Understanding the interplay between RUNX1 and Polycomb proteins in leukaemias with fusion proteins



Dimitra Dimou

St Hugh's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary term 2022

Abstract

RUNX1 is a critical regulator of the development of the haematopoietic system and a frequent target of genetic alteration in leukaemia. In addition, wildtype RUNX1 is required for maintenance of disease in leukaemias with fusion proteins, including MLL-AF4 and RUNX1-RUNX1T1 leukaemias. However, the exact mechanisms of how RUNX1 shapes the chromatin landscape and impacts gene regulation in normal and leukaemic cells remain elusive. In this thesis, it was investigated how RUNX1 binding affects the recruitment of gene-regulatory proteins and impacts the transcriptional regulation of its target genes, and how its binding is affected by the chromatin landscape. Using a refined intracellular tethering system, it was shown that RUNX1 alone is sufficient to recruit co-activators and the transcriptional machinery, and also co-repressors, including Polycomb group (PcG) proteins. In turn, it was shown that the chromatin environment affects the ability of RUNX1 to bind to its binding sites, as gains in the repressive histone marks H3K27me3 and H3K9me3 reduce RUNX1 binding. In addition, the existing chromatin landscape impacts RUNX1-mediated recruitment of co-factors. More specifically, at active sites RUNX1 promotes recruitment of co-activators, while at PcG-repressed genes RUNX1 binding stabilises PcG recruitment. By further investigating RUNX1-mediated recruitment of PcG proteins, it was shown that a key function of RUNX1 is the stabilisation of Polycomb Repressive Complex 1 (PRC1) in MLL-AF4 ALL and RUNX1-RUNX1T1 AML. *PcG* mutations are found in a subset of RUNX1-RUNX1T1 and *RUNX1*-mutant leukaemia patients and are often associated with a worse clinical outcome, but the mechanistic details of the functional RUNX1-PRC1 cooperation have not been investigated. In this thesis, it is shown that the RUNX1-RUNX1T1 fusion protein, and other truncated RUNX1 mutant proteins lacking the RUNX1 C-terminal region, cannot associate with PcG proteins. In these leukaemias, a key role of wildtype RUNX1 is to stabilise PcG binding at PcG-repressed genes, including a subset of genes upregulated in *PcG*-mutant RUNX1-RUNX1T1 patients. Overall, these results demonstrate that an important function of RUNX1 is to maintain PcG repression and that this activity is disrupted

in *RUNX1*-mutant leukaemias, particularly those with *PcG* mutations, thus providing a mechanistic explanation for the positive selection of leukaemic clones with *RUNX1*/*PcG* mutations.

Dedication

To the memory of my father

To my family

Acknowledgements

I would like to thank my supervisors, Tom Milne and Irene Roberts, for their support and encouragement throughout the last four years of my DPhil. Tom and Irene always managed to find time to respond to my questions, trusted my insight and encouraged me to pursue my own scientific interests.

I would also like to thank current and previous members of the Milne lab for teaching me molecular biology techniques, for stimulating discussions and for their advice on scientific techniques. My sincere gratitude goes to Medical Research Council for providing funding for my DPhil studies, and to the researchers in the Weatherall Institute of Molecular Medicine (WIMM), who offered expertise and advice. Special thanks goes to the Greek DPhil students and postdocs in the WIMM, with whom we shared the same background, and whose kind words and willingness to help always made my day.

A big thank you goes to my friends who I met more than ten years ago, during the first semester of medical school, and have encouraged and supported me with every possible way since then. More importantly, my gratitude goes to my family, my mum and Anastasia, for their unconditional love and support.

Declaration

I declare that the work in this thesis is my own research, except when otherwise stated. This thesis has not been submitted, either partially or in full, for another qualification at this University, or for a qualification at any other institution.

Table of contents

Abstract 1

Dedication 3

Acknowledgements 4

Declaration 5

Table of contents 6

List of figures 11

List of tables 15

Abbreviations 16

Chapter 1 – Introduction 24

1.1 Structure of chromatin 24

1.2 Histone modifications and regulation of gene expression 25

1.3 Transcription factors and regulation of gene expression 27

1.4 Function of enhancers in gene regulation and the role of enhancer-promoter interactions 31

1.5 Introduction to Polycomb Group (PcG) and Trithorax Group (TrxG) proteins 37

1.6 Polycomb Group proteins 38

1.6.1 PRC1 complex 38

1.6.2 PRC2 complex 43

1.6.3 Hierarchy in the recruitment of Polycomb complexes 45

1.6.4 H3K27 methylation 47

1.6.5 Role of H3K27me3 49

1.6.6 Malignancies with alterations in global H3K27me3 levels 50

1.6.7 Recruitment of PcG proteins 53

1.6.8 Polycomb repression 59

1.6.9 PRC1 and 3D nuclear structure 61

1.7 MLL proteins 65
1.8 MLL-rearranged leukaemias 68
1.9 RUNX1 70
1.9.1 RUNX1 and the development of the haematopoietic system 70
1.9.2 The role of RUNX1 in leukaemogenesis 73
Aims of the thesis 81
Chapter 2 –Methods 82
2.1 Tissue culture methods 82
2.1.1 Cell lines and growth conditions 82
2.1.2 Mycoplasma testing 83
2.1.3 Storage and recovery of cell lines 83
2.2 RNAi and cell transfection methods 83
2.2.1 siRNA knockdown assays 83
2.2.2 Lipofectamine transfections and generation of stable mESC lines 84
2.3 Molecular and cloning methods 85
2.3.1 Polymerase Chain Reaction (PCR) 85
2.3.2 Restriction enzyme digestion 85
2.3.3 Ligation Independent Cloning (LIC) 86
2.3.4 Bacterial transformations 87
2.3.5 Plasmid amplification and sequencing 88
2.4 Gene expression methods 88
2.4.1 Total RNA extraction 88
2.4.2 Generation of cDNA 88
2.4.3 Real Time Quantitative Polymerase Chain Reaction (RT-qPCR) 89
2.4.4 Nascent RNA-sequencing 90
2.5 Protein methods 90
2.5.1 Whole cell extracts 90

2.5.2 Histone acid extraction 91
2.5.3 Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE) 92
2.6 Cell cycle analysis 94
2.7 ChIP-qPCR and ChIP sequencing 95
2.7.1 Fixation 95
2.7.2 ChIP protocol 95
2.7.3 Reference-normalised ChIP sequencing 97
2.7.4 Library preparation for ChIP sequencing 97
2.8 ATAC sequencing 97
2.8.1 Tagmentation protocol 97
2.8.2 ATAC-seq library preparation 98
2.9 Sequencing 99
2.10 Bioinformatics analysis 99
2.10.1 ChIP-seq/ATAC-seq 99
2.10.2 Nascent RNA-seq 100
Chapter 3 - RUNX1 is sufficient to recruit co-activators and Polycomb repressors 102
3.1 Introduction 102
3.1.1 RUNX1 structure and interaction partners 102
3.1.2 The Tet Repressor (TetR) system 106
3.2 Aims 108
3.3 Results 109
3.3.1 Runx1 binding is sufficient to recruit Trithorax and Polycomb proteins and generate a Polycomb domain. 109
3.3.2 Runx1 is sufficient to recruit co-activators and the transcriptional machinery. 120
3.3.3 Reduction in H2AK119ub levels converts the TetO locus of Runx1-TetR mESCs into a regulatory element with promoter-like features. 125

3.3.4 Runx1-mediated recruitment of PcG proteins results in transcriptional repression at endogenous loci. 129

3.3.5 Runx1-mediated recruitment of co-factors is context-dependent 134

3.4 Discussion 137

Chapter 4 - Understanding RUNX1-mediated PRC1 recruitment in leukaemia 141

4.1 Introduction 141

4.1.1 The role of RUNX1 in leukaemias with fusion proteins 141

4.1.2 Increased frequency of *RUNX1* and *PcG* mutations in leukaemia 142

4.2 Aims 146

4.3 Results 146

4.3.1 RUNX1 binding stabilises PRC1 binding at active and PcG-repressed genes in MLL-AF4 leukaemia, and contributes to PcG-mediated gene repression. 146

4.3.2 RUNX1 mediates recruitment of OGT at active PRC1-bound sites. 156

4.3.3 The C-terminal part of Runx1, lost in RUNX1 truncating mutations and leukaemia-associated RUNX1 translocations, is sufficient for Polycomb recruitment. 158

4.3.4 RUNX1 binding stabilises PRC1 binding at PcG-repressed sites in RUNX1-RUNX1T1 leukaemia. 164

4.3.5 Differential gene expression analysis reveals genes upregulated in *PcG*-mutant RUNX1-RUNX1T1 patients, including RUNX1-bound PcG-repressed genes. 167

4.3.6 Effects of CCND1 upregulation on the cell cycle in RUNX1-RUNX1T1 leukaemia cells 173

4.3.7 Differential gene expression analysis reveals genes upregulated in *PcG*-mutant compared to *PcG*-wildtype RUNX1-mutant patients. 177

4.4 Discussion 180

4.4.1 RUNX1 binding stabilises PRC1 binding in MLL-AF4 cells. 180

4.4.2 RUNX1 binding stabilises PRC1 binding in RUNX1-RUNX1T1 AML 181

4.4.3 Summary and conclusion 185

Chapter 5 - Effects of PRC2 inhibition on the chromatin landscape and RUNX1 binding in MLL-AF4 leukaemia 186

5.1 Introduction 186

5.2 Aims 187

5.3 Results 187

5.3.1 Reduction in H3K27me3 in MLL-AF4 cells results in increased chromatin accessibility and increased RUNX1 binding. 187

5.3.2 Reduction in H3K27me3 levels in MLL-AF4 cells following EZH2/1 inhibition results in reduced TF binding and reduced chromatin accessibility. 192

5.3.3 EZH2/1 inhibition in MLL-AF4 cells results in accumulation of H3K9me3. 201

5.3.4 *EZH1* knockdown in MLL-AF4 cells results in accumulation of H3K9me3. 205

5.4 Discussion 208

Chapter 6 - General discussion and future work 213

6.1 RUNX1 is sufficient to recruit both co-activators and repressors. 213

6.2 RUNX1 contributes to the stabilisation of PRC1 binding in leukaemic contexts. 215

6.3 RUNX1-mediated recruitment of PcG proteins provides a mechanistic explanation for the co-occurrence of *RUNX1* and *PcG* mutations. 217

6.4 Effects of PRC2 inhibition in the chromatin landscape in MLL-AF4 leukaemia 219

6.5 Summary 220

7. Appendix 221

References 225

List of figures

Figure 1.1: Schematic of enhancer-promoter interactions. 36

Figure 1.2: Schematic of PRC1 complexes. 40

Figure 1.3: Schematic of PRC2 complexes. 44

Figure 1.4: Interplay between PRC1 and PRC2 complexes. 46

Figure 1.5: Schematic of long-range Polycomb interactions mediated by canonical PRC1 complexes. 62

Figure 1.6: Schematic of the MLL proteins. 65

Figure 1.7: Schematic of the MLL fusion protein complexes bound to a target gene. 69

Figure 1.8: Schematic of wildtype RUNX1, RUNX1 point mutations and the RUNX1-RUNX1T1 fusion protein. 74

Figure 3.1: Schematic of Runx1 protein structure and interaction partners. 104

Figure 3.2: The TetR system. 107

Figure 3.3: Validation of the Runx1-TetR system. 110

Figure 3.4: Runx1 and interacting TFs in haematopoietic stem and progenitor cells are expressed at low levels in mESCs. 111

Figure 3.5: ChIP-seq correlation analysis in MLL-AF4 cells reveals that RUNX1 binding correlates with the binding of activators and repressors. 112

Figure 3.6: Runx1 is sufficient to recruit Trithorax-group proteins Brg1 and Mll. 114

Figure 3.7: Runx1 is sufficient to recruit PRC1 and generate a Polycomb domain at the TetO locus. 116

Figure 3.8: Runx1 is sufficient to recruit the Bap1 deubiquitinase complex. 120

Figure 3.9: Runx1 is sufficient for the recruitment of p300 and Hdacs, but there is no enrichment in histone acetylation at the TetO locus. 121

Figure 3.10: Inhibition of Hdac activity in Runx1-TetR cells does not result in increased acetylation levels at the TetO locus. 122

Figure 3.11: Reduction in H2AK119ub results in increase in active modifications at the TetO locus of Runx1-TetR cells. 124

Figure 3.12: Runx1 is sufficient to recruit Brd4. 125

Figure 3.13: Runx1 is sufficient for the recruitment of the transcriptional machinery, but no transcription is observed at the TetO locus. 126

Figure 3.14: Reduction in H2AK119ub results in change of the conformational status of RNAPII and transcriptional initiation from the TetO locus of Runx1-TetR cells. 128

Figure 3.15: Runx1-TetR binding to thousands of loci genome-wide in mESCs. 129

Figure 3.16: Runx1 binding results in increased PcG binding at PcG sites. 130

Figure 3.17: Runx1 binding at endogenous Polycomb sites results in increased binding of PcG proteins and transcriptional repression. 132

Figure 3.18: Runx1-TetR binding results in transcriptional repression of lowly expressed Runx1-bound PcG-repressed genes. 133

Figure 3.19: Runx1 binding at active loci results in increased binding of co-activators. 135

Figure 3.20: Runx1 binding at highly expressed active loci results in transcriptional upregulation. 136

Figure 4.1: A subset of RUNX1-RUNX1T1 patients have additional mutations in PcG genes. 143

Figure 4.2: *RUNX1* and *PcG* mutations frequently co-occur in AML/MDS patients. 144

Figure 4.3: Significant chromatin co-occupancy between RUNX1 and RING1B is observed in MLL-AF4 cells 148

Figure 4.4: RUNX1 stabilises RING1B binding at promoters and distal regulatory elements in MLL-AF4 cells 151

Figure 4.5: Canonical and variant PRC1 complexes are bound at RUNX1-bound active promoters and enhancers in MLL-AF4 cells. 153

Figure 4.6: A subset of PcG-repressed genes become upregulated after *RUNX1* siRNA KD in MLL-AF4 cells 155

Figure 4.7: RUNX1 stabilises OGT binding in MLL-AF4 cells. 157

Figure 4.8: Schematic of interaction between RUNX1 mutant proteins and PRC1. 159

Figure 4.9: The N-terminal part of RUNX1 is not sufficient to recruit PRC1. 161

Figure 4.10: The C-terminal part of RUNX1, lost in the RUNX1-RUNX1T1 fusion protein and in truncated RUNX1 mutant proteins, is sufficient to recruit PRC1 and establish a Polycomb domain. 163

Figure 4.11: RUNX1 recruits RING1B at PcG-repressed loci in RUNX1-RUNX1T1 AML cells..... 166

Figure 4.12: Schematic of RUNX1-mediated recruitment of PRC1 at PcG-repressed loci in RUNX1-RUNX1T1 leukaemia..... 167

Figure 4.13: Identification of genes differentially expressed in *PcG*-mutant RUNX1-RUNX1T1 AML patient samples compared to *PcG*-wildtype ones. 169

Figure 4.14: RUNX1 contributes to the stabilisation of RING1B binding at a subset of PcG-repressed genes upregulated in *PcG*-mutant RUNX1-RUNX1T1 AML patients. 171

Figure 4.15: *RUNX1* siRNA KD in SKNO-1 cells results in downregulation of replication-associated genes, but does not lead to upregulation of PcG-repressed genes. 173

Figure 4.16: *CCND1* upregulation results in changes in cell cycling. 175

Figure 4.17: *EPCAM* is upregulated in a subset of *PcG*-mutant RUNX1-RUNX1T1 patients. 176

Figure 4.18: Identification of genes differentially expressed in *PcG*-mutant *RUNX1*-mutant AML patient samples compared to *PcG*-wildtype *RUNX1*-mutant samples. 179

Figure 5.1: GSK126 treatment in SEM cells results in increased RUNX1 binding. 189

Figure 5.2: GSK126 treatment in MLL-AF4 cells results in increased chromatin accessibility. 190

Figure 5.3: GSK126 treatment results in increased accessibility and the emergence of novel distal ATAC-seq peaks. 191

Figure 5.4: UNC1999 treatment in MLL-AF4 cells results in global decrease in H3K27me3..... 193

Figure 5.5: UNC1999 treatment in MLL-AF4 cells results in genome-wide reduction in RUNX1 binding..... 195

Figure 5.6: EZH2/1 inhibition after UNC1999 treatment in SEM cells results in reduced binding of GR and MLL. 196

Figure 5.7: The dual EZH2/1 inhibitor UNC1999 results in reduced chromatin accessibility in MLL-AF4 cells..... 197

Figure 5.8: UNC1999 treatment results in increased H3K27ac levels across the genome and at PcG-repressed genes. 198

Figure 5.9: UNC1999 treatment in MLL-AF4 cells results in upregulation of PcG-repressed genes. 200

Figure 5.10: GO term analysis after UNC1999 treatment in MLL-AF4 cells demonstrates downregulation of cell cycle genes. 201

Figure 5.11: The dual EZH2/1 inhibitor DS-3201b results in reduced RUNX1 binding in MLL-AF4 cells. 202

Figure 5.12: Dual EZH2/1 inhibitors, but not EZH2-specific inhibitors, result in increased H3K9me3 levels in MLL-AF4 cells. 204

Figure 5.13: UNC1999 treatment does not lead to transcriptional changes in genes encoding H3K9 methyltransferases and demethylases. 205

Figure 5.14: *EZH1* siRNA knockdown in MLL-AF4 cells results in increase in H3K9me3. 206

Figure 5.15: The PRC2-EZH1 complex is bound at the promoters of active genes, including cell cycle genes *CDC25A* and *SKP2* 207

Figure 5.16: Schematic of the working hypothesis linking the inhibition of EZH1 activity to cell cycle arrest and accumulation of H3K9me3. 211

Figure 6.1: Schematic of the Runx1 interaction partners validated using the TetR system. 214

Figure 6.2: RUNX1-mediated recruitment of co-factors is context dependent. 215

Figure 6.3: Schematic of chromatin interactions between RUNX1-bound loci, mediated through the recruitment of PcG proteins. 216

List of tables

Table 2.1: Cell line information 82

Table 2.2: PCR reaction program 85

Table 2.3: Restriction digest mix for LIC 86

Table 2.4: LIC reaction 86

Table 2.5: RT-qPCR amplification program 89

Table 2.6: Antibody information 93

Table 2.7: ATAC-seq library preparation reaction 98

Table 2.8: PCR amplification for ATAC-seq library preparation 98

Table 7.1: Accession numbers of ChIP-seq/BioCap datasets in SEM cells 221

Table 7.2: Mutations and GEO accession numbers for RUNX1-RUNX1T1 AML patient RNA-seq datasets analysed in this thesis 222

Table 7.3: Mutations and GEO accession numbers for *RUNX1*-mutant AML patient RNA-seq datasets analysed in this thesis 223

Abbreviations

3C	Chromosome conformation capture
AD	Activation domain
AEBP2	Adipocyte enhancer-binding protein 2
AGM	Aorta-gonad-mesonephros
ALL	Acute lymphoblastic leukaemia
AML	Acute myeloid leukaemia
ATAC	Assay for Transposase-Accessible Chromatin
AUTS2	Autism susceptibility protein 2
BAH	Bromo adjacent homology
BCOR	BCL6 Co-repressor
BCORL1	BCL6 Co-repressor-Like 1
BET	Bromodomain and extraterminal
BMI1	B cell-specific Moloney murine leukemia virus integration site 1
bp	Base pair
BRD	Bromodomain
CBFb	Core-binding factor subunit Beta
CBX	Chromobox homolog
CDB	ChIP Dilution Buffer
CDK	Cyclin-dependent kinase

cDNA	Complementary DNA
CGI	CpG-island
ChIP	Chromatin Immunoprecipitation
CK2	Casein kinase 2
CMML	Chronic myelomonocytic leukaemia
COMPASS	Complex of proteins associated with SET
CPM	Counts per million
cPRC1	Canonical Polycomb Repressive Complex 1
CTD	Carboxy-terminal domain
DBD	DNA binding domain
dCas9	Cas9 Endonuclease Dead, aka dead Cas9
DLBCL	Diffuse large B-cell lymphoma
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNase I	Deoxyribonuclease I
DNMT	DNA methyltransferase
dSet1	*Drosophila* Set1
DSG	Disuccinimidyl glutarate
DTT	1,4-Dithiothreitol

EB	Embryonic body
ECL	Enhanced chemiluminescence
EED	Embryonic ectoderm development
EHT	Endothelial-to-hematopoietic transition
EPOP	Elongin BC and PRC2-associated protein
ESC	Embryonic stem cell
ETP ALL	Early T-cell precursor acute lymphoblastic leukaemia
EZH1/2	Enhancer of Zeste 1 or 2
FA	Formaldehyde
FBRS	Fibrosin
FBRSL1	Fibrosin-like 1
FBS	Foetal Bovine Serum
FDR	False discovery rate
FL	Follicular lymphoma
FPD/AML	Familial platelet disorder with predisposition to acute myeloid leukemia
GAF	GAGA factor
GR	Glucocorticoid Receptor
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HE	Haemogenic endothelium

HOTAIR	HOX antisense intergenic RNA
Hox	Homeotic
hRNPK	Heterogeneous nuclear ribonucleoprotein K
HSC	Haematopoietic stem cell
IDR	Intrinsically disordered region
IMDM	Iscove's Modified Dulbecco's Media
JARID2	Jumonji and AT-rich interaction domain-containing protein 2
JMJD3	Jumonji domain-containing protein D3
kb	Kilobase
KD	Knockdown
KDM2B	Lysine (K)-Specific Demethylase 2B
KDM6A	Lysine (K)-Specific Demethylase 6A
L3MBTL2	Lethal (3) malignant brain tumor-like 2
LB	Lysogeny broth
LCDR	Low-complexity disordered region
LDB1	Lim-domain-binding 1
LIC	Ligation Independent Cloning
lncRNA	Long non-coding ribonucleic acid
MAX	MYC-associated factor X
MDS	Myelodysplastic syndrome

MEFs	Mouse embryonic fibroblasts
MEL18	Melanoma nuclear protein 18
mESC	Mouse embryonic stem cell
MGA	MAX gene-associated protein
MLL	Mixed lineage leukaemia
MLL-FPs	Mixed lineage leukaemia fusion proteins
MLL-r	Mixed lineage leukaemia gene rearranged/rearrangement
MPN	Myeloproliferative neoplasms
ncPRC1	Non-canonical Polycomb Repressive Complex 1
NGS	Next generation sequencing
PAF1	RNA polymerase II-associated factor 1
PAL1/2	PRC2-associated LCOR isoform 1 or 2
PBS	Phosphate Buffered Saline
Pc	Polycomb
PCA	Principal component analysis
PcG	Polycomb group
PCGF1-6	Polycomb group ring finger proteins 1-6
PCL1-3	Polycomb-like subunit 1-3
PCR	Polymerase Chain Reaction
Ph	Polyhomeotic

PHC	Polyhomeotic-like
PHD	Plant homeodomain
Pho	Pleiohomeotic
Phol	Pleiohomeotic-like
PRC	Polycomb Repressive Complex
PRE	Polycomb Response Element
Psc	Posterior Sex Combs
PTD	Partial tandem duplication
P-TEFb	Positive transcription elongation factor B
PVDF	Polyvinylidene fluoride
qPCR	Quantitative polymerase chain reaction
qRT-PCR	Quantitative reverse transcriptase polymerase chain reaction
Rb	Retinoblastoma
RBBP4/7	Retinoblastoma-binding protein 4 or 7
RNA	Ribonucleic acid
RNAPII	Ribonucleic acid polymerase II
RPKM	Reads per Kilobase Million
RPMI	Roswell Park Memorial Institute
RUNX1	Runt-related transcription factor 1
RYBP	RING1/YY1-binding protein

SAHF	Senescence-associated heterochromatic foci
SAM	S-adenosyl-methionine
SAM	Sterile alpha motif
SEC	Super elongation complex
SEM	Standard Error of the Mean
-seq	sequencing
SKP1	S-Phase Kinase-Associated Protein 1
SOC	Super Optimal broth with Catabolite repression
SUZ12	Suppressor of Zeste 12
SWI/SNF	Switch/sucrose non-fermenting
TAD	Topologically associating domain
TALEs	Transcription activator-like effectors
T-ALL	T-cell acute lymphoblastic leukaemia
TBP	TATA-box binding protein
TBS	Tris-buffered saline
TetO	Tetracycline operator
TetR	Tetracycline repressor
TF	Transcription factor
TFDP1	Transcription factor Dp-1
trr	Trithorax-related

trx	Trithorax
TrxG	Trithorax group
TSS	Transcription start site
UTX	Ubiquitously transcribed tetratricopeptide repeat on chromosome X
VEFS	VRN2–EMF2–FIS2–SUZ12 domain
vPRC1	Variant Polycomb Repressive Complex 1
WDR5	WD repeat-containing protein 5
WDR68	WD repeat-containing protein 68
Xist	X inactive specific transcript

Chapter 1 – Introduction

1.1 Structure of chromatin

In eukaryotic cells the long DNA molecules consisting of up to 250 million base pairs, are packaged in the nucleus in the form of chromatin. The basic subunit of chromatin is the nucleosome, which is composed of ~147 bps of DNA wrapped around a histone octamer, consisting of two copies of the canonical histones H2A, H2B, H3 and H4 (Cutter and Hayes, 2015; Kornberg, 1977; Luger et al., 1997).

The C-terminal part of each histone contains a highly conserved 'histone fold' motif, consisting of three α-helices (α1, α2 and α3) connected by two short loops (L1 and L2). Interactions between L1 and L2 result in the formation of heterodimers between H3-H4 and H2A-H2B histones. Additional interactions between the heterodimers result in the tripartite structure of the histone octamer, in which a central H3-H4 dimer is flanked by two H2A-H2B dimers (Arents et al., 1991). In contrast to the C-terminal part, the N-terminal 'tail' domain of the histones is unstructured. Residues in the N-terminal and C-terminal part of each histone are subject to posttranslational modifications, central to epigenetic signalling (Luger et al., 1997).

The chromatin structure does not only function to package DNA inside the cell nuclei; rather the three dimensional folding of the eukaryotic genome is a highly dynamic process ensuring the orchestration of complex biological processes including DNA replication and transcription. Understanding the principles of chromatin folding is essential for elucidating the mechanisms of gene regulation, and advances in microscopy and next generation sequencing during the past decades have resulted in a significant gain in knowledge of the spatial organisation of chromatin. Based on these studies, it has been established that chromosomes occupy specific non-random regions in the nucleus, known as chromosome territories, with gene-rich chromosomes tending to occupy more central positions, while gene-poor chromosomes appearing to be in the nuclear periphery (Cremer and Cremer, 2010; Croft et al., 1999). Within chromosome territories,

chromosomes segregate into large ‘A’ or ‘B’ compartments, which contain active, gene-rich or inactive chromatin respectively (Lieberman-Aiden et al., 2009; Wang et al., 2016b).

If we further zoom in, at the scale of tens to hundreds of kilobases (kbs), we recognise domains known as Topologically Associated Domains (TADs) that correspond to regions within which chromatin contacts occur more frequently compared to interactions with other parts of the genome. Enhancer-promoter interactions are confined within the same TAD (Jin et al., 2013; Sun et al., 2019; Symmons et al., 2014), whose boundaries are defined by two convergently oriented CTCF sites and are enriched in cohesin (Dixon et al., 2012; Hou et al., 2012; Sofueva et al., 2013). Experimental data suggests that the mechanism driving TAD formation is loop extrusion, and more specifically that the ring-shaped cohesin molecules are loaded onto chromatin and start moving along DNA which extrudes as a loop. This loop extrusion is halted when cohesin encounters two CTCF binding sites in a divergent orientation (Sanborn et al., 2015).

Disruption in the formation of TAD boundaries can lead to the formation of inappropriate contacts between promoters and regulatory elements residing in different TADs and deregulation of gene expression, which can promote carcinogenesis or lead to developmental defects and malformations (Flavahan et al., 2016; Hnisz et al., 2016; Lupianez et al., 2015). Thus, given the importance of chromatin spatial organisation and its potential impact on human disease, it is crucial to further understand the precise mechanisms leading to the formation of TADs and the establishment of chromatin interactions.

1.2 Histone modifications and regulation of gene expression

The central globular domain and the N-terminal tails of the histones are subject to multiple covalent posttranslational modifications including acetylation, methylation, ubiquitylation, phosphorylation and other less well studied modifications (Bannister and Kouzarides, 2011). Catalytic enzymes that mediate the deposition and the removal of histone modifications are known as ‘writers’ and

'erasers' respectively, and the proteins that recognise specific histone marks are known as 'readers' or 'effectors' (Baker et al., 2008; Ruthenburg et al., 2007a). Histone modifications affect inter-nucleosomal interactions and regulate the chromatin structure directly and also by promoting (or hindering) the binding of effectors, including chromatin-modifying enzymes. Although the function of histone modifications is still incompletely understood, changes in chromatin structure as a result of changes in histone modifications can have profound impacts on transcriptional regulation.

More specifically, acetylation of lysine residues mediated by histone acetyltransferases (HATs) neutralises the positive charge of lysines and weakens the electrostatic interactions between DNA and histones. This is associated with increased DNA accessibility and binding of protein complexes including transcription factors and the transcriptional machinery (Brownell and Allis, 1996; Shahbazian and Grunstein, 2007). Acetylation even at a single residue, H4K16, has been shown, at least *in vitro*, to inhibit the formation of 30 nm fibre, which consists of compacted nucleosome arrays and has been identified as a compacted state of chromatin *in vitro* (Shogren-Knaak et al., 2006). Furthermore, acetylated lysine residues can be recognised by bromodomain (BRD)-containing proteins (Dhalluin et al., 1999). BRD-containing proteins are clustered into 8 families, and include HATS, nucleosome-remodelling complexes like the BRG1 subunit of the switch/sucrose non-fermenting (SWI/SNF) complex and the bromodomain and extraterminal (BET) family of proteins (Fujisawa and Filippakopoulos, 2017). The BET family consists of BRD2, BRD3, BRD4 and the testis-specific BRDT proteins. Of special importance are BRD4 and BRDT, since it has been shown that they can interact with the positive transcription elongation factor b complex (P-TEFb) (Gaucher et al., 2012; Yang et al., 2005). P-TEFb consists of cyclin T1/2 and cyclin-dependent kinase 9 (CDK9), which phosphorylates the Carboxy-terminal domain (CTD) of RNA polymerase II and promotes transcriptional elongation (Marshall and Price, 1995).

