bröske et al. (2009) showed that DNMT1 deletion resulted in BM failure. A hypomorphic DNMT1 mutation, that is a mutation causing reduced activity or expression was used to elucidate the role of DNMT1 in haematopoiesis by performing in vivo repopulation assays. Stem cell maintenance genes were downregulated reducing self-renewal and lineage commitment was skewed to myeloid as myeloerythroid TFs were up-regulated and lymphoid genes were down-regulated. DNA methylation maintenance appears to be needed for lymphoid but not myeloid differentiation, the literature illustrates that in myeloid cells methylation is lost whilst lymphoid cells actually gain sites of increased methylation (Bock et al. 2012; Bocker et al. 2011; Hodges et al. 2011; Hogart et al. 2012).

In anticipation, committed progenitors show these methylation differences. Methylation changes occur at transcription factors and specific mature cell activity genes (Bock et al. 2012). Methylation also occurs at TF binding sites and enhancers (Bock et al. 2012; Hodges et al. 2011; Shen et al. 2012). In lymphoid cells, myeloid specific gene promoters and sites where their transcription factors bind are methylated, thus preventing myeloid lineage development. Ageing HSCs show a skew towards the myeloid lineage; this may be because hypomethylation of genes in ageing HSCs is seen at genes that are demethylated in myeloid development (Bocker et al. 2011; Pang et al. 2011). While DNA demethylases TET2 and TET3 are simultaneously expressed in peripheral blood, only TET2 is expressed in the BM (Lorsbach et al. 2003). KO of TET2 results in a shift to development of monocytic cells and HSC expansion and leukaemia occurs in young TET2 KO mice (Butler and Dent 2013).

1.3.2 Chromatin Remodelling

Multiple CRCs have been attributed important roles in normal haematopoiesis and their dysregulation often contributes to malignancies. For example, in murine bone marrow, accumulation of erythroid progenitors alongside a reduction in all other progenitors and vast proliferation of HSCs occurs when the catalytic subunit of the Nucleosome Remodelling and Deacetylase (NuRD) complex is deleted (Yoshida et al. 2008). In mice it was found that bone marrow failure could be induced by depletion of BAF53a of the BRG1- or BRM-Associated Factors (BAF) complex (Krasteva et al. 2012). A myeloid maturation block could be induced in vitro by expression of a dominant negative mutant form of *Brg1*, one of the potential catalytic cores of BAF (Vradii et al. 2006).

1.3.3 Histone Acetylation

Higher H3 and H4 acetylation is seen globally in HSCs and early progenitors than in differentiating cells (Chung et al. 2009). Turnover was also more dynamic suggesting that chromatin is open and more malleable. Whilst not needed for differentiation, the KAT enzyme KAT6A as part of the MOZ (monocytic leukaemia zinc finger protein) complex, is required for HSC and progenitor maintenance (Thomas et al. 2006). In mice, knocking out histone deacetylase 1 and histone deacetylase 2 (*Hdac1* and *Hdac2* respectively) in progenitors prevents formation of erythrocytes and megakaryocytes (Wilting et al. 2010).

1.3.4 Bivalent Promoters

Bivalent promoters are modified simultaneously by repressive H3K27me3 and active H3K4me3 marks, this means that whilst transcriptionally suppressed, they are also poised for transcription. Cell type specific genes and those involved in differentiation, including key transcription factors of haematopoiesis frequently possess bivalent promoters (Abraham et al. 2013; Adli et al. 2010; Bernstein et al. 2006). The repressive H3K27me3 mark is lost at various stages of haematopoiesis and regained after cell fate is determined (Abraham et al. 2013).

Untimely derepression of bivalent B cell genes in HSCs can be induced by depletion or mutation of BMI1, the PcG component of PRC1, leading to aberrantly enhanced B cell formation and HSC reduction (Oguro et al. 2010; Park et al. 2003). As one might expect, conversely, HSC self-renewal increases when BMI1 or the KMT of the PRC2 complex, EZH2 are overexpressed causing aberrantly maintained repression of bivalent promoters of stage specific transcription factors (Kamminga et al. 2006). In mice, overexpression of EZH2 additionally increases myeloid growth (Hidalgo et al. 2012) and mutation affects B cell formation (Mochizuki-Kashio et al. 2011). KDMs are also important in haematopoiesis; expression of lineage restrictive genes is controlled by KDM1A via its relationship with the GFI1 repressor (Saleque et al. 2007).

In HSC, H3K4me2 was found to be enriched at promoters of multiple genes important for B cell development. In B cells H3K4me3 was found in place of H3K4me2, whilst in T cells K4 methylation was removed altogether (Maës et al. 2008). Transcriptionally silent promoters with H3K4me2 present and void of H3K4me3 in HSCs were overrepresented by lineage fate determining genes. When erythroid differentiation was initiated, erythroid genes gained H3K4me3, unrelated genes lost methylation (Orford et al. 2008), suggesting that H3K4me2 may be an alternative mark for poised promoters and characteristic of a recurring theme of bivalency as a measure of differentiation potential in HSCs.

1.3.5 Mixed lineage leukaemia protein

Lysine Methyltransferase 2A (KMT2A), also known as Mixed Lineage Leukaemia protein (MLL), is a H3K4 methyltransferase encoded by *KMT2A* located at human chromosome 11q23 (Benjamin et al. 1995). Clustered Homeobox (Hox) genes are among the best studied MLL downstream targets, critical for HSC renewal (Jude et al. 2007; McMahon et al. 2007). Complete MLL KO experiments revealed failed Hox gene expression during embryonic development and the mice died before birth (Benjamin et al. 1995). Later, biochemical studies showed direct binding and transcriptional regulation of Hox genes by MLL (Milne et al. 2002; Yokoyama et al. 2004). As MLL knock out is lethal to embryonic development, conditional deletion models have been utilised to study the role of MLL in haematopoiesis (McMahon et al. 2007; Jude et al. 2007). Firstly McMahon et al. (2007) showed that MLL KO resulted in a decrease of HSCs in the foetal liver. Contrastingly, in adult haematopoiesis they showed that MLL was not essential, however KO HSCs were found to have impaired reconstitution abilities. Using a different model, (Jude et al. 2007) demonstrated that induced KO of MLL in adult mice results in bone marrow failure and death within just 3 weeks.

Following translation of MLL, the protein is proteolytically cleaved into two units, the amino-terminal (MLL^N) and the carboxy-terminal (MLL^C). These fragments, combined with additionally recruited proteins, form a large complex with histone methyltransferase, histone acetyltransferase and chromatin remodelling functions (see Figure 1.2). MLL^C is responsible for enhancing transcription of targets both directly through H3K4 methylation catalysed by the SET (Suppressor of variegation Enhancer of zeste Trithorax) domain and through its associated proteins. H3K4 methylation is further facilitated by ASH2L and RBBP5 which stabilise the MLL complex in its active confirmation (Southall et al. 2009). WRD5 recognises H3K4 methylation and is theorised to promote consecutive histone modification (Dou et al. 2006). H3K16 acetylation is achieved by MOF in association with MLL^C (Dou et al. 2005). MLL^N ensures that activation of transcription is targeted to the correct genes. The CxxC region binds only to unmethylated DNA and therefore targets transcriptional activation through MLL^C to unmethylated regions. Depending on association of CYP33 with the Plant Homeodomain (PHD), the CxxC region also recruits HDAC and PcG proteins including BMI-1 (Xia et al. 2003). MLL^N also contains AT-hook DNA-binding motifs, which bind preferentially to A-T rich kinks in DNA (Zeleznik-Le et al. 1994). Menin, a tumour suppressor binds to MLL^N allowing the interaction of LEDGF, which in turn binds to DNA through its PWWP domain (Yokoyama and Cleary 2008).

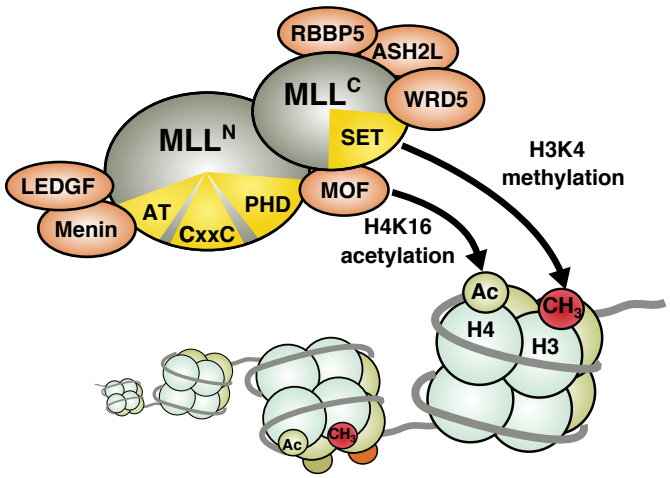

Figure 1.2: The MLL Complex. Adapted from Slany (2009). Following translation, a complex is formed from the proteolysis of the amino-terminal (MLL^N) and carboxy-terminal (MLL^C). The following functional domains are highlighted in yellow: AT=AT hooks, CxxC = motif recognizing unmethylated CpG di-nucleotides. PHD = plant homeodomain, SET = histone methyltransferase active site. Histone acetyltransferase activity requires the cooperation of MOF. Other associated proteins are described in the text.

1.3.6 Transcription Factors

In the bone marrow, myeloid differentiation is initiated by cytokine or growth factor signalling produced mostly by stromal cells (Kleer et al. 2014). In peripheral tissues and blood, cytokines, antigens and other external factors induce signalling cascades in the cell leading to terminal differentiation (Geissmann et al. 2010). Cytokines are required to control lineage-instructive TF expression (Mossadegh-Keller et al. 2013; Rieger et al. 2009). The transcription factors PU.1, CEBPα and GATA1 have been shown to be required for myeloid differentiation (Laslo et al. 2006; Scott et al. 1994). Furthermore, reprogramming of cells or transdifferentiation induced by transcription factors has been widely reported. For example PU.1 and CEBPα can cause fibroblasts to transdifferentiate into macrophages (Feng et al. 2008). Furthermore, CEBPα can transdifferentiate B cells into macrophages, the TF allows B-cells to acquire macrophage-specific gene expression programs (Bussmann et al. 2009).

Transcription factors are of course also responsible for maintaining expression of epigenetic modulators; HDAC7 which represses macrophage-specific genes is down-regulated by CEBPα. Lineage specific TFs also control chromatin modifying enzyme recruitment, priming chromatin for secondary TFs to bind which depend on external stimuli (Heinz et al. 2010; Ostuni et al. 2013). Some of the suggested interactions between epigenetic regulators and TFs in myeloid differentiation are depicted in Figure 1.3A.

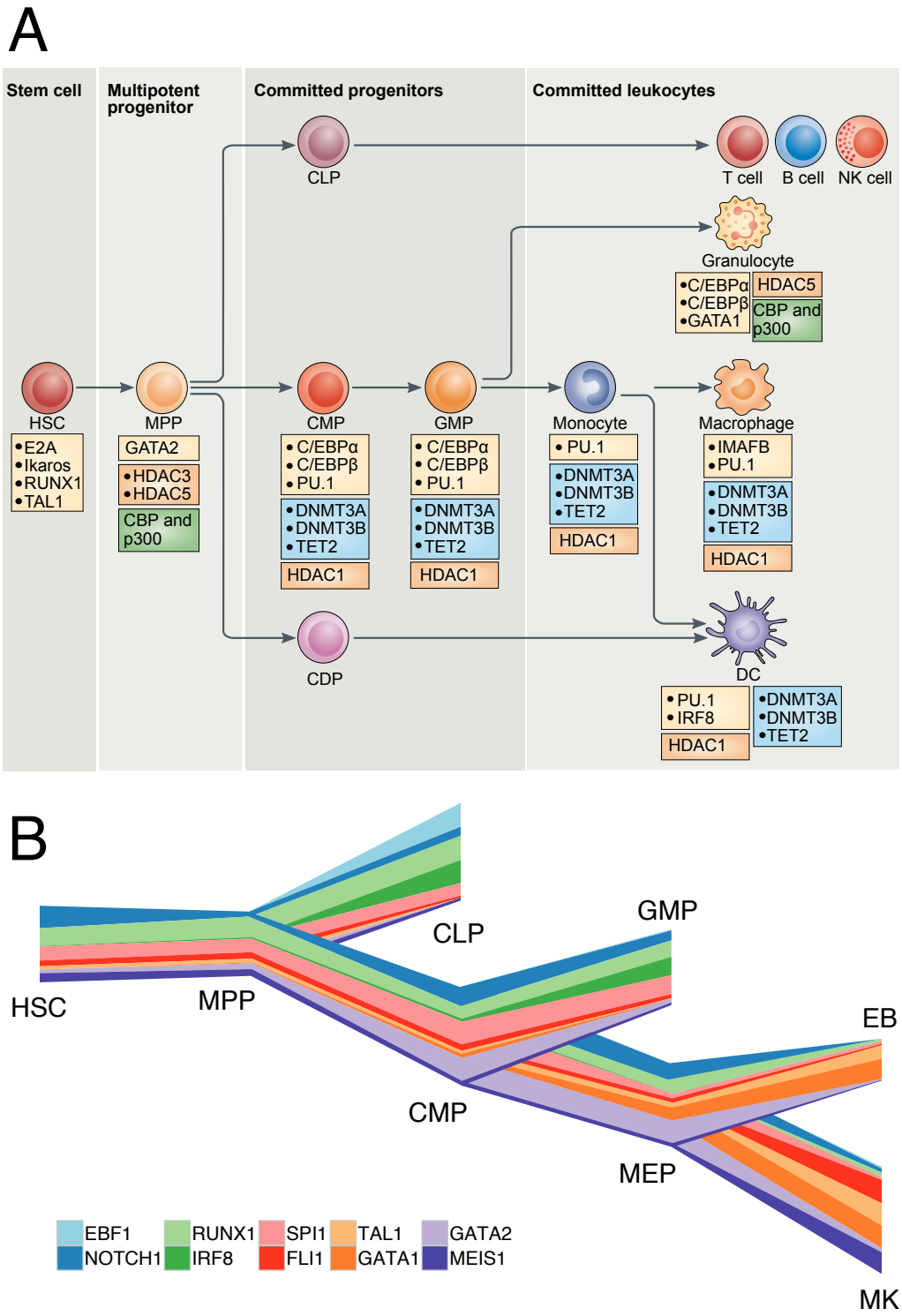

Figure 1.3: Epigenetic control of myeloid differentiation A) Adapted from Álvarez-Errico et al. (2015). Yellow boxes highlight key TFs, blue boxes show DNA methylation enzymes, HDACs in orange Histone Acetyl-Transferases (HATs) in blue. B) Adapted from Chen et al. (2014), a BLUEPRINT project publication. This diagram depicts a riverplot, the thickness of the lines represents the expression level of each of the key transcription factors.

1.3.7 Characterisation of the Epigenome in Haematopoiesis

In recent years numerous genome wide transcriptional and epigenetic sequencing experiments have been conducted in haematopoietic cells to locate epigenetic modifications, TF binding sites TF Binding Site (TFBS) and Long non-coding RNA (LncRNA) binding sites. There are several initiatives which aim to provide next generation sequencing datasets which are easy to obtain and apply to research in haematology and haemato-oncology.

The CODEX project aims to organise publicly available NGS data including histone modification ChIP-Seq, DNase-Seq, which captures open chromatin profiles and RNA Sequencing (RNA-Seq). These data are vetted as to be directly comparable no matter which laboratory they were generated in. CODEX is split into two repositories, ESCODE for embryonic stem cells and HEAMCODE for haematopoiesis (Sánchez-Castillo et al. 2014). The BLUEPRINT project aims to generate more than 100 reference epigenomes, that is ChIP-Seq datasets using defined human samples from both healthy and diseased individuals. At the top of their agenda are blood disorders including autoimmune disease and common leukaemias (BLUEPRINT Consortium 2015). It can be seen from Chromatin Immunoprecipitation (ChIP) sequencing of key histone marks that differentiation is facilitated by activation and inactivation of lineage specific enhancers (Lara-Astiaso et al. 2014).

As it has often been discussed in this section, genetic changes causing epigenetic and transcriptional dysregulation can lead to blocks in differentiation and expansion of progenitors leading to haematopoietic malignancies including leukaemia. Development of drugs targeting epigenetic modifications is a feasible and attractive pursuit as epigenetic regulation can be reversed. Epigenetic marks can also be utilised as prognostic factors or biomarkers. Due to the importance of precise transcriptional and epigenetic control in haematopoiesis, it comes as little surprise that transcription factors and epigenetic modulators are frequently mutated in leukaemia. In the next section, I will describe leukaemia, its subtypes and their causes with an emphasis on epigenetic regulation.

1.4 Leukaemia and its Subtypes

Leukaemia, is a malignant disease indicated by an accelerated expansion of abnormal blood cells or blasts that accumulate in the bone marrow and impede normal blood cell development. In general,

leukaemia can be classified into four primary groups: CML, AML, CLL and ALL. Additionally, other classifications of leukaemia arise from haematological malignancies of mixed linage or ambiguous lineage such as mixed lineage leukaemia (not to be mistaken with the MLL protein) and acute Biphenotypic Leukaemia (ABL). Chronic Myeloid Leukaemia (CML) and acute myeloid leukaemia (AML) are myeloproliferative disorders, meaning that the disease presents with abnormally high numbers of myeloid cells. In CML, abnormal mature cells accumulate in blood and BM. In contrast, in AML the majority of abnormal myeloid growth in peripheral blood and the BM is composed of immature myeloid blast cells. Chronic Lymphocytic Leukaemia (CLL) is an accumulation of lymphocytes. Acute Lymphoblastic Leukaemia (ALL), the most prevalent cancer in children, encompasses B-cell Acute Lymphoblastic Leukaemia/Lymphoma (B-ALL) and T cell acute lymphoblastic leukaemia/lymphoma T-cell Acute Lymphoblastic Leukaemia/Lymphoma (T-ALL). As the most common acute leukaemia in adults is AML, in this section I will focus on AML.

Genetic abnormalities arise during the proliferation and differentiation of haematopoietic stem cells and progenitors such as: mutations, chromosome copy alterations and translocations. These abnormalities may cooperate to alter differentiation, cell death, proliferation and self-renewal resulting in high amounts of abnormal progenitors in the peripheral blood, bone marrow and infiltration of tissues characteristic of AML. Symptoms include anaemia, fatigue, bleeding, and susceptibility to infection (Lowenberg et al. 1999). AML is a heterogeneous disease, therefore it is critical for patients to be further classified prior to treatment (Grimwade et al. 2010).

Originally the French-American-British (FAB) system defined leukaemia as cases of >30% blasts in the bone marrow, AML was then categorised using cell morphology and cytochemistry (Bennett et al. 1976). More recently, that has been modified by the World Health Organisation (WHO) (Vardiman et al. 2002). Now a minimum of 20% blasts in BM are sufficient for AML diagnosis and stratification methods additionally utilise cytogenetics and gene mutations (Vardiman et al. 2009; Vardiman et al. 2002). AML can also occur as a secondary malignancy in patients with unrelated neoplasms. Alkylating agent treatment of other cancers such as breast cancer, can stimulate AML (Leeuwen 1996). Moreover, AML can evolve from other diseases of the blood such as Myelo-Dysplastic Syndrome (MDS), it was also suggested that predisposition to AML can be promoted by germline polymorphisms involving the CEBPα gene (Smith et al. 2004).

Table 1.2: FAB classification of AML

FAB Subtype	Name	Adult AML patients (%)	Prognosis compared to average AML
M0	Undifferentiated acute myeloblastic leukaemia	5%	Worse
M1	Acute myeloblastic leukaemia with minimal maturation	15%	Average
M2	Acute myeloblastic leukaemia with maturation	25%	Better
M3	Acute promyelocytic leukaemia	10%	Best
M4	Acute myelomonocytic leukaemia	20%	Average
M4 eos	Acute myelomonocytic leukaemia with eosinophilia	5%	Better
M5	Acute monocytic leukaemia	10%	Average
M6	Acute erythroid leukaemia	5%	Worse
M7	Acute megakaryoblastic leukaemia	5%	Worse

Table 1.3: WHO classification of AML

Acute myeloid leukaemia with recurrent genetic abnormalities

- Acute myeloid leukaemia with t(8;21)(q22;q22), (AML1/ETO)
- Acute myeloid leukaemia with abnormal bone marrow eosinophils and inv(16) (p13q22) or
- Acute promyelocytic leukaemia with t(15;17)(q22;q12), (PML/RARα) and variants
- Acute myeloid leukaemia with 11q23 MLL abnormalities

Acute myeloid leukaemia with multilineage dysplasia

- Following MDS or MDS/MPD
- Without antecedent MDS or MDS/MPD, but with dysplasia in at least 50% of cells in 2 or more myeloid lineages

Acute myeloid leukaemia and myelodysplastic syndromes, therapy related

- Alkylating agent/radiation-related type
- Topoisomerase II inhibitor-related type (some may be lymphoid)

Acute myeloid leukaemia, not otherwise categorized classify as:

- Acute myeloid leukaemia, minimally differentiated (M1)
- Acute myeloid leukaemia without maturation (M0)
- Acute myeloid leukaemia with maturation (M2)
- Acute myelomonocytic leukaemia (M4)
- Acute monoblastic/acute monocytic leukaemia (M5)
- Acute erythroid leukaemia (M6)
- Acute megakaryoblastic leukaemia (M7)
- Acute basophilic leukaemia
- Acute panmyelosis with myelofibrosis
- Myeloid sarcoma

Whole genome sequencing studies revealed that AML has a much lower mutational background compared to solid tumours (Cancer Genome Atlas Research Network et al. 2013). The Cancer Genome Atlas analysed 200 AMLs consisting of each subtype represented in real-world proportions. On average only 13 mutations were found in AML cells compared to healthy somatic cells, this average dropped in those exhibiting fusion proteins. In contrast, solid tumours average between 30 nonsynonymous mutations in breast cancers and 200 in lung cancer and melanoma (Vogelstein et al. 2013).

Mutations in 23 genes were found to reoccur with statistical significance amongst the 200 AMLs, on average just 5 of these were present in each patient (Cancer Genome Atlas Research Network et al. 2013). AML mutations can be grouped into: myeloid TFs, TF fusions, tumour suppressors, signalling proteins, epigenetic modifiers, further highlighting the impact of transcriptional and epigenetic dysregulation in disease development. Of note, cohesin and spliceosome mutations are also recurrently identified in AML (Cancer Genome Atlas Research Network et al. 2013).

Although these mutations are common in AML the prognostic impact in particular in the case of the mutations in epigenetic modifiers is currently unknown. Of the 200, it was found that 44% had mutations in genes involved in DNA methylation and 30% had mutations in chromatin modification (Cancer Genome Atlas Research Network et al. 2013). In cancer, tumour suppressors are often silenced via hypermethylation of local CpG islands (Sandoval and Esteller 2012). Mutations in DNMT3A occur frequently in AML and confer poor prognosis (Ley et al. 2010; Walter et al. 2011; Yan et al. 2011). Mutations in the gene TET2 are frequent in myeloid leukaemia (Chan and Majeti 2013). Leukaemia, often myeloproliferative, occurs spontaneously in TET2 KO mice (Butler and Dent 2013). Common to AML, IDH1 and IDH2 mutations are mutually exclusive with TET2 mutations. The normal product of IDH1/2 is α-Ketoglutarate (α-KG), mutated IDH1/2 produces 2-hydroxyglutarate instead. TET2 function is dependant on α-KG therefore IDH mutations are functionally equivalent to nonsense mutations in TET2, thus presenting a similar disease (Chan and Majeti 2013).

As mentioned earlier in subsection 1.3.1, normal myeloid differentiation requires DNA demethylation, therefore mutations resulting in hypermethylation could lead to a block in differentiation. On the other hand DNMT3A nonsense mutations also cause methylation, therefore locus specific methylation is important. AML frequently harbours loss of function mutations in the components comprising PRC2 such as EZH2. In striking contradiction, gain of function mutations in EZH2 are reported in ALL (Butler and Dent 2013). Protein arginine

N-methyltransferases (PRMTs) mono- or di-methylate arginine residues including on histones and TFs, some of which are overexpressed in leukaemia (Greenblatt and Nimer 2014).

Myeloid differentiation was preferentially induced with *PRMT4* knockdown in human CD34+ cells and the Kasumi-1 leukaemia cell line suggesting that PRMT4 blocks myeloid differentiation (Vu et al. 2013). Genetic lesions affecting KDMs of the Jumanji C family are also involved in the initiation of leukaemia, for example; translocations leading to fusions between *NUP98* and *JARID1A* (KDM5A) exist in 10% of acute megakaryoblastic leukaemia (Rooij et al. 2013) and inactivating mutations of UTX are frequently present in AML and ALL (Mar et al. 2012).

The most informative standalone prognostic factor is considered to be cytogenetic analysis (Mrózek et al. 1997). Interestingly, in more than one third of AML patients chromosomal rearrangements can be identified and the majority involves Leukaemia Transcription Factors (LATFs). Therefore understanding how aberrant LATF gene fusions function is imperative to developing novel therapies treating this leukaemia. Among the most unfavourable chromosomal genetic lesions are MLL (11q23) rearrangements (Slovak et al. 2000) and in the next section I will describe common LATFs rearrangements with a focus on MLL-fusions.

1.5 Gene Fusions

Recurrent chromosomal translocations, more frequently in haematological cancer than in other cancers, give rise to chimaeric fusion proteins. LATF fusions initiate the transformation of HSCs and myeloid progenitors into AML stem cells. Remarkably, for those diagnosed with AML, 25% of adults and 35% of children harbour 1 of 4 of the most recurring chromosomal translocations: t(8;21) AML1-ETO, inv(16) CBFβ-MYH11, t(15:17) PML-RARα and 11q23 MLL-fusions (Doulatov et al. 2012).

Interestingly one of these translocations (11q23) directly involves the MLL gene with epigenetic modifying functions, and all recurrent fusion proteins invariably recruit epigenetic modifying proteins. Moreover, some of the less recurrent chromosomal translocations often involve genes encoding epigenetic modifying functions such as translocations between the KATs MOZ or MORF and the KATs CBP (Panagopoulos et al. 2001; Rozman et al. 2004) or KMT NSD1 that fuses to NUP98, a nuclear pore protein (Rosati et al. 2002). Acute leukaemia harbouring mixed lineage leukaemia MLL gene translocations has been described as a “prototypical epigenetically

driven cancer" (Neff and Armstrong 2013). Of all haematological cancers, AML with 11q23 translocations is one of the poorest prognostic subtypes, in infants it is particularly severe with 3 year overall survival of 5% (Taki et al. 1996). MLL translocations occur in 5-10% of AML in adults and children and 70% of childhood ALL (Neff and Armstrong 2013).

Intriguingly, more than 70 different MLL translocation partners have been identified in AML patients (Broad Institute 2017; Slany 2009), however 85% of 11q23 translocation cases involve only six of these, see Table 1.4. In an MLL fusion, the C-terminus of MLL including the SET domain is invariably replaced by the fusion partner which is fused in frame immediately after the CxxC domain. At least 4 mechanisms have been proposed for how MLL fusions deregulate transcription in an MLL fusion partner dependant manner (Figure 1.4). These include: histone H3K79 methylation by DOT1L recruitment, histone acetylation by CBP/p300, histone H4R3 methylation by PRMT1 recruitment and dimerisation of the MLL fusion transcriptional complex. All MLL fusions activate the transcription of key downstream targets such as Hox genes (Zeisig et al. 2004) as well as transcriptional programmes such as leukaemic stem cell maintenance (Somervaille et al. 2009). As a result, these epigenetic and transcriptomic changes initiated by MLL fusion proteins lead to leukaemic transformation.

Importantly, other molecules which are not direct targets of the MLL fusion transcriptional complex such as CEBPα and β-catenin have been found to be critical in MLL fusion transformation (Ohlsson et al. 2014; Wang et al. 2010a; Yeung et al. 2010), although the mechanisms of their contribution is currently unknown. While a recent report revealed that CEBPα dependent myeloid differentiation may be a prerequisite to MLL fusion mediated transformation (Ye et al. 2015), no mechanistic insights into the role of β-catenin in MLL transformation have been reported, highlighting the need for further dissection to uncover these molecular mechanisms.

Table 1.4: The 6 most common MLL-fusion partners Adapted from Slany (2009).

Name	Gene Alias	Features
ENL	MLLT1	Binds histone H3, assembles EAP elongation/chromatin modification complex
AF9	MLLT3	ENL homolog, also found in EAP
AF4	AFF1/MLLT2	Founder of AF4 family, member of EAP
ELL		Elongation factor, interacts with a protein related to AF4
AF10	MLLT10	Interacts with DOT1L histone methyltransferase
AF6	MLLT4	Dimerisation domain

Leukaemia appears to be organised in a hierarchical structure similar to normal haematopoiesis, with Leukaemia Stem Cells (LSCs) residing at the apex (Bonnet and Dick 1997; Bonnet 2005), and like normal HSCs, possess ability to self-renew and give rise to differentiated progeny. LSCs are thought to be the source of patient relapse after aggressive treatments and developing drug resistance (Chan and Huntly 2008; Estrov 2009; Konopleva and Jordan 2011). Therefore it is imperative to understand the properties of LSCs and potentially their origin to facilitate direct treatments to them. In the next section I will describe the clinical relevance of LSCs and studies that have been used to investigate their origins.

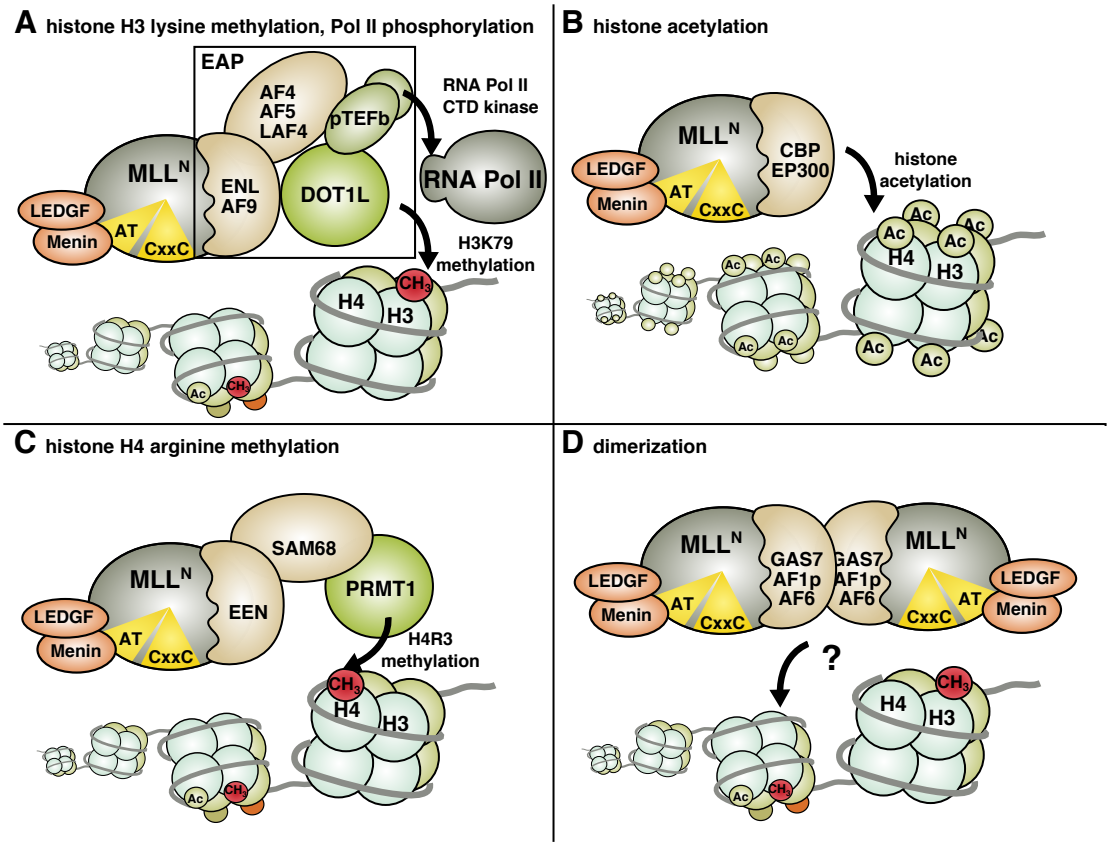

Figure 1.4: Mechanisms of transcriptional regulation by MLL-fusions Adapted from Slany (2009). A) Partners ENL and AF4 form the EAP complex. DOT1 provides H3K79 methylation whilst pTEFb phosphorylates the C-terminus of RNApol II stimulating transcriptional elongation. B) CBP and P300 can fuse to MLL and acetylate histones. C) The adaptor SAM68 binds EEN in MLL-EEN and recruits PRMT1, the H4R3 methyltransferase. D) Aberrant transcription is activated by unknown means when fusion dimerise via dimerisation domains of the fusion partner.

1.6 The Origin of Leukaemia Stem Cells

Most of what we know about Cancer Stem Cells (CSCs) is derived from experiments involving both normal blood and leukaemia, this research was accelerated with the discovery of lineage surface markers (Bonnet 2005). The first identification of CSCs was reported in 1997 when sorted CD34+/CD38- but not CD34+/CD38+ AML cells induced leukaemia in recipient immunodeficient mice possessing the $Prkdc^{scid}$ mutation, (known as NOD SCID or SCID mice); this made AML the first malignancy to demonstrate proof of the cancer stem cell concept (Bonnet and Dick 1997).

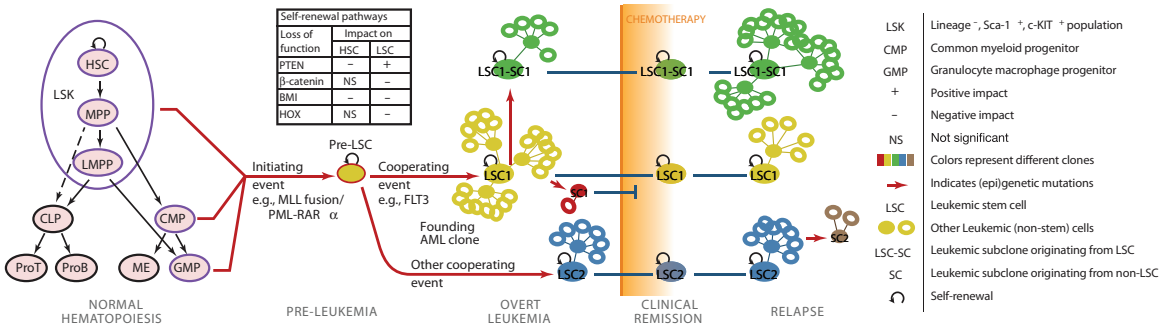

Figure 1.5: Cell-of-origin and clonal evolution model of LSCs Adapted from Zeisig et al. (2012). A stem cell or progenitor cell takes a genetic hit which transforms it into a Pre-Leukaemic Stem Cell (Pre-LSC). A cooperating genetic event, likely a signalling protein, establishes the LSC, this is the "two-hit hypothesis". LSCs progeny can then go on to produce subclones with advantageous mutations. After chemotherapy treatment, the bulk of the tumour, LSC progeny die, however LSCs are allowed to proliferate. This means that after treatment, a different clone may dominate.

CD34+CD38- AML cells are tolerant to drugs and provoke little immune response in vitro (Costello et al. 2000) and also show high drug resistance levels when transplanted into mice (Saito et al. 2010). Although CD34+CD38- frequency is a valuable prognostic factor at diagnosis (Terwijn et al. 2014), it is important to note that also other fractions such as CD34+/CD38+ have recently been shown to contain LSC activity (Majeti and Weissman 2011). Moreover, about 20% of AML constitutes of only CD34- cells (Taussig et al. 2010), thus the LSC activity resides within CD34- AML cells in these patients. Feuring-Buske and Hogge (2001) identified a population of cells enriched in both stem cells and CSCs "Side Population cell", using flow cytometry to measure expulsion of fluorescent Hoechst33342. Additional markers such as ALDH activity as well as gene expression programs have been reported to identify HSCs and LSCs (Eppert et al. 2011; Feller et al. 2013).

Like HSCs, leukaemia stem cells LSCs have self-renewal capacity, their progeny are short lived and undergo some level of differentiation. LSC progeny can go on to become subclone neoplasms with other mutations resulting in a number of different clones in different proportions making up the leukaemia Figure 1.5. Although LSCs share many properties with HSCs, the origin of LSCs remains ambiguous. The transformative property of gene fusions provides an opportunity to investigate the potential cell-of-origin in leukaemia and give insight into their characteristics.

1.6.1 Retroviral Cell-of-Origin Models

Retroviral/lentiviral transduction and transformation assays Retroviral Transduction and Transformation Assay (RTTA) (Zeisig and So 2009) revealed that oncogenic fusion proteins, such as MLL fusions can transform haematopoietic stem/progenitor cells into immortalized so called Pre-Leukaemic Stem Cells with unlimited self-renewal in vitro. Due to a significant increase in latency between Pre-LSCs and secondary transplants, it is believed that Pre-LSCs acquire an additional cooperating event to produce LSCs that can initiate leukaemia in mice (Figure 1.5), such as FMS-like tyrosine kinase 3 (FLT3), internal tandem repeat mutations (FLT3-ITD). For example, while MLL fusions were shown to transform murine HSC, CMP and GMP populations to induce AML (Cozzio et al. 2003; Krivtsov et al. 2013; So et al. 2003), BCR-ABL was only able to transform HSC but not GMPs (Huntly et al. 2004). Similarly, MOZ-TIF2 transformed HSC and GMP, but MN1 only transformed CMP (Heuser et al. 2011; Huntly et al. 2004), suggesting that multiple haematopoietic populations are susceptible to leukaemic transformation in an oncogene/fusion gene dependent manner.

Whilst retroviral transduction models have been instrumental in the study of MLL fusion leukaemia (Cozzio et al. 2003; So et al. 2003; Huntly et al. 2004; Krivtsov et al. 2013; Heuser et al. 2011), the limitations of these models must be considered when interpreting these data. Firstly, as the retrovirus promoter is much stronger than the endogenous MLL promoter and the number of inserted copies is generally poorly controlled, retroviral transduction leads to expression of the fusion which may be much greater than physiological levels. Moreover, retroviral integration has the potential of introducing unintentional mutations by integrating into gene bodies or aberrant expression of genes when inserting into enhancers or promoters.

Curiously, genome wide studies have revealed that retroviruses preferentially insert into actively transcribed regions as they are more available for insertion, in particular this has been observed for the Murine Stem Cell Virus (MSCV) in haematopoietic cells (Aker et al. 2006). Indeed, exploiting this nature of retroviruses and more recently, transposons, lead to the development of insertional mutagenesis models for oncogene discovery (Uren et al. 2005; Rad et al. 2010). This is also a concern for gene therapy strategies utilising retroviruses (Baum et al. 2003). Another caveat of the transduction process is that it allows selection for exceptional cells in the population such as: contaminating cells from FACS impurity or particular mutants. Finally, gamma-retrovirus integration is dependent on active proliferation, consequently quiescent populations such as HSCs can't be targeted without modification.

1.6.2 Knock-in Mouse Cell-of-Origin Models

Mouse models expressing MLL fusions have also been used to investigate LSC cell-of-origin. In contrast to retroviral models, using a knock-in MLL-AF9 model Chen et al. (2008) showed that under endogenous expression, GMPs were unable to produce leukaemia. On further inspection, it was found that the expression of MLL-AF9 and its targets was higher in HSCs than GMPs suggesting that dosage is critical for activation of Hox genes and subsequent leukaemic transformation. Dox-inducible MLL-AF9 (Stavropoulou et al. 2016) and MLL-ENL (Stavropoulou et al. 2018; Ugale et al. 2014) mouse models have also been used to investigate cell-of-origin with varying results. Stavropoulou et al. (2016), showed that whilst expressing much lower levels of the fusion, similarly to what is seen using retroviral models (Krivtsov et al. 2013), Dox-inducible MLL-AF9 transformed HSCs produced a much more aggressive leukaemia than GMPs. On the other hand, Ugale et al. (2014) showed Dox-inducible MLL-ENL transformation was possible in a subpopulation of CMPs, known as precursor GMPs (pGMs) but not fully defined GMPs and surprisingly, not HSCs. Similarly, the inducible Dox-inducible MLL-ENL model developed by Stavropoulou et al. (2018) also showed that GMPs could not be transformed by MLL-ENL. However, in contrast, the mouse model from Stavropoulou et al. (2018) resulted in leukaemic transformation of HSCs.

To create a more physiological model Takacova et al. (2012) inserted MLL-ENL-ERtm at the endogenous promoter. Intriguingly, activation of MLL-ENL with 4-OHT induced a state of chronic stress activating DNA damage and inflammation responses forcing cells into a somewhat quiescent state which could be reversed caffeine treatment. Earlier work using an

in vivo MLL-ENL translocator mouse model (Forster et al. 2003), showed that MLL-ENL did not produce a transplantable leukaemia. These studies show MLL-ENL associated leukaemia as much less aggressive than retroviral models. As described in Figure 1.5, it is likely that MLL-ENL transformed cells go on to acquire cooperating mutations which support leukaemogenesis and disease maintenance. AML samples with MLL-fusions often harbour RAS mutations (Grossmann et al. 2013). Retroviral models were used to show that RAS mutations cooperate with MLL fusions to produce a more aggressive leukaemia (Zuber et al. 2009; Ng et al. 2014).

It remains unclear if cancers arising from different cells of origin inherit epigenetic features of their cell-of-origin type allowing them to switch on different mechanisms of self renewal, prevent differentiation and apoptosis. Critical self-renewal pathways in both HSCs and LSCs include Hedgehog, Hox, polycomb/Mll, PTEN, telomerase, GSK3 and Wnt/β-catenin (Copland et al. 2009). Indeed, MLL fusion expression increases Homeobox A9 *Hoxa9*, Myeloid Ecotropic Viral Integration Site 1 *Meis1* and β-catenin expression levels (Zeisig et al. 2004; Balgobind et al. 2011), suggesting that these pathways may be utilized for MLL fusion mediated transformation. Recent work from our lab and others show that β-catenin has an important role development of MLL fusion LSCs (Yeung et al. 2010; Wang et al. 2010a; Gandillet et al. 2011; Siapati et al. 2011; Griffiths et al. 2010). Interestingly, while β-catenin is essential for other adult stem cells it is dispensable for HSCs maintenance, suggesting that the role of β-catenin in HSCs is different from other tissue stem cells (Cobas et al. 2004). The fact that β-catenin is not essential in adult HSCs also makes it is an attractive target for therapy. Considering MLL as an example of epigenetic disease and β-catenin as regulatory hub protein provides an opportunity to explore cascade of molecular events leading to LSC formation. β-catenin dependent transcription has been shown to be responsible for aberrant self-renewal in other cancers. First, we must understand how β-catenin comes to regulate gene expression.

1.7 Wnt Signalling in Development and Disease

In humans, β-catenin is encoded by the gene Catenin Beta 1 (*Ctnnb1*) at chromosome 3 p21. β-catenin has a dual function as a coordinator of cell adhesion and as the terminal effector of Wnt signalling, activating gene expression (Barker and Clevers 2000; Brault et al. 2001). Wnt signalling is essential in embryonic development, responsible for directing cell fate in tissue morphogenesis, adult stem cells continue to utilise Wnt signalling for directing growth and differentiation to maintain tissue homoeostasis and respond to damage (Holland et al. 2013).

There are two categories of the Wnt pathway: canonical and non-canonical. In the canonical Wnt pathway, Wnt signalling induces transcriptional activation dependent on β-catenin. Therefore canonical Wnt signalling is referred to as Wnt/β-catenin signalling. Non-canonical Wnt pathways act independently of β-catenin, these are split into two branches, categorised by utilising either calcium (Wnt/Ca2+) or the c-Jun N-terminal kinase (Wnt/JNK) to initiate signalling cascades (Lien and Fuchs 2014).

As Wnt/β-catenin signalling is so important for morphogenesis and control of cell proliferation, inappropriate activation causes severe hereditary syndromes and malignancies (Clevers 2006; Lien and Fuchs 2014). β-catenin deficient, mice die in utero due to failed mesoderm and axis formation (Haegel et al. 1995; Huelsken et al. 2000). Conditional inactivation of β-catenin further into development in mice produces lethal brain and craniofacial malformations (Brault et al. 2001). Indeed there is substantial evidence regarding β-catenin's roles in normal and malignant haematopoiesis.

The basic pathway is described in Figure 1.6A. A large number of proteins which interact with β-catenin directly or through multi-subunit complexes have been confirmed in multiple animal model species (Brault et al. 2001), suggesting that Wnt targets are under complex regulatory mechanisms. Co-activators of β-catenin often influence gene expression through perturbation of chromatin conformation, some of these co-activators are illustrated in Figure 1.6B. (Barker et al. 2001) showed that the chromatin remodelling enzymes of type SWI/ SNF (SWItch/Sucrose Non-Fermentable) can be recruited by β-catenin, additionally β-catenin mediated recruitment of the protein ISWI (imitation SWI) was shown by Song et al. (2009). Furthermore, acetyl-transferases have shown to be recruited by β-catenin both acetylating β-catenin itself and modifying surrounding histones, such as p300 (Lévy et al. 2004), CBP (Wolf et al. 2002), and Tip60 (Kim et al. 2005).

Additional co-factors link enhancers via TFs to RNA pol II, through components of the Mediator complex and the Paf1 complex (Carrera et al. 2008; Kim et al. 2006; Mosimann et al. 2009; Parker et al. 2008). Intriguingly, it has been reported that β-catenin can recruit the MLL1 complex to DNA (Sierra et al. 2006; Wend et al. 2013). Sierra et al. (2006) conducted time-course ChIP quantitative PCR (ChIP-qPCR) at Wnt target genes to investigate β-catenin mediated regulation of H3K4 trimethylation. Upon β-catenin stimulation, β-catenin was found to be co-bound to the Myc enhancer with MLL2 along with SET1 subunits Ash2 and Menin (Sierra et al. 2006).

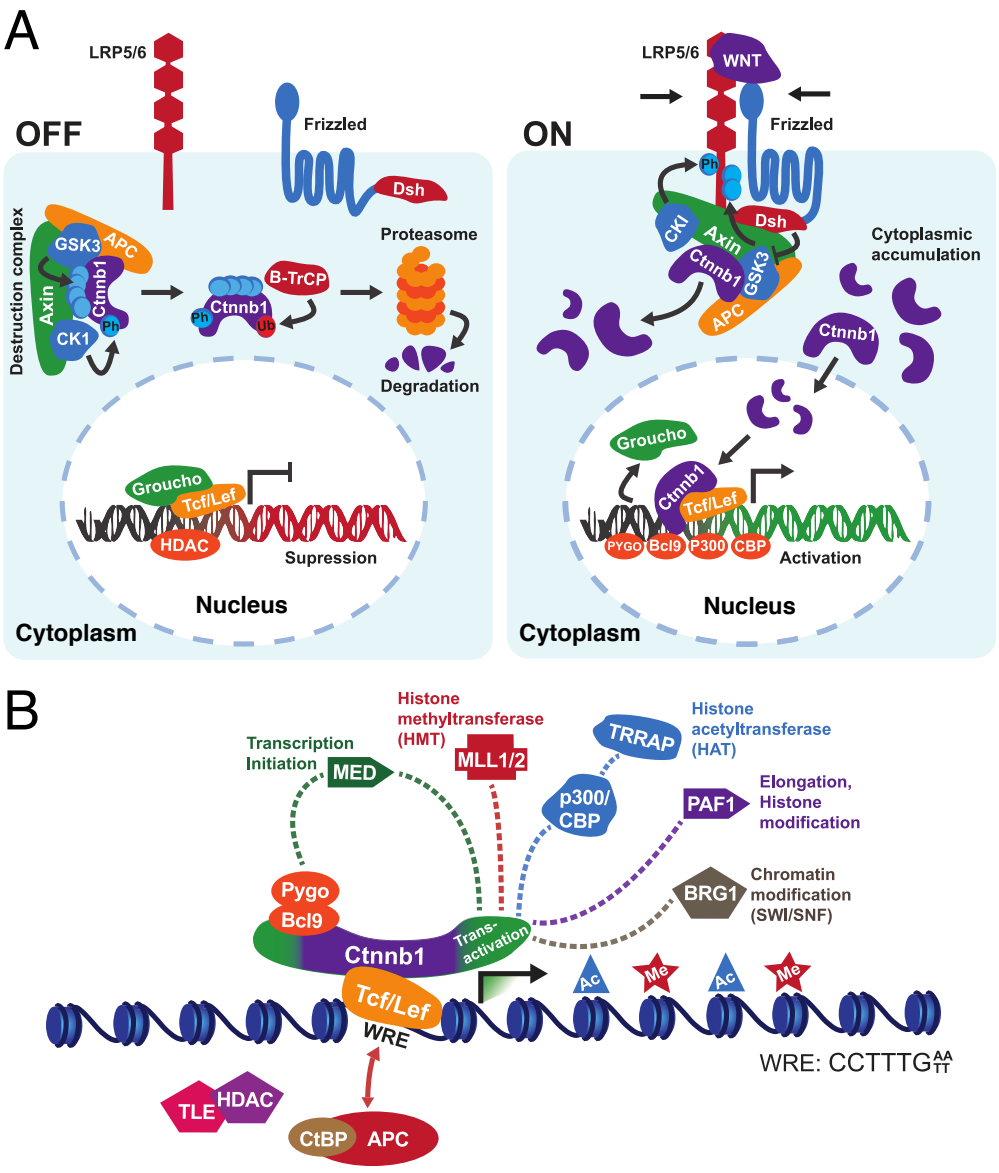

Figure 1.6: Canonical Wnt signalling A) OFF: Groucho is bound to Wnt responsive elements Wnt Responsive Elements (WREs) and recruits HDACs, repressing gene expression. Destruction of the continuously transcribed effector protein β-catenin is carried out by the Axin complex containing Axin, APC, GSK-3 and CKI. β-catenin is phosphorylated in the cytoplasm firstly via GSK-3 followed by CKI, marking it for ubiquitination by b-Trcp and subsequent degradation by the proteasome. APC and Axin act as tumour suppressors suppressing β-catenin, mutations in these genes occur frequently in Familial Adenomatous Polyposis (FAP) and sporadic colorectal cancers (Nishisho et al. 1991; Kinzler and Vogelstein 1996). In the nucleus, WREs are bound by T cell factor/lymphoid enhancer factor TCF/LEF and recruit co-repressor Groucho which recruits HDACs, repressing gene expression (Cavallo et al. 1998; Lien and Fuchs 2014). ON: At the cell surface, Wnt ligand binds to the frizzled receptor (Fz) and its co-receptor lipoprotein receptor related protein (LRP5/6). Subsequently Dishevelled (Dsh) disrupts the Axin complex, β-catenin accumulates in the cytoplasm leading to nuclear translocation. β-catenin competitively displaces Groucho and binds to TCF/LEF TFs via its armadillo domains to activate gene expression (Daniels and Weis 2005; Behrens et al. 1996; Molenaar et al. 1996). β-catenin recruits various cofactors which remodel chromatin for transcription. The lower portion of this figure shows known co-factors of β-catenin which interact with the N-terminal or C-terminal transactivation domains reviewed in Mosimann et al. (2009).

While most studies using a mammalian system utilised epithelial or embryonic stem cells (Mosimann et al. 2009), it is currently unclear if the described Wnt/β-catenin mechanisms are also active in haematopoietic cells. Although β-catenin is not directly bound to chromatin, recently gene expression regulation via β-catenin has been investigated genome wide using ChIP-Seq on several occasions, mostly in human colorectal cancer cells (Bottomly et al. 2010; Schuijers et al. 2014; Watanabe et al. 2014) and more recently in Embryonic Stem Cells (ESCs) (Estarás et al. 2015; Funa et al. 2015). However, quality datasets have not been achieved in healthy and malignant haematopoietic cells in human or mouse. It has been proposed that numerous transcription factors tether β-catenin to DNA besides TCF/LEF (see Table 1.5). However, Schuijers et al. (2014) suggested that β-catenin binding is exclusively determined by TCF/LEF as both β-catenin bound regions with low and high levels of TCF4 binding were diminished when a dominant negative form of TCF4 (DNTCF4) was expressed.

Table 1.5: TFs proposed to recruit β-catenin Adapted from Schuijers et al. (2014).

Protein	Citation(s)
cJun	Nateri et al. (2005)
ERα	Kouzmenko et al. (2004)
Foxo1/4/3a	Essers et al. (2005)
Hif1α	Kaidi et al. (2007)
LRH-1 NR5A2	Botrugno et al. (2004)
Oct4	Kelly et al. (2011), Zhang et al. (2013)
Sox17	Sinner et al. (2004)

Indeed there is substantial evidence regarding β-catenin's roles in normal and malignant haematopoiesis. Wnt signalling was first found to have a role in haematopoiesis when Reya et al. (2000) found proliferation of B lymphocyte progenitors was Wnt dependent by knocking out the Wnt responsive TF Lymphoid Enhancer Binding Factor (*Lef1*) in mice. Thereafter it was reported that expression of constitutively active β-catenin in HSCs lead to maintenance of an immature state (Reya et al. 2003). Moreover Scheller et al. (2006) discovered that expression of constitutively active β-catenin in HSCs leads to death due to bone marrow failure as HSCs were unable to generate differentiated blood cells.

As one may expect, β-catenin has been implicated in AML (Simon et al. 2005). Although activation of β-catenin impairs normal haematopoiesis, β-catenin loss on the other hand, has

been reported to have little effect on HSCs and haematopoiesis (Cobas et al. 2004). β-catenin deletion in adult HSC does not affect self-renewal nor haematopoiesis as HSCs could give rise to all normal haematopoietic cells (Cobas et al. 2004). Jeannet et al. (2008) further showed that this is also true in long term studies. Although Zhao et al. (2007) showed that HSC maintenance decreased upon serial transplantation of β-catenin deleted HSCs, these HSCs remained fully functional. Moreover, β-catenin deletion in mature T cells revealed that memory T cell functions were not affected and showed normal proliferation responses to lymphopenia or infection (Prlic and Bevan 2011).

Because normal adult haematopoiesis does not require β-catenin, this makes it an attractive target for therapy. To exploit its therapeutic potential, we have to understand Wnt/β-catenin mechanisms in both normal as well as malignant haematopoiesis. Whilst dispensable for HSCs, β-catenin is a critical element of MLL LSC maturation. In the following section I will describe what is known about the role of β-catenin in LSCs, this work provides the foundation and rationale behind this thesis.

1.7.1 The Role of β-catenin in Leukaemia Stem Cells

Our group and others have revealed β-catenin as a critical component in LSC maintenance and drug resistance (Gandillet et al. 2011; Griffiths et al. 2010; Siapati et al. 2011; Wang et al. 2010a; Yeung et al. 2010; Rathert et al. 2015; Fong et al. 2015). As previously mentioned, aberrant β-catenin pathway regulation has been studied primarily in epithelial cancers often harbouring mutations in either β-catenin or other proteins involved in the destruction complex leading to increased β-catenin protein. On the other hand, mutations in these genes are very rare in leukaemia, therefore β-catenin must be activated via alternative mechanisms. Possible genetic mechanisms include FLT3-ITD mutations through tyrosine phosphorylation of β-catenin promoting nuclear translocation (Kajiguchi et al. 2007). In the absence of cooperating mutations, β-catenin activation can be promoted through extracellular signalling the bone marrow niche (Lento et al. 2013). Cell-of-origin also plays a part in the activation of β-catenin, HSCs have activated β-catenin compared with GMPs (Lento et al. 2013). It was shown that MLL-AF9 can activate β-catenin in GMPs however HOXA9 MEIS1 transformed GMPs do not produce leukaemia unless constitutively active β-catenin is also expressed (Wang et al. 2010a). Additionally it was demonstrated that β-catenin may be specifically activated in MLL-AF9 cells though G protein-coupled receptor 84 (GPR84) and Gq

protein alpha (Gaq) signalling (Lynch et al. 2016; Dietrich et al. 2014; Yi et al. 2013). Leucine-rich repeat-containing G protein-coupled receptor 4 (*Lgr4*) is highly overexpressed in human MLL-fusion associated AML (Bagger et al. 2013). As R-spondin (*Rspo*)/*Lgr4* signalling has been demonstrated to be critically involved in Wnt/β-catenin signalling, Yi et al. (2013) tested Rspo and Wnt proteins for their β-catenin activation capabilities in MLL-AF9 LSCs. A combination of Rspo3 and Wnt3a were found to significantly increase β-catenin activation (Yi et al. 2013).

Table 1.6: Wnt pathway components expressed in the bone marrow environment Taken from Lento et al. (2013).

Cell source	**Wnt ligands**	**FZD/ LRP**	**sFRP/ Dkk/ WIF**	**Transcription Factors**	**References**
Osteoblasts	Wnt2, Wnt2B2, Wnt3A, Wnt4, Wnt5AQ, Wnt10A, Wnt11	FZD2, FZD3, FZD4, FZD5, FZD6	sFRP1, sFRP2, sFRP3, sFRP4	TCF1, TCF2, TCF3, TCF4, TLE1, CTBP1, β-catenin	Spencer et al. (2006)
Mesenchymal stem Cell	Wnt2B, Wnt5A, Wnt2B, Wnt11	FZD1, FZD2, FZD3, FZD4, FZD5, FZD6	sFRP2, sFRP2, Dkk2	β-catenin, TCF1	Dufourcq et al. (2008), Li et al. (2010) and Qiu et al. (2011)
Endothelial cells	Wnt2, Wnt2b, Wnt3, Wnt5A, Wnt7A, Wnt11, Wnt14	FZD3, FZD4, FZD6, LRP5, LRP6	sFRP1, sFRP3, Dkk1, Dkk2, Dkk3	β-catenin, TCF1, Lef1	Li et al. (2010) and Planutiene et al. (2011)

Increasing evidence has shown that reliance of LSCs on β-catenin is not universal, but is dependent on disease phenotype, oncogenic mutations and cell-of-origin. Response to β-catenin down-regulation was shown to be variable amongst primary patient samples (Gandillet et al. 2011). The BCR-ABL fusion, product of the t(9;21) translocation, produces CML and Philadelphia Chromosome positive acute lymphoblastic leukaemia (Ph+ ALL). While the chronic phase of CML is quite stable and disease development is slow, CML frequently transforms into an acute leukaemia during blast crisis. It was shown that β-catenin deletion in BCR-ABL transduced cells impaired CML formation in transplanted mice, however allowed ALL to develop (Zhao et al. 2007). In this case β-catenin loss also resulted in higher latency of disease progression and impaired self-renewal of CML stem cells, suggesting a potential differential requirement of β-catenin in ALL and CML.

Imatinib, a drug which targets BCR-ABL, can achieve clinical remission in the chronic phase of CML however relapse frequently occurs when treatment is ceased (Hu et al. 2009). It

was found that whilst proliferating BCR-ABL-induced LSCs were sensitive to Imatinib, a subset of drug resistant LSCs remains which rely critically on β-catenin (Heidel et al. 2012; Hu et al. 2009). In AML, it was found that LSC development from HSCs transformed by *HoxA9* and *Meis1* requires β-catenin. It was also found that MLL-AF9 transformed GMPs and HSCs required β-catenin for self-renewal and in vivo disease development (Wang et al. 2010a). However, using a compound β-catenin Loss Of Function (LOF) and KRAS Gain Of Function (GOF) mutant it was found that β-catenin is no longer required for MLL-AF9 LSC development (Ng et al. 2014).

The Glycogen Synthase Kinase 3 family (GSK3), key components of the β-catenin destruction complex (see the previous section 1.7), are often mutated in cancer leading to greater levels of β-catenin (McCubrey et al. 2014). Seemingly paradoxically, treatment of MLL-fusion associated with GSK3 inhibitors have demonstrated promising results (Wang et al. 2008; Banerji et al. 2012; McCubrey et al. 2014). However, in addition being involved in the β-catenin destruction complex, the GSK3 family are involved in self-renewal pathways central to AML such *Hoxa9/Meis1* (Wang et al. 2010b). As previously mentioned, our group reported that β-catenin is activated during the development of LSC from MLL fusion transformed pre-LSCs. Moreover, these LSCs displayed increased drug resistance to GSK3 inhibitors and could be re-sensitized by β-catenin depletion (Yeung et al. 2010). More recently, in addition to GSK3 inhibitor resistance, it has been demonstrated that β-catenin may contribute to acquisition of resistance to BET inhibition (Rathert et al. 2015; Fong et al. 2015).

In addition to disease and genetics, we recently demonstrated that cell-of-origin can dictate β-catenin requirement of LSCs (Siriboonpiputtana et al. 2017). Although GMPs express only low levels of *HoxA9* and β-catenin, they are both needed for transformation by MLL fusions. HSCs express β-catenin and *HoxA9* at higher levels than GMPs, however can still be transformed without one of these. Biologically, while β-catenin deleted GMP-MLL-ENL did not produce leukaemia in transplanted mice, MLL-ENL transformed LSKs did not require β-catenin for LSC development. This result was in contrast with previous studies that showed β-catenin was required for MLL-fusion associated leukaemia regardless of cell-of-origin (Wang et al. 2010a). β-catenin cell-of-origin specific dependency was shown to be specific to MLL-ENL and was not observed for other common MLL fusions (Siriboonpiputtana 2014; Siriboonpiputtana et al. 2017). In the Siriboonpiputtana et al. (2017) study, KO of *Hoxa9*, one of the major targets of MLL-ENL (Zeisig et al. 2004), was also detrimental to GMP-MLL-ENL LSC formation, but not to that of LSK-MLL-ENL, suggesting that β-catenin and *HoxA9* may both regulate pathways critical for LSK-MLL-ENL LSC development.

1.8 Project Aims

The epi/genetic events initiating LSC development critically shape the biology, pathology and response to therapy of AML. Learning from our group's recent research and others, it appears that β-catenin dependency in vivo varies with different LATFs and cell-of-origin. The role of β-catenin in facilitating aberrant gene expression and cancer development has been studied extensively in solid tumour malignancies. However, it remains unclear how β-catenin influences transcription in leukaemia. Understanding the mechanisms by which β-catenin facilitates MLL LSC development in distinct cells of origin may facilitate the design of rational, novel therapies targeting the poor AML subgroup of MLL-fusion associated leukaemia.

The aims of this PhD thesis are as follows:

- To identify a global set of direct/ downstream target genes of β-catenin in origin specific MLL transformed cells.
- To investigate the effects of β-catenin target expression perturbation
- To characterise epigenetic modifications mediated by β-catenin in MLL transformed cells.

Chapter 2

Methods

2.1 Production of β-catenin conditional pre-LSCs

In this section, I will describe the methods that were used to generate both previously published cell-of-origin specific MLL-ENL transformed Pre-LSC cell lines (Siriboonpiputtana et al. 2017) and additional cell lines I produced during this study.

2.1.1 β-catenin conditional mouse lines

To investigate the roles of β-catenin in haematopoietic cells transformed by MLL-ENL in vitro referred to as Pre-LSCs we employed an animal model in which the gene encoding β-catenin, catenin beta 1 (*Ctnnb1*), could be conditionally inactivated. The mouse model was created by interbreeding mice harbouring an allele of *Ctnnb1*, with its critical exons (2-6) flanked by LoxP (floxed) recombination sites and transgenic mice expressing Cre-recombinase fused an oestrogen receptor (Cre-ER). In the absence of its ligand, the synthetic antagonist of oestrogen, 4-hydroxytamoxifen (4-OHT), Cre-ER is bound to heat shock protein-90 and confined to the cytoplasm. 4-OHT bound Cre-ER translocates to the nucleus where Cre recognises LoxP sequences and induces genetic recombination.

Exons 2-6 of *Ctnnb1* encode the N-terminal and 1-4 armadillo repeats, excision of these by Cre-recombinase results in loss of function (LOF) of β-catenin. This mouse was originally produced in the R.Kemler laboratory and is known as *Ctnnb1*$^{\mathrm{Kem/flox}}$ (Brault et al. 2001), the transgene and recombination product is illustrated in Figure 2.1A.

In addition to this, I also crossed mice harbouring an allele of *Ctnnb1* with LoxP sites flanking exon 3 alone, produced by M.M.Taketo's laboratory, *Ctnnb1*$^{\mathrm{Mmt/flox}}$ (Harada et al. 1999) with transgenic mice expressing Cre-ER. Exon 3 of *Ctnnb1* encodes phosphorylation sites which are targeted by GSK3β for ubiquitination and destruction of β-catenin. Thus, deletion of exon 3 results in expression of a permanently active form of β-catenin which cannot be degraded by the β-catenin destruction complex. Constitutively active β-catenin builds up in the cytoplasm leading to subsequent nuclear translocation and target gene expression. The transgene and recombination product of *Ctnnb1*$^{\mathrm{Mmt/flox}}$ is illustrated in Figure 2.1B

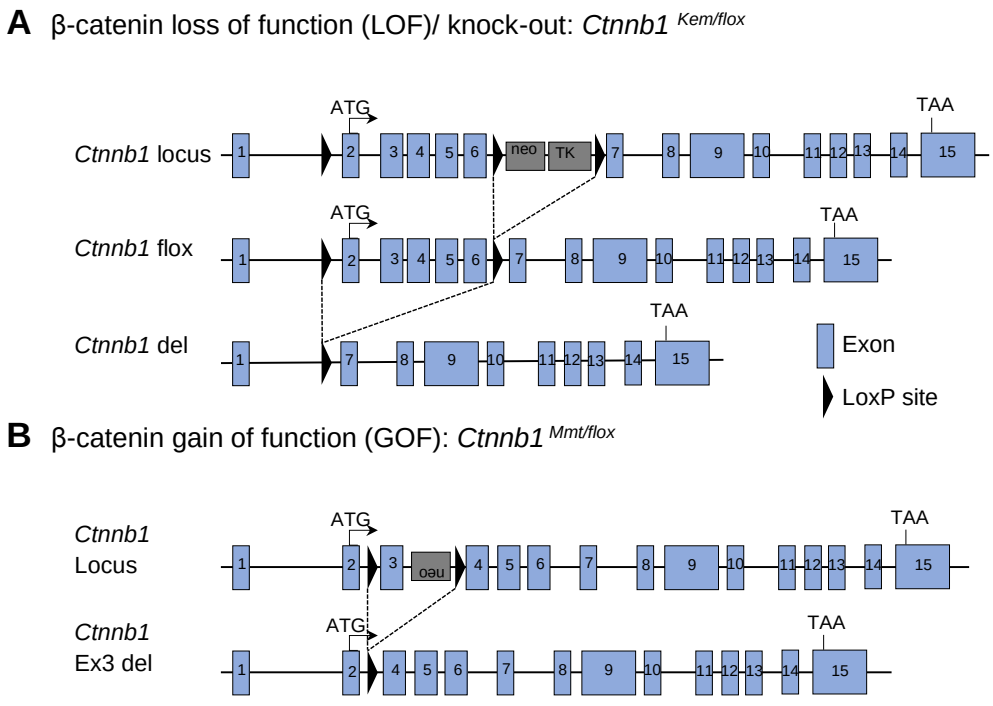

Figure 2.1: β-catenin conditional alleles. A. The *Ctnnb1* Kem/flox LOF allele: In mouse ESCs, a loxP site was inserted into the first intron and a neomycin-TK cassette was inserted into the sixth intron of *Ctnnb1*. Prior to production of chimeric mice, transient Cre-recombinase expression using a vector was used to remove the neomycin-TK cassette leaving exon 2-6 floxed. B. The *Ctnnb1* Mmt/flox GOF allele.

2.1.2 Genotyping PCRs

To determine the genotype of mice, ear biopsies were taken, and DNA was extracted using alkaline lysis buffer (25 mM NaOH; 0.2 mM EDTA) at 95°C for 20-60 minutes. After returning to room temperature an equal volume of Neutralizing Buffer (40 mM Tris-HCl) was added to samples. Once neutralised, the ear biopsy DNA is ready for genotyping Polymerase Chain Reaction (PCR). The DNeasy Blood & Tissue Kit (QIAGEN, catalogue number 69581) was used to extract DNA from in vitro cell lines and the liver, spleen, bone marrow and blood of experimental mice.

Genotyping primer sequences for *Ctnnb1* Kem/flox loss of function LOF were taken from Brault et al. (2001) and are shown in Table 2.1. Primer sequences for *Ctnnb1* Mmt/flox GOF were taken from our collaborator and are shown in Table 2.2. To perform genomic PCR, primers were ordered from Sigma. The PCR reaction mix was prepared with the appropriate primers at 0.5 µM, 0.2 mM of mixed dNTPs (Roche), 2 µL 10 × DreamTaq™ Green buffer (Fermentas), 0.5 µL of DreamTaq™ DNA Polymerase (Fermentas) and ddH20 up to 20 µL for each reaction. PCR was performed using a thermocycler with the following program: 94°C for 5 minutes to denature dsDNA followed by 30 cycles at 94°C, 56°C then 72°C each for 30 seconds. PCR products were separated by electrophoresis on a 1.5% agarose gel. *Ctnnb1* Mmt/flox genotyping PCR resulted in 3 possible discretely visible bands at 221 bp, 324 bp and 500 bp for wildtype, floxed and deleted *Ctnnb1* respectively. Two separate genotyping PCRs for *Ctnnb1* Mmt/flox were used to distinguish the wildtype and mutant alleles detected as a single band at around 100 bp.

Table 2.1: LOF *Ctnnb1*$^{\text{Kem/flox}}$ genotyping primers

Primer	Sequence (5' to 3')
RM41	AAG GTA GAG TGA TGA AAG TTG TT
RM42	CAC CAT GTC CTC TGT CTA TTC
RM43	TAC ACT ATT GAA TCA CAG GGA CTT

Table 2.2: GOF *Ctnnb1*$^{\text{Mmt/flox}}$ genotyping primers

Primer	Reaction	Sequence (5' to 3')
WT Fwd	Wildtype	GCA AGA GCA AGT AGC TGG TAA AG
WT Rev	Wildtype	CCA GCC CTG TCA AGA AAC TTA ATG
Mutant Fwd	Mutant	CGT CCA GCC AAG CTT AGG AT
Mutant Rev	Mutant	CGG TGG CTT GCT GAT TAT TTC AC

2.1.3 Haematopoietic stem cell and progenitor sorting

Bone marrow BM was harvested from mice with the appropriate genotype, and GMP and LSK were Fluorescence-Activated Cell Sorted FACS as previously described (Yeung and So 2009). Briefly, bone marrow cells were flushed from femur and tibias with Sorting Media (SM) containing 0.2% Foetal Bovine Serum (FBS), 1% Penicillin/ Streptomycin in 1× Phosphate-Buffered Saline (PBS) and filtered through a 40 µm nylon cell strainer. Throughout the sorting process, cells were kept on ice to preserve viability. To isolate white blood cells, cells were centrifuged at 1250 rpm for 5 minutes at room temperature and resuspended in 1 mL red cell lysis buffer (155 mM NH4Cl, 12 mM NaHCO3 and 0.1 mM EDTA) and incubated for 10 minutes. The cells were resuspended with SM to wash before centrifugation. Cells were counted with a haemocytometer and a small aliquot of cells were taken to act as a negative control to aid in FACS compensation.

To increase the FACS efficiency and reduce machine processing time, BM samples were depleted of lineage committed cells by magnetic bead selection. Cells were incubated in 0.5 mL SM per mouse harvested with 1:200 of unconjugated rat IgG antibodies targeting the following lineage markers; Ter119, Mac1, gr1, CD3e, CD4, CD8 and B220 for 30 minutes at 4°C. After incubation cells were washed in SM, pelleted by centrifugation and resuspended in SM. A small aliquot was taken at this stage to act as a positive control for lineage markers and to monitor success of lineage depletion. 1 mL of sheep anti-rat IgG Dynabeads® (Invitrogen) were prepared per 1×10^8 cells counted. Lineage-stained cells were washed and then transferred to capped FACS tubes and incubated with prepared Dynabeads® for 30 minutes at 4°C on a rotator to bind lineage positive cells. The cell and bead mix were then applied to a magnet, and the supernatant enriched for lineage negative cells was collected. The beads were washed twice to recover additional unbound cells, recovered cells were then washed and pelleted.

Lineage-depleted and the lineage positive aliquot were then incubated with secondary antibody anti-rat-TR (Texas Red) 1:10 for 30 minutes at 4°C in darkness. After washing the cells with SM, lineage depleted cells were incubated with rat IgG (Sigma) to block for 30 minutes at 4°C. Lineage depleted BM was washed in SM and stained with c-kit-PE (1:100), sca1-PECy-7(1:50), CD34-FITC(1:25), CD16/32-APC (1:75) for 30 minutes at 4°C in darkness. Following incubation, cells were washed again and passed through a 40 µm cell strainer to avoid machine blockage. Lineage positive and lineage depleted cells were resuspended in SM with 1 µg/mL Propidium Iodide (PI) to detect dead cells.

Finally, cells were sorted using a BD Aria cell sorter (BD biosciences, Oxford, U.K.) into pre-warmed R20 (RPMI, 20% FBS and 1% Penicillin/Streptomycin). The LSK compartment cells, which are enriched for HSCs were isolated by the immunophenotype Lin- c-Kit+ Sca-1+, and myeloid progenitors GMP cells with Lin- c-Kit+ Sca-1- CD34+ CD16/32high/+. After sorting up to 2 $\times 10^4$ cells, these were resuspended in 250 µL R10 (RPMI with 10% FBS and 1% Penicillin/Streptomycin) with 20 ng/mL SCF, 10 ng/mL IL3 and 10 ng/mL IL6 added to maintain stemness. Cells were seeded into 96 well plate and incubated at 5% CO2 and 37°C overnight to recover.

2.1.4 Retroviral transduction and transformation assay

To create MLL-ENL transformed cell lines, cells were transduced with MLL-ENL expressing retrovirus. The MLL-ENL fusion was previously cloned into Murine Stem Cell Virus (MSCV) plasmids obtained from our clone database. RTTA was carried out as previously described (Zeisig and So 2009).

Virus packaging

3×10^6 GP2-293 retrovirus packaging cells cells (Clonetech) were seeded in 8 mL of D10 (Dulbecco's Modified Eagle Medium (DMEM), 10% FBS, 1% penicillin and streptomycin) in 10cm plates and were left to adhere overnight. The next day, transfection was either carried out using the Calcium phosphate precipitation method or the polycation polyethylenimine (PEI) method. Both methods work by facilitating DNA to bind to the cell surface which can then be taken up by endocytosis. For Calcium phosphate precipitation, 15 µg of the construct was mixed with 8 µg pVSV-G, 61 µL of 2M CaCl2, and 500 µL 2×HBS. For PEI transfection, 8 µg of the construct was mixed with 2 µg pVSV-G and diluted to 1 mL in culture media containing no serum. 30 µL PEI was added dropwise to the diluted DNA, the mixture was then mixed using a vortex mixer immediately for 30 seconds and incubated for 15 minutes. Finally, the mixture was added dropwise to the plates. 24 hours after transfection the media was aspirated and replaced with fresh D10 to avoid plasmid and reagents from appearing in the virus collection.

Viral supernatants were collected 48h after transfection and filtered with a 0.45 µm filter before concentration with the Optima™ L-100 XP Ultracentrifuge, Beckman Coulter at 25,000 rpm, 4° C for 3 hours and subsequently stored at -80°C. Virus collection was repeated at 72 hours post-transfection.

Viral Transduction and Transformation

Approximately 150 µL of R10 was aspirated from the wells containing LSK and GMP sorted cells leaving only 100 µL. 100-200 µL of concentrated virus, supplemented at a final concentration of 20 ng/mL SCF, 10 ng/mL of IL3, 10 ng/mL of IL6, and polybrene (10 mg/mL) was added to each well. To increase contact between cells and viral particles and thus increase transduction efficiency cells were centrifuged or "spinoculated" for 2 hours at 800 g and 32°C. 24 hours later cells were plated in methylcellulose (MethoCult™ GF M3434) with 20 ng/mL SCF, 10 ng/mL of IL3, IL6, penicillin, 100 µg/mL streptomycin. GM-CSF at 100IU/mL was also added prior to plating to encourage cells to move towards the myeloid lineage. To isolate cells which had been successfully infected with virus containing the MLL-ENL fusion, the media also contained the appropriate antibiotic for which resistance gene was expressed by the construct 1.3 µg/mL and 1 µg/mL was used for Puromycin and Blasticidin respectively. Methylcellulose was dispensed into a 24 well plate with a 18G blunt-end needle and unused wells were filled with water to increase humidity inside the plate and prevent the media from drying out.

Colonies were allowed to grow a further 5-7 days and then their phenotypes were scored. Colonies were extracted by dissolving methylcellulose in PBS and approximately 10,000 viable cells were replated in methylcellulose. After 3 rounds of plating, only cells with enhanced self-renewal capacity form colonies, whereas non-transformed cells exhaust their proliferative capacity in previous rounds of plating. Cell lines were established by growing 3rd round colonies in R20/20 (RPMI1640, 20% FBS, 20% WEHI, 100U/ML penicillin and 100 µg/mL streptomycin). WEHI is a conditioned media, a source of IL3 produced by culturing WEHI cells in RPMI before harvesting and filtering. This resulted in MLL-ENL transformed immortal cell lines of LSK and GMP with conditional deletion of β-catenin, which were used for all experiments in this study (Ben-David et al. 2018).

2.1.5 In vivo Leukaemogenesis assay

Swiss Jim Lambert (SJL) mice or C57BL/6, also known as Black 6 mice were sub-lethally irradiated with two rounds of 1100 rads. 5×10^5 - 1×10^6 Pre-LSCs or LSCs were prepared in 150 µL PBS per mouse for tail vein injection. Mice were monitored daily for loss of condition (symptoms including, weight loss, hunched posture, flat laying ears, piloerection, reduced movement and hind leg paralysis). Mice were euthanised upon signs of leukaemia development

and peripheral blood, spleen, liver and bone marrow were processed for analysis. On first inspection, leukaemic mice often present with enlarged spleen and liver (termed splenomegaly and hepatomegaly respectively) and the bone marrow appears pale alluding to high concentrations of myeloid (white) cells. Depending on progression of the disease, infiltration of leukaemic cells may be visible in the liver as white spots and the blood may appear anaemic (low red blood cell count). Invariably, leukaemia is identified if there are donor cells in the bone marrow, identified by FACS, distinguishable from the recipient cells with CD45 markers. Additionally, cells were stained for c-Kit (D7 clone), Mac-1 (M1/70 clone) and, Gr-1 (RB6-8C5 clone) to confirm myeloid disease.

2.2 Cell Culture and *Ctnnb1* Knock-out

The previously prepared cell lines were retrieved from stock in the cryogenic storage, with two biological replicates taken each for GMP and LSK MLL-ENL transformed cell lines. Cells were briefly rinsed with RPMI (Sigma-Aldrich), span down and then grown in R20/20 media, and cell viability was assessed after 24 hours. Each cell line was treated with Density Gradient Centrifugation (Ficoll 400, Sigma Aldrich) to remove dead cells. These cells were allowed to recuperate for 48 hours and cell viability was assessed again. When the viability reached above 90%, cells were harvested and frozen for future use. Approximately 200 million cells were grown between two T175 flasks for each condition in R20/20 media with 0.25 $\times$ cytokines SCF GM-SCF, IL6 and IL3 at 5% CO2.

To introduce the β-catenin deletion for the selected cell lines, cells were treated with 20 nM 4-Hydroxytamoxifen (4-OHT) in ethanol suspension (deletion) or 0.1% total ethanol (control) for 72 hours. Media were replenished including appropriate concentrations of drug or vehicle as the acidity indicator in the media began to turn yellow. Cells were checked daily for viability using a haemocytometer and Gibco™ trypan blue (Thermo Fisher Scientific), and were discarded if viability declined below 80%. To prepare the cells for extracting RNA, DNA and protein, aliquots of 2-4 million cells were pelleted at 1300 rpm and snap-frozen at -80°C. The remaining cells were harvested for formaldehyde fixation and subsequent ChIP experiments. This procedure guaranteed that ChIP, RNA and DNA sequencing had been performed on exactly the same cells under the same conditions. This approach facilitates meaningful comparisons across the different types of data. A schematic of the experimental set-up is illustrated in Figure 2.2.

2.3 Quantitative Reverse Transcriptase Polymerase Chain Reaction (RT-qPCR)

RNA was extracted with mirVana miRNA Isolation Kit™ (ThermoFisher Scientific, catalogue number AM1560) or the RNEasy kit (Qiagen). cDNA was produced for each sample by adding 1 µg of template RNA to 1 µL oligo dT 100 µM and 1 µL 10 mM dNTPs. The mixture was incubated for five minutes at 65°C and then on ice for one minute. The first strand buffer was added in combination with 0.1M DDT, 1 µL RNAse out and 1 µL SuperScript™ III Reverse Transcriptase (Thermo Fisher). This mixture was then placed in a thermocycler with the following program: 25°C for 5 min 50°C for 60 min and 70°C for 15 min. Previously designed Reverse Transcriptase quantitative PCR (RT-qPCR) primers were used (Brault et al. 2001) and can be found in Table 2.3.

Table 2.3: RT-qPCR primers

Gene	Direction	Sequence (5' to 3')
Gapdh	Forward	AGGTCGGAGTCAACGGATTTG
Gapdh	Reverse	GACGGTGCCATGGAATTTG
Ctnnb1	Forward	ATGGAGCCGGACAGAAAAGC
Ctnnb1	Reverse	CTTGCCACTCAGGGAAGGA
Tgm2	Forward	GATGACCAGGCCAGACCTAC
Tgm2	Reverse	CGAAGGGTGCATCATACTTG
Hoxa9	Forward	GCCTTCTCCGAAAACAATGCCG
Hoxa9	Reverse	TTCCGAGTGGAGCGAGCATGTA
Gadd45a	Forward	GCTGCCAAGCTGCTCAAC
Gadd45a	Reverse	TCGTCGTCTTCGTCAGCA
Cyclin D1	Forward	AAGTGCGTGCAGAAGGAGAT
Cyclin D1	Reverse	TTAGAGGCCACGAACATGC

2.4 Western Blot

Protein was extracted from frozen cell pellets in 100 µL sample buffer per 1×10^6 cells. Samples were placed at 95°C for 5 minutes. 20 µL of each protein sample and 10 µL of protein ladder were ran on 2-20% SDS polyacrylamide gels at 150 V for 90 minutes. Protein was transferred to nitrocellulose membrane (Amersham™ Hybond™ ECL) at 100 V for one hour. The membrane was blocked with PBS with 2% milk and 0.1% tween for one hour at room temperature. 1 in 2,000 dilution of mouse anti-β-catenin antibody (Purified Mouse Anti-β-catenin Clone 14/Beta-Catenin,

BD Transduction Laboratories™) overnight at 4°C. It was washed with PBS-T (5 minutes, 4 times) and incubated with secondary antibody or β-actin antibody conjugated with HRP (Actin (I-19; SC-1616) HRP, Santa Cruz Biotechnology) for one hour at room temperature. Then the membrane was washed with PBS-T (5 minutes, 4 times). Chemiluminescent signal was detected by using ECL Western Blotting Detection System (Amersham™, GE Healthcare, UK). Signals were developed on X-Ray film (Amersham Hyperfilm™, ECL).

2.5 RNA-Sequencing (RNA-Seq) and analysis

2.5.1 A note on experimental design

Differential gene expression analysis aims to uncover statistically significant differences in mRNA levels between conditions of interest. Statistical hypothesis testing allows us to formally reject the null hypothesis, if assuming the null hypothesis is correct and the data approximately fits a known distribution, the probability (P.value) of obtaining the observed data is very low.

However, the likelihood of exceptional data increases as more hypotheses are tested, this is known as the "multiple testing problem". To mitigate this, mathematicians have developed multiple test correction methods such as the Bonferroni correction. The correction is applied by testing each hypothesis at the desired P.value needed to convincingly reject the null hypothesis (often 0.05 as proposed by R.A. Fischer in the 1920's), divided by the total number of hypotheses tested. Nevertheless, high-throughput technologies such as RNA-Seq present an additional hurdle as the Bonferroni correction is found to be too stringent when a large number of variables/hypotheses are tested (many thousands of genes). Furthermore, false discovery rate methods such as the Benjamini Hochberg (BH) method were developed (Benjamini and Hochberg 1995) and are used by popular differential expression analysis software such as EdgeR (Robinson et al. 2010) and DESeq2 (Love et al. 2014).

To determine if a gene is significantly differentially expressed, like any other statistical comparison, we rely on being able to reliably estimate the variability across multiple observations. Making up this variability is both technical; due to samples processing, sequencing and biological: natural variability existing in the population and in response to treatment. In RNA-Seq, technical variance is generally very low compared to the biological variance if experiments are carefully controlled.

Biological variance comes from both the stochastic character of gene expression itself and differences in expression between individuals of the same group. Reducing confounding biological variability can be done where possible by using organisms of the same sex, age and litter at the same time. However, as long as there is a record of potential confounding factors, we may remove some of this during analysis by modelling for known sources of variation. Too much variability can reduce our ability to detect differential gene expression caused by the experimental condition we are interested in.

Biological replicates can be defined depending on the inference to be made by the biological question. To allow us to make generalisations about the effect of a treatment in a particular cell line, biological replicates can be made by processing and treating the same cell line separately. However, if we want to make a generalisation about an experimental condition in a variable population, such as MLL-ENL transformed pre-LSCs, biological replicates from different individuals and are more appropriate. Separate sorting and transduction of biological replicates also encapsulates biological variation introduced by the sorting and MLL-ENL transformation process.

2.5.2 RNA-Seq Libraries

RNA was extracted using the mirVana miRNA Isolation Kit™ (ThermoFisher Scientific, catalogue number AM1560). Importantly, this RNA extraction kit retains micro RNAs. RNA concentration was measured, and purity was assessed using NanoDrop 8000 (Thermo Fisher Scientific). The Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) was used to measure RNA integrity (RIN), all RIN scores were determined to be between 9 and 10. The TruSeq Stranded Total RNA kit with RiboZero Gold (Illumina, Inc., catalogue number RS-122- 2201) was used to produce the library according to manufacturer's protocol. Appropriate size selection was confirmed using Agilent 2100 Bioanalyzer DNA high sensitivity chip (Agilent Technologies, catalogue number 5067-4626). The final library concentrations were measured using Qubit® dsDNA HS Assay Kit (ThermoFisher Scientific, catalogue number Q32854) and lanes were normalised to 2 nM for sequencing.

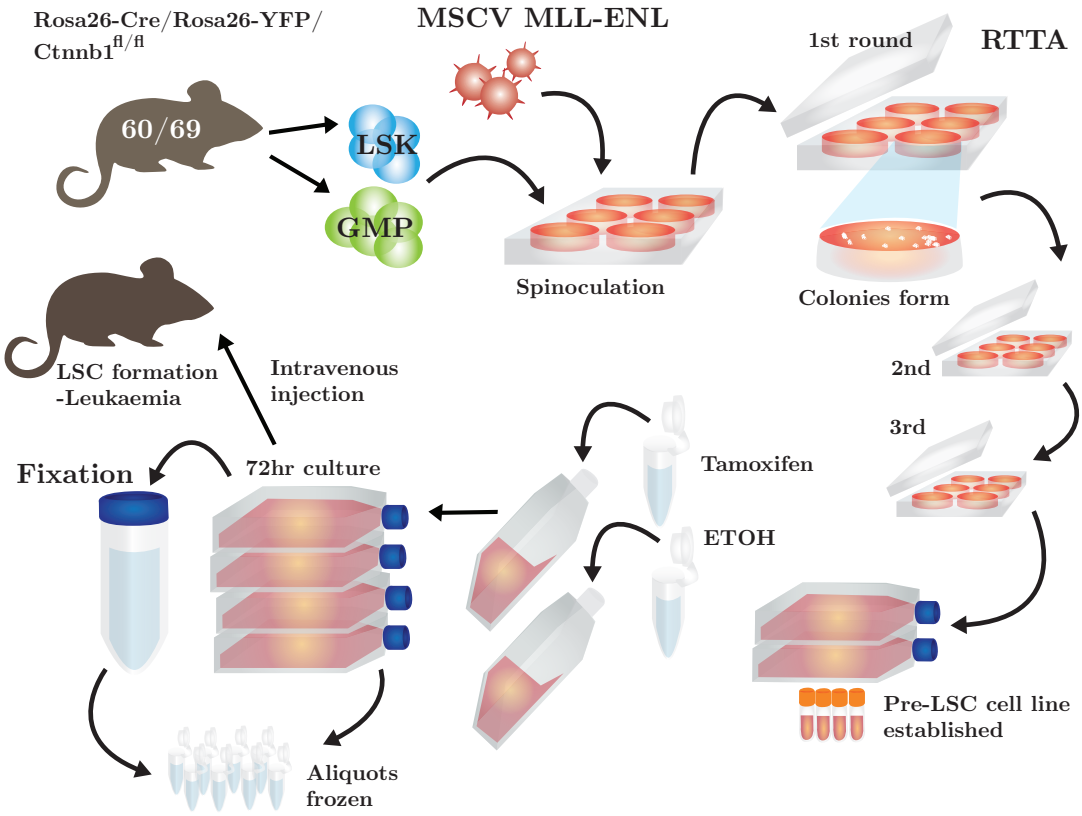

Figure 2.2: Schematic of experimental set-up In this schematic, the experimental set-up of the previous work and currently performed work are described. Firstly, bone marrow was extracted from *Ctnnb1*$^{fl/fl}$ mice, LSKs and GMPs were sorted by flow cytometry. GMPs and LSKs were spinoculated with MSCV expressing MLL-ENL and puromycin resistance. The transduced cells were then grown in methylcellulose containing puromycin. The resulting colonies were extracted and re-plated two times. The third round colonies were grown in R20/20 in the presence of cytokines to establish the cell line. Cells were grown in the presence of either tamoxifen 4-OHT or Ethanol (ETOH) for 72 hours. Previously, *Ctnnb1*$^{fl/fl}$ and *Ctnnb1*$^{-/-}$ Pre-LSCs were transplanted into recipient mice to monitor disease progression. For ChIP experiments, the cells were massively expanded in T175 culture flasks before treatment. A small number of cells were aliquoted for RNA, DNA and protein extraction. The remaining cells, approximately 200 million per experiment, were fixed in 1% formaldehyde, aliquoted and frozen for ChIP.

2.5.3 RNA-Seq Processing

In this section I provide details of the initial processing pipeline for RNA-Seq, further details are provided in subsection 3.2.3. Initially, quality checking of raw reads was performed using the software FASTQC v0.11.4 (Andrews 2015), and reads were trimmed for adapters and low quality bases using Trim Galore! v0.4.0 (Krueger 2015). Trimmed reads were then mapped to the mouse reference genome version GRCm38 (obtained from the Ensembl FTP server `ftp://ftp.ensembl.org/pub/`) using the short read mapper Tophat2 v2.1.0 (Kim et al. 2013).

The GRCm38 genome assembly uses the same scaffold as UCSC's mm10 however they are each based on GENCODE and RefSeq annotation respectively. Notably, when compared, GENCODE transcripts have more exons and therefore describe more gene variants than RefSeq (Frankish et al. 2015). Picard tools MarkDuplicates and CollectRnaSeqMetrics v2.0.1 (Broad Institute 2017) were used to mark duplicates and obtain other RNA-Seq mapping statistics. Read counts per gene were obtained using htseq-count v0.11.1 (Anders et al. 2015) using union mode and the "strandedness" option set as reverse.

2.6 *Gadd45a* overexpression

Mouse *Gadd45a* (NM007836) open reading frame (ORF) clone tagged with DYKDDDDK was purchased from GenScript (Catalog No: OMu16639D). To express *Gadd45a* stably in MLL transformed leukaemic cells, the ORF was cloned into MSCV retroviral expression vectors. Firstly, the *Gadd45a* plasmid was PCR-ed with Easy-A High-Fidelity PCR Cloning Enzyme (Agilent) to introduce the appropriate restriction enzyme sites as shown in Table 2.4. I designed primers against the Open Reading Frame (ORF) and added restriction enzyme site sequences (GAATTC for EcoRI and CTCGAG for XhoI) on each end creating "sticky ends" for ligation. To increase efficiency of restriction cutting "AATT" was also added prior to each restriction site. PCR products were purified using the QIAquick PCR Purification Kit (Qiagen) following manufacturers instructions.

Table 2.4: Primers to amplify *Gadd45a* ORF and add restriction sites

Primer name	Sequence 3' - 5'
EcoRI For	AATTGAATTCGCCACCATGACTTTGGAGGAATTC
EcoRI Rev	AATTGAATTCTTATCACTTATCGTCGTCATCCTTGT
XhoI For	AATTCTCGAGGCCACCATGACTTTGGAGGAATTC
XhoI Rev	AATTCTCGAGTTATCACTTATCGTCGTCATCCTTGT

Briefly, MSCV plasmids with the appropriate antibiotic resistance gene were cut with restriction enzymes (FastDigest, ThermoFisher Scientific) at 37°C for 20 minutes. The digested plasmid was run on a 0.8% agarose gel. The linearised plasmid was cut out of the gel with a clean scalpel and purified using the QIAquick Gel Extraction Kit (Qiagen). The *Gadd45a* PCR product and the linearised plasmid were then ligated overnight at 16°C with T4 ligase (NEB). Competent DH5 alpha cells were transformed with the ligation product and spread onto LB agar plates containing ampicillin. Several single clones were picked and grown in 5 mL LB media at 37°C overnight. DNA was extracted from 4 mL of each clone with the QIAPrep Miniprep Kit and digested with the appropriate enzymes.

The digestion products were run on a 0.8% agarose gel to pick promising clones for sanger sequencing. The remaining cultures which perfectly sequence matched the ORF in the correct orientation were used to prepare the plasmid for virus production with the QIAPrep Maxiprep Kit (Qiagen). Retroviral particles were produced as described in section 2.1.4, using the PEI method. Approximately 3-5 million cells were seeded into a 6 well plate and resuspended with viral supernatant and polybrene at a final concentration of 4 µM. Cells were spinoculated for 2 hours at 800 *g* and 32°C. The next day, the viral supernatant was removed and replaced with fresh R2020 supplemented with cytokines. 48 hours after spinoculation, cells were treated with the appropriate antibiotic in vitro. 72 hours after selection, viable cells were collected for RNA and protein, 5000 cells were plated in methylcellulose and cells were taken for MTT assay, cytospin and RT-qPCR.

2.7 *Gadd45a* knockdown

Short Hairpin RNA (shRNA) sequences against *Gadd45a* were generated using Sigma's online tool and the top 3 predictions were selected. Sequences were added to either end of the oligos to allow for cloning into the pLKO lentivirus backbone (Stewart et al. 2003), the resulting sequences were ordered from sigma and are shown in Table 2.5.

Table 2.5: *Gadd45a* shRNA Sequences

shRNA name	Primer name	Sequence
sh 1	NM_007836.1-575s1c1_Gadd45a_F	CCG GCC TGC CTT AAG TCA ACT TAT TCT CGA GAA TAA GTT GAC TTA AGG CAG GTT TTT G
sh 1	NM_007836.1-575s1c1_Gadd45a_R	AAT TCA AAA ACC TGC CTT AAG TCA ACT TAT TCT CGA GAA TAA GTT GAC TTA AGG CAG G
sh 2	NM_007836.1-251s1c1_Gadd45a_F	CCG GGC TCG GAG TCA GCG CAC CAT TCT CGA GAA TGG TGC GCT GAC TCC GAG CTT TTT G
sh 2	NM_007836.1-251s1c1_Gadd45a_R	AAT TCA AAA AGC TCG GAG TCA GCG CAC CAT TCT CGA GAA TGG TGC GCT GAC TCC GAG C
sh 3	NM_007836.1-550s1c10_Gadd45a_F	CCG GCC CAC ATT CAT CAC AAT GGA ACT CGA GTT CCA TTG TGA TGA ATG TGG GTT TTT G
sh 3	NM_007836.1-550s1c10_Gadd45a_R	AAT TCA AAA ACC CAC ATT CAT CAC AAT GGA ACT CGA GTT CCA TTG TGA TGA ATG TGG G

The first step in the shRNA cloning process was to anneal the oligos to form double stranded DNA. Oligos were resuspended in double distilled water (ddH2O) at 20 µM and then 5 µL of each oligo were combined with 5 µL of 10×NEB buffer. ddH2O was added for a total volume of 50 µL and the mixture was denatured at 95°C for 4 minutes. Oligos were incubated at 70°C for 10 minutes and then allowed to cool to room temperature by slowly reducing the temperature to facilitate annealing.

The pLKO vector was digested by both BshTI and EcoRI restriction enzymes (FastDigest, ThermoFisher Scientific) for 30 minutes at 37°C. The digested plasmid was run on a 0.8% agarose gel. The linearised plasmid (7 kb) was cut out of the gel with a clean scalpel and purified using the QIAquick Gel Extraction Kit (Qiagen). The concentration of the digested vector was measured using the NanoDrop™ 2000 (Thermofisher Scientific). 2 µL of annealed oligo, 20 ng digested pLKO, 0.5 µL T4 DNA ligase (NEB) and 2 µL 10× NEB T4 DNA ligase buffer was combined and the final volume was brought to 20 µL with ddH2O. The ligation reaction was incubated at 16°C for overnight. Competent DH5 alpha cells were transformed with 2 µL ligation mix and spread onto Lysogeny Broth (LB) agar plates containing 100 µg/mL Ampicillin. 4 clones were picked and grown in 5 mL LB media at 37°C overnight. DNA was extracted from 4 mL of each clone with the QIAPrep Miniprep Kit (Qiagen) and was sent for Sanger sequencing. The remaining cultures with perfectly matched sequences of the shRNA in the correct orientation were used to prepare the plasmid for virus production with the QIAPrep Maxiprep Kit (Qiagen).

Lentiviral particles were produced using the Polyethyleneimine (PEI) method, similarly to retroviral production in section 2.1.4, however non-GP2 expressing Hek293 cells were used for packaging and lentiviral packaging plasmids were used. β-catenin conditional cells were transduced with lentiviral particles expressing the 3 shRNAs. RNA was collected for RT-qPCR to check knock down efficiency prior to cre mediated excision of β-catenin. The shRNA which showed the greatest reduction in mRNA was used in further experiments.

2.7.1 MTT assay

Cellular metabolic activity can be reliably investigated by measuring the metabolic reduction of tetrazolium MTT (3-(4, 5-dimethylthiazolyl-2)-2, 5-diphenyltetrazolium bromide). Clear yellow MTT when reduced by cellular dehydrogenases is converted into purple formazan. This colour change can be measured using a spectrophotometric reader and is linearly related to active cell numbers. After seeding equal cell numbers in equal volumes of media for at least 24 hours and avoiding confluency above 80%, 100 µL of culture was aliquoted in triplicate into a 96 well plate. 100 µL of MTT reagent (CT01-5 Sigma-Aldrich), was added to cells and placed back in the incubator for 1-3 hours before measuring spectrophotometric signal. 100 µL of fresh media was used as a background control to normalize the absorbance values.

2.7.2 Cytospin

To examine the morphology of cells, I resuspended 10,000-20,000 cells in 200 µL PBS and centrifuged them at 300 rpm for 5 minutes onto microscope slides using a Cytospin™ 4 Cytocentrifuge (Thermo Scientific™). In order to distinguish nuclei from the cytoplasm I stained the slides with the May-Grünwald stain (Sigma Aldrich) for 3 minutes. After rinsing the slides in water, I proceeded to stain the slides with Giemsa stain (1:20, Sigma Aldrich) for 20 minutes. Following a final wash with water and briefly drying, slides were examined under a microscope and photographs were taken.

2.7.3 Annexin Staining

Approximately 1 $\times$ 10^6 cells were washed in ice cold PBS twice and resuspended in 1 mL of 1$\times$ Binding Buffer (BD Biosciences). 100 µL of cell suspension was transferred to a FACSs tube, 2 µL Propidium Iodide (PI) was added to the cells and the mixture was left to incubate at room temperature for 15 minutes in the dark. After incubating, cells were washed in 1 mL 1$\times$ Binding Buffer. 0.5 µg Streptavidin-Fluorescein Isothiocynate (SAv-FITC, Cat. No. 554060) was diluted in 100 µL 1$\times$ Binding Buffer and added to the cells. Cells were incubated in the dark for another 15 minutes. After incubation, FACS tubes were topped up to 400 µL with 1$\times$ Binding Buffer and analysed by FACS.

2.8 Chromatin Immunoprecipitation coupled to high-throughput sequencing (ChIP-Seq) Generation

To uncover the chromatin configuration in cells with or without β-catenin present, I conducted ChIP-Seq of cells treated with antibodies against histone marks indicating open chromatin: H3K4me1, H3K4me3, and H3K27ac or associated with repressed chromatin: H3K27me3. An outline of the ChIP-Seq methodology is illustrated in Figure 2.3. Previously prepared cell cultures (section 2.2) were subjected to fixation at 1% formaldehyde for 10 minutes at room temperature. The cross linking reaction was quenched using 0.25M Glycine for 2 minutes at room temperature. Cells were rinsed in ice cold PBS, aliquoted and snap-frozen at -80°C. ChIP was then carried out as described below using an optimised protocol derived from the Agilent and Upstate protocol for ChIP on ChIP and modified by Lab Pierre Ferrier, Centre d'Immunologie de Marseille Luminy.

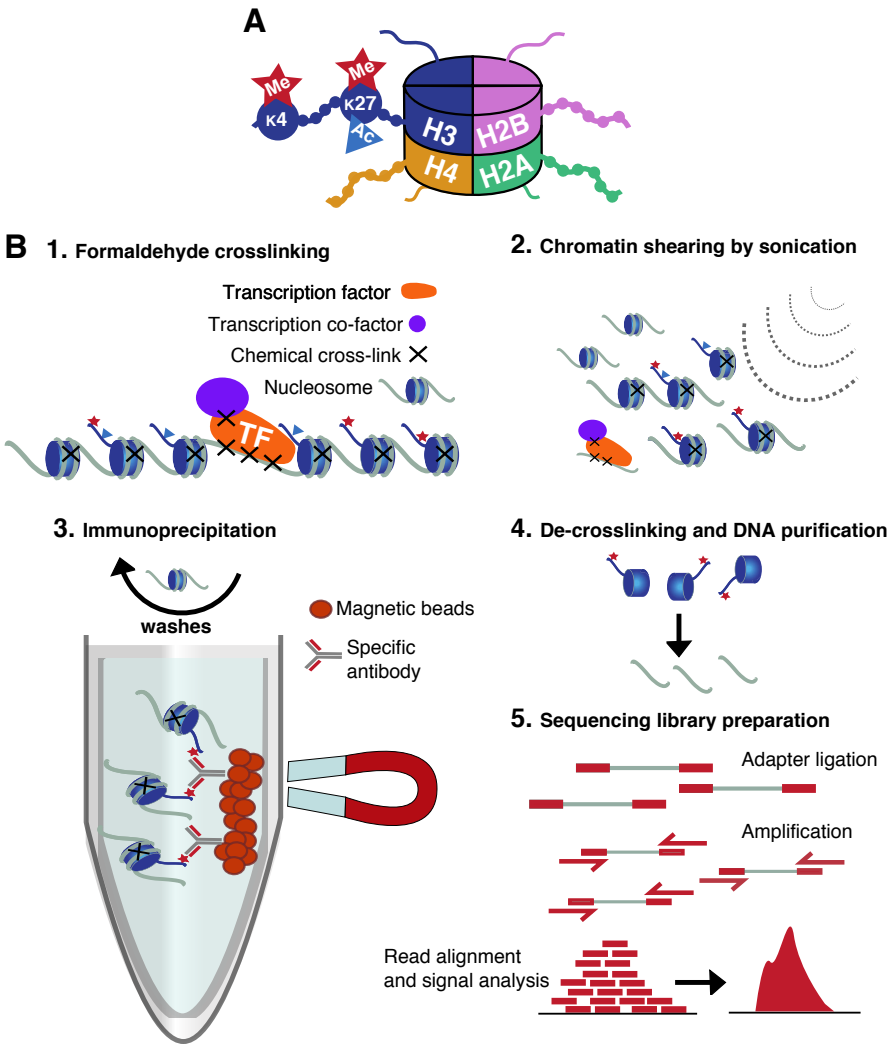

Figure 2.3: ChIP-Sequencing Schematic A) The histone octamer. Each histone protein has a tail in which amino acids, such as lysine at position 27 or 4, may be covalently modified for example by addition of an acetyl or methyl group. B) Schematic of ChIP-Seq method: 1. Firstly, cells are subjected to fixation with formaldehyde or another cross-linking agent. This permanently links bound proteins such as histones or transcription factors to the DNA molecule. Theoretically, cross-linking should also fix factors bound to transcription factors, however often cells are often subjected to several fixation agents to increase likelihood of this. 2. Next, cell lysate is subjected to mechanical shearing by sonication. This may also be done enzymatically. 3. Immunoprecipitation: Specific antibodies bound by magnetic beads are incubated with sheared cross-linked chromatin. Exposed epitopes bind to specific antibodies. DNA crosslinked to the protein of interest are separated from the solution with a magnet. Unbound DNA is removed by washing the beads. 4. Following immunoprecipitation, chromatin is de-cross-linked and purified. 5. Library preparation: Adapters are ligated to the DNA and the library is amplified, sequenced and analysed for signal enrichment.

Table 2.6: ChIP Lysis Buffers

Lysis Buffer 1	Lysis Buffer 2	Lysis Buffer 3
50 mM Hepes (pH 7.5)	200 mM NaCl	1 mM EDTA (pH 8)
140 mM NaCl	1 mM EDTA (pH 8)	0.5 mM EGTA (pH 8)
1 mM EDTA (pH 8)	0.5 mM EGTA (pH 8)	10 mM Tris-HCL (pH 8)
10% Glycerol	10 mM Tris-HCL (pH 8)	100 mM NaCl
0.75% Triton X-100		0.1% Na-Deoxycholate
0.25% NP-40		0.5% N-Lauroylsarcosine

Table 2.7: ChIP Wash Buffers

Low Salt Wash	High Salt Wash	LiCl Wash	TE Wash
0.1% SDS	0.1% SDS	250 mM LiCl	1 mM EDTA
1% Triton X-100	1% Triton X-100	1% NP 40	10 mM Tris (pH 8)
2 mM EDTA	2 mM EDTA	1% Sodium Deoxycholate	
20 mM Tris-HCL (pH 8)	20 mM Tris-HCL (pH 8)		
150 mM NaCl	500 mM NaCl		

The ChIP-Seq protocol is initiated by a series of lysis steps, aiming to isolate the cell nuclei, and then break up the chromatin so that it is accessible to the antibodies; cells from the previously frozen cell pellets were resuspended in Lysis Buffer 1 (Table 2.6) and were mixed at 4°C for 60 minutes. Cells were centrifuged at 1350 g for 5 minutes and resuspended in Lysis Buffer 2 (Table 2.6), then mixed on a rotating platform for 20 minutes. Cells were pelleted again at 1350 g and resuspended in Lysis Buffer 3 (Table 2.6) containing 2% SDS and mixed for 60 minutes. The nuclear fraction was then sonicated using the Bioruptor® Pico sonication device from Diagenode for 10 cycles (30 seconds on/off). The sonicated chromatin was span at 17900 rpm with 1% Triton-X-100 to pellet remaining cell debris.

For histone modification ChIP, 3 µg of histone antibody and 15 µL of Protein G Dynabeads® (Invitrogen) was used per 5 million cells. 15 µL of Protein G Dynabeads® were blocked with 0.5% BSA and 1× protease inhibitor cocktail. Antibodies used were Anti-Histone H3 (mono methyl K4) antibody (abcam cat no. ab8895), Anti-Histone H3 (tri methyl K4) antibody (abcam cat no. ab8580), Anti-Histone H3 (tri methyl 27) antibody (Merck Millipore 07-449) and Anti-Histone H3 (acetyl K27) antibody (abcam cat no. ab4729) (abcam, Cambridge, UK). Antibody was incubated with beads on a rotational wheel with beads for 4-12 hours at 4°C. Approximately 5-10 million fixed cells were used for each histone ChIP.

50 µL Protein G magnetic beads per 10 million cells of sonicated supernatant were blocked with 0.5% BSA. Beads were coupled with chromatin for 60 minutes to "pre-clear" chromatin of DNA which non-specifically binds to the beads. To act as quantitative control for both ChIP-qPCR and next generation sequencing, a small fraction (5-10%) of chromatin was taken from each sonication at this point and elution buffer was added and left at 65°C for 12 hours for reverse cross-linking. Whilst sonication causes DNA to fragment in a mostly randomised fashion, there is a slight bias towards breakage at open chromatin. This fragmentation bias, depending on the chromatin confirmation of the cells and the method of fixation and sonication, can cause false enrichment at open chromatin in ChIP-Seq. We can compensate for this bias by sequencing input DNA from each condition to facilitate enrichment cut-offs used in subsequent data analysis.

Remaining chromatin was added to the antibody conjugated beads for at least 12 hours at 4°C. To remove low affinity binding chromatin from the ChIP sample, chromatin and beads were washed sequentially for 10 minutes on a magnetic stand with the wash buffers listed in Table 2.7 the first wash,"Low Salt Wash" was repeated twice. Chromatin was then eluted and reverse cross-linked as the input sample was. Any remaining RNA from the input or ChIP sample was removed from the DNA by incubating it in 0.2 µg/mL RNAse for one hour at 37°C and 0.2 µg/mL Protease K for at least 2 hours at 55°C. QIAquick PCR purification (Qiagen) was performed on all sample ChIP DNA.

2.8.1 ChIP q-PCR

Percentage input, the proportion of total input DNA pulled down by the ChIP assay, was calculated using the ΔCt method. The first step of the ΔCt method is to calculate the average Ct value for the ChIP-ed DNA and a known fraction of the total input DNA. As input samples represent just 5% of the whole DNA input in the ChIP assay, the next step of the calculation is to adjust the Ct value obtained for the input to 100%. Assuming the input fraction and ChIP DNA are eluted in the same volume, we do this by subtracting the logarithm base 2 of the dilution factor from the average Ct value, in this case the dilution factor for 5% is 20. Next, the average Ct value for the ChIP DNA is subtracted from the adjusted input Ct to get ΔCt. Finally, we calculate 2 to the power of ΔCt and multiply this by 100 to get the percentage input. Primers used for ChIP-qPCR are listed in Table 2.8.

Table 2.8: ChIP-qPCR primers

Primer	Orientation	Sequence (5' to 3')
Meis1 TSS	Forward	GCGTCGCTGAGCATAAAAGC
Meis1 TSS	Reverse	CCTCCCACTGCTGTCTTGGT
Meis1 Enhancer 9	Forward	TTTCCAAGCTCAATGGAATGC
Meis1 Enhancer 9	Reverse	TTGCTGAGGCCAAGGGTAAC
Hoxa9 TSS	Forward	GAATTTGCAGGGAAAGGAAA
Hoxa9 TSS	Reverse	GGCAGGAAGAAGAAAGTGGT
Hoxa9 TSS region 2	Forward	TTGATGTTGACTGGCGATTTTC
Hoxa9 TSS region 2	Reverse	ATCTGTATGCCTAGTCCCGCTCC
p16 TSS	Forward	CTTAGAGTTACAGAAAGGGCTGGA
p16 TSS	Reverse	GAATTTCAAGGAAGTGCTACCCTA
Gene Desert	Forward	CCACTCTTCTTATAGGACCCTTTG
Gene Desert	Reverse	CCTGTCTACCTGTTCTTTACATTCT

2.8.2 ChIP-Seq library preparation

ChIP-Seq libraries were produced using the NEXTFLEX ChIP-Seq library kit (Bioo Scientific, catalogue number 5143-01). For H3K4me3, 10 ng of ChIP-ed DNA and 10 ng input DNA were used as starting material. For H3K4me1, H3K27me3 and H3K27ac, smaller amounts ranging between 2.5-6 ng were used. The ratio of beads to sample in the protocol were altered to select for a larger size range (200-400 bp) using Agencourt AMPure XP beads (Beckman Coulter, A63881). This process was optimised prior to library production using excess sheared fixed chromatin and analysis using an Agilent 2100 bioanalyser. The final average library size was determined to be 350 bp for all libraries. The libraries were pooled and normalised to 2 nM using Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific). ChIP-qPCR primers for the *Meis1* putative promoter were designed using primer3 software (Untergasser et al. 2012). ChIP-qPCR primers for the *Meis1* enhancer were taken from Wang et al. (2014).

2.9 ChIP-Seq processing and quality assessment

In this section I provide details of the initial processing pipeline for ChIP-Seq analysis which are briefly referred to in the results section 6.2. Specific analysis used to address biological questions are detailed within a separate methods section in section 6.2. Firstly, to remove adapters and low quality bases, I trimmed the resulting reads with Trim Galore! v0.4.0 (Krueger 2015). Next, I individually aligned each sequencing run to the mouse genome assembly GRCm38 with Bowtie2 v2.2.6 (Langmead and Salzberg 2012). Alignments were filtered for properly mapped pairs with Samtools v1.2 (Li et al. 2009). To distinguish reads originating from different runs and avoid excessive duplicate removal, before merging runs by sample, I annotated each Sequence Alignment Map (SAM) file with a unique read group reference using Samtools v1.2 re-header.

During the ChIP-Seq library preparation, ChIP DNA fragments are amplified to achieve DNA concentrations required for sequencing. In the case of over-amplification of the library or low complexity libraries, this may result in PCR duplicates which due to PCR bias can create false enrichment. PCR duplicates are defined as reads with exactly the same start and end position in the genome, these reads are unlikely to exist by chance due to the mostly randomised nature of sonication. Additionally, optical duplicates may be sequenced due incorrect multiple identifications of a single cluster on the flow cell. To identify and remove both PCR and optical duplicates I used Picardtools v2.0.1 (Broad Institute 2017).

I called regions of enrichment (peaks) using MACS2 v2.1.0.2 (Zhang et al. 2008) in paired-end mode using broad peak settings with a broad-cutoff of 0.1, keeping duplicates as they had already been removed using a more sensitive method. It has been discovered that some genomic regions accumulate ChIP-Seq reads, and create false peaks (Carroll et al. 2014). These regions have been complied into a blacklist by Anshul Kundaje as part of the ENCODE project (ENCODE Project Consortium and others 2012; Kundaje 2014) and can be found at: `https://sites.google.com/site/anshulkundaje/projects/blacklists`. The blacklist was provided in Mm9 coordinates, therefore to make these compatible with my alignment files, I converted these to GRCm38 coordinates using UCSC LiftOver tools (Hinrichs et al. 2006). The peaks were then filtered to remove those overlapping blacklisted regions using BedTools v2.27.1 (Quinlan and Hall 2010). Initially, I used the R package ChIPQC v1.6.1 (Carroll et al. 2014) to determine the quality of ChIP-Seq enrichment.

Peaks were visualised using Integrative Genomics Viewer (IGV) v2.3 (Thorvaldsdóttir et al. 2013), coverage tracks normalised to counts per million total reads were generated using deepTools bamCoverage v2.5.4 (Ramírez et al. 2014; Ramírez et al. 2016). The R package ChIPSeeker v1.6.2 (Yu et al. 2015) was used to assign peaks to genes and retrieve general metrics of binding patterns in relation to genes. To plot average profiles of histone marks around the TSS, I calculated library size normalized genomic coverage using BEDTools genomecov v2.27.1 (Quinlan and Hall 2010). To get normalised coverage, I multiplied the number of reads by a scale factor, calculated by dividing 1×10^6 by the total number of properly paired mapped reads after duplicate removal.

2.10 Immunofluorescence

Cytospins were prepared as described in subsection 2.7.2. Cytospins were fixed in 4% paraformaldehyde for 10 minutes at room temperature. After fixation, cells were washed in cold PBS three times and then simultaneously permeablise and blocked by incubation for 10 minutes in blocking buffer (1× PBS /5% FBS/0.3% Triton TM X-100). Following permeabilisation, cells were washed 3 times for 5 minutes in PBS. Next, to contain a of droplet containing primary antibody, an outline was drawn around the cells using a wax pen. Primary antibody (Purified Mouse Anti-β-catenin Clone 14/Beta-Catenin, BD Transduction Laboratories™) was diluted in 1:100 with antibody dilution buffer (1× PBS/ 1% BSA/0.3% Triton TM X-100) and 200 µL was pipetted on to each slide within the wax outline. Slides were incubated overnight at 4°C in a covered box. The next day, slides were washed 3 times for 5 minutes in PBS. The secondary antibody (FITC-conjugated anti-mouse goat 4409s from Cell Signalling) was diluted at 1:200 and was added to the slides and protected from light at room temperature for 1-4 hours. Slides were washed 3 times for 5 minutes in PBS. Finally, after drying, mounting medium and cover slips were placed over the cells before viewing with a fluorescence microscope. The resulting images were taken with an attached camera and processed in the Leica Application Suite X (LAS X) by Leica™.

Chapter 3

Transcriptional profiling of β-catenin conditional cell lines: Combined Batch Analysis

3.1 Introduction

The cell populations from which LSCs can arise have been widely debated, the discussion has also been confounded by the differences in mouse and human haematopoiesis (Doulatov et al. 2012). Given their similar properties, HSCs seem the most plausible candidates for LSCs origin, but it has also been proposed that more differentiated haematopoietic stem cells such as GMPs may become LSCs (Cozzio et al. 2003; So et al. 2003; Huntly et al. 2004; Krivtsov et al. 2013; Heuser et al. 2011). The key to leukaemogenesis arising from GMPs has been suggested to be aberrant β-catenin upregulation (Yeung et al. 2010; Siriboonpiputtana et al. 2017). β-catenin is both a key regulator of cell adhesion and the terminal effector protein of Wnt signalling. Activated β-catenin translocates to the nucleus where it binds to transcription factors and by recruiting epigenetic modifiers, modulates gene expression (Mosimann et al. 2009). β-catenin is a key regulator of both normal stem cell and cancer stem function, activating genes involved in self-renewal (Lien and Fuchs 2014).

Others have shown that β-catenin is required for MLL-fusion driven leukaemia, regardless of cell-of-origin (Wang et al. 2010a). The study from Siriboonpiputtana et al. (2017), which formed the foundation of my PhD, showed that in contrast with other MLL fusions, the in vivo requirement of β-catenin for leukaemic transformation and maintenance of MLL-ENL transduced LSCs is cell-of-origin specific; Pre-LSCs derived from progenitors (GMPs) are dependent on β-catenin whilst Pre-LSCs derived from a HSC enriched population (LSKs) are β-catenin independent. RNA-Seq analysis revealed that despite the genome-wide revolution of transcription introduced by MLL-ENL, some transcriptional differences between GMPs and LSKs are maintained after MLL-ENL leukaemic transformation, including several Hox genes and *Meis1*. We hypothesised that these are indicators of self-renewal programs in LSKs which remain after transformation and compensate for the loss of β-catenin.

Similarly β-catenin, LSKs were shown not to require *HoxA9* for MLL-ENL leukaemic transformation. Following the hypothesis that *HoxA9* is responsible for conferring differential β-catenin requirement Siriboonpiputtana et al. (2017), deleted both *HoxA9* and β-catenin in tandem, resulting in abolishment of leukaemia development. To investigate potential molecular links between the effects of *HoxA9* and β-catenin deletion, we looked for patterns of gene expression showing synergistic effects; genes whose expression change following compound KO is greater than the additive effect of single gene knock-outs. Further RNA-Seq analysis revealed amongst synergistically affected genes there where several *Hoxa9/Meis1* targets including protein arginine methyltransferase 1 (*Prmt1*) responsible for methylating arginine residues in proteins including histones. The therapeutic potentials are further discussed in (Cheung et al. 2016).

Furthermore, following on from the Siriboonpiputtana et al. (2017) study which focused mainly on mechanisms of β-catenin redundancy in MLL-ENL transformed LSK (LSK^{ME}), in this chapter I study the transcriptional effects of β-catenin depletion in cell-of-origin specific MLL transformed haematopoietic cells. As our ultimate aim is to characterise the potential epigenetic effects of β-catenin depletion it is necessary that the RNA-Seq data are directly comparable with ChIP-Seq data in this study. Therefore, in this chapter, I generate new RNA-Seq data from stored cells originally used in the Siriboonpiputtana et al. (2017) study. This allowed me to not only increase the sample size and reproducibility of these data, but also obtain RNA-Seq data from the identical cells that were used for ChIP-Seq with a strict experimental design. Using linear modelling to control for batch effects, I perform differential expression analysis combining RNA-Seq from both studies, the suitability of this method is discussed.

3.2 Methods

3.2.1 RNA-Seq Generation

For this study, we chose to use existing cell lines resulting from two independent Fluorescence-activated cell sorting FACS experiments of pooled bone marrow cells transduced with the same construct of MLL-ENL and at the time of the original experiments, were shown to have cell-of-origin specific dependency on β-catenin in vivo Table 3.1. These two experiments were referred to as experiment "60" and "69" throughout Dr. Siriboonpiputtana's doctoral thesis (Siriboonpiputtana 2014) and this is how I will refer to them throughout this thesis. Each experiment resulted in two MLL-ENL transformed β-catenin conditional cell lines originating from either LSK or GMP populations. As the LSK and GMP cell lines were produced in paired, independent experiments originating from a different mixture of individuals, we may treat cell lines derived from the same sorted population as biological replicates. These replicates encapsulate the variation introduced by the sorting procedure and the RTTA, however, variation between individual mice was lost.

Briefly, as described in section 2.2, I recovered the aforementioned cell lines from liquid nitrogen storage and following 24 hours recovery in liquid culture, I removed dead cells by centrifugation with Density Gradient Centrifugation (Ficoll 400, Sigma Aldrich) for 30 minutes at 800 rcf. To generate new RNA-Seq samples I deleted β-catenin with 20 nM 4-OHT in liquid culture for 72 hours, exactly following the methods of the Siriboonpiputtana et al. (2017) study. To obtain RNA for sequencing, I pelleted approximately 3 million cells from each condition, lysed and snap froze them at -80°C overnight before extracting RNA. RNA was extracted and RNA-Seq libraries were constructed as described in subsection 2.5.2. Importantly, at the same time, I harvested cells from the same culture vessels, fixed them with formaldehyde and snap froze these at -80°C for ChIP-Seq as described in section 2.8. Performing formaldehyde fixation for ChIP within hours of collecting cells for RNA ensured that the RNA-Seq and ChIP-Seq would be conducted on cell populations which were as close to identical as possible.

3.2.2 Confirming β-catenin depletion

For the purpose of our experiment, it was important to confirm that β-catenin had been completely knocked-out in the vast majority of cells, so I measured the KO using three different methods: 1. Genotyping PCR, 2. RT-qPCR and 3. Western Blot. Firstly, I extracted genomic DNA from cell pellets using the DNeasy Blood & Tissue Kit (QIAGEN, catalogue number 69581) and performed PCR using previously designed genotyping primers (see subsection 2.1.2, Table 2.1) followed by gel electrophoresis with a 1.5% polyacrylamide gel. To further confirm a depletion in β-catenin mRNA I then produced cDNA from the same RNA used for sequencing and measured β-catenin mRNA by RT-qPCR as described in section 2.3 using previously designed RT-qPCR primers (see Table 2.3. Finally western blot was carried out as described in section 2.4.

Additionally, to confirm cell line identity, I checked the mRNA levels of genes shown to be differentially expressed in different cells of origin (*Hoxa9*) and after β-catenin KO (*Tgm2*) by RT-qPCR. To further confirm KO after sequencing, I calculated Reads Per Kilobase of transcript per Million mapped reads (RPKM) normalised coverage using deepTools v2.5.4 (Ramírez et al. 2014; Ramírez et al. 2016) and visualised RNA-Seq reads at β-catenin exons 2-6 in the Integrative genomic viewer (IGV) v2.4.19 (Thorvaldsdóttir et al. 2013).

3.2.3 RNA-Seq Analysis

In total 24 RNA-Seq samples were analysed, with at least two replicates for each condition of interest, cell-of-origin and *Ctnnb1* KO (Table 3.1). After read counts per gene and quality control metrics were obtained as detailed in subsection 2.5.3, I conducted downstream analysis in R using Bioconductor packages (R Core Team 2015; Huber et al. 2015). Quality control metrics for RNA-Seq generated in this study are shown in Table 3.2.

Table 3.1: All RNA-Seq samples produced in our lab and analysed in this chapter

SRA sample	Date*	Experiment**	Cell-of-Origin	Ctnnb1 Status***	MLL-ENL Construct****	Antibiotic
ERX994611	Sep-14	1	GMP	N/A	N/A	N/A
ERX994613	Sep-14	2	GMP	N/A	N/A	N/A
ERX994631	Sep-14	1	LSK	N/A	N/A	N/A
ERX994633	Sep-14	2	LSK	N/A	N/A	N/A
ERX994603	Feb-14	51	GMP	fl/fl	short	Blasticidin
ERX994595	Feb-14	51	GMP	-/-	short	Blasticidin
ERX994607	Feb-14	65	GMP	fl/fl	short	Blasticidin
ERX994599	Feb-14	65	GMP	-/-	short	Blasticidin
ERX994605	Feb-14	60	GMP	fl/fl	long	Puromycin
ERX994597	Feb-14	60	GMP	-/-	long	Puromycin
ERX994609	Feb-14	69	GMP	fl/fl	long	Puromycin
ERX994601	Feb-14	69	GMP	-/-	long	Puromycin
ERX994627	Apr-14	60	LSK	fl/fl	long	Puromycin
ERX994615	Apr-14	60	LSK	-/-	long	Puromycin
ERX994629	Apr-14	69	LSK	fl/fl	long	Puromycin
ERX994617	Apr-14	69	LSK	-/-	long	Puromycin
Unpublished-1	Dec-15	60	LSK	fl/fl	long	Puromycin
Unpublished-2	Dec-15	60	LSK	-/-	long	Puromycin
Unpublished-3	Dec-15	60	GMP	fl/fl	long	Puromycin
Unpublished-4	Dec-15	60	GMP	-/-	long	Puromycin
Unpublished-5	Dec-15	69	LSK	fl/fl	long	Puromycin
Unpublished-6	Dec-15	69	LSK	-/-	long	Puromycin
Unpublished-7	Dec-15	69	GMP	fl/fl	long	Puromycin
Unpublished-8	Dec-15	69	GMP	-/-	long	Puromycin

* I produced the samples from Dec-2015, all others were published in (Siriboonpiputtana et al. 2017).
** Experiment refers to retroviral transduction and/or FACS experiment.
*** *Ctnnb1* Status; "N/A" means wild-type; fl/fl (floxed) indicates floxed alleles; -/- indicates double deletion/knock-out.
**** "NA" means not transduced; "short" refers to the MLL-ENL construct containing the minimum transformation element; "long" refers to the full length MLL-ENL fusion.

Table 3.2: Quality Control metrics for 2015 RNA-Seq samples

Sample	Total read pairs	Proper pairs	% uniquely mapped	% Genic	% Intergenic	% Ribosomal
Unpublished-1	21,802,759	13272709	74.10%	86.9%	2.31%	0.98%
Unpublished-2	24,589,580	15301022	77.7%	92.5%	1.60%	1.24%
Unpublished-3	21,482,841	16710697	79.8%	91.0%	1.72%	0.89%
Unpublished-4	24,385,310	17978289	75.3%	93.6%	1.43%	0.95%
Unpublished-5	24,941,668	18999975	73.8%	91.0%	1.62%	0.66%
Unpublished-6	25,023,990	15034344	69.4%	91.4%	2.13%	1.30%
Unpublished-7	18,761,283	15840581	74.4%	90.5%	1.86%	0.51%
Unpublished-8	20,230,699	16183220	79.1%	91.3%	1.74%	0.51%

Differential expression analysis for the effects of cell-of-origin and *Ctnnb1* KO in both LSK^{ME} and MLL-ENL transformed GMP (GMP^{ME}) was performed using EdgeR v3.28.0 (Robinson et al. 2010; McCarthy et al. 2012; Zhou et al. 2014). First, genes were filtered for those which were expressed with a Count Per Million (CPM) value greater than one in more than two samples. To allow visualisation of the raw data, the variance stabilising transformation function from DESeq2 v1.22.2 (Love et al. 2014) was applied to library size normalised counts. Next, sources of variation in these data were assessed by Principal Component Analysis (PCA) of the top 1000 most variable genes. Subsequently, generalised linear models were fitted on library size normalised counts with design formulas to account for known batch effects discovered in PCA and described in the results using the glmFit function (EdgeR). Finally, differential expression was assessed using the glmLRT function (EdgeR) which applies a Log Ratio Test (LRT). P.values were corrected for multiple testing using the Benjamini-Hochberg (BH) method with the TopTags function (EdgeR) and differentially expressed genes were taken with an False Discovery Rate (FDR) cut-off of less than 0.05.

Although modelling for batch effects allows us to adjust for batches when making statistical inferences, it does so without modifying the data, therefore batch effects are still visible if expression values are plotted. To assess the effectiveness of modelling known batch effects, I used the removeBatchEffect function from limma v3.42.0 (Ritchie et al. 2015) on variance stabilised counts. The removeBatchEffect function effectively works by modelling the given design including batches and conditions of interest, it then subtracts the specified covariate effect from these data. Following batch correction, the results of each model were visualised in PCA, heatmaps and smearplots using functions from pheatmap v1.0.12 and ggplot2 v3.1.1 (Kolde 2019; Wickham 2016).