In contrast to acetylation, the addition of methyl groups by histone methyltransferases on lysine and arginine residues does not seem to affect histone charges and directly cause dramatic changes

to the chromatin structure. Instead, methylated histone residues have been reported to function as docking sites for reader proteins, and a large number of protein domains have been identified to bind to methylated lysine residues, including plant homeodomain (PHD), tudor, double/tandem tudor, chromo, WD40 and PWWP domains (Yun et al., 2011). For example, H3K27me3 deposited by the PRC2 complexes can be recognised by the chromodomain of chromobox (CBX) subunits of PRC1 complexes (Fischle et al., 2003; Min et al., 2003). In turn, PRC1 complexes have the ability to polymerize through the CBX and Polyhomeotic-like (PHC) components and result in local chromatin compaction and long-range interactions between Polycomb-bound loci (Isono et al., 2013; Kim et al., 2002; Kim and Kingston, 2020; Kundu et al., 2017; Lau et al., 2017; Wani et al., 2016), as will be discussed in greater detail below. Another example of recognition of methylated lysine residues is the binding of deacetylase Rpd3 through its PHD and chromodomains to H3K36me3 across gene bodies. It has been reported that Rpd3 removes acetyl marks during transcriptional elongation, preventing inappropriate intragenic transcription (Carrozza et al., 2005; Li et al., 2007).

These examples illustrate the cross-talk between histone modifications and downstream effectors, and highlight how changes in histone marks can impact transcription and downstream biological outputs. The changes in chromatin accessibility as a result of the deposition of specific histone marks and the recognition of histone-modified nucleosomes by reader proteins seem to be central for the regulation of chromatin-associated processes, and more research is needed to better understand the cross-talk between histone marks, for example understand the cross-talk between active histone modifications (Worden et al., 2019) or the mechanistic details of how repressive histone marks inhibit the deposition of active ones (Fursova et al., 2021).

1.3 Transcription factors and regulation of gene expression

Ensuring that genes are expressed at the correct time and in the correct tissues is a highly regulated process that requires the coordinated activity of transcription factors (TFs). TFs can bind to both

promoters and enhancers and their association with co-activators (or co-repressors) leads to the recruitment of multi-protein complexes including the basal transcriptional machinery, and to activation (or repression) of target genes (Lambert et al., 2018; Spitz and Furlong, 2012).

Most TFs include a structured DNA binding domain (DBD), that typically recognises small 6-12 bp-long DNA sequences, and an activation domain (AD), which is usually unstructured (Keegan et al., 1986; Luscombe et al., 2000; Mitchell and Tjian, 1989). Through their DBDs, TFs bind to their cognate sites on DNA; it has been suggested that missense variations in TF DBDs that affect their DNA binding affinity or specificity are prevalent in humans and are associated with phenotypic variation. For example, mutations in the DBD of *HOXD13* are associated with limb malformations (Barrera et al., 2016). In addition, somatic mutations or variants that lead to the acquisition of novel TF binding sites or conversely result in loss of TF binding sites can change the binding profile of TFs and lead to human pathologies, including cancer and obesity (Claussnitzer et al., 2015; Mansour et al., 2014).

Despite the initial assumptions describing stable TF binding, it has emerged that the binding of TFs is a highly dynamic process (Hager et al., 2009). It has been suggested that TFs search for their binding sites along the DNA through facilitated diffusion, which includes non-specific stochastic collisions with DNA and more stable binding to their cognate DNA binding sites (Hager et al., 2009; von Hippel and Berg, 1989). It has been estimated that the residence times of TFs on DNA are in the range of seconds, and that at any given time, more than half of the TF molecules are bound to chromatin (Chen et al., 2014a; Lu and Lionnet, 2021; Xie et al., 2017). For example, single molecule tracking experiments examining the binding of Sox2 to its target sites in mouse embryonic stem cells (mESCs) have shown that Sox2 uses a trial-and-error target search mechanism until it engages with its cognate DNA binding sites (Chen et al., 2014a). Transient non-specific collisions last for approximately 0.7 seconds, whereas during more stable binding events, Sox2 binds to its cognate DNA binding sites for approximately 12 seconds. The residence time in the scale of seconds

underlines the dynamic behaviour of TF binding, especially given the stability of chromatin, i.e. the slow exchange rate of histones, with only a small fraction of core histones H3 and H4 being exchanged per hour (Kimura and Cook, 2001). The search time, i.e. the time that a TF spends between two stable binding events, is, among other factors, dependent on the concentration of the TF binding sites (Reisser et al., 2018) and the 3D conformation of the genome (Cortini and Filion, 2018; Lu and Lionnet, 2021). Biophysical studies propose that the residence times of TFs influence transcription bursts. More specifically, longer residence times of artificial repressors consisting of a TF fused to the DNA binding domain of transcription activator-like effectors (TALEs) resulted in reduced burst frequency through antagonism with endogenous activators (Clauss et al., 2017). In contrast, longer residence time of the activator Rap1 led to increased transcription, while fast turnover of Rap1 was associated with low transcriptional output (Lickwar et al., 2012). In addition, it has been shown that variations in TF residence time had more significant effect on burst frequency than variations in TF concentration (Popp et al., 2020)

The ADs of TFs are characterised by the presence of intrinsically disordered regions (IDRs). The majority of eukaryotic TFs have IDRs and the degree of disorder in the ADs is significantly higher compared to that in DBDs (Liu et al., 2006). The IDRs become structured upon interaction with other molecules (Frankel and Kim, 1991). Despite the traditional models of protein-protein domain interactions that were described as 'lock-and-key' interactions, the presence of IDRs could provide an explanation for the ability of TFs to interact with multiple co-factors. This results in increased tunability and versatility, which are required for gene regulation in eukaryotic cells (Staby et al., 2017). In addition, IDRs are considered to be key for the formation of phase separated condensates, which are nuclear membraneless structures that have been proposed to be integral for transcriptional activation (Boija et al., 2018; Sabari et al., 2018), as will be discussed in section 1.4.

It has been shown that for most TFs to bind to the DNA sequence they recognise, a certain degree of accessibility and the presence of active histone marks are required (Guertin and Lis, 2010;

Guertin and Lis, 2013; John et al., 2011). For example, chromatin accessibility in combination with the delayed accumulation of the repressive H3K27me3 mark on nascent DNA following replication in ESCs and haematopoietic progenitors provides a "window" for the binding of lineage-specific TFs (Petruk et al., 2017a; Petruk et al., 2017b). In addition, the repressive H3K9me3 modification impedes TF binding (Soufi et al., 2012). The levels of DNA methylation seem to have a different effect on the recruitment of TF; some TFs preferentially bind to methylated CpGs, while the binding of other TFs is inhibited by methylated DNA (Cusack et al., 2020; Yin et al., 2017).

Interestingly, a subset of TFs have been characterised as pioneer TFs, as they have the ability to compete with nucleosomes for binding to DNA. Unlike most TFs that are unable to bind to nucleosomal DNA, pioneer TFs can bind to partial DNA motifs exposed on the surface of a nucleosome and facilitate the cooperative binding of other TFs (Roberts et al., 2021; Zaret, 2020). Among the most well-studied pioneer factors are FOXA1 (Cirillo et al., 2002; Cirillo et al., 1998), and the pioneer factors OCT4, KLF4 and SOX2 which can induce reprogramming of fibroblasts into induced pluripotent stem cells. Unlike the majority of TFs, the binding of which is impeded at closed heterochromatic regions (Soufi et al., 2012), the pioneer factor TCF-1 has been shown to bind to nucleosomes enriched in H3K27me3 and H3K9me3, and its binding is followed by gain in H3K27ac (Johnson et al., 2018). Similarly, the TF Pax7 is able to bind to sites with intermediate H3K9me3 and high H3K9me2 levels, but not at sites with high H3K9me3 and high H3K9me2 levels and at insulator sites (Mayran et al., 2018).

The robust control of gene expression is controlled through the concerted action of TFs and chromatin modifying complexes. TFs physically interact with, and recruit, chromatin proteins at their binding sites including acetyltransferases, the Mediator complex and nucleosome-remodelling complexes, or chromatin complexes that function as repressors (Fournier et al., 2016; Hu et al., 2011; Reed-Inderbitzin et al., 2006). In turn, these chromatin complexes modify the chromatin landscape by affecting the histone modifications and by altering accessibility, which

further affects the binding of additional TFs. The association of TFs with co-activators or co-repressors can lead to gene activation or repression respectively. In line with this, some TFs, including Runx1 and Pax5, have been characterised to function both as activators and as repressors, even in the same cell type, depending on the chromatin complexes they associate with (Bonifer et al., 2017; McManus et al., 2011). The presence of IDRs on a TF could explain their association both with co-activators and co-repressors, in turn dictating their function as activators or repressors in a gene-specific or cell-specific fashion.

1.4 Function of enhancers in gene regulation and the role of enhancer-promoter interactions

Enhancers were first described as genomic elements that stimulate gene transcription in a time- and tissue-specific manner, even when they are located thousands of kilobases away from their target gene and irrespective of their orientation (Banerji et al., 1981; Benoist and Chambon, 1981). Enhancers contain binding sites for TFs, including lineage-specific TFs (Spitz and Furlong, 2012), and they were first identified through functional assays, based on their ability to increase the transcription of a reporter gene with a minimal promoter. Initial efforts to identify enhancers genome-wide focused on identifying non-coding sequences highly conserved between the different species (Allende et al., 2006; Pennacchio et al., 2006), but conservation is neither sufficient nor necessary to predict whether a non-coding sequence functions as an enhancer (Blow et al., 2010; Visel et al., 2008).

The advent of next generation sequencing has enabled the genome-wide identification of enhancers across different cell types based on the presence of histone modifications (Calo and Wysocka, 2013), and more specifically, it was shown that enhancers are enriched for H3K4me1, whereas promoters are enriched for H3K4me3 (Heintzman et al., 2007). Further studies showed that the presence of H3K4me1 alone at distal elements marks not only active, but also poised, enhancers and it is the presence of H3K27ac that distinguishes poised versus active enhancer

elements (Creyghton et al., 2010; Rada-Iglesias et al., 2011). H3K27ac is catalysed by the acetyltransferases CBP/P300 (Bannister and Kouzarides, 1996; Delvecchio et al., 2013; McManus and Hendzel, 2003; Ogryzko et al., 1996). CBP/P300 are recruited to enhancers via their interaction with TFs (Rada-Iglesias et al., 2011; Visel et al., 2008). Interestingly, fusion of nuclease-deficient Cas9 (dCas9) with the acetyltransferase domain of p300 resulted in an increase of H3K27ac and to transcriptional activation when targeted with guide RNAs to endogenous enhancers and promoters (Hilton et al., 2015). In addition, targeting the H3K4me1/2 demethylase LSD1 fused to TALE or dCas9 to enhancers resulted in reduction of H3K4me2 and H3K27ac and repression of gene expression (Kearns et al., 2015; Mendenhall et al., 2013), suggesting the role of these histone marks for enhancer function.

However, further studies also identified enhancers that lack H3K27ac and are instead marked with different histone modifications, including H4K16ac (Taylor et al., 2013) and H3K122ac (Pradeepa et al., 2016; Tropberger et al., 2013), which suggests that the open chromatin conformation that is mediated by acetylation of histone residues is a vital feature of enhancer function. This further suggests that other histone modifications should also be taken into consideration for the annotation of enhancers.

The mammalian MLL3 and MLL4 methyltransferases, homologs of *Drosophila* trithorax-related (trr), are the methyltransferases responsible for catalysing H3K4me1 at enhancers (Herz et al., 2012; Hu et al., 2013a), and are discussed in more detail in section 1.7. The deposition of H3K4me1 by MLL3/4 has been reported to lead to enhancer priming, which is followed by acetylation of H3K27 by CBP/p300 on active enhancers (Lai et al., 2017; Wang et al., 2016a). However, other work has suggested that although MLL3 /4 are essential for enhancer function, their methyltransferase activity is not crucial (Dorighi et al., 2017). Since no physical interaction between MLL3/4 and lysine acetyltransferases has been identified, the mechanistic details of how H3K4me1 leads to enhancer priming and deposition of H3K27ac still has to be further elucidated (Lai et al., 2017). It has been

suggested that H3K4me1 is preferentially recognised by the chromodomain of acetyltransferase TIP60, in contrast to other chromodomain and PHD domain-containing proteins that preferentially bind to H3K4me3 over H3K4me1 (Jeong et al., 2011). The addition of acetyl marks by TIP60 could result in a more open chromatin conformation which could in turn facilitate binding of TFs that interact with CBP/P300 (Calo and Wysocka, 2013; Jeong et al., 2011).

Other characteristics of enhancers include DNAase I hypersensitivity (Gross and Garrard, 1988), DNA hypomethylation (Lister et al., 2009) and the production of non-coding enhancer RNAs (eRNAs) of varying polyadenylation status and strand specificity (Kim et al., 2010; Orom et al., 2010). Enhancers are, similarly to promoters, bound by TFs, co-factors including p300 (Rada-Iglesias et al., 2011; Visel et al., 2008), components of the Mediator complex (Loven et al., 2013; Whyte et al., 2013), BRD4, ATP-dependent nucleosome-remodelling enzymes, including the SWI/SNF complex, general transcription factors (GTFs) and RNA polymerase (Koch et al., 2011). A class of enhancers called super-enhancers have been identified, which are defined as clusters of enhancers with high levels of H3K27ac, H3K4me1 and are bound by BRD4, Mediator and important developmental TFs, and regulate the expression of key developmental genes (Loven et al., 2013; Whyte et al., 2013). However, beyond the identification of super-enhancers, they have been difficult to functionally validate, and to show how they function differently from non super-enhancers. Interestingly, despite the classic divide between enhancers and promoters, it can be argued that enhancers share many characteristics with promoters. Given that promoters can interact with other promoters (Li et al., 2012) and can function as enhancers (Dao et al., 2017; Medina-Rivera et al., 2018; Rennie et al., 2018), and conversely, that intragenic enhancers can act as alternative promoters (Kowalczyk et al., 2012), the perception that enhancers and promoters are different entities has been challenged (Andersson, 2015). Understanding these distinctions continues to be an active area of research for the field.

How the binding of tissue-specific TFs at enhancers affects gene expression at promoters of target genes, sometimes located up to 1 Mb away (Lettice et al., 2003), is under intensive investigation.

The prevailing model is that TF-bound enhancers come in close proximity with the transcriptional machinery at their target promoters, and co-activators mediate transcriptional activation. The mechanistic details explaining the spatial proximity between enhancers and promoters are incompletely understood. Enhancer-promoter interactions are confined within the same TAD (Jin et al., 2013; Sun et al., 2019; Symmons et al., 2014). Enhancer-promoter proximity mediated by loop formation, i.e. by looping out the intervening DNA, is considered the prevalent mechanism explaining how enhancer-promoter communication is achieved over long distances within TADs. Looping is supported by chromosome conformation capture techniques (3C) and its derivatives, and elegant studies show that enforcing looping between enhancers and promoters results in gene activation (Bartman et al., 2016; Deng et al., 2012; Deng et al., 2014). It should be noted, however, that there is experimental data reporting enhancer-dependent transcription without enhancer-promoter proximity (Alexander et al., 2019; Benabdallah et al., 2019), suggesting that there is much more that needs to be understood about this complex process.

Several proteins have been reported to have an architectural role in mediating enhancer-promoter looping, including TFs (Drissen et al., 2004; Nolis et al., 2009; Vakoc et al., 2005; Weintraub et al., 2017); the Lim-domain-binding 1 (LDB1), which binds to TFs and can also form homodimers and acts as a bridge between TF-bound enhancers and promoters (Deng et al., 2012; Krivega et al., 2014; Krivega and Dean, 2017; Matthews and Visvader, 2003) (Figure 1.1 *top panel*); and Mediator (Kagey et al., 2010). In addition, it has also been suggested that cohesin functions as an architectural protein in mediating enhancer-promoter interactions (El Khattabi et al., 2019) (Figure 1.1 *bottom panel*). More specifically, it has been reported that at CTCF-independent sites cohesin co-localises with tissue-specific TFs (Schmidt et al., 2010). A subset of those sites are found at enhancers and promoters of active genes and are co-bound by Mediator (Kagey et al., 2010). It has been further shown that degradation of cohesin resulted in decrease of enhancer-promoter interactions (El

Khattabi et al., 2019). These results suggest that different factors can be responsible for mediating a subset of enhancer-promoter interactions in different cellular contexts/experimental systems.

Emerging evidence suggests that at enhancers in close proximity to each other and at super-enhancers, the intrinsically disordered ADs of TFs are found together with Mediator and RNA polymerase II in phase-separated condensates that form distinct non-membrane compartments in cells (Boija et al., 2018; Cho et al., 2018; Sabari et al., 2018). Liquid-liquid phase separation could provide an attractive mechanism to explain chromatin interactions, but given that the disruption of phase condensates induces transcriptional changes without affecting enhancer-promoter interactions (Crump et al., 2021), the role of phase separation on chromatin interactions needs to be further evaluated.

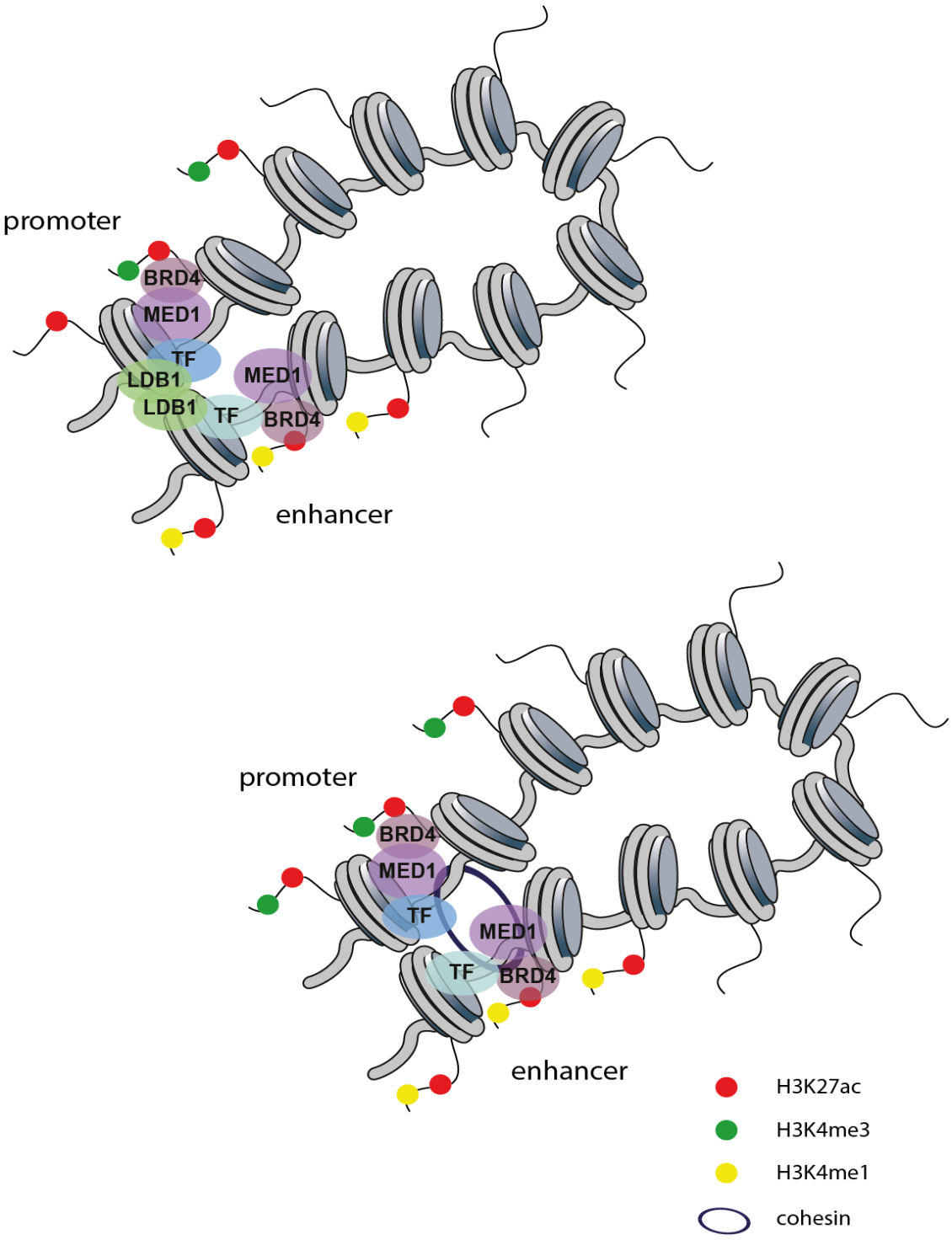


Figure 1.1: Schematic of enhancer-promoter interactions. Enhancers (marked by H3K27ac and H3K4me1 histone modifications) and promoters (marked by H3K27ac and H3K4me3 modifications) are bound by TFs and co-activators including BRD4 and the Mediator subunit MED1. Physical proximity between enhancers and promoters can be mediated through the dimerization of the Lim-domain-binding 1 (LDB1) protein *(top panel)* or by the cohesin molecule *(bottom panel)*, among other mechanisms.

1.5 Introduction to Polycomb Group (PcG) and Trithorax Group (TrxG) proteins

Polycomb group (PcG) proteins were initially identified in *Drosophila*, as regulators of fly body development (Duncan and Montgomery, 2002). The name Polycomb was given due to the phenotype that arises from their deficiency, which is the appearance of extra sex combs in the second and third pairs of thoracic legs of adult male flies, while wildtype males only have sex combs on the first pair of legs (Lewis and Mislove, 1947; Slifer, 1942). PcG proteins control the development of body segments in *Drosophila* by functioning as repressors of the *Homeotic* (*Hox*) genes (Lewis, 1978). In turn, the *Hox* genes, which form two clusters, the Antennapedia and the Bithorax clusters, control fly body development along the anterior-posterior (A-P) axis (Kaufman et al., 1980; Lewis, 1978). As a result, mutations in the *Polycomb* group of genes cause ectopic expression of *Hox* gene products along the A-P axis (McKeon and Brock, 1991; Simon et al., 1992).

PcG proteins are not required for initiation of silencing of *Hox* genes, but for maintenance of their transcriptional status in later stages of embryonic development (Jones and Gelbart, 1990; Simon et al., 1992; Struhl and Akam, 1985). The initial establishment of *Hox* gene expression patterns is determined by the concerted action of transiently expressed TFs encoded by the gap and pair-rule genes. Once the proteins encoded by these genes disappear, it is the expression of the PcG proteins that maintains the transcriptional repression of the *Hox* genes (Casares and Sanchez-Herrero, 1995; Harding and Levine, 1988; Ingham and Martinez-Arias, 1986; Jack and McGinnis, 1990; Reinitz and Levine, 1990). In embryos with mutations in genes encoding for PcG proteins, the expression of *Hox* genes is established correctly, but then inappropriately spreads beyond the normal segments (McKeon and Brock, 1991; Simon et al., 1992; Struhl and Akam, 1985; Wedeen et al., 1986).

The PcG proteins form multiprotein complexes, the Polycomb Repressive Complexes (PRCs), and the best studied are the PRC1 and PRC2 complexes, discussed in more detail in the next sections.

Gene silencing by PcG proteins at *Hox* genes and other loci is antagonised by the Trithorax Group proteins (TrxG), whose function is to sustain gene expression at active genes (Hanson et al., 1999; Kassis et al., 2017). The first *TrxG* gene to be identified was *trithorax* (*trx*) and its mutations cause homeotic transformations of the thoracic and abdominal segments in *Drosophila* (Ingham and Whittle, 1980). Other members of the TrxG genes were identified by mutations that mimic the loss-of-function phenotypes of the *Hox* genes. The TrxG proteins are more heterogeneous than the PcG proteins, can be divided in two broad classes based on their function, the SWI/SNF chromatin remodelling complexes and the COMPASS (complex of proteins associated with Set), also known as the MLL (Mixed Lineage Leukemia) proteins that function as H3K4 methyltransferases (Cenik and Shilatifard, 2021; Krogan et al., 2002; Miller et al., 2001; Stern et al., 1984). The latter will be discussed in section 1.7.

1.6 Polycomb Group proteins

1.6.1 PRC1 complex

In *Drosophila,* the PRC1 complex contains the core components Polycomb (Pc), that recognises H3K27me3 (Min et al., 2003); dRing, which is the catalytic subunit responsible for catalysing ubiquitylation of histone H2A on lysine 118 (H2AK118ub); Posterior Sex Combs (Psc) that mediates chromatin compaction *in vitro* (Francis et al., 2004); and Polyhomeotic (Ph) (Shao et al., 1999).

In mammalian cells, for each *Drosophila* Polycomb protein there are several mammalian homologs, which associate with each other in a combinatorial fashion, resulting in a great degree of diversity in the subunit composition of PRC1 complexes (Gao et al., 2012). The central components of all mammalian PRC1 complexes are the catalytic subunits RING1A and RING1B (homologous to dRing in *Drosophila*), which are mutually exclusive, and one of the six Polycomb group ring finger proteins (PCGF1-PCGF6), which determine the identity of the PRC1 sub-complexes (PRC1.1-PRC1.6) (Gao et

al., 2012). RING1A and RING1B are E3 ubiquitin ligases that catalyse ubiquitylation of histone H2A on lysine 119 (H2AK119ub) in mammalian cells (Li et al., 2006; Wang et al., 2004a). The PCGF proteins, homologs of the *Drosophila* Psc, are also mutually exclusive and association of RING1A/B with the PCGF proteins enhances their catalytic activity.

PRC1 complexes are subdivided into canonical and non-canonical complexes based on the proteins that associate with RING1A and RING1B (Figure 1.2). In canonical PRC1 complexes (cPRC1), RING1A and RING1B associate with the CBX proteins (CBX2, CBX4, CBX6, CBX7 and CBX8) and the Polyhomeotic-like proteins (PHC1, PHC2, PHC3) which are homologs of *Drosophila* Pc and Ph respectively. The PCGF proteins found in the cPRC1 complexes are PCGF2 (also known as MEL18 - melanoma nuclear protein 18) and PCGF4 (also known as BMI1 - B cell-specific Moloney murine leukemia virus integration site 1) (Gao et al., 2012) (Figure 1.2). The canonical PRC1 complex makes little contribution to H2AK119ub and gene repression (Blackledge et al., 2014; Fursova et al., 2019; Rose et al., 2016), but seems to play a role in mediating higher-order chromatin structures (Isono et al., 2013; Kim et al., 2002; Kim and Kingston, 2020; Kundu et al., 2017; Lau et al., 2017; Wani et al., 2016).

The canonical PRC1 complexes are bound to chromatin through the association of CBX proteins with H3K27me3 (Bernstein et al., 2006b). Similarly, in *Drosophila*, the Pc protein directs the localisation of the PRC1 complexes to H3K27me3 (Fischle et al., 2003; Min et al., 2003). The domain responsible for this association is the chromodomain (chromatin organization modifier domain), which shares sequence similarity with the chromodomain in the HP1 protein, which binds to H3K9me3 and is involved in constitutive heterochromatin silencing (Paro and Hogness, 1991). Elegant experiments showed that substitution of the Pc chromodomain with that of HP1 was sufficient to recruit Pc to constitutive heterochromatin, and conversely, substitution of the chromodomain of HP1 with the Pc chromodomain led to HP1 relocalisation to Polycomb sites (Fischle et al., 2003; Platero et al., 1995).

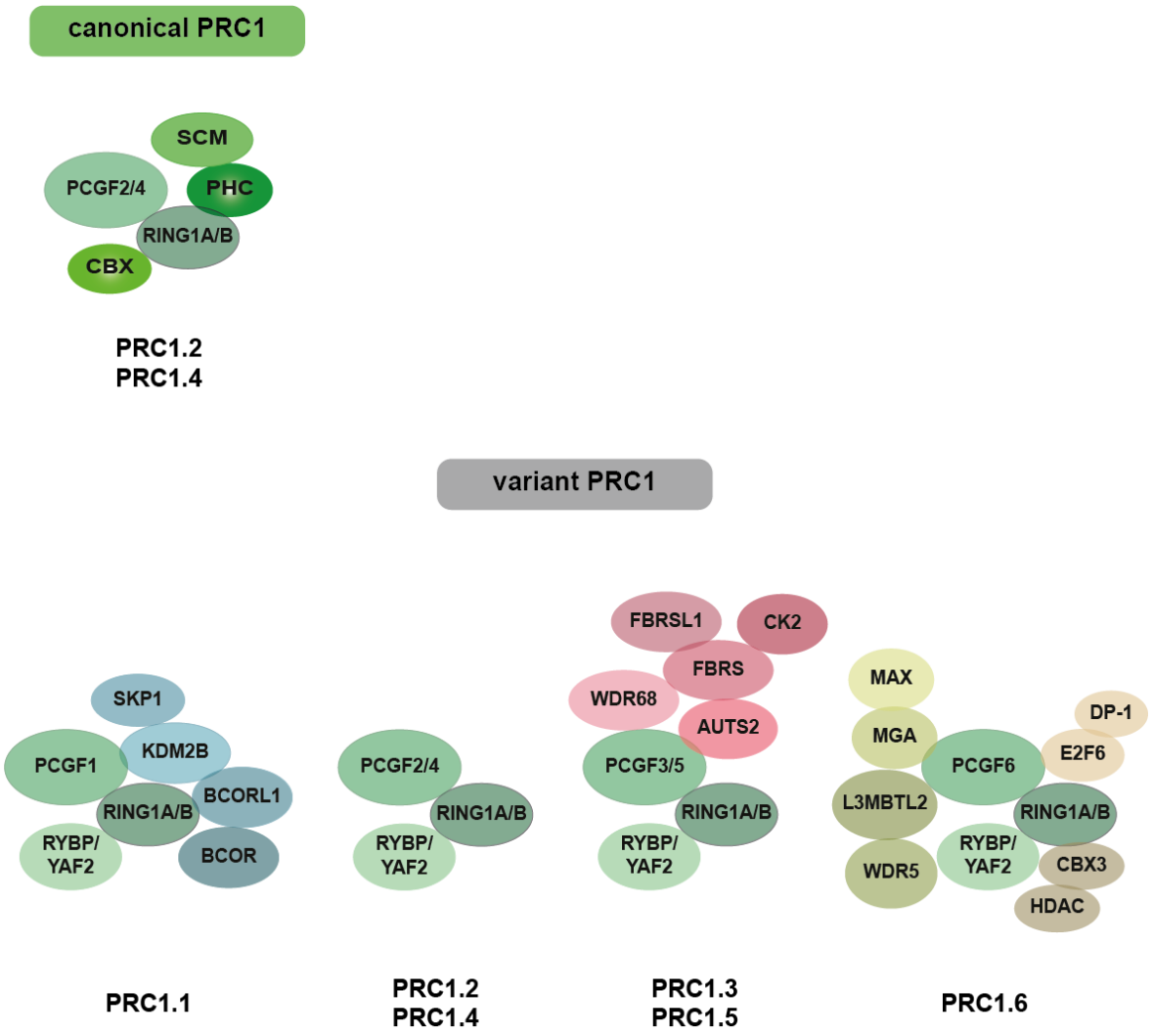


Figure 1.2: Schematic of PRC1 complexes. *(Top panel)*: Schematic of canonical PRC1 complexes PRC1.2 and PRC1.4. *(Bottom panel)*: Schematic of variant PRC1 complexes PRC1.1, PRC1.2/4, PRC1.3/5 and PRC1.6.

According to the traditional model of Polycomb recruitment, upon PRC2 recruitment and deposition of H3K27me3, this histone mark gets recognised by the chromodomain of the PRC1 CBX proteins. In turn, upon its recruitment, PRC1 catalyses the ubiquitylation of histone H2A on lysine 119 (H2AK119ub) in mammals and on lysine 118 (H2AK118ub) in *Drosophila*. However, seminal findings during the last decade have dramatically changed our understanding about the crosstalk between PRC1 and PRC2 complexes and the order of events during Polycomb recruitment. More

specifically, it has been demonstrated that PRC1 binding and deposition of H2AK119ub can be the first step towards the generation of a Polycomb domain, followed by PRC2 binding (the hierarchy of PRC1/2 recruitment will be discussed in more detail below) (Blackledge et al., 2014; Cooper et al., 2014; Fursova et al., 2019).

In contrast to the canonical PRC1 complexes, the non-canonical complexes (ncPRC1) lack the CBX and the PHC proteins, and the catalytic subunits RING1A and RING1B associate with either RING1/YY1-binding protein (RYBP) (Garcia et al., 1999) or YY1- associated factor 2 (YAF2) (Kalenik et al., 1997), and one of the PCGF proteins (Figure 1.2). RYBP and YAF are included in the complex in a mutually exclusive manner, as they interact with the same domain of RING1B (Gao et al., 2012). The non-canonical complexes have increased enzymatic activity as the association of RING1B with RYBP and YAF2 leads to a conformation change of RING1B (Gao et al., 2012; Rose et al., 2016). Interestingly, PCGF2 and PCGF4 can also associate with RYBP/YAF2, but the association with RYBP/YAF2 excludes canonical components (CBX and PHF proteins) from the complex (Gao et al., 2012), consistent with structural data showing that RYBP and CBX7 bind to the same binding site on the C-terminal part of RING1B (Wang et al., 2010b). This demonstrates that RYBP/YAF2 can associate with all PCGF proteins, and furthermore that PRC1.2 and PRC1.4 complexes can form two types of sub-complexes, a canonical one with incorporation of canonical subunits (CBX and PHC proteins) and a non-canonical one when they interact with the RYBP/YAF2 subunits (Gao et al., 2012) (Figure 1.2).

The non-canonical complex PRC1.1, except for PCGF1, RYBP/YAF2 and RING1A/B subunits, also includes proteins KDM2B (Lysine (K)-Specific Demethylase 2B), BCOR (BCL6 Co-repressor), BCORL1 (BCL6 Co-repressor-Like 1) and SKP1 (S-Phase Kinase-Associated Protein 1) (Figure 1.2). Of note, KDM2B is an H3K36 demethylase, and recruits PRC1.1 complex to unmethylated CpG islands through its CXXC domain which recognises unmethylated CpG dinucleotides (Farcas et al., 2012; He et al., 2013; Wu et al., 2013b), as it will be discussed in section 1.6.7.

As far as the PRC1.6 complex is concerned, proteomics studies have identified that it consists of additional components and more specifically, transcription factors E2F6, TFDP1 (transcription factor Dp-1) and MAX (MYC-associated factor X), and proteins MGA (MAX gene-associated protein), CBX3, L3MBTL2 (lethal (3) malignant brain tumor-like 2) and WDR5 (WD repeat-containing protein 5) (Figure 1.2). It has been shown that the heterodimers MGA-MAX and E2F6-TFDP1 are responsible for recruiting the PRC1.6 complex to a subset of Polycomb targets enriched for germ cell-related genes (Endoh et al., 2017; Scelfo et al., 2019; Stielow et al., 2018; Trojer et al., 2011).

The PRC1.3 and PRC1.5 sub-complexes, have almost identical subunit composition, and more specifically the PCGF3 and PCGF5 subunits associate with proteins AUTS2 (Autism susceptibility protein 2), FBRS (Fibrosin), FBRSL1 (Fibrosin-like 1), WDR68 (WD repeat-containing protein 68) and CK2 (Casein kinase 2) (Gao et al., 2012) (Figure 1.2). The recruitment of these complexes is less well understood, but they seem to be the initiating Polycomb complexes to be recruited during X chromosome inactivation (Almeida et al., 2017; Pintacuda et al., 2017). More specifically, it has been demonstrated that the RNA binding protein hnRNPK (heterogeneous nuclear ribonucleoprotein K) is the bridge between *Xist* RNA (X inactive specific transcript) and PRC1.3/PRC1.5 complexes and that PCR1.3/PRC1.5-mediated deposition of H2AK119ub is followed by the recruitment of the other PRC1 sub-complexes and of PRC2 (Almeida et al., 2017; Pintacuda et al., 2017). In addition, it has been recently suggested that a potential mechanism for recruitment of PRC1.3 is through interaction with the Upstream Transcription factor 1 (USF1) (Scelfo et al., 2019).

This heterogeneity in the composition of PRC1 complexes results in functional specificity and explains the complexity of Polycomb recruitment and Polycomb-mediated transcriptional repression. Despite extensive work, the detailed mechanisms of Polycomb recruitment and the role of PcG proteins and the corresponding histone modifications to repression remain incompletely understood.

1.6.2 PRC2 complex

The mammalian PRC2 complex comprises the core subunits Enhancer of Zeste 1 or 2 (EZH1/2), embryonic ectoderm development (EED), Suppressor of Zeste 12 (SUZ12) and retinoblastoma-binding protein 4 or 7 (RBBP4/7) (Kuzmichev et al., 2002; Muller et al., 2002) (Figure 1.3). Additional auxiliary subunits that bind sub-stoichiometrically to the core PRC2 complex stimulate the catalytic activity of PRC2 and mediate its recruitment to chromatin (Healy et al., 2019; Hojfeldt et al., 2019).

The functional integrity of the PRC2 complex during development is crucial, as knockout of the core PRC2 components *Ezh2*, *Eed* and *Suz12* in mice results in embryonic lethality (Faust et al., 1995; O'Carroll et al., 2001; Pasini et al., 2004; Pasini et al., 2010). PRC2 forms two distinct lobes, a catalytic and a targeting lobe (Chen et al., 2018; Ciferri et al., 2012; Glancy et al., 2021) (Figure 1.3). The catalytic one is comprised of EZH1/2, EED and the VEFS (VRN2–EMF2–FIS2–SUZ12) domain of SUZ12, which together mediate the deposition of H3K27 methylation. The targeting lobe consists of the N-terminal part of SUZ12, RBBP4/7 and a set of auxiliary proteins (Chen et al., 2018; Glancy et al., 2021), discussed below. The highly conserved SET domain of the catalytic subunits EZH1 and EZH2 mediates the deposition of all three forms of H3K27 methylation (H3K27me1/2/3) (Ferrari et al., 2014; Kuzmichev et al., 2002; Margueron et al., 2008). H3K27me3 is the most well studied modification of the three and is integral to Polycomb-mediated transcriptional repression (Pengelly et al., 2013). EZH2 is the main H3K27me3 methyltransferase and is more efficient than EZH1 in catalysing this histone mark (Lavarone et al., 2019; Lee et al., 2018; Shen et al., 2008). Association of EZH2 with EED and SUZ12 is necessary for its catalytic activity, as their absence leads to EZH2 autoinhibition, and the binding of EED to H3K27me3 leads to the allosteric activation of PRC2 enzymatic activity (Margueron et al., 2009).

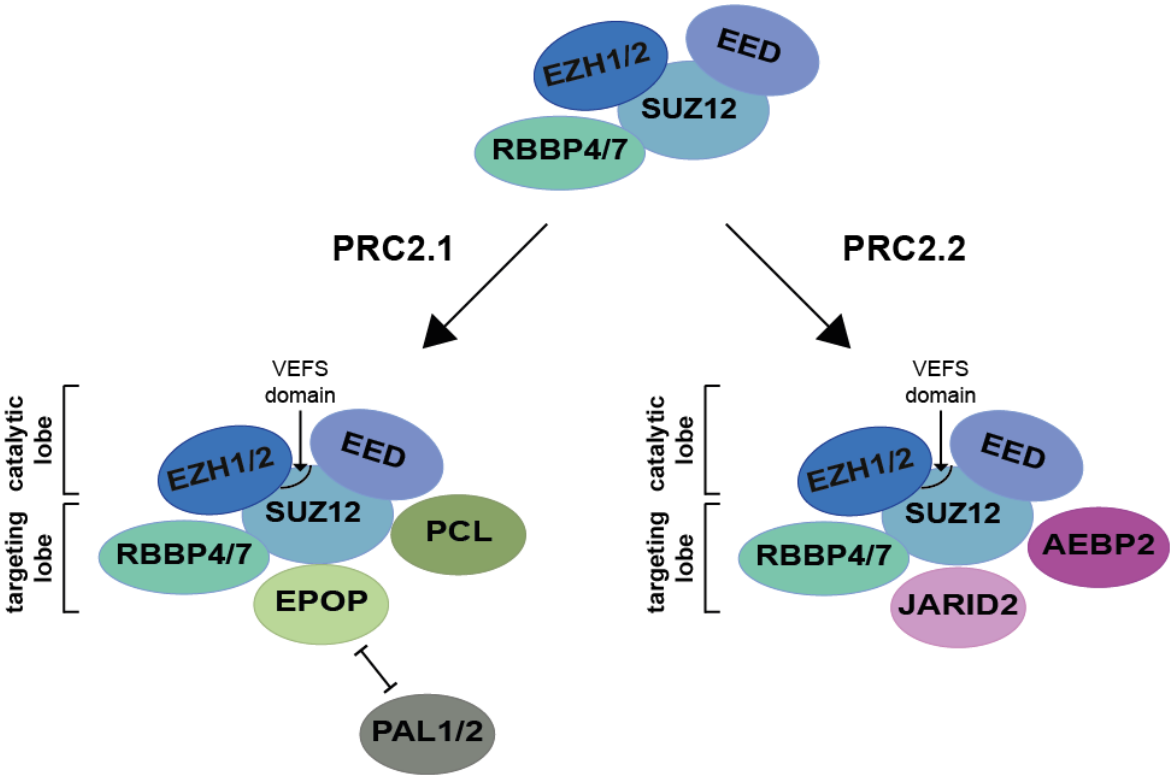


Figure 1.3: Schematic of PRC2 complexes. *(Top panel)*: Schematic of the core PRC2 subunits EZH1/2, EED, SUZ12 and RBBP4/7. *(Bottom panel)*: Schematic of the PRC2 subcomplexes PRC2.1, in which the core subunits associate with the auxiliary subunits PCL1-3 (also known as PHF1, MTF2 and PHF19 respectively) and either EPOP or PAL1/2, and PRC2.2, in which the core subunits associate with the auxiliary subunits JARID2 and AEBP2. The catalytic lobe of PRC2, consisting of EZH1/2, EED and the VEFS domain of SUZ12, and the targeting lobe, consisting of the N-terminal part of SUZ12, RBBP4/7 and the auxiliary subunits, are highlighted.

The auxiliary sub-stoichiometric PRC2 subunits identified by mass spectrometry experiments are JARID2 (Jumonji and AT-rich interaction domain-containing protein 2), AEBP2 (Adipocyte enhancer-binding protein 2), PCL1-3 (Polycomb-like subunit 1-3, also known as PHF1, MTF2 and PHF19 respectively), EPOP (Elongin BC and PRC2-associated protein) and PALI1/2 (PRC2-associated LCOR isoform 1 or 2). These proteins bind in a mutually exclusive fashion to the N-terminal part of SUZ12 and affect PRC2 recruitment and enzymatic activity to target genes. More specifically, the association of the core PRC2 subunits with the auxiliary ones gives rise to distinct PRC2 subcomplexes, PRC2.1 comprising of one of the PCL proteins PCL1-3 together with EPOP or PALI1/2, and PRC2.2, which consists of JARID2 and AEBP2 (Chen et al., 2018; Healy et al., 2019; Hojfeldt et

al., 2019; Kasinath et al., 2021; Lee et al., 2018; Sanulli et al., 2015) (Figure 1.3). Together the non-core auxiliary PRC2 subunits are required for PRC2 targeting to Polycomb-repressed sites, and their combined loss leads to PRC2 mis-localisation (Healy et al., 2019; Hojfeldt et al., 2019; Li et al., 2017).

1.6.3 Hierarchy in the recruitment of Polycomb complexes

It is well established that there is crosstalk between PRC1 and PRC2 complexes and extensive overlap of the two complexes on chromatin. The traditional hierarchical model for the recruitment of PcG proteins had placed PRC1 recruitment downstream of PRC2, which has been shown to be recruited to unmethylated CpG islands through the PCL proteins (Li et al., 2017). In turn, the recruitment of PRC1 is mediated through its binding to H3K27me3 via the chromodomain of CBX proteins (Wang et al., 2004b) (Figure 1.4a). However, experimental evidence during the last decade challenged this long-standing model, as it has been shown that PRC1 can be recruited independently of H3K27me3 (Gao et al., 2012; Pasini et al., 2007; Schoeftner et al., 2006; Tavares et al., 2012), for example through the association of KDM2B with unmethylated CpG islands (Farcas et al., 2012; He et al., 2013; Wu et al., 2013b). In addition, it has been demonstrated that H2AK119ub deposited by PRC1 can also recruit PRC2 (Blackledge et al., 2014; Cooper et al., 2014; Kalb et al., 2014). More specifically, elegant tethering experiments have shown that H2AK119ub catalysed by ncPRC1 complexes is sufficient to recruit PRC2, which in turn catalyses H3K27me3 (Blackledge et al., 2014; Cooper et al., 2014) (Figure 1.4b). At the same time, *in vitro* experiments showed that PRC2 associated with JARID2 and AEBP2 subunits can bind to nucleosomes bearing H2AK119ub, and binding to this histone mark enhances its ability to catalyse H3K27me3 (Kalb et al., 2014). In recent experiments the structural details of PRC2 binding to H2AK119ub have been further elucidated. In particular, the cryo-EM structure of the PRC2-JARID2-AEBP2 complex bound to a nucleosome with H2AK119ub shows that a region at the N-terminal part of JARID2 binds between one ubiquitin molecule and the H2A-H2B acidic patch and AEBP2 interacts with the other ubiquitin molecule and the surface of the other H2A-H2B dimer in the same nucleosome (Kasinath

et al., 2021). Following these studies, it has been established that each Polycomb complex can bind to the histone modifications deposited by the other complex, further reinforcing Polycomb binding and repression at target genes.

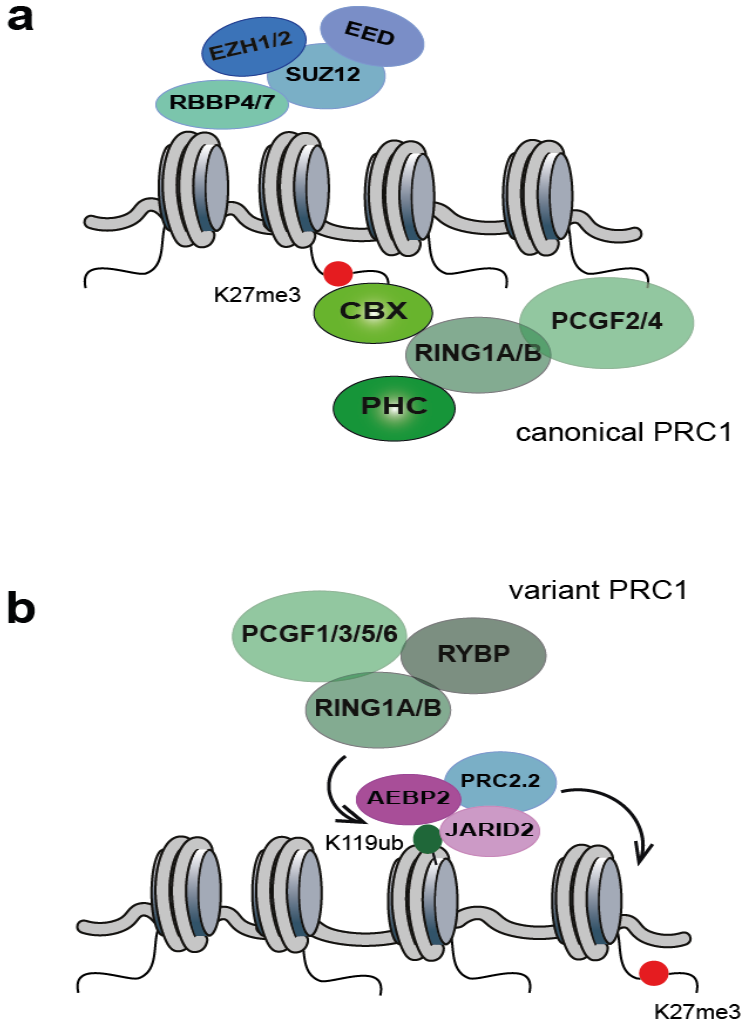


Figure 1.4: Interplay between PRC1 and PRC2 complexes. (a) PRC2 complexes deposit H3K27me3, which is recognised by the CBX subunits of canonical PRC1 complexes. (b) Variant PRC1 complexes catalyse H2AK119ub, which is recognised by the PRC2.2 complex subunits JARID2 and AEBP2. In turn, PRC2 mediates the deposition of H3K27me3.

1.6.4 H3K27 methylation

PRC2 catalyses mono-, di- and trimethylation of lysine 27 at histone H3 (H3K27me1, H3K27me2 and H3K27me3) (Cao et al., 2002; Ferrari et al., 2014; Kurzhals et al., 2008; Kuzmichev et al., 2002; Margueron et al., 2008). It has been shown that the three different forms of H3K27 methylation occupy distinct regions across the genome and more specifically, H3K27me1 is found at the gene bodies of actively transcribed genes; H3K27me2 is spread at intergenic and intragenic regions across the genome; and H3K27me3 is found at the promoters of PcG target genes (Barski et al., 2007; Ernst and Kellis, 2010; Ferrari et al., 2014; Juan et al., 2016; Wang et al., 2008). Binding of PcG proteins overlaps with H3K27me3 sites but not with H3K27me1/me2 sites. Taking into consideration *in vitro* studies which have shown that PRC2 preferentially methylates H3K27me0 and H3K27me1 compared to H3K27me2 (McCabe et al., 2012a; Sneeringer et al., 2010), we can conclude that at H3K27me3 sites, trimethylation of H3K27 is attributed to stable PRC2 binding at its nucleation sites. In contrast, regions enriched in H3K27me1 or H3K27me2 represent genomic sites where PRC2 binding is not stabilised, but rather the deposition of these histone modifications could be attributed to a transient interaction between nucleosomes and PRC2 via a 'hit & run' mechanism.

H3K27me1, found at the gene bodies of actively transcribed genes, overlaps with the active histone modification H3K36me3, and mass spectrometry analysis has shown that H3K27me1 and H3K36me3 co-exist at the same histone tail. In mESCs knockdown of Setd2, the main enzyme that catalyses H3K36me3, results in reduction of H3K27me1 and increase of H3K27me2, which suggests that deposition of H3K36me3 by Setd2 antagonises PRC2-mediated addition of a methyl-mark to H3K27me1 (Ferrari et al., 2014). Consistent with that, further studies have shown that PRC2 fails to methylate histones in the presence of active marks including H3K4me3 and H3K36me3 (Schmitges et al., 2011). Recent studies in *Drosophila* attempted to shed light on whether H3K27me1 is an intermediate modification deposited by PRC2, enriched at active loci where transcription and activating histone modifications do not allow stable binding of PRC2 and its conversion to

H3K27me3; or whether it has a functional role in promoting gene activation. *In Drosophila,* an *E(z)* mutant with a point mutation at the catalytic SET domain that converts E(z) to a monomethyltransferase resulted in derepression of Polycomb target genes that, interestingly, exceeded the derepression observed upon expression of the catalytically inactive E(z) protein (Wang et al., 2018). This suggests that H3K27me1 could have a role in gene activation in contrast to H3K27me3 which is associated with gene repression.

H3K27me2 is the most abundant form of H3K27 methylation, present on 70% of total histone H3. Research in the Pasini lab using *Eed* KO mESCs, suggested that H3K27me2 protects from inappropriate activation of non-tissue specific enhancers. More specifically, in *EED* KO mESCs they observed the appearance of new H3K27ac peaks at distal regulatory sites following the reduction of H3K27me3 and H3K27me2 levels. H3K4me1 ChIP-seq in *Eed* KO cells compared to wild-type cells showed an increase in H3K4me1 at the novel H3K27ac peaks, and RNA-seq revealed an increase in transcription of genes in close proximity to the novel distal H3K27ac sites (Ferrari et al., 2014). Similarly, using a temperature-sensitive *Drosophila* cell line with a homozygous inactivating *E(z)* mutation, the Pirrotta lab showed that upon *E(z)* inactivation, H3K27me2/3 levels decrease. As a result, they observed a general transcriptional derepression accompanied with increase in H3K27ac and H3K4me1 (Lee et al., 2015).

In mESCS undergoing differentiation, a significant proportion of H3K27me2-marked promoters and enhancers gain either H3K27me3 or H3K27ac. Murine ESCs with a point mutation in Ezh2 that enhances its ability to trimethylate H3K27me3, show a reduction in the H3K27me2 levels and upregulation of genes of neural cell fate. These results suggest that the H3K27me2/H3K27me3 ratio can have an important role for cell fate determination in mESCs (Juan et al., 2016).

H3K27me2/3 marks are removed by the demethylases UTX (Ubiquitously transcribed tetratricopeptide repeat on chromosome X), also known as KDM6A (Lysine demethylase 6A); and by JMJD3 (Jumonji domain-containing protein D3), also known as KDM6B (Agger et al., 2007; De

Santa et al., 2007; Lan et al., 2007; Lee et al., 2007b). The two proteins share highly homologous Jumonji C (JmjC) domains and mediate the transition from H3K27me3 to H3K27me2 and from H3K27me2 to H3K27me1, with H3K27me3 being the preferential substrate (Agger et al., 2007). The association between UTX and H3K4 methyltransferases MLL3/4 (Cho et al., 2007; Issaeva et al., 2007; Lee et al., 2007b) shows a mechanism for coupling the removal of repressive H3K27 methylation marks with the addition of active H3K4 methylation marks and transcriptional activation at enhancers (Agger et al., 2007; Lee et al., 2007b). The opposing actions of PRC2 and H3K27 demethylases, in combination with histone turnover, dictate the patterns of H3K27 methylation across the genome (Chory et al., 2019).

1.6.5 Role of H3K27me3

Although it is established that H3K27me3 is found at repressed genes, its exact role in transcriptional repression is not fully understood. In *Drosophila*, H3K27R and H3K27A mutations that abolish H3K27 methylation, result in derepression of PcG target genes and give rise to the same phenotype as *E(z)* mutations (McKay et al., 2015; Pengelly et al., 2013). In addition, in *Drosophila,* the R741K mutation in the SET domain of *E(z)* has been described (Bajusz et al., 2001), which results in enhanced ability of *E(z)* to trimethylate H3K27, and consequently to increased levels of H3K27me3 and decrease in H3K27me1/2 and H3K27ac. In this case, the increase in H3K27me3 results in inappropriate repression of PcG target genes and a phenotype that mimics the loss of function mutations in *trx* (Stepanik and Harte, 2012). These results suggest that H3K27me3 is central to PRC2-mediated gene silencing.

H3K27me3 is recognised by the CBX proteins in mammalian cells, and Pc in *Drosophila*, suggesting that it serves as a docking site for canonical PRC1 complexes. Indeed, reduction of H3K27me3 after knockdown and knockout of PRC2 components leads to reduction in PRC1 binding, dispersion of Polycomb bodies and transcriptional derepression in a subset of Polycomb target genes (Boyer et

al., 2006; Cao et al., 2002; Hernandez-Munoz et al., 2005). Consistent with that, knockdown of H3K27me2/3 demethylase UTX in HEK293 cells and in a human embryonic teratocarcinoma cell line leads to an increase in H3K27me3 and enrichment in PRC1 binding at *HOX* genes, which become repressed (Agger et al., 2007; Lee et al., 2007b). Based on these observations, it seems that, at least one of the functions of H3K27me3 is to repress gene expression through the recruitment of canonical PRC1 complexes. These complexes in turn lead to chromatin compaction and long-range interactions through the ability of the CBX and PHC subunits to polymerise (Figure 1.4), as will be discussed in section 1.6.9. In addition, it has been recently reported that the bromo adjacent homology (BAH) domain in BAHD1 and BAHCC1 proteins is a novel reader of the H3K27me3 mark. These proteins in turn contribute to gene repression through the recruitment of co-repressors including HDACs (Fan et al., 2021; Fan et al., 2020; Zhao et al., 2016)

However, it should be noted that the disruption of PRC2 complexes and the consequent decrease in H3K27me3 in mESCs is not accompanied by dramatic transcriptional changes compared to wild-type cells (Lavarone et al., 2019; Pasini et al., 2004; Riising et al., 2014). Nevertheless, the failure of mESCs with PRC2 KO to differentiate into embryoid bodies (EBs) potentially suggests that maintaining transcriptional repression by PRC2 is important for cells in the process of differentiating, rather than for cells with defined cellular identity and pluripotent cells (Chamberlain et al., 2008). In contrast, in mESCs, H2AK119ub is central to PcG-mediated gene repression (Fursova et al., 2019), as distinct from *Drosophila*, in which PcG repression seems to be more dependent on H3K27me3 rather than H2AK119ub (Pengelly et al., 2013; Pengelly et al., 2015).

1.6.6 Malignancies with alterations in global H3K27me3 levels

The role of H3K27me3 in the regulation of gene expression is further underlined by the identification of *EZH2* and oncohistone mutations that induce alterations in global H3K27me3 levels. More specifically, in a significant proportion of patients with follicular lymphoma (FL) and

diffuse large B-cell lymphoma (DLBCL) (Morin et al., 2010), and in a small subset of melanoma cases (Hodis et al., 2012), activating mutations in EZH2 have been identified. These mutations increase the ability of EZH2 to catalyse H3K27me3 and consequently lead to increased levels of this histone mark (Sneeringer et al., 2010; Yap et al., 2011). The first EZH2 mutation identified was a Tyrosine 641 substitution in the catalytic SET domain (Morin et al., 2010), and further activating mutations involving amino acids in the SET domain have been reported (Majer et al., 2012; McCabe et al., 2012a; Ott et al., 2014). Interestingly, ChIP-seq for H3K27me3 in melanoma and lymphoma cells with the $Ezh2^{Y641F}$ mutant protein showed that, despite the increase in H3K27me3, there are also loci where H3K27me3 is reduced (Souroullas et al., 2016). This suggests that the mutant EZH2 protein results in redistribution of H3K27me3 rather than a genome-wide increase in this histone mark.

In lymphomas, it is believed that gene repression due to increased levels of H3K27me3 contributes to lymphomagenesis through impairment of B cell differentiation (Beguelin et al., 2013) and downregulation of tumour suppressors including *PRDM1* and *SESTRIN1* (Caganova et al., 2013; Oricchio et al., 2017). Some studies support that the mutant Ezh2 protein alone is not sufficient to induce lymphoma in mouse models and that additional mutations are required (Beguelin et al., 2013; Berg et al., 2014), whereas in other reports expression of the mutant protein $Ezh2^{Y641F}$ from the endogenous locus was sufficient to induce B-cell lymphoma (Souroullas et al., 2016).

It has been suggested that treatment with small molecule inhibitors that abrogate the catalytic activity of EZH2 results in reversal of differentiation arrest (Beguelin et al., 2013) and impaired growth of lymphoma and melanoma cells *in vitro* and in xenograft models (Knutson et al., 2012; McCabe et al., 2012b; Qi et al., 2012; Souroullas et al., 2016; Tiffen et al., 2015). These inhibitors are currently in clinical trials for cancer patients including patients with non-Hodgkin lymphomas.

In addition, EZH2 overexpression has been reported in a number of solid malignancies including breast (Collett et al., 2006; Kleer et al., 2003), colorectal (Wang et al., 2010a), pancreatic (Chen et

al., 2014b; Ougolkov et al., 2008) and prostate cancer (Bachmann et al., 2006; Varambally et al., 2002), and high expression levels of EZH2 correlate with poor prognosis. Interestingly, it has been shown that degradation of EZH2 *in vivo* and in mouse xenografts of triple-negative breast cancer, rather than the inhibition of the EZH2 catalytic activity suppresses tumour growth, which suggests that EZH2 may have additional regulatory roles rather than functioning solely as a methyltransferase (Ma et al., 2020).

Apart from mutations in *EZH2*, mutations in genes encoding histones have been identified that cause alterations to the levels and distribution of H3K27me3. More specifically, H3K27M mutations have been identified in paediatric glioblastoma and in diffuse intrinsic pontine glioma. H3K27M mostly occurs in the histone H3.3 variant and to a lesser degree in the histone H3.1 variant. H3K27M oncohistones cause global reduction in H3K27me3, and it has been reported through *in vitro* experiments that the H3K27M peptide binds to the active site of EZH2 and inhibits its catalytic activity (Justin et al., 2016; Lewis et al., 2013). However, recent studies using H3.3 K27M-expressing neural stem cell cultures have suggested that the mutant H3.3 K27M histone is mainly incorporated into active enhancers and disrupts their activity, but is not sufficient to sequester PRC2 to enhancer elements; in contrast, its incorporation to a subset of PRC2-repressed promoters results in increased PRC2 binding and aberrant repression of a subset of neurodevelopmental genes (Brien et al., 2021).

Despite the fact that gain-of-function mutations of *EZH2* and histone variants have an oncogenic role in a subset of malignancies, in other forms of cancer, EZH2 functions as a tumour suppressor. For example, in myelodysplastic syndromes (MDS) and myeloproliferative neoplasms (MPN), *EZH2* deletions and frameshift and missense mutations are frequent and lead to a decrease in H3K27me3 levels and aberrant upregulation of PcG-repressed genes (Ernst et al., 2010; Nikoloski et al., 2010). Similarly, in a significant proportion of patients with T-cell acute lymphoblastic leukaemia (T-ALL), *EZH2* and *SUZ12* loss-of-function mutations have been identified (Ntziachristos et al., 2012), and in

patients with early T-cell precursor acute lymphoblastic leukaemia (ETP ALL), it has been shown that *EZH2* mutations lead to a decrease in H3K27me3 and derepression of PcG targets (Danis et al., 2016; Zhang et al., 2012). *EZH2* mutations in myeloid neoplasms frequently co-occur with *RUNX1* and *TET2* mutations, and mouse models have shown that *Ezh2* loss accelerates the development of MDS if combined with these mutations (Muto et al., 2013; Sashida et al., 2014).

Deregulation of H3K27me3 levels through mutations in *EZH2* and genes encoding histones emphasizes the significance of this histone modification in the regulation of gene expression, but it is still not clearly understood how H3K27me3 contributes to gene repression.