To elucidate whether the transcriptional response to β-catenin deletion was specific to LSK^{ME} or GMP^{ME} I calculated the Pearson's correlation using the cor function and a hypergeometric test with the phyper function in R. For the hypergeometric test, the size of the gene universe was the number of expressed genes (those with a CPM value of 1 or greater) in this study. To investigate if there were any patterns of concerted expression changes among groups of genes with similar functions, I performed hypergeometric test based functional profiling with the bioconductor packages BiomaRt v2.42.0 (Durinck et al. 2009) and clusterProfiler v3.14.3 (Yu et al. 2012).

3.3 Results

3.3.1 Validation of β-catenin knock-out

To evaluate the global transcriptional targets of β-catenin, it was vital that complete β-catenin KO was achieved. Therefore, I used three different methods to confirm complete functional KO by excision of β-catenin exons 2-6. Firstly, genotyping PCR and gel electrophoresis (Figure 3.1A) showed that in 4-OHT treated samples, only the deleted band at 500 bp is present, confirming homozygous genetic knock-out. Notably, gel electrophoresis reveals β-catenin deleted bands are also present in all untreated cell lines (Figure 3.1A). This indicates that in the absence of 4-OHT treatment, Rosa-Cre is spuriously activated leading to excision of β-catenin. Next, the absence of β-catenin protein following KO was confirmed by western blot (Figure 3.1B). Additionally, I confirmed mRNA depletion of β-catenin with RT-qPCR (Figure 3.1C).

Transglutaminase 2 (*Tgm2*) was identified to be one of the most significant down-regulated genes after β-catenin deletion regardless of cell-of-origin (Siriboonpiputtana et al. 2017). To further prove the effects of β-catenin KO, I showed significant downregulation of *Tgm2* transcription by RT-qPCR (Figure 3.1E). As mentioned in section 3.1, we previously identified *Hoxa9* as being more highly expressed in LSK than GMP-derived MLL-ENL transformed cells. Consistently I observed higher *Hoxa9* expression in the LSK^{ME} cell lines, regardless of β-catenin status, confirming their cell-of-origin identity (Figure 3.1D). Finally, RNA-Seq is a very sensitive method to detect even low levels of remaining intact β-catenin expression, but we could not observe any reads on exons 2-6 (Figure 3.1F). In conclusion, all methods concur that we observe a very-near complete functional KO of β-catenin in these cells after 4-OHT treatment.

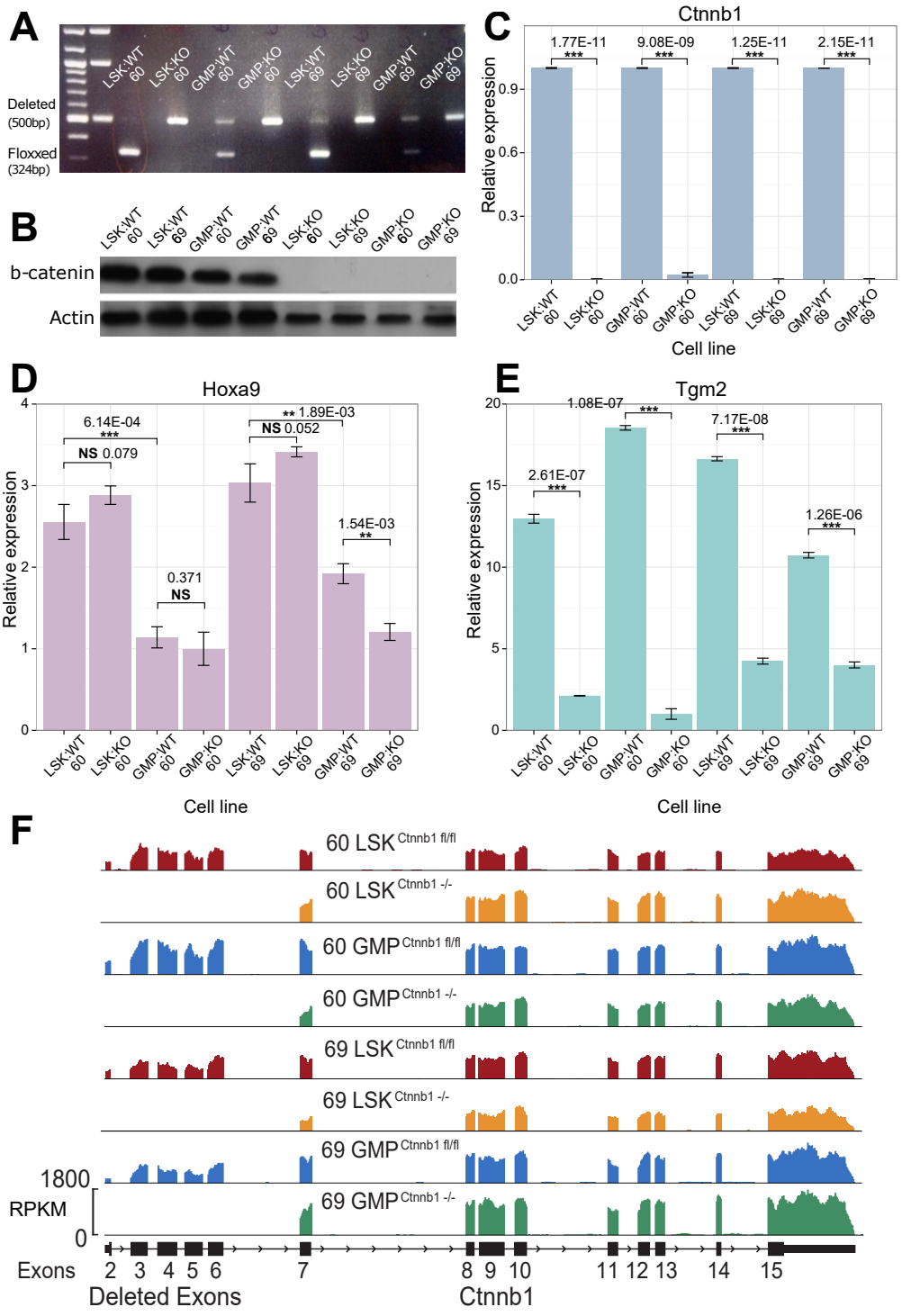

Figure 3.1: Validation of RNA-Seq samples A) Genotyping PCR for β-catenin. Floxed bands are present at 324 bp and deleted are at 500 bp. B) Western blot for β-catenin and Actin. C) RT-qPCR for β-catenin expression. Values are normalised to *Gapdh* expression and are shown relative to the floxed sample for each line. D-E) RT-qPCR for *Hoxa9* and *Tgm2* before and after β-catenin knock. out. F) IGV profile showing KO of exons 2-6.

3.3.2 Principal Component Analysis reveals batch effects and bias

Aiming to understand the sources of variance in these data, I visualised each experimental factor in a series of Principal Component Analysis (PCA) plots. PCA plots show the relationship between samples by reducing data into dimensions (principal components) with maximal variation, revealing the overall similarities and differences in global gene expression between samples. Upon MLL-ENL transformation, PCA reveals that gene expression between samples from different cells-of-origin become more similar and transformed cells separate altogether from their normal cells-of-origin (Figure 3.2A, PC1: 48% of variance). In line with previously published data (Lara-Astiaso et al. 2014), the expression profiles of normal LSK and GMP separate clearly from each other in PCA (Figure 3.2A, PC2: 21% of variance). Observing only transformed cells in Figure 3.2B, contrary to expectation, samples do not clearly separate by cell-of-origin, or treatment. Instead, there is a strong clustering effect from experiment, with PC1 and PC2 separating the samples into 4 clusters, each containing samples from a distinct experiment (Figure 3.2B).

Isolating only the data used in the Siriboonpiputtana et al. (2017) study, samples separate by virus used (Full length MLL-ENL with puromycin or the minimal transformation element of MLL-ENL with blasticidin) explaining 39% of the variance (Figure 3.2C, PC1). Notably, LSK^{ME} replicates in the 2014 study have very little variance compared to GMP^{ME}s. Comparing this with freshly generated sequencing (Figure 3.2D) we can see that LSK^{ME} cell lines display a similar in-group variation to GMP^{ME}, clustering by cell line rather than treatment or cell-of-origin. I conclude that the strongest determinant of global gene expression is cell-of-origin, followed by introduction of the fusion. Treatment and cell-of-origin in transformed cells have modest influence on variance, and there are effects from several other factors, including batch effects and the transformation element/ antibody used. It is therefore important to model these additional factors in differential expression analysis.

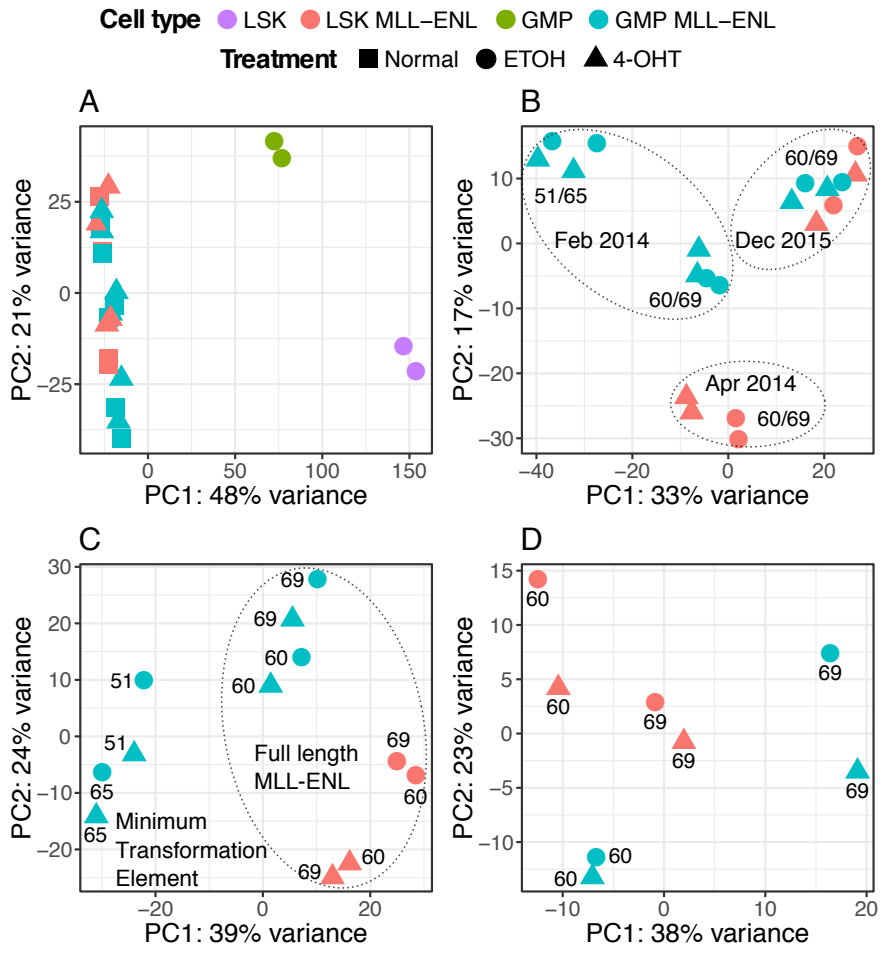

Figure 3.2: Experimental variables and batch effects separate samples in principal component analysis (PCA) A) All RNA-Seq samples in this study including normal cell compartments. B) All MLL-ENL transformed samples labelled with experiment reference. Samples with the same sequencing date are labelled and contained within an ellipse. C) Samples included in Siriboonpiputtana et al. (2017). An ellipse indicates those samples with the full length MLL-ENL fusion. Samples outside the ellipse utilise the minimum transformation element. D) Samples independently sequenced in 2015.

3.3.3 Batch Correction for the effect of cell-of-origin

As identified in the previous subsection 3.3.2, there are multiple known covariates which confound the effect of cell-of-origin. These covariates are, as shown in Table 3.1: Date, Experimental Reference, MLL-ENL construct and Antibiotic. In a perfectly balanced experiment, we would attempt model all of these covariates. Unfortunately, some of these covariates are perfectly confounded and therefore cannot be modelled separately. In order to attempt correction for any of the observed batch effects, we must remove all samples utilising the minimum transformation element or Blasticidin as this covariate is unbalanced.

In the design models I will refer to the cell-of-origin effect as "celltype". Differential expression analysis for the effect of cell-of-origin using only cell-of-origin in the design model (*~celltype*) yields few differentially expressed genes as samples are poorly separated (Figure 3.3A, 58 up and 44 down). Although it has been shown that sequencing runs on the same machine does not have a large impact on differential expression analysis, separate RNA-Seq library processing of RNA has been shown to significantly affect differential expression analyses (Peixoto et al. 2015; Danielsson et al. 2015). In Table 3.1, we can see that all 2014 LSK^{ME} samples were sequenced in April 2014 whilst 2014 GMP^{ME} samples were sequenced in February of 2014. It is not known if the 2014 samples were also processed separately. As "date" perfectly confounds cell-of-origin we are unable to model "date" and must assume that the effect is much smaller than cell-of-origin. However, it is certain that the 2015 samples were processed separately so we may attempt to model for this effect using sequencing "year" instead. Modelling for *~year+celltype* does not result in clear separation by cell-of-origin however significantly increases differentially expressed gene detection (Figure 3.3B, 152 up and 163 down). Next, we may attempt to model for experiment, which combines the variation introduced by both FACS experiment and transduction experiment (Table 3.1). Modelling for *~experiment+celltype* does not result in clear separation by cell-of-origin however mildly increases differentially expressed gene detection over the minimal model (Figure 3.3C, 77 up and 56 down).

We may also attempt to correct for both year and experiment by adding both of these covariates to the design model. Modelling for *~year+experiment+celltype* only somewhat improves clustering in PCA and hierarchical clustering but significantly increases differentially expressed gene detection (Figure 3.3D, 224 up and 237 down). Modelling for year and experiment as an interaction (*~year*experiment+celltype*), gives similar results to the additive model and does not improve clustering (Figure 3.3E, 233 up and 232 down). In conclusion, modelling unwanted covariates for the cell-of-origin improved differential expression detection. However, it is unclear if the batch effects have been removed effectively as separation is still incomplete.

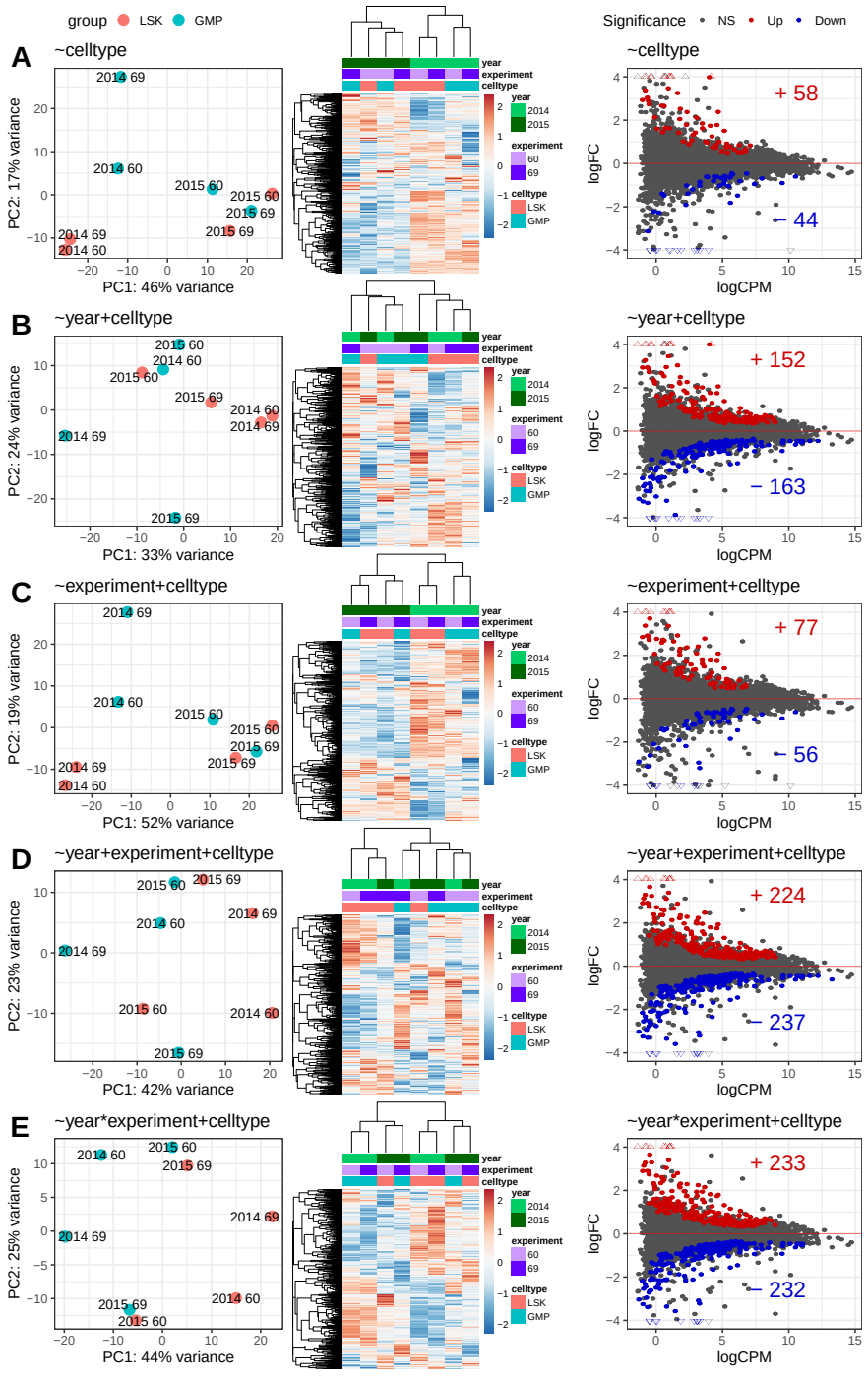

Figure 3.3: Batch correction to extract the cell-of-origin effect A-E) Differential expression analysis for the effect of cell-of-origin using different models to account for unwanted covariates. The design model for each row is specified above the plots in the left and right columns. For each row, all effects in the indicated model with the exception of "celltype" were removed. Left column: PCA plots for the 1000 most variable genes, cell-of-origin is indicated by colour specified by the key (top). Points are labelled with their sequencing year and experimental reference. Middle column: Hierarchical clustered heatmaps for the 1000 most variable genes. Groups are specified in the key (right). Right column: Smear plots indicating the results of differential expression analysis. The number of significantly (FDR < 0.05) upregulated or downregulated genes are indicated in red and blue respectively.

3.3.4 Biological themes related to the combined cell-of-origin signature

For the purpose of understanding cell-of-origin related genes that remain differentially expressed between LSK^{ME} and GMP^{ME}, I performed pathway analysis on the differentially expressed genes obtained by modelling for year and experiment (*~year+experiment+celltype*). Figure 3.4 illustrates Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enriched in the combined cell-of-origin signature. Most of the differentially expressed pathways between LSK^{ME} and GMP^{ME} are upregulated in GMP^{ME}. Amongst these, there are cancer and cell cycle related pathways including the Wnt signalling pathway which are upregulated in GMP^{ME} (Figure 3.4A). Downregulated pathways in GMP^{ME} relate to metabolism (Figure 3.4B).

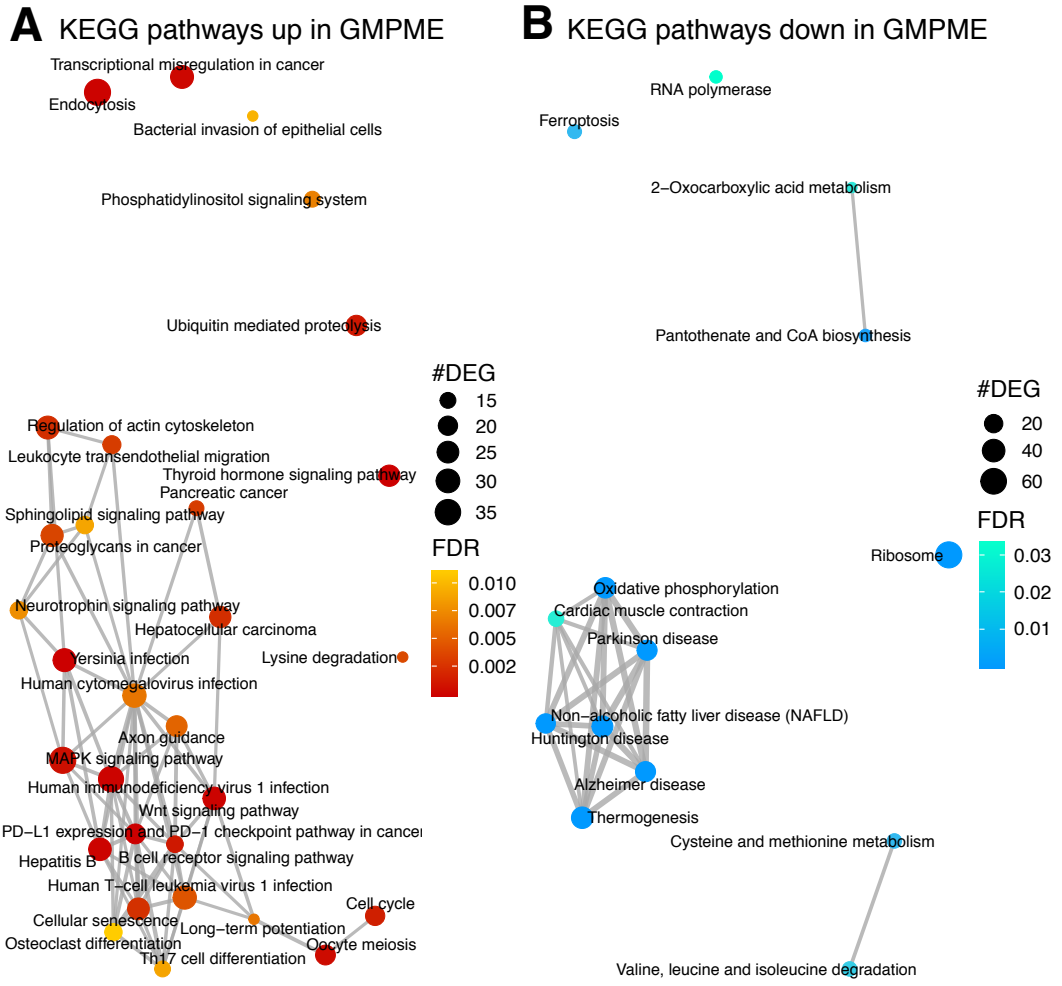

Figure 3.4: Biological themes of cell-of-origin signature Networks between KEGG pathways enriched for significantly up-regulated (A) and down-regulated (B) genes in GMP^{ME} compared to LSK^{ME}. The size of nodes represents the number of differentially expressed genes, indicated by the key (right) within the labelled pathway. The thickness of the edges between nodes denotes the degree of similarity by shared genes.

3.3.5 Batch Correction for the effect of β-catenin deletion in GMP^{ME}

Similarly to subsection 3.3.3, in this section I attempted to model for unwanted covariates to extract the effect of β-catenin deletion in GMP^{ME}. In the design models, I refer to the β-catenin deletion effect as "genotype". Without any batch effect correction (Figure 3.5A, *~genotype*), it is clear that samples separate by both year (PC1, 50%) and experiment (PC2, 24%). Addition of year to the design model does not improve clustering in PCA or hierarchical clustering and detects few differentially expressed genes (Figure 3.5B, *~year+genotype*). Similarly, addition of experiment alone does not improve clustering or differential gene expression detection (Figure 3.5C, *~year+genotype*).

Modelling for both year and experiment together shows significant improvement in PCA clustering and detection of differentially expressed genes (Figure 3.5D, PC2 21%, 24 up and 21 down). However, samples do not separate by genotype in hierarchical clustering (Figure 3.5D, middle). Modelling year and experiment as interacting terms finally separates samples clearly in PCA and hierarchical clustering with a considerable increase in differential gene expression detection (Figure 3.5E).

3.3.6 Batch Correction for the effect of β-catenin deletion in LSK^{ME}

In this section I attempt to model for unwanted covariates to extract the effect of β-catenin deletion in LSK^{ME}. Without any batch effect correction, samples separate clearly by year (PC1, 63%), but not by experiment (Figure 3.6A, *~genotype*). Unlike GMP^{ME}, addition of year alone to design appears to improve clustering in PCA, and has considerable effect on the number of differentially expressed genes (Figure 3.6B, *~year+genotype*, 208 up and 71 down). On the other hand, addition of experiment alone appears to improve clustering in PCA but not hierarchical clustering and has little effect on the number of differentially expressed genes (Figure 3.6C, *~experiment+genotype*, 37 up and 25 down).

Modelling for both year and experiment as additive effects allows clear separation in PCA (Figure 3.6D *~year+experiment+genotype*, PC1, 61%) and hierarchical clustering and continues to increase the number of differentially expressed genes (336 up and 116 down). However, as with GMP^{ME}, modelling for the interaction between year and experiments yields the clearest separation in clustering and the greatest number of differentially expressed genes (Figure 3.6E, 449 up and 158 down).

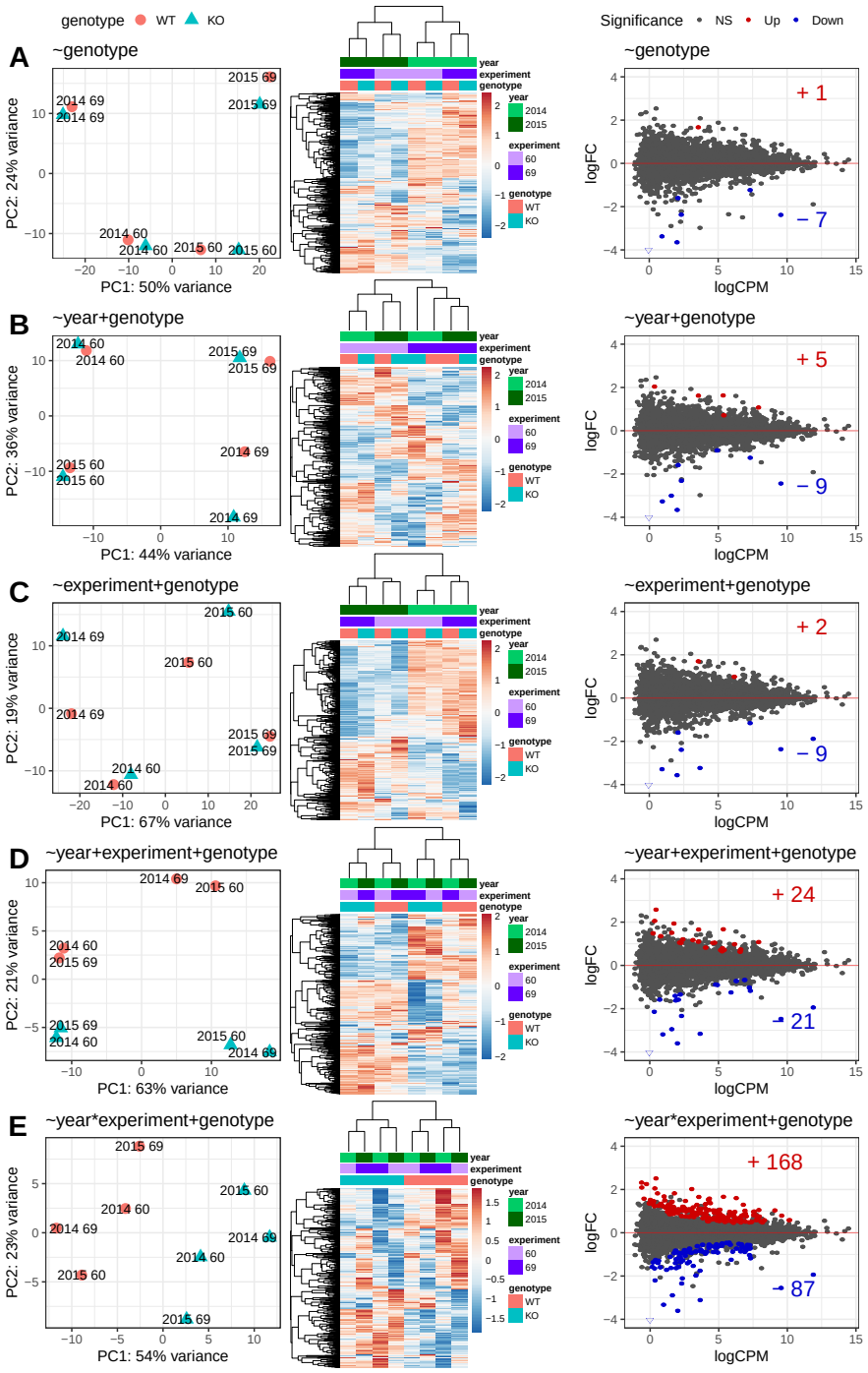

Figure 3.5: Batch correction for loss of β-catenin in GMP^{ME} A-E) Differential expression analysis for the effect of β-catenin loss in GMP^{ME} using different models to account for unwanted covariates. The design model for each row is specified above the plots in the left and right columns. For each row, with the exception of genotype all effects in the indicated model were removed. Left: PCA plots for the 1000 most variable genes, genotype is indicated by colour specified by the key (top). Middle: Hierarchical clustered heatmaps for the 1000 most variable genes. Groups are specified in the key (right). Right: Smear plot indicating the results of differential expression analysis. The number of significantly (FDR < 0.05) upregulated or downregulated genes are indicated in red and blue respectively.

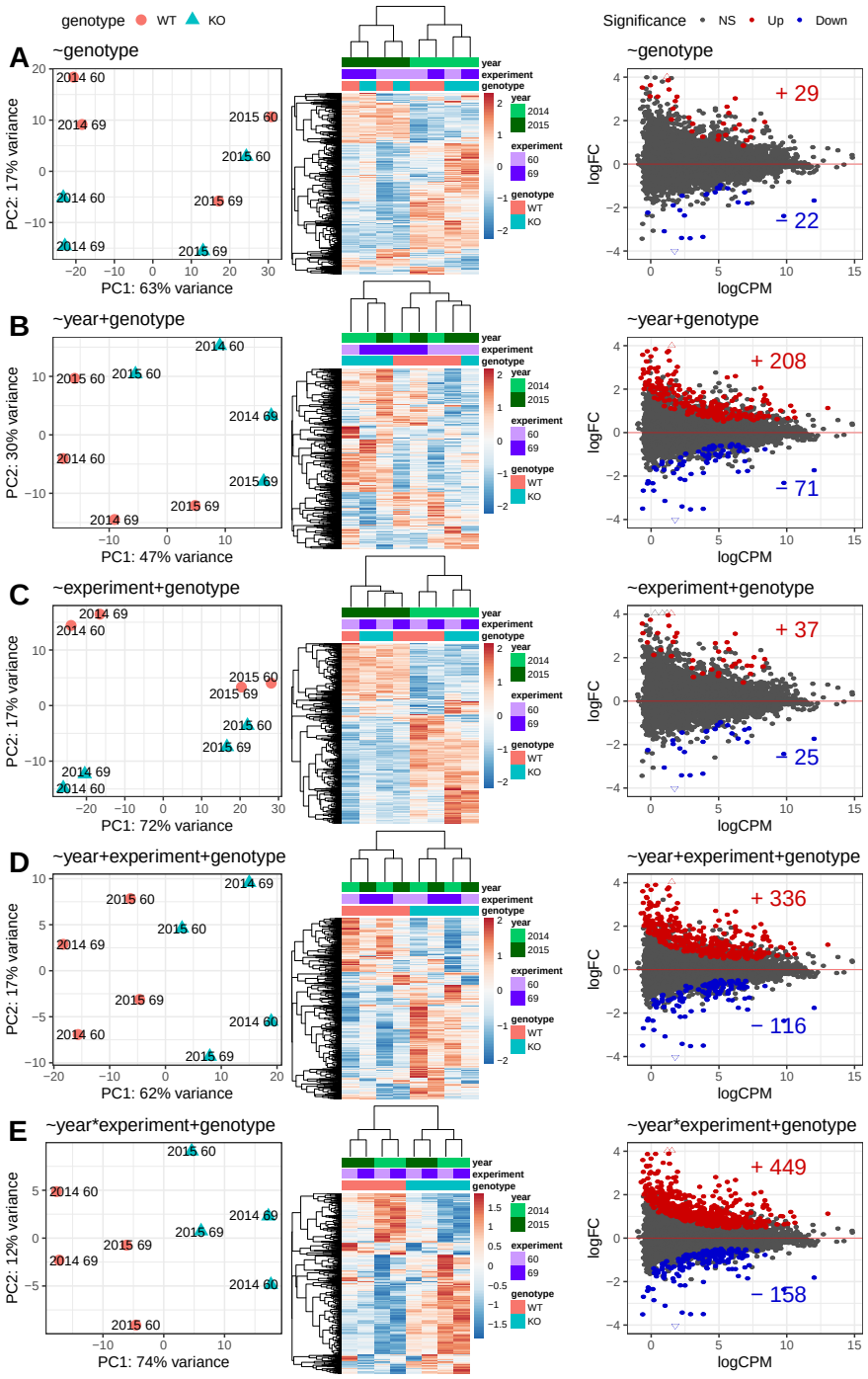

Figure 3.6: Batch correction for loss of β-catenin in LSK^{ME} A-E) Differential expression analysis for the effect of β-catenin loss in LSK^{ME} using different models to account for unwanted covariates. The design model for each row is specified above the plots in the left and right columns. For each row, with the exception of genotype all effects in the indicated model were removed. Left: PCA plots for the 1000 most variable genes, genotype is indicated by colour specified by the key (top). Middle: Hierarchical clustered heatmaps for the 1000 most variable genes. Groups are specified in the key (right). Right: Smear plot indicating the results of differential expression analysis. The number of significantly (FDR < 0.05) upregulated or downregulated genes are indicated in red and blue respectively.

3.3.7 Putative β-catenin target genes in MLL-ENL pre-LSCs are largely independent of cell-of-origin

β-catenin is a multi-functional protein with context dependent roles in cell function (Valenta et al. 2012), so we wanted to determine if β-catenin downstream targets were the same, or different in LSK-origin and GMP-origin cells. Immediately, we observe that the majority of deregulated genes following β-catenin KO are upregulated (Figure 3.7A and previously in Figure 3.5D & 5.6D). As would be expected, the overlaps between β-catenin affected genes in GMP^{ME} and LSK^{ME} are statistically significant as determined by hyper-geometric analysis (Figure 3.7D). As there is great variation between replicate experiments, it would be quite arbitrary to use FDR cut-offs to determine cell-of-origin unique β-catenin targets. Linear regression for the log fold change of all genes after β-catenin deletion in LSK^{ME} and GMP^{ME} shows a strongly positive correlation (Figure 3.7B, R2=0.444).

The similarities and variations in LSK^{ME} and GMP^{ME} expression in response to β-catenin deletion are illustrated further in a heatmap (Figure 3.7C). By colour coding the genes into common, and "cell-of-origin unique" signature genes we can see that despite not reaching significance cut-offs in the opposing group, the trend of gene expression of "unique" targets is remarkably similar in LSK^{ME} and GMP^{ME}. Furthermore, hierarchical clustering of the signature genes shown in Figure 3.7C showed clear separation of upregulated and downregulated genes despite across-group variation. I conclude that these data provide little evidence to support the hypothesis that β-catenin targets different genes in LSK and GMP derived MLL-ENL Pre-LSCs.

3.3.8 Biological themes related to the combined β-catenin knock-out signature

In order to understand the biological effects of β-catenin deletion, I performed Gene Ontology (Gene Ontology (GO)) analysis for GMP^{ME} and LSK^{ME} targets, the results are shown in the form of networks in Figure 3.8. From this analysis it was found that the majority of enriched GO terms are shared between LSK^{ME} and GMP^{ME} upregulated signatures (Figure 3.8A). These positively enriched GO terms are related to innate immune response and regulation of cell adhesion. However, both LSK^{ME} and GMP^{ME} downregulated signatures are enriched for unique GO terms related to immune response (Figure 3.8B). Additionally, downregulated genes in LSK^{ME} exhibit unique enrichment for negative regulation of the cell cycle (Figure 3.8B).

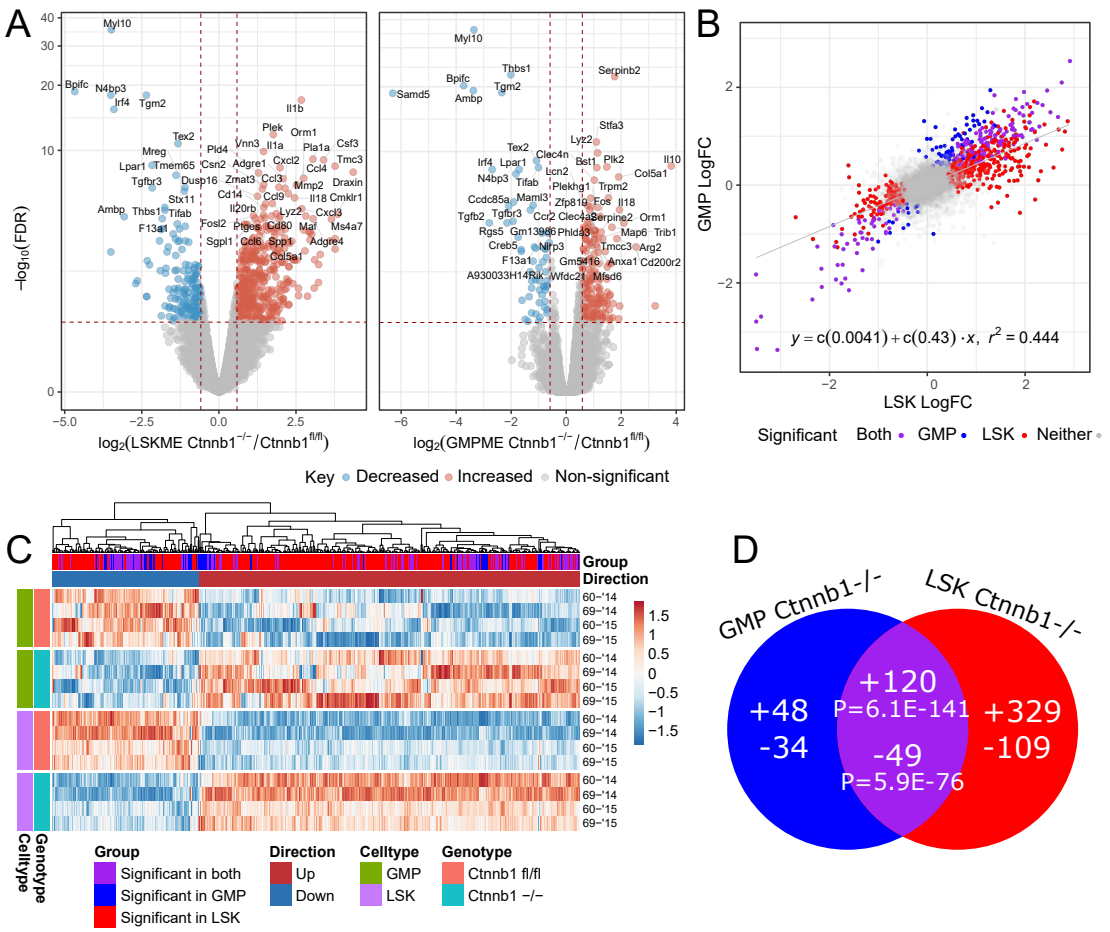

Figure 3.7: Most of the transcriptional effects of β-catenin deletion are independent of cell-of-origin Differential expression analysis was performed with the design model: *~year*experiment+genotype*. A) Volcano plots show the distribution of differentially expressed genes after β-catenin KO in LSK^{ME} (left) and GMP^{ME} (right). The top 50 significant genes are labelled with their gene symbol. The colour of the points reflects the direction of change and significance (FDR <0.05). B) Scatter plot to visualise the correlation between log fold changes of gene expression in LSK^{ME} β-catenin KO (x-axis) versus GMP^{ME} β-catenin KO (y-axis). The plot is labelled with the Pearson's correlation. Points are coloured by the group in which they are significant as indicated by the key (bottom). C) Heatmap visualising all genes deregulated after β-catenin KO in GMP^{ME} or LSK^{ME}. Genes (columns) are ordered by hierarchical clustering. The dendrogram (top) indicates the relationships between clusters. The direction of gene expression and the comparison in which they are significant in is indicated by colour bars at the top and are explained by the key below. The rows are grouped by cell type of origin and genotype as indicated by the colour bars on the left and the key below. D) Venn diagram showing overlaps between differentially expressed genes after β-catenin KO LSK^{ME} and GMP^{ME}. "+" and "-" indicates upregulated and downregulated genes respectively. P.Values are as labelled and calculated by hypergeometric test.

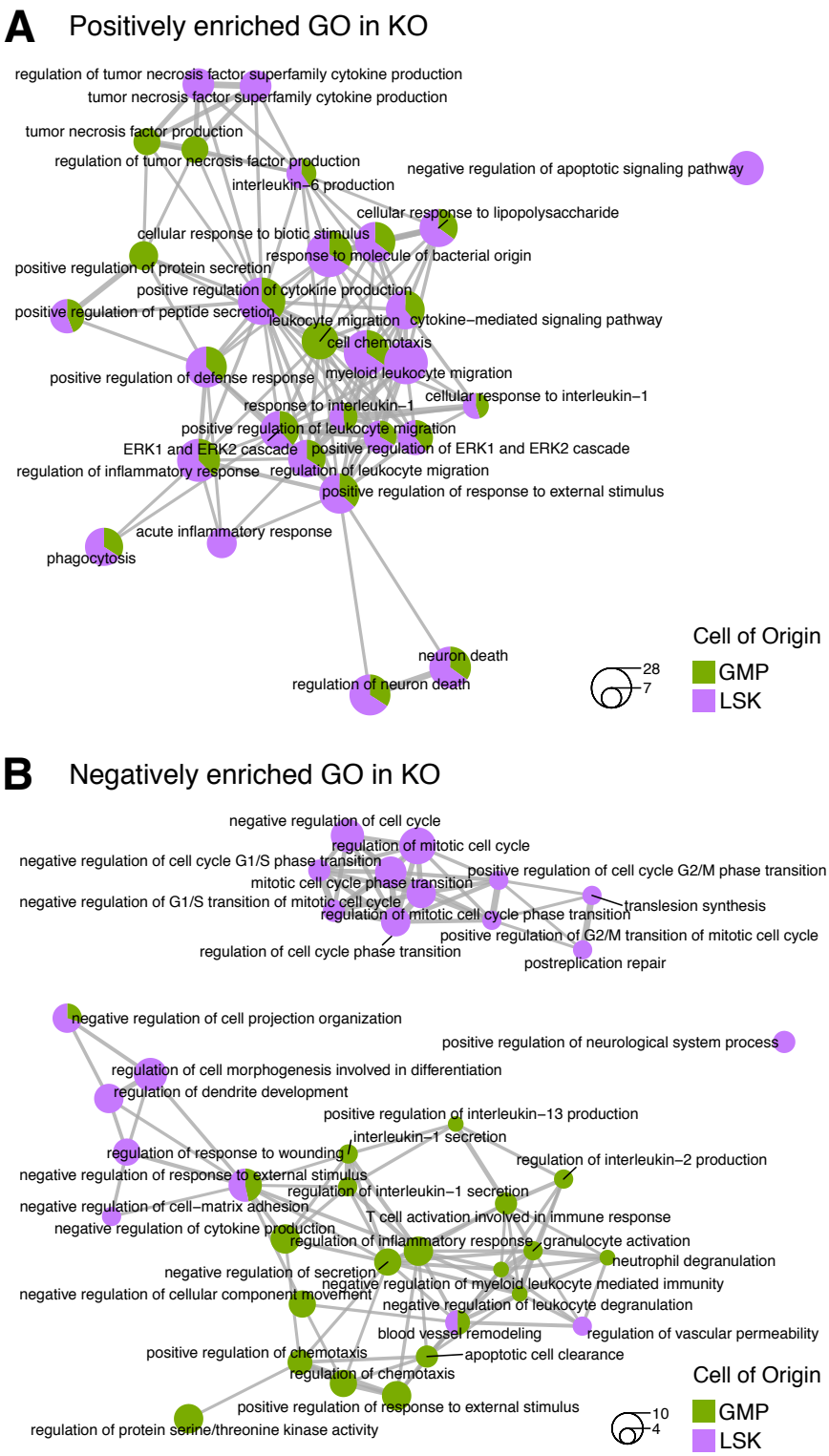

Figure 3.8: GO enrichment for genes deregulated after β-catenin knock-out Differential expression analysis was performed with the design model: *~year*experiment+genotype*. Networks between KEGG pathways enriched for significantly upregulated (A) and regulated (B) genes in GMP^{ME} compared to LSK^{ME}. The size of nodes represents the number of differentially expressed genes, indicated by the key (bottom right) within the labelled pathway. The thickness of the edges between nodes denotes the degree of similarity by shared genes. Each node is coloured like a pie chart, indicating the proportion of significant genes belonging to the LSK^{ME} or the GMP^{ME} signature (purple and green respectively, as indicated by the key labelled "Cell of Origin", bottom right)

3.4 Discussion

In this chapter I produced high quality RNA-Seq data (Table 3.2) which is directly compatible with the ChIP-Seq data required to address our eventual aim to understand the epigenetic modifications mediated by β-catenin in MLL-ENL associated AML. Using multiple methods, I showed that good β-catenin genetic KO was achieved (Figure 3.1). However, comparing all RNA-Seq samples revealed batch effects and biases (Figure 3.2). There was additional variation seen in the 2015 dataset, especially between LSK^{ME} cells which could have arisen from multiple technical and biological sources such as heterogeneity of the cell culture and even stochastic variation. RNA-Seq experiments are often plagued with batch effects due to experimental design choices which can greatly confound the biological effects of the condition of interest (Conesa et al. 2016).

First, I attempted to model for the effect of year and cell line to extract a cell-of-origin signature from the combined dataset. Although greater differentially expressed genes were detected when modelling for these effects, it is unclear if batch effects were modelled effectively (Figure 3.3B). The number of differentially expressed genes is markedly reduced when compared to Siriboonpiputtana et al. (2017), this is likely due to the increases in variation seen in LSK^{ME} in the 2015 study (Figure 3.2D). Upregulation of cancer pathways and cell cycle genes in GMP^{ME} appears contrary to expectation (Figure 3.4A), as LSK derived pre-LSCs are known to be more aggressive (Wang et al. 2010a; Siriboonpiputtana et al. 2017). However, contradictory pathways related to senescence, differentiation and immune response are also enriched (Figure 3.4A).

Examining our potential β-catenin targets, we find the majority of these are upregulated after KO (Figure 3.5, 3.6 & 3.7A). As β-catenin is mostly known for positively regulating gene expression, therefore we may assume that the upregulation of many of these genes are downstream events. On the other hand, there is some evidence that β-catenin can also recruit transcriptional repressors (Bauer et al. 2000; Weiske et al. 2007; Song et al. 2009; Liu et al. 2008). However, there appears to be greater variance of deregulation amongst upregulated genes as can be seen in Figure 3.7C which could suggest that this cluster contains more non-specific effects.

Trying to understand if β-catenin KO yields different downstream effects in LSK^{ME} and GMP^{ME}, I compared the transcriptional effects of β-catenin KO in LSK^{ME} and GMP^{ME} utilising all available RNA-Seq data. Targets were shown to be mostly the same between LSK^{ME} and GMP^{ME} with a strong positive correlation/pattern whether or not genes reached significance cut-offs (Figure 3.7B & 3.7C) leading to the conclusion that β-catenin targets of LSK^{ME} and GMP^{ME} are largely equivalent. In a similar vein, GO analysis revealed terms related to immune response and cell adhesion which were mostly overlapping in LSK^{ME} and GMP^{ME} signatures (Figure 3.8).

When performing differential expression analysis it is typically advised to fit the design model on RNA-Seq samples from all groups together. The reason for this is that a dispersion parameter is estimated for each individual gene. This provides us with deviation from the mean that is to be expected given the variability within groups (Love et al. 2014). Generally, the more data we have, the more likely that this dispersion parameter will be accurate, decreasing the rate of false positive discovery and increasing true positive discovery. However, for whatever reason, the within-group variability may be much higher for some groups. If this is the case, sub-setting the dataset into groups may be more sensitive. For example, LSK^{ME}s from 2015 have greater in-group variability than the LSK^{ME}s from 2014. This means that the per-gene dispersion estimate would be higher in the 2015 group than the 2014 group.

Another issue when combining the two datasets is that it changes our original concept of biological replication. Originally, our biological replicates could be used to estimate the combined variability between different individuals and independent MLL-ENL transduction and transformation. If we compare the 2015 and 2014 studies, these would be known as "isogenic replicates" (ENCODE Project Consortium 2019). Although they are not truly biologically distinct, the samples were grown separately with two separate deletions. For the purpose of this chapter, I merged the two data sets into one, however when comparing these data with its corresponding ChIP-Seq they must be separated. In the next chapter, I will take a deeper look into the individual batch experiments and compare their results.

Chapter 4

Transcriptional profiling of β-catenin conditional cell lines: Separate Batch Analysis

4.1 Introduction

Given the difficulty of batch correction demonstrated in chapter 3 and conflicting results with the published dataset (Siriboonpiputtana et al. 2017), we continued to examine these datasets further with separate analysis and comparison. The first experimental aim of this chapter is to try to replicate the cell-of-origin signature from (Siriboonpiputtana et al. 2017). In particular for this study, it is necessary for the effects of β-catenin depletion to be equivalent. Therefore, the next aim of this chapter is to understand if it is possible to observe similar β-catenin KO signatures in both existing and newly generated data sets. Finally, as on first inspection the effects of β-catenin KO do not show deregulation of classical β-catenin transcriptional targets, I aim to gain deeper insight into the specificity of transcriptional effects of β-catenin depletion in this study compared to others using the same mouse model.

4.2 Methods

4.2.1 Differential expression analysis

After read counts per gene and quality control metrics were obtained as detailed in subsection 2.5.3, I conducted downstream analysis in R using Bioconductor packages (R Core Team 2015; Huber et al. 2015). The differential expression methods and design models used are explicitly stated in the figure legends and main body text of the results section 4.3. Initially, to allow accurate comparison with the published cell-of-origin signature (Siriboonpiputtana et al. 2017), DESeq2 v1.22.2 with Wald statistics (Love et al. 2014) was used to perform differential expression analysis between LSK^{ME} and GMP^{ME}. DESeq2 assumes a negative binomial distribution and log fold changes are shrunken using empirical Bayes statistics. Following the comparison of 2014 and 2015 datasets differential expression for the cell-of-origin signature was repeated using EdgeR v3.28.0 (Robinson et al. 2010; McCarthy et al. 2012; Zhou et al. 2014). All further analysis of the cell-of-origin signature was done with the differential expression results obtained with EdgeR using the design model: $\sim$*experiment+celltype*.

Differential expression analysis for the effects of *Ctnnb1* KO in both LSK^{ME} and GMP^{ME} was performed using EdgeR. First, genes were filtered for those which were expressed in more than 2 samples. Then, differential expression analysis was assessed using an LRT with a blocked design to account for batch effects ($\sim$*experiment+genotype*). To investigate if there were any patterns of concerted expression changes among groups of genes with similar functions, I performed hypergeometric test based functional profiling and Gene Set Enrichment Analysis (GSEA) with the bioconductor packages biomaRt v2.42.0 (Durinck et al. 2009) and clusterProfiler v3.14.3 (Yu et al. 2012). Additional plots were generated using pheatmap v1.0.12 and ggplot2 v3.1.1 (Kolde 2019; Wickham 2016).

4.2.2 Comparison of targets with published β-catenin knock-out expression data

To shortlist genes whose expression is likely to change as a result of β-catenin knock-out, I gathered all publicly available β-catenin KO data I could identify originating from the same mouse model listed in Table 4.1, regardless of which cell-type the experiment had occurred in. For each of the datasets, where possible, differential expression results were taken from the supplementary tables of their original manuscript. Those datasets without reported signatures (GSE49248 and GSE7430) were analysed using the GEO2R online tool (NCBI 2019) using default settings.

To compare published transcriptomic signatures associated with β-catenin KO with our β-catenin KO signatures, I applied GSEA analysis with the GSEA function from clusterProfiler v3.14.3 (Yu et al. 2012) with 10000 permutations. Genes were ranked by their signed P-value, this was calculated by transforming the log10 P.value by the sign of the log2 Fold Change. It has been noted that for the GSEA algorithm, normalisation is not accurate for very small (smaller than 15) or very large (greater than 500) gene sets (Subramanian et al. 2005). Therefore, to make comparisons with published datasets fair, the top 250 most significantly upregulated genes and top 250 downregulated genes were used to create custom gene sets for GSEA.

Table 4.1: Published β-catenin deletion transcriptomic data These studies utilised the same β-catenin mouse model and were used for comparison with our β-catenin KO signature

GEO ID	Tissue	Platform	Citation
GSE9629	Embryonic kidneys	Affymetrix Mouse Genome 430 2.0 Array	(Bridgewater et al. 2008)
GSE49248	MLL-AF9 or MLLAF9/ KrasG12D AML	Affymetrix Mouse Genome 430 2.0 Array	(Ng et al. 2014)
GSE8818	Intestinal crypts	Affymetrix Mouse Genome 430 2.0 Array	(Fevr et al. 2007)
GSE67463	Embyronic gonads	Affymetrix Mouse Genome 430 2.0 Array	(Nicol and Yao 2015)
GSE7430	Embryonic pancreas	Affymetrix Mouse Genome 430 2.0 Array	(Wells et al. 2007)
GSE50620	Hematopoietic stem cells	Affymetrix Mouse Gene 1.0 ST Array	(Lento et al. 2014)

4.2.3 Comparison of targets with other studies utilising Rosa-Cre-ER

To evaluate if the β-catenin targets found in this study are specific or could be explained by a treatment-specific effect, I gathered RNA-Seq data from within our group which utilized the Rosa-Cre-ER system in vitro (listed in Table 4.2). The significance of overlaps between studies was calculated using both a permutation test and a hypergeometric test. Both hypergeometric tests and permutation tests rely on establishing the possible gene universe. A gene universe was created by simply taking those genes which had a CPM value of at least one in at least two samples. Hypergeometric tests of gene list overlaps were computed using the phyper function in R.

A permutation test was performed by creating a function in R to simulate re-sampling. On each iteration, a random sample of the length of each signature to be compared was taken from the gene universe using the sample function and the overlap of these samples was calculated with the intersect function. Over 10000 iterations, the number of instances in which the overlap of the random samples was equal to or exceeded the actual observed overlap of the signatures was calculated. This number was then divided by the total number of iterations (10000) to calculate a P.Value.

Table 4.2: In-house 4-OHT treatment transcriptomic data Studies which were compared with β-catenin KO signature (Unpublished).

Gene KO	Oncogene/Fusion	Date	Treatment	Replicates
Bmi1	CDX2	Feb-2018	in vitro 4-OHT	2
Bmi1	MLL-AF9	Feb-2018	in vitro 4-OHT	2
Prmt1	MLL-GAS7	Mar-2018	in vitro 4-OHT	2
Cbfb	AML1-ETO	Jan-2017	in vitro 4-OHT	2

4.3 Results

4.3.1 The observable cell-of-origin transcriptomic signature is reduced in 2015 samples

Seeing that the separation between cell-of-origin is less clear in my samples (see subsection 3.3.2, Figure 3.2), I went on to individually compare the 2014 and 2015 cell-of-origin or "celltype"

signatures. To keep the comparison in this section as close to the results of Siriboonpiputtana et al. (2017) as possible, differential expression between LSK^{ME} and GMP^{ME} was conducted using DESeq2 with the design model: *~experiment+celltype*. As a major component of our paper (Siriboonpiputtana et al. 2017), we describe a cell-of-origin signature as a possible reason for the difference of β-catenin requirement in LSK and GMP derived MLL-ENL transformed cells. In the original study, we show that although most transcriptional differences between LSK and GMP are abolished following MLL-ENL transformation, some differences in expression remain. The differentially expressed genes between normal LSK and GMP overlapped significantly with differentially expressed genes between LSK^{ME} and GMP^{ME}. Therefore, it was argued that the common signature could be due to an inherited transcriptional "memory" from their cell-of-origin.

One way of showing the distribution of differentially expressed genes is by plotting the log2 Fold Change of gene expression against the mean average gene expression, known as an MA plot. If we couple this with significance threshold dependent colouring, this allows us to observe the relationship between the level of transcript expression and differential gene expression detection. Similarly to what we show in (Siriboonpiputtana et al. 2017), Figure 4.1A shows MA plots between: 1. Normal FACS sorted LSK and GMP, 2. 2014 transformed LSK^{ME} and GMP^{ME}, 3. Additionally I show LSK^{ME} and GMP^{ME} from my 2015 experiment. Compared to normal LSK and GMP (Figure 4.1A, top), MLL-ENL transformation decreases the dispersion of expression values (Figure 4.1A, centre)., and in 2015 the dispersion is even lower (Figure 4.1A, bottom).

With an FDR cut-off of <0.05, very few genes in the so-called cell-of-origin "memory" signature from (Siriboonpiputtana et al. 2017) remain differentially expressed (Figure 4.1B), one of those being *Hoxa9* (Figure 4.1A, bottom). Comparing enriched gene ontologies and pathways across normal cells, 2014 and 2015 transformed cells (Figure 4.1C), it is apparent that while many of the enriched pathways in normal cells are also enriched in the 2014 transformed data, none of the enriched pathways are shared between normal cells and the 2015 transformed data. The only common cell-of-origin related pathways shared between the 2014 and the 2015 dataset are related to the ribosome and nonsense mediated decay (Figure 4.1C). I conclude that we are unable to observe a comparable cell-of-origin signature in the 2015 RNA-Seq data set when analysed separately with DESeq2.

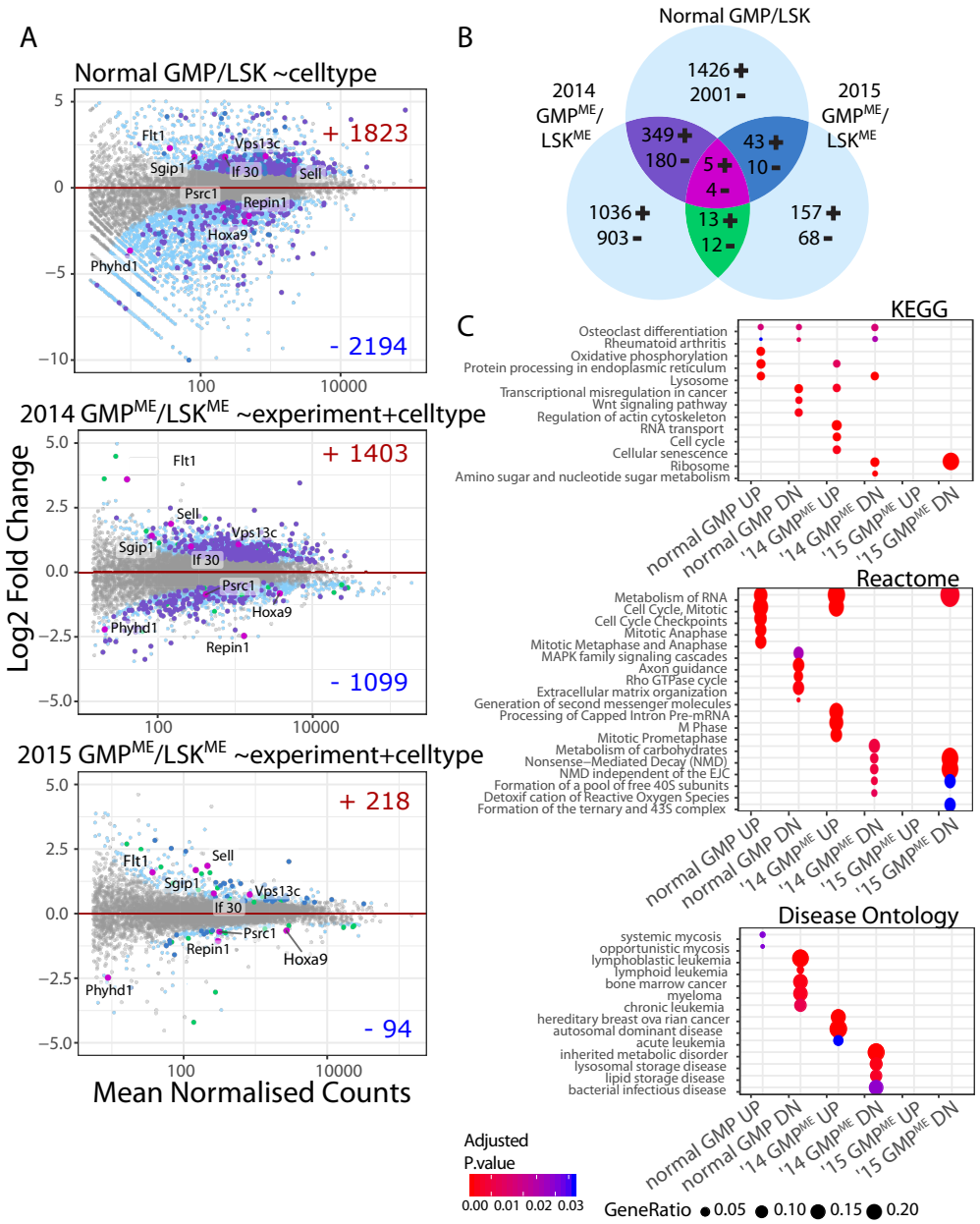

Figure 4.1: Detectable cell-of-origin signature is markedly reduced in 2015 samples Differential expression analysis for both 2014 and 2015 datasets using DESeq2. A) MA plots showing the distribution of gene expression between LSKs and GMPs in normal cells, 2014 MLL-ENL and 2015 MLL-ENL. Colours of points in the MA plots correspond to groups shown in the Venn diagram (B). Points are coloured depending on in which comparisons they are significant as indicated by the colours of the overlapping segments in the Venn diagram in B. Non-significant genes are plotted in grey. Genes which are significant in all comparisons are highlighted in pink are also labelled with their gene symbol. B) Venn diagram of overlaps of significantly differentially expressed genes from the three comparisons. C) Dot plots comparing pathway enrichment (rows) for up and downregulated genes from each comparison (columns). Each plot is titled with the source of the enriched pathways: KEGG, Reactome and Disease Ontology. The labels on the x-axes indicate the transcriptional signature. The enriched terms are listed on the y-axes. The size of each dot indicates the gene ratio; the ratio between the total number of genes in the enriched term and the number of significantly differentially expressed genes in the signature in that term. The colour of the dots indicates the adjusted P.value as shown in the key (bottom).

4.3.2 β-catenin signature genes significantly overlap with those in the published dataset

Following on from the differences seen in the cell-of-origin signature, I went on to compare the effects of β-catenin KO in the published dataset with my own data set. In the published study (Siriboonpiputtana et al. 2017), differential analysis was conducted using the Limma package. However, as the results were highly comparable, I chose to continue using EdgeR for this analysis. The same design model was used for each differential expression comparison to account for the cell line or "experiment" ("60" or "69"): $\sim experiment + genotype$.

In the published dataset (Siriboonpiputtana et al. 2017), when we knocked-out β-catenin in LSK^{ME} we obtained over 1000 differentially expressed genes. In contrast with LSK^{ME}, β-catenin KO in GMP^{ME} allowed detection of less than 200 genes (Figure 4.2A, top & 4.2B). This difference may be explained by greater biological variation between GMP^{ME} replicates than LSK^{ME} replicates, obscuring the transcriptional differences (see subsection 3.3.2, Figure 3.2C). Similarly, LSK^{ME} replicates in the 2015 study exhibit greater within-group variance than the LSK^{ME} replicates in the 2014 study (see subsection 3.3.2, Figure 3.2C & D). Thus, there are fewer detectable differentially expressed genes for LSK^{ME} in the 2015 study than the 2014 study. However, although differences in significant genes may be explained by reduced statistical power, it cannot explain the disconnect of log fold changes between the 2014 and 2015 LSK^{ME} datasets shown by an R2 value of 0.09 Figure 4.2A (top). There is also a reduction of observable targets in GMP^{ME}, but less obviously so (Figure 4.2A bottom & Figure 4.2B).

Nonetheless, despite their differences, many of the same enriched pathways remain in the upregulated genes following β-catenin KO (Figure 4.2C). Upregulated genes in all comparisons overwhelmingly highlight pathways related to immune response (Figure 4.2C). However, enriched pathways for downregulated genes such as the HIF-1 signalling pathway in GMP^{ME} and cell cycle pathways in LSK^{ME} for 2014 samples were no longer detectable in 2015 samples. As would be expected, overlaps between the 2014 and 2015 β-catenin KO signatures were found to be highly significant by permutation test and hypergeometric test (Table 4.3). Interestingly, comparison of genes affected by β-catenin loss in GMP^{ME} but not LSK^{ME} in a Venn diagram (Figure 4.2B) revealed Growth arrest and DNA damage-inducible protein alpha (*Gadd45a*) to be consistent between the two studies.

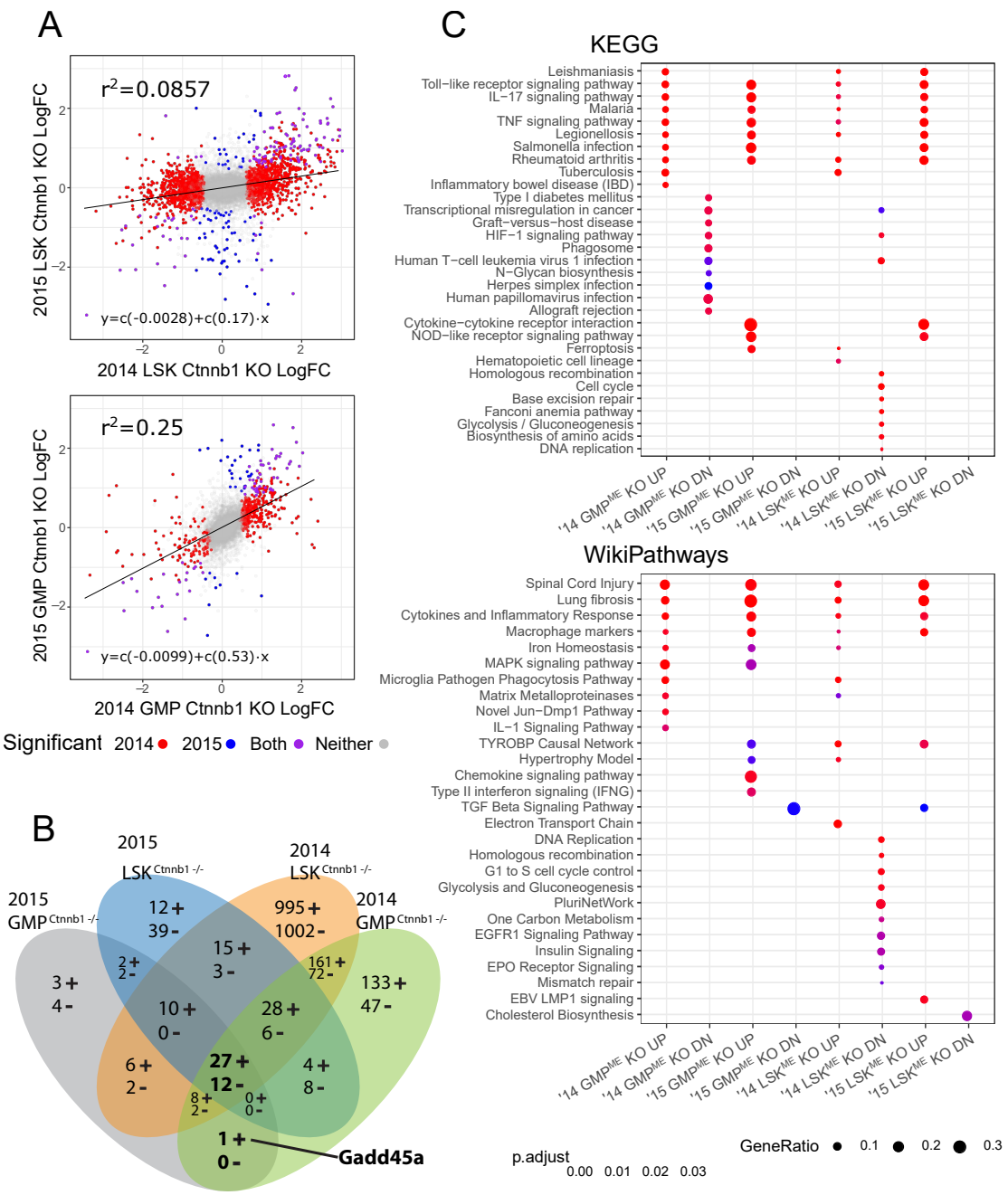

Figure 4.2: Comparison of transcriptomic effects of β-catenin KO across experiments A) A comparison of log-fold changes between 2014 and 2015 studies across all genes in LSK^{ME} (top) and GMP^{ME} (bottom). As indicated by the key (bottom): Grey points are non-significant, red points are significant genes unique to the 2014 dataset, blue points are unique to the 2015 dataset and purple points are significant in both. B) Venn diagram showing the overlaps between comparisons in each dataset. *Gadd45a* is highlighted as a target unique to GMP which is consistently found if the 2014 and 2015 datasets are analysed separately. C) A comparison of the enriched KEGG pathways (top) and WikiPathways (bottom) for significant genes in each signature. The labels on the x-axes indicate the transcriptional signature. The enriched terms are listed on the y-axes. The size of each dot indicates the gene ratio as indicated by the key (bottom). The colour of the dots indicates the multiple adjusted P.value as shown in the key (bottom).

Table 4.3: Significance of overlaps of differentially expressed genes following β-catenin deletion in LSK^{ME} & GMP^{ME} Significance of overlaps calculated using a permutation test and hypergeometric distribution. "N >= obs." stands for the number of overlaps greater than the observed in 10000 permutations. "P.val" is the P.value calculated from a permutation test with 1000 iterations. "phyper" is the P.value determined by hypergeometric test.

	Significantly Up			Significantly Down		
Comparison	N >= obs.	P.val	phyper	N >= obs.	P.val	phyper
14 GMP^{ME} Ctnnb1 | 15 GMP^{ME} Ctnnb1	0	<1E-04	6.1E-54	0	<1E-04	8.3E-39
14 GMP^{ME} Ctnnb1 | 15 LSK^{ME} Ctnnb1	0	<1E-04	1.5E-84	0	<1E-04	1.2E-41
14 GMP^{ME} Ctnnb1 | 14 LSK^{ME} Ctnnb1	0	<1E-04	5E-136	0	<1E-04	7.7E-61
15 GMP^{ME} Ctnnb1 | 15 LSK^{ME} Ctnnb1	0	<1E-04	1.4E-92	0	<1E-04	3.6E-33
15 GMP^{ME} Ctnnb1 | 14 LSK^{ME} Ctnnb1	0	<1E-04	4.2E-52	0	<1E-04	1.1E-19
15 LSK^{ME} Ctnnb1 | 14 LSK^{ME} Ctnnb1	0	<1E-04	1.2E-71	0	<1E-04	4.9E-09

To confirm the cell-of-origin difference in the upregulation of *Gadd45a* expression following β-catenin deletion, I performed RT-qPCR with 3 independent 4-OHT treatments in both biological replicates used for sequencing. As expected, KO of β-catenin in GMP induced higher expression of *Gadd45a* in GMP^{ME} but not in LSK^{ME} cells (Figure 4.3). In conclusion, whilst there is an overall reduction in detectable differential expression due to β-catenin knock-out, crucially, much of the same signature genes remain in GMP^{ME} derived cells including *Gadd45a*.

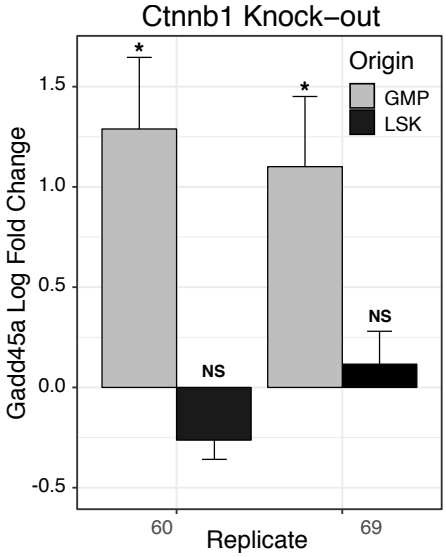

Figure 4.3: *Gadd45a* upregulation is a unique effect of β-catenin deletion in GMP^{ME} RT-qPCR validation of *Gadd45a* upregulation in GMP^{ME} (n=3, P.values are 0.0145 and 0.0231 for replicate 60 and 69 respectively) and insignificant deregulation in LSK^{ME} following β-catenin. (n=3, P.values are 0.3804 and 0.4785 for replicate 60 and 69 respectively)

4.3.3 Putative β-catenin targets are largely unique to MLL-ENL transformed GMP and LSK

The main gene ontology themes in both LSK^{ME} and GMP^{ME} appear to involve inflammatory response, cell adhesion, cell cycle and DNA-damage response (Figure 3.8 & 4.2). β-catenin has previously been shown to be involved in cell adhesion and cell cycle processes (Valenta et al. 2012). However, when I compared our potential targets with curated experimentally validated β-catenin target genes from the Stanford Wnt homepage (Nusse 2019), I did not see common genes in the expected pattern (Figure 4.4). Indeed, there is a small overlap from the LSK^{ME} signature, but they are regulated in the opposite direction than expected. For example, Matrix Metallopeptidase 2/9 (*Mmp2/9*) were up-regulated after KO but their expression is known to be activated by β-catenin. Furthermore, GSEA analysis of our transcriptomic signatures did not reveal enrichment for known Wnt/β-catenin pathways (Table 4.5).

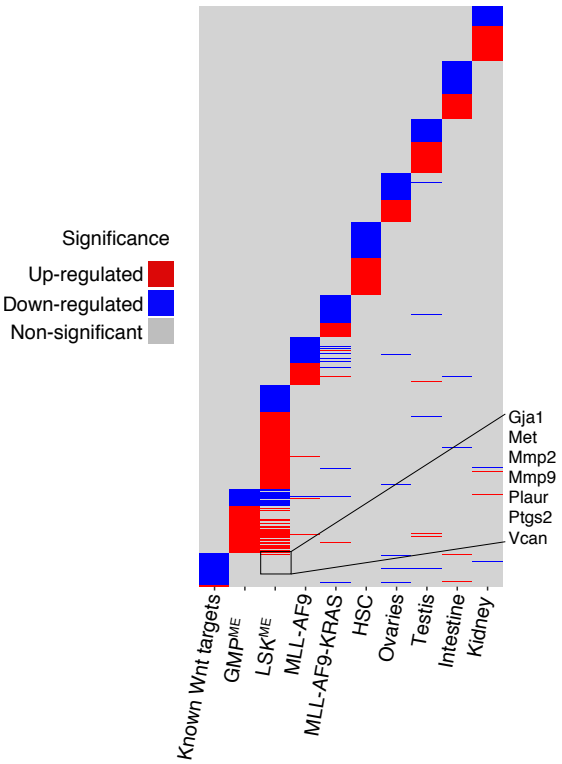

Figure 4.4: Comparison of β-catenin knock-out signatures with known targets and across tissues Simplified heatmap showing the incidence of differentially expressed genes responding to β-catenin deletion across different tissues. Each column represents a different experiment with the exception of the first which includes the Wnt/β-catenin targets taken from the Wnt Homepage (Nusse 2019). Each row is a gene which is coloured by significance in the experiment as decoded by the key (left). Genes from the LSK^{ME} signature which overlap Wnt Homepage genes are labelled (right).

Comparing genes which respond to β-catenin deletion in experiments using the same transgenic mouse model in different tissues and in similar systems such as MLL-AF9 LSCs (Ng et al. 2014), listed in Table 4.1, there is little overlap (Figure 4.4). In a similar vein, GSEA analysis does not reveal any enrichment of these transcriptomic signatures in our differential expression analysis (Figure 4.5). Together, the findings reported in this section suggest that β-catenin deletion results in a highly tissue and context dependent response.

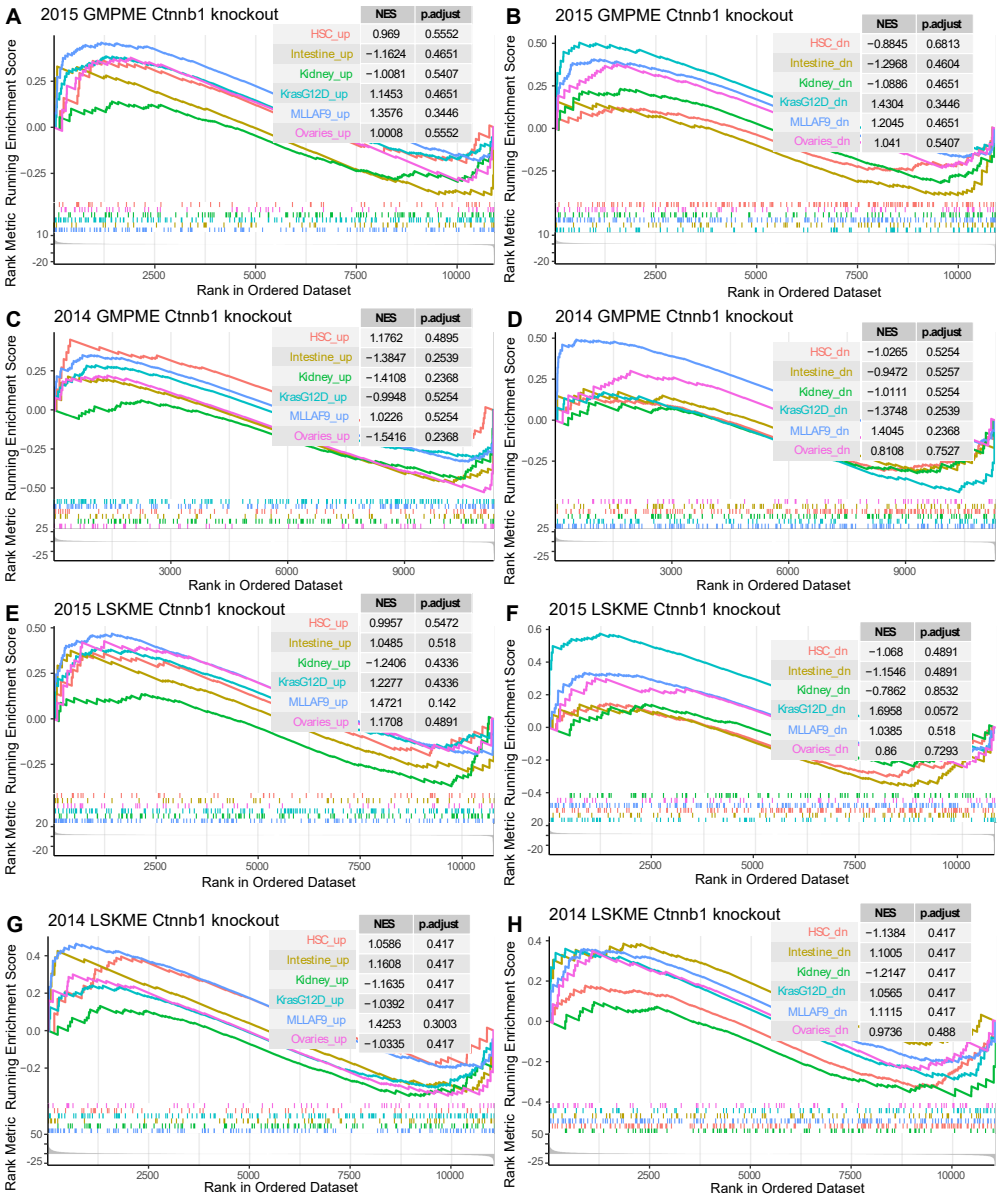

Figure 4.5: GSEA enrichment for genes deregulated after β-catenin knock-out. A-H) The plots are titles (top) by the comparison for which the GSEA was performed. The line graphs show the running enrichment score for a given gene set. Below this are barcode plots showing where the genes in fall in the ranked list. At the bottom, a histogram shows the rank values. The colour of the line and barcode plot indicates the gene set, as decoded by the table on the right. The Normalised Enrichment Score (NES) and adjusted P.value (p.adjust) for each gene set is given in the table (right). A, C, E & G are upregulated. B, D, F & H are downregulated.

Table 4.4: GSEA analysis of Wnt/β-catenin related pathways NES: Normalised Enrichment score. P.adj: Benjamini-Hochberg adjusted P.Values.

		14 LSK^{ME} Ctnnb1		15 LSK^{ME} Ctnnb1		14 GMP^{ME} Ctnnb1		15 GMP^{ME} Ctnnb1	
Source	Pathway Description	NES	P.adj	NES	P.adj	NES	P.adj	NES	P.adj
KEGG	Wnt signaling pathway	1.039	0.644	1.051	0.700	0.662	0.984	1.112	0.728
Reactome	Degradation of beta-catenin by the destruction complex	1.407	0.312	1.178	0.581	0.225	1.000	0.151	0.992
Reactome	Signaling by WNT	1.331	0.245	1.331	0.364	1.424	0.387	1.146	0.992
Reactome	Beta-catenin phosphorylation cascade	-0.633	0.999	1.040	0.811	0.394	1.000	0.394	0.992
Reactome	TCF dependent signaling in response to WNT	1.206	0.468	1.307	0.492	1.413	0.387	1.122	0.992
Reactome	Formation of the beta-catenin:TCF transactivating complex	0.838	0.978	1.219	0.523	1.254	0.811	0.349	0.992
Reactome	Deactivation of the beta-catenin transactivating complex	0.624	0.999	-1.169	0.668	0.382	1.000	0.414	0.992
Reactome	Beta-catenin independent WNT signaling	1.559	0.137	1.239	0.523	1.076	1.000	1.108	0.992
Reactome	WNT5A-dependent internalization of FZD4	1.015	0.833	0.422	0.946	1.750	0.963	0.472	0.992
WikiPathways	Wnt Signaling Pathway and Pluripotency	1.019	0.732	0.976	0.704	0.683	0.989	1.155	0.794
WikiPathways	Wnt Signaling Pathway NetPath	1.095	0.596	1.235	0.354	-0.717	0.989	1.128	0.794

4.3.4 Many of our β-catenin targets are also seen in experiments using the Cre-ER system in our lab

As I showed in subsection 4.3.3 that the transcriptional effects of β-catenin KO are largely unique to this study, even when compared with very similar cell lines, I wanted to investigate if there were any treatment related effects in our β-catenin KO signature. During my Ph.D., I have had the opportunity to analyse plentiful RNA-Seq datasets from different transgenic leukaemia models generated in our lab. From my experience I have found that some genes are recurrently significantly differentially expressed following 4-OHT treatment in vitro, regardless of genetic background or floxed allele. To see if any of our β-catenin targets could be explained by the treatment required to deplete β-catenin, I gathered RNA-Seq data from the lab which required 4-OHT treatment in vitro, these are listed in Table 4.2. In Figure 4.6B I show the β-catenin dependent signatures alongside multiple other knock-outs in different leukaemia cell lines generated in our group. Surprisingly, a significant amount of these genes overlap suggesting there may be a non-specific treatment effect. A summary of the overlaps is represented in a diagram (Figure 4.6A) and their statistics are listed in Table 4.5. These overlapping genes form part of the transcriptomic signature involving the immune system and extracellular matrix. I conclude that although there may be unique effects of β-catenin deletion in our system, we must interpret these with caution.

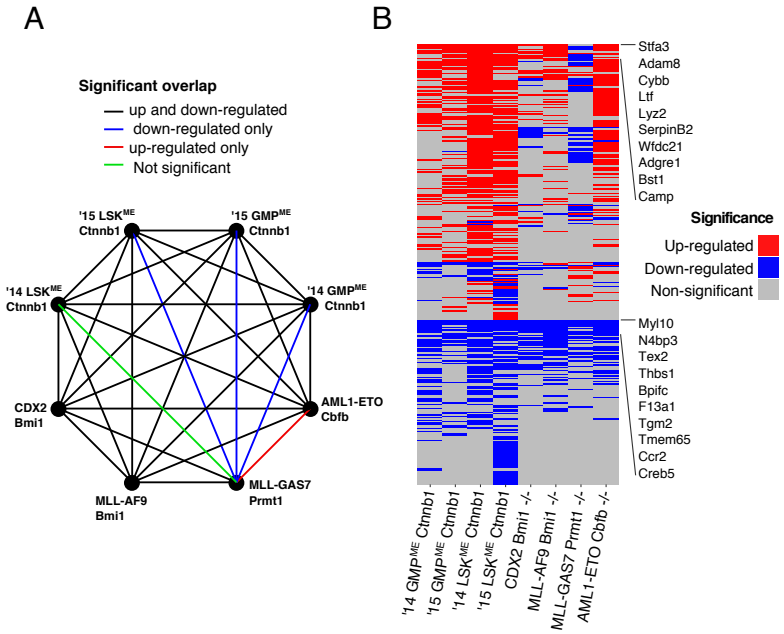

Figure 4.6: Specificity of β-catenin target genes compared to others using 4-OHT The diagram shows whether differentially expressed genes overlap between two studies was significant as assessed by permutation test. Significance is indicated by the colour of the line (key above). B) Plot showing differentially expressed genes in other studies in our lab which use 4-OHT in vitro. Significance is decoded in the key (right). Some of the most recurrent genes are named by their gene symbol.

Table 4.5: Significance of overlaps of differentially expressed genes following 4-OHT treatment in other studies Significance of overlaps calculated using a permutation test and hypergeometric distribution. "obs. overlap" is the number of genes which overlap in the comparison defined in the first column. "N >= obs." stands for the number of overlaps greater than the observed in 1000 permutations. "P.val" is the P.value calculated from the permutation test. "phyper" is the P.value determined by hypergeometric test.

	Significantly Up-regulated				Significantly Down-regulated			
Comparison	obs. overlap	N >= obs.	P.val	phyper	obs. overlap	N >= obs.	P.val	phyper
14 GMP^{ME} Ctnnb1 | CDX2 BMI1	27	0	<1E-04	4.4E-11	40	0	<1E-04	3.6E-23
14 GMP^{ME} Ctnnb1 | MLL-AF9 BMI1	71	0	<1E-04	2.4E-65	48	0	<1E-04	8.6E-49
14 GMP^{ME} Ctnnb1 | MLL-GAS7 PRMT1	16	3844	0.3844	0.4154	28	0	<1E-04	2.2E-09
14 GMP^{ME} Ctnnb1 | AML1ETO Cbfb	184	0	<1E-04	9.1E-79	42	0	<1E-04	6.8E-12
15 GMP^{ME} Ctnnb1 | CDX2 BMI1	8	0	<1E-04	4.9E-06	11	0	<1E-04	1.8E-09
15 GMP^{ME} Ctnnb1 | MLL-AF9 BMI1	12	0	<1E-04	2.0E-11	17	0	<1E-04	8.7E-24
15 GMP^{ME} Ctnnb1 | MLL-GAS7 PRMT1	1	9562	0.9562	0.8332	9	0	<1E-04	4.7E-06
15 GMP^{ME} Ctnnb1 | AML1ETO Cbfb	37	0	<1E-04	2.0E-17	13	0	<1E-04	2.6E-07
15 LSK^{ME} Ctnnb1 | CDX2 BMI1	12	0	<1E-04	3.8E-07	25	0	<1E-04	1.6E-14
15 LSK^{ME} Ctnnb1 | MLL-AF9 BMI1	15	0	<1E-04	2.5E-11	29	0	<1E-04	1.3E-28
15 LSK^{ME} Ctnnb1 | MLL-GAS7 PRMT1	2	9692	0.9692	0.9216	21	0	<1E-04	1.8E-08
15 LSK^{ME} Ctnnb1 | AML1ETO Cbfb	64	0	<1E-04	5.1E-28	27	0	<1E-04	6.0E-08
14 LSK^{ME} Ctnnb1 | CDX2 BMI1	40	1	1.00E-04	1.4E-05	78	0	<1E-04	3.5E-08
14 LSK^{ME} Ctnnb1 | MLL-AF9 BMI1	51	0	<1E-04	1.6E-14	45	0	<1E-04	1.7E-08
14 LSK^{ME} Ctnnb1 | MLL-GAS7 PRMT1	34	9979	0.9979	0.9952	59	3687	0.3687	0.3651
14 LSK^{ME} Ctnnb1 | AML1ETO Cbfb	304	0	<1E-04	8.7E-47	168	0	<1E-04	5.1E-14
CDX2 BMI1 | MLL-AF9 BMI1	28	0	<1E-04	3.6E-20	128	0	<1E-04	3E-122
CDX2 BMI1 | MLL-GAS7 PRMT1	33	0	<1E-04	1.4E-10	93	0	<1E-04	1.0E-26
CDX2 BMI1 | AML1ETO Cbfb	47	0	<1E-04	1.6E-05	65	5	5.00E-04	2.2E-03
MLL-AF9 BMI1 | MLL-GAS7 PRMT1	17	7	7.00E-04	7.7E-04	37	0	<1E-04	1.3E-09
MLL-AF9 BMI1 | AML1ETO Cbfb	78	0	<1E-04	1.5E-28	43	0	<1E-04	3.3E-06
MLL-GAS7 PRMT1 | AML1ETO Cbfb	73	66	0.0066	0.0709	47	8682	0.8682	0.9792

4.4 Discussion

The first experimental aim of this chapter was to replicate the cell-of-origin signature in Siriboonpiputtana et al. (2017). On first inspection, it could be observed that dispersion of expression values was markedly reduced between LSK and GMP origin cells in the 2015 study (Figure 4.1A). With the exception of *Hoxa9*, most of the original cell-of-origin signature genes were not differentially expressed in the 2015 study (Figure 4.1B). Congruently, no cell-of-origin differentially regulated pathways common to normal and transformed cells from the 2015 study were found (Figure 4.1C). However, due to the variance in the 2015 dataset, the experiment was statistically underpowered with only 2 replicates. Indeed, RNA-Seq experiments with few replicates lead to results which are highly sensitive to analysis methods (Baccarella et al. 2018). However, when both datasets are combined and the effects of experiment are modelled, some interesting cell-of-origin differences remain detectable, such as genes involved in cancer and cell cycle (Figure 3.4).

Next, I investigated whether the effects of β-catenin were similar to those reported in Siriboonpiputtana et al. (2017). Although fewer genes could be identified as statistically significant in both LSK^{ME} and GMP^{ME} (Figure 4.2A & B), most of these genes were also found in the published study (Siriboonpiputtana et al. 2017), including *Gadd45a* (Figure 4.2B & 4.3). The reduction in number of differentially expressed genes following β-catenin KO between the 2014 and 2015 studies could be explained by a number of factors. Firstly a it can be observed that there is greater within-group variance of LSK^{ME} replicates in the 2015 study than the 2014 study. The reduction of genes responding to β-catenin deletion in both LSK^{ME} and GMP^{ME} could also be explained by "leaky" Cre-recombinase activation as demonstrated in Figure 3.1A. Even very low expression of Cre-recombinase will result in an accumulation of irreversible deletions in the cell population. Over culture time, cells may adapt to the altered gene expression caused by deletions and minor changes in the culturing environment.

On first examination, it was found that our β-catenin targets do not overlap with functionally validated canonical Wnt targets (Figure 4.4) which could suggest β-catenin performs unique roles in leukaemia. In addition, applying GSEA revealed that Wnt/β-catenin pathways were not enriched in our signatures (Table 4.4). In order to understand how general the transcriptional effects of β-catenin KO are, I compared our results with other studies (Figure 4.4 & 4.5). However, there was generally little consensus across experiments using the same mouse model in different tissues (Figure 4.4). In fact, only embryonic ovaries showed reduced expression in a well known Wnt/ β-catenin target, Leucine-rich repeat-containing G-protein coupled receptor 5 (*Lrg5*). Most surprisingly, β-catenin targets do not overlap with those shown in MLL-AF9 transformed c-kit cells (Ng et al. 2014). The lack of consensus between our β-catenin KO experiment and other tissues was also confirmed by GSEA (Figure 4.5).

Some of the differences in transcriptomic signatures between published studies may be explained by the context in which β-catenin was deleted. For example, Ng et al. (2014) transplanted MLL-AF9 transformed haematopoietic cells into recipient mice and treated them with polyinosinic:polycytidylic acid (pIpC) to induce β-catenin deletion. RNA was obtained from GFP positive sorted cells after leukaemia development. As previously mentioned in section 1.7, LSCs show increased activation of β-catenin over Pre-LSCs and also show an increase in conventional Wnt signalling target expression (Yeung et al. 2010). This may be due to secondary genetic events in LSCs such as the FLT3-ITD mutation which has been shown to increase β-catenin activity in AML (Kajiguchi et al. 2007). β-catenin dependency was shown to be conditional in vivo however, deletion of β-catenin and subsequent RNA-Seq was conducted in vitro with Pre-LSCs (Siriboonpiputtana et al. 2017). It is important to note that although β-catenin is constitutively expressed in most tissues, it is the activation of β-catenin and suppression of its destruction which is critical for its involvement in transcription (see Figure 1.6). The lack of canonical target effects could be due to insufficient β-catenin activation in pre-LSCs. Our β-catenin signature may be further reduced by depletion of β-catenin through leakiness of the Cre-recombinase in the absence of tamoxifen (Figure 3.1A). However, as the RNA-Seq in Ng et al. (2014) also failed to show expected Wnt targets (Figure 4.4), it remains unclear which expression signature is more relevant.

Furthermore, the absence of known targets in gene expression signatures or unexpectedly modest phenotypic effects after gene KO is not a phenomenon unique to β-catenin deletion. It has been hypothesized that this is in part due to transcriptional redundancy mechanisms (El-Brolosy and Stainier 2017). This effect was first demonstrated in (Rossi et al. 2015) by comparing Epidermal Growth Factor-Like Protein 7 (*egfl7*) mutants and morphants in zebrafish. The study found that genetic knock-outs elicited upregulation of *egfl7* target genes whilst knockdown did not. Very recently, further compelling arguments for caution against the use of certain KO models for unpicking molecular mechanisms of transcription factors were published in Nature. It was shown that transcription of faulty mRNA in a process called Nonsense Induced Transcriptional Compensation (NITC) causes increases in the expression of the transcription factor target genes (Ma et al. 2019). Indeed, a NITC response may by triggered in the β-catenin model we utilise in this study as a truncated mRNA is still transcribed following KO (see subsection 3.3.1, Figure 3.1F). The implications of this could also extend to the study of β-catenin in normal haematopoiesis and may partially explain discrepancies between the effect of Lymphoid enhancer-binding factor-1 (*Lef1*) KO (Reya et al. 2000) and β-catenin deletion (Cobas et al. 2004; Jeannet et al. 2008). This research will hopefully influence the design of new models to avoid this response which may be responsible for widespread irreproducibility in molecular analyses of single gene knock-outs (Ma et al. 2019).

As mentioned in section 1.3, β-catenin has dual functionality, as well as its diverse roles in transcription, β-catenin is critically involved in cell adhesion. Consistent with this, GO terms related to the extracellular matrix were enriched amongst β-catenin targets (Figure 3.8). Abolishment of homing ability of was shown not be the cause of β-catenin dependency of GMP^{ME} (Siriboonpiputtana et al. 2017). Impaired cell adhesion ability has already been put forward as a possible reason for requirement of β-catenin in LSCs (Chung et al. 2002; Lane et al. 2011). Surprisingly however, some validated Wnt targets, such as Matrix Metallopeptidase 2 and 9 (*Mmp2/9*) and Claudin 1 (*Cldn1*) are unexpectedly upregulated after knock-out in LSK^{ME} (Figure 4.4). *Mmp2/9* belong to the matrix metallopeptidase family of enzymes involved in the degradation of the extracellular matrix which have been shown to be critically involved in metastasis (Conlon and Murray 2019). *Cldn1* is involved in formation of tight junctions and like MMPs are involved in the spread of cancers (Tabariès and Siegel 2017). A possible explanation could be a non-specific effect of 4-OHT, as Nilsson et al. (2007) showed that *Mmp9* and *Mmp2* are positively regulated by tamoxifen in human cells in culture.

With great consistency, deletion of β-catenin seems to cause activation of immune response genes (Figure 3.8 & 4.2C). Intriguingly, it was recently shown that the Wnt/ β-catenin pathway may be hijacked by solid tumours to escape and increase tolerance to the host immune response (Pai et al. 2017). However, I had observed similar gene signatures in differential analysis of other projects using the same Cre-ER system in vitro in the lab and decided to investigate this further (Figure 4.6 and Table 4.5). It could be argued that immune response genes are genuine frequent common targets of transcription factors involved in leukaemia. Indeed, regulation of inflammation is critical to leukaemia development (Giles et al. 2014). However, Cre-recombinase activation has been shown to cause toxicity to cells in vitro due to DNA damage and presence of recombination products triggering an interferon response in vitro (Janbandhu et al. 2014; Liu et al. 2018; Loonstra et al. 2001; Pepin et al. 2016; Higashi et al. 2009). One particular study has shown that we see this immune response regardless of the presence of floxed alleles (Janbandhu et al. 2014). Specifically, Cre-ER toxicity has been noted in haematopoietic cells (Higashi et al. 2009). These studies highlight the possibility that molecular effects of β-catenin depletion may be overshadowed by Cre-mediated excision side-effects. There are several strategies suggested to mitigate these issues including self excising retroviral vectors and chemical inhibitors of the interferon response (Silver and Livingston 2001).

Whether or not the observed expression changes in response to β-catenin deletion are due to loss of transcriptional regulation mediated by β-catenin or model system effects, the in vivo consequences of β-catenin deletion in GMP^{ME} remain. As I am to attempt to functionally validate my findings in the same cell lines used for RNA-Seq, I can only count on genes which are consistent between studies. For functional analysis it would be logical to pursue genes which a uniquely perturbed in GMP^{ME}s. Simply comparing the overlaps between differentially expressed genes highlighted *Gadd45a* as a consistent GMP^{ME} unique target (Figure 4.2B & 4.3). *Gadd45a* is also a biologically plausible target, as it is a cell cycle arrest mediator regulated by p53, which has been shown to interact with β-catenin in the context of MLL leukaemia (Lynch et al. 2016). Additionally, methylation of *Gadd45a* has been shown as a poor prognostic factor in human leukaemia (Perugini et al. 2013). In the next chapter I will continue investigating *Gadd45a* expression in GMP derived MLL-ENL leukaemia as a possible cause for β-catenin dependency in vivo.

Chapter 5

The role of *Gadd45a* in GMP derived MLL-ENL LSCs

5.1 Introduction

In the previous chapter, I highlighted Growth arrest and DNA damage-inducible protein alpha *Gadd45a* as a β-catenin target unique to MLL-ENL pre-LSCs of GMP origin. As β-catenin has been shown to be required for leukaemia development from MLL-ENL transformed cells of GMP origin (Siriboonpiputtana et al. 2017), GMP specific β-catenin targets should provide us with greater information of the mechanism by which β-catenin is required in vivo.

The primary function of *Gadd45a* is a growth arrest mediator, working in collaboration with p53, therefore it makes sense that *Gadd45a* alone would act as a tumour suppressor. Intriguingly, *Gadd45a* is also involved in the re-activation of genes silenced by DNA methylation through (Barreto et al. 2007). Particularly in the context of cancer *Gadd45a* is also important for maintaining cell-cell adhesion (Ji et al. 2007). In fact, silencing of *Gadd45a* in human leukaemia via DNA methylation has been demonstrated as a factor for poor prognosis (Perugini et al. 2013). Suppression *Gadd45a* was demonstrated to cause a block in myeloid differentiation (Perugini et al. 2009). *Gadd45a* has also been shown to directly interact with β-catenin in leukaemia, reducing its activity by translocation of β-catenin to the mitochondria (Voli et al. 2015; Lynch et al. 2016). *Gadd45a* loss can protect MLL-fusion transformed Pre-LSCs from mitochondrial reactive oxygen species (ROS) and cell death (Lynch et al. 2016).

However, based on our RNA-Seq data, I hypothesise that the transcription of *Gadd45a* may in fact be controlled by β-catenin itself in a feedback loop. I set out to test the importance of *Gadd45a* in GMP derived MLL-ENL leukaemia experimentally using a two pronged approach: Overexpression of *Gadd45a* in GMP derived LSCs to see if I can eliminate leukaemia and phenocopy β-catenin KO in GMP^{ME}. In addition I used shRNA to knockdown *Gadd45a* in GMP derived cells with a β-catenin KO background to see if I could rescue the leukaemic phenotype. Finally, I tested the effect of *Gadd45a* overexpression on LSK derived cells to determine if *Gadd45a* exerts any cell-of-origin specific function.

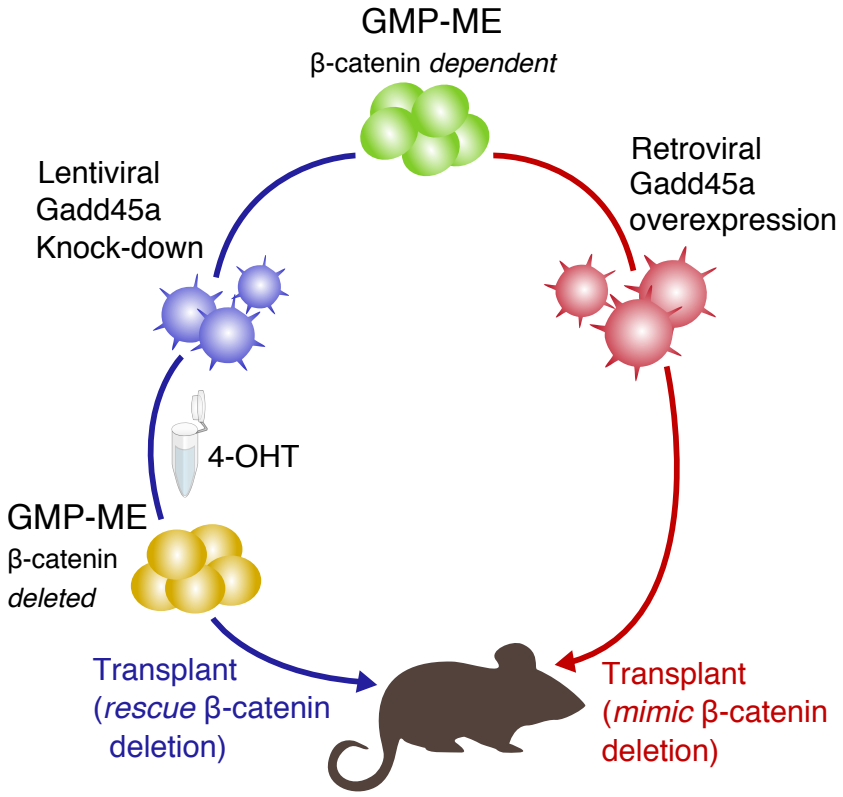

Figure 5.1: Schematic illustrating the experimental set up for this chapter.

5.2 Methods

To reduce the time span of transplant experiments I used pre-existing origin specific MLL-ENL transformed LSCs (previously transplanted into mice). As explained in section 1.6, Figure 1.5, Pre-LSCs require further genetic or epigenetic events to occur in vivo to produce leukaemia in vivo. However, it was shown in (Siriboonpiputtana et al. 2017) that β-catenin is still required in GMP derived MLL-ENL transformed LSCs, therefore the readout of in vivo experiments should theoretically be equivalent to Pre-LSCs. To confirm the expression of *Gadd45a*, I calculated

RT-qPCR using the delta CT method. Where available, log2 fold changes are used to allow display of standard error in a meaningful way. P.values for RT-qPCR are calculated with a two-tailed t-test on delta CT values. Detailed methods for cloning are provided in section 2.6 & 2.7. Retroviral/ lentiviral transduction is explained in subsection 2.1.4.

Prior to transplanting cells, I performed in vitro assays to characterize the effect of *Gadd45a* perturbation. Firstly, I used a colourimetric assay to measure the metabolic activity of the cells (MTT assay) as described in subsection 2.7.1. Then, to measure the proliferation capacity I performed a colony formation assay; I seeded 5000 cells into methylcellulose medium which is a thick medium allowing each isolated cell to form distinct colonies. Counting the number of colonies allows us to measure the proliferative capacity of the cells. If there were any differences found using these assays, I then moved on to check the cell morphology by Cytospin followed by Wright-Giemsa staining as described in subsection 2.7.2. Additionally, I used Annexin FACS analysis as described in subsection 2.7.3 to measure the number of apoptotic cells.

To investigate the effects of *Gadd45a* overexpression or knockdown in leukaemia development, cells were transplanted into SJL or Black 6 recipient mice as described in subsection 2.1.5. Mice were sacrificed when they began showing signs of leukaemia or 6 months after transplant. Following necropsy, I recorded the appearance, size and weight of the spleen and liver. I took samples of peripheral blood to perform bloodcounts and smears. To confirm leukaemia, I took white cells from the blood, bone marrow, spleen and liver and performed immunophenotyping as described in subsection 2.1.5 and analysed the resulting FACS data using FlowJo software, using the R package CytoML to pull the sample information from the XML files. Note that in the *Gadd45a* overexpression experiment, CD45.1 positive cells were transplanted into Black 6 mice (Figure 5.3D). In contrast, CD45.2 positive cells were transplanted into SJL mice for the *Gadd45a* knockdown experiment (Figure 5.6B).

5.3 Results

5.3.1 *Gadd45a* overexpression shows little phenotypic effect in vitro

After transduction with *Gadd45a* expressing retrovirus and antibiotic selection, I tested the expression of *Gadd45a* with RT-qPCR, achieving overexpression between 10 and 50 fold (Figure 5.2A). To assess the phenotypic effect of *Gadd45a* in vitro I performed MTT and colony

forming assays. The MTT showed a mild,but insignificant decrease in proliferation after transduction with pMSCV-*Gadd45a* (Figure 5.2B, p=0.322). No significant change in colony formation was observed after overexpression either (Figure 5.2C, p=0.173).

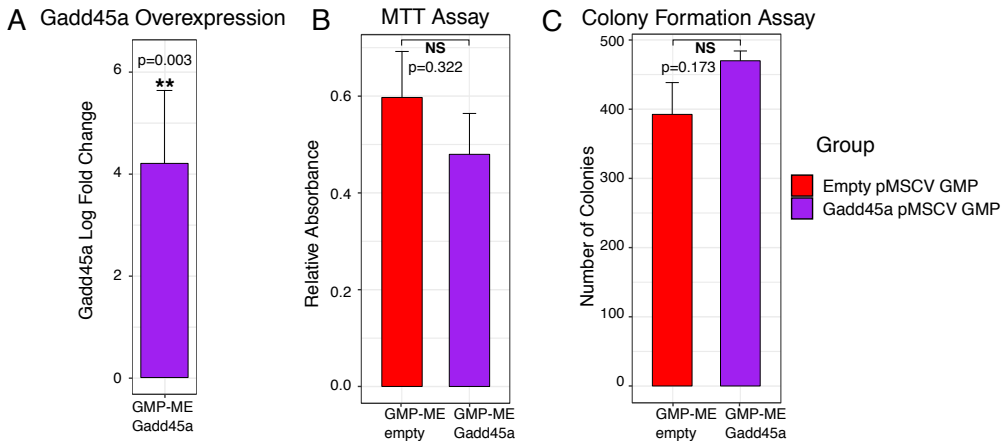

Figure 5.2: In vitro Overexpression of *Gadd45a* A) RT-qPCR showing overexpression of *Gadd45a*. B) MTT assay for pMSCV empty vector or pMSCV-*Gadd45a* transduced cells. Relative absorbance is calculated by deducting the absorbance of fresh media from the absorbance of the cells. C) Colony forming assay of pMSCV empty vector or pMSCV-*Gadd45a* transduced cells after antibiotic selection. P.values are as indicated and were calculated with a two-tailed t-test.

5.3.2 *Gadd45a* overexpression in GMP^{ME} eliminates leukaemia in vivo

MLL-ENL transformed cells of GMP origin with confirmed *Gadd45a* overexpression, alongside negative control GMP^{ME} cells transduced with empty pMSCV vector were transplanted into mice. 5 mice were transplanted each for empty pMSCV vector controls and *Gadd45a* overexpression. This experiment was independently repeated once resulting in a combined total of 10 mice per condition (Figure 5.3).

A leukaemic phenotype was confirmed by splenomegaly (Figure 5.3B & C) engraftment in the bone marrow and infiltration into the spleen and liver (Figure 5.3D & E). A skew towards the myeloid lineage was seen in immunophenotyping (Figure 5.3D & E). Mice that lived beyond 6 months were sacrificed and the absence of engrafted donor cells was confirmed by FACS analysis. Remarkably, whilst showing little effect in vitro (Figure 5.2), *Gadd45a* overexpression in GMP^{ME} eliminated leukaemia in vivo (Figure 5.3A).

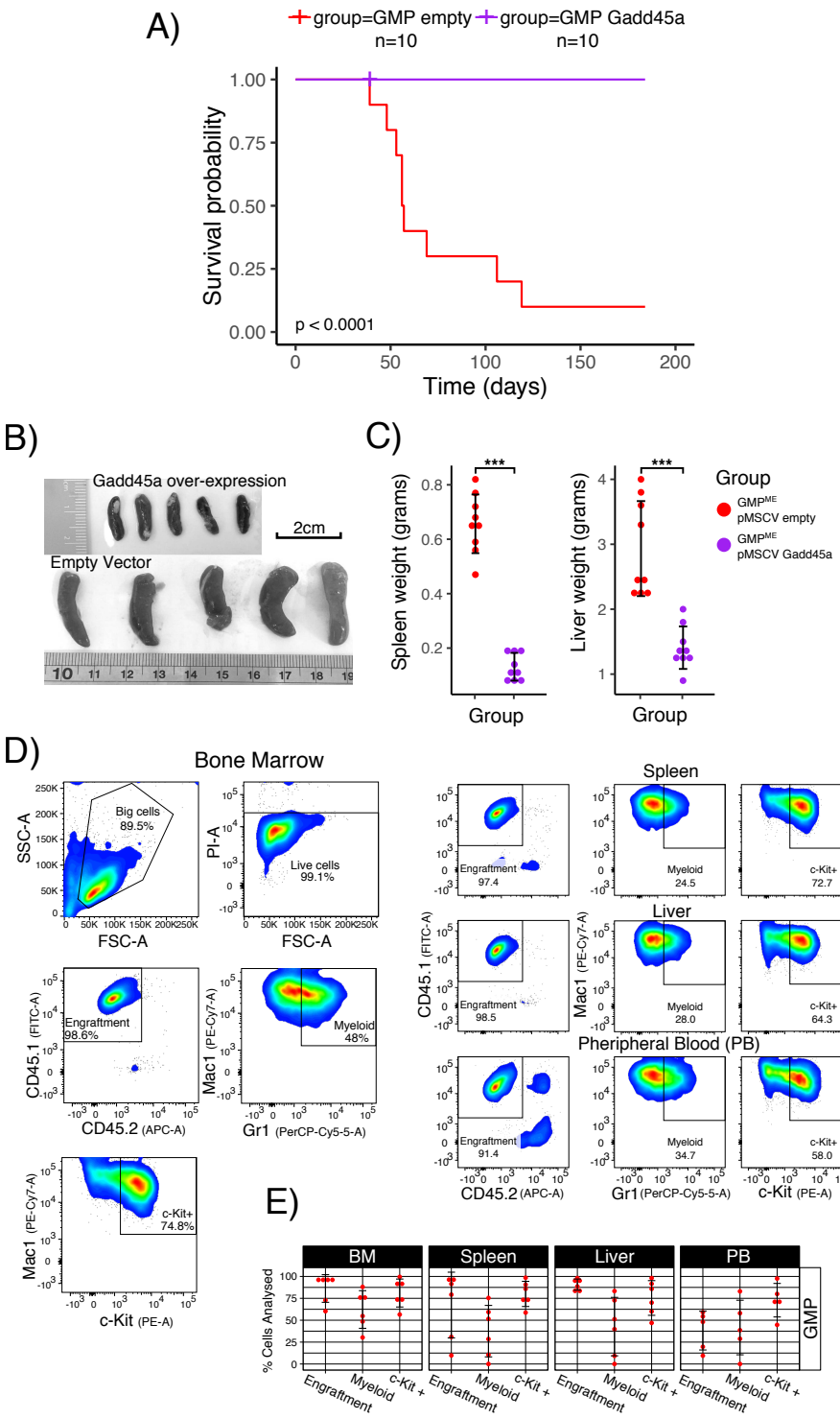

Figure 5.3: *Gadd45a* overexpression eliminates GMP derived MLL-ENL leukaemia
A) Survival curves for transplant of both GMP^{ME} with *Gadd45a* overexpression (purple) or empty pMSCV vector (red). Censored values are indicated by a vertical line. B) Photograph of spleens at necropsy. C) Spleen and liver weights for each group at necropsy. P.values for decrease in spleen and liver weight when *Gadd45a* is overexpressed are 1.1375E-06 and 3.218E-05 respectively. D) Immunophenotyping of leukaemia cells from a representative individual transplanted with GMP^{ME} cells transduced with pMSCV empty vector. E) Dotplot showing a summary of FACS analysis across samples. The y axis indicates the percentage of the parent gate. Gates are as shown in D.

5.3.3 *Gadd45a* knock-down causes an increase in proliferation and self-renewal

To knockdown *Gadd45a* expression, I cloned 3 different shRNA sequences into the pLKO lentiviral backbone. Following lentiviral production and transduction of the shRNA, I tested knock-down efficiency by RT-qPCR (Figure 5.4A). I found that one shRNA (sh*Gadd45a* 3) had markedly more effective knockdown and was therefore selected for subsequent experiments. Interestingly, when *Gadd45a* expression was reduced there was a noticeable and significant increase in proliferation with or without deletion of β-catenin shown by colony forming assay (Figure 5.4C), this increase in proliferation was also seen by the MTT assay, but it was statistically insignificant (Figure 5.4B). I conclude that *Gadd45a* knock down causes a modest increase in cell proliferation regardless of β-catenin status and may partially compensate for the effects of β-catenin in vitro.

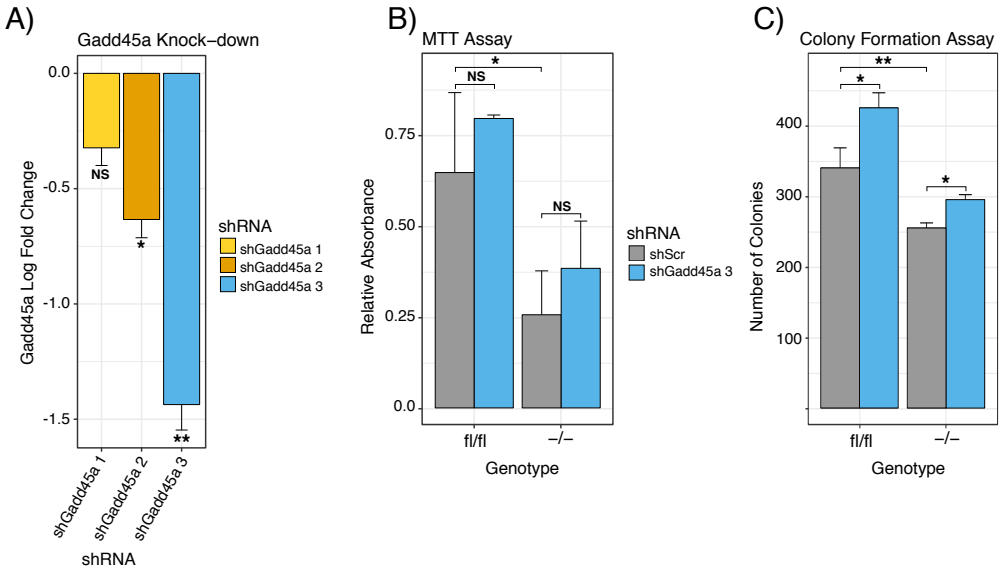

Figure 5.4: In vitro knockdown of *Gadd45a* A) Log fold change of *Gadd45a* measured by RT-qPCR following shRNA transduction. P. values are 0.761, 0.045, 0.001 for shRNA-1, shRNA-2 and shRNA-3 respectively. B) MTT assay in the context of β-catenin deletion. The absorbance between shScr and sh*Gadd45a* treated floxed cells was insignificant (p=0.294). The absorbance between shScr treated floxed and deleted cells was significant (p=0.026). The absorbance difference between shScr and sh*Gadd45a* treated deleted cells was insignificant (p=0.078). C) Colony forming assay in the context of β-catenin deletion. sh*Gadd45a* expression showed a significant increase in colony numbers in both floxed (p=0.020) and deleted cells (p=0.013). β-catenin deletion conferred a significant decrease in colony numbers (p=0.007)

5.3.4 β-catenin deletion results in apoptosis and differentiation in vitro

As we observed an in vitro phenotype in response to β-catenin deletion (Figure 5.4), which was not described in Siriboonpiputtana et al. (2017), I went on to examine this effect further. To investigate this, I monitored transformed β-catenin conditional cells by MTT every 24 hours for 72 hours following addition of 20 nM 4-OHT. The metabolic activity of LSK^{ME} is reduced after 24 hours but recovers after 48h (Figure 5.5B, top). This suggests that LSK^{ME} suffers toxicity in response to 4-OHT or β-catenin knock-out, but later recovers without β-catenin present. GMP^{ME} on the other hand continues to decline after 72 hours of 4-OHT treatment (Figure 5.5B, bottom). Annexin FACS analysis (Figure 5.5C) also shows a greater number of dead or apoptotic cells after 72 hours with a greater effect seen in GMP^{ME}s (bottom) than in LSK^{ME}s (top). In Figure 5.5A left it can be observed that the normal phenotype of LSK^{ME} and GMP^{ME} shows mostly large round mono-nuclei cells also known as blasts. However, after addition of 4-OHT dead cells and cells with lobed nuclei indicating differentiation are observed more frequently (Figure 5.5A, right). I conclude that GMP^{ME} cell lines used in this study are sensitive to β-catenin deletion in vitro, exhibiting differentiation and apoptosis.

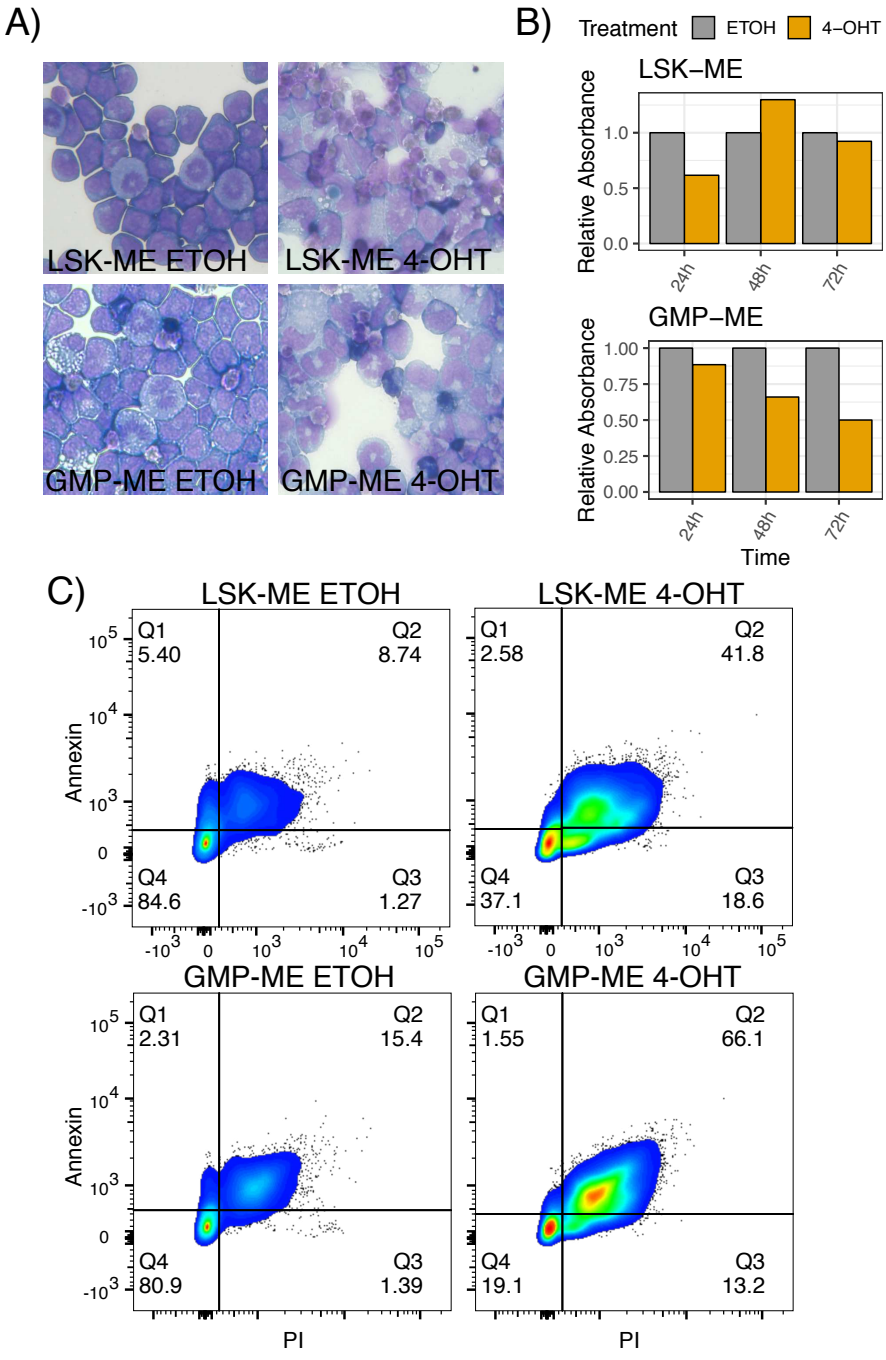

Figure 5.5: Effect of β-catenin KO with 20 nM 4-OHT. A) Cytospin with Wright-Giemsa staining in response to 20 nM 4-OHT. Microscope images are labelled with their sample and treatment. B) MTT assay at time points after 4-OHT treatment. Statistical significance is not shown as only technical replicates were used for this assay. C) Annexin FACS analysis to measure apoptosis. The x axis shows PI staining for dead cells. The y axis shows Annexin staining for apoptotic cells.

5.3.5 β-catenin deletion resistance emerges in GMP^{ME} in vivo

To investigate whether *Gadd45a* knockdown could rescue leukaemia in GMP^{ME} β-catenin knock-out, I transplanted GMP^{ME}s following confirmation of *Gadd45a* knockdown and β-catenin knock-out. 5 mice were transplanted for both β-catenin KO with scrambled shRNA and β-catenin KO with shRNA targeted to *Gadd45a*. For positive β-catenin floxed controls, 3 mice were transplanted. This study was independently repeated once resulting in a total of 10 mice for each β-catenin KO condition and 6 mice for β-catenin floxed cells. In these transplant studies the β-catenin conditional cell lines displayed incomplete penetrance with only 40-60% of each group coming down with leukaemia (Figure 5.6A). Surprisingly, β-catenin KO cells expressing only the negative scrambled shRNA control was the deadliest of all groups (Figure 5.6A). However, although there was greater incidence of leukaemia there was a reduction in c-kit cells (Figure 5.6B) suggesting that there is some effect on the phenotype of the resulting disease.

As LSCs are assumed to acquire addition mutations (see section 1.6, Figure 1.5) in vivo, it may be speculated that some LSCs become resistant to β-catenin deletion through accumulation of oncogenic mutations. However, additional experiments also showed that GMP^{ME} Pre-LSCs can develop in vivo resistance to β-catenin KO (Figure 5.7A). In all cases of β-catenin knock-outs producing leukaemia, I confirmed that the leukaemias were not caused by β-catenin cells which escaped deletion by performing genotyping PCR for β-catenin on DNA extracted from the tissues of these mice (Figure 5.7C). As I was unable to replicate in vivo β-catenin dependency of GMP^{ME}, the effect of *Gadd45a* knockdown on β-catenin dependent MLL-ENL transformed cells remains inconclusive. Following these data, I wanted to test whether the effect of *Gadd45a* overexpression was specific to GMP^{ME}. In order to do this I repeated the overexpression of *Gadd45a* and subsequent transplant experiments in LSK^{ME} cells.

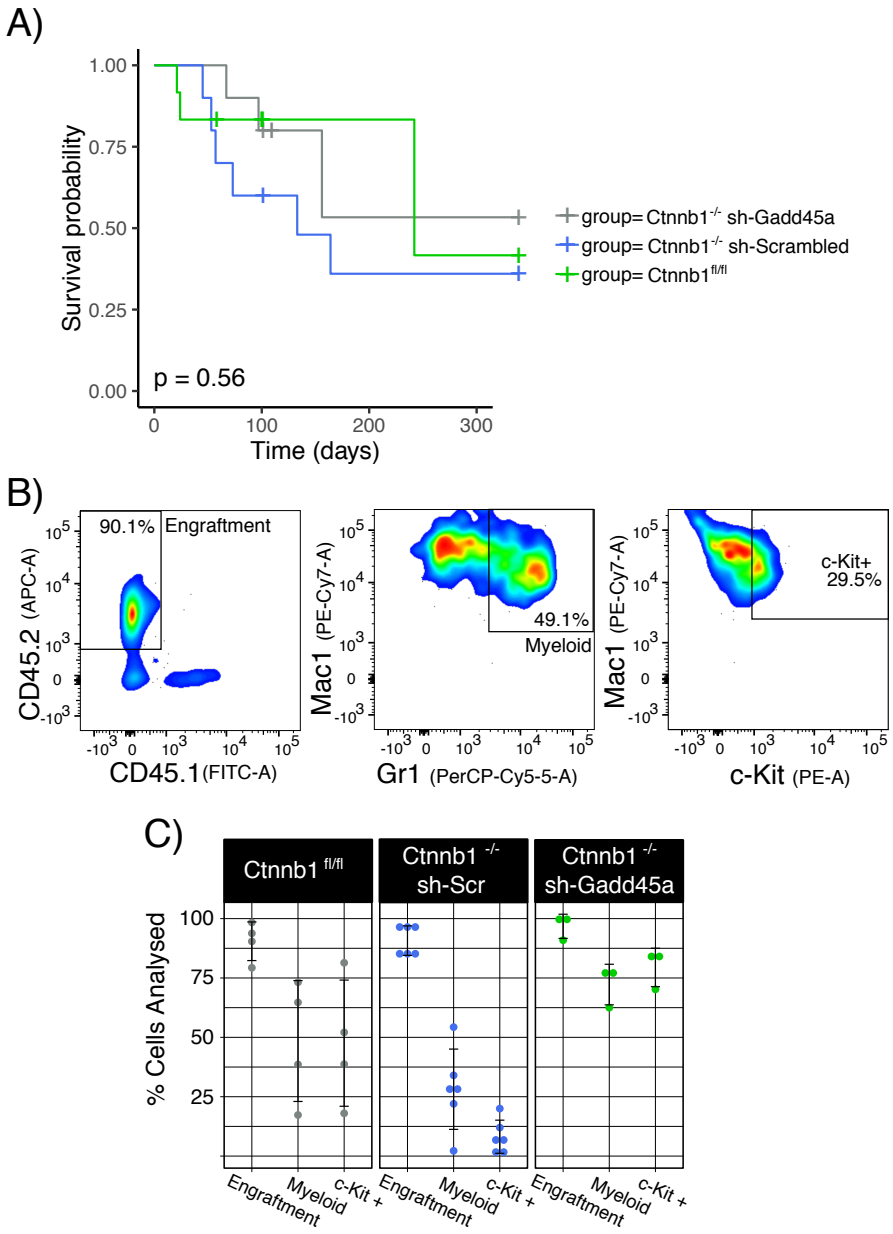

Figure 5.6: β-catenin KO cells showed no significant difference in two transplant experiments. A) Survival curves for groups in shRNA study indicated by the key (right). Censored values are indicated by a vertical line. 10 mice were transplanted for each group across 2 independent experiments. B) Representative FACS analysis showing myeloid leukaemia arising from β-catenin deleted GMP^{ME} cells. C) Percentage of FACS populations in bone marrow of leukaemic mice for each experimental group.