1.6.7 Recruitment of PcG proteins

In *Drosophila*, PcG proteins bind to promoter-like sequences, called Polycomb Response Elements (PREs) (Chan et al., 1994; Chiang et al., 1995; Schmitt et al., 2005; Simon et al., 1993). PREs contain motifs for DNA binding proteins, including Pleiohomeotic (Pho) (Fritsch et al., 1999), Pleiohomeotic-like (Phol) (Brown et al., 2003), GAGA factor (GAF) and Pipsqueak (Psq) (Hodgson et al., 2001; Strutt et al., 1997), as well as Zeste (Hur et al., 2002; Saurin et al., 2001), which in turn recruit PcG proteins. In mammalian cells PREs have not been identified (Bauer et al., 2016), although it has been reported that unmethylated CpG islands correspond to PREs and that several DNA binding TFs can recruit PcG complexes to their target sites, as will be discussed below. Other mechanisms that have been proposed for PcG recruitment in mammalian cells are through long non-coding RNAs and through their binding to unmethylated CpG islands, discussed in more detail below.

1.6.7.1 Recruitment of PcG proteins by transcription factors

The most robust evidence for TF-mediated recruitment of PcG proteins in mammalian cells comes from studies showing that the ncPRC1.6 complex is recruited to chromatin via its association with

TF heterodimers E2F6-TFDP1 and MGA-MAX. These TFs mediate the localisation of the complex in mESCs to targets enriched in meiosis- and germ cell-related genes. In turn, the catalytic subunit of PRC1.6 catalyses H2AK119ub which leads to their repression (Endoh et al., 2017; Scelfo et al., 2019; Stielow et al., 2018; Trojer et al., 2011).

Several other TFs have been reported to recruit PcG complexes to their target sites in mammalian cells, including YY1, REST, Zfp277 and RUNX1. More specifically, it has been indicated that YY1 is the homolog of Pho and Phol (Brown et al., 2003; Brown et al., 1998), which are two of the most well studied TFs driving recruitment of PcG proteins to PREs in *Drosophila*. However, in mESCs biochemical and genome-wide ChIP-seq studies showed that, at least in that cell type, YY1 does not seem to play a role in the recruitment of PcG proteins (Vella et al., 2012). Zfp277 is a zinc finger domain-containing protein, and through co-immunoprecipitation experiments it has been shown to interact with Bmi1 in mouse embryonic fibroblasts (MEFs). Upon its knockout, Bmi1 binding at the *Ink4a/Arf* locus is reduced and transcriptional derepression is observed (Negishi et al., 2010). Similarly, co-immunoprecipitation and knockdown experiments indicate that the REST TF physically and functionally interacts with the canonical PRC1 complex and with the PRC2 complex in human cell lines and in mESCs. In addition, a number of papers have indicated interaction between RUNX1 and its co-factor CBFb with the PRC1 components RING1B and Bmi1 in megakaryocytes and murine thymocytes, and in human cell lines bearing the chromosome 16 inversion, which gives rise to the CBFb-SMMHC fusion protein (Cordonnier et al., 2020; Pulikkan et al., 2018; Yu et al., 2012).

It should be noted that for most of these TFs, it is reported that they interact with Bmi1 and other components of the cPRC1 complex. Given that the cPRC1 complexes recognise H3K27me3, it cannot be excluded that their co-localisation with TFs is not due to physical interaction, but rather an indirect consequence of transcriptional repression. More specifically, TF-mediated transcriptional repression could result in a block in transcription and in the indirect recruitment of PRC2 to repressed genes (Riising et al., 2014), as discussed below, followed by deposition of

H3K27me3 and stabilisation of the binding of cPRC1 subunits including Bmi1. As a result, much of the published literature does not allow us to discern whether cPRC1 components are recruited to chromatin through direct interaction with TFs or their binding is stabilised due to gene repression mediated by TFs that function as repressors.

1.6.7.2 Recruitment of PcG proteins by long non-coding RNAs

Another proposed mechanism for PRC2 binding has been through interaction with long non-coding RNAs (lncRNAs), with the most well-known examples being the recruitment of PRC2 by *Xist* RNA during X chromosome inactivation (Erhardt et al., 2003; Mak et al., 2002; Plath et al., 2003; Silva et al., 2003), and by *HOTAIR* lncRNA (HOX antisense intergenic RNA). *HOTAIR* is transcribed from the *HOXC* locus, and has been reported to drive the recruitment of PRC2 *in trans* to the *HOXD* locus, resulting in its silencing (Rinn et al., 2007). However, recent evidence has challenged the notion that PRC2 is directly recruited by these lncRNAs. More specifically, *HOTAIR*-mediated repression appears to be independent of PRC2, whose binding is a consequence of gene silencing (Portoso et al., 2017). In addition, in the case of *Xist*, super-resolution microscopy experiments have demonstrated that *Xist* and PRC2 do not co-localise on the inactive X chromosome (Cerase et al., 2014). In contrast, it seems that the first step towards Polycomb recruitment during X chromosome inactivation is the interaction between *Xist* and PRC1 sub-complexes PRC1.3/PRC1.5, mediated by RNA binding protein hnRNPK (heterogeneous nuclear ribonucleoprotein K) (Almeida et al., 2017; Pintacuda et al., 2017). This results in deposition of H2AK119ub, and a cascade of enrichment of PcG proteins and transcriptional repression through the recruitment of the other ncPRC1 sub-complexes and the PRC2 complex that further deposits H3K27me3.

1.6.7.3 Recruitment of PcG proteins to CpG islands and responsive model of PcG binding

CpG islands (CGIs) are defined as DNA sequences approximately 1 kb in length that contain a high frequency of conserved CpG sites which tend to be unmethylated (Deaton and Bird, 2011). In contrast, the vast majority of CpG dinucleotides in mammalian genomes are methylated at the C5 position of the cytosine ring and their representation across the genome is low due to the mutagenic properties of methylcytosines, which get deaminated and converted to thymine (Bird, 1980). The enzymes that are responsible for catalysing methylation of cytosine residues are the DNA methyltransferases DNMT1, DNMT3A and DNMT3B (Lyko, 2018). Half of the CGIs associate with annotated gene promoters, while the other half of CGIs are localised away from annotated gene promoters, at inter- and intragenic sites, and are termed “orphan CGIs”. Orphan CGIs often produce transcripts, and are more likely to be methylated compared to CGIs associated with TSSs (Illingworth et al., 2010).

Binding of PcG proteins (and a subset of proteins of the MLL family) to unmethylated CGIs is considered to be the main mechanism for their recruitment to chromatin, and it has been suggested that CpG islands are the equivalent of PREs in mammalian cells. PRC2 binds to unmethylated CpG islands through the PCL proteins (Li et al., 2017), and PRC1 associates with unmethylated CpG islands through the interaction of the CXXC domain of KDM2B with unmethylated CpG dinucleotides (Farcas et al., 2012; He et al., 2013; Wu et al., 2013b). In fact, more than 10 different proteins have been shown to include a ZF-CXXC domain that directs them to CpG islands. Among these proteins are, in addition to KDM2B, KDM2A; the histone methyltransferases MLL, MLL2, CFP1; and also DNMT1 which catalyses methylation of DNA; and the TET proteins TET1 and TET3 that hydroxylate methylated cytosines (Long et al., 2013). The affinity of interaction between these proteins and CpG DNA is low (Kd in the range of a few μM) (Xu et al., 2018), a lot weaker compared to the affinity of protein-protein interactions and the interaction between DNA and TFs. For example, the binding affinity of the MLL CXXC domain to CpGs has been estimated to be 4.3 μM (Allen et al., 2006), while the affinity of MLL interaction with

its interaction partner Menin is a lot stronger (estimated to be 10 nM) (Grembecka et al., 2010). Potentially this demonstrates that for stable binding of these proteins to their target loci interactions with other factors is required, or that at large CpG islands the sum of low affinity interactions results in high affinity, a phenomenon known as multivalency (Rothbart et al., 2013; Ruthenburg et al., 2011).

Research in the Bernstein lab showed that insertion of a CpG island in a gene-poor region in mESCs was sufficient for PRC2 recruitment and deposition of H3K27me3 (Mendenhall et al., 2010). In this study, they identified a CpG island that overlaps with the promoters of two ubiquitously expressed genes. At its endogenous locus this CpG island is marked by H3K4me3 and there is no evidence of Polycomb binding. One half of the CpG island has TF activating motifs, while the other half does not. In elegant experiments they also showed that tethering the fragment of the CpG island that lacked activating motifs in a gene-poor region was sufficient to recruit PRC2, whereas tethering the CpG island fragment encompassing the activating motifs led to deposition of H3K4me3. These results suggest that in the absence of transcription and activating TF motifs, a GC-rich sequence can be bound by PRC2.

In agreement with these observations, it has been proposed that the binding of PcG proteins to target loci in mammalian cells is mainly responsive to the transcriptional status of target genes (in addition to recruitment mediated by TFs or ncRNAs that subsequently causes repression of target genes). More specifically, it has been suggested that PcG proteins sample chromatin and their binding is stabilised at loci with repressive histone marks where no active transcription is taking place (Klose et al., 2013). In support of this, experiments in mESCs investigating the effect of transcription on the recruitment of PcG proteins showed that transcriptional inhibition at individual loci is sufficient to cause PRC2 binding and deposition of H3K27me3 (Hosogane et al., 2016). Similarly, blocking the function of RNAPII with chemical inhibitors leads to PRC2 binding to thousands of CpG islands genome-wide (Riising et al., 2014).

Moreover, consistent with the responsive model for Polycomb binding, it has been demonstrated that PRC2 promiscuously binds to nascent RNA and this interaction precludes PRC2 binding to actively transcribed genes. More specifically, it has been shown that the association between PRC2 and RNA or nucleosomes is mutually antagonistic, and RNA degradation results in PRC2 binding at active genes enriched in CpG islands (Beltran et al., 2016). Thus, it seems that exclusion of PRC2 binding at actively transcribed genes by nascent RNA is one of the mechanisms through which PRC2 can sense transcriptional status. Similarly, experimental data in *Drosophila* have shown that recruitment of PcG proteins to PREs is inhibited by the presence of the TrxG proteins and active transcription (Cavalli and Paro, 1998, Klymenko and Müller, 2004, Schmitt et al., 2005). In addition, work *in vitro* has highlighted that a PRC2 interaction with histone H3 is inhibited by the presence of H3K4me3 (Schmitges et al., 2011) and PRC2 catalytic activity is repressed by the active histone marks H3K4me3 and H3K36me2/3 (Schmitges et al., 2011; Yuan et al., 2011). Recent structural studies have demonstrated that unmodified H3K36 promotes threading of the N-terminus of H3 into the active site of EZH2, providing mechanistic insight into the inhibition of PRC2 activity by H3K36me3 at active genes (Finogenova et al., 2020).

Based on these observations we can conclude that upon binding of PcG proteins to silenced loci with repressive chromatin marks, and in the absence of antagonistic activating marks and activators, they maintain transcriptional silencing of target genes. In that sense, it seems that PcG proteins can sense the absence of transcription, and function to maintain transcriptional repression rather than to initiate gene silencing. This is consistent with the early studies in *Drosophila* that demonstrated that PcG proteins are not required for initiation of silencing of *Hox* genes, but for maintenance of their transcriptional status in later stages of embryonic development (Jones and Gelbart, 1990, Simon et al., 1992, Struhl and Akam, 1985). Overall, in mammalian cells PcG proteins can be recruited to target loci by TFs and ncRNAs (instructive model for PcG recruitment), and also, through chromatin sampling they can establish repressive Polycomb domains in the absence of transcription and activating signals (Klose et al., 2013).

1.6.8 Polycomb repression

Despite the extensive studies of PcG proteins, the mechanistic details of how PcG complexes maintain repression of target genes is still not fully understood. PcG proteins have been proposed to negatively affect every step in the cascade of events leading to transcription, i.e. to decrease accessibility (Fitzgerald and Bender, 2001; Zink and Paro, 1995), inhibit the assembly of the pre-initiation complex, impede transcriptional initiation (Dellino et al., 2004; Nakagawa et al., 2008) or inhibit the activity of RNAPII post initiation (Brookes et al., 2012; Brookes and Pombo, 2009; Stock et al., 2007).

More specifically, it has been shown that Ezh1 can compact chromatin *in vitro* and that *in vivo* it causes reduced nuclease accessibility in a Gal4-luciferase reporter assay (Margueron et al., 2008). Similarly, electron microscopy studies have shown that RING1B can compact nucleosome arrays *in vitro* and further *in vivo* studies demonstrated that it mediates chromatin compaction independent of its catalytic activity (Eskeland et al., 2010). In addition, previous studies have shown that the PRC2-mediated H3K27me3 mark is central to PcG repression (Pengelly et al., 2013). Furthermore, work in *Drosophila* has suggested that chromatin compaction mediated by canonical PRC1 components Pc and Ph is required for gene silencing. More specifically, work in the Cavalli lab, aiming to investigate the order of chromatin compaction mediated by canonical PRC1 complexes, as will be discussed in the next session, and gene silencing, showed that in *Pc* and *Ph* mutants in *Drosophila*, chromatin decompaction is observed, followed by derepression of *Hox* genes (Cheutin and Cavalli, 2018). These results, in combination with observations that link chromatin compaction mediated by cPRC1 components with gene silencing (Francis et al., 2004; Lau et al., 2017) suggest that the higher-order chromatin folding and chromatin compaction by cPRC1 is key to PcG repression.

However, it should be noted that despite these studies showing the central role of H3K27me3 and cPRC1 components in PcG repression during *Drosophila* and mouse development (Lau et al., 2017;

Pengelly et al., 2013; Pengelly et al., 2015), in other model systems it has been shown that reduction in H3K27me3 or removal of cPRC1 components does not significantly affect PcG repression (Lavarone et al., 2019), and that vPRC1 components have a more important role in mediating gene silencing. More specifically, it has been proposed that the deposition of H2AK119ub by vPRC1 is central to PcG repression, at least in mESCs (Fursova et al., 2019; Tamburri et al., 2020). The exact mechanisms of how H2AK119ub contributes to gene silencing are also unknown, but there are data showing that H2AK119ub impedes the deposition of transcription-associated histone modifications (Fursova et al., 2021; Nakagawa et al., 2008) and that it inhibits transcriptional initiation (Nakagawa et al., 2008), potentially by affecting the frequency of transcription bursts (Dobrinic et al., 2021).

Despite the fact that these and other studies have shown that PcG binding leads to chromatin compaction and reduced accessibility (Fitzgerald and Bender, 2001), other studies have shown that the presence of PcG proteins does not impede the binding of TFs and the transcriptional machinery (Dellino et al., 2004; McCall and Bender, 1996) and that loss of PcG binding does not affect chromatin accessibility (King et al., 2018). For example, after artificial insertion of a PRE upstream of the *hsp26* promoter in *Drosophila,* it was shown that upon PcG binding and PcG-mediated repression, RNAPII and TBP were still able to bind, but the formation of the transcription bubble (Holstege et al. 1997; Revyakin et al. 2004) during the step of transcriptional initiation was impaired (Dellino et al., 2004). These observations are in agreement with studies in mammalian cells, as it has been shown that bivalent genes in mESCs that are bound by PcG proteins are also bound by RNAPII, but these genes are not transcribed or are transcribed at very low levels (Brookes et al., 2012; Stock et al., 2007). Given that the binding of PcG proteins is dynamic rather than stable (Ficz et al., 2005; Huseyin and Klose, 2021; Vandenbunder et al., 2014), steric hindrance of RNA polymerase binding and binding of co-activators does not seem to be the main mechanism of repression. This suggests that PcG proteins and their repressive histone marks affect other stages of transcription. As mentioned above, H2AK119ub has been reported to inhibit transcriptional initiation (Dobrinic et al., 2021; Nakagawa et al., 2008), while other studies at bivalent domains

have suggested that the deposition of H2AK119ub impairs transcription after the transcriptional initiation step. More specifically, it has been reported that PcG-mediated deposition of H2AK119ub is associated with a different conformation status of RNAPII. In particular, RNAPII is found in a poised state, recognised by antibodies specific for Ser5 phosphorylation of the CTD, but not with antibodies that recognise the Ser2 phosphorylation of the CTD, that is associated with transcriptional elongation. In those studies, reduction of H2AK119ub after knockout of *Ring1a/b* resulted in derepression of bivalent genes (Brookes et al., 2012; Stock et al., 2007).

Overall, despite the fact that the mechanisms of Polycomb-mediated gene repression are not completely understood, it seems that in some contexts the activity of PRC1 is central to Polycomb silencing through the deposition of H2AK119ub, while in other contexts, for example for the maintenance of the repressed state of *HOX* genes in *Drosophila* melanogaster, Polycomb repression is dependent on H3K27me3.

1.6.9 PRC1 and 3D nuclear structure

PcG proteins alter the chromatin structure by affecting local chromatin compaction and long-range interactions (Denholtz et al., 2013; Eagen et al., 2017; Grau et al., 2011; Kundu et al., 2017; Lau et al., 2017; Pachano et al., 2019; Plys et al., 2019). In mESCs engineered to encompass an array of lac operator repeats, tethering EZH2 fused to the *Escherichia coli* Lac repressor results in the generation of a Polycomb domain, and to novel chromatin interactions with other PcG-repressed loci (Wijchers et al., 2016). The PcG proteins that seem to mediate chromatin compaction and long-range inter- and intra-chromosomal interactions between PcG-bound loci are the CBX and PHC proteins, components of canonical PRC1 complexes (Grau et al., 2011; Kundu et al., 2017; Lau et al., 2017) (Figure 1.5). Deletion of PRC2 subunits also reduces long-range chromatin interactions (Denholtz et al., 2013; Hernandez-Munoz et al., 2005; Wijchers et al., 2016), but the effects of PRC2

are potentially indirect, due to reduction of H3K27me3 and subsequent loss of canonical PRC1 binding.

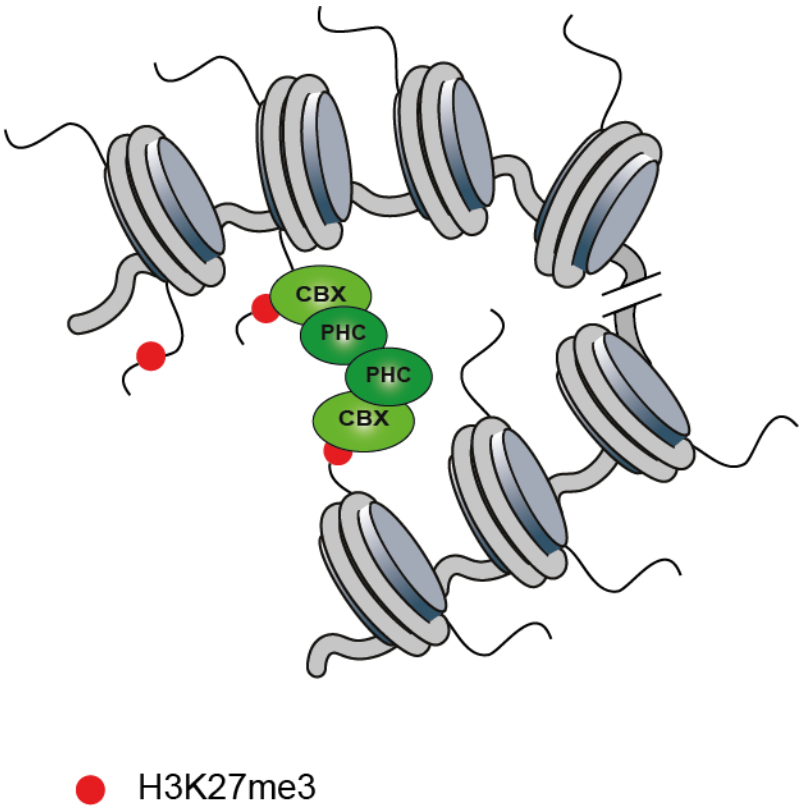


Figure 1.5: Schematic of long-range Polycomb interactions mediated by canonical PRC1 complexes. Canonical PRC1 complexes are bound to PcG-repressed loci through the association of the CBX proteins to H3K27me3 mark. The polymerisation of the sterile alpha motif (SAM) domains of the PHC proteins of canonical PRC1 complexes results in long-range interactions between PcG-bound loci.

More specifically, the CBX2 protein of canonical PRC1 complexes has the ability to mediate local chromatin compaction *in vitro* (Grau et al., 2011), and to form liquid–liquid phase separated structures in solution (Plys et al., 2019). CBX2 includes a central, positively charged disordered region, which is not shared with the rest of the CBX proteins. This domain, named as the low-complexity disordered region (LCDR) (Plys et al., 2019) is reported to be responsible for chromatin compaction and formation of phase-separated condensates (Grau et al., 2011; Kim and Kingston, 2020; Plys et al., 2019). Mutation of lysine and arginine residues to alanine in this region results in disruption of phase separation *in vitro* and to impaired formation of PRC1 condensates *in vivo* (Plys

et al., 2019), which appear as nuclear puncta in imaging studies and are known as Polycomb bodies (Saurin et al., 1998). In addition, these mutations in the positively charged residues impair transcriptional repression of PcG target genes in mESCs (Lau et al., 2017), and mice homozygous for the mutant proteins exhibit anterior-to-posterior transformations in the axial skeleton (Lau et al., 2017), similar to the axial patterning defects observed in *Cbx2* knockout mice (Core et al., 1997).

Despite the fact that the other CBX proteins do not share the compaction domain of CBX2, proteins CBX4, CBX6 and CBX8 also have a high net positive charge and disordered domains (Grau et al., 2011; Kim and Kingston, 2020), and their ability to form puncta in cells and repress PcG targets has been demonstrated. This suggests that the other CBX proteins could also have the potential to compact chromatin and form phase separated structures (Kim and Kingston, 2020). In contrast, CBX7 has a much shorter disordered domain, is not as positively charged compared to the other CBX proteins (Grau et al., 2011; Kim and Kingston, 2020) and cannot compact nucleosome arrays *in vitro* (Grau et al., 2011; Jaensch et al., 2021). Cbx7 is the main Cbx protein expressed in mESCs where it represses the expression of the other Cbx proteins and contributes to the maintenance of pluripotency. In contrast, during differentiation, the *Cbx7* gene becomes PcG-repressed, and Cbx7 is replaced by the other Cbx proteins (Kloet et al., 2016; Morey et al., 2012; O'Loghlen et al., 2012). Similarly to mESCs, CBX7 is highly expressed in haematopoietic stem cells (HSCs) and is downregulated upon differentiation (Klauke et al., 2013). The structural differences between CBX7 and the other CBX proteins could contribute to the difference in chromatin compaction in differentiated cells compared to undifferentiated stem cells and progenitors, in which Polycomb sites reside in the active 'A' compartment (Vieux-Rochas et al., 2015) and can be accessed by TFs and the transcriptional machinery. In fact, ectopic introduction of the Cbx2 disordered domain that is associated with phase separation in Cbx7 in mESCs leads to chromatin compaction and impaired differentiation (Jaensch et al., 2021).

In addition to CBX proteins, it has been shown that cPRC1 components PHC1 and PHC2, homologs of *Drosophila* Ph protein, contribute to the formation of long-range interactions between PcG-bound loci (Isono et al., 2013; Kundu et al., 2017; Wani et al., 2016). These interactions are mediated through their sterile alpha motif (SAM) domains which have been shown to have the ability of head to tail polymerization *in vitro* (Isono et al., 2013; Kim et al., 2002) (Figure 1.5). Mutations in the SAM domain of Ph protein disrupts the formation of PcG clusters (Wani et al., 2016).

The formation of Polycomb long-range interactions between PcG-repressed promoters and poised enhancers, through the mechanisms described above, has been proposed to lead to the establishment of a regulatory topology that is required for future gene activation. More specifically, the binding of PcG proteins results in physical proximity between poised enhancers and promoters of target genes, which is required during later stages of development, when the appropriate activation signals lead to the eviction of PcG proteins and to induction of gene expression (Cruz-Molina et al., 2017; Gentile et al., 2019; Kondo et al., 2014).

Interestingly, apart from their binding at repressed loci, PcG complexes have been described to be bound at active sites as well, both in *Drosophila* and in mammalian cells, where they can even facilitate transcription (Cohen et al., 2018; Cohen et al., 2019; Frangini et al., 2013; Gao et al., 2014; Loubiere et al., 2020; Schaaf et al., 2013; Zhang et al., 2021). Although the mechanisms of their recruitment at active sites have not been fully delineated, it has recently been suggested that the presence of cPRC1 components at active sites promotes the formation of loops and the establishment of chromatin interactions between active loci, including between active promoters and enhancers (Loubiere et al., 2020; Zhang et al., 2021). More specifically, it has been proposed that PcG proteins contribute to the establishment of a regulatory topology that can be further "exploited" by TFs, which in turn take advantage of this PRC1-dependent enhancer-promoter physical proximity to control the expression of target genes (Loubiere et al., 2020). These

observations further suggest that PcG complexes probably have a far broader regulatory role rather than functioning just as repressors.

1.7 MLL proteins

The mammalian family of MLL or COMPASS proteins contains six proteins, MLL1/2, MLL3/4 and SETD1A/B, that are homologs of the *Drosophila* proteins Trithorax (trx), Trithorax-related (trr) and *Drosophila* Set1 (dSet1) respectively (Crump and Milne, 2019) (Figure 1.6). All proteins of the family share a SET domain, responsible for H3K4 methylation (Milne et al., 2002), and knockout of each of the genes encoding the proteins of the MLL family in mice results in embryonic lethality (Bledau et al., 2014; Glaser et al., 2006; Lee et al., 2013; Yu et al., 1995).

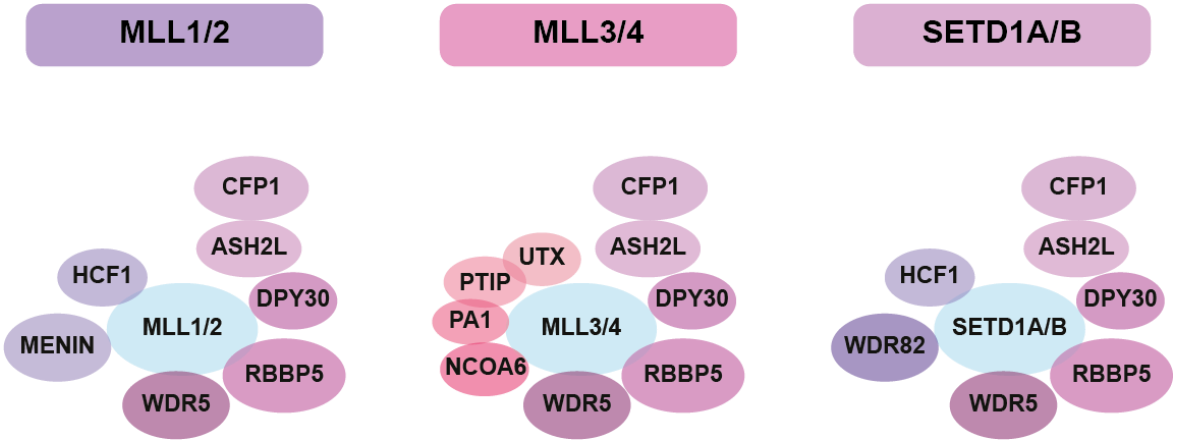


Figure 1.6: Schematic of the MLL proteins. *(Left panel)*: Schematic of MLL1/2 complexes, *(middle panel)*: Schematic of MLL3/4 complexes and *(right panel)*: Schematic of SETD1A/B complexes.

The first protein of the family, MLL or MLL1 (also known as KMT2A) is encoded by the *MLL* gene on the long arm of chromosome 11 (11q23). This was initially identified after cloning the breakpoint of chromosomal translocations involving *MLL* in a subset of very aggressive leukaemias (Gu et al., 1992; Tkachuk et al., 1992). To date, translocations have been described that fuse *MLL* in frame

with over 120 translocation partners (Meyer et al., 2018). *MLL* is a homolog of the *Drosophila* gene *trithorax* (*trx*), which encodes a protein involved in the regulation of the expression of *Homeotic* (*Hox*) genes (Gu et al., 1992; Tkachuk et al., 1992). Similarly to *Drosophila*, *Mll* knockout in mice results in embryonic lethality with skeletal defects due to deregulation of *Hox* genes (Yu et al., 1995). It was further shown that the SET domain of MLL is responsible for catalysing methylation of H3K4 to a subset of gene promoters including the promoters of *HOX* genes (Milne et al., 2002). The recruitment of MLL is mediated through the association of its CXXC domain with a subset of unmethylated CpG islands (Milne et al., 2010), as well as through the association of its AT hooks with AT-rich DNA (Zeleznik-Le et al., 1994), and also the interaction of its third PHD finger (PHD3) with H3K4me2/3 (Milne et al., 2010; Wang et al., 2010c). In addition, the association between the MLL CXXC domain with the RNA polymerase II-associated factor 1 (PAF1) elongation complex seems to have a role in recruiting MLL to active targets (Milne et al., 2010; Muntean et al., 2010).

As far as Mll2 is concerned, *Mll2* knockout in mice results in early embryonic lethality, without any obvious skeletal defects. *Mll2* knockout embryos display a general growth retardation, potentially due to widespread apoptosis (Glaser et al., 2006). *Mll2* knockdown in mESCs results in reduction of H3K4me3 at the promoters of bivalent genes, which represent a small subset of genes (Denissov et al., 2014; Hu et al., 2013b). In contrast, in oocytes, MLL2 is the main methyltransferase for H3K4me2/3, and loss of MLL2 expression results in global reduction of these histone modifications (Andreu-Vieyra et al., 2010).

MLL3 and MLL4 have been identified as the enzymes responsible for catalysing H3K4me1 in mammalian cells, similarly to their homolog trr in *Drosophila* (Herz et al., 2012; Hu et al., 2013a). *Mll3* knockout mice die around birth with no obvious morphological defects, and *Mll4* knockout mice die around day E9.5 (Lee et al., 2013). Unlike the other mammalian MLL proteins, which bind to unmethylated DNA through a CXXC domain (MLL1, MLL2 and MLL fusion proteins (Allen et al., 2006; Ayton et al., 2004; Birke et al., 2002)), or through their interaction with the CXXC domain-

containing protein CFP1 (SETD1A and SETD1B) (Lee and Skalnik, 2005; Lee et al., 2007a; Thomson et al., 2010), MLL3 and MLL4 lack such a domain. This explains why these proteins and H3K4me1 are not found at CpG islands at gene promoters, as the other proteins of the MLL family, but instead at enhancers (Hu et al., 2013a). This further suggests that other mechanisms are responsible for the recruitment of MLL3/4 to chromatin, which remain to be investigated, so as to examine whether the MLL3/4-mediated deposition of H3K4me1 is the first step towards generating enhancers. One proposed mechanism is through the interaction of MLL3/4 with TFs, including HOXA9 and CEBPa and the pioneer factor FOXA1 (Serandour et al., 2011; Sun et al., 2018), although other studies have shown that the presence of H3K4 methylation precedes FOXA1 binding and chromatin remodelling (Lupien et al., 2008). Recent studies have reported that the disruption of the catalytic activity of MLL3/4, followed by reduction in H3K4me1, did not result in significant changes in gene expression, in contrast to double knockout of *MLL3/4* which resulted in transcriptional changes. This suggests that the presence of MLL3/4 complexes may be required for enhancer function independent of their catalytic activity (Dorighi et al., 2017)**.**

Lastly, SETD1A and SETD1B are homologs to the Set1 protein in *S. cerevisiae* and to the *Drosophila* protein dSet1, which is the major H3K4me3 methyltransferase in *Drosophila* (Ardehali et al., 2011; Hallson et al., 2012). *Setd1a* and *Setd1b* knockout in mice results in embryonic lethality, similarly to knockouts of the other proteins of the MLL family. Consistent with observations in *Drosophila,* deletion of *Setd1a* in mESCs results in global decrease in H3K4me1/2/3 but not in complete loss of these histone modifications. In contrast, no reduction in H3K4 methylation is seen after *Setd1b* deletion in mESCs (Bledau et al., 2014). These observations on the role of MLL proteins in catalysing H3K4 methylation suggest that there is some redundancy among the different members of the MLL family, but that each protein also seems to have distinct roles.