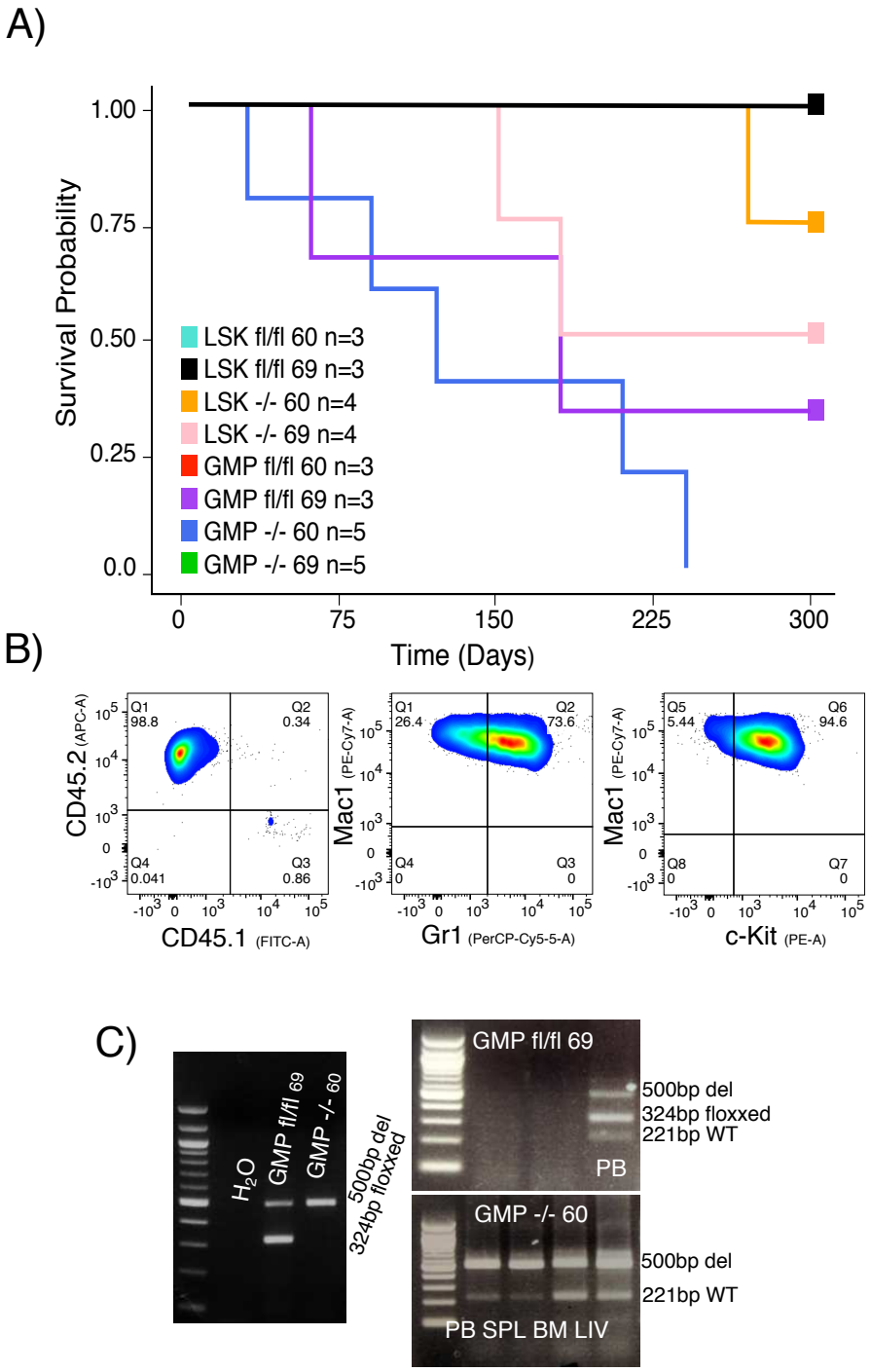

Figure 5.7: Survival curve for transplant of cells used for sequencing in this study A) Survival curve for pre-LSCs used in this study. Groups which are not visible all survived past 300 days. B) Representative example FACS analysis showing engraftment of β-catenin deleted Pre-LSCs cells. C) PCR to confirm KO is still complete following transplant.

5.3.6 *Gadd45a* overexpression in LSK^{ME} eliminates leukaemia in vivo

Interestingly, *Gadd45a* overexpression was found to eliminate leukaemia in LSK derived cells (Figure 5.8D). As previously described, GMP derived leukaemia was shown to have a greater latency than LSK (Figure 5.3A & 5.8D). Interestingly, compared with LSK^{ME}, increased variability in GMP^{ME} is also seen in FACS analysis (Figure 5.3E and Figure 5.8F) with surprisingly less infiltration into organs and peripheral blood and a reduction in gr1 and ckit positive cells in all tissues (Figure 5.3E and Figure 5.8F). It was also apparent that consistently higher overexpression was achieved in LSK^{ME} than in GMP^{ME} (Figure 5.2A and Figure 5.8A). I conclude that the leukaemia eliminating effects of high levels of *Gadd45a* overexpression are not specific to GMP^{ME}.

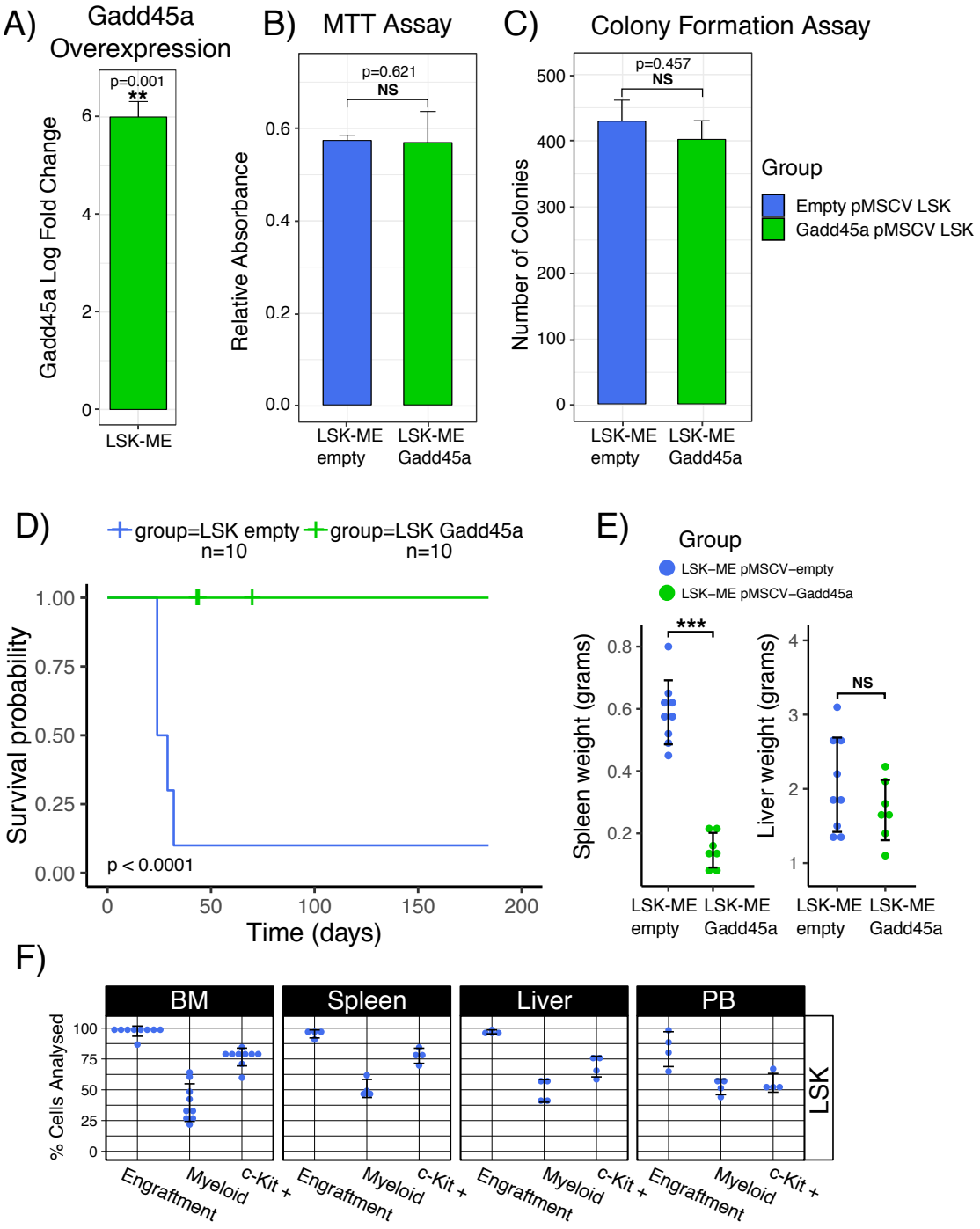

Figure 5.8: *Gadd45a* overexpression eliminates LSK derived MLL-ENL leukaemia
A) RT-qPCR showing overexpression of Gadd45a. B) MTT assay for pMSCV empty vector or pMSCV-*Gadd45a* transduced cells. C) Colony forming assay of pMSCV empty vector or pMSCV-*Gadd45a* transduced cells after antibiotic selection. D) Survival curves for transplant of both LSK^{ME} with *Gadd45a* overexpression (green) or empty pMSCV vector (blue). Censored values are indicated by a vertical line. E) Spleen and liver weights for each group at time of sacrifice. Spleen weight was significantly greater in control experiments (p= 7.012E-08). Liver weight was not significantly different (p=0.237). F) Immunophenotyping of leukaemia cells from GMP^{ME} cells transduced with pMSCV empty vector. E) Dotplot showing a summary of FACS analysis across samples.

5.4 Discussion

The aim of this chapter was to investigate whether *Gadd45a* is a key mediator of GMP-origin MLL-ENL leukaemia development. *Gadd45a* overexpression did not show a significant phenotype in vitro (Figure 5.2B & C). It is interesting to note that from 3 independent experiments, *Gadd45a* overexpression was markedly higher in LSK^{ME} than GMP^{ME} (Figure 5.2A and Figure 5.8A). In these experiments the same conditions for infection and the same batch of virus was used. Therefore, one could speculate that GMP^{ME} is more susceptible to high expression of *Gadd45a* and these cells are lost before the end of antibiotic selection. Another explanation could be that GMP^{ME} derived cells are in general more resistant to retroviral integration.

Whilst no significant phenotype is detectable in vitro (Figure 5.2B & C, Figure 5.8B & C), *Gadd45a* eliminated leukaemia in vivo (Figure 5.3A & 5.8D). Although both LSK and GMP derived leukaemia was eliminated by *Gadd45a* overexpression, there are significant differences in disease latency (Figure 5.3A compared to 5.8D) and organ infiltration (Figure 5.3E compared to 5.8F) which has been demonstrated before (Krivtsov et al. 2013). Organ size was markedly increased in GMP^{ME} compared with LSK^{ME} (Figure 5.3C compared to 5.8E), however the values are not normalized by animal weight so this may be due to ageing. More work is required to underpin the molecular mechanism for disease elimination via *Gadd45a* overexpression in vivo and its drug resistance properties and to link β-catenin depletion to its transcription. Moreover, it would be interesting to see if the effects of *Gadd45a* overexpression is specific to MLL-ENL leukaemia or if other fusions or mutations would also be affected. Notably, retroviral overexpression was not comparable to physiological changes in expression (Figure 5.2A & 5.8A compared to Figure 4.3). One could argue that we may see a difference in cell-of-origin effect if we were to overexpress *Gadd45a* at more physiological levels. Further experiments could be improved by titrating the virus in combination with using a vector with a weaker promoter to obtain a similar level of overexpression to that seen following β-catenin deletion. However, the fact that *Gadd45a* overexpression eliminates leukaemia in LSK^{ME} does not discredit the hypothesis that *Gadd45a* may confer β-catenin dependency specificity in GMP^{ME}.

As there was a significant reduction in viability after β-catenin knock-out (Figure 5.4B, C & 5.5), I decided to transduce cells with *Gadd45a* shRNA prior to β-catenin knock-out. The effect of β-catenin deletion was also markedly worse in GMP^{ME} than LSK^{ME} (Figure 5.5). However, it is possible that this ordering may be critical to the phenotype produced by *Gadd45a* knock down in vitro. Moreover, this result was unexpected as previously, no significant differences

in cell cycle or proliferation were reported following β-catenin KO (Siriboonpiputtana et al. 2017). Interestingly, knocking down *Gadd45a* mildly increases proliferation of GMP^{ME} cells, and some of this advantage was retained after β-catenin KO (Figure 5.4C) suggesting downregulation of *Gadd45a* unlocks greater proliferative capacity, consistent with its roles in leukaemia (Perugini et al. 2013).

Despite the exaggerated phenotype observed in response to β-catenin deletion in vitro (Figure 5.5), β-catenin KO GMP^{ME} LSCs and Pre-LSCs unexpectedly and repeatedly gave rise to leukaemia in vivo (Figure 5.6 & 5.7). This suggests that while originally sensitive to β-catenin deletion, mutations or epigenetic states may be acquired in vitro which allow GMP^{ME}s to overcome their β-catenin requirement in vivo. It could be argued that development of tolerance to β-catenin KO may have been encouraged via leakiness of Cre-recombinase, which is demonstrated in Figure 3.1A and Figure 5.7C. As discussed in section 4.4, this could be due to transcriptomic acclimatisation. Because of the inconsistencies in β-catenin requirement for leukaemia development from GMP^{ME} with Siriboonpiputtana et al. (2017), as can be seen in Figure 5.6 & 5.7, I was unable to test if *Gadd45a* knockdown could rescue the β-catenin deletion phenotype in GMP^{ME}.

To determine if the phenotypic differences seen between leukaemias with distinct cell-of-origin could be due to selective mutations, Siriboonpiputtana et al. (2017) investigated genic and intergenic Single Nucleotide Polymorphisms (SNPs). This analysis showed that the number of SNPs between cell lines of different cellular origins wasn't significantly different to the number of SNPs found between cell lines derived from different FACS experiments and MLL-ENL transductions. However, this analysis was limited in that we did not have the original genome sequence of the cells-of-origin preceding transformation. Additionally, the functional consequences of individual SNPs were not explored. Genomic DNA used for SNP analysis in the Siriboonpiputtana et al. (2017) study was taken from the batch of cells that I had previously cultured for RNA-Seq.

Cultured cell lines are the lifeblood of cancer research leading to many cornerstone discoveries. However, whilst original studies had shown that cell lines were relatively stable in vitro, recent research reveals that cancer cell lines and even short-term cultured primary tumour cells are highly genetically and transcriptomically unstable (Ben-David et al. 2017; Ben-David et al. 2018; Fong et al. 2015). Meta-analysis comparing cancer cell lines with that of primary patient samples found that there was often more differential expression between cell lines than cancer type. This suggests that over time, culturing conditions can lead transcriptional programs

away from those of their primary origin (Gillet et al. 2011). Recently, sophisticated bar-coding studies of single clones demonstrated how positive selection leads to cell line evolution fine-tuned to culture conditions (Ben-David et al. 2018). Indeed, the rapid adaptive nature of leukaemia cells was recently exploited to create an in vitro model of drug resistance through exposure to stress (Fong et al. 2015). These studies highlight the need for greater genomic characterization of cell lines and give some insight behind the growing irreproducibility crisis in cancer research.

In cancer, genetic or transcriptional dependency doesn't have a binary definition, rather dependency may be scored by effect (Tsherniak et al. 2017). Whilst β-catenin has recurrently been shown to be critical for MLL-fusion transformation, the degree of dependency shows variation (Siriboonpiputtana et al. 2017; Gandillet et al. 2011; Griffiths et al. 2010; Siapati et al. 2011; Wang et al. 2010a; Yeung et al. 2010; Ng et al. 2014). With the exception of our paper (Siriboonpiputtana et al. 2017), β-catenin dependency in these studies can be interpreted as a significant reduction of Leukaemia Initiating Cells (LICs) rather than complete leukaemia abolishment. Nevertheless, even if there is a 1000-fold reduction in LIC it is still possible that leukaemia will still arise with greater latency. For example, Ng et al. (2014) showed that leukaemia would arise from MLL-AF9 transformed LSK β-catenin deleted cells with a similar latency to β-catenin intact cells if over 1×10^4 cells were transplanted. It is noteworthy that in this case, β-catenin deletion was induced in vivo whereas in my experiments and (Siriboonpiputtana et al. 2017) deletion was induced in vitro. However, unexpectedly in both of my experiments, β-catenin deleted GMP^{ME}s were the first group to show signs of leukaemia (Figure 5.6A and 5.7A). Additionally, the failure of floxed cell lines to induce leukaemia seen in Figure 5.7A highlights possible technical issues with this particular experiment. Genomic deletion of retrovirally introduced transgenes has been frequently observed (Migliaccio et al. 2000), it is therefore possible that failure to induce leukaemia could be due to decreases in MLL-ENL expression over time.

Furthermore, analysis including both samples from (Siriboonpiputtana et al. 2017) and RNA-Seq produced independently for this study showed that the effect of β-catenin deletion is largely equivalent in LSK^{ME} and GMP^{ME} (see subsection 3.3.7 & 3.3.8, Figure 3.7 & 3.8). Therefore, we may expect that the transcriptomic and epigenomic effects of at least β-catenin KO should remain equivalent regardless of the conflicting dependency observed. Considering the potential implications of the findings in this chapter, I proceeded to study the epigenome of these cells.

Chapter 6

Integrative ChIP-Seq analysis of cell-of-origin specific β-catenin conditional cell lines

6.1 Introduction

Abnormal DNA methylation and chromatin modification are crucial features of CSCs, in addition to mutational and transcriptional profiling, increasingly, we are exploring the epigenomes of cancer cells in our search for therapeutic strategies targeting CSCs (Mohammad et al. 2019). Chromatin precipitation coupled with next generation sequencing (ChIP-Seq) allows us to sequence DNA which is bound by specific proteins. β-catenin has been shown to associate with epigenetic factors such as histone modifying enzymes (see section 1.7 and Figure 1.6). To elucidate mechanisms of transcriptional control mediated by β-catenin and explore cell-of-origin differences, I performed ChIP-Seq and analyses for key histone modifications. Histone modifications were explicitly chosen for their known effects on transcription: H3K4 tri-methylation for active promoters, H3K4 mono-methylation H3K27 acetylation for enhancers and H3K27 tri-methylation for silent promoters. Specifically, H3K4me3 is crucial for the initiation of transcription whilst H3K27me3 inhibits transcription elongation.

Before the biological questions of this chapter can be addressed it is paramount that we assess the ChIP-Seq quality in terms of read quality, mappability, duplication and enrichment over noise. Following confirmation of successful ChIP-Seq, to ensure antibody specificity, I should observe histone modification patterns congruent with the literature. In a similar vein, I seek the expected correlations between histone modification and RNA-Seq data. Furthermore, I will investigate if based on these histone modification data, I can predict cis-regulatory elements with known characteristics. Once I have characterised these data, I will quantify changes in histone modification at cis-regulatory elements dependent on cell-of-origin and β-catenin status. I will determine whether there are common biological themes of deregulated promoters and enhancers; pathways of coordinated regulation or motifs which may shed light on the transcription factors involved. Finally, I will attempt to correlate any of these changes with direct β-catenin binding.

6.2 Methods

6.2.1 ChIP-Seq Generation

With the aim to maintain compatibility with RNA-Seq, I performed ChIP on cell pellets which had been fixed and frozen at -80°C on the same day that cells were collected for RNA extraction. A large number of aliquots were fixed as occasionally multiple attempts were required to achieve comparable sonication profiles within the same experiment. Consistent sonication was found to be the most crucial step of the assay to control non-specific DNA pull-down. As thoroughly described in section 2.8, I performed ChIP for: H3K4me3, H3K4me1, H3K27ac and H3K27me3.

Before library preparation and sequencing, I first tested for enrichment using positive and negative control regions by ChIP-qPCR. Percentage input, the proportion of total input DNA pulled down by the ChIP assay, was calculated using the ΔCT method. ChIP-qPCR reactions were set up in triplicate to adjust for technical variance. Following ChIP-qPCR validation, I sequenced samples at low coverage (4-5 million reads) to determine enrichment at the genomic level; with the exception of H3K4me3 which was sequenced in a single run. Following satisfactory results, I re-sequenced libraries with the aim to obtain approximately 20 million reads per sample. For each histone modification ChIP, sequencing libraries from all samples were multiplexed together in the same lane to avoid introducing technical batch effects. Details of all sequencing runs are shown in Table 6.1.

Table 6.1: ChIP-Seq samples generated for this study

Cell-of-Origin	Factor	β-catenin	Rep	Run ID(s)	Date(s)
LSK	input	fl/fl	69	241	Jan-16
LSK	input	-/-	69	242	Jan-16
GMP	input	fl/fl	69	243	Jan-16
GMP	input	-/-	69	244	Jan-16
LSK	H3K4me3	fl/fl	60	229	Jan-16
LSK	H3K4me3	-/-	60	230	Jan-16
LSK	H3K4me3	fl/fl	69	233	Jan-16
LSK	H3K4me3	-/-	69	234	Jan-16
GMP	H3K4me3	fl/fl	60	231	Jan-16
GMP	H3K4me3	-/-	60	232	Jan-16
GMP	H3K4me3	fl/fl	69	235	Jan-16
GMP	H3K4me3	-/-	69	236	Jan-16
LSK	H3K4me1	fl/fl	60	239, 279, 452	Jan16,Apr16,May16
LSK	H3K4me1	-/-	60	277, 450	Apr16,May16
LSK	H3K4me1	fl/fl	69	273, 446	Apr16,May16
LSK	H3K4me1	-/-	69	274, 447	Apr16,May16
GMP	H3K4me1	fl/fl	60	240, 280, 453	Jan16,Apr16,May16
GMP	H3K4me1	-/-	60	278, 451	Apr16,May16
GMP	H3K4me1	fl/fl	69	275, 448	Apr16,May16
GMP	H3K4me1	-/-	69	276, 449	Apr16,May16
LSK	H3K27me3	fl/fl	60	723, 770, 790	Apr17,May17
LSK	H3K27me3	-/-	60	724, 771, 791	Apr17,May17
LSK	H3K27me3	fl/fl	69	727, 774, 794	Apr17,May17
LSK	H3K27me3	-/-	69	728, 775, 795	Apr17,May17
GMP	H3K27me3	fl/fl	60	725, 772, 792	Apr17,May17
GMP	H3K27me3	-/-	60	726, 773, 793	Apr17,May17
GMP	H3K27me3	fl/fl	69	729, 776, 796	Apr17,May17
GMP	H3K27me3	-/-	69	730, 777, 797	Apr17,May17
LSK	H3K27ac	fl/fl	60	731, 778, 798	Apr17,May17
LSK	H3K27ac	-/-	60	732, 779, 799	Apr17,May17
LSK	H3K27ac	fl/fl	69	735, 782, 802	Apr17,May17
LSK	H3K27ac	-/-	69	736, 783, 803	Apr17,May17
GMP	H3K27ac	fl/fl	60	733, 780, 800	Apr17 May17
GMP	H3K27ac	-/-	60	734, 781, 801	Apr17,May17
GMP	H3K27ac	fl/fl	69	737, 784, 804	Apr17,May17
GMP	H3K27ac	-/-	69	738, 785, 805	Apr17,May17

6.2.2 Correlating ChIP-Seq with RNA-Seq

Initial processing steps of ChIP-Seq analysis are detailed in section 2.9. ChIP heatmaps were generated using the R package EnrichedHeatmap v1.16.0 (Gu et al. 2018). To simplify these data for visualisation, replicates for each condition were merged into an average value. To quantify the relationship connecting gene expression and individual histone modifications, I used countOverlaps from the GenomicRanges v1.38.0 package (Lawrence et al. 2013) to count the number of reads falling at the promoter or gene body. To ensure the RNA-Seq and ChIP-Seq were annotated consistently, coordinates for genomic features were created using the same annotation file used for RNA-Seq processing. To ensure that there was a single ChIP-Seq value for each histone mark for each gene, in this analysis I defined promoters as 1 kb upstream and 2 kb downstream of the most upstream TSS of each gene. Using base R functions, I calculated the Spearman's correlation between normalised gene expression and ChIP-Seq tags per million mapped reads at the promoter or gene body. To perform PCA and calculate the global Pearson's correlation between ChIP-Seq samples I used csaw v1.20.0 (Lun and Smyth 2016) to count reads falling into 10 kb non-overlapping bins across the entire genome. Counts were normalised using trimmed mean of M-values (TMM) normalisation.

6.2.3 Regulatory state segmentation

ChromHMM (Ernst and Kellis 2017) uses a multivariate Hidden Markov Models (HMM) to learn and characterise patterns in genomic data. Markov models are suited to this task, as the regulation of transcription is fundamentally a stochastic process. The first step in the process required that aligned reads are converted to binarised data (0 or 1) depending on the presence or absence of ChIP-Seq tags above a threshold within bins of pre-determined length. I trained ChromHMM v1.20 (Ernst and Kellis 2017) to find 4-16 chromatin states based on the four histone modifications in 200 bp bins; after training ChromHMM to discover up to 16 states, I decided that 6 states were sufficient to segment the genome into non-redundant functional units.

To compare chromatin states between conditions, I split the segments back into 200 bp non-overlapping bins. Slight shifts of the boundaries of regulatory elements which could simply due to fractionation differences may be falsely detected as chromatin state transitions. Attempting to eliminate these false positives, I merged transitions where there was a break of under 400 bp. Chromatin states are predicted with ChromHMM for each sample individually.

To look at transitions between chromatin states depending on experimental conditions I took only chromatin states which were consistent between replicates. I then compared consistent states across replicates for each 200 bp bin and represented transitions as chord diagrams using the circlize v0.4.9 package (Gu et al. 2014). I plotted heatmaps for ChromHMM states using EnrichedHeatmap v1.16.0 (Gu et al. 2018).

6.2.4 Gene-centric quantitative differential ChIP-Seq analysis

The most widespread pipelines for ChIP-Seq analysis depend on characterizing enriched regions, or "peaks", to call high confidence DNA-protein binding events. Following peak calling, we may qualitatively or quantitatively compare peak regions between conditions. In comparison to transcription factor ChIP-Seq, histone modification data rarely show discrete patterns especially when observing heterogeneous cell populations. Histone modification differences between experimental conditions are more likely to show continuous changes in signal and differing spatial deposition patterns. Using peaks to identify quantitative differences between conditions in predominantly continuous data reduces the resolution for differential signal detection (Lun and Smyth 2016).

Considering this, I chose to use analysis workflows provided in the csaw v1.20.0 package (Lun and Smyth 2016), which can be used to quantify reads in sliding windows and calculate significantly differentially bound regions using edgeR v3.28.0 (Robinson et al. 2010). After calculating window-based statistics, windows can be grouped based on areas of interest, such as genes or promoters. This allows us to keep the resolution of observed differences, while maintaining error control. Firstly, I counted reads into windows between 150 bp and 2 kb. Next, I counted 10 kb bins across the genome to determine global background. I filtered for windows with a read count at least 2 log fold above the global background signal. To account for systematic differences in enrichment between samples I applied trimmed mean of M-values (TMM) normalisation for enriched regions. Differential expression analysis for ChIP-Seq between LSK^{ME} and GMP^{ME} or "celltype" was performed using a design model to account for experimental replicate: $\sim$*experiment+celltype*. Differential expression analysis for ChIP-Seq between β-catenin floxed and β-catenin deleted or "genotype" was performed for LSK^{ME} and GMP^{ME} separately using a design model to account for experimental replicate: $\sim$*experiment+genotype*.

To group windows of different sizes, I used the consolidateWindows function from csaw. Windows were aggregated over five different region categories: 1. "Genes" Gene-bodies minus the promoter region. 2. "Naïve" regions created by merging nearby windows of high enrichment. 3. "peaks" MACS2 peaks. 4. "Promoter by gene" regions 2 kb upstream and 1 kb downstream of the first annotated TSS. 5. "Promoter by transcript" regions 2 kb upstream and 1 kb downstream of every annotated TSS. To identify common themes amongst differentially bound regions I subjected the genes associated with significantly differentially modified regions to pathway analysis with clusterProfiler v3.14.3 (Yu et al. 2012). Ensembl gene identifiers were converted to Entrez gene IDs with Ensembl biomaRt v2.42.0 (Durinck et al. 2009). To create genomic coverage tracks for visualisation, I used deepTools bamCoverage v2.5.4 (Ramírez et al. 2014; Ramírez et al. 2016), the signal was normalised using a scale factor. The scale factor was calculated by first multiplying the total number of reads by the TMM normalisation factor calculated with csaw, 1×10^6 was then divided by this number. Coverage tracks were plotted using functions from rtracklayer v1.46.0 (Lawrence et al. 2009) and Gviz v1.30.0 (Hahne and Ivanek 2016).

6.2.5 Enhancer Analysis

Putative enhancers were crudely defined as H3K4me1 peaks present in both replicates for each condition. To filter out promoters, peaks overlapping 2 kb upstream and 1 kb downstream of TSSs were removed. Putative enhancer regions were then filtered for those co-localised with H3K27ac peaks to identify active enhancers. For identifying quantitative histone modification differences at putative enhancers, I used DiffBind v2.14.0 (Stark and Brown 2011), which uses total count numbers for specified regions, such as peaks or promoter regions. To look for de novo enhancers I also compared peak locations using DiffBind.

6.2.6 Motif Enrichment

Histone ChIP binding sites as opposed to transcription factor ChIP, are not usually appropriate for de novo motif discovery as it is unlikely that a single motif is present in most or all of the sequences. Therefore, I used the R package PWMEnrich v4.18.0 (Stojnic and Diez 2018) to discover the locations of known motifs within differentially bound regions. PWMEnrich scans input sequences and calculates their enrichment over expected occurrence using a precomputed background within mouse promoters.

6.2.7 Comparisons with published β-catenin ChIP-Seq

To elucidate whether the changes in ChIP-Seq signal could be directly linked to β-catenin occupancy I collected all available β-catenin ChIP-Seq data I could identify from public databases (GEO, ENA, SRA and DRA), these are listed in Table 6.2. I processed the ChIP-Seq data as previously described in section 2.8 and calculated quality control metrics. Peak locations from β-catenin ChIP-Seq experiments in mice with sufficient quality were compared with promoters exhibiting differential histone modification levels following knock-out. The significance of overlaps between differentially modified regions and β-catenin ChIP-Seq peaks was calculated using permutation tests with functions from regioneR v1.18.0 (Gel et al. 2016).

Table 6.2: Available published β-catenin ChIP-Seq data

GSE/SRA	Antibody	Species	Tissue	β-catenin Stimulation/ Activation	Citation
GSE63682	610154, BD Biosciences	mouse	ckit+ MLL-AF9	NA	(Fong et al. 2015)
GSE43565	bCatenin-FLAG	mouse	ESC	GSK3 inhibitor	(Zhang et al. 2013)
GSE39837	bCatenin-BIOTIN B-cat Cell Signalling 4176	mouse	Mesenchymal nephron progenitors	GSK3 inhibitor (BIO)	(Park et al. 2012)
GSE35213	unknown	mouse	Transgenic hepatocytes	Apc knock-out	(Gougelet et al. 2014)
GSE84456*	H102; Santa Cruz	mouse	Colon crypt	NA	(Schuijers et al. 2015)
SRP002018	610154, BD Biosciences	human	HCT116	known APC mutant	(Bottomly et al. 2010)
GSE64758	β-catenin (Santa Cruz sc7199)	human	H1 hESCs	Wnt3a	(Estarás et al. 2015)
GSE58476	β-catenin (Santa Cruz sc7199)	human	Human embryonic stem cells	GSK3 inhibitor (CHIR)	(Funa et al. 2015)
GSE53927	β-catenin (Santa Cruz sc7199)	human	SW480 colon cancer	known APC mutant	(Watanabe et al. 2014)
GSE49199	9562 (Cell signalling technology	human	Primary regulatory T cells	GSK3 inhibitor (BIO)	(Loosdregt et al. 2013)
GSE117944	71–2700; ThermoFisher	human	K562	Wnt3a	(Shooshtarizadeh et al. 2019)

* Raw data not available.

6.2.8 β-catenin ChIP in constitutively active β-catenin cell lines

Attempts to achieve β-catenin ChIP enrichment were unsuccessful using the β-catenin floxed MLL-ENL transformed cell lines in this study. To overcome this and get insight into β-catenin binding in MLL-ENL transformed cells, I utilised a β-catenin activation model. I crossed mice harbouring an allele of *Ctnnb1* with LoxP sites flanking exon 3, with transgenic mice expressing Cre-ER and produced MLL-ENL transformed cells from their bone marrow using the RTTA method described in section 2.1. As explained in subsection 2.1.1, deletion of exon 3 of *Ctnnb1* results in expression of a constitutively active form of β-catenin.

Exon 3 of *Ctnnb1* was excised by treating cell lines with 20 nM 4-OHT. To detect presence of the truncated β-catenin protein I used western blot and immunofluorescence as described in section 2.4 & 2.10 respectively. Using anti-β-catenin antibodies from BD Biosciences (#610154) and Cell Signalling Technology® (#8480) and anti-TCF4/TCF7L2 antibody from Cell Signalling Technology® (#2569), I performed ChIP as described in section 2.8 followed by ChIP-qPCR. β-catenin binding sites were obtained from literature involving β-catenin ChIP and are shown in Table 6.3. The expression of a well known β-catenin transcriptional target, Cyclin D1 (*Cnnd1*) (Nusse 2019) was quantified by RT-qPCR in ETOH treated control cells and those expressing constitutively active β-catenin.

Table 6.3: β-catenin ChIP-qPCR primers

Primer	Orientation	Sequence (5' to 3')
Myc TSS	Forward	GTCACCTTTACCCCGACTCA
Myc TSS	Reverse	TCCAGGCACATCTCAGTTTG
Myc Enhancer	Forward	TCTTTGATGGGCTCAATGGT
Myc Enhancer	Reverse	TTCCCTTCACCTGATGAACC
Ccnd1 TSS	Forward	CCGGCTTTGATCTCTGCTTA
Ccnd1 TSS	Reverse	CGCGGAGTCTGTAGCTCTCT
Axin2 TSS	Forward	GAGCGCCTCTGTGATTGG
Axin2 TSS	Reverse	GCGAACGGCTGCTTATTTT

6.3 Results

6.3.1 Successfully enriched ChIP was obtained

Before analysis of histone modifications can proceed, quality assessment of these data should be performed to ensure that meaningful results may be extracted. Quality control steps are carried out both before and after sequencing. Prior to sequencing, I performed ChIP-qPCR for positive and negative control regions. Homeobox genes: Myeloid Ecotropic Viral Integration Site 1 (*Meis1*) and Homeobox A9 (*Hoxa9*), are targets of MLL fusions and therefore were used as positive controls of active promoters and enhancers.

Consistently, the active chromatin marks H3K4me3 and H3K27ac were found to be enriched at these promoters (Figure 6.1). H3K4me1 was shown to have enrichment for both the *Meis1* promoter and a known *Meis1* enhancer element. Cyclin Dependent Kinase Inhibitor 2A, *Cdkn2a*, also known as p16, was used as a positive control for repressed promoters and showed significant enrichment for the repressive histone mark H3K27me3. Notably, H3K27me3 was found at *Meis1*, however, following additional of ChIP-qPCR for *Hoxa9* promoter regions which showed little enrichment, it was determined that pull-down at the *Meis1* promoter was likely to be real signal rather than non-specific pull-down. Primers designed on a gene desert region were used as a negative control for all histone modifications.

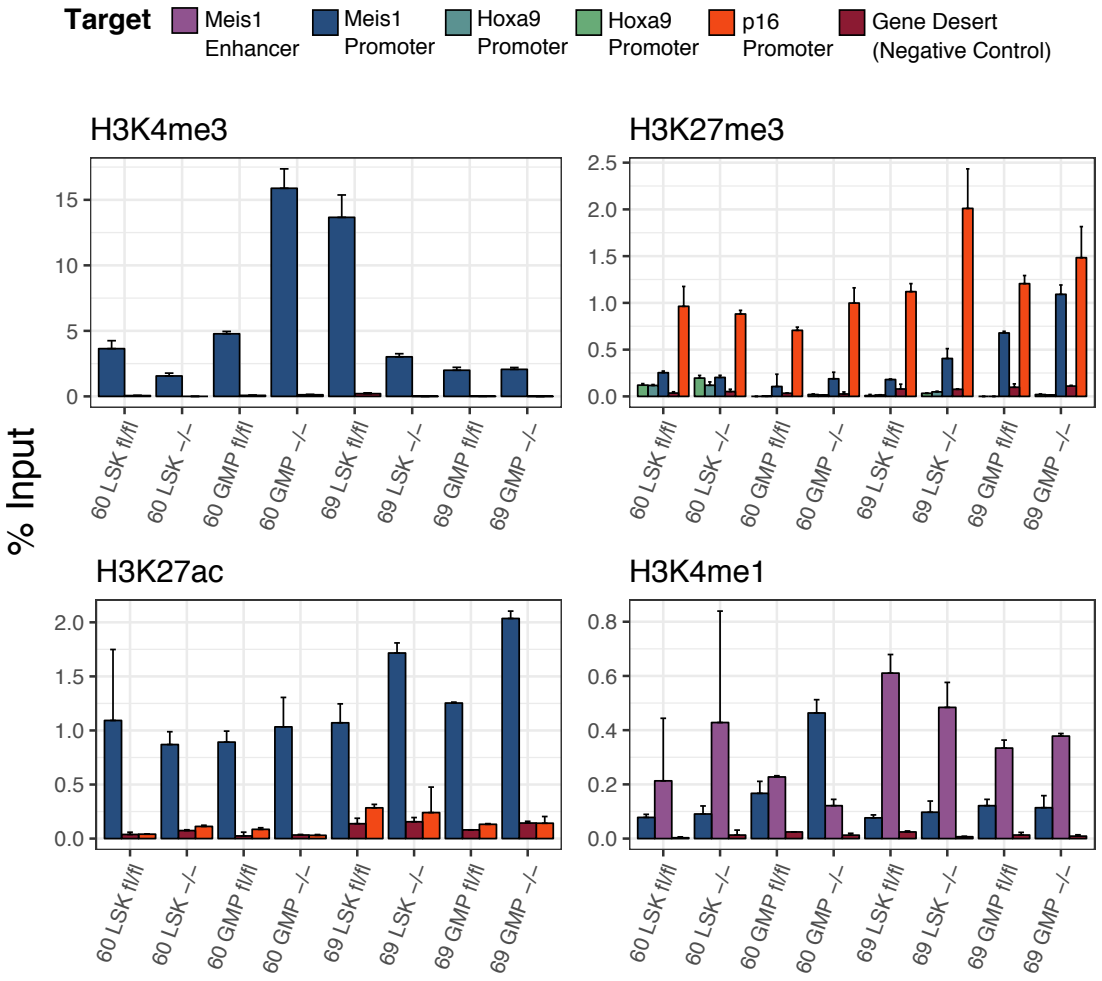

Figure 6.1: ChIP-qPCR for positive and negative control regions. Four bar charts showing the percentage of input chromatin pulled down by immunoprecipitation from each ChIP experiment which were sent for sequencing. The key at the top indicates the quantified target region.

6.3.2 Successfully enriched ChIP-Seq data were obtained

After receiving and processing ChIP-Seq reads through initial mapping and peak calling stages, I assessed the quality of ChIP-Seq data using ChIPQC v1.22.0 (Carroll et al. 2014), these quality metrics are listed in Table 6.4. Firstly, I found that through multiplexing error some samples received fewer reads than others, namely H3K4me1 for fl/fl replicate "60" LSK^{ME} and GMP^{ME}. However, the duplication rates for these samples were low and therefore following filtering, the number of usable reads was acceptable. The duplication rate varies from as low as 6% to as high as 43%, however this range is still acceptable within ENCODE standards (ENCODE Project Consortium 2019).

The percentage of reads in peaks (RiP%) is also sufficiently high to indicate successful enrichment in all samples. The SSD score is the standard deviation of read pileup along the genome. In a successfully enriched ChIP-Seq library, a small proportion of the genome should account for most the of signal pileup giving a high SSD score. It can be seen in Table 6.4 that all ChIP samples have a greater SSD than input DNA and more punctuate marks (H3K27ac and H3K4me3) have a greater SSD than broader marks (H3K27me3 and H3K4me1). Once quality of enrichment and reads was confirmed, I then compared the global patterns of these histone modifications with their published characteristics to verify antibody specificity.

Table 6.4: Quality control results for ChIP-Seq generated for this study

Origin	Factor	βc	Rep	Read Pairs	Dup%	Frag Len	Rel CC	SSD	RiP%
LSK	input	fl/fl	69	18449538	11.1	247	2.2	0.9	NA
LSK	input	-/-	69	17013527	10.7	267	13	0.9	NA
GMP	input	fl/fl	69	19299982	12.1	247	1.8	0.9	NA
GMP	input	-/-	69	18371420	10.9	247	1.9	0.9	NA
LSK	H3K4me3	fl/fl	60	16449673	10.4	249	1.6	4.3	45
LSK	H3K4me3	-/-	60	16170088	11	247	1.7	4.6	50
LSK	H3K4me3	fl/fl	69	19253018	7.6	247	1.6	7.5	68
LSK	H3K4me3	-/-	69	16160071	10.4	250	1.6	5.8	60
GMP	H3K4me3	fl/fl	60	16888912	7.6	248	1.6	6.9	66
GMP	H3K4me3	-/-	60	14058856	8.8	247	1.7	5.8	62
GMP	H3K4me3	fl/fl	69	15618923	10.9	250	1.6	7.5	73
GMP	H3K4me3	-/-	69	16193481	9	247	1.6	6.2	60
LSK	H3K4me1	fl/fl	60	8510117	8	250	1.5	1.4	38
LSK	H3K4me1	-/-	60	40208082	30.4	249	3.5	2.7	57
LSK	H3K4me1	fl/fl	69	38251913	24.8	249	2.3	2.9	61
LSK	H3K4me1	-/-	69	34279331	25	247	2.7	2.7	61
GMP	H3K4me1	fl/fl	60	9519836	5.7	247	1.7	1.4	39
GMP	H3K4me1	-/-	60	27513021	40.5	247	4.4	2.4	50
GMP	H3K4me1	fl/fl	69	25017964	26.1	248	4	2	42
GMP	H3K4me1	-/-	69	33049539	37.4	247	6	2.2	41
LSK	H3K27me3	fl/fl	60	24306814	43.8	247	5.9	2.4	24
LSK	H3K27me3	-/-	60	29062920	38.2	247	7.2	2.2	27
LSK	H3K27me3	fl/fl	69	30501030	16.9	247	2.8	1.7	17
LSK	H3K27me3	-/-	69	31246898	43.8	249	2.2	2	18
GMP	H3K27me3	fl/fl	60	35660713	13.1	248	6.7	1.3	11
GMP	H3K27me3	-/-	60	28725499	40.1	247	5.1	2.2	22
GMP	H3K27me3	fl/fl	69	27135151	16	247	6.2	1.7	18
GMP	H3K27me3	-/-	69	24394960	17.4	247	6.6	1.6	14
LSK	H3K27ac	fl/fl	60	29834972	7.6	247	1.8	4.4	22
LSK	H3K27ac	-/-	60	33963197	9.3	248	2	3.7	19
LSK	H3K27ac	fl/fl	69	29881504	11	249	1.8	5.1	24
LSK	H3K27ac	-/-	69	25277815	12.3	247	1.9	4.2	26
GMP	H3K27ac	fl/fl	60	37007168	8.3	247	1.8	6.2	27
GMP	H3K27ac	-/-	60	28407232	12.2	247	1.8	4.9	24
GMP	H3K27ac	fl/fl	69	25508436	14.4	249	1.9	6.4	27
GMP	H3K27ac	-/-	69	28712079	15.9	249	1.4	7.2	32

6.3.3 Histone modification data demonstrates expected patterns at the TSS

The histone modifications studied in this chapter have well established patterns or profiles in relation to gene architecture and gene transcription (Heintzman et al. 2007; Barski et al. 2007; Zhou et al. 2011; ENCODE Project Consortium and others 2012). These patterns are demonstrated in Figure 6.2; average signal across all samples for each histone modification are plotted for 2 kb upstream of the TSS and 10 kb downstream. To allow assembly of protein complexes which facilitate transcription, nucleosomes are displaced creating an accessible nucleosome free region (NFR) at enhancers and promoters, flanked by their relevant histone modifications. H3K4me3 is found at active promoters; just upstream of the TSS resides a NFR causing a drop in H3K4me3 signal. Nucleosomes just downstream of the TSS are generally highly localised which appears as high H3K4me3 signal in active promoters often spreading far into the gene body of highly expressed genes. Notably, H3K4me3 is consistently found at many inactive promoters (Guenther et al. 2007).

Although associated with positive regulation of transcription, H3K4me-di and mono are progressively distanced from the TSS of active genes (Barski et al. 2007). The width of the dip in H3K4me1 signal surrounding the TSS correlates with higher gene expression signifying greater displacement with H3K4-tri methylation. H3K27ac is found at active enhancers and also at active promoters showing a similar pattern to H3K4me3 but more punctuate in appearance. H3K27me3 binding is low at active promoters and broad regions of modest enrichment are seen at silenced promoters. In this section I have confirmed that the histone modification ChIP-Seq in this study demonstrates profiles supported by the literature.

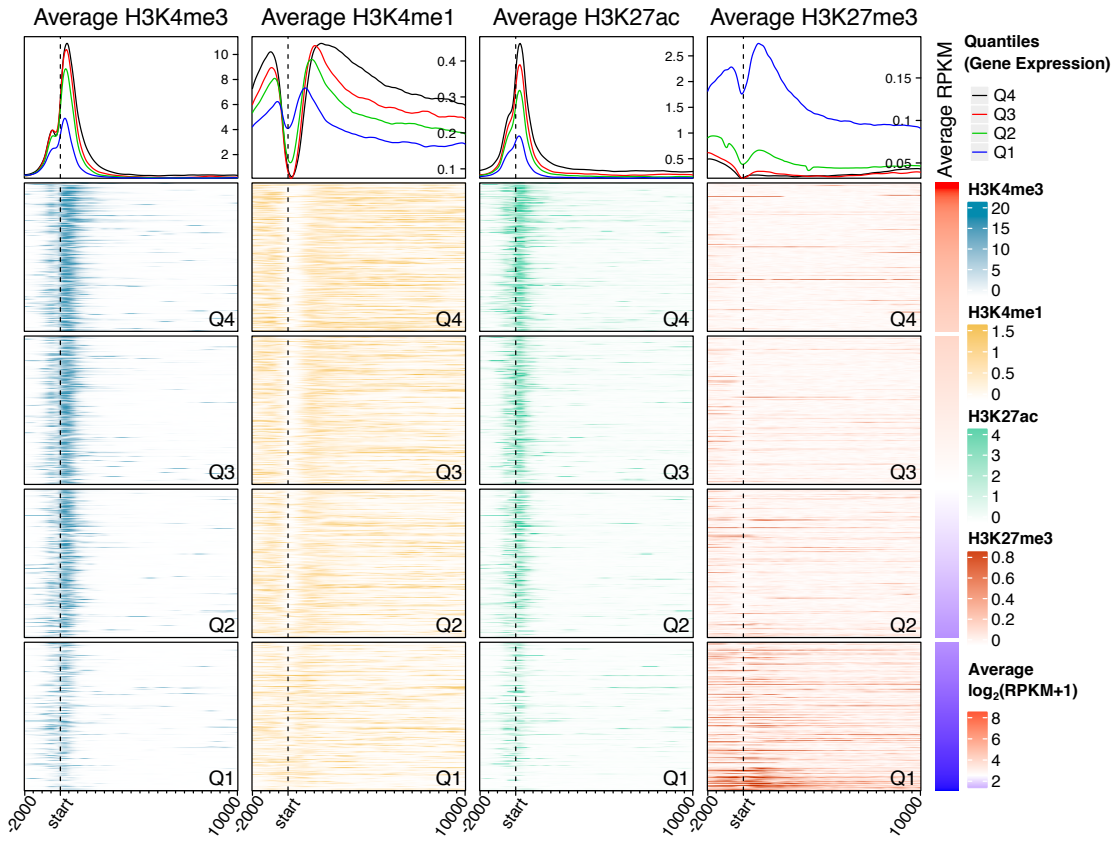

Figure 6.2: Heatmaps and average profiles for all histone marks from the TSS to 10 kb into the genebody Approximately 12,000 genes are displayed in this heatmap, genes were selected with at least one RNA-Seq count per million reads in at least two samples. Each horizontal line represents a single promoter, the intensity of colour indicates the read density at that location. The location of the TSS is indicated by "start" on the scale at the bottom and by a dotted vertical line. Genes/promoters are ordered by average RPKM gene expression. To simplify each gene to one promoter, the most upstream TSS of each gene was chosen, despite that this may not necessarily be the most active TSS. To show a generalised pattern, ChIP signal is plotted as an average value for all samples in the study. Gene regions are split into 4 quartiles labelled Q1-4 based on their average RNA-Seq RPKM value with the most highly expressed genes in Q4 and the lowest in Q1. Above each histone modification is a line graph depicting the average signal profile of each quartile represented by a different colour line as indicated by the key (top right).

6.3.4 Histone modification signal at gene loci correlates with corresponding gene expression data

To further illustrate the relationship between histone marks and gene expression, I calculated the Spearman's correlation between all individual ChIP-Seq and RNA-Seq samples. RPKM values were calculated for the average transcript length (the sum of exons) or the promoter and gene body (including introns, exons and Untranslated Regions (UTRs)) for RNA-Seq and ChIP-Seq respectively, the resulting correlation matrix is shown as a heatmap in Figure 6.3.

Firstly, as expected, all histone modifications group separately by hierarchical clustering. Clusters do not clearly resolve by cell-of-origin or genotype and have a short branch distance. Both in the promoter (Figure 6.3A) and gene body (Figure 6.3B), active marks H3K4 tri-methylation and H3K27 acetylation are positively correlated with transcription. On the other hand H3K27 methylation is inversely correlated with transcription. H3K4 mono-methylation also shows a mildly negative correlation with transcription. The only mark which shows statistically weak correlations below the cut-off is H3K4me1 indicating a more complex relationship with gene expression. The greatest correlation in the matrices presented in Figure 6.3 is the positive relationship between H3K4me3 and H3K27ac. Out of all histone marks, H3K4me3 across the gene body correlates most closely to transcription (Figure 6.3B), however if only the promoter region is considered, H3K27ac has a mildly stronger correlation with transcription (Figure 6.3A). In this section I confirmed that our ChIP-Seq for histone modifications show the expected directional relationships with gene expression.

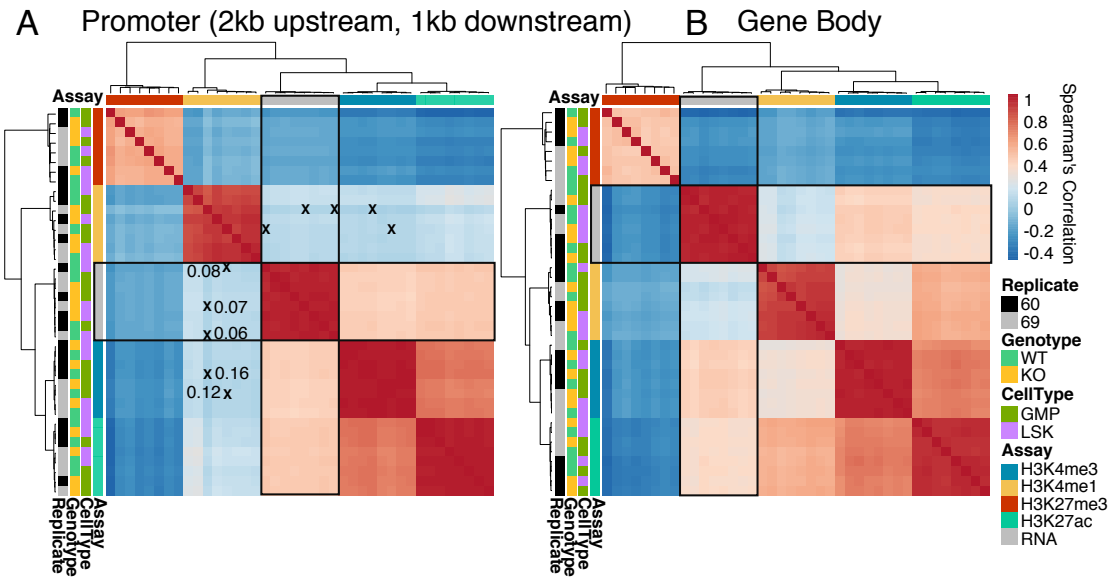

Figure 6.3: Correlation between RNA-Seq and ChIP-Seq Spearman's correlation between log RPKM normalised gene expression and counts at the promoter (A) or gene body (B). Each square represents the correlation between any two individual samples and its colour indicates the correlation value. Samples are ordered by hierarchical clustering, indicated by the dendrograms along the top and left sides. Note that the ordering is different in A and B. The colour block annotation on the top of the heatmaps and the innermost left indicate the assay. Three other factors, "celltype" (cell-of-origin), genotype and replicate are indicated by colour bars on the left. All correlations depicted are statistically significant (P <0.05) with exception of those marked with a cross, which are labelled with their actual P.value.

6.3.5 Variation and Correlation between ChIP-Seq samples

As previously demonstrated in Figure 6.3, genomewide PCA also shows that ChIP-Seq samples cluster by assay (Figure 6.4A). It is interesting to note that H3K27ac samples appear to cluster most closely with H3K4me1 (Figure 6.4A) possibly due to their coexistence outside of promoter regions. As highlighted in chapter 5, there are observable qualitative differences in phenotype between replicate samples. We therefore may speculate that there are differences at the chromatin level. We previously observed that there is greater transcriptomic variation amongst GMP^{ME} replicates than LSK^{ME}s (see subsection 3.3.1, Figure 3.2), this can also be seen in the PCA plot for ChIP-Seq samples in Figure 6.4B. Samples show some separation between different cells-of-origin, however within-group variation between replicates greater than treatment (Figure 6.4).

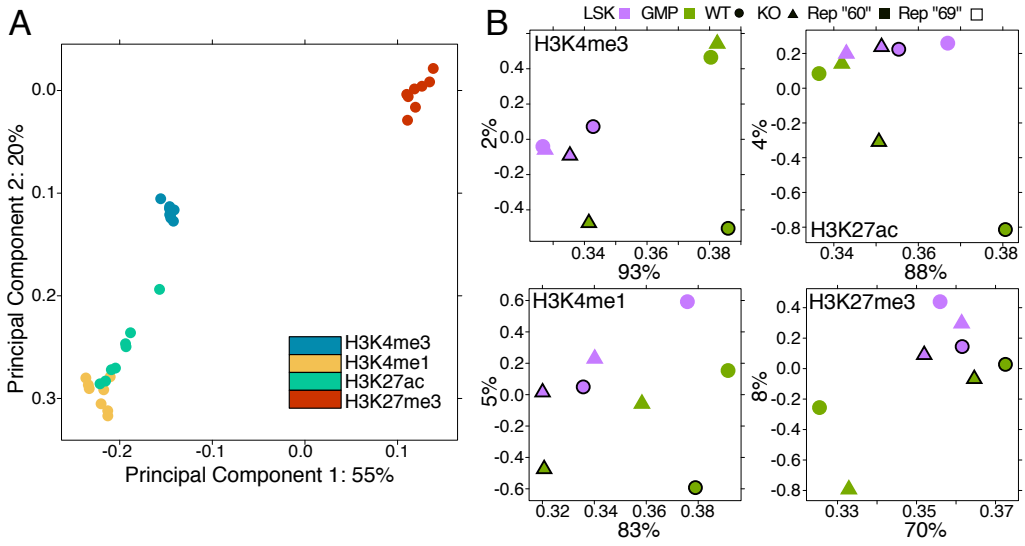

Figure 6.4: Relationships between ChIP-Seq samples A) Principal component analysis PCA plots. A) All histone modification data coloured by modification. B) Separate PCA plots for each modification, colour and shape define their sample identity.

To further visualize the relationships between different biological samples, I plotted their genome wide correlations in Figure 6.5. All samples are highly similar, all correlations had R^2 values greater than 0.9 and P.values below 0.001. The correlation between ETOH treated biological replicates is weaker than the correlation between ETOH and 4-OHT treated samples. In the majority of cases, the correlation between ETOH treated biological replicates is equal to or weaker than biological replicates with different cells-of-origin (LSK^{ME} versus GMP^{ME}). The only exception is that the correlation between H3K27me3 signal LSK^{ME} and GMP^{ME} from replicate "60" is slightly less correlated than LSK^{ME} biological replicates (Figure 6.5, bottom right, R^2 values: 0.93, 0.95).

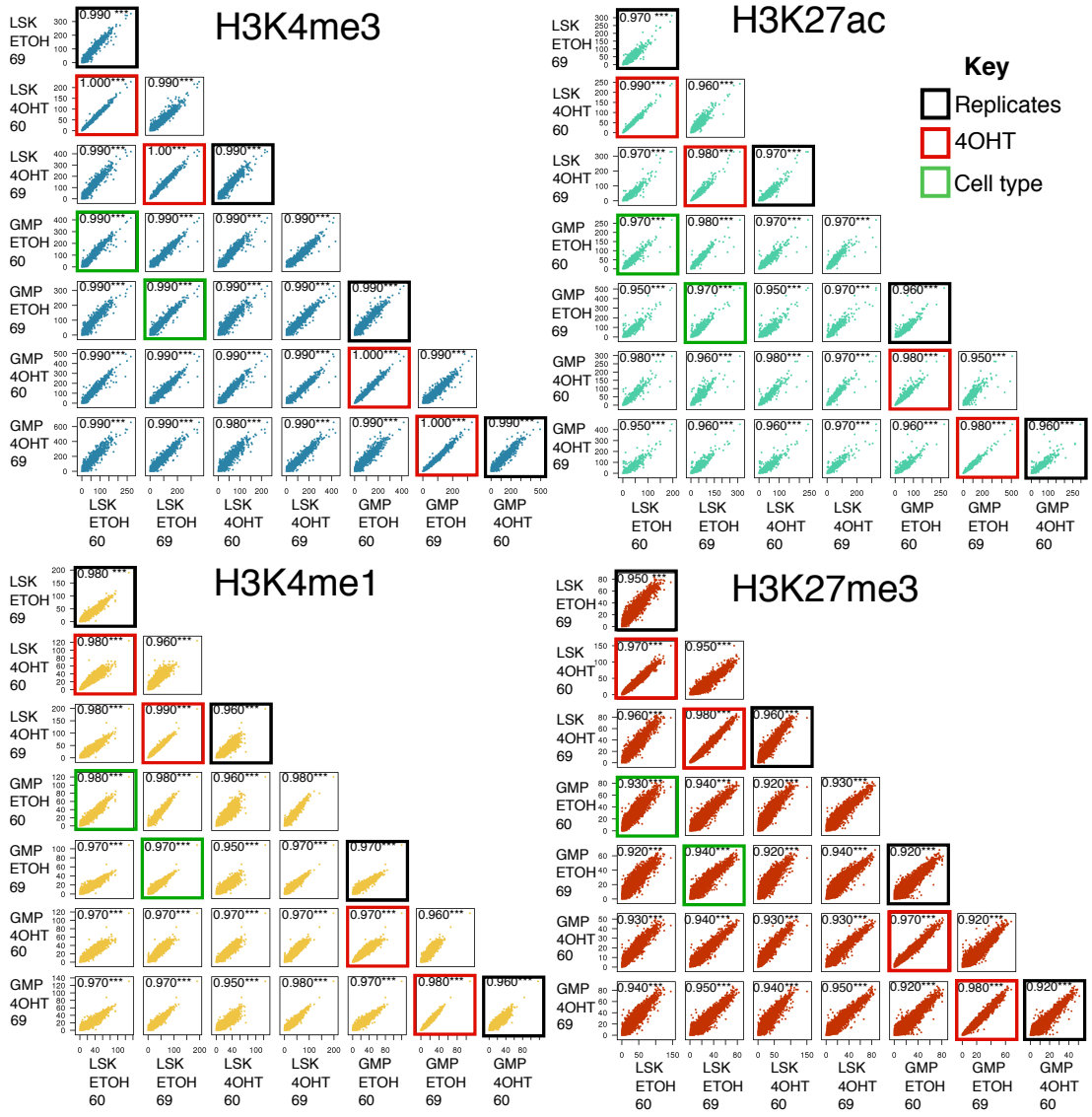

Figure 6.5: Global correlation between ChIP-Seq samples This figure shows the correlation between all samples as scatter plots. Values were obtained by counting the number of reads in 10 kb bins across the entire genome. Notable comparisons are highlighted with a coloured border. As indicated by the key on in the top right: A black border signifies biological replicates of ETOH controls. A red border highlights the plots comparing individual replicates after β-catenin deletion. Green borders highlight the comparison between cell types. In the top left corner of each plot is the Pearson's correlation coefficient, the asterisks indicate significance level. (***P.value <0.001)

6.3.6 Histone modification data segment the genome into biologically relevant chromatin states

To take advantage of all histone modification data in symphony, I employed ChromHMM to segment the genome into different chromatin states. With just a 6-state model, we were able to see separation of the genome into segments with predictable biological functions (Figure 6.6A). Figure 6.6A shows the probability of a mark residing in each state. No histone modification implies heterochromatin regions (Het). The presence of only H3K27me3 signifies polycomb repression (ReprPC). The presence of only H3K4me1 identifies possible enhancers, additional presence of H3K27ac transforms these enhancer regions from bivalent (EnhBiv) to active (EnhA). Finally, H3K4me3 combined with H3K27ac or H3K27me3 defines active TSS (TssA) or bivalent TSS (TssBiv) respectively. As expected, most of the genome is unmodified and therefore heterochromatin (Figure 6.6B). Following heterochromatin, the most abundant chromatin states are bivalent enhancers, which require only the presence of H3K4me1. These regions may be sensitive to the binarisation threshold, therefore may contain noisy regions. Active enhancers and polycomb repressed regions are about equally represented. Next comes active TSSs and finally bivalent TSS is the rarest state.

The total estimated size of the mouse genome, omitting centromeres is 160 mega-bases (Mb). Considering this, we can interpret 1.6Mb in the scale in Figure 6.6C-E as being 1/100 of the genome. As would be expected only a small proportion of states in the genome change between conditions. The vast majority of these changes in both cell-of-origin and β-catenin deletion are between Heterochromatin ("Het") and single modification states ("ReprPC" and "EnhBiv"). To take a closer look at the more interpretable states (between active and inactive promoters and enhancers) I removed these from the plot. Most promoters remain stable to change, however more segments belonging to enhancers transition. To elucidate if there are qualitative differences between biological replicates at the chromatin level, I plotted the chromatin states of detected transitions at the TSS in a heatmap form (Figure 6.7). Although there are some transitions which appear legitimate, I was unable to filter out noisy transitions using the methods I developed. The results of these analyses were difficult to interpret and were concluded to be unsuitable for the comparisons in this study. However, chromatin state segmentation further validates the expected patterns of histone modification in these data.

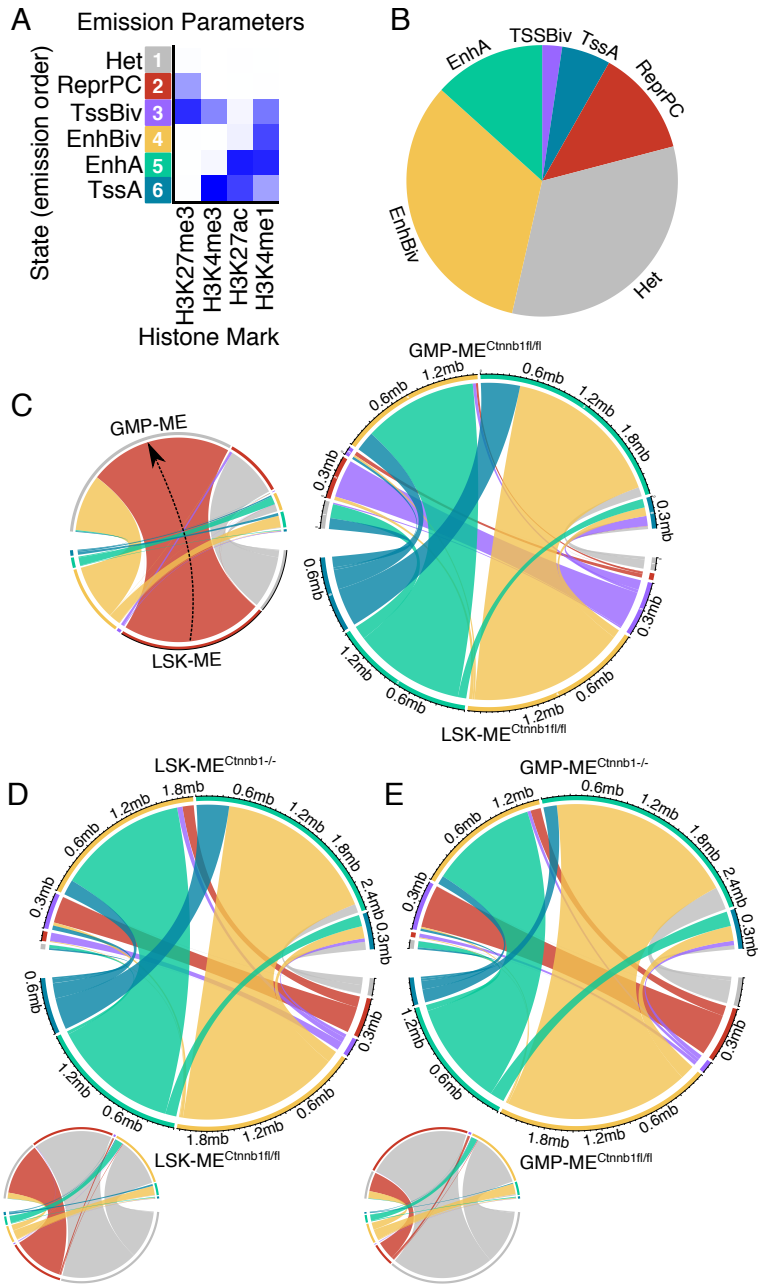

Figure 6.6: Global chromatin state assignment and transitions between experimental conditions A) Emission parameters for assigned states. The intensity of each cell indicates the probability of the presence of the indicated modification (bottom) in each state (Left): heterochromatin or unmodified (Het), polycomb repressed (ReprPC), bivalent TSS (TssBiv), bivalent enhancers (EnhBiv), active enhancers (EnhA) and active TSS (TssA). B) Pie chart showing the proportion of assigned states across the genome. C-E) Chord diagrams showing flow of transitioning states between experimental conditions. The connections of the segments between the bottom and top of the semi-circular axis represent the transition from one state to another. Scales are shown in mega-bases (mb) of DNA. C) Chromatin state transitions between different cells-of-origin. Left: Simplified chord diagram including all transitional states, an arrow is drawn to demonstrate the direction of transition. Right: Chord diagram with dominant transitions removed (EnhBiv & ReprPC) to reveal remaining interactions in greater detail. D:E) Chromatin transitions following *Ctnnb1* KO in LSK^{ME} (D) and in GMP^{ME} (E). Layout for D:E is as described in C.