1.8 MLL-rearranged leukaemias

As mentioned above, the *MLL* gene is often a target of mutations and translocations. *MLL* translocations fuse the N-terminal part of the *MLL* gene in-frame to one of more than 120 different gene partners, resulting in the generation of the MLL fusion proteins (Meyer et al., 2018). The vast majority of the MLL fusion proteins result from the fusion of the *MLL* gene to 6 other genes (*AF4, AF9, ENL, ELL, AF10* or *AF6*) (Meyer et al., 2009). Among them, the MLL-AF4 fusion protein, generated from the *t(4;11)* translocation, is a major cause of infant acute lymphoblastic leukaemia (ALL), and is associated with a very poor prognosis (Sanjuan-Pla et al., 2015).

In contrast to other forms of leukaemia, MLL-rearranged (MLL-r) leukaemias are characterised by a low frequency of additional mutations (on average 1.3 non-silent mutations for infant MLL-r ALL). Mutations in infant ALL are mostly activating mutations of the *RAS* pathway, but they are usually subclonal and are often lost in relapse samples (Andersson et al., 2015). Even in adult AML patients, that tend to carry a heavier mutational burden, MLL-r leukaemias stand out as having a particularly low mutation rate and no recurring cooperating mutations (Cancer Genome Atlas Research et al., 2013). Together, this suggests that the MLL fusion proteins function as very potent oncogenes that can drive leukaemogenesis in the absence of additional mutations.

The molecular mechanisms by which MLL-AF4 and the other MLL fusion proteins drive leukaemogenesis have been studied extensively. These studies show that they activate downstream target genes through the recruitment of transcriptional elongation complexes and chromatin proteins, including the H3K79 methyltransferase DOT1L (Slany, 2020; Takahashi and Yokoyama, 2020) (Figure 1.7). More specifically, MLL-AF4 retains the N-terminal part of MLL that includes AT hooks and the CXXC domain, while the C-terminal domain that encompasses the SET methyltransferase domain is lost. The CXXC domain of the fusion proteins can associate with unmethylated CpG islands (Milne et al., 2010; Okuda et al., 2014), and has also been shown to interact with the PAF1 complex (Milne et al., 2010; Muntean et al., 2010) (Figure 1.7). In addition,

the association of the N-terminal part of MLL with Menin is retained in MLL fusion proteins (Yokoyama et al., 2005) (Figure 1.7). Menin in turn interacts with LEDGF (Yokoyama and Cleary, 2008), which has been shown to associate with the active marks H3K36me2/3 at active loci via its PWWP domain (Okuda et al., 2014; Pradeepa et al., 2012; Wang et al., 2020).

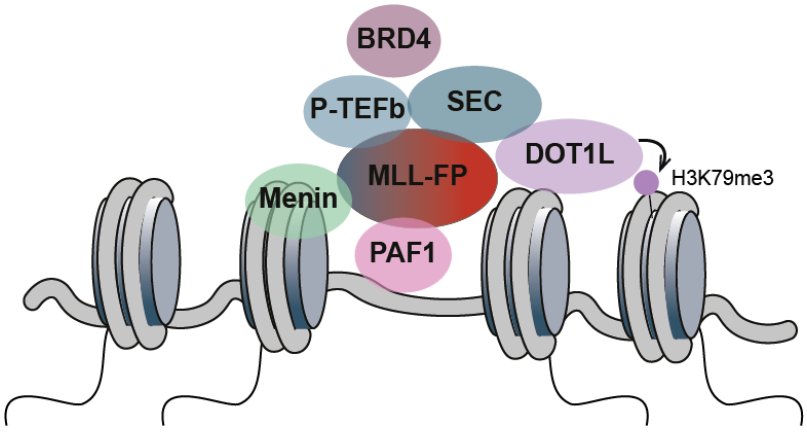


Figure 1.7: Schematic of the MLL fusion protein complexes bound to a target gene. MLL fusion proteins (MLL-FPs) bound to target loci recruit DOT1L that deposits H3K79 methyl marks , the super elongation complex (SEC), P-TEFb, PAF1 and BRD4, promoting transcriptional elongation.

Upon binding of the fusion proteins at their target sites, they cause transcriptional activation of the corresponding genes. This is in part due to the recruitment of P-TEFb (Yokoyama et al., 2010) (Figure 1.7), which phosphorylates the Ser2 of the CTD of RNAPII and promotes transcriptional elongation (Marshall and Price, 1995). Pharmacological inhibition of CDK9, a component of P-TEFb, has been shown to reduce the proliferation of MLL-r leukaemic cells (Mueller et al., 2009). In addition, the most common fusion partners of MLL are members of the Super elongation complex (SEC) (Biswas et al., 2011; Lin et al., 2010; Yokoyama et al., 2010), suggesting that the aberrant activation of target genes of the MLL fusion proteins is through the recruitment of complexes involved in transcriptional elongation. Furthermore, inhibition of BRD4 binding, which has been shown to recruit P-TEFb (Yang et al., 2005) and interact with the SEC and PAF complexes (Dawson et al., 2011), has also been demonstrated to inhibit the growth of leukaemias with MLL

rearrangements in *in vitro* and *in vivo* models (Dawson et al., 2011). In addition, the MLL fusion proteins recruit the H3K79 methyltransferase DOT1L, which has been shown to interact with AF9, ENL and AF10 (Biswas et al., 2011; Monroe et al., 2011; Okada et al., 2005; Yokoyama et al., 2010), that are frequent MLL fusion partners, and indeed, targets of the MLL fusion proteins have increased levels of H3K79 methylation (Bernt et al., 2011; Krivtsov et al., 2008) (Figure 1.7).

Overall, the MLL fusion proteins activate target genes through the recruitment of co-activators and complexes associated with transcriptional elongation, leading to transcriptional deregulation. Among the direct downstream targets of MLL fusion proteins is the *RUNX1* gene (Prange et al., 2017; Wilkinson et al., 2013), a key gene for the development of the haematopoietic system, and also a very frequent target of genetic alteration in human leukaemias.

1.9 RUNX1

1.9.1 RUNX1 and the development of the haematopoietic system

RUNX1 (Runt-related transcription factor 1) is a sequence-specific DNA binding protein that belongs to the Runt-related transcription factor family, along with transcription factors RUNX2 and RUNX3 in mammalian cells, mainly involved in neurogenesis and osteogenesis respectively (Otto et al, 1997, Inoue et al, 2002). These proteins are homologs of the *Drosophila* protein runt, which is encoded by a primary pair-rule gene and lies at the top of the hierarchy of proteins regulating fly body segmentation. The name runt was given due to the smaller size of *runt* mutant flies compared to the wild-type ones, because of lack of body segments (Gergen and Wieschaus, 1985; Nusslein-Volhard and Wieschaus, 1980). Besides *runt*, the Runx gene family in *Drosophila* includes genes *RunxA, RunxB* and *lozenge* (Rennert et al., 2003). The *Drosophila* and the mammalian proteins share a highly conserved 128 aa region near the N-terminus, the runt domain, which contains the residues that are responsible for DNA binding. RUNX1 and the other members of the RUNX family form a

heterodimer with the ubiquitously expressed protein Core-binding factor subunit Beta (CBFb). CBFb does not bind to DNA, but interacts with the runt domain and enhances the DNA binding affinity of RUNX1, and also protects it from proteasome-mediated degradation (Huang et al., 2001; Wang et al., 1993; Wang and Speck, 1992).

RUNX1 is considered to be a master regulator of the development of the haematopoietic system. *Runx1* knockouts result in embryonic lethality by E12.5 due to haemorrhage in the central nervous system (Okuda et al., 1996; Wang et al., 1996a). The co-factor of Runx1, Cbfb, is also essential for haematopoietic development, as its knockout results in embryonic lethality with the same phenotype as the *Runx1* knockout (Sasaki et al., 1996; Wang et al., 1996b). During embryogenesis, HSCs arise from specialised endothelium, termed the haemogenic endothelium (HE) via a process known as endothelial-to-hematopoietic transition (EHT). More specifically, RUNX1-expressing endothelial cells lining the dorsal aorta in the aorta-gonad-mesonephros (AGM) region and other blood vessels give rise to HSC and progenitor cells which migrate to the foetal liver (Jaffredo et al., 1998; Lancrin et al., 2009; Mukouyama et al., 2000). RUNX1 is crucial for EHT (Chen et al., 2009; Lancrin et al., 2009; North et al., 1999), which requires the suppression of the endothelial transcriptional program and the initiation of the haematopoietic transcriptional program (Swiers et al., 2013). More specifically, the correct dosage of Runx1 is required for EHT (Lie et al., 2020), as Runx1 overexpression also blocks the emergence of haematopoietic cells in human and mESC differentiation systems (Chen et al., 2017; Lie et al., 2018). Further studies have suggested that the TFs GFI1 and GFI1B are downstream targets of RUNX1, and their expression in the absence of RUNX1 leads to morphological changes characteristic of EHT, while their deficiency maintains the expression of endothelial genes (Lancrin et al., 2012), suggesting that RUNX1, at least in part, functions by regulating their expression.

In contrast to RUNX1 being required for the generation of definitive HSCs during foetal life, it is not required for the maintenance of HSCs in adult haematopoiesis, but it is indispensable for the

terminal differentiation of haematopoietic progenitors of megakaryocytic and lymphoid lineages (Ichikawa et al., 2004). More specifically, targeted excision of *Runx1* in HSC and progenitor cells in adult mice showed that Runx1 is essential for the maturation of megakaryocytes and the differentiation of B and T cells. In addition, a significant expansion of myeloid progenitors was observed, which showed increased replication potential, consistent with a myeloproliferative phenotype. However, these mice did not develop full-blown leukaemia, which would potentially require the acquisition of additional mutations (Growney et al., 2005; Ichikawa et al., 2004).

During haematopoiesis RUNX1 has been reported to reshape the epigenetic landscape. Upon activation, RUNX1 leads to the deposition of H3K9ac (Lichtinger et al, 2012) and H4K5ac and to redistribution of the TFs it interacts with towards its binding sites (Gilmour et al, 2018). Work in the Bonifer lab has shown that induction of RUNX1 expression in mESCs cultured in conditions promoting haematopoietic differentiation results in increased acetylation at novel RUNX1 binding sites followed by binding of BRD4, CDK9, Mediator and LDB1 (Gilmour et al, 2018). Given that the last two proteins have been reported to promote enhancer-promoter interactions, it has been suggested that RUNX1 binding could establish enhancer-promoter interactions through their recruitment; however, experimental data validating this concept is lacking. Previous experiments in HSCs that focused on a limited number of loci, have indicated that loss of RUNX1 binding or mutation of its binding sites can lead to a reduced frequency of enhancer-promoter interactions (Levantini et al., 2011; Staber et al., 2014). More research is needed to elucidate whether RUNX1 impacts chromatin interactions and more specifically enhancer-promoter interactions, and whether this is mediated through the recruitment of specific co-factors.

Apart from the co-factors mentioned above, RUNX1 has been reported to associate with co-activators, including the acetyltransferases p300, CBP and MOZ, the nucleosome remodelling complex SWI-SNF, the MLL methyltransferase and the TET2 DNA demethylase, and also with co-repressors including histone deacetylases, the H3K9me3 methyltransferase SUV39H1 and PRC1

(Bakshi et al., 2010; Herglotz et al., 2013; Huang et al., 2011; Kitabayashi et al., 1998; Reed-Inderbitzin et al., 2006; Suzuki et al., 2017; Taniuchi et al., 2002; Yu et al., 2012). RUNX1 has been described to function both as an activator and as a repressor, for example during T-cell development (Sato et al., 2005; Taniuchi et al., 2002). Its function as an activator or repressor seems to depend on its preferential association with co-activators or co-repressors respectively, which in turn has been suggested to be influenced by the existing chromatin landscape and by the posttranslational modifications of RUNX1 (Blumenthal et al., 2017; Bonifer et al., 2017). However, it is still not completely understood exactly how RUNX1 could mediate both gene repression and activation, and how this "switch" is precisely controlled at gene targets.

1.9.2 The role of RUNX1 in leukaemogenesis

Given the crucial role of RUNX1 in the development of the haematopoietic system, it is not surprising that *RUNX1* is one of the most common targets of genetic alteration in leukaemia (Sood et al., 2017). Interestingly, the previous name for RUNX1 is AML1, which stands for Acute Myeloid Leukaemia 1, as it was one of the first mutated genes to be identified in acute myeloid leukaemia (AML) (Miyoshi et al., 1991). More specifically, *RUNX1* somatic mutations have been identified in AML, ALL, myelodysplastic syndromes (MDS) and in chronic myelomonocytic leukaemia (CMML). Furthermore, *RUNX1* germline mutations lead to the development of familial platelet disorder with predisposition to acute myeloid leukemia (FPD/AML), a rare syndrome characterised by mild thrombocytopenia and increased risk for the development of AML/MDS (Song et al., 1999; Sood et al., 2017).

RUNX1 mutations are usually point mutations in the DNA binding domain that result in defective DNA binding, or nonsense mutations occurring in the region after the DNA binding domain (Figure 1.8 a,b). Nonsense mutations result in the generation of truncated mutant proteins and likely affect

the interactions between RUNX1 and its interaction partners (Michaud et al., 2002; Osato, 2004). These point mutations of *RUNX1* contribute to the development of MDS and AML (Christiansen et al., 2004; Harada et al., 2004; Osato et al., 1999). The significance of *RUNX1* mutations in AML is highlighted by the fact that *RUNX1*-mutant AML seems to be a biologically distinct AML subgroup with a worse prognosis compared to other AML subgroups (Gaidzik et al., 2016; Mendler et al., 2012; Schnittger et al., 2011), and is listed as a distinct entity in the 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukaemia (Arber et al., 2016).

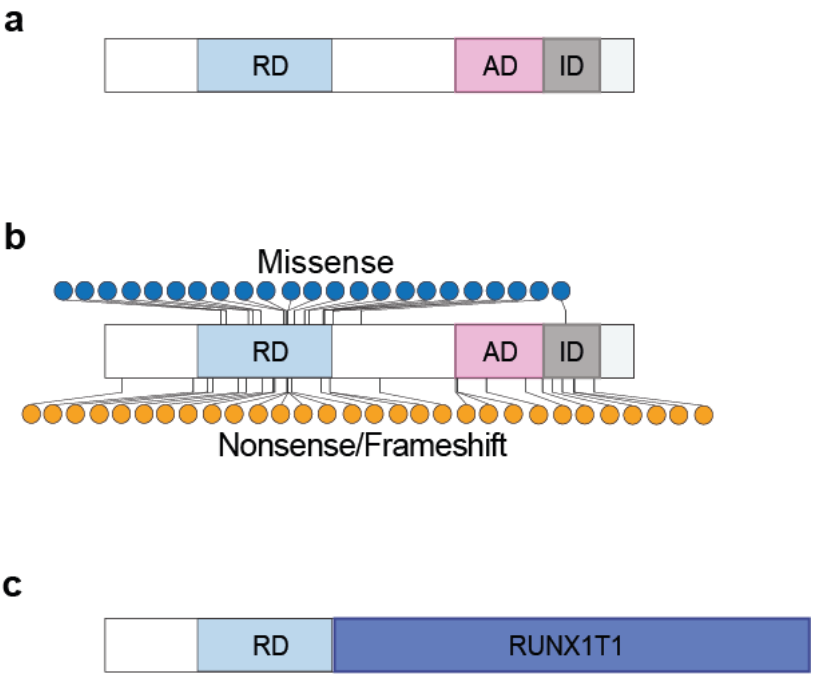


Figure 1.8: Schematic of wildtype RUNX1, RUNX1 point mutations and the RUNX1-RUNX1T1 fusion protein. (a) Schematic of wildtype RUNX1 protein. The DNA binding domain, also known as the Runt domain (RD) is shaded in light blue, the activation domain (AD) in pink and the inhibitory domain (ID) in grey. (b) Schematic of missense *RUNX1* mutations (highlighted in blue) and nonsense and frameshift mutations (highlighted in yellow) from AML patients. Adapted from (Simon et al., 2017). (c) Schematic of the RUNX1-RUNX1T1 fusion protein.

In *RUNX1*-mutant AML there is also an increased frequency of additional mutations in epigenetic modifiers, such as *IDH2* and *MLL,* including the *MLL-PTD (*partial tandem duplication), which results in an in-frame duplication of part of the N-terminus of MLL and has been associated with an adverse clinical outcome (Hinai et al., 2019). In addition, *RUNX1* mutations co-occur with mutations in *PcG*

genes (*ASXL1, EZH2, BCOR)*, with genes encoding components of the spliceosome complex (*SRSF2*, *SF3B1*) and also with mutations in the gene *PHF6* and the gene encoding the cohesin component STAG2 (Gaidzik et al., 2016; Ochi et al., 2020). By contrast, *RUNX1* mutations in AML seem to be mutually exclusive with *NPM1* mutations (Mendler et al., 2012). *RUNX1* AML mutations are generally associated with reduced overall survival and reduced event-free survival compared to other AML subgroups (Gaidzik et al., 2016; Mendler et al., 2012), and it has been shown that among *RUNX1*-mutant AML patients, those with additional *ASXL1, PHF6* and *SRSF2* mutations have worse clinical outcomes (Gaidzik et al., 2016).

The proportion of AML patients with *RUNX1* mutations varies in the different patient cohorts and ranges from 6% to over 30% (Gaidzik et al., 2011; Schnittger et al., 2011). In a study by the Cancer Genome Atlas Research Network consortium (TCGA) in which ~ 200 *de novo* AML patient samples were analysed, *RUNX1* mutations were identified in ~ 10% of adult AML patients (Cancer Genome Atlas Research et al., 2013). Further analysis of the clonal architecture of leukaemic samples showed that *RUNX1* mutations were identified in the founding clone, suggesting that they function as initiating mutations (Miller et al., 2013). Studies examining the transcriptional profile of cells from *RUNX1*-mutant AML patients compared to AML patients without *RUNX1* mutations have shown upregulation of genes expressed in primitive haematopoietic cells and B-cell progenitors (*DNTT*, *BAALC*, *HGF, GNABLNK*, *CD109*, *RBPMS*, *FLT3)* in *RUNX1*-mutant samples and downregulation of genes that promote myeloid differentiation, such as *CEBPA* and genes encoding components of neutrophil granules (*MPO, AZU1, CTSG*) (Greif et al., 2012; Mendler et al., 2012; Simon et al., 2017). In addition, the miRNA *miR-223* which promotes granulopoiesis, and miRNAs of the let-7 tumour suppressor family, which contribute to stem cell differentiation, were downregulated in *RUNX1*-mutant AML patients (Mendler et al., 2012) consistent with impaired myeloid differentiation in *RUNX1*-mutant AML. In agreement with this, *in vitro* experiments have shown that forced expression of wildtype RUNX1 in $CD34^+$ cord blood cells results in myeloid

differentiation, while RUNX1 mutant proteins identified in AML/MDS patients fail to induce myeloid differentiation of $CD34^+$ cells *in vitro* (Goyama et al., 2013).

In addition to point mutations, the *RUNX1* gene is involved in more than 50 translocations that lead to the generation of fusion proteins and the development of AML, ALL and/or MDS (Sood et al., 2017). The most notable translocations are the *t(8;21)* and the *t(12;21)* translocations. The *t(8;21)* translocation leads to the generation of the RUNX1-RUNX1T1 fusion protein (Miyoshi et al., 1993; Miyoshi et al., 1991), a driver of paediatric and adult AML (discussed in more detail in the next section), while the *t(12;21)* translocation gives rise to the ETV6-RUNX1 fusion protein and is the most common genetic aberration in childhood ALL (Sun et al., 2017). Furthermore, the co-factor of RUNX1, CBFb, is involved in the *inv(16)* translocation, which leads to the generation of the CBFb-SMMHC fusion protein and the development of AML (Liu et al., 1993).

The wild-type RUNX1 protein is critical for the survival of leukaemic cells with the RUNX1-RUNX1T1 and the CBFb-SMMHC fusion proteins (Ben-Ami et al., 2013; Goyama et al., 2013). Similarly, wild-type RUNX1 protein is a key protein in other forms of leukaemias in which the *RUNX1* gene is not involved in mutations or translocations, including leukaemias with rearrangements of the *MLL* gene (Goyama et al., 2013; Wilkinson et al., 2013) and in leukaemias with the E2A-PBX fusion protein (Pi et al., 2020). In the case of MLL-AF4 leukaemias, *RUNX1* is a direct downstream target of the MLL-AF4 fusion protein, and data from our lab has shown that RUNX1 is necessary for the survival of leukaemic cells (Wilkinson et al., 2013) (see section 1.9.2.2)

1.9.2.1 RUNX1 in RUNX1-RUNX1T1 leukaemia

The RUNX1-RUNX1T1 fusion protein (also known as AML1-ETO and RUNX1-ETO protein) results from the *t(8;21)* translocation between the *RUNX1* locus on chromosome 21 and the *RUNX1T1* locus on chromosome 8. More specifically, breakpoints within *RUNX1* intron 5 and *RUNX1T1* intron 1 result in the generation of a chimeric protein, consisting of the first 177aa of RUNX1, including

the DNA-binding RUNT domain, fused to the entire RUNX1T1 protein (previously known as ETO protein) (Miyoshi et al., 1993; Miyoshi et al., 1991; Nucifora et al., 1993) (Figure 1.8 c).

The traditional view of how RUNX1-RUNX1T1 promotes leukaemogenesis is that the fusion protein functions as a transcriptional repressor that inhibits the activity of wild-type RUNX1 and blocks myeloid differentiation. More specifically, it has been traditionally considered that RUNX1-RUNX1T1 has a dominant negative effect over wild-type RUNX1. Mice heterozygous for a R*UNX1-RUNX1T1* knock-in allele display embryonic lethality due to defects in foetal liver-derived haematopoiesis and haemorrhages, a phenotype similar to *Runx1*$^{-/-}$ and Cbfb$^{-/-}$ mice (Okuda et al., 1998; Yergeau et al., 1997). The function of RUNX1-RUNX1T1 as a repressor has been ascribed to the interaction of RUNX1T1 with the co-repressors NCoR/SMRT, mSin3a, and HDACs (Gelmetti et al., 1998; Lutterbach et al., 1998; Wang et al., 1998). This has been supported by data showing recruitment of co-repressors at RUNX1 binding sites, and inhibition of RUNX1-mediated transcriptional upregulation of critical myeloid genes, such as *CEBPA* by the fusion protein RUNX1-RUNX1T1 (Pabst et al., 2001; Ptasinska et al., 2014) resulting in a differentiation block (Ptasinska et al., 2014; Wang et al., 1998). In addition to downregulating key myeloid genes, RUNX1-RUNX1T1 has been reported to cause profound transcriptional deregulation by affecting the expression of miRNAs (Li et al., 2013; Li et al., 2008); through deregulation of key signalling pathways including the thrombopoietin/MPL (Chou et al., 2012; Pulikkan et al., 2012) and Wnt signalling pathways (Muller-Tidow et al., 2004); and by promoting mutagenesis through the downregulation of key DNA repair genes (Forster et al., 2016; Krejci et al., 2008). Despite traditionally being considered as a transcriptional repressor, the fusion protein has also been shown to activate a subset of downstream target genes (Li et al., 2016; Ptasinska et al., 2012), potentially through its association with co-activators (Wang et al., 2011). Therefore this suggests a context- and target-dependent function of the RUNX1-RUNX1T1 fusion protein that should be further investigated.

Mouse models of RUNX1-RUNX1T1 leukaemia have shown that the fusion protein alone is not sufficient to induce leukaemogenesis, and additional mutations are needed for the development of leukaemia (Higuchi et al., 2002; Yuan et al., 2001). This is further supported by clinical studies showing that patients with RUNX1-RUNX1T1 leukaemia have on average 2.2 additional secondary cooperating mutations (Christen et al., 2019). More specifically, recurrent mutations have been identified in genes of tyrosine kinase signalling pathways (*KIT*, *NRAS*, *KRAS*, *FLT3*), epigenetic modifiers (*ASXL1*, *ASXL2*, *EZH2*, *BCOR*, *TET2*), cohesin components (*RAD21*, *SMC1A*, *SMC3*, *STAG2*) and the spliceosome (*SRSF2*), and in addition, in a significant proportion of patients the loss of a sex chromosome is observed (Bolouri et al., 2018; Christen et al., 2019; Duployez et al., 2016). Despite the fact that RUNX1-RUNX1T1 leukaemias are generally considered to have a favourable prognosis, a significant proportion of patients relapse after receiving treatment (Jourdan et al., 2013; Yin et al., 2012). Minimal residual disease (MRD) (Hollein et al., 2018) and the presence of the *KIT D816* mutation (Krauth et al., 2014; Schnittger et al., 2006) are associated with reduced survival in RUNX1-RUNX1T1 patients, while the presence of *ASXL1* mutations is associated with reduced relapse-free survival (Hollein et al., 2018; Krauth et al., 2014; Micol et al., 2014).

Interestingly, despite the fact that *RUNX1* is very frequently mutated in leukaemia, in RUNX1-RUNX1T1 leukaemias, the RUNX1 allele that is not involved in the translocation is not mutated. This led to the hypothesis that wild-type RUNX1 protein promotes leukaemogenesis (Goyama and Mulloy, 2011), and indeed experimental data have shown that knockdown (KD) of wild-type RUNX1 inhibits the survival of RUNX1-RUNX1T1 leukaemic cells, by affecting the expression levels of apoptotic and cell signalling genes. More specifically, *RUNX1* siRNA KD in RUNX1-RUNX1T1 cells results in impaired G2/M-to-G1 and G1-to-S transitions, downregulation of mitotic genes and caspase-induced apoptosis. Interestingly, apoptotic cell death is dependent on RUNX1-RUNX1T1, as additional KD of the fusion protein after *RUNX1* siRNA KD attenuates the effect of *RUNX1* KD on reduced cell survival (Ben-Ami et al., 2013). This suggests that wild-type RUNX1 has an anti-apoptotic role, while the fusion protein promotes apoptosis. Consistent with that, other studies

have also reported that RUNX1-RUNX1T1 has a pro-apoptotic activity (Burel et al., 2001; Lu et al., 2006), and it can be concluded that the wild-type RUNX1 in part promotes survival of cancer cells by counteracting the pro-apoptotic function of the fusion protein.

RUNX1 and RUNX1-RUNX1T1 ChIP-seq has shown that there is a significant chromatin co-occupancy between the two proteins, although there are also RUNX1 ChIP-seq peaks not bound by the fusion protein, and conversely unique RUNX1-RUNX1T1 peaks (Ben-Ami et al., 2013; Li et al., 2016; Ptasinska et al., 2012). Peaks uniquely occupied by RUNX1 are more likely to be found at transcriptional start sites, and are enriched for the canonical RUNX motif, in contrast to the RUNX1-RUNX1T1 unique peaks, which are enriched for the variant RUNX motif (Ben-Ami et al., 2013). These findings in combination with recent observations that wild-type RUNX1 forms a complex with the fusion protein, and their relative signals determine whether a target gene will be repressed or activated (Li et al., 2016), highlight that an intricate balance between the fusion protein and wild-type RUNX1 is needed for maintenance of the leukaemogenic potential in RUNX1-RUNX1T1 leukaemias (Ben-Ami et al., 2013).

1.9.2.2 RUNX1 in MLL-AF4 leukaemia

RUNX1 has been shown to be necessary for the survival of MLL-rearranged (MLL-r) leukaemias (Goyama et al., 2013; Harman et al., 2021; Wilkinson et al., 2013), and increased *RUNX1* expression among MLL-r ALL patients correlates with a poor clinical outcome (Wilkinson et al., 2013). *RUNX1* is a direct downstream target of MLL-AF4 and other MLL fusion proteins in ALL and AML (Prange et al., 2017; Wilkinson et al., 2013), and it is an example of a spreading target of MLL fusion proteins. Spreading targets are characterised by the spreading of the fusion protein, the protein Menin that interacts with the fusion protein, and the H3K79me3 mark into the gene body, coupled to increased transcription (Kerry et al., 2017). *RUNX1* is overexpressed in MLL-r ALL cases compared to other ALL subgroups, and among MLL-r leukaemias its expression is higher in MLL-AF4 ALL cases

(Wilkinson et al., 2013). MLL-AF4 binding to the *RUNX1* locus results in activation of *RUNX1*, and reduction in MLL-AF4 binding is followed by reduction in RUNX1 protein levels. RUNX1 itself activates target genes in MLL-AF4 leukaemias, at least partly through its association with the MLL wildtype protein and the AF4-MLL protein which in turn interacts with P-TEFb (Benedikt et al., 2011; Wilkinson et al., 2013). These results are in agreement with previous observations showing that RUNX1 associates with the C-terminal of MLL (Huang et al., 2011), retained in the AF4-MLL protein but lost in the MLL-AF4 fusion protein.

Despite the critical role of RUNX1 in the establishment of definitive haematopoiesis and for leukaemic maintenance in leukaemias with fusion proteins, including MLL-AF4 and RUNX1-RUNX1T1 leukaemias (Ben-Ami et al., 2013; Goyama et al., 2013; Harman et al., 2021; Pi et al., 2020; Wilkinson et al., 2013), the exact mechanisms of how RUNX1 shapes the chromatin landscape and impacts gene regulation in normal and leukaemic cells remain elusive and need to be further investigated.

Aims of the thesis

Precise spatiotemporal control of gene expression is controlled through the concerted action of TFs and chromatin modifying complexes. TFs physically interact with and recruit chromatin proteins at their binding sites and upon their recruitment, chromatin complexes modify the chromatin landscape further affecting the binding of TFs. RUNX1 is a crucial TF for the development of the haematopoietic system that is frequently implicated in leukaemogenesis and leukaemia maintenance; however, the molecular mechanisms of how RUNX1 impacts gene regulation are incompletely understood.

The aim of this thesis is to understand the interplay between the binding of TF RUNX1 and the chromatin landscape, and more specifically to answer the following three questions:

- What are the histone modifying enzymes / nucleosome remodelling complexes that can be recruited by RUNX1?
- How does RUNX1 regulate the expression of target genes through the recruitment of chromatin-modifying complexes in MLL-AF4 and RUNX1-RUNX1T1 leukaemias?
- How do changes in the chromatin environment affect the binding of RUNX1?

Chapter 2 –Methods

2.1 Tissue culture methods

2.1.1 Cell lines and growth conditions

Cell lines used in this thesis are listed in the table below.

Table 2.1: Cell line information

Cell line	Species	Medium	Supplements	Notes
SEM	Human	Iscove's Modified Dulbecco's Medium (IMDM)	10% Foetal Bovine Serum (FBS), 1% Glutamax (#A12860-01, Gibco, Life Technologies)	MLL-AF4 pre-B ALL
SKNO-1	Human	Roswell Park Memorial Institute (RPMI)	10% FBS, 1% Glutamax, 10 ng/mL GM-CSF	AML1-ETO AML FAB M2
Kasumi-1	Human	RPMI	10% FBS, 1% Glutamax	AML1-ETO AML FAB M2
Embryonic stem cells	Mouse	Dulbecco's Modified Eagle Medium (DMEM)	10% FBS, 1% Glutamax, LIF, non-essential amino acids, β-mercaptoethanol	E14, TOT2N variant (Blackledge et al, 2014)
Schneider 2 (S2) cells	*Drosophila melanogaster*	Schneider's *Drosophila* medium	10% FBS	Derived from late stage (20–24 hours old) *Drosophila melanogaster* embryos

All mammalian cell lines were maintained at 37°C, 5% CO_2. Cells grown in suspension (SEM, SKNO-1, Kasumi-1) were maintained between 0.5 x 10^6 – 2 x 10^6 cells/mL, and were passaged every 2-3 days. Adherent mESCs were grown in 10 cm dishes and passaged at 80-90% confluency to 1:6 dilution. In brief, DMEM medium was aspirated from confluent mESCs in 10 cm dishes, cells were washed once with Phosphate Buffered Saline (PBS) and 1x trypsin (0.5% Trypsin-EDTA (10X), #15400-054, Gibco, Life Technologies) diluted in PBS was added. Cells were incubated at 37°C for 4

min, and trypsin was inactivated with DMEM supplemented with FBS. Cells were added in 1:6 dilution to new 10 cm dishes, to which 0.1% gelatin in PBS was added for 10 min and then aspirated.

Drosophila S2 cells were grown at 26°C; they were maintained between 0.5 x 10^6 – 2 x 10^6 cells/mL as a semi-adherent monolayer in T-25cm^2 flasks, and were passaged every 3-4 days.

2.1.2 Mycoplasma testing

All cell lines were tested for *Mycoplasma* contamination using the MycoAlert Mycoplasma Detection Kit, as per manufacturer's instructions. Cells used in this thesis were free from *Mycoplasma.*

2.1.3 Storage and recovery of cell lines

For storage, 10^7 cells were spun down at 1,000 rpm for 5 min and resuspended in medium plus 10% Dimethyl sulfoxide (DMSO). Cells were aliquoted in 1 mL into freezing vials and transferred at -80 °C in Cryo 1˚C Freezing Containers overnight. The next day, freezing vials were transferred to liquid nitrogen.

For recovery, cells were thawed at room temperature and transferred into warm medium. In order to remove DMSO, cells were spun at 1,000 rpm for 5 min and resuspended in the appropriate volume of medium before being placed in the incubator.

2.2 RNAi and cell transfection methods

2.2.1 siRNA knockdown assays

siRNA knockdown assays were performed using a rectangle pulse EPI 2500 electroporator (Fischer Heidelberg, Germany (Martinez et al. 2004)). 3 x 10^7 cells were spun at 1,000 rpm for 5 min and resuspended in 600 µL fresh medium appropriate for each cell line and transferred to a 2 mm

cuvette. 10 μL of 20 μM stock siRNA was added, and electroporation was performed for 0.1 sec at 330 V for SEM cells and at 350 V for the cell lines with the *RUNX1-RUNX1T1* translocation (SKNO-1 cells and Kasumi-1 cells). After the electroporation, cells were incubated at room temperature for 15 min, and were diluted in the appropriate for each cell line medium, pre-warmed at 37^{o}C, for a final concentration 10^{6} cells/mL. 48h after the electroporation cells were harvested for downstream applications, or a second electroporation was performed, and cells were harvested 48h afterwards (96h after the first electroporation).

2.2.2 Lipofectamine transfections and generation of stable mESC lines

mESCs were plated in 10 cm dishes and lipofectamine transfection was performed when cells were at 60-70% confluency. 75 μL of Lipofectamine 2000 reagent were diluted in 500 μL Opti-MEM medium (#51985-026, Gibco, Life Technologies) in an Eppendorf tube, and 28 μg of plasmid were added to 500 μL of Opti-MEM medium in a separate Eppendorf tube. The two mixtures were combined, and after a 5 min incubation at room temperature, the solution was added dropwise to mESCs. 24h post transfection cells were harvested for downstream applications, or 1 μg/mL puromycin was added for the selection of transfected cells, in order to generate stable cell lines. Puromycin selection was continued for 7-12 days, until the formation of visible colonies. Individual colonies were picked and placed in the wells of a 96-well dish (#3595, Corning Incorporated), in medium supplemented with puromycin. Cells were expanded and sequentially transferred to the wells of 24-well, 12-well and 6-well dishes (#3526, #3513, #3516 respectively, Corning Incorporated). Western blot analysis was employed to identify the clones expressing the proteins of interest.

2.3 Molecular and cloning methods

2.3.1 Polymerase Chain Reaction (PCR)

Polymerase Chain Reaction (PCR) was performed to amplify genomic loci of interest. For PCR reactions, the KAPA HiFi HotStart mix (#KK2601, Roche) was used. 10 ng of plasmid was added to 25 µL reaction mix (12.5 µL 2x KAPA HiFi HotStart, 1.5 µL 10 µM forward/reverse primers and water for a total volume of 25 µL). The steps of the PCR reaction are described in the table below.

Table 2.2: PCR reaction program

Temperature (°C)	Time (min)	Cycles
95	3:00	1
98	0:20	repeat 30 times
65	0:15	
72	1:00	
72	2:00	1
4	Hold	-

The PCR products were purified using the Qiagen PCR purification kit (#28104) or the Qiagen gel extraction kit (#28704), depending on downstream applications.

2.3.2 Restriction enzyme digestion

For restriction digests, 1 µg DNA was digested with the appropriate enzyme and buffer according to the manufacturer's instructions (New England Biosciences). The products of restriction digest were run on an agarose gel (the concentration of agarose varied according to the size of DNA), and purified using a Qiagen gel extraction kit (#28704).

For ligation of vector and insert with sticky ends, different ratios of insert: vector were used (e.g. 3:1, 2:1). For the ligation reaction, insert and vector were mixed with 1 µL T4 DNA Ligase 400,000 units / mL (#M0202S, NEB) and 2 µL buffer 10x T4 DNA Ligase Buffer (#B0202S, NEB) in a total volume of 20 µL at room temperature for 2 hours. Ligated DNA was used for bacterial transformation, as will be described in section 3.4.

2.3.3 Ligation Independent Cloning (LIC)

For the generation of plasmids for the TetR system, the method of Ligation Independent Cloning (LIC) was used. More specifically, the TetR vector (provided by the Klose lab) was digested overnight at 25°C with the BaeI enzyme, according to the reaction described in the table below.

Table 2.3: Restriction digest mix for LIC

Component	**Volume**
Vector DNA	Volume corresponding to 5 µg
10x CutSmart Buffer (#B7204S, NEB)	5 µL
BaeI enzyme (#R0613S, NEB)	1 µL
SAM (#B9003S, NEB)	1 µL
H2O	up to 50 µL

Following the overnight incubation with BaeI, an additional 1 µL BaeI and 1 µL SAM were added and the sample was left for two additional hours at 25°C. The linearised vector was gel purified using the Qiagen gel extraction kit (#28704).

The desired insert DNA was PCR amplified using primers with 15bp LIC sites, with sequence complementary to the LIC sites flanking the BaeI digestion sites in the TetR vector. Following PCR amplification, the insert DNA with the correct band size was gel purified. The vector and the insert were incubated with T4 DNA polymerase (#M0203L, NEB) and dGTPs (#R0161, ThermoFisher Scientific) or dCTPs (#R0151, ThermoFisher Scientific) respectively, as follows:

Table 2.4: LIC reaction

Component	**Volume**
Vector or insert	10 µL
T4 DNA polymerase (#M0203L, NEB)	3 µL
dGTPs (#R0161) for vector or dCTPs (#R0151, ThermoFisher Scientific) for insert	2 µL
NEBuffer 2.1 (#B7202S, NEB)	5 µL
H2O	25 µL

The reaction was carried out for 30 min at 22°C, followed by 20 min at 75°C for heat inactivation of the enzyme. The reaction uses the 3' to 5' exonuclease activity of DNA polymerase. More specifically, DNA polymerase removes the nucleotides of vector and insert that correspond to the 15bp LIC sites. The removal of nucleotides continues until DNA polymerase reaches a guanine (vector) or a cytosine (insert), as the addition of dGTPs and dCTPs respectively inhibits further digestion. This results in the generation of 15bp overhangs in the vector and the insert, which are complementary. Purification of the DNA polymerase-treated vector and insert is performed using the Qiagen PCR purification kit (#28104).

Ligation is performed for 30 min at room temperature, without the addition of DNA ligase, by mixing the vector and the insert using different molar ratios. Afterwards, the ligation product is used for bacterial transformation, as described in the section below.

2.3.4 Bacterial transformations

For bacterial transformations, Stellar competent cells (#636763, Takara), an *E. coli* HST08 strain with high transformation efficiency, were used. Bacterial cells were thawed on ice for 30 min and 50 µL of cells were mixed with 5 µL of DNA from ligation reactions or 0.5 µL maxiprep /miniprep DNA. Bacterial cells were heat-shocked at 42°C for 45 seconds and then incubated on ice for 2 min. 450 µL Super Optimal broth with Catabolite repression (SOC) medium was added to the bacteria and they were incubated for 1 h at a shaker at 37°C. Following that, 100 µL were added to agar plates with the appropriate antibiotic. Plates were incubated at 37°C overnight and assessed for the presence of colonies the following day.

2.3.5 Plasmid amplification and sequencing

Bacterial colonies identified in agar plates were picked and incubated overnight in a shaker at 37°C in 3 mL Lysogeny broth (LB) supplemented with 50 μg/mL of the appropriate antibiotic. The next day plasmid DNA was purified using the Qiagen MiniPrep Kit (#27104), according to the manufacturer's instructions. Correct DNA sequences were verified by sending the purified plasmid DNA together with appropriate primers for Sanger sequencing. Similarly, for maxiprep, 1.5 mL from liquid bacterial cultures were grown in 200 mL LB medium supplemented with 50 μg/mL of the appropriate antibiotic, in a shaker at 37°C overnight. Plasmid DNA was purified the next day using the Qiagen EndoFree Plasmid Maxi Kit (#12362), according to the manufacturer's instructions.

2.4 Gene expression methods

2.4.1 Total RNA extraction

10^6 cells were used for RNA extraction. The RNeasy Mini Kit (#74106, Qiagen) was used according to the manufacturer's instructions. RNA quantification was performed using the NanoDrop instrument (ThermoFisher Scientific).

2.4.2 Generation of cDNA

For the generation of cDNA, 8 μL of RNA was used, to which 1 μL dNTP mix and 1 μL random hexamers were used. In order to compare gene expression across different conditions, the same amount of RNA (up to 4 μg) was used and nuclease-free water was added for a total volume of 8 μL. The mixture of RNA, dNTPs and random hexamers was placed at 65°C for 5 min and then snap-cooled on ice for 2 min. 10 μL of reaction mixture consisting of 4 μL 5x First-Strand Buffer (#02321, Invitrogen), 2 μL DTT (0.1 M, #00147, Invitrogen), 1 μL RNaseOUT (40 units/μL , #51535, Invitrogen) and 1 μL SuperScript III Reverse Transcriptase (200 units/μL , #56575, Invitrogen) and 2 μL water. PCR reaction was performed using the following parameters: 25°C for 5 min (primer extension),

50°C for 60 min (cDNA synthesis), and 70°C for 15 min (termination). cDNA was diluted to 200 µL with the addition of water and was stored at -20°C.

2.4.3 Real Time Quantitative Polymerase Chain Reaction (RT-qPCR)

Real Time Quantitative Polymerase Chain Reaction (RT-qPCR) was performed to assess gene expression (cDNA) or enrichment of a histone modification / binding of a transcription factor at a specific genomic locus (ChIP DNA). For SYBR Green primers, 5 µL of sample was added to 15 µL of reaction mixture (10 µL Fast SYBR® Green Master Mix (#4385612, Applied Biosystems), 0.25 µL forward primer (10 µM), 0.25 µL reverse primer (10 µM), 4.5 µL water) in a well of a 96-well PCR plate (MicroAmp Fast Optical 96-well Reaction Plate, #4346906, Applied Biosystems). Similarly, for Taqman primers, 5 µL of sample was added to 15 µL of reaction mixture (10 µL TaqMan® Fast Advanced Master Mix (#4444557, Applied Biosystems), 1 µl primer/probe mix (Invitrogen), 4 µL water). Plates were spun for 1 min and PCR reactions were performed on an RT-qPCR machine (7500 Fast Real Time PCR System, Applied Biosystems), with parameters Comparative Ct (ΔΔCt), Fast. The conditions of PCR reactions are described in the table below.

Table 2.5: RT-qPCR amplification program

Temperature (°C)	Time (sec)	Cycles
95	20	1 (initial denaturation)
95	3	40 (denaturation, annealing and extension)
60	30	
95	15	1 (melt curve stage for SYBR Green primers)
60	60	
95	15	
60	15	

SYBR Green primers for ChIP DNA were designed using the IDT primer quest tool (https://eu.idtdna.com/PrimerQuest/Home/Index). PCR reactions with serial 10-fold dilutions of ChIP input material were performed to check the efficiency of the primers, and melt curve analysis

was performed to ensure the generation of a single amplification product and the absence of primer dimers.

2.4.4 Nascent RNA-sequencing

$3x10^7$-$5x10^7$ cells were incubated with 50 µM 4-thiouridine (#T4509, Sigma) for 1 hour. Cells were harvested and RNA was extracted with Trizol (#15596026, ThermoFisher Scientific) and precipitated with ethanol. RNA was treated with DNAase using the TURBO DNA-free kit (#AM1907, ThermoFisher Scientific), according to the manufacturer's instructions. In brief, RNA was incubated for 30 min at 37°C with 10 µL TURBO DNase buffer and 3 µL TURBO DNase, and the reaction was stopped by adding 10 µL DNase Inactivation Resin. RNA was quantified and a biotinylation reaction was performed, in which 100 µg 4-thiouridine-treated RNA was incubated with 1 mg/mL EZ-Link™ HPDP-Biotin (#21341, ThermoFisher Scientific) for 90 min at room temperature. Labelled RNA was chloroform extracted and isolated using streptavidin beads (µMACS Streptavidin Kit, #130-074-101, Miltenyi Biotec). Beads were washed on a µMACS magnetic stand and RNA was eluted using 100 mM DTT. RNA was purified using the RNeasy MiniElute Cleanup Kit (#74204, Qiagen). Library preparation was performed using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (#E7760, NEB), according to the manufacturer's instructions. Sequencing and bioinformatics analysis was performed as described in sections 2.9 and 2.10.2 respectively.

2.5 Protein methods

2.5.1 Whole cell extracts

10^6 cells were spun at 1,000 rpm for 5 min and rinsed with PBS. Pellets were resuspended in 50 µL BC300 (20 mM Tris-HCL pH7.5, 20% glycerol, 300 mM KCl, 5 mM EDTA) plus 0.5% NP-40 plus protease inhibitor complex in 1:100 concentration. Cells were incubated on ice for 30 min, and were spun down at 13,000 rpm for 10 min at 4°C. The supernatant was transferred to a new

Eppendorf tube, to which 50 µL of 2x NuPAGE LDS sample buffer (#1633780, Life Technologies) plus 0.1% β-mercaptoethanol was added. Samples were heated at 95°C for 5 min and a small aliquot was loaded onto pre-cast gels or samples were stored at -20°C. When required, the pellet obtained after the centrifugation step was used for histone acid extraction as described below.

2.5.2 Histone acid extraction

Pellets retained after the centrifugation step during the preparation of whole cell extracts (described in section 2.5.1) were used for histone acid extraction. 50 µL of 0.4M HCl were added to each pellet and the samples were incubated on ice for 1h. Samples were spun at 13,000 rpm at 4°C for 10 min, the supernatant was transferred to a new Eppendorf tube and 1 mL ice cold acetone was added. Samples were stored at -20°C overnight, and on the following day were spun at 13,000 rpm at 4°C for 20 min. The pellets were then washed twice with 1 mL ice cold acetone and spun at 13,000 rpm at 4°C for 10 min. Acetone was removed and the pellets were air-dried at room temperature. The pellets were resuspended in 50 µL 1x LDS NuPAGE loading buffer with 4M urea plus 0.5% β-mercaptoethanol, and then the samples were stored at -20°C or were loaded onto pre-cast Bis-Tris gels for sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), as described below.

2.5.3 Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

For the detection of proteins, except for histone proteins, pre-cast 4-12% Bis-Tris gels (#NP0323BOX, ThermoFisher Scientific) and 1x MOPS SDS Running buffer (#2395813, Life Technologies) were used. For the detection of histones, pre-cast 12% Bis-Tris gels (#NP0321BOX, ThermoFisher Scientific) and 1x MES SDS Running buffer (#2062909, Life Technologies) were used. Aliquots of the samples were added to the wells of the gels, and the Precision Plus Protein Kaleidoscope ladder (#110375, Biorad) was used to monitor electrophoretic separation for SDS-PAGE gels and determine molecular weight. Gels were run for approximately 1 h at 180 V.

2.5.4 Western blotting

After electrophoresis the proteins on the gel were transferred to a Polyvinylidene fluoride (PVDF) membrane (Immobilon-P Transfer Membrane, #IPVH00010, EMD Millipore) using a wet transfer setup. For the activation of the PVDF membrane 100% methanol was added for 1 min, and then the membrane was rinsed with transfer buffer (192 mM glycine; 25 mM Trizma® base; 10% methanol). The gel and the membrane were sandwiched between filter paper and sponges pre-soaked in transfer buffer, and were submerged in a tank filled with transfer buffer. Transfer was performed for 1 h at 100V at 4^{o}C.

After transfer, membranes were incubated with Ponceau stain for 5 min and then were washed three times with 100% of methanol. Membranes were dried and blocked in 5% w/v milk in 1x Tris-buffered saline (TBS) with 0.05% Tween-20 for 1 h. Afterwards the membranes were incubated overnight at 4^{o}C with primary antibody diluted in 5% w/v milk or 5% w/v BSA in 1x TBS - 0.05% Tween, according to the manufacturer's instructions. The next day the membranes were washed three times for 5 min with 1x TBS - 0.05% Tween-20, and were incubated with secondary antibody, diluted at the appropriate concentration in 5% w/v milk in 1x TBS - 0.05% Tween-20, for 1 h at room temperature. Membranes were washed again three times for 5 min with 1x TBS - 0.05% Tween-20.

For the detection of the proteins the Enhanced chemiluminescence (ECL) reagent (ECL Prime Western Blotting Detection Reagent, #RPN2232, Amersham, GE Healthcare) was used. More specifically, ECL solutions A and B were mixed in a 1:1 ratio and added to the membrane for 5 min. Excess reagent was removed and the membrane was developed using chemiluminescence film (Amersham Hyperfilm ECL, #28906837, GE Healthcare) and an X-ray developing machine. The antibodies used for Western blot (and ChIP) experiments in this thesis are listed in the table below.

Table 2.6: Antibody information

Antibody name	**Company**	**Species**	**Working concentrations**
Anti-Rabbit HRP (2°)	Sigma, #A6667	Goat	1:10000 WB
Anti-Mouse HRP (2°)	Sigma, #A4416	Goat	1:10000 WB
BAP1	Bethyl, #A302-242A	Rabbit	1:330 ChIP
BRD4	Bethyl, #A301-985A	Rabbit	1:500 ChIP
BRG1 (SMARCA4)	Bethyl, #A300-813A	Rabbit	1:500 ChIP
CBFβ	Cell Signaling, #62184	Rabbit	1:5000 WB, 1:500 ChIP
CBX4	Abcam, #ab242149	Rabbit	1:330 ChIP
CTCF	Sigma, #07-729	Rabbit	1:500 ChIP
EZH2	Cell Signaling, #5246	Rabbit	1:5000 WB, 1:500 ChIP
GAPDH	Bethyl, #A300-641A	Rabbit	1:10000 WB
GR (glucocorticoid Receptor)	Cell Signaling, #12041	Rabbit	1:500 ChIP
H2AK119ub	Cell Signaling, #8240	Rabbit	1:500 ChIP
H3	Abcam, #ab1791	Rabbit	1:1000000 WB
H3K4me1	Diagenode, #C15410194	Rabbit	1:500 ChIP
H3K4me3	Diagenode, #C15410003	Rabbit	1:500 ChIP
H3K9me3	Abcam, #ab8898	Rabbit	1:5000 WB, 1:500 ChIP
H3K27ac	Diagenode, #C15410196	Rabbit	1:10000 WB, 1:500 ChIP
H3K27me3	Sigma, #07-449	Rabbit	1:5000 WB, 1:500 ChIP
H4 ac	Milipore, #06-866	Rabbit	1:10000 WB, 1:500 ChIP
HDAC2	Cell Signaling, #57156	Rabbit	1:500 ChIP

MLL1	Bethyl, #A300-086A	Rabbit	1:500 ChIP
OGT	Novus Biologicals, #NBP1-89844	Rabbit	1:500 ChIP (human cells)
OGT	Cell Signaling, #5368	Rabbit	1:500 ChIP (mouse cells)
P300	Bethyl, #A300-358A	Rabbit	1:500 ChIP
PCGF1	gift from Klose lab	Rabbit	1:330 ChIP
PCGF6	Novus Biologicals, #NBP1-55372	Rabbit	1:330 ChIP
PHC1	Cell Signaling, #13768	Mouse	1:330 ChIP
PHC2	Merck, #ABE1069	Rabbit	1:330 ChIP
RING1B	Cell Signaling, #5694	Rabbit	1:2500 WB, 1:500 ChIP
RNA Polymerase II	Santa Cruz, #sc-899	Rabbit	1:500 ChIP
RNA Polymerase II	Millipore, #05-623	Rabbit	1:500 ChIP
RUNX1	Abcam, #ab23980	Rabbit	1:5000 WB, 1:500 ChIP (human cells)
RUNX1	Cell Signaling, #4336	Rabbit	1:5000 WB, 1:500 ChIP (human cells)
RUNX1 (mouse preferred)	Cell Signaling, #8529	Rabbit	1:500 ChIP (mouse cells)
SUZ12	Cell Signaling, #3737	Rabbit	1:500 ChIP
TAF1C	Bethyl, #A303-697A	Rabbit	1:330 ChIP
TBP	Bethyl, #A301-229A	Rabbit	1:500 ChIP
TetR (FS2)	gift from Klose lab	Rabbit	1:1000000 WB, 1:500 ChIP
VINCULIN	Abcam, #129002	Rabbit	1:20000 WB

2.6 Cell cycle analysis

For cell cycle analysis, 10^6 cells were fixed in 70% ethanol and incubated at 4 °C overnight, then stained with FxCycle PI/RNase staining solution (#F10797, ThermoFisher Scientific). Cells were analysed using 532 nm excitation and collected using 610/20 bandpass (by Dr. Natalina Elliott). Analysis was performed with the FlowJo software.

2.7 ChIP-qPCR and ChIP sequencing

2.7.1 Fixation

ChIP-qPCR and ChIP-seq experiments were performed using 10^7 cells and $4x10^7$-10^8 cells respectively. For fixation, 10^7 cells were harvested, spun down for 5 min at 1,000 rpm and washed with PBS. Cells were either single-fixed for the detection of histone modifications, or double-fixed for the detection of transcription factor binding. More specifically, for single fixation, cells were resuspended in 1 mL 1% formaldehyde (FA) and rotated for 10 min at room temperature. For double fixation, cells were resuspended in 1 mL 2 mM disuccinimidyl glutarate (DSG) and rotated for 30 min at room temperature. Following that, cells were spun at 10,000 rpm for 1 min and DSG was removed. Pellets were resuspended in 1 mL 1% FA for 30 min. After FA fixation, double-fixed and single-fixed cells were spun at 10,000 rpm for 1 min, pellets were rinsed in PBS and were frozen on dry ice.

2.7.2 ChIP protocol

Day I

Single-fixed or double-fixed pellets were thawed on ice, resuspended in 120 µL lysis buffer (1% SDS, 10mM EDTA, 50mM Tris, pH 8.1) plus 1:1,000 protease inhibitor cocktail, and syringe-passaged through a 27-gauge needle. Lysed cells were transferred to Covaris tubes, and sonication was performed using a Covaris sonicator (S220, Covaris Inc.). The Covaris SonoLite software was used with parameters: Duty cycle = 10%, Intensity = 5, Cycles / burst = 200, Duration = 300 s. This was performed once (300 s) for single-fixed chromatin and twice (600 s) for double-fixed chromatin. Sonicated chromatin was then transferred to Eppendorf tubes and spun at 13,000 rpm for 10 min at 4^{o}C. Supernatant was diluted 1:10 in ChIP dilution buffer (CDB - 16.7 mM Tris-HCl pH 8.1, 167 mM NaCl; 1.2 mM EDTA; 0.01% SDS; 1.1% Triton X-100). Protein A/G Dynabeads resuspended in CDB were added in concentration 1:500 to the diluted sonicated chromatin for pre-clearing.

Samples were incubated at 4°C for 30 min with rotation, and then spun for 2 min at 2,000 rpm. Cleared chromatin in CDB was split into 1 mL aliquots, and 2 x 50 µL aliquots were taken as input samples and for checking the efficiency of sonication. Antibody was added to the 1 mL aliquots and samples were left overnight at 4°C with rotation. The antibodies used for ChIP experiments in this thesis are listed in the table in section 2.5.4.

Day II

Next day, 15 µL of A/G Dynabeads resuspended in CDB were added to the 1 mL ChIP samples and they were incubated for 3-5 hours at 4°C. Following that, beads were washed 3 times with 1 mL RIPA buffer (50 mM HEPES-KOH pH7.6, 500 mM LiCl, 1 mM EDTA, 1% NP-40, 0.7% Na-Deoxycholate) on a magnetic stand, and Eppendorf tubes were changed after the first wash in order to reduce background signal. Beads were rinsed in 1 mL TE-50 mM NaCl buffer, and spun at 3,200 rpm for 3 min. TE buffer was removed and beads were resuspended in 105 µL elution buffer (1% SDS, 10mM EDTA, 50mM Tris, pH 8.1) and incubated at 65 °C for 1 hour. Following the 1 hour incubation, samples were spun at 10,000 rpm for 1 min, 100 µL of the supernatant were transferred to a new Eppendorf tube and were incubated at 65°C overnight in order to reverse cross-linking. 50 µL elution buffer was added to each one of the 50 µL input samples and they were incubated at 65°C overnight as well.

Day III

1 µL of RNAse A (10 mg/mL, #EN0531, ThermoFisher Scientific) was added to each sample and they were placed at 37°C for 1 hour. Following that, 1 µL of proteinase K (20mg/mL, #EO0491, ThermoFisher Scientific) was added to each sample and they were incubated at 45°C for 1 hour. DNA was extracted using the ChIP DNA Clean & Concentrator kit (#D5205, Zymo Research), and eluted in 200 µL elution buffer for ChIP-qPCR or 55 µL for ChIP-seq. Samples were stored at -20°C.

2.7.3 Reference-normalised ChIP sequencing

Reference-normalised ChIP-seq experiments were performed exactly as described above, with the addition of *Drosophila* S2 cells during the lysis step of the ChIP protocol. S2 cells were fixed as described before (single-fixed for histone modification ChIP-seq or double-fixed for transcription factor ChIP-seq), and were added to 1:5 ratio to human / mouse cells. The ChIP protocol described above was followed from this point onwards. The ratio of *Drosophila* to human / mouse cells in the input and IP samples was used for normalisation of ChIP-seq reads, as will be described in section 2.10.1.

2.7.4 Library preparation for ChIP sequencing

The efficiency of ChIP sequencing was validated with qPCR at selected loci and DNA was quantified using the Qubit dsDNA HS Assay kit (#Q32851, ThermoFisher Scientific). Library preparation was performed using the NEBNext Ultra II DNA Library Prep Kit for Illumina (#E7645S, NEB) and NEBNext Multiplex Oligos for Illumina (#E7335L, #E7500L, #E7710L, #E7730L, NEB), as per manufacturer's instructions. In brief, the steps of the library preparation protocol include end repair of fragmented DNA, adaptor ligation and clean-up without size selection using AMPure XP Beads (Beckman Coulter, Inc. #A63881). This is followed by PCR amplification of adaptor-ligated DNA for 8-9 PCR cycles, and clean-up of amplified DNA with AMPure XP Beads. Libraries were stored at -20^{o}C before proceeding to next generation sequencing (discussed in section 2.9).

2.8 ATAC sequencing

2.8.1 Tagmentation protocol

The protocol for ATAC sequencing (ATAC-seq) was adapted from (Buenrostro et al., 2013). In brief, approximately 50,000 cells were pelleted, rinsed in cold PBS and resuspended in cold lysis buffer

(10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% IGEPAL CA-630). Cells were spun at 500 g for 10 min at 4°C and the pellets were resuspended in a transposase reaction mix (25 µL 2x Tagment DNA Buffer, 2.5 µL Tagment DNA Enzyme (#20034197, Illumina) and 22.5 µL nuclease-free water, for a total volume of 50 µL). Samples were incubated at 37°C for 30 min. DNA was purified using a Qiagen MinElute Kit, and it was eluted in 10 µL 10 mM Tris pH 8 and stored at -20°C before using it for ATAC-seq library preparation, as described in the section below.

2.8.2 ATAC-seq library preparation

For PCR amplification of purified transposed DNA the following components were added to the PCR reaction:

Table 2.7: ATAC-seq library preparation reaction

Purified DNA	10 µL
Customised universal primer	2.5 µL
Customised index primer	2.5 µL
2x NEBNext high-fidelity 2x PCR master mix	25 µL
Nuclease-free water	10 µL

The PCR amplification programme is described in table below. PCR-amplified DNA was purified using Qiagen PCR clean up kit and eluted in 20µL elution buffer (10mM Tris Buffer, pH 8).