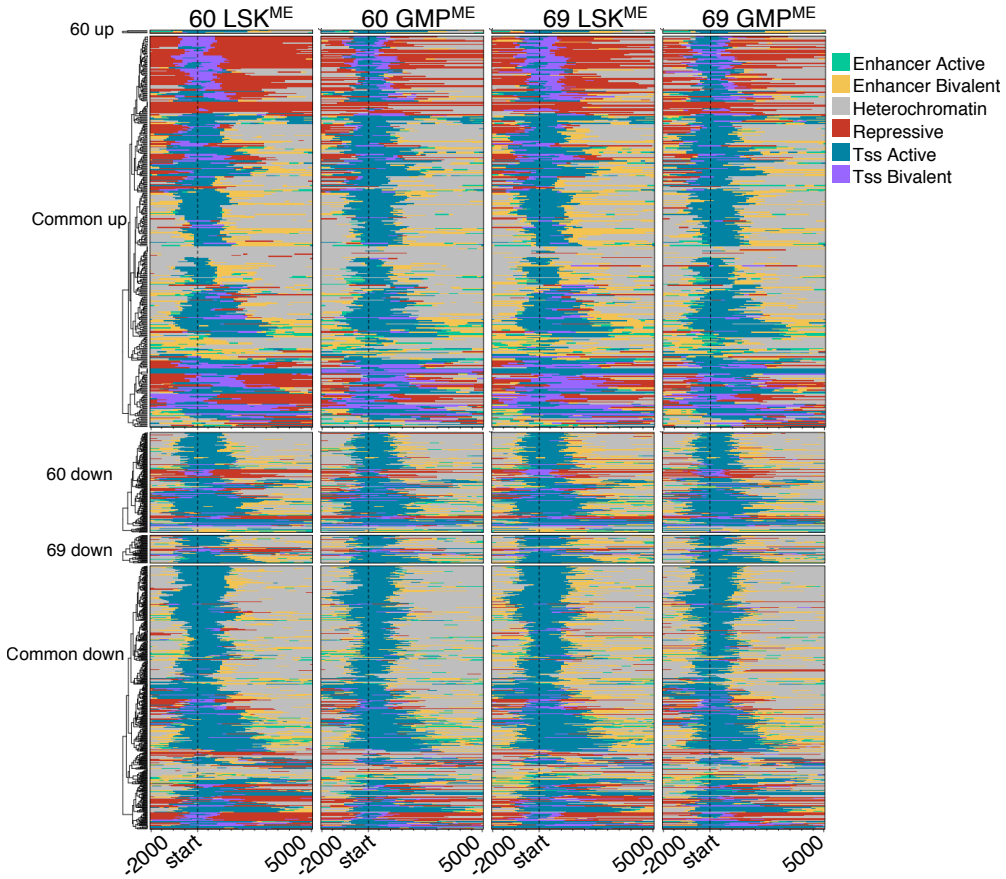

Figure 6.7: Chromatin state changes between conditions at the TSS. Heatmaps showing the states called by ChromHMM 2 kb upstream and 5 kb downstream of the TSS which were detected to transition between LSK^{ME} and GMP^{ME}. The location of the TSS is indicated by "start" on the scale at the bottom and by a dotted vertical line. The colour of each state is as indicated in the key on the right. Regions are shown which transition between TssA to TssBiv or ReprPC. The labels on the left indicate in which replicate the state change was found and the direction of that change in relation to promoter activity (up or down).

6.3.7 Variation of histone modification observed at the HoxA locus

As described in subsection 1.2.1, the Homeobox A (HoxA) locus is a critical target in MLL-fusion-transformed leukaemia (Zeisig et al. 2004; Ayton and Cleary 2003; Faber et al. 2009; Milne et al. 2010). Activity at the HoxA locus may therefore confer the "level" of transformation by MLL fusions. Interestingly, visualisation of normalized ChIP signal at the HoxA locus reveals considerable differences in histone modification between cell-of-origin and replicate (Figure 6.8). Whilst both LSK^{ME} replicates appear similar, GMP^{ME} from replicate "60" shows unique extension of H3K27me3 into the HoxA locus and H3K4me1 reduction (Figure 6.8A). This sample also shows reduction in H3K4me3 and H3K27ac at *Hoxa9* and flanking regions resulting in lower transcription of *Hoxa9* (Figure 6.8C). These changes could imply a less extreme transformation of

the cells or a more heterogeneous population. GMP^{ME} replicate "69" exhibits unique enrichment for H3K4me3 and H3K27ac and a decrease in H3K27me3 resulting in expression of Homeobox A11 (*Hoxa11*) (Figure 6.8B). Interestingly, both GMPs "60" and "69" show increased active marks and transcript expression of Homeobox A2 *Hoxa2* (Figure 6.8D). For simplicity, only β-floxed samples are plotted, however the described patterns remain after knock-out.

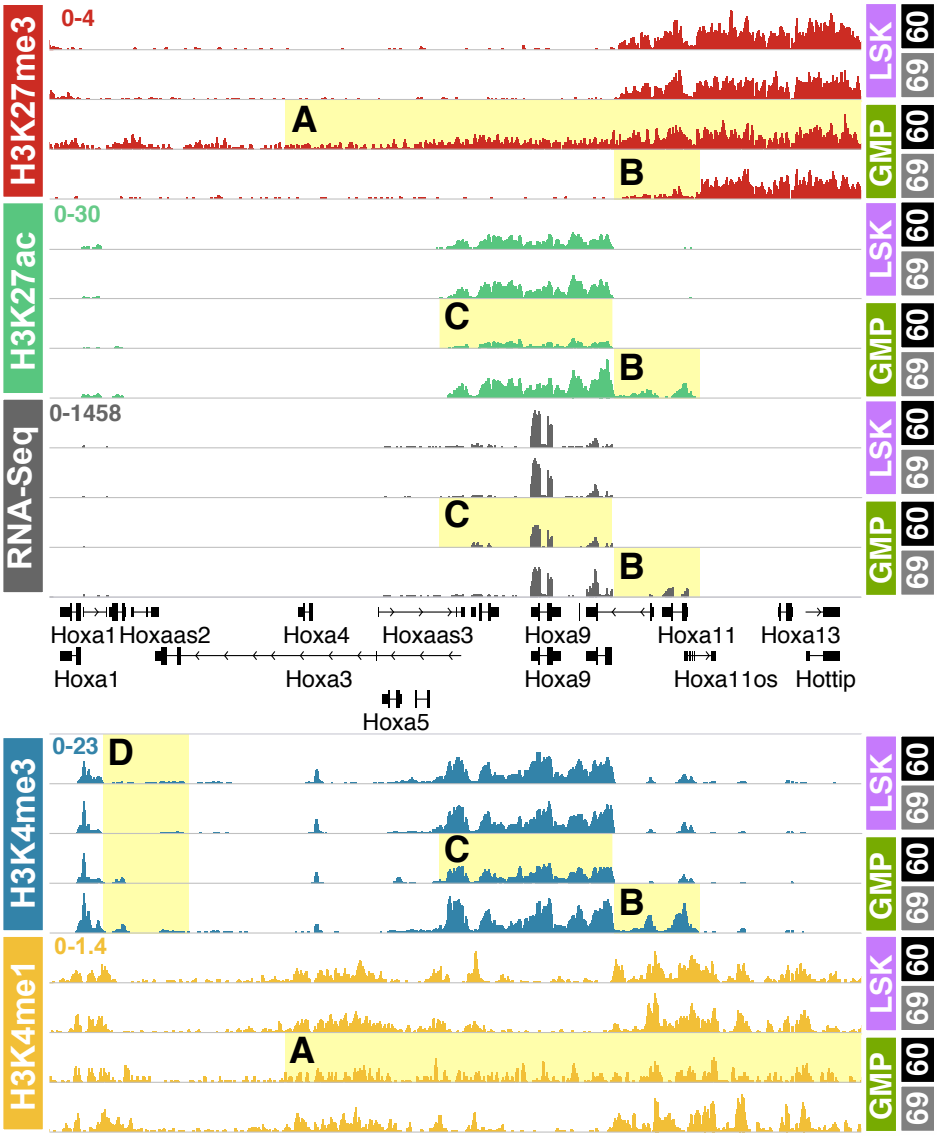

Figure 6.8: Transcriptomic and epigenetic variation at the HoxA locus in MLL-ENL transformed cells Normalised signal for all four histone marks and RNA-Seq. The assay shown in each track is labelled on the left. The identity of the biological sample is labelled on the right. Replicate dependent differences (A-C) and cell-of-origin dependent (D) differences are highlighted and described in the main text. The scales on the right are in TMM normalised CPM.

To compare the observed differences at the HoxA locus with the modification of this region in normal sorted populations I plotted the normalised coverage from publicly available data in Figure 6.9, (Bernt et al. 2011; Lara-Astiaso et al. 2014). Interestingly, like GMP^{ME} replicate "60", there is greater H3K27me3 across the entire HoxA locus in GMP compared to LSK (Figure 6.9A). Also similarly to GMP^{ME} replicate "60", there is reduced H3K4me3 and H3K27ac in GMP versus HSC (Figure 6.9C). Conversely, at the *Hoxa11* locus there is higher H3K27ac and H3K4me3 in HSC than in GMP (Figure 6.9B), which is the opposite trend to what is seen in GMP^{ME} replicate "69" versus LSK^{ME}. I conclude that there is considerable variation between replicates at the HoxA locus which may correspond to biologically important differences in their normal cell-of-origin.

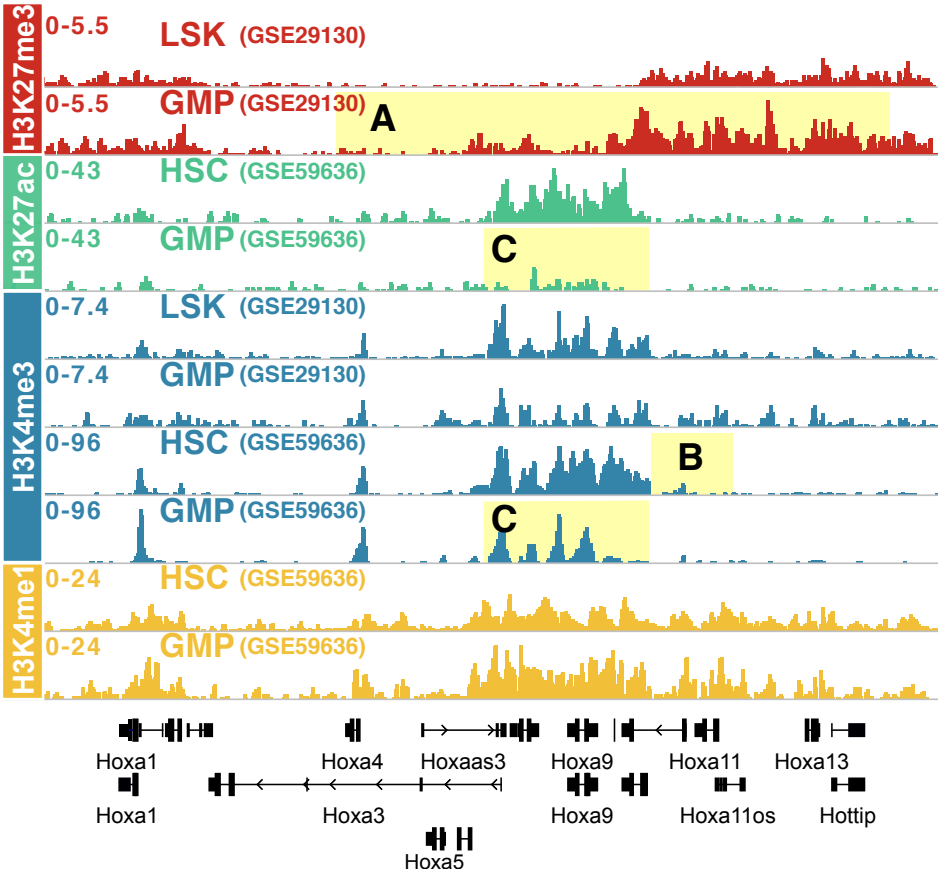

Figure 6.9: Transcriptomic and epigenetic variation at the HoxA locus in normal cells Normalised signal taken from the Gene Expression Omnibus (GEO). Tracks are labelled with their cell type and GEO accession. GSE29130 and GSE59636 were taken from Bernt et al. (2011) and Lara-Astiaso et al. (2014), respectively. A-C are the same regions highlighted in Figure 6.8.

6.3.8 Gene-centric quantitative analysis of histone modification between samples with different cells-of-origin

Statistically differentially epigenetically modified regions between LSK^{ME} and GMP^{ME} were detected with csaw. Cell-of-origin dependent differences in histone modification and overlaps with differentially expressed genes are summarised in Figure 6.10. The general trend in histone mark changes for cell-of-origin reflect RNA-Seq; the majority of differentially modified genes are upregulated in GMP^{ME} compared with LSK^{ME}. Activating marks, H3K4me3 and H3K27ac, are mostly upregulated in GMP^{ME} compared with LSK^{ME} (Figure 6.10A). The repressive mark H3K27me3 is mostly downregulated in GMP^{ME} compared with LSK^{ME} (Figure 6.10A).

Of all histone modifications probed, H3K4me3 shows the most cell-of-origin related differences. 322 unique genes are differentially modified by H3K4me3 (Figure 6.10A), over 30% of these genes were also found to be differentially expressed in RNA-Seq (Figure 6.10B). 79 unique genes are differentially modified by H3K27ac and over 50% of these are also differentially expressed (Figure 6.10B). Just 27 regions are differentially modified by H3K27me3 however over 60% (17) of these are also differentially expressed (Figure 6.10B). Only 15 genes are differentially modified by H3K4me1 and over 70% (11) of these are also differentially expressed (Figure 6.10B). None of the regions differentially modified by H3K4me1 were identified at the promoter and lie somewhere in the gene body (Figure 6.10A). Using the "Naïve" method to group adjacent windows gave similar numbers or less than gene centric analysis for H3K4me3 and H3K27me3 (Figure 6.10A). On the other hand, there were many additional changes for H3K4me1 and H3K27ac found lying outside gene bodies (Figure 6.10A). This implies there are significant changes at other regulatory features such as enhancer regions.

To visualise the relationship between differential H3K4me3, H3K27ac and gene expression, I plotted the H3K4me3 and H3K27ac signal at the TSS of differentially modified and/or differentially expressed genes (Figure 6.11). There are few genes differentially modified by H3K27ac and differentially expressed. The H3K27ac differential regions undetected in RNA-Seq appear inconsistent between replicates (Figure 6.11A). There are many genes with low H3K27ac signal which do not exhibit changes but are differentially expressed (Figure 6.11A). However, of these, the most extreme differences in H3K27ac appear to correlate with differential expression. The H3K4me3, differential regions undetected in RNA-Seq appear inconsistent between replicates (Figure 6.11B). Additionally for H3K4me3, like H3K27ac, the most extreme changes reflect the expression of differentially expressed genes in RNA-Seq although it is less clear (Figure 6.11B).

To visualise individual events of cell-of-origin differential ChIP-Seq, I plotted normalised signal tracks for selected gene loci in Figure 6.12. Myb/SANT DNA Binding Domain Containing 3 *Msandt* and Pre-B-Cell Leukaemia Homeobox 3 *Pbx3* were found to have decreased HK27me3, increased H3K27ac, H3K4me3 and gene expression in GMP^{ME} compared to LSK^{ME} (Figure 6.12A). Zinc finger protein 951 *Zfp951* was found to have decreased H3K27ac, H3K4me3 and gene expression in GMP^{ME} compared to LSK^{ME} (Figure 6.12A). Acyl-CoA Synthetase Short Chain Family Member 2 *Acss2* was found to have increased H3K27ac, H3K4me3 and gene expression in GMP^{ME} compared to LSK^{ME} (Figure 6.12A). Of note, although there was no difference in gene expression, SMAD Family Member 4 (*Smad4*) and Nuclear Receptor Subfamily 4 Group A Member 2 (*Nr4a2*) exhibited significant decreases in H3K27me3 in GMP^{ME} compared with LSK^{ME} (Figure 6.12B). Complementary to the variation shown at the HoxA locus in subsection 6.3.7, Figure 6.8, notable replicate specific differences in histone modification are shown in Figure 6.12C. *Hoxa11* and Frizzled Class Receptor 1 *Fzd1* exhibit increased H3K27ac, H3K4me3 and gene expression in GMP^{ME} replicate "69". CD48 molecule (*CD48*, also known as Signalling Lymphocytic Activation Molecule 2) exhibits increased H3K27ac, H3K4me3 and gene expression in GMP^{ME} replicate "60".

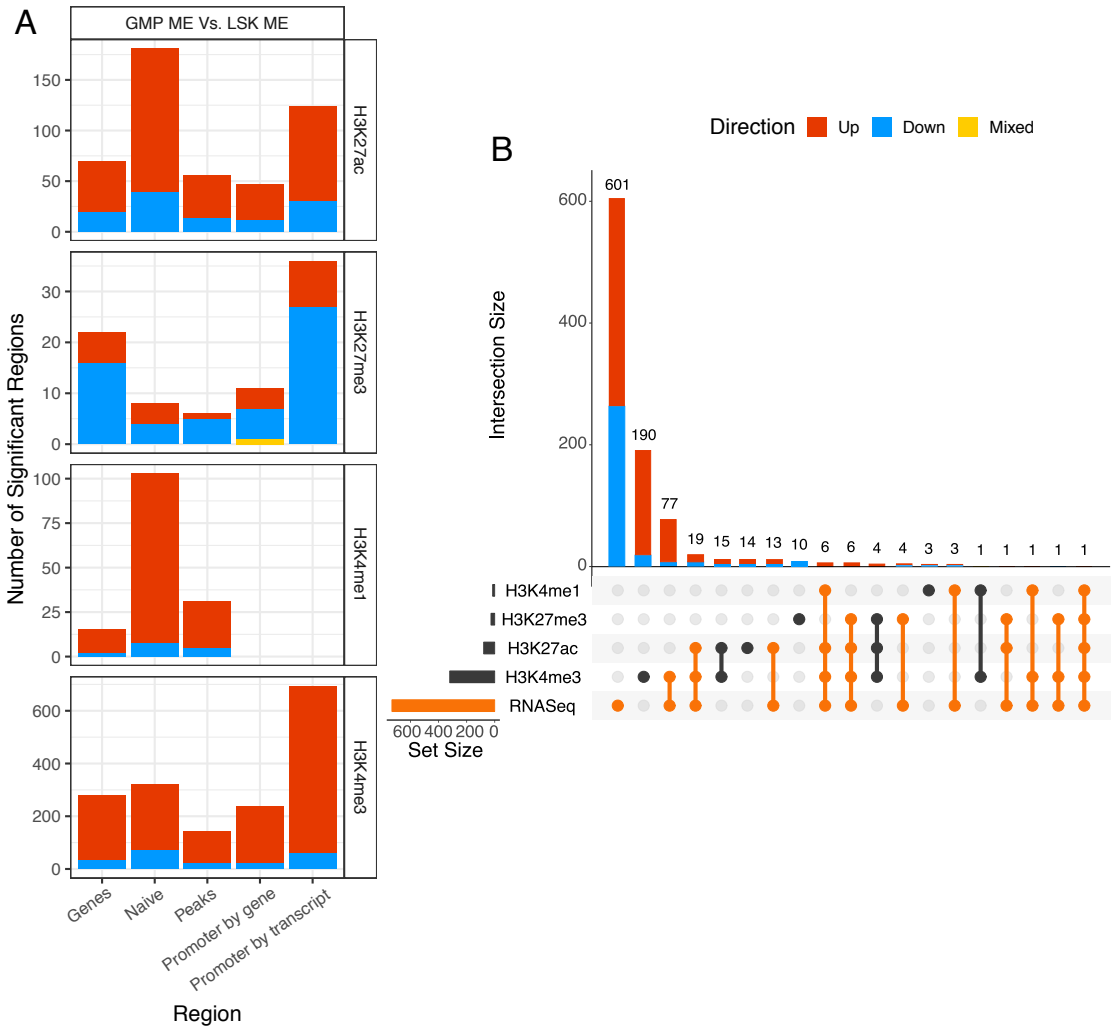

Figure 6.10: Summary of significantly differentially modified regions between LSK^{ME} and GMP^{ME} The colour of the vertical bars represents the proportion of regions showing up-regulation (red), down-regulation (blue) and mixed regulation (yellow). This is indicated by the colour key at the top, right. A) The number of significantly differentially modified regions for each histone modification detected by csaw with an FDR cut-off of 0.05. The histone modification for each bar chart is labelled on the left. The x-axis describes which region category was examined for each bar. Genes: whole gene bodies. Peaks: peaks detected with MACS2. Naïve: regions created by simply merging adjacent windows with significant ChIP-Seq enrichment. Promoter by gene: the most upstream promoter region for a given gene. Promoter by transcript: promoters at all annotated TSSs. B) The overlaps between histone mark changes and RNA-Seq. This diagram is equivalent to a 5 set Venn diagram. The horizontal histogram shows the total number of significantly different genes for each sequencing assay. The vertical histogram shows the number of significantly different genes in each intersect. The assay(s) belonging to the members of each set are defined in the matrix below. The shaded dots indicate which assays the genes in the intersection belong to. Any overlaps with RNA-Seq are highlighted in orange.

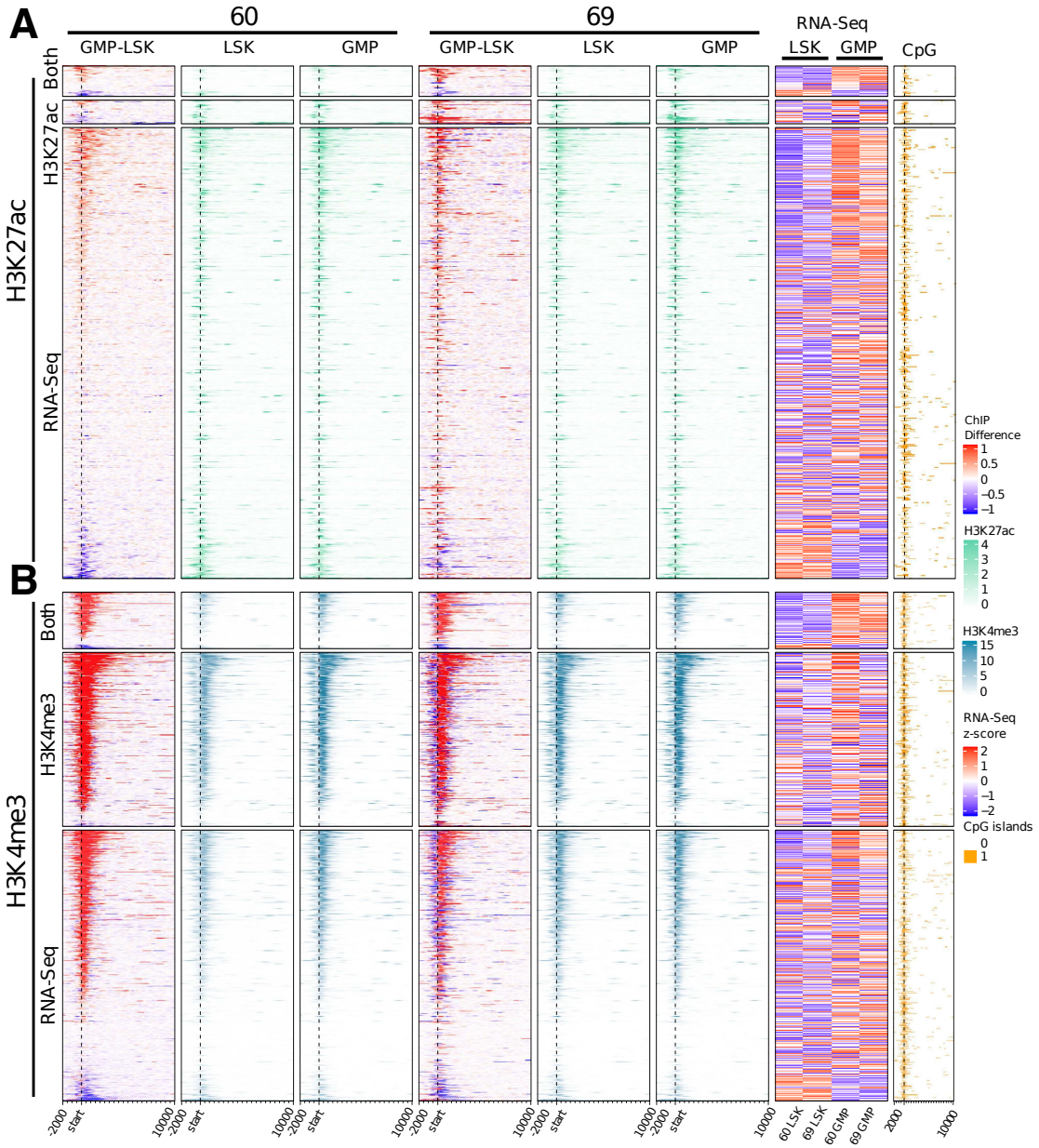

Figure 6.11: H3K27ac, H3K4me3 and gene expression at differentially modified and/or expressed genes between LSK^{ME} and GMP^{ME} A) Differential H3K27ac compared with RNA-Seq. B) Differential H3K4me3 compared with RNA-Seq. Heatmaps are plotted 2 kb upstream and 10 kb downstream of the TSS. The location of the TSS is indicated by "start" on the scale at the bottom and by a dotted vertical line. The biological replicate ("60" or "69") is indicated by labels at the top. Heatmaps labelled "GMP-LSK" show the difference in ChIP signal between GMP and LSK derived cell lines. Heatmaps labelled "LSK" or "GMP" show the normalised signal for the indicated modification for the LSK^{ME} and GMP^{ME} sample respectively. The penultimate heatmap shows normalised gene expression. The final heatmap shows the location of CpG islands in yellow. The heatmaps are clustered into 3 groups: 1. "Both": genes which are significantly different in ChIP-Seq and RNA-Seq. 2. "H3k27ac" or "H3K4me3": significant in ChIP-Seq only. 3."RNA-Seq": significant in RNA-Seq only. Within each cluster, rows are ordered by difference in histone modification between LSK^{ME} and GMP^{ME}.

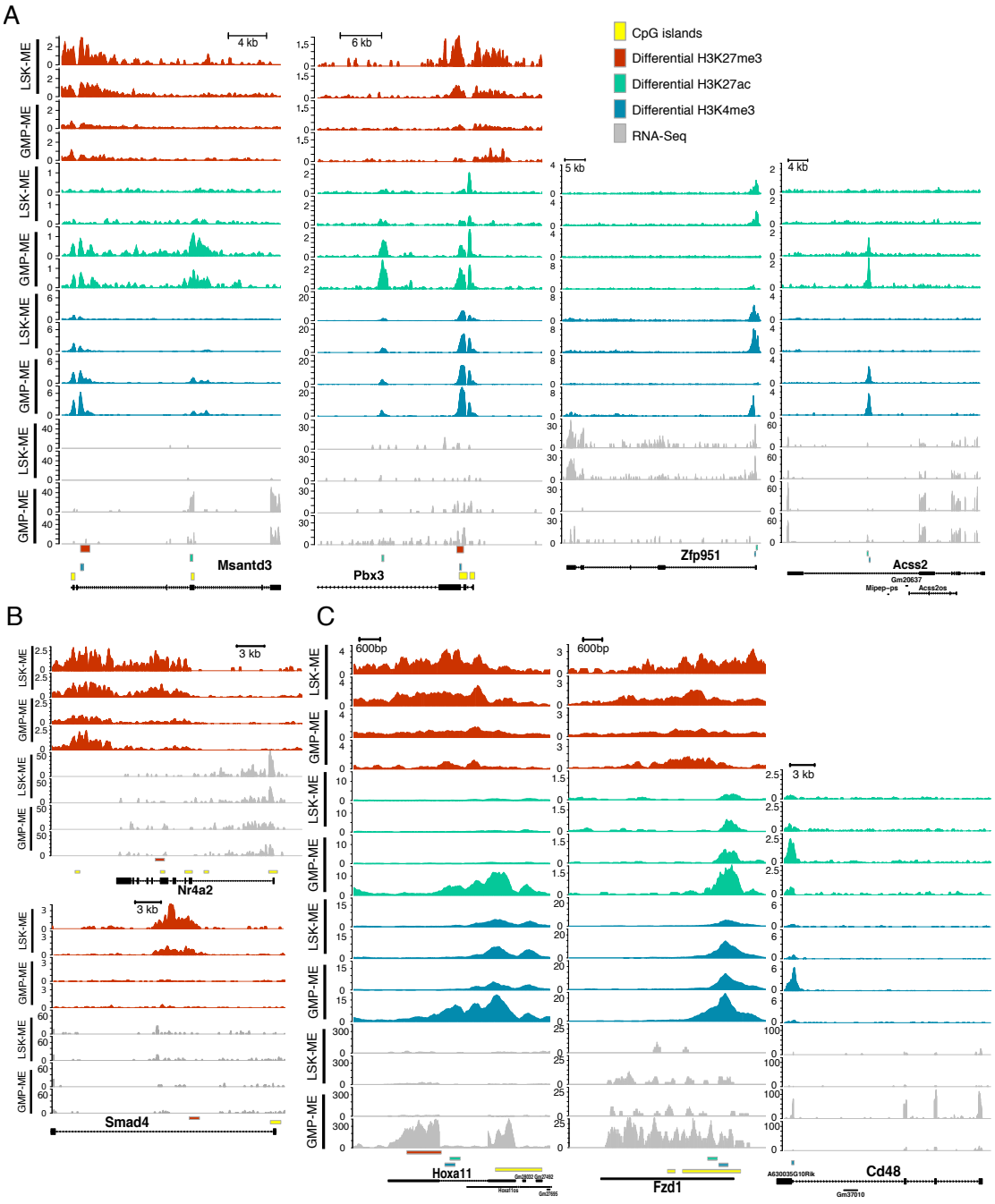

Figure 6.12: Signal tracks of loci exhibiting cell-of-origin dependent changes in histone modification The blocks underneath the tracks signify statistically significant changes in modification or CpG islands, as indicated by the key (top right). The colours in the key also correspond to the assay shown in each signal track. Gene models are also shown beneath the tracks. Each plot has a key indicating the distance shown. Sample identity is indicated on the left. The signal scale is TMM normalised CPM. A) Consistently differentially expressed genes which also show changes in histone modification. B) Changes in H3K27me3 which do not correlate with gene expression. C) Cell line specific changes in histone modification and gene expression.

6.3.9 Pathway analysis of differentially modified genes between different cells-of-origin

To investigate if there were coordinated changes in histone marks at promoter regions I performed pathway analysis for their associated genes (Figure 6.13). H3K27me3 differences were enriched in developmental terms, including *Smad4* and *Nr4a2* (Figure 6.13A). *H3K4me3* and *H3K27ac* mirrored some of the pathways which were found in differential expression analyses (see subsection 3.3.4, Figure 3.4) such as "Transcriptional Missregulation in Cancer" and "Cellular Senescence" (Figure 6.13B & C). As there were so few genes differentially marked by H3K4me1, there were no enriched GO terms or KEGG pathways for H3K4me1.

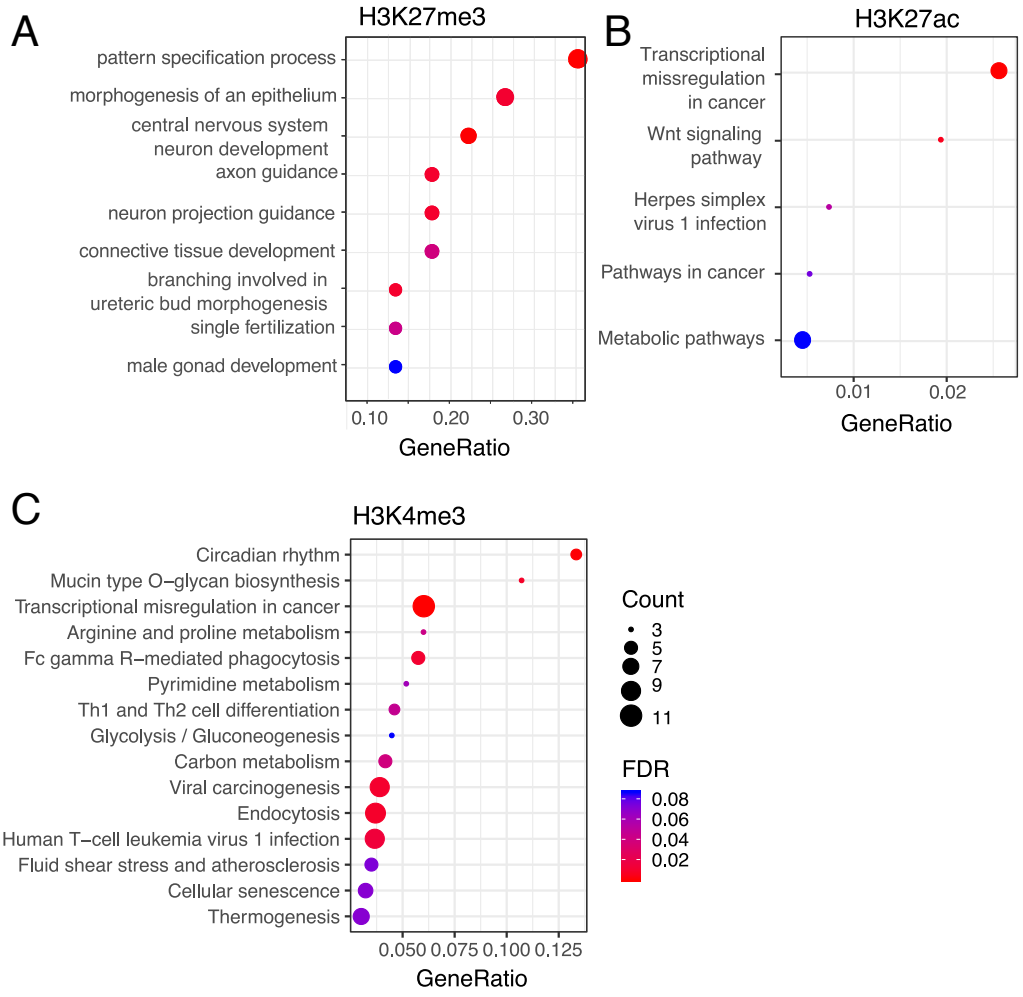

Figure 6.13: GO and KEGG analysis for genes with differentially bound promoters between LSK^{ME} and GMP^{ME} A) Enriched biological process GO terms for promoters differentially bound by H3K27me3. B) Enriched KEGG pathways for promoters differentially bound by H3K27ac. C) Enriched KEGG pathways for promoters differentially bound by H3K4me3. The x axes show the ratio between the number of genes in the pathway and number of differentially marked genes in the enriched term. The size and colour of the dots indicate the number of differential genes (Count) and FDR respectively.

6.3.10 Cell-of-Origin associated Enhancer analysis

As highlighted previously in Figure 6.8, there were more epigenetic differences outside of the promoter region between cells of different origin. Therefore, it would be informative to investigate these regions and attempt to uncover their potential biological functions giving rise to the phenotypic differences seen between LSK^{ME} and GMP^{ME}. Enhancers were crudely defined as H3K4me1 peak regions outside of the promoter region which were defined as 1 kb upstream 2 kb downstream of TSSs. Putative enhancers which appeared in more than one sample were catalogued. To detect changes in enhancer activity between LSK^{ME} and GMP^{ME} differential expression of H3K27ac at enhancers was calculated with DiffBind. De novo enhancers were determined by comparing overlaps of active enhancers exhibiting H3K27ac enrichment between conditions. Enriched motifs found in de novo active enhancers or enhancers with differential activity between different cells-of-origin are shown in Figure 6.14.

Enhancers unique to GMPME: n=50

Target	PWM	Motif ID	P–value	In top motifs
Sox2		Sox2	0.033	23 %
Sox6		Sox6	0.0124	21 %
Sox30		SCI09/Sox30	0.00599	17 %
Bhlha15		M0218_1.02	0.0462	15 %
Mafg		M0308_1.02	2.46e–07	15 %
Atf4		M5988_1.02	0.00348	15 %
Dlx2		jolma2013 Dlx2	0.00426	15 %
Hoxb3		Cell08/Hoxb3_1720.2	0.0124	15 %

Enhancers unique to LSKME: n=342

Target	PWM	Motif ID	P–value	In top motifs
Erg		Erg	3.8e–07	15 %
Myb		Myb	0.000173	14 %
Tcf12		Tcf12	0.0453	13 %
Mafg		M0308_1.02	9.45e–19	13 %
Meis3		jolma2013-Meis3	3e–07	12 %
Cebpd		M0314_1.02	3.96e–13	12 %
NFATC2		NFATC2	0.000499	12 %
Cebpg		M0313_1.02	1.8e–13	12 %

Increased enhancer activity in GMPME: n=58

Target	PWM	Motif ID	P–value	In top motifs
Cebpg		M0313_1.02	2.69e–06	22 %
Cebpd		M0274_1.02	1.32e–05	21 %
Cebpb		M0314_1.02	1.14e–07	19 %
RUNX1		RUNX1	0.000155	19 %
Cebpa		M0304_1.02	5.57e–07	17 %
Ets1		Ets1	7.02e–05	17 %
Sox17		Sox17	0.00686	17 %
Cebpb		M0315_1.02	3.88e–07	16 %

Increased enhancer activity in LSKME: n=12

Target	PWM	Motif ID	P–value	In top motifs
Mafk		M0290_1.02	0.00268	42 %
Cebpg		M0313_1.02	3.46e–05	42 %
Tef		M0293_1.02	0.000907	33 %
Cebpb		M0314_1.02	0.000662	33 %
Mycn		M1927_1.02	0.0219	33 %
Arid5a		M0108_1.02	0.0108	25 %
Setbp1		M0130_1.02	0.00248	25 %
Cebpd		M0274_1.02	0.000211	25 %

Figure 6.14: Motif analysis of enhancers exhibiting cell-of-origin specific activity

6.3.11 β-catenin related changes in histone modifications

Statistically differentially epigenetically modified regions between β-catenin floxed β-catenin deleted GMP^{ME} and LSK^{ME} were detected with csaw. β-catenin dependent differences in histone modification and overlaps with differentially expressed genes are summarised in Figure 6.15. Interestingly, no significant changes were found for H3K27me3 (Figure 6.15A). Very few significant changes were detected for H3K4me1 (Figure 6.15A). In GMP^{ME} β-catenin KO, the only significant change in H3K4me1 was a decrease found at the gene: Non Imprinted In Prader-Willi/Angelman Syndrome 1 (*Nipa1*). Following β-catenin deletion in LSK^{ME}, decreases in H3k4me1 were found at 6 genes: *Nipa1*, NLR Family pyrin domain-containing 3 (*Nlrp3*), Bactericidal/Permeability-Increasing Fold-Containing Family C (*Bpifc*), Early Endosome Antigen 1 *Eea1*, Mitochondrial Ribosomal Protein S26 (*Mrps26*) and *Ctnnb1*.

In line with RNA-Seq analyses, H3K4me3 is mostly upregulated in β-catenin deleted samples compared with β-catenin floxed (Figure 6.15A). β-catenin deletion in LSK^{ME} resulted in significant H3K4me3 changes associated with 100 genes, almost 50% were also differentially expressed in RNA-Seq (Figure 6.15B). Furthermore, 147 genes were differentially marked by H3K4me3 following β-catenin deletion in GMP^{ME}, however less than 30% of these were also found in RNA-Seq (Figure 6.15B). In contrast with RNA-Seq analyses, H3K27ac is mostly downregulated after β-catenin deletion (Figure 6.15A). In LSK^{ME}, 119 genes were differentially modified by H3K27ac, only 15 of these were also differentially expressed (Figure 6.15B). Similarly in GMP^{ME}, 100 genes were differentially modified by H3K27ac and only 9 of these were also differentially expressed (Figure 6.15B). Using the "Naïve" method to group adjacent windows gave similar numbers or less than gene centric analysis for H3K4me3 and H3K27Ac (Figure 6.15A). This suggests that there is little detectable differential modification outside of genic regions.

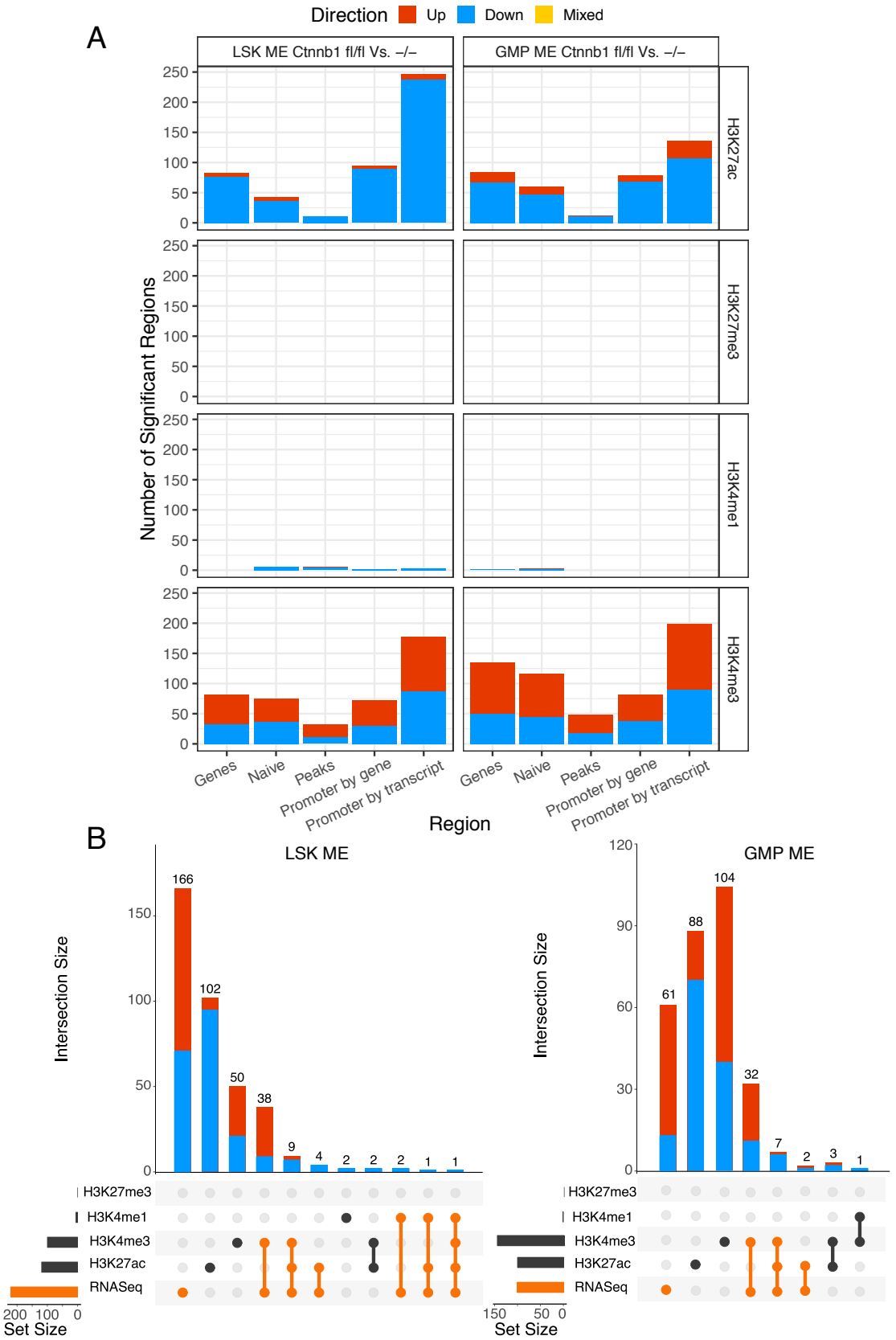

Figure 6.15: Summary of significantly differentially modified regions following β-catenin knock-out A) Significantly differentially bound regions detected with csaw with a FDR cut-off of 0.05. B) The overlaps between significant histone mark changes and RNA-Seq. Left: LSK^{ME}, Right: GMP^{ME}. The axes and layout of these plots are as described in the legend of Figure 6.10.

To visualise the relationship between differential H3K4me3, H3K27ac and gene expression, I plotted the H3K4me3 and H3K27ac signal at the TSS of differentially modified and/or differentially expressed genes in LSK^{ME} (Figure 6.16A) and GMP^{ME} (Figure 6.16B). When ordered by differential H3K27ac, most significantly differentially expressed genes detected in RNA-Seq after β-catenin KO in both LSK^{ME} and GMP^{ME} visually segregate into upregulated and downregulated genes (Figure 6.16A & B, top). However, genes which exhibit significant differential H3K27ac but were not detected as differentially expressed in RNA-Seq are largely incongruent with transcription (Figure 6.16A & B, top). Amongst these there are a collection of histone protein genes. (Figure 6.16A & B, top). On the other hand, changes in H3K4me3 are mostly congruent with gene expression whether or not they are differentially expressed (Figure 6.16A & B, bottom), this is particularly clear in GMP^{ME} (Figure 6.16B, bottom). Interestingly, there appeared to be an enrichment for CpG islands in the promoters of genes which were differentially modified (especially for H3K27ac) and a depletion of CpG in promoters which were differentially expressed but not differentially modified (Figure 6.16).

As *Gadd45a* was found to be uniquely differentially expressed in GMP^{ME} upon β-catenin deletion, it would be interesting if we could observe congruent changes in epigenetic modification. Despite changes in gene expression, no statistically significant differences in histone modifications were detected within 10mb of the *Gadd45a* locus. Normalised signal tracks of all histone marks around the *Gadd45a* locus are visualised in Figure 6.17. To visualise individual events of β-catenin conditional differential ChIP-Seq, I plotted normalised signal tracks for selected gene loci in Figure 6.18. *Tgm2* and Thrombospondin 1 (*Thbs1*) exhibited decreases in H3K27ac, H3K4me3 and transcript expression (Figure 6.18A). Cyclin Dependent Kinase Inhibitor 1A *Cdkn1a* was detected as differentially expressed in LSK^{ME} KO but not in GMP^{ME} KO. At *Cdkn1a*, significant differences in H3K27ac and H3K4me3 were detected in LSK^{ME} KO but only differential H3K4me3 was detected in GMP^{ME} KO (Figure 6.18A, right). H3K27ac and RNA-Seq were plotted at clusters of histone genes (Figure 6.18B). As previously seen in Figure 6.16, there is a disconnect between gene expression and changes in H3K27ac at these regions (Figure 6.18B).

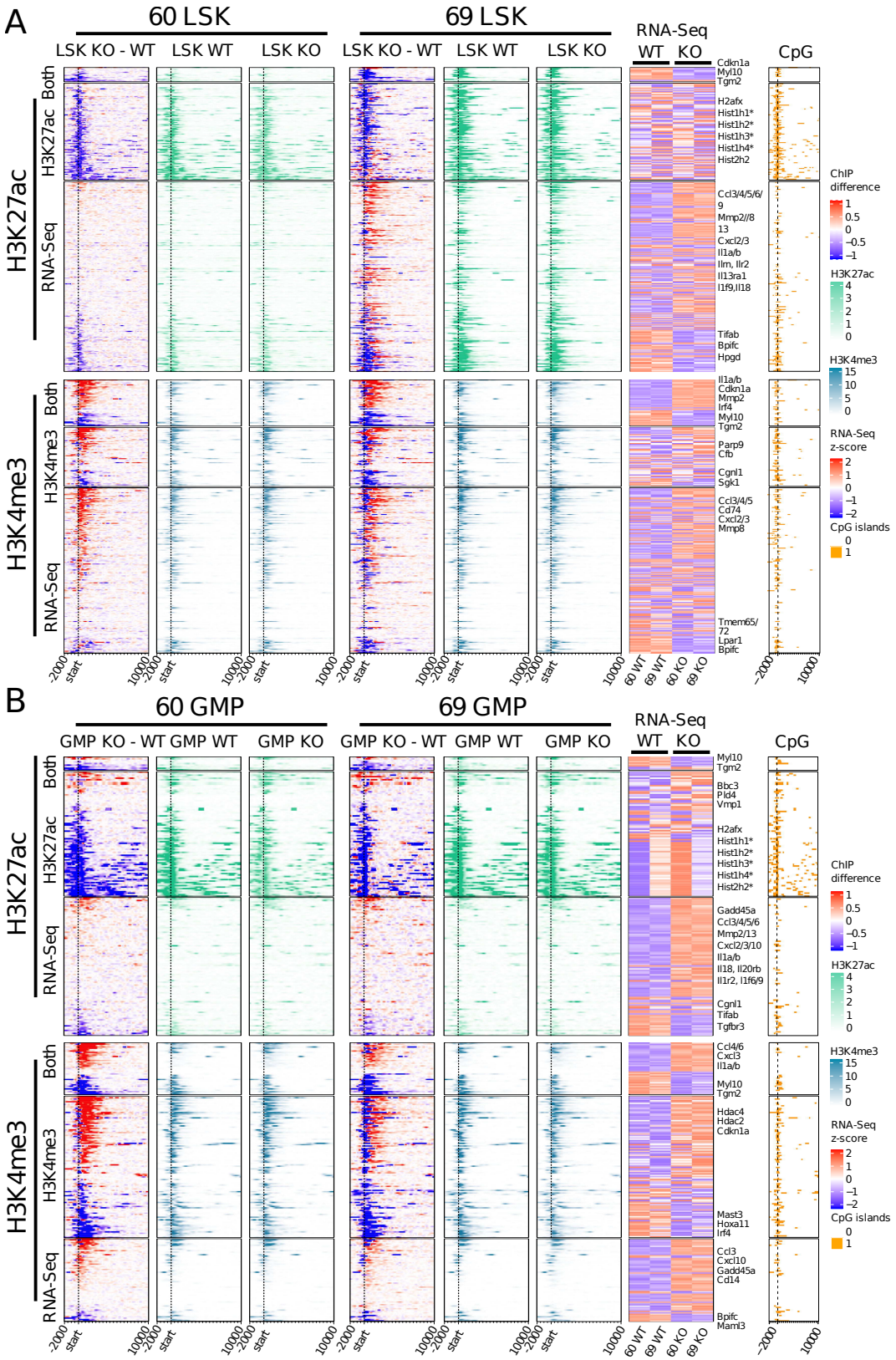

Figure 6.16: H3K4me3 and H3K27ac signal and gene expression at differentially modified or expressed genes following β-catenin knock-out A) β-catenin related changes in LSK^{ME}. B) β-catenin related changes in GMP^{ME}. The layout of these plots is the same as outlined in the legend of Figure 6.11. Within each cluster, rows are ordered by difference in histone modification between β-catenin floxed and deleted. Notable genes are labelled on the right of the RNA-Seq heatmap.

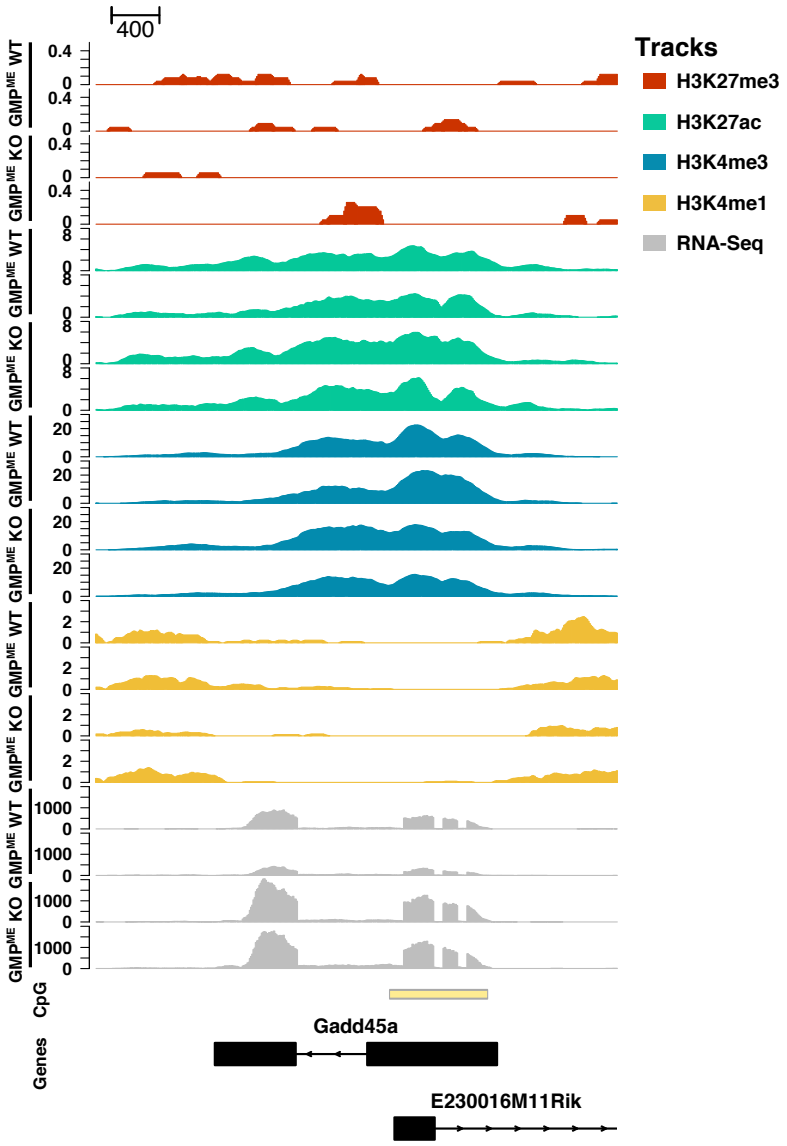

Figure 6.17: Histone modification and RNA-Seq tracks at the *Gadd45a* locus The biological condition for each track is labelled along the left side. The scale of the tracks is the TMM normalised CPM. The assay for each track is indicated by its colour and is decoded by the key (top right). The distance key (top left) specifies the length of 400 bp. Below the signal tracks, the location of a CpG island is indicated by a yellow bar. At the bottom, the position and exons of gene models are drawn.

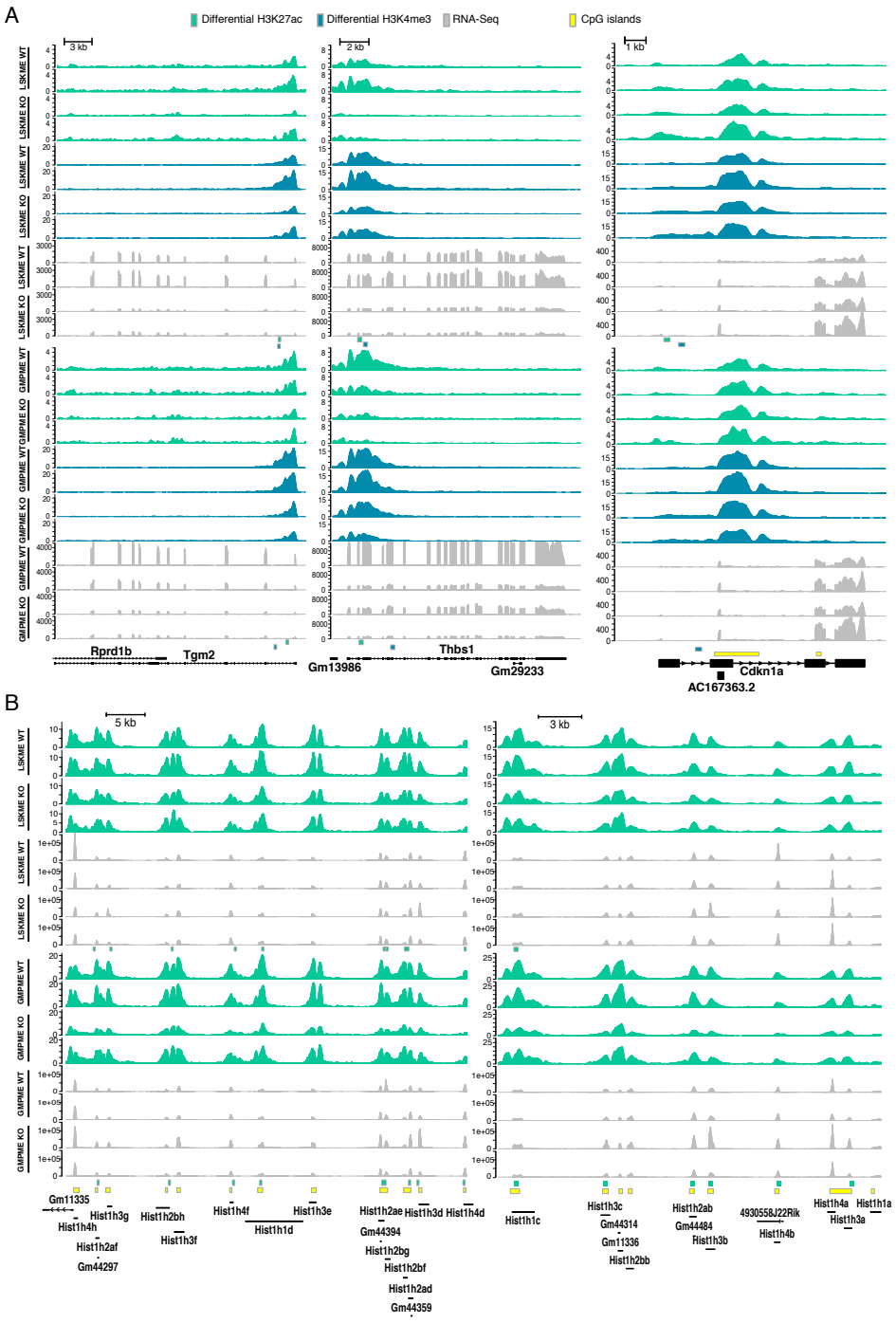

Figure 6.18: ChIP signal at selected significantly differentially modified regions following β-catenin knock-out The scale of the tracks is the TMM normalised CPM. Below signal tracks, blocks indicate the location of CpG islands and windows with significantly differential signal as decoded by the key (top). The colour key also indicates the assay of the signal tracks. At the bottom of each plot are the gene models at the given loci. Each plot has a key indicating the region size at the top. A) H3K27ac, H3K4me3 and RNA-Seq signal at *Tgm2*, *Thbs1* and *Cdkn1a*. B) H3K27ac and RNA-Seq signal across clusters of histone genes.

6.3.12 β-catenin related changes in histone modifications: pathways

To elucidate whether β-catenin conditional changes in histone modification are coordinated into pathways, I performed GO and KEGG pathway analysis for their associated genes. Genes with increased H3K4me3 in both LSK^{ME} and GMP^{ME} β-catenin KO are enriched in GO terms involving the inflammatory response (Figure 6.19A & C). Enriched KEGG pathways include infectious diseases, the p53 signalling pathway and the Toll-like receptor signalling pathway (Figure 6.19B & D). These terms are in line with what was seen in RNA-Seq (Figure 3.8 & 4.2). Genes with significant H3K4me3 reduction were not significantly enriched in any GO terms or KEGG pathways.

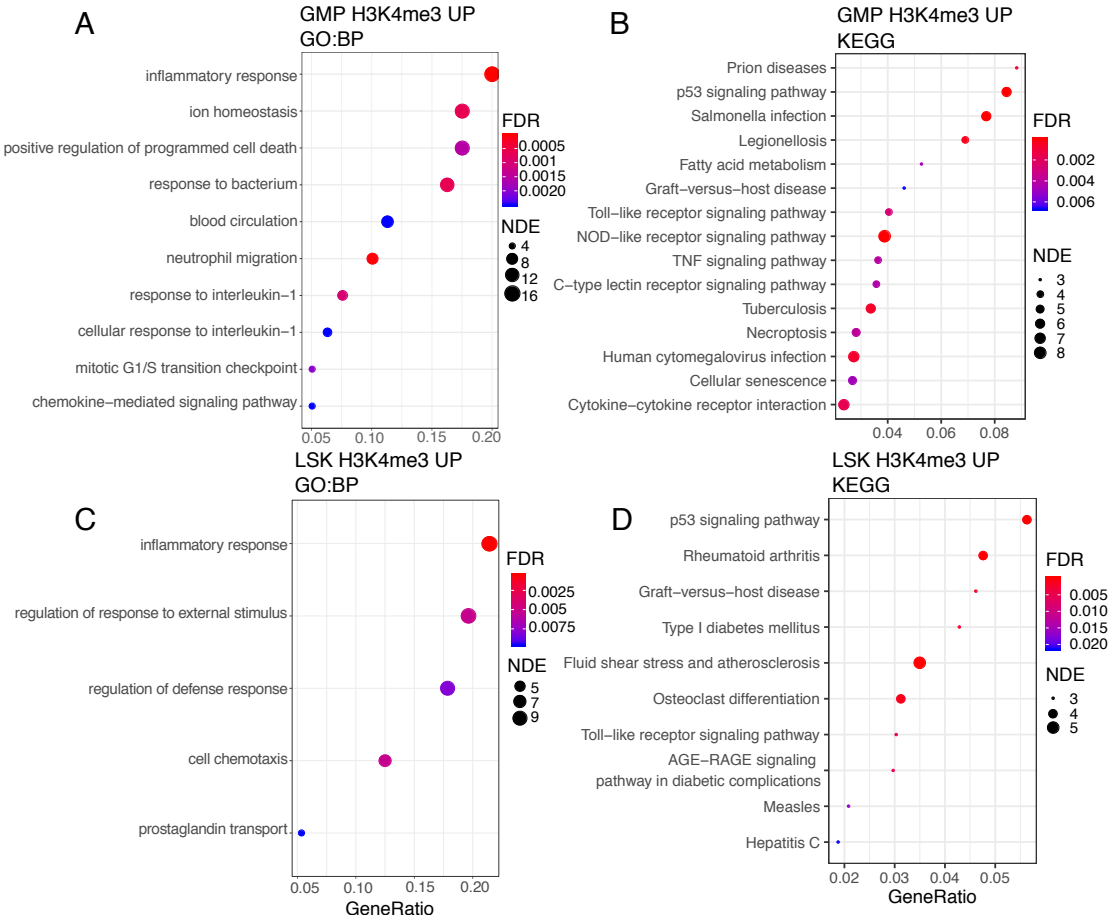

Figure 6.19: Gene ontology and pathways associated with increases in H3K4me3 following β-catenin deletion A & C) Enriched gene ontology terms for increases in H3K4me3 at the promoter in GMP^{ME} and LSK^{ME} respectively. B & D) Enriched KEGG pathways for increases in H3K4me3 at the promoter in GMP^{ME} and LSK^{ME} respectively. The y axes are labelled by the enriched GO term or KEGG pathway. The x axes show the ratio between the number of genes in the pathway and number of differentially marked genes in the enriched term. The size and colour of the dots indicate the number of differential genes (NDE) and FDR respectively.

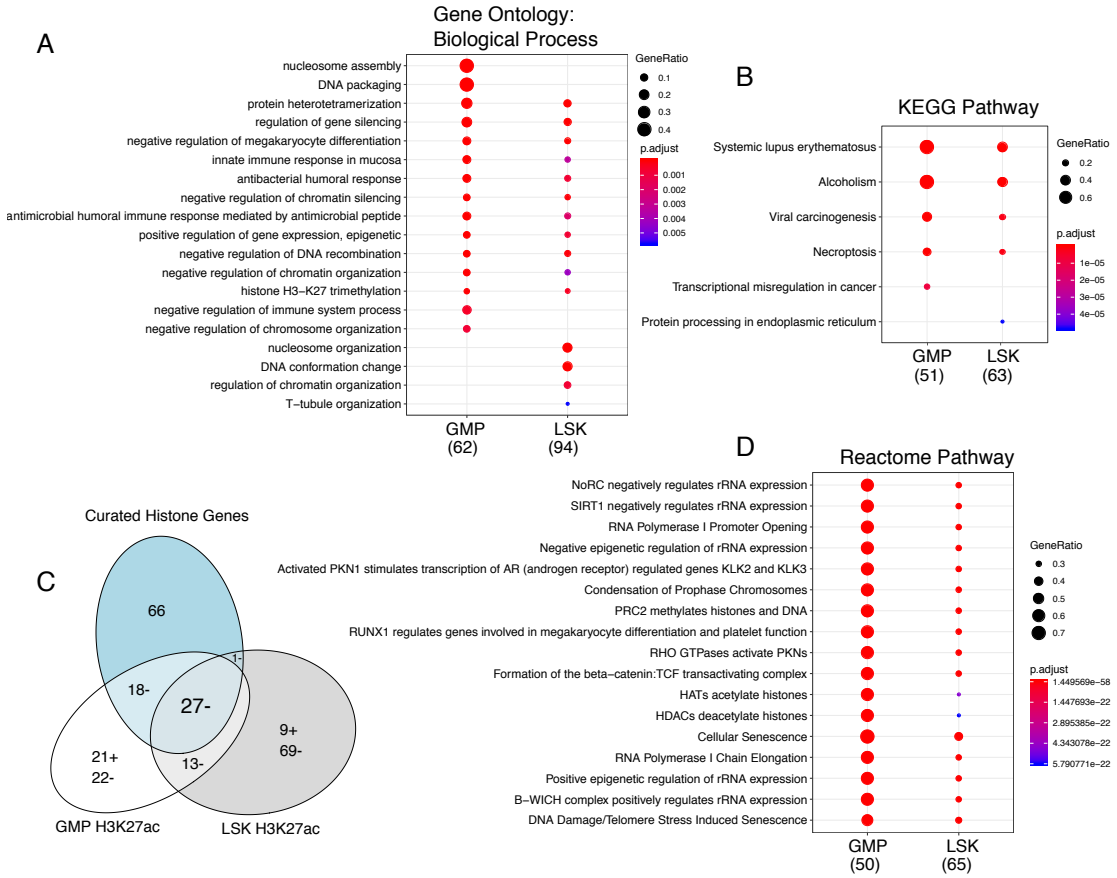

Figure 6.20: Gene ontology and pathways associated with decreases in H3K27ac following β-catenin deletion Comparison of enrichment of gene ontology terms (A), KEGG pathway (B) and Reactome pathways (D) in LSK and GMP following β-catenin deletion. C) A Euler diagram to show the overlaps between differentially acetylated promoter in LSK and GMP with a curated list of histone genes.

Intriguingly, the vast majority of genes with de-acetylation at the promoter are histone proteins Figure 6.20C. Peculiarly, almost all genes in the enriched Reactome pathways Figure 6.20D including "Formation of the β-catenin: TCF transactivation complex", are the same group of histone genes. Histone genes are expressed in G1-Phase of the cell cycle and are critical in progressing to S phase, a reduction of these genes is consistent with cell cycle arrest. GMP^{ME} has even more histone genes significantly de-acetylated than LSK^{ME} and this is reflected in additional GO-terms related to nucleosomes, these histone genes are also responsible for the KEGG pathway: Transcriptional missregulation in cancer. Specifically, non-histone genes which were uniquely deacetylated in GMP^{ME} play roles in: the cytoskeleton (Ankyrin 3: *Ank3*, Microtubule-associated proteins 1A/1B light chain 3B: *Map1lc3b*) RNA binding (Phax) ring/zinc finger (RING finger protein 10: *Rnf10*, Zinc finger with KRAB and SCAN domains 17: *Zkscan17*), small nuclear RNAs (U3B small nuclear RNA complex: *Rnu3b*, U3B small nuclear RNA 3: *Rnu3b3*) and metabolism (Alpha-Tocopherol Transfer Protein: *Ttpa*, Aldehyde

Dehydrogenase 7 Family Member A1: *Aldh7a1*, Metabolism Of Cobalamin Associated C: *Mmachc*). In LSK^{ME} uniquely downregulated genes are mostly involved in the endoplasmic reticulum and cell cycle. Common de-acetylated genes which were not histone proteins included genes involved in formation of the cytoskeleton (FYVE, RhoGEF And PH Domain Containing 4: *Fgd4*, Myosin Light Chain 10: *Myl10*) protein processing (Hypoxia up-regulated protein 1: *Hyou1*, Methyltransferase-like protein 1: *Mettl1*, Stromal Cell-Derived Factor 2-Like 1: *Sdf2l1*) ion transport (Retinoic Acid Induced 1: *Rai1*,*Nipa1*) and innate immune response (NOD-, LRR- and *Nlrp3*). As shown in RNA-Seq and H3K4me3 ChIP-Seq, Transglutaminase 2 (*Tgm2*) and Thrombospondin 1 (*Thsb1*) were also significantly deacetylated, these genes have been found to have oestrogen responsive elements at their promoters and therefore could respond directly to 4-OHT treatment.

As there were few promoters with greater H2K27ac after KO (9 in LSK^{ME}, 21 in GMP^{ME}, see Figure 6.15), these genes did not return any enriched terms by hypergeometric analysis. Although there is no overlap between H3K27ac upregulated genes in LSK^{ME} and GMP^{ME}, KO shows similar biological themes in both LSK^{ME} and GMP^{ME}. In GMP^{ME}, upregulated targets included genes which are pro-apoptotic (BCL2 Binding Component 3: *Bbc3*, Thioredoxin Interacting Protein: *Txnip*), protein processing (Exostosin Glycosyltransferase 1: *Ext1*, Glioma Pathogenesis-Related Protein 1: *Glipr1*, Sterol Carrier Protein 2: *Scp2*), zinc finger-domain containing (Zinc Finger And BTB Domain Containing 2: *Zbtb2*, Zinc finger protein 958: *Zfp958*, Forkhead box protein P1: *Foxp1*) immune response (Interleukin 1 family, member 6 *Il1f6*, Tumor necrosis factor alpha-induced protein 2: *Tnfaip2*, Interferon-induced transmembrane protein 10: *Ifitm10*, Interleukin-1 receptor-associated kinase 3: *Irak3*, Vacuole membrane protein 1: *Vmp1*, Phospholipase D Family Member 4: *Pld4*) and cell-cell interaction (Podocan Like 1: *Podnl1*, Signal Regulatory Protein Alpha: *Sirpa*, Netrin G2: *Ntng2*, Sialic Acid-Binding Ig-Like Lectin 3: *CD33*). In LSK^{ME} we also see pro-apoptotic/p53 regulated genes (Pleckstrin Homology Like Domain Family A Member 3: *Phlda3*, Cyclin Dependent Kinase Inhibitor 1A: *Cdkn1a*), protein processing (Integrator complex subunit 12: *Ints12*), zinc-finger (Zinc Finger Protein 106: *Zfp106*), cytoskeleton (Myosin-6: *Myo6*, Annexin A1:*Anxa1*, Tubulin Gamma Complex Associated Protein 3: *Tubgcp3*), immune response (T-Lymphocyte Activation Antigen: *Cd86*, *Anxa1*) and cell-cell interaction (Carbohydrate Sulfotransferase 11: *Chst11*). In summary, genes with increased acetylation in both LSK^{ME} and GMP^{ME} tell a similar tale to up-regulated H3K4me3 and RNA-Seq: β-catenin deletion in vitro with 4-OHT treatment causes apoptosis, immune response and alterations to cell adhesion.

6.3.13 β-catenin related changes in histone modifications motifs

To see if there are enriched motifs for known interactors of β-catenin, I performed gene-based motif analysis using the web-based tool, opossum (Kwon et al. 2012). There were no enriched motifs found at gene differentially modified by H3K4me3. Interestingly, for de-acetylated regions I found that the E2F1 motif was highly enriched (Table 6.5). Transcription factor E2F1 expression has been shown to be a direct target of the Wnt/β-catenin pathway (Zhou et al. 2008). HDAC1 is recruited to E2F1 target genes and reduces their transcriptional activity through de-acetylation of histones. E2F1 has been shown to cooperate with deacetylases and regulate the expression of chromatin components (Gokhman et al. 2013). This is a possible mechanism for how β-catenin deletion is implicated in cell cycle arrest through loss of E2F1 expression. However, for critical genes such as histone proteins there is transcription factor redundancy, *Sox10* and *Smad4* are also highly enriched at these genes (Table 6.5). *Smad4* is particularly interesting as it was found that *Smad4* is uniquely polycomb repressed in LSK^{ME}, (Figure 6.12). Smad4 competes with β-catenin for binding and represses β-catenin target genes (Salazar et al. 2013).

Table 6.5: Enriched motifs found within promoter regions of differentially H3K27ac marked genes

GMP^{ME}	Name	P-value	Adjusted p-value
	E2F1 (mouse)	1.96E-06	0.0004988
	E2F4 (human)	0.00002822	0.003584
	SOX10 (mouse)	0.00006366	0.00539
	SMAD4 (mouse)	0.0007246	0.04601
LSK^{ME}	Name	P-value	Adjusted p-value
	SOX10 (mouse)	8.04E-15	1.95E-12
	E2F1 (mouse)	4.13E-11	5.02E-09
	E2F4 (human)	0.0004416	0.02146
	CPEB1 (mouse)	0.0004114	0.02146
	NFYB (mouse)	0.0004114	0.02146
	NFYA (mouse)	0.001091	0.04417

6.3.14 Comparison to publicly available β-catenin ChIP-Seq

In our cell lines, β-catenin ChIP-Seq was not successful. As explained in section 1.7, β-catenin does not bind directly to DNA and cooperates with transcription factors to activate gene expression. Although cross-linking theoretically would anchor all proteins directly or indirectly bound to DNA, there is a greater chance that the epitopes will be unavailable to antibodies decreasing the likelihood of immunoprecipitation. Most of the successful high-quality ChIP-Seq data for β-catenin is shown in cell lines with a β-catenin activating mutation, APC loss or drug/ Wnt stimulation to elevate β-catenin's transcriptional activity (Table 6.2 & 6.6). Landt et al. (2012) recommends a FRiP score at least 1% where there are 1000s to 10,000s of binding sites, on the other hand, Carroll et al. (2014) suggests at least 5%. Indeed, the most relevant β-catenin ChIP-Seq data in mouse was reported by Fong et al. (2015). However it was found that this dataset was very poor quality and could not be compared with our dataset (Table 6.6). Nevertheless, we compared the putative binding sites from these analyses with significantly differentially modified regions found in this study following β-catenin deletion. We did not find any overlaps which were statistically significant compared to randomly generated regions.

6.3.15 β-catenin exon three cell lines could be used for further work

As it was discovered that the majority of successful β-catenin ChIP-Seq assays required some form of β-catenin activation (Table 6.2 & 6.6), I sought a method to activate β-catenin in MLL-ENL transformed cells. I produced cell lines using an available mouse model which carries a conditionally active β-catenin allele, the *Ctnnb1* Mmt/flox GOF allele, as described in section 2.1, Figure 2.1B. I crossed these mice with mice expressing Cre-ER to allow conditional activation of *Ctnnb1*. From these mice I sorted bone marrow into LSK and GMP populations and performed retroviral transduction and transformation with MLL-ENL as described in section 2.1.

Expression of the constitutively active form of β-catenin following 4-OHT treatment and excision of exon 3 of *Ctnnb1* was confirmed by western blot (Figure 6.21A). The band at 97 kDa indicates the expression of the full length β-catenin protein, whereas the 66 kDa band shows expression of β-catenin with exon 3 removed (Figure 6.21A). It is notable that there is a weak band at 66 kDa inETOH controls, indicating leaky Cre-recombinase (Figure 6.21A). In line with this, mice harbouring the *Ctnnb1* Mmt/flox GOF allele and Cre-ER frequently exhibited developmental deformities indicating spontaneous recombination (data not shown). To observe β-catenin in

the nucleus I employed immune fluorescence staining (Figure 6.21B). The permeabilisation technique used only allows visible foci in the nucleus of cells as it destroys cell membrane proteins. Very little β-catenin can be seen in β-catenin floxed or deleted cells (Figure 6.21B). However, β-catenin is visible in the nucleus of β-catenin GOF cells (Figure 6.21B). Using the Cell Signalling Technology #8480 β-catenin antibody, ChIP enrichment for the promoters of *Axin2*, *Cnnd1*, *Myc* and the *Myc* enhancer was significantly greater than the negative control region (Figure 6.21C). Interestingly in β-catenin GOF cells, we also see significantly increased expression of the known β-catenin target, *Cnnd1* (Figure 6.21D).

Table 6.6: Published β-catenin ChIP QC GEO: project data were derived from. Condition/Tissue/Factor: Metadata associated to sample. Reads: Number of reads in sample. Dup%: Percentage of reads marked as duplicates. FragLen: Estimated fragment length by cross-coverage method. FragLenCC: Cross-Coverage score at the fragment length. RelCC: Cross-coverage score at the fragment length over Cross-coverage at the read length. SSD: SSD score. RIP%: Percentage of reads within peaks. Wnt 'act' refers to Wnt stimulation, Wnt 'null' refers to a lack of Wnt stimulation. 'Res' refers to BET inhibitor resistance, 'Sen' refers to BET inhibitor sensitivity and DKK refers to treatment with the Wnt antagonist DKK1. 'GSKi' refers to GSK inhibitor

GEO	Condition	Tissue	Reads	Peaks	Dup%	RelCC	SSD	RiP%
GSE39837	GSKi	Nephron progenitors	27634935	5203	40%	4.1	0.83	0.95
GSE39837	GSKi	Nephron progenitors	28988702	7420	49%	4.6	0.97	0.84
GSE35213	Wnt3a	Hepatocytes	10101444	228	23%	2.2	0.66	0.12
GSE63682	Sen	ckit+ MLL-AF9	17883129	108	74%	0.55	1.3	0.021
GSE63682	Res	ckit+ MLL-AF9	17941355	9	16%	5.6	0.78	0.0039
GSE63682	Res DKK	ckit+ MLL-AF9	15020797	172	85%	0.31	2	0.083
GSE43565	No GSKi	ESCs	20970288	472	45.6%	2.49	1.57	0.0618
GSE43565	GSKi	ESCs	24924737	11647	36 %	2.33	1.62	6.17
GSE43565	GSKi	ESCs	29223738	10736	36 %	2.33	1.01	3.61
GSE84456*	-	Colon crypt	-	2915	-	-	-	-

* Raw data not available.

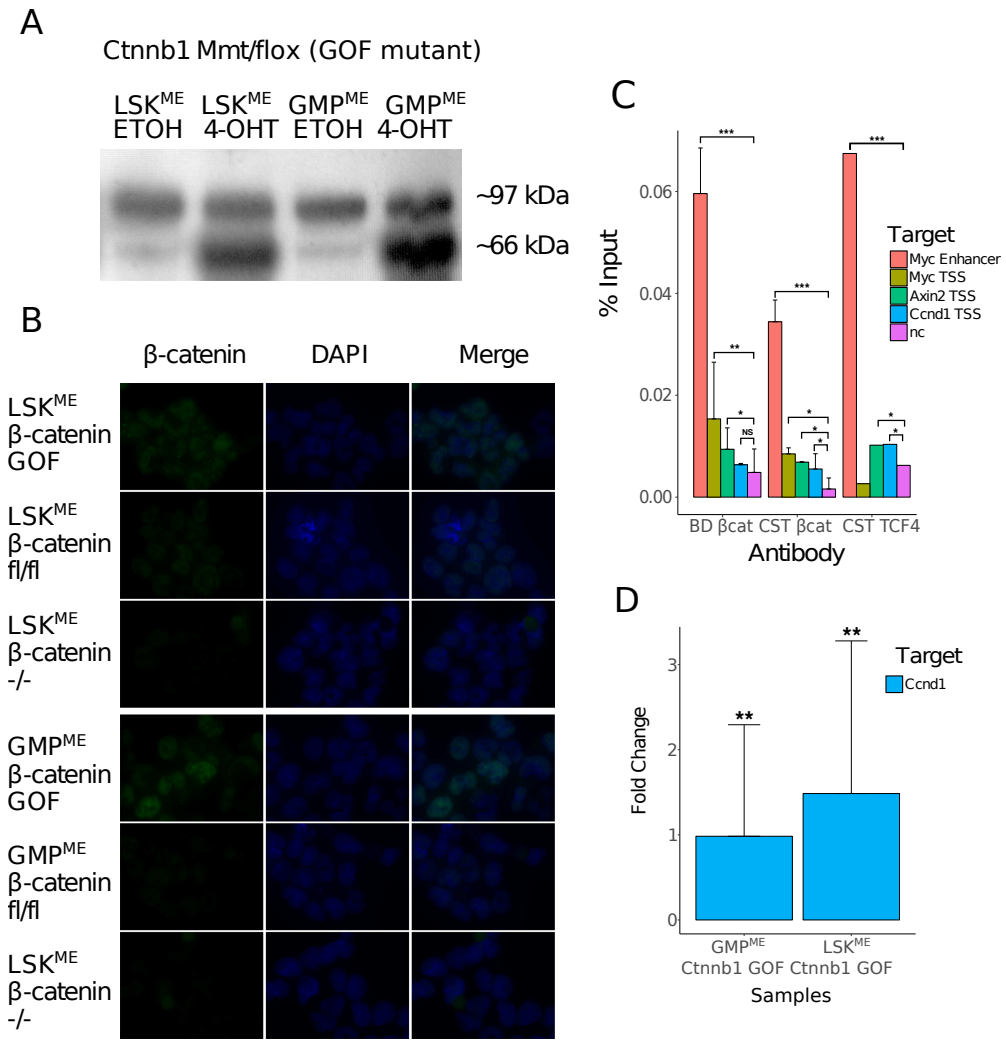

Figure 6.21: Validation of conditionally activated β-catenin MLL-ENL transformed cell lines A) Western blot for β-catenin in LSK^{ME} and GMP^{ME} Ctnnb1 Mmt/flox (GOF) heterozygous cells with the addition of ethanol ETOH or 4-OHT. Bands at 97kDa are full length β-catenin. Bands at 66kDa represent β-catenin with exon 3 excised. B) Immunofluorescence images for β-catenin (green) and DAPI (blue) in β-catenin activated (Ctnnb1 Mmt/flox with exon 3 excised), β-catenin floxed and β-catenin deleted MLL-ENL transformed cells. C) ChIP-qPCR at known β-catenin binding sites. Chromatin was precipitated using three antibodies, BD Biosciences Bcat, CST Bcat and CST TCF4. D) RT-qPCR for cyclin D1. Log fold changes are shown between ethanol controls and 4-OHT treated Ctnnb1 Mmt/flox LSK^{ME}s and GMP^{ME}s. (*P $<$ 0.05, **P $<$ 0.01,***P $<$ 0.001, two tailed t-test)

6.4 Discussion

The aim of this chapter was to uncover epigenetic changes following β-catenin KO and elucidate potential cell-of-origin "epigenetic memory". Firstly it was neccessary to determine the quality of ChIP-Seq enrichment. We obtained good quality ChIP-Seq data (Table 6.4) which was shown to follow expected patterns of spatial deposition around the transcription start sites of genes (Figure 6.2) and correlates with gene expression (Figure 6.2 & Figure 6.3. The only highlighted issues in quality were the unexpectedly high duplication rate of libraries of broad modifications Table 6.4. Of all the marks H3K4me1 appeared to be the least optimal (Table 6.4). Despite some noticable variation in enrichment between histone marks and samples, all samples were acceptable by ENCODE standards (ENCODE Project Consortium 2019).

Further confirmation of the expected arrangement and relationships between different histone marks was realised by unsupervised pattern discovery with ChromHMM (Figure 6.6A). A six state model allowed us to cluster the genome into the main known functional transcriptionally regulatory elements; promoters and enhancers in activated, silenced, poised or bivalent form (Figure 6.6). A limitation of this method is that it uses binarised data and Boolean logic, presence or absence of signal is determined by a thershold cut-off. Therefore, if segmentation is used to compare between experimental condtions we will not detect quantitative differences in histone modification. The calling of regulatory states is also highly sensitive to the chosen threshold. This was demonstrated by frequent detection of switches between heterochromatin and states requiring only one mark (Figure 6.6 C-E). Most detectable chromatin state transitions between experimental condtions could be attributed to boundary shifts which can be seen in Figure 6.7.

To overcome the limitations of binarisation others have developed programs such as "Segway"(Chan et al. 2018) which uses continuous signal values and Dynamic Bayesian networks to segment the genome. However, although some methods predicting chromatin state changes between experimental conditions have been developed, interpretation requires a lot of additional ad hoc analyses (Sohn et al. 2015; Yen and Kellis 2015; Zeng et al. 2013). I was unable to differentiate between true state switches and boundary shifts with simple filtering methods. It was therefore assertained that this method was not suitable for this study and that large switches in chromatin state were not observed.

The HoxA locus is crucial to MLL-ENL leukaemic transformation (Zeisig et al. 2004), in normal cells, the HoxA locus confers stemness (Lawrence et al. 2005) of HSCs. Interestingly, we found that GMP^{ME} replicate "60" has similar differences with LSK^{ME} at the HoxA locus (Figure 6.8) as seen between normal GMP and HSC/LSK (Figure 6.9). On the other hand, replicate GMP^{ME} "69" shows patterns of histone modification which more closely resemble normal HSC than GMP (Figure 6.8 & 6.9). Based on modification at the HoxA locus, it be could argued that GMP^{ME} "60" is GMP-like and GMP^{ME} "69" is LSK-like. However, this assumption is not supported by our in vivo data. In the previous chapter, subsection 5.3.5, Figure 5.7, it appears that GMP^{ME} "60" does not require β-catenin in vivo and therefore based on the findings of Siriboonpiputtana et al. (2017), is acting like LSK derived cell lines.

There are two main approaches for analysing multiple ChIP-Seq samples across conditions of interest. Firstly, you may establish binding sites for your protein of interest by comparing the signal between biological replicates. When it is established that biological replicates are sufficiently similar, they may be merged to increase the sensitivity of binding site discovery. Following this you may qualitatively compare presence or absence of binding across the conditions of interest. The qualitative approach relies on sufficient sequencing depth to ensure true binding sites are captured. This method is not suited to the analysis of histone marks as they are often ubiquitously detectable, for example, H3K4me3 is found at most inactive promoters (Guenther et al. 2007). Therefore, a quantitative method is more appropriate to study the dynamics of histone modifications across conditions. Like differential expression in RNA-Seq, quantitative ChIP-Seq analysis relies on adequate biological replication for sufficient statistical power and accurate variance estimation.

Using quantitative analysis methods, in line with RNA-Seq data, activating marks were generally upregulated in GMP^{ME} whilst the suppressive mark, H3K27me3 was upregulated in LSK^{ME}. Between LSK^{ME} and GMP^{ME}, H3K4me3 exhibited the most significant changes (Figure 6.10A). However, many of these differences were not concordant with changes in gene expression (Figure 6.10B & 6.11B). Nevertheless, a number of common pathways were observed (Figure 6.13). There was a small number of genes which were significantly differentially marked by H3K27me3 between LSK^{ME} and GMP^{ME} (Figure 6.10). Quantitative analysis revealed many significant differences in H3K4me1 and H3K27ac outside of genic regions (Figure 6.10A).

To predict transcription factors which may influence cell-of-origin dependent gene expression, motif analysis was performed on cell-of-origin specific activated enhancers (Figure 6.14). Enhancers which were more active in GMP^{ME} were uniquely enriched for SOX family genes. Enhancers which were more active in LSK^{ME} were uniquely enriched for oncogenes including MYB, ERG and MYCN. Interestingly, TCF12, a downstream β-catenin target was also enriched in uniquely active LSK^{ME} enhancers. These results are in line with LSK^{ME} utilising additional self renewal pathways to GMP^{ME} (Siriboonpiputtana et al. 2017) and producing a more aggressive leukaemia (Krivtsov et al. 2013).

Quantitative analysis of histone modification changes following β-catenin revealed changes in H3K4me3 which were highly congruent with transcriptional changes (Figure 6.17 & Figure 6.18) and identified similar pathways to RNA-Seq analysis (Figure 6.19). On the other hand, many changes in H3K27ac were detected which did not correlate with gene expression (Figure 6.17 & 6.18). Interestingly, most of the genes found to be differentially marked by H3K27ac were histone genes (Figure 6.20. From this set of genes, the E2F motif was found to be highly enriched (Table 6.5). Given its roles in cancer and known interactions with Wnt/β-catenin signalling (Zhou et al. 2008; Kent and Leone 2019), it would be interesting to speculate that following β-catenin deletion, E2F transcription factor activity may decrease, leading to recruitment of deacytylases and reduction of cell cycle progression through repression of transcription for chromatin components and other genes (Gokhman et al. 2013).

For normalisation I used the TMM on enriched regions over background to adjust for efficiency bias. However, it is possible that efficiency bias is trended with high-abundance binding sites, in this case, a single linear scaling factor in will not correct for this. In order to effectively control for trended biases, a non-linear method such as quantile normalisation or cyclic loess would need to be applied (Lun and Smyth 2016). It could be that histone genes, as they are particularly highly expressed are affected by trended biases leading to false detection of differential binding. Therefore, further validation may be required to confirm the specificity of differential H3K27ac binding.

The technology of ChIP-Seq has moved swiftly in the last 5 years and increasingly it is becoming popular to perform ChIP-Seq and RNA-Seq with small populations of cells or single cells to consider the dynamics of the whole cell population (Shema et al. 2019). The current procedure used in this study requires a large bulk of cells as we are limited by the antibody efficiency and require at least 1 ng of ChIP DNA to produce a ChIP-Seq library with sufficient complexity. Because we require several million cells, we are likely to come across some heterogeneity which may reduce our ability to detect effects of treatment. Many are moving to ATAC-Seq (Assay for Transposase-Accessible Chromatin using sequencing) as a way to interogate gene expression control with good predictive capabilities and higher signal-to-noise ratio. It is also easier to perform on smaller cell numbers than ChIP-Seq (Jiang and Mortazavi 2018).

The only way to determine direct regulation of gene expression via β-catenin genome wide is with β-catenin ChIP-Seq. In this study, I was unable to produce β-catenin in LSK^{ME} or GMP^{ME} floxed *Ctnnb1* LOF cells. I was also unable to correlate any changes in histone modification with binding sites found in publicly available data (subsection 6.3.14). However, I was able to obtain enriched ChIP in LSK^{ME} and GMP^{ME} *Ctnnb1* activated GOF cells (Figure 6.21). This implies that with β-catenin activation, β-catenin ChIP-Seq may be possible in MLL-ENL Pre-LSCs. However, in the absence of β-catenin ChIP-Seq, ATAC-Seq may also be used to find potential β-catenin binding sites. High depth ATAC-Seq has been used to find protected sites within nucleosome free regions in order to detect transcription factor binding sites. However, this method, known as "footprinting" is still in its infancy and requires very high sequencing depth to give statistical assurance (Jiang and Mortazavi 2018).

In this study, we do not find widespread epigenetic deregulation following β-catenin depletion. If the transcriptional regulatory role of β-catenin plays a major part in the maintenance of leukaemia we will may not see it in vitro under normal tissue culture conditions. Generally, histone modification data provides further evidence for differences in transcription.

Chapter 7

Conclusion

Particularly aggressive subtypes of acute leukaemia harbouring MLL gene translocations have been described as "prototypical epigenetically driven cancer" (Neff and Armstrong 2013). It has been widely shown that β-catenin is critically involved in the initiation, progression and drug resistance of MLL fusion leukaemia (Gandillet et al. 2011; Griffiths et al. 2010; Siapati et al. 2011; Wang et al. 2010a; Yeung et al. 2010). We hypothesised that β-catenin, which interacts with many epigenetic modulators plays a key role in the epigenetic transformation of MLL-fusion associated leukaemia. Whilst others have investigated the functional role of β-catenin in leukaemia, to our knowledge, the potential involvement of β-catenin in fatefully shaping the epigenome of leukaemia cells had not been explored.

Despite the importance of β-catenin demonstrated in models of leukaemia, treatment of human leukaemia with inhibitors of β-catenin shows variable results (Gandillet et al. 2011). β-catenin dependency varies by disease, driver mutations and as suggested by our group, specifically in the case of MLL-ENL, cell-of-origin (Siriboonpiputtana et al. 2017). We used existing β-catenin conditional MLL-ENL transformed Pre-LSC cell lines which previously demonstrated β-catenin dependence in vivo. These cell lines were of particular interest as their in vivo dependency was specific to their cell-of-origin: whilst stem cell derived LSK Pre-LSCs were independent, GMP derived Pre-LSCs were dependent. In this thesis we aimed to characterise the transcriptional, epigenetic, and functional targets of β-catenin in MLL-ENL transformed Pre-LSCs applying a β-catenin conditional loss of function (LOF) model.

Firstly, to examine downstream transcriptional targets of β-catenin we performed RNA-Seq analyses in floxed and deleted Pre-LSCs. β-catenin loss was shown to have modest effects on transcription and did not reveal any patterns of expression in line with published results. When compared with pre-existing RNA-Seq from the same cell lines (Siriboonpiputtana et al. 2017), similar pathways were perturbed however the effect was more subtle in this repeat experiment. This was likely in-part due to accumulation of irreversible β-catenin deletions in untreated cells by leaky-Cre activation and potentially in vitro cell line evolution. Additionally, much of the previously described cell-of origin "transcriptional memory" was no longer detectable. Nevertheless, pathway analysis identified activation of innate immune pathways as the primary transcriptional readout of β-catenin deletion, with similar effects in both LSK^{ME} and GMP^{ME}.

It was initially thought that the absence of known β-catenin targets in our knock-out transcriptional signature signified unique roles of β-catenin in AML. Interestingly, comparing transcriptomic data of studies utilising the same β-catenin deletion model revealed that the effects of β-catenin are highly context dependent, even in MLL fusion-associated leukaemia. Another explanation for the lack of classical β-catenin targets seen in β-catenin knockout studies as compared with β-catenin activation studies is transcriptional compensatory mechanisms activated by the expression of faulty β-catenin mRNA (Ma et al. 2019). Although suppression of the innate immune system is an intriguing role for β-catenin in AML, similar non-specific signatures had been found in studies using the same Cre-ER system for different knockouts which may reflect the cytotoxic effects of Cre activation. Therefore, we must interpret these results with caution and further validating work with appropriate Cre-negative controls needs to be done to confirm that this response is specific. To support that suppression of the immune response is a potential reason for β-catenin requirement in GMP^{ME}, one suggestion could be to use anti-inflammatory cytokines to attempt to rescue the phenotype. Similarly to how caffeine treatment was used to inhibit the DNA damage response caused by MLL-ENL expression in a knock-in mouse model (Takacova et al. 2012).

With these RNA-Seq data we then sought to examine potentially functionally important β-catenin targets through expression perturbation. Because β-catenin was required in GMP^{ME} specifically, it was expected that downstream β-catenin targets specific to GMP^{ME} would have greater therapeutic potential. *Gadd45a* was found to be a unique β-catenin target in GMP^{ME} and was validated by qPCR. However it was found that *Gadd45a* overexpression indiscriminately eliminates leukaemia arising from GMP^{ME} and LSK^{ME} in vivo. *Gadd45a* is already known to be a tumour suppressor, therefore it is not too surprising that extreme overexpression would

impair leukaemia. We may be able to observe a cell-of-origin dependent response to *Gadd45a* by reducing its overexpression to more physiological levels by using a retroviral vector with a weaker promoter and titrating the virus. To illuminate the function of our β-catenin targets it was originally planned to probe our entire β-catenin knockout signature with an in vivo shRNA screen. As shown in chapter 5, we were unable to replicate the cell-of-origin specific dependency of β-catenin in vivo or even significantly delay leukaemia development. Additionally, for in vivo experiments in this study, there were issues with inconsistent engraftment. It was revealed that β-catenin dependency may be lost with prolonged culture, possibly due to transcriptional adaptation. This phenomenon greatly limited our ability to address our biological questions.

Finally, to investigate potential epigenetic modifications mediated by β-catenin we produced ChIP-Seq for histone modifications. In line with RNA-Seq analyses, ChIP-Seq revealed subtle epigenetic responses to β-catenin loss. Whilst only modest statistically significant changes were detected, H3K4me3 and H3K27ac were the most dynamic and generally mirrored RNA-Seq regardless of FDR cut-offs. The innate immune response signatures revealed in RNA-Seq were further confirmed via epigenetic analyses. Intriguingly, H3K27ac appeared to be perturbed at replication dependent histone clusters which were enriched for E2F motifs. It is interesting to speculate that this could reveal a mechanism by which β-catenin modulates cell cycle progression (Gokhman et al. 2013). However, these changes did not correlate with gene expression, therefore it is difficult to interpret their biological readout. The only way to detect β-catenin direct targets with confidence is via ChIP-Seq. In this study I was unable to produce ChIP-Seq for β-catenin, I speculate that this is largely due to the low presence of nuclear β-catenin which may be exacerbated by leaky Cre reducing the amount of available β-catenin. We were also unable to correlate differential ChIP-Seq regions with binding sites from β-catenin ChIP in other tissues. In this study, ChIP-Seq did not shed light on how β-catenin modulates the expression of its targets in MLL-fusion associated leukaemia. Similar results were likely to have been obtained with ATAC-Seq which would be more time and cost affective.

In summary, major transcriptomic or epigenetic changes in MLL pre-LSCs following β-catenin loss were not detectable using our methods. Inconsistent behaviour of the cell lines in relation to β-catenin dependency casts considerable doubt over the biological relevance of our transcriptomic signatures. Genomic sequencing technologies such as ChIP-Seq and RNA-Seq are extremely sensitive which makes them ideal for uncovering systemic patterns explaining underlying molecular mechanisms for the observable biology. However, the sensitivity of genomewide studies means that all variation from unwanted sources will be captured. To avoid losing biologically

interesting signal to confounding effects, we need to reduce variation as much as possible whilst maximising the biological signal being investigated without becoming too artificial.

7.1 Suggestions for Future Work

To address the original biological question to elucidate the transcriptional and epigenetic changes mediated by β-catenin we could use either a β-catenin LOF model or a GOF model. As mutations in the β-catenin pathway such as *Axin2* or *APC* are rare in leukaemia, a LOF was deemed to be more physiologically relevant for this study. However the effect of deletion by 4-OHT administration was greatly reduced by decreasing level of β-catenin due to leaky Cre-recombinase activity and may also be affected by NITC. Moreover, insufficient molecular groundwork was done to determine that β-catenin was activated highly enough in these Pre-LSCs to influence transcription before it was deleted. It is important to note that Pre-LSC were previously reported not to show a phenotype related to β-catenin deletion in vitro (Siriboonpiputtana 2014).

It is known that although MLL fusions alone can somewhat activate β-catenin in a subpopulation of cells, the activation of β-catenin is much higher in fully fledged LSCs (Yeung et al. 2010). To counteract these problems, one proposal could be to repeat molecular analyses using LSCs rather than Pre-LSCs and to use a self-excising Cre (Silver and Livingston 2001) to delete β-catenin. However, the mechanisms for β-catenin activation in MLL-fusion transformed LSCs are not clearly defined, these could be is genetic, epigenetic or due to external signalling from the bone marrow niche. LSCs will develop different mutations and undergo clonal expansion in the mouse adding to unwanted variation to the study. To avoid obfuscating effects, it may be better to explore β-catenin perturbation methods to create a cleaner model which mimics how β-catenin is activated in clinically relevant leukaemia.

Most studies with good quality β-catenin ChIP-Seq data and most RNA-Seq datasets studying β-catenin transcriptional control use some form of β-catenin stimulation, either through chemical inhibitors of the β-catenin destruction complex, Wnt ligand or by genetic mutations known to disrupt the destruction complex. To this end I created MLL-ENL transformed cell lines expressing constitutively active β-catenin and was able to obtain β-catenin ChIP. It could be argued however that this is not physiologically relevant as β-catenin mutations are generally not seen in leukaemia. As mentioned in section 1.5, FLT3-ITD mutations have been shown to increase β-catenin protein by direct phosphorylation, Indeed MLL-ENL and FLT3-ITD models

have been created with retroviral transformation (Barabé et al. 2007; Kajiguchi et al. 2007). Alternatively, activation of β-catenin could be achieved by use of controllable external factors such as a combination of Rspo3 and Wnt3a as suggested by Yi et al. (2013) which was found to activate β-catenin in MLL-AF9 LSCs.

One of the other limitations in this study was heterogeneity in the cell lines used. Firstly, it could be seen that cell lines became more variable over culture time and therefore culture time should be strictly minimised for in vitro molecular and in vivo studies. It is known that the cell lines were derived from multiple mice whose bone marrow was pooled (Siriboonpiputtana 2014). Pooling individuals will increase the heterogeneity in each cell line and could contribute to the variance seen in this study. It is possible to get enough material from a single mouse to produce retrovirally transformed cell lines from LSKs or GMPs and where possible, litter mates should be used as replicates. Furthermore avoiding retroviral transformation by use of MLL-ENL mouse models could greatly improve the detection of β-catenin's molecular signature. However, these models would not produce a GMPs derived transplantable leukaemia.

The greatest improvement for this study would be to increase the number of biological replicates for RNA-Seq and ChIP-Seq. It is not uncommon to find inconsistencies between RNA-Seq studies with few biological replicates (Baccarella et al. 2018; Peixoto et al. 2015; Conesa et al. 2016; Liu et al. 2014; Schurch et al. 2016). The number of replicates required depends on the variance of the biological system; the greater the variance the more replicates are needed to maintain error control. It has been recommended that at least six biological replicates are used for RNA-Seq for reproducible results (Schurch et al. 2016). In this study, we obtained approximately 10-20 million mapped reads for RNA-Seq. However, by perturbing read depth and replicates in analysis Baccarella et al. (2018) found that differential expression analysis performance was not negatively affected above 2 million reads. On the other hand, they determined that a loss of performance was shown when less than seven biological replicates were used for each condition (Baccarella et al. 2018). Considering this, I would recommend that future studies require at least 6 replicates. Similarly, differential ChIP-Seq analyses using count data will benefit from greater replication over sequencing depth.

The multidisciplinary perspective of this thesis provides unique insight into the interpretation of multi-omics data. Taken together, this study contributes to the scientific community by highlighting key limitations in of the use of genomics to explain the molecular mechanisms of transcription factor activity in the context of leukaemia. The need for strict experimental design control, replicates and use of appropriate models is discussed.

References

Abraham, B.J., Cui, K., Tang, Q., and Zhao, K. (2013). Dynamic regulation of epigenomic landscapes during hematopoiesis. BMC Genomics *14*.1, 193.

Adli, M., Zhu, J., and Bernstein, B.E. (2010). Genome-wide chromatin maps derived from limited numbers of hematopoietic progenitors. Nature Methods *7*.8, 615–618.

Adolfsson, J., Månansson, R., Buza-Vidas, N., Hultquist, A., Liuba, K., Jensen, C.T., Bryder, D., Yang, L., Borge, O.-J., Thoren, L.A., Anderson, K, Sitnicka, E, Sasaki, Y, Sigvardsson, M, and Jacobsen, S. (2005). Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential: a revised road map for adult blood lineage commitment. Cell *121*.2, 295–306.

Aker, M., Tubb, J., Miller, D.G., Stamatoyannopoulos, G., and Emery, D.W. (2006). Integration bias of gammaretrovirus vectors following transduction and growth of primary mouse hematopoietic progenitor cells with and without selection. Molecular Therapy: the Journal of the American Society of Gene Therapy *14*.2, 226–235.

Allfrey, V.G. and Mirsky, A.E. (1964). Structural Modifications of Histones and their Possible Role in the Regulation of RNA Synthesis. Science *144*.3618, 559.

Álvarez-Errico, D., Vento-Tormo, R., Sieweke, M., and Ballestar, E. (2015). Epigenetic control of myeloid cell differentiation, identity and function. Nature Reviews Immunology *15*.1, 7–17.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq–A Python framework to work with high-throughput sequencing data. Bioinformatics *31*.2, 166–169.

Andersson, R. et al. (2014). An atlas of active enhancers across human cell types and tissues. Nature *507*.7493, 455–461.

Andrews, S. (2015). *FastQC*. [Online; accessed Dec-2015; version 0.11.2]. URL: `http://www.bioinformatics.babraham.ac.uk/projects/fastqc`.

Ayton, P.M. and Cleary, M.L. (2003). Transformation of myeloid progenitors by MLL oncoproteins is dependent on Hoxa7 and Hoxa9. Genes & Development *17*.18, 2307.

Baccarella, A., Williams, C.R., Parrish, J.Z., and Kim, C.C. (2018). Empirical assessment of the impact of sample number and read depth on RNA-Seq analysis workflow performance. BMC Bioinformatics *19*.1, 1–12.

Bagger, F.O., Rapin, N., Theilgaard-Mönch, K., Kaczkowski, B., Thoren, L.A., Jendholm, J., Winther, O., and Porse, B.T. (2013). HemaExplorer: a database of mRNA expression profiles in normal and malignant haematopoiesis. Nucleic acids research *41*.D1, D1034–D1039.

Balgobind, B.V., Zwaan, C.M., Pieters, R., and Heuvel-Eibrink, M.M. Van den (2011). The heterogeneity of pediatric MLL-rearranged acute myeloid leukemia. Leukemia *25*.8, 1248.

Banerji, V., Frumm, S.M., Ross, K.N., Li, L.S., Schinzel, A.C., Hahn, C.K., Kakoza, R.M., Chow, K.T., Ross, L., Alexe, G., Tolliday, N., Inguilizian, H., Galinsky, I., Stone, R.M., DeAngelo, D.J., Roti, G., Aster, J.C., Hahn, W.C., Kung, A.L., and Stegmaier, K. (2012).

The intersection of genetic and chemical genomic screens identifies GSK-3α as a target in human acute myeloid leukemia. The Journal of clinical investigation *122*.3, 935–947.

Bannister, A.J. and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Research *21*.3, 381–395.

Barabé, F., Kennedy, J.A., Hope, K.J., and Dick, J.E. (2007). Modeling the initiation and progression of human acute leukemia in mice. Science *316*.5824, 600–604.

Barker, N. and Clevers, H. (2000). Catenins, Wnt signaling and cancer. Bioessays *22*.11, 961–965.

Barker, N., Hurlstone, A., Musisi, H., Miles, A., Bienz, M., and Clevers, H. (2001). The chromatin remodelling factor Brg-1 interacts with β-catenin to promote target gene activation. The EMBO Journal *20*.17, 4935–4943.

Barreto, G., Schaefer, A., Marhold, J., Stach, D., Swaminathan, S.K., Handa, V., Doederlein, G., Maltry, N., Wu, W., Lyko, F., and Niehrs, C. (2007). Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. Nature *445*.7128, 671–675.

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-Resolution Profiling of Histone Methylations in the Human Genome. Cell *129*.4, 823–837.

Bauer, A., Chauvet, S., Huber, O., Usseglio, F., Rothbächer, U., Aragnol, D., Kemler, R., and Pradel, J. (2000). Pontin52 and Reptin52 function as antagonistic regulators of β-catenin signalling activity. The EMBO Journal *19*.22, 6121–6130.

Baum, C., Düllmann, J., Li, Z., Fehse, B., Meyer, J., Williams, D.A., and Kalle, C. von (2003). Side effects of retroviral gene transfer into hematopoietic stem cells. Blood *101*.6, 2099–2114.

Behrens, J. urgen, Kries, J.P. von, uhl, M.K., Bruhn, L., Wedlich, D., Grosschedl, R., and Birchmeier, W. (1996). Functional interaction of β-catenin with the transcription factor LEF-1. Nature *382*.6592, 638–642.

Ben-David, U., Ha, G., Tseng, Y.-Y., Greenwald, N.F., Oh, C., Shih, J., McFarland, J.M., Wong, B., Boehm, J.S., Beroukhim, R., and Golub, T.R. (2017). Patient-derived xenografts undergo mouse-specific tumor evolution. Nature Genetics *49*.11, 1567–1575.

Ben-David, U., Siranosian, B., Ha, G., Tang, H., Oren, Y., Hinohara, K., Strathdee, C.A., Dempster, J., Lyons, N.J., Burns, R., Nag, A., Kugener, G., Cimini, B., Tsvetkov, P., Maruvka, Y.E., O'Rourke, R., Garrity, A., Tubelli, A.A., Bandopadhayay, P., Tsherniak, A., Vazquez, F., Wong, B., Birger, C., Ghandi, M., Thorner, A.R., Bittker, J.A., Meyerson, M., Getz, G., Beroukhim, R., and Golub, T.R. (2018). Genetic and transcriptional evolution alters cancer cell line drug response. Nature *560*.7718, 325–330.

Benjamin, D.Y., Hess, J.L., Horning, S.E., Brown, G.A., and Korsmeyer, S.J. (1995). Altered Hox expression and segmental identity in Mll-mutant mice. Nature *378*.6556, 505–508.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Statistical Methodology) *57*.1, 289–300.

Bennett, J.M., Catovsky, D., Daniel, M.-T., Flandrin, G., Galton, D., Gralnick, H.R., and Sultan, C. (1976). Proposals for the Classification of the Acute Leukaemias French-American-British (FAB) Co-operative Group. British Journal of Haematology *33*.4, 451–458.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S.L., and Lander, E.S. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*.2, 315–326.

Bernt, K.M., Zhu, N., Sinha, A.U., Vempati, S., Faber, J., Krivtsov, A.V., Feng, Z., Punt, N., Daigle, A., Bullinger, L., Pollock, R.M., Richon, V.M., Kung, A.L., and Armstrong, S.A.

(2011). MLL-Rearranged Leukemia Is Dependent on Aberrant H3K79 Methylation by DOT1L. Cancer Cell *20*.1, 66–78.

Bird, A. (2007). Perceptions of epigenetics. Nature *447*.7143, 396–398.

BLUEPRINT Consortium (2015). *A BLUEPRINT of Haematopoietic Epigenomes*. [Online; accessed Dec-2015]. URL: `http://www.blueprint-epigenome.eu`.

Bock, C., Beerman, I., Lien, W.-H., Smith, Z.D., Gu, H., Boyle, P., Gnirke, A., Fuchs, E., Rossi, D.J., and Meissner, A. (2012). DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. Molecular Cell *47*.4, 633–647.

Bocker, M.T., Hellwig, I., Breiling, A., Eckstein, V., Ho, A.D., and Lyko, F. (2011). Genome-wide promoter DNA methylation dynamics of human hematopoietic progenitor cells during differentiation and aging. Blood *117*.19, e189.

Bonnet, D. (2005). Normal and leukaemic stem cells. British Journal of Haematology *130*.4, 479.

Bonnet, D. and Dick, J.E. (1997). Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nature Medicine *3*.7, 737.

Botrugno, O., Fayard, E., Annicotte, J.-S., Haby, C., Brennan, T., Wendling, O., Tanaka, T., Kodama, T., Thomas, W., Auwerx, J., and Schoonjans, K. (2004). Synergy between LRH-1 and β-Catenin Induces G1 Cyclin-Mediated Cell Proliferation. Molecular Cell *15*, 499–509.

Bottomly, D., Kyler, S.L., McWeeney, S.K., and Yochum, G.S. (2010). Identification of beta-catenin binding regions in colon cancer cells using ChIP-Seq. Nucleic Acids Research *38*.17, 5745.

Brault, V., Moore, R., Kutsch, S., Ishibashi, M., Rowitch, D.H., McMahon, A.P., Sommer, L., Boussadia, O., and Kemler, R. (2001). Inactivation of the (β)-catenin gene by Wnt1-Cre-mediated deletion results in dramatic brain malformation and failure of craniofacial development. Development *128*.8, 1253–1264.

Bridgewater, D., Cox, B., Cain, J., Lau, A., Athaide, V., Gill, P.S., Kuure, S., Sainio, K., and Rosenblum, N.D. (2008). Canonical WNT/beta-catenin signaling is required for ureteric branching. Developmental Biology *317*.1, 83–94.

Broad Institute (2017). *Picard Tools*. URL: `http://broadinstitute.github.io/picard/`.

Bröske, A.-M., Vockentanz, L., Kharazi, S., Huska, M.R., Mancini, E., Scheller, M., Kuhl, C., Enns, A., Prinz, M., Jaenisch, R., Nerlov, C., Leutz, A., Andrade-Navarro, M.A., Jacobsen, S.E.W., and Rosenbauer, F. (2009). DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. Nature Genetics *41*.11, 1207–1215.

Brownell, J.E. and Allis, C.D. (1995). An activity gel assay detects a single, catalytically active histone acetyltransferase subunit in Tetrahymena macronuclei. Proceedings of the National Academy of Sciences of the United States of America *92*.14, 6364–6368.

Bussmann, L.H., Schubert, A., Vu Manh, T.P., De Andres, L., Desbordes, S.C., Parra, M., Zimmermann, T., Rapino, F., Rodriguez-Ubreva, J., Ballestar, E., and Graf, T. (2009). A robust and highly efficient immune cell reprogramming system. Cell Stem Cell *5*.5, 554–566.

Butler, J. and Kadonaga, J. (2002). The RNA polymerase II core promoter: a key component in the regulation of gene expression. Genes & Development *16*.20, 2583–2592.

Butler, J.S. and Dent, S.Y. (2013). The role of chromatin modifiers in normal and malignant hematopoiesis. Blood *121*.16, 3076–3084.

Calo, E. and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? Molecular Cell *49*.5, 825–837.

Cancer Genome Atlas Research Network et al. (2013). Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. The New England Journal of Medicine *368*.22, 2059–2074.

Cancer Research UK (2015). *Acute myeloid leukaemia (AML) statistics*. [Online; accessed Dec-2015]. URL: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/leukaemia-aml.

Carrera, I., Janody, F., Leeds, N., Duveau, F., and Treisman, J.E. (2008). Pygopus activates Wingless target gene transcription through the mediator complex subunits Med12 and Med13. Proceedings of the National Academy of Sciences of the United States of America *105*.18, 6644–6649.

Carroll, T.S., Liang, Z., Salama, R., Stark, R., and Santiago, I. de (2014). Impact of artifact removal on ChIP quality metrics in ChIP-seq and ChIP-exo data. Frontiers in Genetics *5*, 75.

Catherine, B.M. (2013). Organizing the genome with H2A histone variants. Biochemical Journal *449*.3, 567–579.

Cavallo, R.A., Cox, R.T., Moline, M.M., Roose, J., Polevoy, G.A., Clevers, H., Peifer, M., and Bejsovec, A. (1998). Drosophila Tcf and Groucho interact to repress Wingless signalling activity. Nature *395*.6702, 604–608.

Challen, G.A., Sun, D., Jeong, M., Luo, M., Jelinek, J., Berg, J.S., Bock, C., Vasanthakumar, A., Gu, H., Xi, Y., Liang, S., Lu, Y., Darlington, G.J., Meissner, A., Issa, J.-P.J., Godley, L.A., Li, W., and Goodell, M.A. (2012). Dnmt3a is essential for hematopoietic stem cell differentiation. Nature Genetics *44*.1, 23–31.

Chan, R.C., Libbrecht, M.W., Roberts, E.G., Bilmes, J.A., Noble, W.S., and Hoffman, M.M. (2018). Segway 2.0: Gaussian mixture models and minibatch training. Bioinformatics *34*.4, 669–671.

Chan, S.M. and Majeti, R. (2013). Role of DNMT3A, TET2, and IDH1/2 mutations in pre-leukemic stem cells in acute myeloid leukemia. International Journal of Hematology *98*.6, 648–657.

Chan, W.-I. and Huntly, B.J. (2008). Leukemia stem cells in acute myeloid leukemia. Seminars in oncology *35*.4, 326–335.

Chen, L. et al. (2014). Transcriptional diversity during lineage commitment of human blood progenitors. Science *345*.6204, 1251033.

Chen, W., Kumar, A.R., Hudson, W.A., Li, Q., Wu, B., Staggs, R.A., Lund, E.A., Sam, T.N., and Kersey, J.H. (2008). Malignant transformation initiated by Mll-AF9: gene dosage and critical target cells. Cancer Cell *13*.5, 432–440.

Cheung, N., Fung, T.K., Zeisig, B.B., Holmes, K., Rane, J.K., Mowen, K.A., Finn, M.G., Lenhard, B., Chan, L.C., and So, C.W.E. (2016). Targeting Aberrant Epigenetic Networks Mediated by PRMT1 and KDM4C in Acute Myeloid Leukemia. Cancer Cell *29*.1, 32–48.

Chung, E.J., Hwang, S.-G., Nguyen, P., Lee, S., Kim, J.-S., Kim, J.W., Henkart, P.A., Bottaro, D.P., Soon, L., Bonvini, P., Lee, S.-J., Karp, J.E., Oh, H.J., Rubin, J.S., and Trepel, J.B. (2002). Regulation of leukemic cell adhesion, proliferation, and survival by beta-catenin. Blood *100*.3, 982–990.

Chung, Y.S., Kim, H.J., Kim, T.-M., Hong, S.-H., Kwon, K.-R., An, S., Park, J.-H., Lee, S., and Oh, I.-H. (2009). Undifferentiated hematopoietic cells are characterized by a genome-wide undermethylation dip around the transcription start site and a hierarchical epigenetic plasticity. Blood *114*.24, 4968–4978.

Clevers, H. (2006). Wnt/β-catenin signaling in development and disease. Cell *127*.3, 469–480.

Cobas, M., Wilson, A., Ernst, B., Mancini, J.C., MacDonald, H.R., Kemler, R., and Radtke, F. (2004). beta-catenin is dispensable for hematopoiesis and lymphopoiesis. Journal of Experimental Medicine *199*.2, 229.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szczesniak, M.W., Gaffney, D.J., Elo, L.L., Zhang, X., and Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. Genome Biology *17*, 13.

Conlon, G.A. and Murray, G.I. (2019). Recent advances in understanding the roles of matrix metalloproteinases in tumour invasion and metastasis. The Journal of pathology *247*.5, 629–640.

Copland, M., Michie, A., and Holyoake, T. (2009). Stem cells in Leukemia and other hematological malignancies. In : Majumder S. (eds) Stem Cells and Cancer, S. Majumder, ed. (New York, NY: Springer), pp. 111–136.

Costa, G., Kouskoff, V., and Lacaud, G. (2012). Origin of blood cells and HSC production in the embryo. Trends in Immunology *33*.5, 223.

Costello, R.T., Mallet, F., Gaugler, B., Sainty, D., Arnoulet, C., Gastaut, J.-A., and Olive, D. (2000). Human acute myeloid leukemia CD34+/CD38- progenitor cells have decreased sensitivity to chemotherapy and Fas-induced apoptosis, reduced immunogenicity, and impaired dendritic cell transformation capacities. Cancer Research *60*.16, 4403–4411.

Cozzio, A., Emmanuelle, P., Ayton, P.M., Karsunky, H., Cleary, M.L., and Weissman, I.L. (2003). Similar MLL-associated leukemias arising from self-renewing stem cells and short-lived myeloid progenitors. Genes & Development *17*.24, 3029–3035.

Daniels, D.L. and Weis, W.I. (2005). β-catenin directly displaces Groucho/TLE repressors from Tcf/Lef in Wnt-mediated transcription activation. Nature Structural & Molecular Biology *12*.4, 364–371.

Danielsson, F., James, T., Gomez-Cabrero, D., and Huss, M. (Mar. 2015). Assessing the consistency of public human tissue RNA-seq data sets. Briefings in Bioinformatics *16*.6, 941–949.

Deaton, A.M., Webb, S., Kerr, A.R., Illingworth, R.S., Guy, J., Andrews, R., and Bird, A. (2011). Cell type–specific DNA methylation at intragenic CpG islands in the immune system. Genome Research *21*.7, 1074–1086.

Dietrich, P.A., Yang, C., Leung, H.H., Lynch, J.R., Gonzales, E., Liu, B., Haber, M., Norris, M.D., Wang, J., and Wang, J.Y. (2014). GPR84 sustains aberrant β-catenin signaling in leukemic stem cells for maintenance of MLL leukemogenesis. Blood, The Journal of the American Society of Hematology *124*.22, 3284–3294.

Dou, Y.L., Milne, T.A., Tackett, A.J., Smith, E.R., Fukuda, A., Wysocka, J., Allis, C.D., Chait, B.T., Hess, J.L., and Roeder, R.G. (2005). Physical association and coordinate function of the H3K4 methyltransferase MLL1 and the H4K16 acetyltransferase MOF. Cell *121*.6, 885.

Dou, Y., Milne, T.A., Ruthenburg, A.J., Lee, S., Lee, J.W., Verdine, G.L., Allis, C.D., and Roeder, R.G. (2006). Regulation of MLL1 H3K4 methyltransferase activity by its core components. Nature Structural & Molecular Biology 8, 713–719.

Doulatov, S., Notta, F., Laurenti, E., and Dick, J.E. (2012). Hematopoiesis: A Human Perspective. Cell Stem Cell *10*.2, 136.

Dufourcq, P., Descamps, B., Tojais, N.F., Leroux, L., Oses, P., Daret, D., Moreau, C., Lamazière, J.-M.D., Couffinhal, T., and Duplaa, C. (2008). Secreted frizzled-related protein-1 enhances mesenchymal stem cell function in angiogenesis and contributes to neovessel maturation. Stem Cells *26*.11, 2991–3001.

Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/ package biomaRt. Nature Protocols *4*.8, 1184–1191.

El-Brolosy, M.A. and Stainier, D.Y.R. (2017). Genetic compensation: A phenomenon in search of mechanisms. PLoS Genetics *13*.7, e1006780.

ENCODE Project Consortium (2019). *Encode Data Standards: Terms and Definitions*. [Online; accessed May-2019]. URL: `https://www.encodeproject.org/data-standards/terms/`.

ENCODE Project Consortium and others (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*.7414, 57–74.

Eppert, K., Takenaka, K., Lechman, E.R., Waldron, L., Nilsson, B., Galen, P. van, Metzeler, K.H., Poeppl, A., Ling, V., Beyene, J., Canty, A.J., Danska, J.S., Bohlander, S.K., Buske, C., Minden, M.D., Golub, T.R., Jurisica, I., Ebert, B.L., and Dick, J.E. (2011). Stem cell gene expression programs influence clinical outcome in human leukemia. Nature Medicine *17*.9, U91.

Ernst, J. and Kellis, M. (2017). Chromatin-state discovery and genome annotation with ChromHMM. Nature Protocols *12*, 2478–2492.

Essers, M., Vries-Smits, L. de, Barker, N, Polderman, P., Burgering, B., and Korswagen, H. (2005). Functional interaction between beta-catenin and FOXO in oxidative stress signaling. Science *308*.5725, 1181–1184.

Estarás, C., Benner, C., and Jones, K.A. (2015). SMADs and YAP compete to control elongation of β-catenin:LEF-1-recruited RNAPII during hESC differentiation. Molecular Cell *58*.5, 780–793.

Estrov, Z. (2009). "The leukemia stem cell". Acute Myelogenous Leukemia, (Springer, New York, NY), pp. 1–17.

Faber, J., Krivtsov, A.V., Stubbs, M.C., Wright, R., Davis, T.N., Heuvel-Eibrink, M. van den, Zwaan, C.M., Kung, A.L., and Armstrong, S.A. (2009). HOXA9 is required for survival in human MLL-rearranged acute leukemias. Blood *113*.11, 2385.

Feller, N., Velden, V. van der, Brooimans, R., Boeckx, N., Preijers, F., Kelder, A., Greef, I de, Westra, G, Te Marvelde, J., Aerts, P., Wind, H, Leenders, M, Gratama, J., and Schuurhuis, G. (2013). Defining consensus leukemia-associated immunophenotypes for detection of minimal residual disease in acute myeloid leukemia in a multicenter setting. Blood Cancer Journal *3*.8, e129.

Feng, R., Desbordes, S.C., Xie, H., Tillo, E.S., Pixley, F., Stanley, E.R., and Graf, T. (2008). PU. 1 and C/EBPα/β convert fibroblasts into macrophage-like cells. Proceedings of the National Academy of Sciences *105*.16, 6057–6062.

Feuring-Buske, M. and Hogge, D.E. (2001). Hoechst 33342 efflux identifies a subpopulation of cytogenetically normal CD34+ CD38- progenitor cells from patients with acute myeloid leukemia. Blood *97*.12, 3882–3889.

Fevr, T., Robine, S., Louvard, D., and Huelsken, J. (2007). Wnt/beta-catenin is essential for intestinal homeostasis and maintenance of intestinal stem cells. Molecular and Cellular Biology *27*.21, 7551–7559.

Fong, C.Y., Gilan, O., Lam, E.Y.N., Rubin, A.F., Ftouni, S., Tyler, D., Stanley, K., Sinha, D., Yeh, P., Morison, J., Giotopoulos, G., Lugo, D., Jeffrey, P., Lee, S.C.-W., Carpenter, C., Gregory, R., Ramsay, R.G., Lane, S.W., Abdel-Wahab, O., Kouzarides, T., Johnstone, R.W., Dawson, S.-J., Huntly, B.J.P., Prinjha, R.K., Papenfuss, A.T., and Dawson, M.A. (2015). BET inhibitor resistance emerges from leukaemia stem cells. Nature *525*.7570, 538–542.

Forster, A., Pannell, R., Drynan, L.F., McCormack, M., Collins, E.C., Daser, A., and Rabbitts, T.H. (2003). Engineering de novo reciprocal chromosomal translocations associated with Mll to replicate primary events of human cancer. Cancer cell *3*.5, 449–458.

Frankish, A., Uszczynska, B., Ritchie, G.R.S., Gonzalez, J.M., Pervouchine, D., Petryszak, R., Mudge, J.M., Fonseca, N., Brazma, A., Guigo, R., and Harrow, J. (2015). Comparison of GENCODE and RefSeq gene annotation and the impact of reference geneset on variant effect prediction. BMC Genomics *16 Suppl 8*, S2.

Funa, N.S., Schachter, K.A., Lerdrup, M., Ekberg, J., Hess, K., Dietrich, N., Honoré, C., Hansen, K., and Semb, H. (2015). β-Catenin Regulates Primitive Streak Induction through Collaborative Interactions with SMAD2/SMAD3 and OCT4. Cell Stem Cell *16*.6, 639–652.

Gandillet, A., Park, S., Lassailly, F., Griessinger, E., Vargaftig, J., Filby, A., Lister, T.A., and Bonnet, D. (2011). Heterogeneous sensitivity of human acute myeloid leukemia to beta-catenin down-modulation. Leukemia *25*.5, 780.

Geissmann, F., Manz, M.G., Jung, S., Sieweke, M.H., Merad, M., and Ley, K. (2010). Development of monocytes, macrophages, and dendritic cells. Science *327*.5966, 656–661.

Gel, B., Diez-Villanueva, A., Serra, E., Buschbeck, M., Peinado, M.A., and Malinverni, R. (2016). regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. Bioinformatics *32*.2, 289–291.

Giles, F.J., Krawczyk, J., O'Dwyer, M., Swords, R., and Freeman, C. (2014). The role of inflammation in leukaemia. Advances in Experimental Medicine and Biology *816*, 335–360.

Gillet, J.-P., Calcagno, A.M., Varma, S., Marino, M., Green, L.J., Vora, M.I., Patel, C., Orina, J.N., Eliseeva, T.A., Singal, V., Padmanabhan, R., Davidson, B., Ganapathi, R., Sood, A.K., Rueda, B.R., Ambudkar, S.V., and Gottesman, M.M. (2011). Redefining the relevance of established cancer cell lines to the study of mechanisms of clinical anti-cancer drug resistance. Proceedings of the National Academy of Sciences *108*.46, 18708–18713.

Gokhman, D., Livyatan, I., Sailaja, B.S., Melcer, S., and Meshorer, E. (2013). Multilayered chromatin analysis reveals E2f, Smad and Zfx as transcriptional regulators of histones. Nature Structural & Molecular Biology *20*.1, 119–126.

Gougelet, A., Torre, C., Veber, P., Sartor, C., Bachelot, L., Denechaud, P.-D., Godard, C., Moldes, M., Burnol, A.-F., Dubuquoy, C., Terris, B., Guillonneau, F., Ye, T., Schwarz, M., Braeuning, A., Perret, C., and Colnot, S. (2014). T-cell factor 4 and β-catenin chromatin occupancies pattern zonal liver metabolism in mice. Hepatology *59*.6, 2344–2357.