Table 2.8: PCR amplification for ATAC-seq library preparation

Step	**Time (min)**	**Temperature (°C)**
1	5:00	72
2	0:30	98
3	0:10	98
4	0:30	63
5	0:10	72 (repeat 12x steps 3-5)
6	Hold	4

2.9 Sequencing

After library preparation quality control of the libraries was performed using a 2200 TapeStation instrument (Agilent technologies) according to the manufacturer's instructions. For library quantification the KAPA Universal Complete Kit (#KK4824, Kappa Biosystems) was used. Sequencing was performed using NextSeq 500/550 v2.5 Kits from Illumina (75 cycles high kits were used for ChIP-seq and ATAC-seq, 150 cycles high kits were used for nascent RNA-seq). Libraries were sequenced using the NextSeq 500 platform (Illumina), and for all experiments paired end reads were obtained.

2.10 Bioinformatics analysis

2.10.1 ChIP-seq/ATAC-seq

ChIP-seq and ATAC-seq results were analysed using an in-house pipeline (Telenius et al., 2018). Briefly, the quality of the fastq files was assessed with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and then mapped to the corresponding genome assembly (hg19 or mm10, and dm3 for ChIP-rx). Reads that were not mapped, were trimmed with trim-galore (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), and short unmapped reads that did not map after trimming, were combined using Flash (Magoc and Salzberg, 2011) and mapped again. PCR duplicates were filtered using samtools (Li et al., 2009), and reads mapped to UCSC blacklisted regions were removed. The bam files that were generated were used to make Tag directories with the Homer makeTagDirectory tool (Heinz et al., 2010). The Homer tool makeBigWig.pl was used for the creation of BigWig files by normalizing tag counts to tags per ten million, for data visualisation in the UCSC genome browser. For ChIP-rx experiments, for the normalisation of tag counts, the ratio of *Drosophila* to human reads (or mouse reads) was calculated for the input and IP samples across conditions. A normalisation factor was derived for

each condition so that the resulting *Drosophila* signal was equilibrated across conditions, as described in (Orlando et al., 2014), as the same number of *Drosophila* S2 cells was added across conditions. More specifically, the normalisation factor was calculated as 1/number of reads mapping to the *Drosophila* genome per million, after having taken into consideration the ratio of *Drosophila* to human (or mouse) reads in the input samples, to correct for variations in the number of the *Drosophila* cells that were spiked-in across conditions.

To determine ChIP-seq and ATAC-seq peaks, the Homer tool findPeaks was used, with the parameters – style histone for histone ChIP-seq and – style factor for TF ChIP-seq and ATAC-seq, and the input files were used for background correction. To identify overlapping peaks between the different datasets, the Homer command mergePeaks was used. Metagene profiles were created using the Homer command annotatePeaks.pl. Correlation maps were created using the deeptools commands multiBigwigSummary and plotCorrelation. Heatmaps were created using the deeptools commands computeMatrix and plotHeatmap, with default parameters (Ramirez et al., 2016).

2.10.2 Nascent RNA-seq

For the analysis of nascent RNA-seq data, the quality of the fastq files was assessed with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and the reads were aligned to the corresponding genome assembly (hg19 or mm10) using STAR (Dobin et al., 2013). The featureCounts function of the Subread package (Liao et al., 2013) was used for the quantification of gene expression levels, by assigning the mapped reads to genes using default parameters. The read counts were then used for differential gene expression analysis with the EdgeR package (Robinson et al., 2010), and genes were considered differentially expressed between conditions when they had an adjusted p-value of less than 0.05.

For publicly available RNA-seq data from RUNX1-RUNX1T1 and *RUNX1*-mutant AML patients (chapter 4), the bioinformatics analysis was performed as above, and additionally unwanted variation between samples (batch effect, library preparation) was removed using the RUVg function of the RUVseq package (Risso et al., 2014). Following this, differential gene expression analysis was performed using EdgeR as described above.

Chapter 3 - RUNX1 is sufficient to recruit co-activators and Polycomb repressors

3.1 Introduction

3.1.1 RUNX1 structure and interaction partners

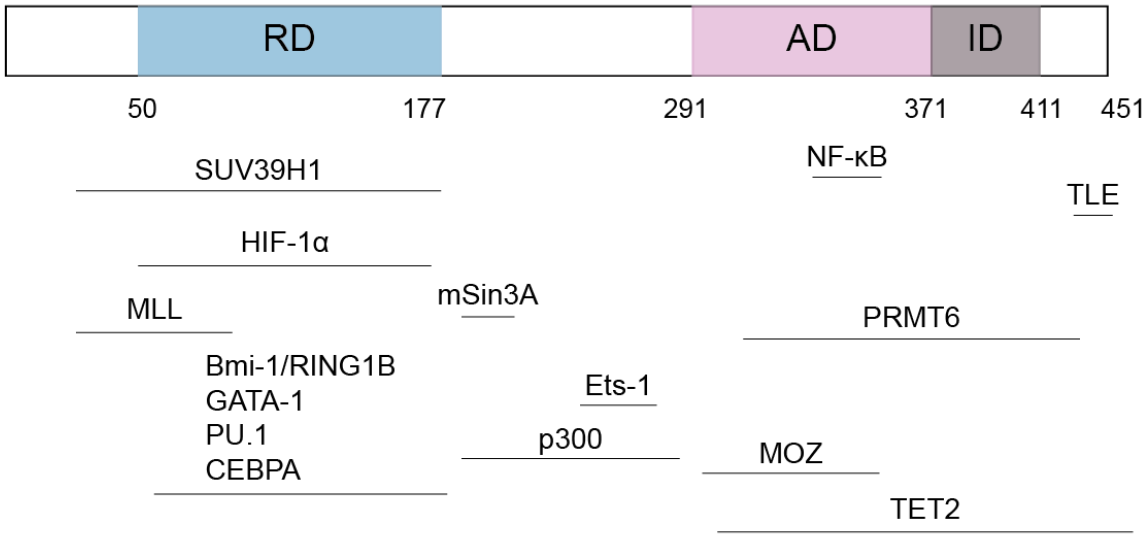
RUNX1 and its co-factor CBFb are indispensable for the development of the haematopoietic system, as their expression is critical for emergence of haematopoietic stem and progenitor cells from the haemogenic endothelium during foetal life (Chen et al., 2009; Lie et al., 2014). RUNX1 is also necessary for the maturation of megakaryocytes and for lymphocyte differentiation (Ichikawa et al., 2004). It is therefore not surprising that *RUNX1* is often a target of genetic alteration in myeloid and lymphoid malignancies (Sood et al., 2017). In order to understand how RUNX1 orchestrates the development of haematopoietic tissues and how its mutations promote leukaemogenesis, it is crucial that we better understand the mechanisms of RUNX1-mediated gene regulation. Here, we focus on better understanding how RUNX1 binding affects the recruitment of gene-regulatory proteins at regulatory elements to shape the chromatin landscape and how this impacts gene regulation.

Due to the observation that RUNX1 can interact with both with activators and repressors, including complexes with opposite activities (e.g. histone acetyltransferases as well as deacetylases), it has been concluded that RUNX1 can function both as an activator and as a repressor depending on the cellular and gene context. For example, during the endothelial-haematopoietic transition (EHT) that leads to the emergence of the first definitive blood cells (Ottersbach, 2019), RUNX1 binding has been reported to lead to an increase in acetylation and increased binding of co-activators (Gilmour et al., 2018). In addition, its function both as co-activator and co-repressor is required for T-cell maturation, and more specifically for the transition from double negative thymocytes to the emergence of $CD8^+$ or $CD4^+$ T cells. For the emergence of $CD8^+$ T cells, RUNX1 has been shown to

be required for binding at enhancer elements controlling the expression of *CD8*, and also is required for the repression of *CD4* (Sato et al., 2005; Taniuchi et al., 2002).

More specifically, previous studies have suggested that RUNX1 co-localises on chromatin and interacts with other TFs, including the GATA proteins, PU.1, SCL/TAL1 and other haematopoietic TFs (Lichtinger et al., 2012; Wilson et al., 2010). In addition, it has been reported that RUNX1 interacts with many gene-regulatory and chromatin remodelling proteins, including p300 and MOZ acetyltransferases, the MLL methyltransferase, the SWI/SNF complex, the TET2 demethylase and also with co-repressors, including the RING1B/BMI1 components of the PRC1 complex, the histone arginine methyltransferase PRMT6, the SUV39H1 methyltransferase that catalyses methylation of H3K9 and histone deacetylases (HDACs) that remove acetyl marks from histones (Bakshi et al., 2010; Herglotz et al., 2013; Huang et al., 2011; Kitabayashi et al., 1998; Reed-Inderbitzin et al., 2006; Suzuki et al., 2017; Taniuchi et al., 2002; Yu et al., 2012) (Figure 3.1 a). How is RUNX1 capable of binding to such a diverse array of regulatory partners? With the exception of the DNA binding domain, RUNX1 is predicted to be made up of mainly intrinsically disordered regions (Figure 3.1 b). The presence of intrinsically disordered regions enables multiple different conformation states rather than a fixed tertiary structure, and facilitates the interaction of a protein with a plethora of different co-factors, providing both flexibility and versatility (Staby et al., 2017). The question then becomes, how does RUNX1 use this capability of binding to a diverse range of different activating and repressing partners to produce specific transcriptional outputs at target loci?

a



Protein	Activity
MLL	H3K4 methyltransferase
SWI/SNF	Nucleosome-remodelling complex
RING1B	E3 ubiquitin ligase (deposition of H2AK119ub)
SUV39H1	H3K9 methyltransferase
p300, MOZ	Histone acetyltransferases
HDACs	Histone deacetylases
TET2	DNA demethylase
PRMT6	Arginine methyltransferase

b

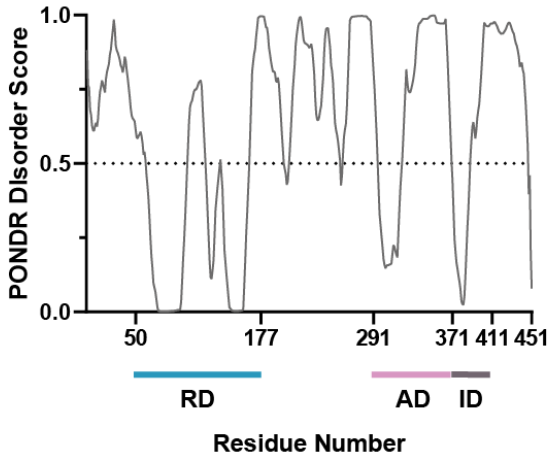


Figure 3.1: Schematic of Runx1 protein structure and interaction partners. (a) Diagram of Runx1 protein structure and its conserved domains. The DNA binding domain, also known as the Runt

domain (RD) is shown in blue, the activation domain (AD) in pink and the inhibitory domain (ID) in grey. Numbers correspond to amino acid residues. Below the schematic of Runx1 structure are shown proteins that have been described to interact with Runx1, and black lines denote the proposed interaction sites. The list of Runx1 interaction partners is not exhaustive. Adapted from (Koh et al., 2013). In the table below the schematic, the activity of the chromatin proteins identified as Runx1 interaction partners is mentioned. (b) Disorder score for Runx1 (mRunx1b isoform - NCBI reference sequence: NP_001104493.1), calculated using Predictor of Natural Disordered Regions (PONDR). Score above 0.5 indicates regions that are predicted to be disordered. The amino acids corresponding to the Runt domain (RD), activation domain (AD) and inhibitory domain (ID) are highlighted.

The dual function of RUNX1 as an activator and as a repressor has been assumed to depend on the existing chromatin landscape, the different co-factors recruited by RUNX1 (Bonifer et al., 2017), and also on different posttranslational modifications of RUNX1 residues by its interaction partners (Blumenthal et al., 2017). However, in these cellular contexts RUNX1 is in complex with other lineage-specific TFs and as a result, it cannot be excluded that the described role of RUNX1 as an activator or as a repressor is instead dependent on its association with other TFs. In addition, it is unclear how the chromatin environment and more specifically how the histone modifications, accessibility and the set of co-factors bound at a locus affect RUNX1-mediated recruitment of binding partners. Similarly, whether RUNX1 preferentially associates with different co-factors in different chromatin contexts has not been investigated in detail.

The first important step is to verify which proteins and protein complexes can be directly recruited to chromatin by RUNX1 binding. The interaction partners of RUNX1 mentioned above have been identified mainly through co-immunoprecipitation experiments and validated by knockdown assays, where *RUNX1* knockdown is followed by reduced binding of an interaction partner to an endogenous locus. However, co-immunoprecipitation assays have several disadvantages. They lack a chromatin context and it is difficult to separate out weak interactions that are crucial for anchoring proteins to chromatin, from weak non-specific interactions that can occur with proteins in solution. In addition, since they lack a gene context, they can't be used to prove that a TF is directly capable of mediating the recruitment of a gene-regulatory protein to chromatin. Reduction

of the binding of a gene-regulatory protein after the knockdown of a TF is consistent with the possibility that the TF is necessary for the recruitment of that gene-regulatory protein. However, interpretation of such results is complicated by the multitude of parallel regulatory processes taking place at endogenous sites bound by multiple protein complexes, and the possibility of indirect effects from the overall change in transcription. Thus, in order to determine if TF binding, in this case RUNX1, is sufficient to recruit additional protein complexes, neutral experimental approaches that can assess recruitment removed from the complexities of endogenous gene regulation are needed. A powerful system to examine protein recruitment *in vivo* without the complications from endogenous regulatory elements is the TetR system, which was developed by the lab of Rob Klose (Blackledge et al., 2014) and will be described in the next section.

3.1.2 The Tet Repressor (TetR) system

The Tet Repressor (TetR) system has been previously used to determine the recruitment of PcG proteins and MLL-AF4 (Blackledge et al., 2014; Kerry et al., 2017) and allows the investigation of protein interactions in a cellular context. The system takes advantage of the interaction between the bacterial protein TetR and TetO DNA sequences in order to target proteins of interest to a gene-poor region in mESCs. More specifically, fourteen TetO sequences separated by ~20 bp DNA sequences (the TetO locus), flanked by human DNA sequences from a gene desert on human chromosome 7, were introduced into a BAC and integrated in a gene-poor region of chromosome 8 in mESCs (Figure 3.2 a). The TetO locus is surrounded by chromatin that lacks genes, unmethylated CpGs and CpG islands and is not bound by RNA polymerase nor enriched in active/repressive histone modifications (H3K4me3, H3K9ac, H3K36me3, H3K27me3, H2AK119ub1, H4K20me3, and H3K9me3) (Blackledge et al., 2014). As a result, this system can be used to investigate *de novo* protein recruitment in a chromatin context removed from the complications of ongoing transcription and the presence of endogenous regulatory elements.

Fusion of the TetR protein to a protein (or protein domain) of interest (e.g. RUNX1) results in its recruitment to the TetO locus in mESCs; ChIP-qPCR can then be used to detect the presence of potential interacting proteins to determine if the protein of interest is sufficient for their recruitment. Furthermore, as the nucleosomes surrounding the TetO locus are not decorated with active/repressive histone marks, the recruitment of a histone-modifying enzyme can be investigated through examining the presence of histone modifications around the TetO locus. An additional advantage of the system is that upon treatment with doxycycline, the TetR-TetO interaction is disrupted, resulting in reduced binding at the TetO locus of the protein of interest fused to the TetR and its interaction partners (Figure 3.2 b). This further confirms that the presence of a protein at the TetO locus is due to its recruitment by the TetR fusion protein. A cell line expressing the TetR only is also used as a control to validate that the presence of additional proteins is not a result of the binding of TetR itself to the TetO locus.

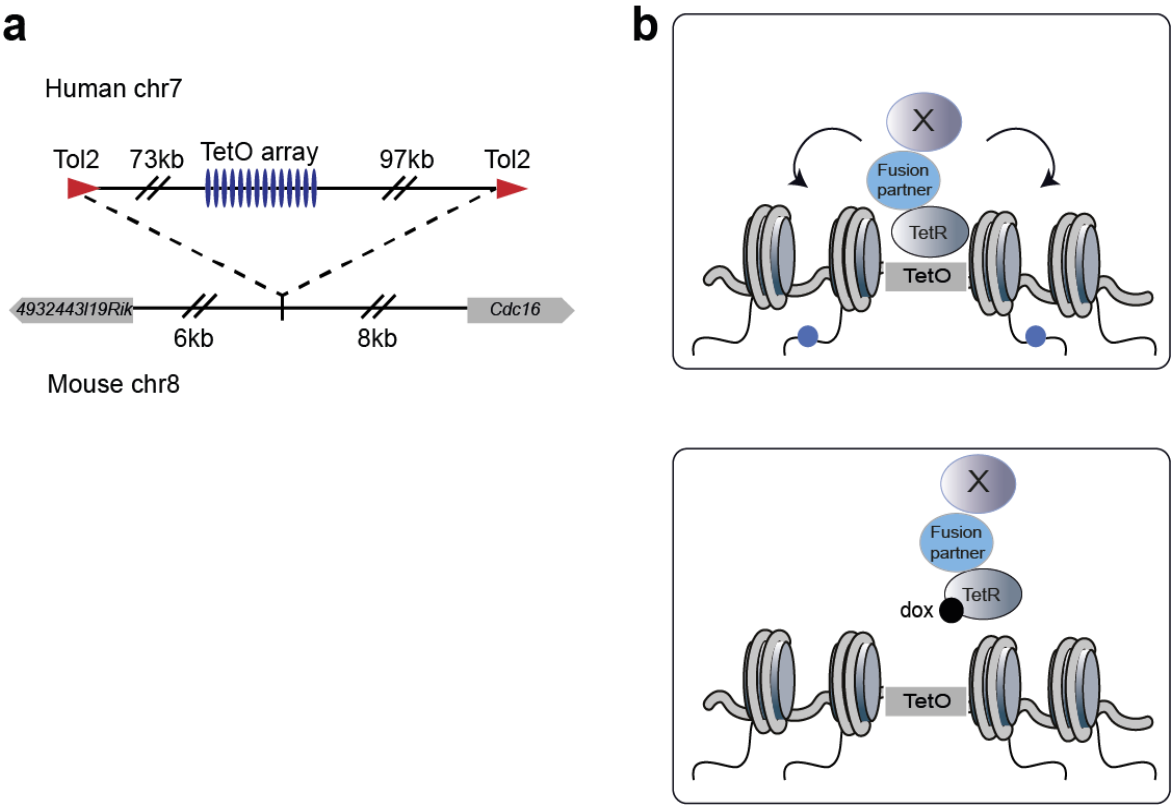


Figure 3.2: The TetR system. (a) Schematic of the TetO array (flanked by human DNA sequences from a gene desert on human chromosome 7) at its integration site on mouse chromosome 8. The figure was adapted from (Blackledge et al., 2014). (b) The protein of interest fused to the TetR

(fusion partner) is localised at the TetO locus. Protein X, an interaction partner of our protein of interest, is also recruited to the TetO locus. Where X is a histone-modifying enzyme, its recruitment to the TetO locus may be indirectly determined by the deposition of the corresponding histone modifications at the nucleosomes flanking the TetO locus (*top panel)*. Upon treatment with doxycycline, the TetR-TetO interaction is disrupted, resulting in reduced binding of the protein of interest and its interaction partners at the TetO locus.

As the detection of interacting proteins in the TetR system is performed by ChIP-qPCR, the results are dependent on antibody quality and crosslinking conditions. As a result, the inability to detect a protein of interest at the TetO locus could be a false negative result due to crosslinking conditions, rather than due to the lack of recruitment. In this case, ChIP-qPCR signals at endogenous control regions can help discriminate between the efficiency of the ChIP experiment and recruitment of the TetO locus. For example, the inability to detect a particular protein at the TetO locus that can be detected at endogenous control loci, supports the conclusion that this protein cannot be recruited by our protein of interest at the TetO region or that is recruited at very low levels. Furthermore, the fusion of a protein or protein domain of interest to the TetR could result to protein misfolding and inability to detect its interaction partners. In this case, ChIP for known interaction partners of the protein of interest can be used to confirm its correct folding. Another limitation of the TetR system is that it cannot reveal whether the interaction partners of our protein of interest are recruited through direct or indirect interactions. For this question to be adequately answered, other experimental methods that detect protein-protein interactions are needed, for example the Biacore system which also provides information about the binding affinity of the interaction.

3.2 Aims

- To determine the chromatin complexes that can be recruited by RUNX1
- To determine whether the chromatin environment affects RUNX1-mediated recruitment of co-factors

3.3 Results

3.3.1 Runx1 binding is sufficient to recruit Trithorax and Polycomb proteins and generate a Polycomb domain.

Firstly, a plasmid consisting of mouse Runx1b (mRunx1b) fused to the N-terminal of TetR (mRunx1b-TetR) was inserted into mESCs containing the TetO locus on chromosome 8. As Runx1b, consisting of 451 amino acids, shares approximately 96% amino acid identity to the human RUNX1b isoform (Protein Blast), it is expected that the interaction partners of mRunx1b can be extrapolated for human RUNX1. Stable cell lines expressing the mRunx1b-TetR fusion protein were created (Figure 3.3 a), and ChIP-qPCR confirmed binding of TetR, and Runx1 at the TetO locus in the Runx1-TetR cells (Figure 3.3 b,c). The ChIP-qPCR signal for these proteins showed that Runx1-TetR is localised at the TetO locus, and ChIP-qPCR signal for these proteins is significantly reduced at a distance of 0.5 kb from the TetO locus (Figure 3.3 b,c). In contrast, in the TetR-only control cell lines TetR was detected by ChIP-qPCR at the TetO locus, but Runx1 (which is PcG-repressed in mESCs and expressed at very low levels) (Figure 3.4 a) was not detected (Figure 3.3 b,c).

To test whether Runx1 fused to the TetR is functional and can interact with its partners, we performed ChIP-qPCR for its co-factor Cbfb in Runx1-TetR cells and TetR-only cells, and showed that Cbfb is only recruited to the TetO locus of Runx1-TetR cells (Figure 3.3 d). This observation suggests that Runx1 is functional and sufficient to recruit its interaction partners. Furthermore, recruitment of TetR, Runx1 and Cbfb to the TetO locus in Runx1-TetR mESCs was lost upon treatment with doxycycline for 6h (Figure 3.3 b,c,d). Having established a system where the recruited Runx1-TetR fusion protein at the TetO locus is functional, I next went on to test whether Runx1 is sufficient to recruit activators and/or repressors. It should be noted that the experiments are performed in undifferentiated mESCs, in which genes encoding important haematopoietic TFs that cooperate with Runx1 in haematopoietic blood and progenitor cells (Wilson et al., 2010) are

PcG-repressed and expressed at low levels or not expressed. This allows the effects of Runx1

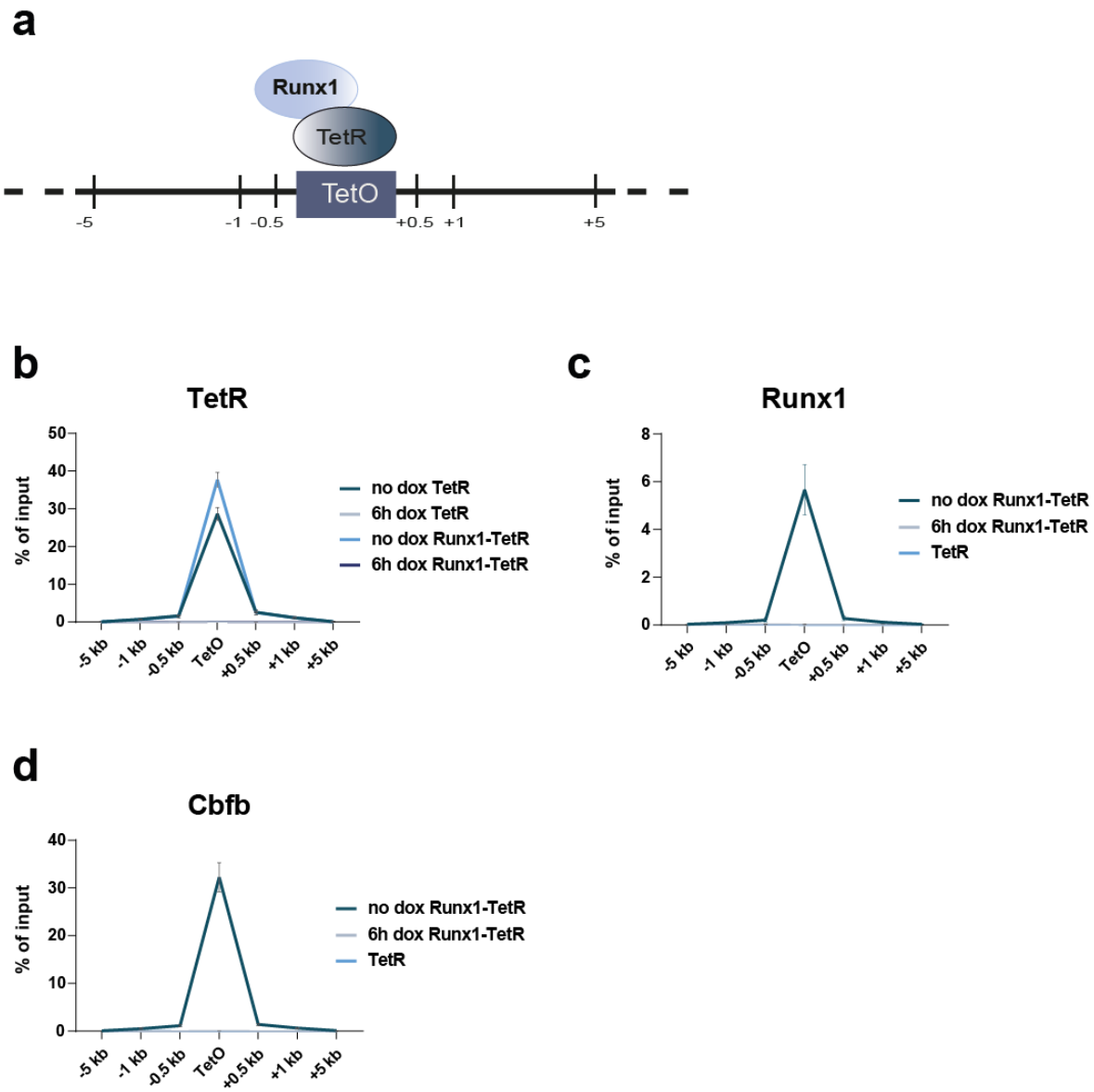


Figure 3.3: Validation of the Runx1-TetR system. (a) Schematic diagram of the TetO locus in mESCs expressing the Runx1-TetR fusion protein. Numbers indicate the qPCR primer positions (in kbs) with respect to the TetO locus. (b) ChIP-qPCR for TetR (antibody recognising the TetR FLAG STREPx2 (FS2) tag), (c) Runx1 and (d) Cbfb at the TetO locus of Runx1-TetR mESCs with/without treatment with doxycycline (grey and blue lines respectively) and cells expressing only the TetR (light blue line), showing recruitment of TetR, Runx1 and Cbfb at the TetO locus of Runx1-TetR cells, and recruitment only of TetR (but not Runx1 and Cbfb) at the TetO locus of TetR-only cells. Numbers on the x axis show the qPCR primer position with respect to the TetO locus. ChIP experiments were performed in 3 biologicals replicates and error bars show the standard error of the mean (S.E.M.).

binding, rather than the effects of binding of the protein complexes consisting of Runx1 and additional haematopoietic TFs, to be determined. Indeed, nascent RNA-seq results in Runx1-TetR

compared to TetR-only cells showed that the genes encoding the TFs Lyl1, Lmo2, Erg, Fli1 and Gfi1b, previously described as Runx1 interaction partners (Wilson et al., 2010), are not expressed. In addition, TFs Tal1, Gata2, Meis1 and Sfpi1 are expressed at very low levels in TetR cells compared to highly expressed genes (e.g. *Sox2)*, and their expression in Runx1-TetR cells is further reduced (Figure 3.4 b).

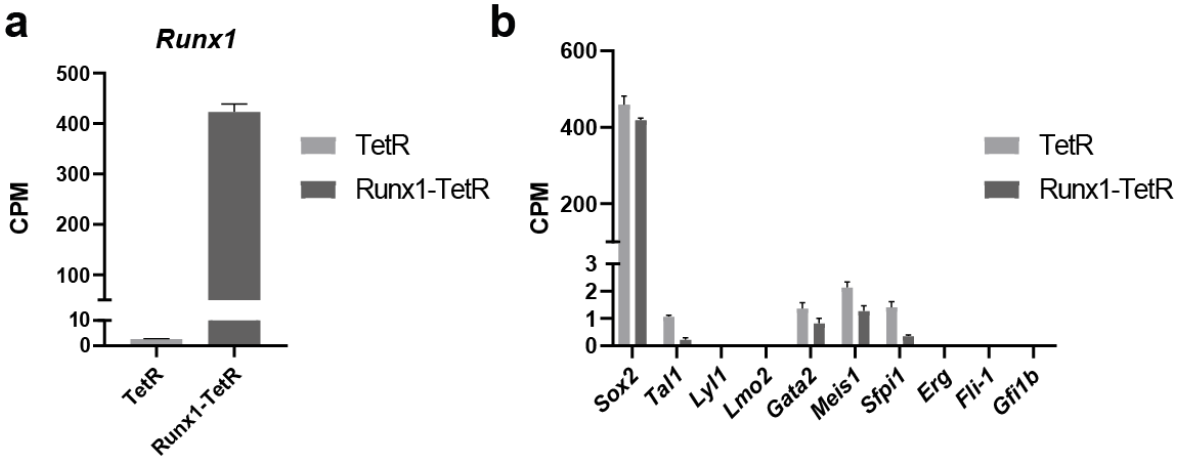


Figure 3.4: Runx1 and interacting TFs in haematopoietic stem and progenitor cells are expressed at low levels in mESCs. (a) Nascent RNA-seq results showing *Runx1* expression levels in TetR-only (light grey) and Runx1-TetR mESCs (dark grey). (b) Nascent RNA-seq results showing expression levels of genes encoding key haematopoietic TFs (*Tal1, Lyl1, Lmo2, Gata2, Meis1, Sfpi1, Erg, Fli1 and Gfi1b*) compared to a highly expressed gene (*Sox2)* in TetR-only (light grey) and Runx1-TetR mESCs (dark grey). These TFs are found in the same complex as Runx1 in blood stem cells and progenitor cells (Wilson et al., 2010), and in mESCs they are PcG-repressed and expressed at very low levels or not expressed. Genes encoding TFs that are expressed at low levels in TetR-only mESCs (*Tal1, Gata2, Meis1, Sfpi1)* are further repressed in Runx1-TetR mESCs. Nascent RNA-seq experiments were performed in 3 biological replicates. Error bars represent the standard error of the mean (S.E.M.).