Greenblatt, S.M. and Nimer, S.D. (2014). Chromatin modifiers and the promise of epigenetic therapy in acute leukemia. Leukemia *28*.7, 1396–1406.

Griffiths, E.A., Gore, S.D., Hooker, C., McDevitt, M.A., Karp, J.E., Smith, B.D., Mohammad, H.P., Ye, Y., Herman, J.G., and Carraway, H.E. (2010). Acute myeloid leukemia is characterized by Wnt pathway inhibitor promoter hypermethylation. Leukemia & Lymphoma *51*.9, 1711–1719.

Grimwade, D., Hills, R.K., Moorman, A.V., Walker, H., Chatters, S., Goldstone, A.H., Wheatley, K., Harrison, C.J., Burnett, A.K., and Leukaemia, N.C.R.I.A. (2010). Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. Blood *116*.3, 365.

Grossmann, V, Schnittger, S, Poetzinger, F, Kohlmann, A, Stiel, A, Eder, C, Fasan, A, Kern, W, Haferlach, T, and Haferlach, C (2013). High incidence of RAS signalling pathway mutations in MLL-rearranged acute myeloid leukemia. Leukemia *27*.9, 1933–1936.

Gu, Z., Eils, R., Schlesner, M., and Ishaque, N. (2018). EnrichedHeatmap: an R/Bioconductor package for comprehensive visualization of genomic signal associations. BMC Genomics *19*.1, 234.

Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). circlize Implements and enhances circular visualization in R. Bioinformatics *30*.19, 2811–2812.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*.1, 77–88.

Haar, J.L. and Ackerman, G.A. (1971). Phase and Electron Microscopic Study of Vasculogenesis and Erythropoiesis in Yolf Sac of Mouse. Anatomical Record *170*.2, &.

Haegel, H., Larue, L., Ohsugi, M., Fedorov, L., Herrenknecht, K., and Kemler, R. (1995). Lack of beta-catenin affects mouse development at gastrulation. Development *121*.11, 3529–3537.

Hahne, F. and Ivanek, R. (2016). "Statistical Genomics: Methods and Protocols". E. Mathé and S. Davis, ed. (New York, NY: Springer New York). Chap. Visualizing Genomic Data Using Gviz and Bioconductor, pp. 335–351.

Harada, N., Tamai, Y., Ishikawa, T., Sauer, B., Takaku, K., Oshima, M., and Taketo, M.M. (1999). Intestinal polyposis in mice with a dominant stable mutation of the beta-catenin gene. The EMBO Journal *18*.21, 5931–5942.

Hargreaves, D.C. and Crabtree, G.R. (2011). ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. Cell Research *21*.3, 396–420.

Heidel, F.H., Bullinger, L., Feng, Z., Wang, Z., Neff, T.A., Stein, L., Kalaitzidis, D., Lane, S.W., and Armstrong, S.A. (2012). Genetic and Pharmacologic Inhibition of beta-Catenin Targets Imatinib-Resistant Leukemia Stem Cells in CML. Cell Stem Cell *10*.4, 424.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., Wang, W., Weng, Z., Green, R.D., Crawford, G.E., and Ren, B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nature Genetics *39*.3, 311–318.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular Cell *38*.4, 576–589.

Heuser, M., Yun, H., Berg, T., Yung, E., Argiropoulos, B., Kuchenbauer, F., Park, G., Hamwi, I., Palmqvist, L., Lai, C.K., Malina, L., Lin, G., Chaturvedi, A., Thakur, B.K., Iwasaki, M., Bilenky, M., Thiessen, N., Robertson, G., Hirst, M., Kent, D., Wilson, N., Göttgens, B., Eaves, C., Cleary, M.L., Marra, M., Ganser, A., and Humphries, R. (2011). Cell of origin in AML: susceptibility to MN1-induced transformation is regulated by the MEIS1/AbdB-like HOX protein complex. Cancer Cell *20*.1, 39–52.

Hidalgo, I., Herrera-Merchan, A., Ligos, J.M., Carramolino, L., nez, J.N., Martinez, F., Dominguez, O., Torres, M., and Gonzalez, S. (2012). Ezh1 is required for hematopoietic stem cell maintenance and prevents senescence-like cell cycle arrest. Cell Stem Cell *11*.5, 649–662.

Higashi, A.Y., Ikawa, T., Muramatsu, M., Economides, A.N., Niwa, A., Okuda, T., Murphy, A.J., Rojas, J., Heike, T., Nakahata, T., Kawamoto, H, Kita, T, and Yanagita, M (2009). Direct hematological toxicity and illegitimate chromosomal recombination caused by the systemic activation of CreERT2. The Journal of Immunology *182*.9, 5633–5640.

Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T.S., Harte, R.A., Hsu, F., Hillman-Jackson, J., Kuhn, R.M., Pedersen, J.S., Pohl, A., Raney, B.J., Rosenbloom, K.R., Siepel, A., Smith, K.E., Sugnet, C.W., Sultan-Qurraie, A., Thomas, D.J., Trumbower, H., Weber, R.J., Weirauch, M., Zweig, A.S., Haussler, D., and Kent, W.J. (2006). The UCSC Genome Browser Database: update 2006. Nucleic Acids Research *34*, D590–D598.

Hodges, E., Molaro, A., Dos Santos, C.O., Thekkat, P., Song, Q., Uren, P.J., Park, J., Butler, J., Rafii, S., McCombie, W.R., Smith, D.A., and J. Hannon, G. (2011). Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. Molecular Cell *44*.1, 17–28.

Hogart, A., Lichtenberg, J., Ajay, S.S., Anderson, S., Margulies, E.H., Bodine, D.M., and NIH Intramural Sequencing Center (2012). Genome-wide DNA methylation profiles in hematopoietic stem and progenitor cells reveal overrepresentation of ETS transcription factor binding sites. Genome Research *22*.8, 1407–1418.

Holland, J.D., Klaus, A., Garratt, A.N., and Birchmeier, W. (2013). Wnt signaling in stem and cancer stem cells. Current Opinion in Cell Biology *25*.2, 254–264.

Hu, M., Krause, D., Greaves, M., Sharkis, S., Dexter, M., Heyworth, C., and Enver, T. (1997). Multilineage gene expression precedes commitment in the hemopoietic system. Genes & Development *11*.6, 774–785.

Hu, Y., Chen, Y., Douglas, L., and Li, S. (2009). β-Catenin is essential for survival of leukemic stem cells insensitive to kinase inhibition in mice with BCR-ABL-induced chronic myeloid leukemia. Leukemia *23*.1, 109–116.

Huber et al. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. Nature Methods *12*.2, 115–121.

Huelsken, J., Vogel, R., Brinkmann, V., Erdmann, B., Birchmeier, C., and Birchmeier, W. (2000). Requirement for β-catenin in anterior-posterior axis formation in mice. The Journal of Cell Biology *148*.3, 567–578.

Huntly, B.J., Shigematsu, H., Deguchi, K., Lee, B.H., Mizuno, S., Duclos, N., Rowan, R., Amaral, S., Curley, D., Williams, I., Akashi, K, and Gilliland, D. (2004). MOZ-TIF2, but not BCR-ABL, confers properties of leukemic stem cells to committed murine hematopoietic progenitors. Cancer Cell *6*.6, 587–596.

Illingworth, R.S. and Bird, A.P. (2009). CpG islands–'a rough guide'. FEBS Letters *583*.11, 1713–1720.

Janbandhu, V.C., Moik, D., and Fässler, R. (2014). Cre recombinase induces DNA damage and tetraploidy in the absence of loxP sites. Cell Cycle *13*.3, 462–470.

Jeannet, G., Scheller, M., Scarpellino, L., Duboux, S., Gardiol, N., Back, J., Kuttler, F., Malanchi, I., Birchmeier, W., Leutz, A., Huelsken, J., and Held, W. (2008). Long-term, multilineage hematopoiesis occurs in the combined absence of beta-catenin and gamma-catenin. Blood *111*.1, 149.

Ji, J., Liu, R., Tong, T., Song, Y., Jin, S., Wu, M., and Zhan, Q. (2007). Gadd45a regulates β-catenin distribution and maintains cell-cell adhesion/contact. Oncogene *26*.44, 6396–6405.

Jiang, S. and Mortazavi, A. (2018). Integrating ChIP-seq with other functional genomics data. Briefings in functional genomics *17*.2, 104–115.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature Reviews Genetics *13*.7, 484–492.

Jude, C.D., Climer, L., Xu, D., Artinger, E., Fisher, J.K., and Ernst, P. (2007). Unique and independent roles for MLL in adult hematopoietic stem cells and progenitors. Cell Stem Cell *1*.3, 324–337.

Kaidi, A., Williams, A.C., and Paraskeva, C. (2007). Interaction between beta-catenin and HIF-1 promotes cellular adaptation to hypoxia. Nature Cell Biology *9*.2, 210–U113.

Kajiguchi, T, Chung, E., Lee, S, Stine, A, Kiyoi, H, Naoe, T, Levis, M.J., Neckers, L, and Trepel, J. (2007). FLT3 regulates β-catenin tyrosine phosphorylation, nuclear localization, and transcriptional activity in acute myeloid leukemia cells. Leukemia *21*.12, 2476.

Kamminga, L.M., Bystrykh, L.V., Boer, A. de, Houwer, S., Douma, J., Weersing, E., Dontje, B., and Haan, G. de (2006). The Polycomb group gene Ezh2 prevents hematopoietic stem cell exhaustion. Blood *107*.5, 2170–2179.

Kelly, K., Ng, D., Jayakumaran, G., Wood, G., Koide, H., and Doble, B. (2011). β-Catenin Enhances Oct-4 Activity and Reinforces Pluripotency through a TCF-Independent Mechanism. Cell Stem Cell *8*, 214–27.

Kent, L.N. and Leone, G. (2019). The broken cycle: E2F dysfunction in cancer. Nature Reviews Cancer *19*.6, 326–338.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology *14*.4.

Kim, J.H., Kim, B., Cai, L., Choi, H.J., Ohgi, K.A., Tran, C., Chen, C., Chung, C.H., Huber, O., Rose, D.W., Sawyers, C.L., Rosenfeld, M.G., and Baek, S.H. (2005). Transcriptional regulation of a metastasis suppressor gene by Tip60 and β-catenin complexes. Nature *434*.7035, 921–926.

Kim, S., Xu, X., Hecht, A., and Boyer, T.G. (2006). Mediator is a transducer of Wnt/β-catenin signaling. Journal of Biological Chemistry *281*.20, 14066–14075.

Kinzler, K.W. and Vogelstein, B. (1996). Lessons from hereditary colorectal cancer. Cell *87*.2, 159–170.

Kleer, I.D., Willems, F., Lambrecht, B., and Goriely, S. (2014). Ontogeny of myeloid cells. Frontiers in Immunology *5*.

Kleff, S., Andrulis, E.D., Anderson, C.W., and Sternglanz, R. (1995). Identification of a gene encoding a yeast histone H4 acetyltransferase. Journal of Biological Chemistry *270*.42, 24674–24677.

Klose, R.J. and Bird, A.P. (2006). Genomic DNA methylation: the mark and its mediators. Trends in Biochemical Sciences *31*.2, 89–97.

Kolde, R. (2019). Pheatmap: pretty heatmaps. R package version *1*.12.

Konopleva, M.Y. and Jordan, C.T. (2011). Leukemia stem cells and microenvironment: biology and therapeutic targeting. Journal of Clinical Oncology *29*.5, 591–599.

Kouzmenko, A., Takeyama, K, Ito, S, Furutani, T, Sawatsubashi, S, Maki, A, Suzuki, E, Kawasaki, Y, Akiyama, T, Tabata, T, and Kato, S (2004). Wnt/beta-catenin and estrogen signaling converge in vivo. Journal of Biological Chemistry *279*.39, 40255–40258.

Krasteva, V., Buscarlet, M., Diaz-Tellez, A., Bernard, M.-A., Crabtree, G.R., and Lessard, J.A. (2012). The BAF53a subunit of SWI/SNF-like BAF complexes is essential for hemopoietic stem cell function. Blood *120*.24, 4720–4732.

Krivtsov, A.V., Figueroa, M.E., Sinha, A.U., Stubbs, M.C., Feng, Z., Valk, P.J.M., Delwel, R., Doehner, K., Bullinger, L., Kung, A.L., Melnick, A.M., and Armstrong, S.A. (2013). Cell of origin determines clinically relevant subtypes of MLL-rearranged AML. Leukemia *27*.4, 860.

Krueger, F. (2015). Trim Galore! 0.3.7. [Online; accessed Dec-2015].

Kundaje, A. (2014). *(2014) mod/mouse/humanENCODE: Blacklisted genomic regions for functional genomics analysis*. [Online; accessed Dec-2015]. URL: `https://sites.google.com/site/anshulkundaje/projects/blacklists`.

Kurdistani, S.K., Tavazoie, S., and Grunstein, M. (2004). Mapping global histone acetylation patterns to gene expression. Cell *117*.6, 721–733.

Kwon, A.T., Arenillas, D.J., Hunt, R.W., and Wasserman, W.W. (2012). oPOSSUM-3: advanced analysis of regulatory motif over-representation across genes or ChIP-Seq datasets. G3: Genes, Genomes, Genetics *2*.9, 987–1002.

Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P., Chen, Y., DeSalvo, G., Epstein, C., Fisher-Aylor, K.I., Euskirchen, G., Gerstein, M., Gertz, J., Hartemink, A.J., Hoffman, M.M., Iyer, V.R., Jung, Y.L., Karmakar, S., Kellis, M., Kharchenko, P.V., Li, Q., Liu, T., Liu, X.S., Ma, L., Milosavljevic, A., Myers, R.M., Park, P.J., Pazin, M.J., Perry, M.D., Raha, D., Reddy, T.E., Rozowsky, J., Shoresh, N., Sidow, A., Slattery, M., Stamatoyannopoulos, J.A., Tolstorukov, M.Y., White, K.P., Xi, S., Farnham, P.J., Lieb, J.D., Wold, B.J., and Snyder, M. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Research *22*.9, 1813–1831.

Lane, S.W., Wang, Y.J., Celso, C.L., Ragu, C., Bullinger, L., Sykes, S.M., Ferraro, F., Shterental, S., Lin, C.P., Gilliland, D.G., Scadden, D.T., Armstrong, S.A., and Williams, D.A. (2011). Differential niche and Wnt requirements during acute myeloid leukemia progression. Blood *118*.10, 2849–2856.

Langmead, B. and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods *9*.4, 357–359.

Lara-Astiaso, D., Weiner, A., Lorenzo-Vivas, E., Zaretsky, I., Jaitin, D.A., David, E., Keren-Shaul, H., Mildner, A., Winter, D., Jung, S., Friedman, N., and Amit, I. (2014). Immunogenetics. Chromatin state dynamics during blood formation. Science *345*.6199, 943–949.

Laslo, P., Spooner, C.J., Warmflash, A., Lancki, D.W., Lee, H.-J., Sciammas, R., Gantner, B.N., Dinner, A.R., and Singh, H. (2006). Multilineage transcriptional priming and determination of alternate hematopoietic cell fates. Cell *126*.4, 755–766.

Lawrence, H.J., Christensen, J., Fong, S., Hu, Y.L., I., G., S.W., Humphries, R.K., and Largman, C. (2005). Loss of expression of the Hoxa-9 homeobox gene impairs the proliferation and repopulating ability of hematopoietic stem cells. Blood *106*.12, 3994.

Lawrence, M., Gentleman, R., and Carey, V. (2009). rtracklayer: an R package for interfacing with genome browsers. Bioinformatics *25*, 1841–1842.

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for Computing and Annotating Genomic Ranges. PLoS Computational Biology *9*.8, e1003118.

Leeuwen, F.E.V. (1996). Risk of acute myelogenous leukaemia and myelodysplasia following cancer treatment. Baillière's Clinical Haematology *9*.1, 57–85.

Lento, W., Ito, T., Zhao, C., Harris, J.R., Huang, W., Jiang, C., Owzar, K., Piryani, S., Racioppi, L., Chao, N., and Reya, T. (2014). Loss of β-catenin triggers oxidative stress and impairs hematopoietic regeneration. Genes & Development *28*.9, 995–1004.

Lento, W.E., Congdon, K.L., Voermans, C., Kritzik, M.R., and Reya, T. (2013). Wnt signaling in normal and malignant hematopoiesis. Cold Spring Harbor Perspectives in Biology *5*.2, a008011.

Lévy, L., Wei, Y., Labalette, C., Wu, Y., Renard, C.-A., Buendia, M.A., and Neuveut, C. (2004). Acetylation of *β*-catenin by p300 regulates *β*-catenin-Tcf4 interaction. Molecular and Cellular Biology *24*.8, 3404–3414.

Ley, T.J., Ding, L., Walter, M.J., McLellan, M.D., Lamprecht, T., Larson, D.E., Kandoth, C., Payton, J.E., Baty, J., Welch, J., Harris, C.C., Lichti, C.F., Townsend, R.R., Fulton, R.S., Dooling, D.J., Koboldt, D.C., Schmidt, H., Zhang, Q., Osborne, J.R., Lin, L., O'Laughlin, M., McMichael, J.F., Delehaunty, K.D., McGrath, S.D., Fulton, L.A., Magrini, V.J., Vickery, T.L., Hundal, J., Cook, L.L., Conyers, J.J., Swift, G.W., Reed, J.P., Alldredge, P.A., Wylie, T., Walker, J., Kalicki, J., Watson, M.A., Heath, S., Shannon, W.D., Varghese, N., Nagarajan, R., Westervelt, P., Tomasson, M.H., Link, D.C., Graubert, T.A., DiPersio, J.F., Mardis, E.R., and Wilson, R.K. (2010). DNMT3A mutations in acute myeloid leukemia. New England Journal of Medicine *363*.25, 2424–2433.

Li, H., Daculsi, R., Grellier, M., Bareille, R., Bourget, C., and Amedee, J. (2010). Role of neural-cadherin in early osteoblastic differentiation of human bone marrow stromal cells cocultured with human umbilical vein endothelial cells. American Journal of Physiology-Cell Physiology *299*.2, 422–430.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Proc,G.P.D. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*.16, 2079.

Lien, W.-H. and Fuchs, E. (2014). Wnt some lose some: transcriptional governance of stem cells by Wnt/beta-catenin signaling. Genes & Development *28*.14, 1532.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.-M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A.H., Thomson, J.A., Ren, B., and Ecker, J.R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*.7271, 315–322.

Liu, Y.I., Chang, M.V., Li, H.E., Barolo, S., Chang, J.L., Blauwkamp, T.A., and Cadigan, K.M. (2008). The chromatin remodelers ISWI and ACF1 directly repress Wingless transcriptional targets. Developmental Biology *323*.1, 41–52.

Liu, Y., Liu, Y., Wu, J., Roizman, B., and Zhou, G.G. (2018). Innate responses to gene knockouts impact overlapping gene networks and vary with respect to resistance to viral infection. Proceedings of the National Academy of Sciences *115*.14, e3230–e3237.

Liu, Y., Zhou, J., and White, K.P. (2014). RNA-seq differential expression studies: more sequence or more replication? Bioinformatics *30*.3, 301–304.

Loonstra, A., Vooijs, M., Beverloo, H.B., Allak, B.A., Drunen, E. van, Kanaar, R., Berns, A., and Jonkers, J. (2001). Growth inhibition and DNA damage induced by Cre recombinase in mammalian cells. Proceedings of the National Academy of Sciences *98*.16, 9209–9214.

Loosdregt, J. van, Fleskens, V., Tiemessen, M.M., Mokry, M., Boxtel, R. van, Meerding, J., Pals, C.E.G.M., Kurek, D., Baert, M.R.M., Delemarre, E.M., Gröne, A., Koerkamp, M.J.A.G., Sijts, A.J.A.M., Nieuwenhuis, E.E.S., Maurice, M.M., Es, J.H. van, Berge, D.T., Holstege, F.C., Staal, F.J.T., Zaiss, D.M.W., Prakken, B.J., and Coffer, P.J. (2013). Canonical Wnt signaling negatively modulates regulatory T cell function. Immunity *39*.2, 298–310.

Lorenz, E., Uphoff, D., Reid, T.R., and Shelton, E. (1951). Modification of irradiation injury in mice and guinea pigs by bone marrow injections. Journal of the National Cancer Institute *12*.1, 197–201.

Lorsbach, R.B., Moore, J., Mathew, S., Raimondi, S.C., Mukatira, S.T., and Downing, J.R. (2003). TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). Leukemia *17*.3, 637–641.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology *15*.12, 550.

Lowenberg, B., Downing, J.R., and Burnett, A. (1999). Acute myeloid leukemia. New England Journal of Medicine *341*.14, 1051–1062.

Luger, K., ader, A.W.M., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature *389*.6648, 251–260.

Lun, A.T.L. and Smyth, G.K. (2016). csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. Nucleic Acids Research *44*.5, e45.

Lynch, J.R., Yi, H., Casolari, D.A., Voli, F., Gonzales-Aloy, E., Fung, T.K., Liu, B., Brown, A., Liu, T., Haber, M., Norris, M.D., Lewis, I.D., So, C.W.E., D'Andrea, R.J., and Wang, J.Y. (2016). Gaq signaling is required for the maintenance of MLL-AF9-induced acute myeloid leukemia. Leukemia *30*.8, 1745–1748.

Ma, Z., Zhu, P., Shi, H., Guo, L., Zhang, Q., Chen, Y., Chen, S., Zhang, Z., Peng, J., and Chen, J. (2019). PTC-bearing mRNA elicits a genetic compensation response via Upf3a and COMPASS components. Nature *568*.7751, 259–263.

Maës, J., Maleszewska, M., Guillemin, C., Pflumio, F., Six, E., André-Schmutz, I., Cavazzana-Calvo, M., Charron, D., Francastel, C., and Goodhardt, M. (2008). Lymphoid-affiliated genes are associated with active histone modifications in human hematopoietic stem cells. Blood *112*.7, 2722–2729.

Majeti, R. and Weissman, I.L. (2011). Human acute myelogenous leukemia stem cells revisited: there's more than meets the eye. Cancer Cell *19*.1, 9–10.

Mar, B.G., Bullinger, L., Basu, E., Schlis, K., Silverman, L.B., ohner, K.D, and Armstrong, S.A. (2012). Sequencing histone-modifying enzymes identifies UTX mutations in acute lymphoblastic leukemia. Leukemia *26*.8, 1881–1883.

Maximow, A.A. (1909). Der lymphozyt als gemeinsame stammzelle der verschiedenen blutelement in der embryonalen entwicklung und im postfetalen leben der säugetiere. Folia Haematologica (Leipzig) *8*, 125–141.

McCarthy, D.J., Chen, Y., and Smyth, G.K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Research *40*.10, 4288–4297.

McCubrey, J.A., Steelman, L., Bertrand, F.E., Davis, N.M., Abrams, S.L., Montalto, G., D'Assoro, A.B., Libra, M., Nicoletti, F., Maestro, R., Basecke, J, Cocco, L, Cervello, M, and Martelli, A. (2014). Multifaceted roles of GSK-3 and Wnt/β-catenin in hematopoiesis and leukemogenesis: opportunities for therapeutic intervention. Leukemia *28*.1, 15–33.

McMahon, K.A., Hiew, S.Y.-L., Hadjur, S., Veiga-Fernandes, H., Menzel, U., Price, A.J., Kioussis, D., Williams, O., and Brady, H.J. (2007). Mll has a critical role in fetal and adult hematopoietic stem cell self-renewal. Cell Stem Cell *1*.3, 338–345.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., Gnirke, A., Jaenisch, R., and Lander, E.S. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*.7205, 766–770.

Migliaccio, A.R., Bengra, C., Ling, J., Pi, W., Li, C., Zeng, S., Keskintepe, M., Whitney, B., Sanchez, M., Migliaccio, G., and Tuan, D. (2000). Stable and unstable transgene integration sites in the human genome: extinction of the Green Fluorescent Protein transgene in K562 cells. Gene *256*.1-2, 197–214.

Milne, T.A., Briggs, S.D., Brock, H.W., Martin, M.E., Gibbs, D., Allis, C.D., and Hess, J.L. (2002). MLL targets SET domain methyltransferase activity to Hox gene promoters. Molecular Cell *10*.5, 1107–1117.

Milne, T.A., Kim, J., Wang, G.G., Stadler, S.C., Basrur, V., Whitcomb, S.J., Wang, Z., Ruthenburg, A.J., Elenitoba-Johnson, K.S.J., Roeder, R.G., and Allis, C.D. (2010). Multiple Interactions Recruit MLL1 and MLL1 Fusion Proteins to the HOXA9 Locus in Leukemogenesis. Molecular Cell *38*.6, 863.

Miyamoto, T., Iwasaki, H., Reizis, B., Ye, M., Graf, T., Weissman, I.L., and Akashi, K. (2002). Myeloid or lymphoid promiscuity as a critical step in hematopoietic lineage commitment. Developmental Cell *3*.1, 137–147.

Mochizuki-Kashio, M., Mishima, Y., Miyagi, S., Negishi, M., Saraya, A., Konuma, T., Shinga, J., Koseki, H., and Iwama, A. (2011). Dependency on the polycomb gene Ezh2 distinguishes fetal from adult hematopoietic stem cells. Blood *118*.25, 6553–6561.

Mohammad, H.P., Barbash, O., and Creasy, C.L. (2019). Targeting epigenetic modifications in cancer therapy: erasing the roadmap to cancer. Nature medicine *25*.3, 403–418.

Molenaar, M., Wetering, M. van de, Oosterwegel, M., Peterson-Maduro, J., Godsave, S., Korinek, V., Roose, J., Destrée, O., and Clevers, H. (1996). XTcf-3 transcription factor mediates β-catenin-induced axis formation in Xenopus embryos. Cell *86*.3, 391–399.

Mosimann, C., Hausmann, G., and Basler, K. (2009). β-catenin hits chromatin: regulation of Wnt target gene activation. Nature Reviews Molecular Cell Biology *10*.4, 276–286.

Mossadegh-Keller, N., Sarrazin, S., Kandalla, P.K., Espinosa, L., Stanley, E.R., Nutt, S.L., Moore, J., and Sieweke, M.H. (2013). M-CSF instructs myeloid lineage fate in single haematopoietic stem cells. Nature *497*.7448, 239–243.

Mrózek, K., Heinonen, K., Chapelle, A De la, and Bloomfield, C.D. (1997). Clinical significance of cytogenetics in acute myeloid leukemia. In Seminars in oncology, vol. 24. 1. (), pp. 17–31.

Narlikar, G.J., Sundaramoorthy, R., and Owen-Hughes, T. (2013). Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. Cell *154*.3, 490–503.

Nateri, A., Spencer-Dene, B, and Behrens, A (2005). Interaction of phosphorylated c-Jun with TCF4 regulates intestinal cancer development. Nature *437*.7056, 281–285.

NCBI (2019). *GEO2R*. [Online; accessed Jan-2019]. URL: `https://www.ncbi.nlm.nih.gov/geo/geo2r/`.

Neff, T. and Armstrong, S.A. (2013). Recent progress toward epigenetic therapies: the example of mixed lineage leukemia. Blood *121*.24, 4847–4853.

Ng, C.E.L., Sinha, A., Krivtsov, A., Dias, S., Chang, J., Armstrong, S.A., and Kalaitzidis, D. (2014). KRasG12D-evoked leukemogenesis does not require β-catenin. Leukemia *28*.3, 698–702.

Nicol, B. and Yao, H.H.C. (2015). Gonadal Identity in the Absence of Pro-Testis Factor SOX9 and Pro-Ovary Factor Beta-Catenin in Mice. Biology of Reproduction *93*.2, 35.

Nilsson, U.W., Garvin, S., and Dabrosin, C. (2007). MMP-2 and MMP-9 activity is regulated by estradiol and tamoxifen in cultured human breast cancer cells. Breast Cancer Research and Treatment *102*.3, 253–261.

Nishisho, I., Nakamura, Y., Miyoshi, Y., Miki, Y., Ando, H., Horii, A., Koyama, K., Utsunomiya, J., Baba, S., and Hedge, P. (1991). Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients. Science *253*.5020, 665–669.

Notta, F., Zandi, S., Takayama, N., Dobson, S., Gan, O.I., Wilson, G., Kaufmann, K.B., McLeod, J., Laurenti, E., Dunant, C.F., McPherson, J.D., Stein, L.D., Dror, Y., and Dick, J.E. (2015). Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. Science, aab2116.

Nusse, R. (2019). *The Wnt Homepage*. [Online; accessed Jan-2019]. URL: `https://web.stanford.edu/group/nusselab/cgi-bin/wnt/target_genes`.

Oguro, H., Yuan, J., Ichikawa, H., Ikawa, T., Yamazaki, S., Kawamoto, H., Nakauchi, H., and Iwama, A. (2010). Poised lineage specification in multipotential hematopoietic stem and progenitor cells by the polycomb protein Bmi1. Cell Stem Cell *6*.3, 279–286.

Ohlsson, E., Hasemann, M.S., Willer, A., Lauridsen, F.K.B., Rapin, N., Jendholm, J., and Porse, B.T. (2014). Initiation of MLL-rearranged AML is dependent on C/EBP alpha. Journal of Experimental Medicine *211*.1, 13.

Okada, S., Nakauchi, H., Nagayoshi, K., Nishikawa, S., Y., and T., S.M. (1992). Invivo And Invitro Stem-Cell Function Of C-Kit-Positive And Sca-1-Positive Murine Hematopoietic-Cells. Blood *80*.12, 3050.

Orford, K., Kharchenko, P., Lai, W., Dao, M.C., Worhunsky, D.J., Ferro, A., Janzen, V., Park, P.J., and Scadden, D.T. (2008). Differential H3K4 methylation identifies developmentally poised hematopoietic genes. Developmental Cell *14*.5, 798–809.

Orkin, S.H. and Zon, L.I. (2008). Hematopoiesis: an evolving paradigm for stem cell biology. Cell *132*.4, 631–644.

Ostuni, R., Piccolo, V., Barozzi, I., Polletti, S., Termanini, A., Bonifacio, S., Curina, A., Prosperini, E., Ghisletti, S., and Natoli, G. (2013). Latent enhancers activated by stimulation in differentiated cells. Cell *152*.1, 157–171.

Pai, S.G., Carneiro, B.A., Mota, J.M., Costa, R., Leite, C.A., Barroso-Sousa, R., Kaplan, J.B., Chae, Y.K., and Giles, F.J. (2017). Wnt/beta-catenin pathway: modulating anticancer immune response. Journal of Hematology & Oncology *10*.1, 101–12.

Panagopoulos, I., Fioretos, T., Isaksson, M., Samuelsson, U., om, R.B., ombeck, B.S., Mitelman, F., and Johansson, B. (2001). Fusion of the MORF and CBP genes in acute myeloid leukemia with the t (10; 16)(q22; p13). Human Molecular Genetics *10*.4, 395–404.

Pang, W.W., Price, E.A., Sahoo, D., Beerman, I., Maloney, W.J., Rossi, D.J., Schrier, S.L., and Weissman, I.L. (2011). Human bone marrow hematopoietic stem cells are increased in

frequency and myeloid-biased with age. Proceedings of the National Academy of Sciences *108*.50, 20012–20017.

Park, I. kyung, Qian, D., Kiel, M., Becker, M.W., Pihalja, M., Weissman, I.L., Morrison, S.J., and Clarke, M.F. (2003). Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. Nature *423*.6937, 302–305.

Park, J.-S., Ma, W., O'Brien, L.L., Chung, E., Guo, J.-J., Cheng, J.-G., Valerius, M.T., McMahon, J.A., Wong, W.H., and McMahon, A.P. (2012). Six2 and Wnt regulate self-renewal and commitment of nephron progenitors through shared gene regulatory networks. Developmental Cell *23*.3, 637–651.

Parker, D.S., Ni, Y.Y., Chang, J.L., Li, J., and Cadigan, K.M. (2008). Wingless signaling induces widespread chromatin remodeling of target loci. Molecular and Cellular Biology *28*.5, 1815–1828.

Peixoto, L., Risso, D., Poplawski, S.G., Wimmer, M.E., Speed, T.P., Wood, M.A., and Abel, T. (July 2015). How data analysis affects power, reproducibility and biological insight of RNA-seq studies in complex datasets. Nucleic Acids Research *43*.16, 7664–7674.

Pepin, G., Ferrand, J., Honing, K., Jayasekara, W.S.N., Cain, J.E., Behlke, M.A., Gough, D.J., Williams, B.R.G., Hornung, V., and Gantier, M.P. (2016). Cre-dependent DNA recombination activates a STING-dependent innate immune response. Nucleic Acids Research *44*.11, 5356–5364.

Perugini, M, Iarossi, D.G., Kok, C.H., Cummings, N, Diakiw, S.M., Brown, A.L., Danner, S, Bardy, P, To, L.B., Wei, A.H., Lewis, I.D., and D'andrea, R.J. (2013). GADD45A methylation predicts poor overall survival in acute myeloid leukemia and is associated with IDH1/2 and DNMT3A mutations. Leukemia *27*.7, 1588–1592.

Perugini, M., Kok, C., Brown, A., Wilkinson, C., Salerno, D., Young, S., Diakiw, S., Lewis, I., Gonda, T., and D'andrea, R. (2009). Repression of Gadd45α by activated FLT3 and GM-CSF receptor mutants contributes to growth, survival and blocked differentiation. Leukemia *23*.4, 729–738.

Planutiene, M., Planutis, K., and Holcombe, R.F. (2011). Lymphoid enhancer-binding factor 1, a representative of vertebrate-specific Lef1/Tcf1 sub-family, is a Wnt-beta-catenin pathway target gene in human endothelial cells which regulates matrix metalloproteinase-2 expression and promotes endothelial cell invasion. Vascular Cell *3*.1, 28.

Prlic, M. and Bevan, M.J. (2011). Cutting edge: β-catenin is dispensable for T cell effector differentiation, Memory formation, and recall responses. The Journal of Immunology *187*.4, 1542–1546.

Qiu, W., Chen, L., and Kassem, M. (2011). Activation of non-canonical Wnt/JNK pathway by Wnt3a is associated with differentiation fate determination of human bone marrow stromal (mesenchymal) stem cells. Biochemical and Biophysical Research Communications *413*.1, 98–104.

Quinlan, A.R. and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*.6, 842.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. (Vienna, Austria). URL: `https://www.R-project.org/`.

Rad, R., Rad, L., Wang, W., Cadinanos, J., Vassiliou, G., Rice, S., Campos, L.S., Yusa, K., Banerjee, R., Li, M.A., Rosa, J. de la, Strong, A., Lu, D., Ellis, P., Conte, N., Yang, F.T., Liu, P., and Bradley, A. (2010). PiggyBac transposon mutagenesis: a tool for cancer gene discovery in mice. Science *330*.6007, 1104–1107.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic acids research *42*.W1, W187–W191.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Research *44*.W1, W160–W165.

Rando, O.J. (2012). Combinatorial complexity in chromatin structure and function: revisiting the histone code. Current Opinion in Genetics & Development *22*.2, 148–155.

Rathert, P., Roth, M., Neumann, T., Muerdter, F., Roe, J.-S., Muhar, M., Deswal, S., Cerny-Reiterer, S., Peter, B., Jude, J., Hoffmann, T., Boryń, Ł.M., Axelsson, E., Schweifer, N., Tontsch-Grunt, U., Dow, L.E., Gianni, D., Pearson, M., Valent, P., Stark, A., Kraut, N., Vakoc, C.R., and Zuber, J. (2015). Transcriptional plasticity promotes primary and acquired resistance to BET inhibition. Nature *525*.7570, 543–547.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.-W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., and Jenuwein, T. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*.6796, 593–599.

Reya, T., Duncan, A.W., Ailles, L., Domen, J., Scherer, D.C., Willert, K., Hintz, L., Nusse, R., and Weissman, I.L. (2003). A role for Wnt signalling in self-renewal of haematopoietic stem cells. Nature *423*.6938, 409–414.

Reya, T., O'Riordan, M., Okamura, R., Devaney, E., Willert, K., Nusse, R., and Grosschedl, R. (2000). Wnt signaling regulates B lymphocyte proliferation through a LEF-1 dependent mechanism. Immunity *13*.1, 15–24.

Rieger, M.A., Hoppe, P.S., Smejkal, B.M., Eitelhuber, A.C., and Schroeder, T. (2009). Hematopoietic cytokines can instruct lineage choice. Science *325*.5937, 217–218.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research *43*.7, e47.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*.1, 139–140.

Rooij, d.J., Hollink, I., Arentsen-Peters, S., Galen, J. van, Beverloo, H.B., Baruchel, A., Trka, J, Reinhardt, D., Sonneveld, E, Zimmermann, M., Alonzo, T., Pieters, R, Meshinchi, S, Heuvel-Eibrink, v.d.M., and Zwaan, C. (2013). NUP98/JARID1A is a novel recurrent abnormality in pediatric acute megakaryoblastic leukemia with a distinct HOX gene expression pattern. Leukemia *27*.12, 2280–2288.

Rosati, R., Starza, R.L., Veronese, A., Aventin, A., Schwienbacher, C., Vallespi, T., Negrini, M., Martelli, M.F., and Mecucci, C. (2002). NUP98 is fused to the NSD3 gene in acute myeloid leukemia associated with t (8; 11)(p11. 2; p15). Blood *99*.10, 3857–3860.

Rossi, A., Kontarakis, Z., Gerri, C., Nolte, H., Hölper, S., Krüger, M., and Stainier, D.Y.R. (2015). Genetic compensation induced by deleterious mutations but not gene knockdowns. Nature *524*.7564, 230–233.

Rozman, M., Camós, M., Colomer, D., Villamor, N., Esteve, J., Costa, D., Ana, C., Aymerich, M., Aguilar, J.L., Domingo, A., Solé, F., Gomis, F., Florensa, L., Montserrat, E., and Campo, E. (2004). Type I MOZ/CBP (MYST3/CREBBP) is the most common chimeric transcript in acute myeloid leukemia with t(8; 6)(p11; p13) translocation. Genes, Chromosomes and Cancer *40*.2, 140–145.

Russo, V.E., Martienssen, R.A., and Riggs, A.D. (1996). Epigenetic mechanisms of gene regulation (Cold Spring Harbour, NY: Cold Spring Harbour Laboratory Press).

Saito, Y., Uchida, N., Tanaka, S., Suzuki, N., Tomizawa-Murasawa, M., Sone, A., Najima, Y., Takagi, S., Aoki, Y., Wake, A., Taniguchi, S., Shultz, L.D., and Ishikawa, F. (2010). Induction of cell cycle entry eliminates human leukemia stem cells in a mouse model of AML. Nature Biotechnology *28*.3, 275–280.

Salazar, V.S., Zarkadis, N., Huang, L., Watkins, M., Kading, J., Bonar, S., Norris, J., Mbalaviele, G., and Civitelli, R. (2013). Postnatal ablation of osteoblast Smad4 enhances proliferative responses to canonical Wnt signaling through interactions with β-catenin. Journal of Cell Science *126*.24, 5598–5609.

Saleque, S., Kim, J., Rooke, H.M., and Orkin, S.H. (2007). Epigenetic regulation of hematopoietic differentiation by Gfi-1 and Gfi-1b is mediated by the cofactors CoREST and LSD1. Molecular Cell *27*.4, 562–572.

Sánchez-Castillo, M., Ruau, D., Wilkinson, A.C., Ng, F.S., Hannah, R., Diamanti, E., Lombard, P., Wilson, N.K., and Gottgens, B. (2014). CODEX: a next-generation sequencing experiment database for the haematopoietic and embryonic stem cell communities. Nucleic Acids Research *43*.D1, D1117–D1123.

Sandoval, J. and Esteller, M. (2012). Cancer epigenomics: beyond genomics. Current Opinion in Genetics & Development *22*.1, 50–55.

Scheller, M., Huelsken, J., Rosenbauer, F., Taketo, M.M., Birchmeier, W., Tenen, D.G., and Leutz, A. (2006). Hematopoietic stem cell and multilineage defects generated by constitutive β-catenin activation. Nature Immunology *7*.10, 1037–1047.

Schuijers, J., Junker, J.P., Mokry, M., Hatzis, P., Koo, B.-K., Sasselli, V., Flier, L.G. van der, Cuppen, E., Oudenaarden, A. van, and Clevers, H. (2015). Ascl2 acts as an R-spondin/Wnt-responsive switch to control stemness in intestinal crypts. Cell Stem Cell *16*.2, 158–170.

Schuijers, J., Mokry, M., Hatzis, P., Cuppen, E., and Clevers, H. (2014). Wnt-induced transcriptional activation is exclusively mediated by TCF/LEF. The EMBO Journal *33*.2, 146–156.

Schurch, N.J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G.G., Owen-Hughes, T., et al. (2016). How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? Rna *22*.6, 839–851.

Scott, E.W., Simon, M.C., Anastasi, J., and Singh, H. (1994). Requirement of transcription factor PU. 1 in the development of multiple hematopoietic lineages. Science *265*.5178, 1573–1577.

Shema, E., Bernstein, B.E., and Buenrostro, J.D. (2019). Single-cell and single-molecule epigenomics to uncover genome regulation at unprecedented resolution. Nature genetics *51*.1, 19–25.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V., and Ren, B. (2012). A map of the cis-regulatory sequences in the mouse genome. Nature *488*.7409, 116–120.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*.7, 941–953.

Shooshtarizadeh, P., Helness, A., Vadnais, C., Brouwer, N., Beauchemin, H., Chen, R., Bagci, H., Staal, F.J.T., Coté, J.-F., and Möröy, T. (2019). Gfi1b regulates the level of Wnt/β-catenin signaling in hematopoietic stem cells and megakaryocytes. Nature Communications *10*.1, 1270.

Siapati, E.K., Papadaki, M., Kozaou, Z., Rouka, E., Michali, E., Savvidou, I., Gogos, D., Kyriakou, D., Anagnostopoulos, N.I., and Vassilopoulos, G. (2011). Proliferation and bone marrow engraftment of AML blasts is dependent on β-catenin signalling. British Journal of Haematology *152*.2, 164–174.

Sierra, J., Yoshida, T., Joazeiro, C.A., and Jones, K.A. (2006). The APC tumor suppressor counteracts β-catenin activation and H3K4 methylation at Wnt target genes. Genes & Development *20*.5, 586–600.

Silver, D.P. and Livingston, D.M. (2001). Self-Excising Retroviral Vectors Encoding the Cre Recombinase Overcome Cre-Mediated Cellular Toxicity. Molecular Cell *8*.1, 233–243.

Simon, J.A. and Kingston, R.E. (2013). Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. Molecular Cell *49*.5, 808–824.

Simon, M., Grandage, V.L., Linch, D.C., and Khwaja, A. (2005). Constitutive activation of the Wnt/β-catenin signalling pathway in acute myeloid leukaemia. Oncogene *24*.14, 2410–2420.

Sinner, D., Rankin, S., Lee, M., and Zorn, A.M. (2004). Sox17 and β-catenin cooperate to regulate the transcription of endodermal genes. Development *131*.13, 3069–3080.

Siriboonpiputtana, T. (2014). "The role of beta-catenin in development of origin-specific leukaemia stem cells". PhD thesis. Guy's, King's and St. Thomas's School of Medicine.

Siriboonpiputtana, T., Zeisig, B.B., Zarowiecki, M., Fung, T.K., Mallardo, M., Tsai, C.-T., Lau, P.N.I., Hoang, Q.C., Veiga, P., Barnes, J., Lynn, C., Wilson, A., Lenhard, B., and So, C.W.E. (2017). Transcriptional memory of cells of origin overrides β-catenin requirement of MLL cancer stem cells. The EMBO Journal *36*.21, 3139–3155.

Slany, R.K. (2009). The molecular biology of mixed lineage leukemia. Haematologica *94*.7, 984–993.

Slovak, M.L., Kopecky, K.J., Cassileth, P.A., Harrington, D.H., Theil, K.S., Mohamed, A., Paietta, E., Willman, C.L., Head, D.R., Rowe, J.M., Forman, S., and FR., A. (2000). Karyotypic analysis predicts outcome of preremission and postremission therapy in adult acute myeloid leukemia: a Southwest Oncology Group/Eastern Cooperative Oncology Group Study. Blood *96*.13, 4075–4083.

Smallwood, A. and Ren, B. (2013). Genome organization and long-range regulation of gene expression by enhancers. Current Opinion in Cell Biology *25*.3, 387–394.

Smith, M.L., Cavenagh, J.D., Lister, T.A., and Fitzgibbon, J. (2004). Mutation of CEBPA in familial acute myeloid leukemia. New England Journal of Medicine *351*.23, 2403–2407.

So, C.W., Karsunky, H., Passegue, E., Cozzio, A., IL, and ML, C.W. (2003). MLL-GAS7 transforms multipotent hematopoietic progenitors and induces mixed lineage leukemias in mice. Cancer Cell *3*.2, 171.

Sohn, K.-A., Ho, J.W., Djordjevic, D., Jeong, H.-h., Park, P.J., and Kim, J.H. (2015). hiHMM: Bayesian non-parametric joint inference of chromatin state maps. Bioinformatics *31*.13, 2066–2074.

Somervaille, T.C.P., Matheny, C.J., Spencer, G.J., Iwasaki, M., Rinn, J.L., Witten, D.M., Chang, H.Y., Shurtleff, S.A., Downing, J.R., and Cleary, M.L. (2009). Hierarchical Maintenance of MLL Myeloid Leukemia Stem Cells Employs a Transcriptional Program Shared with Embryonic Rather Than Adult Stem Cells. Cell Stem Cell *4*.2, 140.

Song, H., Spichiger-Haeusermann, C., and Basler, K. (2009). The ISWI-containing NURF complex regulates the output of the canonical Wingless pathway. EMBO Reports *10*.10, 1140–1146.

Southall, S.M., Wong, P.-S., Odho, Z., Roe, S.M., and Wilson, J.R. (2009). Structural Basis for the Requirement of Additional Factors for MLL1 SET Domain Activity and Recognition of Epigenetic Marks. Molecular Cell *33*.2, 181–191.

Spangrude, G., Heimfeld, S., and Weissman, I. (1988). Purification And Characterization Of Mouse Hematopoietic Stem-Cells. Science *241*.4861, 62.

Spencer, G.J., Utting, J.C., Etheridge, S.L., Arnett, T.R., and Genever, P.G. (2006). Wnt signalling in osteoblasts regulates expression of the receptor activator of NFκB ligand and inhibits osteoclastogenesis in vitro. Journal of Cell Science *119*.7, 1283–1296.

Stark, R and Brown, G (2011). DiffBind: differential binding analysis of ChIP-Seq peak data. URL: http://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf.

Stavropoulou, V., Almosailleakh, M., Royo, H., Spetz, J.-F., Juge, S., Brault, L., Kopp, P., Iacovino, M., Kyba, M., Tzankov, A., Stadler, M., Cazzaniga, G., Peters, A., and Schwaller, J. (2018). A novel inducible mouse model of MLL-ENL-driven mixed-lineage acute leukemia. HemaSphere *2*.4, e51.

Stavropoulou, V., Kaspar, S., Brault, L., Sanders, M.A., Juge, S., Morettini, S., Tzankov, A., Iacovino, M., Lau, I.-J., Milne, T.A., Royo, H., Kyba, M., Valk, P.J.M., Peters, A.H.F.M., and Schwaller, J. (2016). MLL-AF9 Expression in Hematopoietic Stem Cells Drives a Highly Invasive AML Expressing EMT-Related Genes Linked to Poor Outcome. Cancer Cell *30*.1, 43–58.

Stewart, S.A., Dykxhoorn, D.M., Palliser, D., Mizuno, H., Yu, E.Y., An, D.S., Sabatini, D.M., Chen, I.S., Hahn, W.C., Sharp, P.A., et al. (2003). Lentivirus-delivered stable gene silencing by RNAi in primary cells. Rna *9*.4, 493–501.

Stojnic, R. and Diez, D. (2018). PWMEnrich: PWM enrichment analysis. R package version *4*.18.0.

Strahl, B.D. and Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*.6765, 41–45.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences *102*.43, 15545–15550.

Tabariès, S. and Siegel, P.M. (2017). The role of claudins in cancer metastasis. Oncogene *36*.9, 1176–1190.

Takacova, S., Slany, R., Bartkova, J., Stranecky, V., Dolezel, P., Luzna, P., Bartek, J., and Divoky, V. (2012). DNA Damage Response and Inflammatory Signaling Limit the MLL-ENL-Induced Leukemogenesis In Vivo. Cancer Cell *21*.4, 517–531.

Taki, T, Ida, K, Bessho, F, Hanada, R, Kikuchi, A, Yamamoto, K, Sako, M, Tsuchida, M, Seto, M, Ueda, R, and Hayashi, Y (1996). Frequency and clinical significance of the MLL gene rearrangements in infant acute leukemia. Leukemia *10*.8, 1303–1307.

Tan, M., Luo, H., Lee, S., Jin, F., Yang, J.S., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N., Lu, Z., Ye, Z., Zhu, Q., Wysocka, J., Ye, Y., Khochbin, S., Ren, B., and Zhao, Y. (2011). Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. Cell *146*.6, 1016–1028.

Taunton, J., Hassig, C.A., and Schreiber, S.L. (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. Science *272*.5260, 408–411.

Taussig, D.C., Vargaftig, J., Miraki-Moud, F., Griessinger, E., Sharrock, K., Luke, T., Lillington, D., Oakervee, H., Cavenagh, J., Agrawal, S.G., Lister, T.A., Gribben, J.G., and Bonnet, D. (2010). Leukemia initiating cells from some acute myeloid leukemia patients with mutated nucleophosmin reside in the CD34-fraction. Blood *115*.10, 1976–1984.

Terwijn, M., Zeijlemaker, W., Kelder, A., Rutten, A.P., Snel, A.N., Scholten, W.J., Pabst, T., Verhoef, G., Löwenberg, B., Zweegman, S., Ossenkoppele, G., and Schuurhuis, G. (2014). Leukemic stem cell frequency: a strong biomarker for clinical outcome in acute myeloid leukemia. PloS One *9*.9, e107587.

Thein, M.S., Ershler, W.B., Jemal, A., Yates, J.W., and Baer, M.R. (2013). Outcome of older patients with acute myeloid leukemia. Cancer *119*.15, 2720–2727.

Thomas, T., Corcoran, L.M., Gugasyan, R., Dixon, M.P., Brodnicki, T., Nutt, S.L., Metcalf, D., and Voss, A.K. (2006). Monocytic leukemia zinc finger protein is essential for the development of long-term reconstituting hematopoietic stem cells. Genes & Development *20*.9, 1175–1186.

Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in Bioinformatics *14*.2, 178–192.

Till, J.E. and McCulloch, E.A. (1980). Hemopoietic stem cell differentiation. Biochimica et Biophysica Acta - Reviews on Cancer *605*.4, 431–459.

Trowbridge, J.J., Snow, J.W., Kim, J., and Orkin, S.H. (2009). DNA methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells. Cell Stem Cell *5*.4, 442–449.

Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., Meyers, R.M., Ali, L., Goodale, A., Lee, Y., Jiang, G., Hsiao, J., Gerath, W.F.J., Howell, S., Merkel, E., Ghandi, M., Garraway, L.A., Root, D.E., Golub, T.R., Boehm, J.S., and Hahn, W.C. (2017). Defining a Cancer Dependency Map. Cell *170*.3, 564–576.

Ugale, A., Norddahl, G.L., Wahlestedt, M., Säwén, P., Jaako, P., Pronk, C.J., Soneji, S., Cammenga, J., and Bryder, D. (2014). Hematopoietic Stem Cells Are Intrinsically Protected against MLL-ENL-Mediated Transformation. Cell Reports *9*.4, 1246–1255.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3 new capabilities and interfaces. Nucleic Acids Research *40*.15, e115.

Uren, A.G., Kool, J., Berns, A., and Lohuizen, M. van (2005). Retroviral insertional mutagenesis: past, present and future. Oncogene *24*.52, 7656–72.

Valenta, T., Hausmann, G., and Basler, K. (2012). The many faces and functions of β-catenin. The EMBO Journal *31*.12, 2714–2736.

Vardiman, J.W., Harris, N.L., and Brunning, R.D. (2002). The World Health Organization (WHO) classification of the myeloid neoplasms. Blood *100*.7, 2292–2302.

Vardiman, J.W., Thiele, J., Arber, D.A., Brunning, R.D., Borowitz, M.J., Porwit, A., Harris, N.L., Le Beau, M.M., Hellström-Lindberg, E., Tefferi, A., and Bloomfield, C.D. (2009). The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. Blood *114*.5, 937–951.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., and Kinzler, K.W. (2013). Cancer genome landscapes. Science *339*.6127, 1546–1558.

Voli, F., Yi, H., Martinez, B., D'Andrea, R.J., and Wang, J.Y. (2015). Gadd45a Loss Enhances MLL Leukemogenesis By Promoting Beta-Catenin Mitochondrial Translocation and Inhibiting Apoptosis. Blood *126*.23, 1251.

Vradii, D., Wagner, S., Doan, D.N., Nickerson, J.A., Montecino, M., Lian, J.B., Stein, J.L., Wijnen, A.J.V., Imbalzano, A.N., and Stein, G.S. (2006). Brg1, the ATPase subunit of the SWI/SNF chromatin remodeling complex, is required for myeloid differentiation to granulocytes. Journal of Cellular Physiology *206*.1, 112–118.

Vu, L.P., Perna, F., Wang, L., Voza, F., Figueroa, M.E., Tempst, P., Erdjument-Bromage, H., Gao, R., Chen, S., Paietta, E., Deblasio, T., Melnick, A., Liu, Y., Zhao, X., and Nimer, S.D. (2013). PRMT4 blocks myeloid differentiation by assembling a methyl-RUNX1-dependent repressor complex. Cell Reports *5*.6, 1625–1638.

Waddington, C.H. (1957). The Strategy of the Genes; a Discussion of Some Aspects of Theoretical Biology (London, UK: Allen & Unwin).

Walter, M.J., Ding, L., Shen, D., Shao, J., Grillot, M., McLellan, M., Fulton, R., Schmidt, H., Kalicki-Veizer, J., O'Laughlin, M., Kandoth, C., Baty, J., Westervelt, P., DiPersio, J.F.,

Mardis, E.R., Wilson, R.K., Ley, T.J., and Graubert, T.A. (2011). Recurrent DNMT3A mutations in patients with myelodysplastic syndromes. Leukemia *25*.7, 1153–1158.

Wang, Q f, Li, Y j, Dong, J f, Li, B., Kaberlein, J.J., Zhang, L., Arimura, F.E., Luo, R.T., Ni, J., He, F., Wu, J., Mattison, R., Zhou, J., Wang, C z, Prabhakar, S., Nobrega, M.A., and Thirman, M.J. (2014). Regulation of MEIS1 by distal enhancer elements in acute leukemia. Leukemia *28*.1, 138–146.

Wang, Y., Krivtsov, A.V., Sinha, A.U., North, T.E., Goessling, W., Feng, Z., Zon, L.I., and Armstrong, S.A. (2010a). The Wnt/β-catenin pathway is required for the development of leukemia stem cells in AML. Science *327*.5973, 1650–1653.

Wang, Z., Iwasaki, M., Ficara, F., Lin, C., Matheny, C., Wong, S.H., Smith, K.S., and Cleary, M.L. (2010b). GSK-3 promotes conditional association of CREB and its coactivators with MEIS1 to facilitate HOX-mediated transcription and oncogenesis. Cancer cell *17*.6, 597–608.

Wang, Z., Smith, K.S., Murphy, M., Piloto, O., Somervaille, T.C., and Cleary, M.L. (2008). Glycogen synthase kinase 3 in MLL leukaemia maintenance and targeted therapy. Nature *455*.7217, 1205–1209.

Watanabe, K., Biesinger, J., Salmans, M.L., Roberts, B.S., Arthur, W.T., Cleary, M., Andersen, B., Xie, X., and Dai, X. (2014). Integrative ChIP-seq/Microarray Analysis Identifies a CTNNB1 Target Signature Enriched in Intestinal Stem Cells and Colon Cancer. PLoS One *9*.3, e92317.

Weiske, J., Albring, K.F., and Huber, O. (2007). The tumor suppressor Fhit acts as a repressor of β-catenin transcriptional activity. National Academy of Sciences of the United States of America *104*.51, 20344–20349.

Weissman, I.L., Anderson, D.J., and Gage, F. (2001). Stem and progenitor cells: origins, phenotypes, lineage commitments, and transdifferentiations. Annual Review of Cell and Developmental Biology *17*.1, 387–403.

Wells, J.M., Esni, F., Boivin, G.P., Aronow, B.J., Stuart, W., Combs, C., Sklenka, A., Leach, S.D., and Lowy, A.M. (2007). Wnt/β-catenin signaling is required for development of the exocrine pancreas. BMC Developmental Biology *7*.1, 4.

Wend, P., Fang, L., Zhu, Q., Schipper, J.H., Loddenkemper, C., Kosel, F., Brinkmann, V., Eckert, K., Hindersin, S., Holland, J.D., Lehr, S., Kahn, M., Ziebold, U., and Birchmeier, W. (2013). Wnt/β-catenin signalling induces MLL to create epigenetic changes in salivary gland tumours. The EMBO Journal *32*.14, 1977–1989.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (New York: Springer-Verlag). ISBN: 978-3-319-24277-4. URL: `https://ggplot2.tidyverse.org`.

Wilting, R.H., Yanover, E., Heideman, M.R., Jacobs, H., Horner, J., Torre, J. van der, DePinho, R.A., and Dannenberg, J.-H. (2010). Overlapping functions of Hdac1 and Hdac2 in cell cycle regulation and haematopoiesis. The EMBO Journal *29*.15, 2586–2597.

Wolf, D., Rodova, M., Miska, E.A., Calvet, J.P., and Kouzarides, T. (2002). Acetylation of β-catenin by CREB-binding protein (CBP). Journal of Biological Chemistry *277*.28, 25562–25567.

Xia, Z.-B., Anderson, M., Diaz, M.O., and Zeleznik-Le, N.J. (2003). MLL repression domain interacts with histone deacetylases, the polycomb group proteins HPC2 and BMI-1, and the corepressor C-terminal-binding protein. Proceedings of the National Academy of Sciences *100*.14, 8342–8347.

Xie, W., Barr, C.L., Kim, A., Yue, F., Lee, A.Y., Eubanks, J., Dempster, E.L., and Ren, B. (2012). Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. Cell *148*.4, 816–831.

Yamamoto, R., Morita, Y., Ooehara, J., Hamanaka, S., Onodera, M., Rudolph, K.L., Ema, H., and Nakauchi, H. (2013). Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. Cell *154*.5, 1112–1126.

Yan, X.-J., Xu, J., Gu, Z.-H., Pan, C.-M., Lu, G., Shen, Y., Shi, J.-Y., Zhu, Y.-M., Tang, L., Zhang, X.-W., Liang, W.-X., Mi, J.-Q., Song, H.-D., Li, K.-Q., Chen, Z., and Chen, S.-J. (2011). Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. Nature Genetics *43*.4, 309–315.

Ye, M., Zhang, H., Yang, H., Koche, R., Staber, P.B., Cusan, M., Levantini, E., Welner, R.S., Bach, C.S., Zhang, J., Krivtsov, A., Armstrong, S.A., and Tenen, D.G. (2015). Hematopoietic Differentiation Is Required for Initiation of Acute Myeloid Leukemia. Cell Stem Cell *17*.5, 611–623.

Yen, A. and Kellis, M. (2015). Systematic chromatin state comparison of epigenomes associated with diverse properties including sex and tissue type. Nature communications *6*.1, 1–13.

Yeung, J., Esposito, M.T., Gandillet, A., Zeisig, B.B., Griessinger, E., Bonnet, D., and So, C.W.E. (2010). β-Catenin Mediates the Establishment and Drug Resistance of MLL Leukemic Stem Cells. Cancer Cell *18*.6, 606–618.

Yeung, J. and So, C.W.E. (2009). Identification and characterization of hematopoietic stem and progenitor cell populations in mouse bone marrow by flow cytometry. Methods in Molecular Biology *538*.2, 15.

Yi, H., Wang, J., Kavallaris, M., and Wang, J.Y. (2013). Lgr4-mediated potentiation of Wnt/β-catenin signaling promotes MLL Leukemogenesis via an Rspo3/Wnt3a-Gnaq pathway in leukemic stem cells. Blood *122*, 887.

Yokoyama, A., Wang, Z., Wysocka, J., Sanyal, M., Aufiero, D.J., Kitabayashi, I., Herr, W., and Cleary, M.L. (2004). Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression. Molecular and Cellular Biology *24*.13, 5639–5649.

Yokoyama, A. and Cleary, M.L. (2008). Menin Critically Links MLL Proteins with LEDGF on Cancer-Associated Target Genes. Cancer Cell *14*.1, 36–46.

Yoshida, T., Hazan, I., Zhang, J., Ng, S.Y., Naito, T., Snippert, H.J., Heller, E.J., Qi, X., Lawton, L.N., Williams, C.J., and Georgopoulos, K. (2008). The role of the chromatin remodeler Mi-2β in hematopoietic stem cell self-renewal and multilineage differentiation. Genes & Development *22*.9, 1174–1189.

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. Omics: A Journal of Integrative Biology *16*.5, 284–287.

Yu, G., Wang, L.-G., and He, Q.-Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics *31*.14, 2382–2383.

Zeisig, B.B., Milne, T., Garcia-Cuellar, M.P., Schreiner, S., Martin, M.E., Fuchs, U., Borkhardt, A., Chanda, S.K., J., R., S.W., Hess, J.L., and Slany, R.K. (2004). Hoxa9 and Meis1 are key targets for MLL-ENL-mediated cellular. Molecular and Cellular Biology *24*.2, 617–628.

Zeisig, B.B., Kulasekararaj, A.G., Mufti, G.J., and So, C.W.E. (2012). SnapShot: Acute myeloid leukemia. Cancer Cell *22*.5, 698.

Zeisig, B.B. and So, C.W.E. (2009). Retroviral/Lentiviral Transduction and Transformation Assay. In : Eric So C.W. (eds) Leukemia. Methods in Molecular Biology™ (Methods and Protocols), vol. 538. (Humana Press), pp. 207–229.

Zeleznik-Le, N.J., Harden, A.M., and Rowley, J.D. (1994). 11q23 translocations split the "AT-hook" cruciform DNA-binding region and the transcriptional repression domain from the activation domain of the mixed-lineage leukemia (MLL) gene. Proceedings of the National Academy of Sciences of the United States of America *91*.22, 10610–10614.

Zeng, X., Sanalkumar, R., Bresnick, E.H., Li, H., Chang, Q., and Keleş, S. (2013). jMOSAiCS: joint analysis of multiple ChIP-seq datasets. Genome biology *14*.4, R38.

Zentner, G.E. and Henikoff, S. (2013). Regulation of nucleosome dynamics by histone modifications. Nature Structural & Molecular Biology *20*.3, 259–266.

Zhang, X., Peterson, K.A., Liu, X.S., McMahon, A.P., and Ohba, S. (2013). Gene regulatory networks mediating canonical Wnt signal-directed control of pluripotency and differentiation in embryo stem cells. Stem Cells *31*.12, 2667–2679.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biology *9*.9, R137.

Zhao, C., Blum, J., Chen, A., Kwon, H.Y., Jung, S.H., Cook, J.M., Lagoo, A., and Reyal, T. (2007). Loss of beta-catenin impairs the renewal of normal and CML stem cells in vivo. Cancer Cell *12*.6, 528–541.

Zhou, F., Zhang, L., Gong, K., Lu, G., Sheng, B., Wang, A., Zhao, N., Zhang, X., and Gong, Y. (2008). LEF-1 activates the transcription of E2F1. Biochemical and biophysical research communications *365*.1, 149–153.

Zhou, V.W., Goren, A., and Bernstein, B.E. (2011). Charting histone modifications and the functional organization of mammalian genomes. Nature Reviews Genetics *12*.1, 7–18.

Zhou, X., Lindsay, H., and Robinson, M.D. (2014). Robustly detecting differential expression in RNA sequencing data using observation weights. Nucleic Acids Research *42*.11, e91.

Zilberman, D., Coleman-Derr, D., Ballinger, T., and Henikoff, S. (2008). Histone H2A. Z and DNA methylation are mutually antagonistic chromatin marks. Nature *456*.7218, 125–129.

Zuber, J., Radtke, I., Pardee, T.S., Zhao, Z., Rappaport, A.R., Luo, W., McCurrach, M.E., Yang, M.-M., Dolan, M.E., Kogan, S.C., Downing, J., and Lowe, S. (2009). Mouse models of human AML accurately predict chemotherapy response. Genes & development *23*.7, 877–889.