Next, in order to validate previously published interactions and determine if RUNX1 binding correlated with factors from both activation and repression complexes, I performed ChIP-seq correlation analysis at RUNX1 peaks using ChIP-seq data that had been previously generated in our lab for TFs, chromatin-modifying enzymes and histone modifications in an MLL-AF4 ALL cell line (in Appendix table 7.1 the accession numbers of published data, and information about unpublished data is included). It has been previously shown that *RUNX1* is a direct target of the MLL-AF4 fusion

protein and it promotes survival of the MLL-AF4 leukaemic cells (Harman et al., 2021; Wilkinson et al., 2013). ChIP-seq correlation analysis showed that RUNX1 binding correlates with the binding of co-activators (BRD4, TrxG proteins MLL and BRG1) and active histone modifications and it also correlates with the binding of Polycomb repressor RING1B (Figure 3.5 a,b).

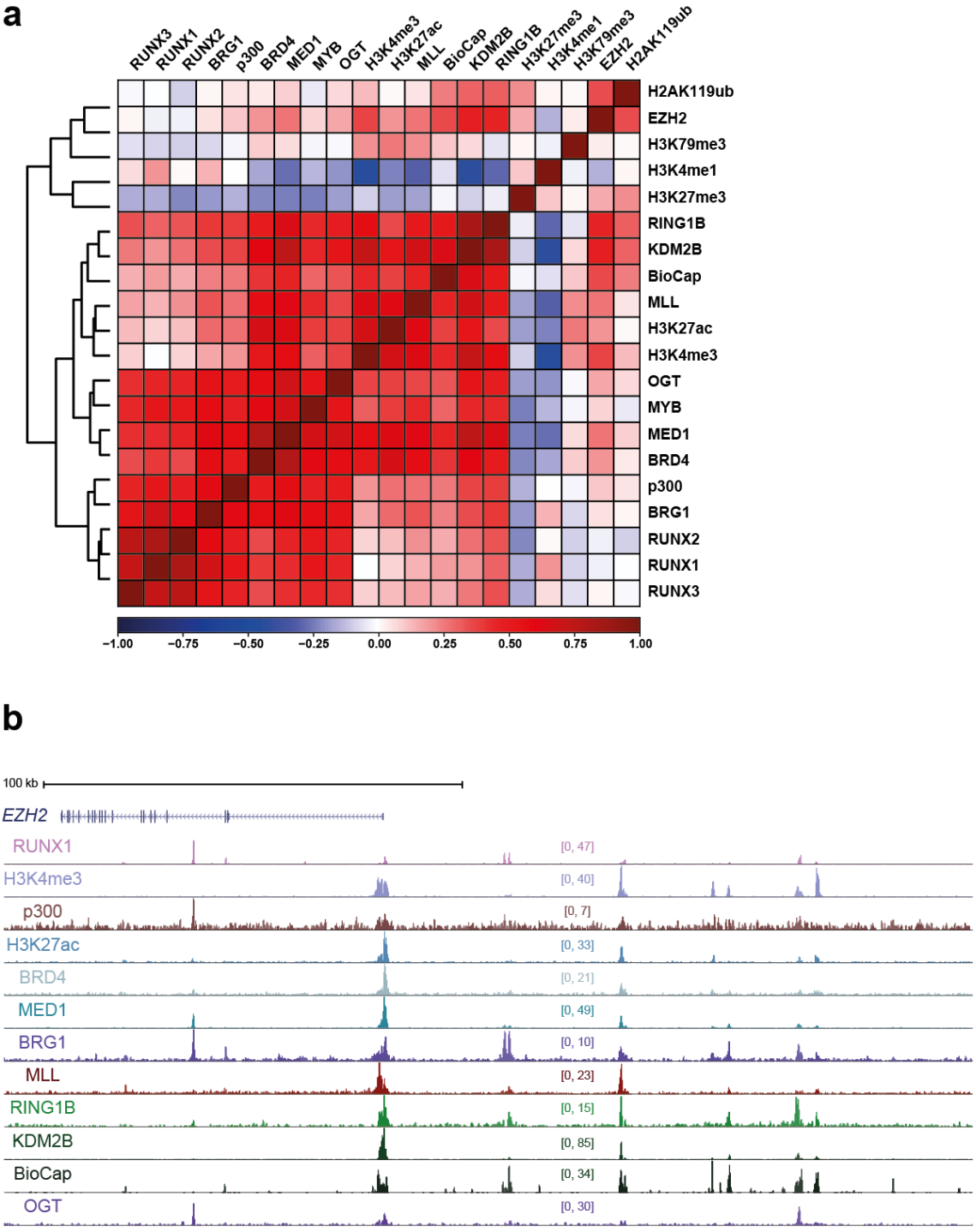


Figure 3.5: ChIP-seq correlation analysis in MLL-AF4 cells reveals that RUNX1 binding correlates with the binding of activators and repressors. (a) Correlation analysis for ChIP-seq data (ChIP-seq

for RUNX1, RUNX2, RUNX3, RING1B, KDM2B, MLL, BRG1, BRD4, MED1, MYB, p300, EZH2, H3K27ac, H3K27me3, H3K4me3, H3K4me1, H2AK119ub, H3K79me3, OGT) and Bio-Cap data at RUNX1 peaks in the MLL-AF4 cell line SEM, showing that RUNX1 binding correlates with the binding of activators (MLL, BRD4, p300, OGT component of the BAP1 deubiquitinase complex) and active histone modifications (H3K27ac, H3K4me3), and also with Polycomb repressor RING1B (please refer to Appendix table 7.1 for accession numbers of published datasets and for information for unpublished ones). (b) UCSC ChIP-seq tracks for RUNX1, H3K4me3, p300, H3K27ac, BRD4, MED1, BRG1, MLL, RING1B, KDM2B, BioCap and OGT at the *EZH2* locus.

In order to precisely determine whether RUNX1 can recruit activators and repressors, and validate whether RUNX1 is sufficient to recruit the previously described interaction partners and the proteins identified in the correlation analysis, the TetR system was employed. Firstly, to check whether Runx1 is sufficient to recruit TrxG and/or PcG proteins, I performed ChIP-qPCR for these proteins using the Runx1-TetR mESCs. Indeed, ChIP-qPCR for Brg1, a component of the SWI/SNF complex (Wang et al., 1996c), showed that it is recruited to the TetO locus of Runx1-TetR mESCs, but not TetR-only cells (Figure 3.6 a). Upon treatment with doxycycline for 6h, reduction in Runx1-TetR binding at the TetO locus resulted in reduced binding of Brg1. Similarly to Runx1 and CbFb binding, the Brg1 ChIP signal is higher at the TetO locus and drops significantly as the distance from the TetO locus increases (Figure 3.6 a).

In addition, by performing ChIP-qPCR at the TetO locus for Mll, which is expressed at low levels in mESCs, I was able to show that Runx1 is sufficient to recruit Mll to the TetO locus of Runx1-TetR cells (Figure 3.6 b), and further confirmed this by showing that treatment with doxycycline results in reduced Mll binding to the TetO locus of Runx1-TetR mESCs. To test whether Runx1-mediated recruitment of Mll is sufficient to result in an increase in the levels of H3K4me3 at the nucleosomes surrounding the TetO locus, I performed ChIP-qPCR for H3K4me3 in Runx1-TetR cells compared to TetR-only mESCs, which showed a slight increase in the levels of H3K4me3 at the TetO locus of Runx1-TetR mESCs (Figure 3.6 c). However, the levels of H3K4me3 at the TetO locus are significantly lower compared to the H3K4me3 levels at the promoter of *Oct4*. The low levels of H3K4me3 at the

TetO locus are probably due to the fact that Mll is expressed at very low levels in mESCs, while at

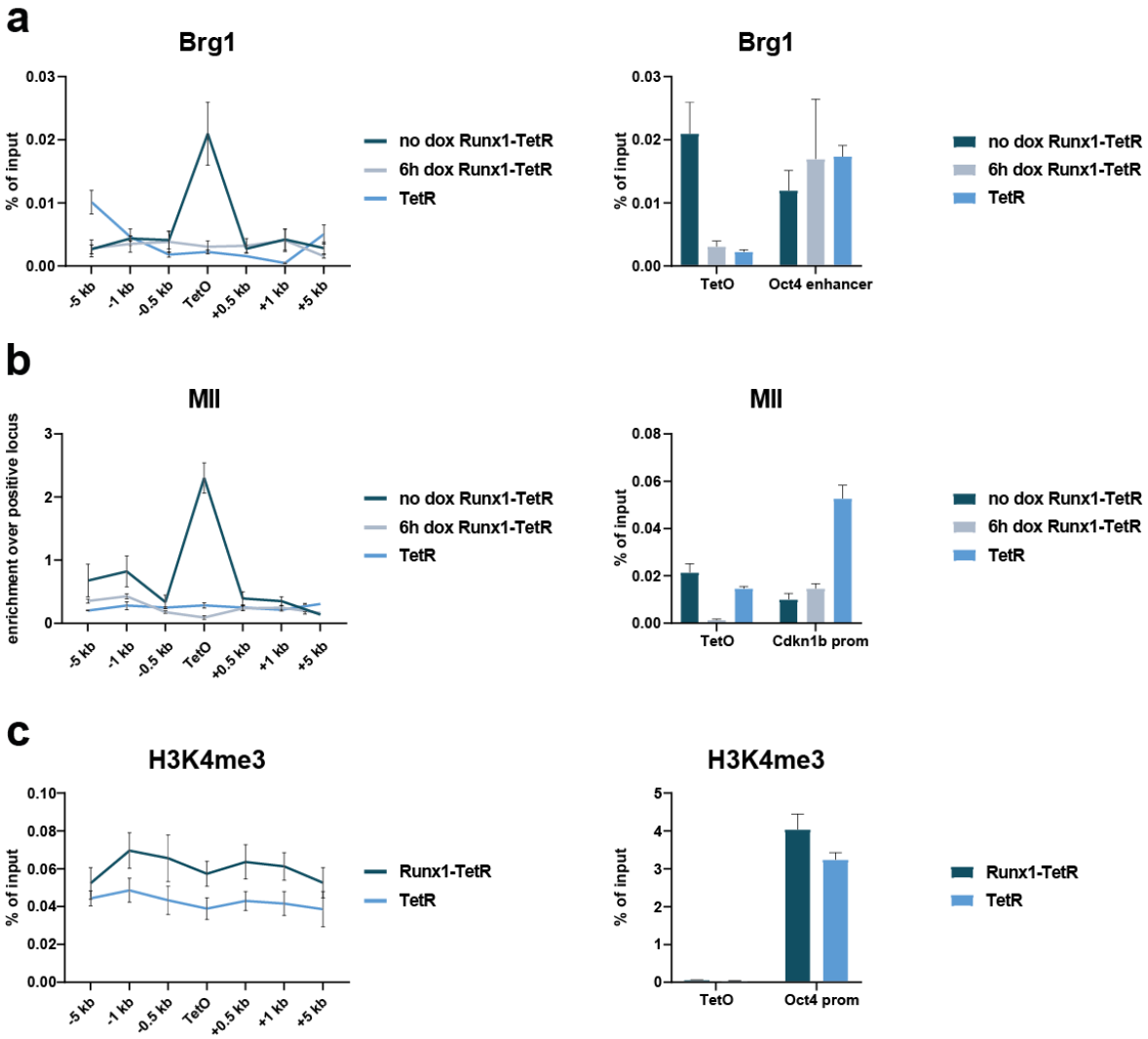


Figure 3.6: Runx1 is sufficient to recruit Trithorax-group proteins Brg1 and Mll. (a) ChIP-qPCR for Brg1, component of the SWI/SNF complex at the TetO locus of Runx1-TetR cells with/without doxycycline treatment and TetR-only cells showing that Runx1 is sufficient to recruit Brg1. (b) ChIP-qPCR for Mll at the TetO locus of Runx1-TetR cells with/without doxycycline treatment and TetR-only cells, showing that Runx1 is sufficient to recruit the Mll methyltransferase. (c) ChIP-qPCR for H3K4me3 at the TetO locus of Runx1-TetR cells and TetR-only cells showing that Runx1-mediated recruitment of Mll results in a slight increase in H3K4me3. ChIP experiments were performed in 3 biologicals replicates and error bars show the standard error of the mean (S.E.M.).

endogenous loci other H3K4 methyltransferases that are more highly expressed, e.g. Setd1a and Setd1b, could be responsible for the deposition of this histone mark (Cenik and Shilatifard, 2021).

As ChIP-seq correlation analysis in MLL-AF4 cells showed that RUNX1 binding also correlates with the binding of the RING1B repressor, the catalytic component of PRC1 that catalyses H2AK119ub (Li et al., 2006; Wang et al., 2004a), I used the TetR system to examine whether RUNX1 is sufficient for the recruitment of RING1B, or whether their chromatin co-occupancy in MLL-AF4 cells is not due to RUNX1-mediated recruitment of PRC1. Indeed, ChIP-qPCR for Ring1b in Runx1-TetR mESCs showed that Ring1b is detected at the TetO locus of Runx1-TetR mESCs, but not TetR-only mESCs. Additionally, its recruitment is lost upon treatment with doxycycline which abolishes the binding of Runx1-TetR to the TetO locus (Figure 3.7 a). Ring1b recruitment is followed by an increase in the levels of H2AK119ub at the nucleosomes flanking the TetO locus of Runx1-TetR cells compared to cells expressing the TetR only (Figure 3.7 b). More specifically, the increase in H2AK119ub is mainly observed at the nucleosomes that are located 0.5 kb and 1 kb around the TetO locus. In contrast, as expected, at the nucleosome-free TetO locus (Blackledge et al., 2014), there was no increase in H2AK119ub. In addition, treatment with doxycycline for 6h reduced the levels of H2AK119ub in Runx1-TetR cells, suggesting that maintenance of H2AK119ub levels is a dynamic process dependent on the activity of PRC1 and the counteracting activity of deubiquitinase enzymes.

As it has been extensively shown that H2AK119ub is recognised by PRC2.2 complexes (Blackledge et al., 2014; Cooper et al., 2014; Cooper et al., 2016; Kalb et al., 2014; Kasinath et al., 2021), I checked whether the increased levels of H2AK119ub resulted in binding of PRC2 components in the Runx1-TetR system. This showed enrichment of Ezh2 binding around the TetO locus (Figure 3.7 c), and the binding pattern of Ezh2 matched the enrichment of H2AK119ub, i.e. Ezh2 is mostly bound to the nucleosomes located 1 kb away from the TetO locus, where the H2AK119ub levels are higher. The pattern of Ezh2 binding further suggests that, unlike Ring1b, Runx1 is not sufficient to recruit Ezh2, and its recruitment to the TetO locus is secondary, following Ring1b-mediated deposition of H2AK119ub. Ezh2 binding at the TetO locus is a lot lower compared to endogenous sites (Figure 3.7 c), despite the fact that H2AK119ub levels at the TetO locus are comparable to its levels at endogenous PcG-repressed loci (Figure 3.7 b). This suggests that H2AK119ub levels at the

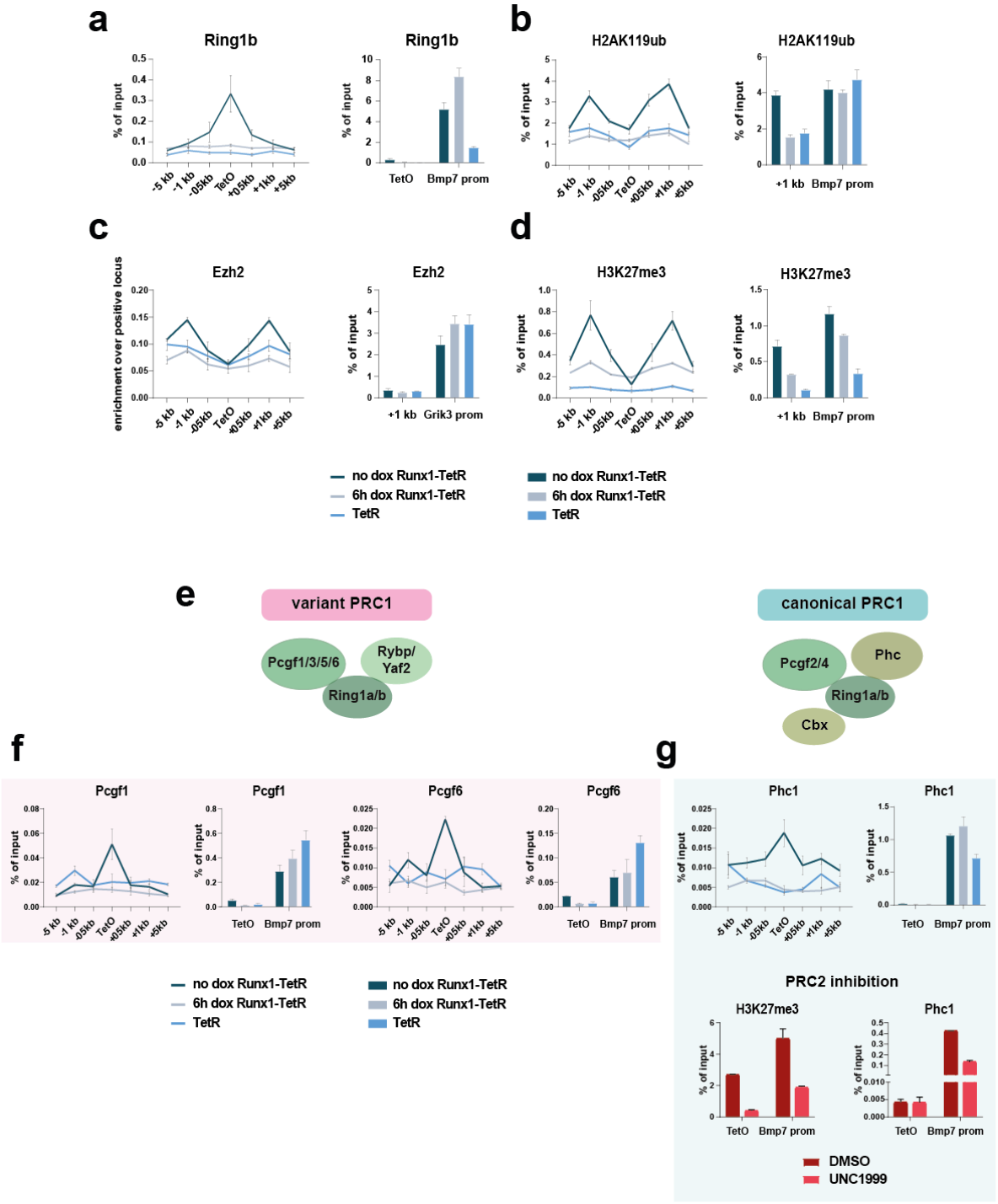


Figure 3.7: Runx1 is sufficient to recruit PRC1 and generate a Polycomb domain at the TetO locus. ChIP-qPCR for Ring1b (a) and H2AK119ub (b) at the TetO locus of Runx1-TetR cells with/without doxycycline treatment (grey and dark blue lines respectively) and TetR-only cells (light blue line), showing that Runx1 is sufficient to recruit Ring1b which deposits the H2AK119ub mark at the nucleosomes flanking the TetO locus. (c) ChIP-qPCR for Ezh2 at the TetO locus shows enrichment for Ezh2 binding at nucleosomes decorated with H2AK119ub around the TetO locus. (d) ChIP-qPCR for H3K27me3 at the TetO locus shows that Ezh2 in turn catalyses H3K27me3 at nucleosomes flanking the TetO locus. This suggests that tethering Runx1 at the TetO locus is sufficient to generate a Polycomb domain. (e) Schematic diagram of the core components of the variant and the canonical PRC1 complexes. (f) ChIP-qPCR for Pcgf1, Pcgf6 (components of variant complexes PRC1.1 and PRC1.6 respectively) in Runx1-TetR mESCs (with/without doxycycline treatment) and

in TetR-only mESCs, showing that Runx1 is sufficient to recruit variant PRC1 complexes. (g)-*top panel:* ChIP-qPCR for Phc1 (component of canonical PRC1 complexes) in Runx1-TetR mESCs (with/without doxycycline treatment) and in TetR-only mESCs showing that Runx1 is sufficient to recruit canonical PRC1 complexes. *Bottom panel*: ChIP-qPCR for H3K27me3 and Phc1 in Runx1-TetR mESCs at the TetO locus and an endogenous locus after treatment with the PRC2 inhibitor UNC1999 (5 μM for 120h), showing that reduction in H3K27me3 does not result in reduced Phc1 binding at the TetO locus. ChIP-qPCR experiments were performed in 3 biologicals replicates and error bars show the standard error of the mean (S.E.M.). ChIP-qPCR at endogenous loci is also shown for comparison of ChIP signal.

TetO locus are sufficient to recruit Ezh2, but Ezh2 seems to bind transiently in a 'hit-and-run' fashion, unlike at endogenous PcG-repressed sites where additional elements like CpG-islands contribute to stable PRC2 binding. In turn, PRC2 binding at the nucleosomes flanking the TetO locus in Runx1-TetR cells is sufficient for the deposition of H3K27me3, which is increased compared to TetR-only mESCs (Figure 3.7 d). Treatment with doxycycline reduced H3K27me3 levels, following the reduction in H2AK119ub levels and the reduction of Ezh2 binding. Overall, these data support a model whereby tethering Runx1 at this gene-poor region, which lacks CpG islands and unmethylated CpGs, is sufficient to recruit PRC1 which deposits H2AK119ub, in turn recognised by PRC2 which catalyses H3K27me3. Thus, tethering Runx1 at the TetO locus leads to the generation of a Polycomb domain that spreads several kbs from the anchoring site, in accordance to the hierarchical model of PcG recruitment (Blackledge et al., 2014; Cooper et al., 2014; Fursova et al., 2019).

I then decided to investigate whether Runx1 is sufficient to recruit variant and/or canonical PRC1 complexes (Figure 3.7 e). As discussed in the introduction, variant and canonical PRC1 complexes have different complex composition and mechanisms of recruitment (Gao et al., 2012). More specifically, in canonical PRC1 complexes the catalytic subunits Ring1a/b associate with Cbx and Phc proteins, bind to nucleosomes decorated with H3K27me3 (Bernstein et al., 2006b; Cao et al., 2002; Fischle et al., 2003; Min et al., 2003) and mediate local chromatin compaction and long-range PcG interactions through the polymerisation of Cbx and Phc subunits (Isono et al., 2013; Kim et al., 2002; Kim and Kingston, 2020; Kundu et al., 2017; Lau et al., 2017; Wani et al., 2016). In contrast,

in variant PRC1 complexes, Ring1a/b associates with Rybp/Yaf2 (Gao et al., 2012) and the complexes are recruited to chromatin through the association of specific subunits with unmethylated CpG islands (through the Kdm2b subunit of PRC1.1 complex (Farcas et al., 2012; He et al., 2013; Wu et al., 2013b)) or through interaction with TFs (TFs E2F6-DP1/MGA-MAX for PRC1.6 complex (Endoh et al., 2017; Scelfo et al., 2019; Stielow et al., 2018; Trojer et al., 2011)). In order to test whether Runx1 can recruit variant and/or canonical PRC1 complexes, I performed ChIP-qPCR at the TetO locus for variant PRC1 proteins PCGF1 and PCGF6 (components of PRC1.1 and PRC1.6 complexes respectively) and for Phc1 protein, a component of canonical PRC1 complexes, which together with the other Phc proteins (Phc2/3) is responsible for mediating long-range interactions between PcG-occupied sites (Isono et al., 2013; Kundu et al., 2017). My results show that Runx1 is indeed sufficient to recruit all these PRC1 sub-complexes at the TetO locus, (albeit at lower levels compared to PcG-repressed endogenous loci) (Figure 3.7 f,g). Interestingly, the pattern of Phc1 binding at the TetO locus suggests that it is bound mainly to the nucleosome-free TetO locus and to a lesser degree to the H3K27me3-decorated nucleosomes flanking the TetO locus (Figure 3.7 g, *top panel*). In addition, reduction in H3K27me3 following treatment with the PRC2 inhibitor UNC1999, which inhibits the catalytic activity of PRC2, does not lead to a reduction in Phc1 binding at the TetO locus (in contrast to endogenous loci, where the reduction in H3K27me3 levels is followed by a reduction in Phc1 binding), suggesting that the recruitment of canonical PRC1 complexes to the TetO locus is largely independent of H3K27me3 (Figure 3.7 g, *bottom panel*). Overall, these data suggest Runx1-mediated recruitment of variant PRC1 complexes and of canonical PRC1 complexes is independent of H3K27me3.

My ChIP-seq correlation analysis in MLL-AF4 cells showed correlation of RUNX1 binding not only with RING1B, but also with OGT, which has been described to be component of many complexes (Yang and Qian, 2017), including the BAP1 deubiquitinase complex (Misaghi et al., 2009; Ruan et al., 2012) that functions by removing H2AK119ub (Scheuermann et al., 2010). The BAP1 complex antagonises PcG activity through removal of H2AK119ub, and recently it has been also shown that

BAP1 activity prevents the widespread accumulation of H2AK119ub that leads to inappropriate gene silencing and compromises the proper localisation of PcG proteins (Conway et al., 2021; Fursova et al., 2021). To determine whether RUNX1 contributes to the recruitment of this complex, I performed ChIP-qPCR for complex components Bap1 and Ogt at the TetO locus of Runx1-TetR mESCs, and showed that Runx1 is indeed sufficient to recruit the Bap1 deubiquitinase complex (Figure 3.8 a,b). These results demonstrate that Bap1 and Ogt are bound to the TetO locus, rather than to the nucleosomes decorated with H2AK119ub flanking the TetO region. In addition, reduction of Bap1 and Ogt binding to the TetO locus of Runx1-TetR mESCs after doxycycline treatment, and their lack of binding to the TetO locus of TetR-only cells further validates that they are recruited through their (direct or indirect) association with Runx1. Despite the fact that in mESCs it has been reported that PcG complexes and the Bap1 complex are bound at distinct loci (Conway et al., 2021), my data support Runx1-mediated recruitment of both complexes which have opposite activities. At the TetO locus, which lacks ongoing transcription and activating marks, PcG proteins prevail and their binding leads to the establishment of a repressive Polycomb domain.

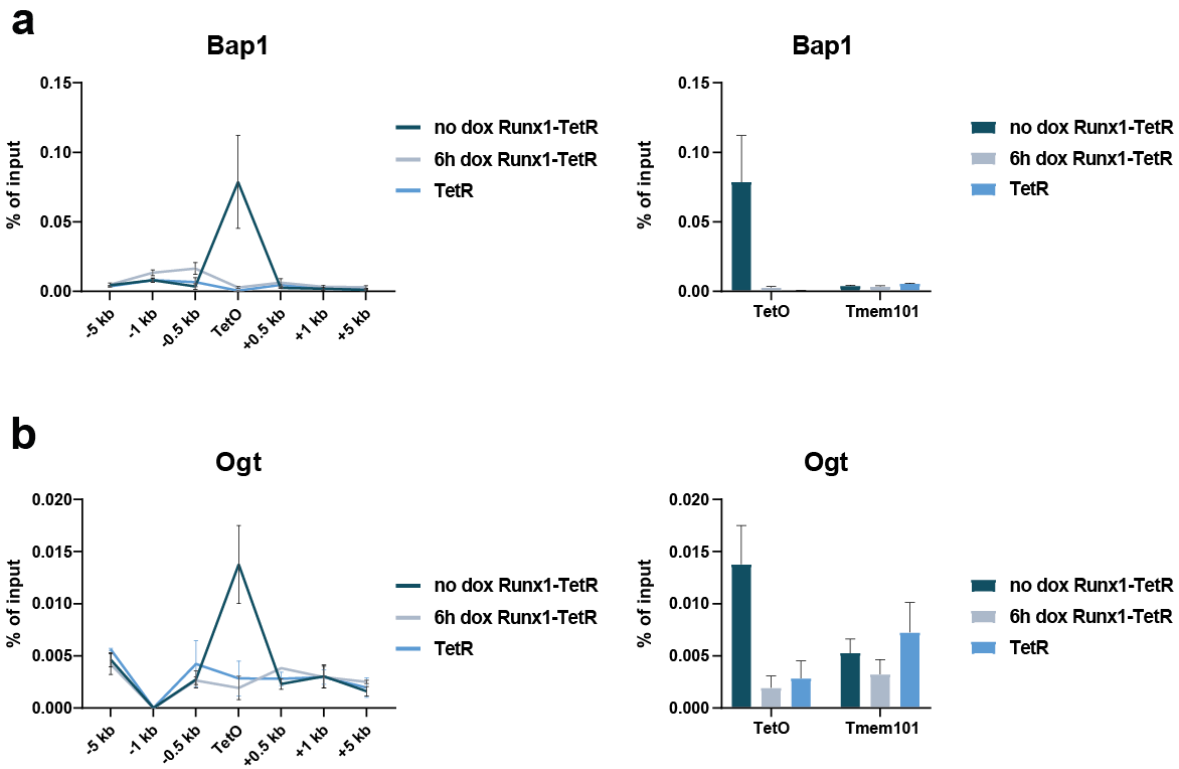


Figure 3.8: Runx1 is sufficient to recruit the Bap1 deubiquitinase complex. ChIP-qPCR for Bap1 (a) and Ogt (b) at the TetO locus of Runx1-TetR cells with/without doxycycline treatment (grey and blue lines respectively) and TetR-only cells (light blue line), showing that Runx1 is sufficient to recruit the Bap1 deubiquitinase complex. ChIP-qPCR experiments were performed in 3 biologicals replicates and error bars show the standard error of the mean (S.E.M.).

3.3.2 Runx1 is sufficient to recruit co-activators and the transcriptional machinery.

To determine whether Runx1 is sufficient to recruit other co-activators, in addition to TrxG complexes, I went on to perform ChIP-qPCR for p300 in Runx1-TetR cells and TetR-only cells. p300 was previously reported to be an interaction partner of Runx1 (Kitabayashi et al., 1998). My results show that Runx1 is sufficient to recruit p300, and in fact, the levels of p300 binding at the TetO locus are higher compared to endogenous p300-bound active regions (Figure 3.9 a). Treatment with doxycycline abolishes the binding of p300, confirming that its presence at the TetO locus is due to Runx1-mediated recruitment. To test whether the binding of p300 results in the deposition of acetyl marks on the histones at the nucleosomes flanking the TetO locus, I performed ChIP for H3K27ac and H4ac (with an antibody recognising acetylated K5, K8, K12 and K16 residues on histone H4) in Runx1-TetR cells and TetR-only cells. Surprisingly, ChIP-qPCR for these marks showed that, despite the recruitment of p300, the levels of H3K27ac and H4ac at the TetO locus of Runx1-TetR cells are not increased compared to TetR-only cells, in which p300 is not recruited to the TetO locus (Figure 3.9 b,c).

To test whether lack of acetylation was due to the counteracting function of Hdacs, which function by removing acetyl marks, ChIP for Hdac2 was performed. HDACs have been previously described as interaction partners of RUNX1 (Reed-Inderbitzin et al., 2006), and ChIP-qPCR at the TetO locus revealed that Hdac2 is bound to the TetO locus of Runx1-TetR mESCs. In addition, upon treatment with doxycycline Hdac2 binding decreased, confirming that Runx1-TetR mediates its binding at the TetO locus (Figure 3.9 d).

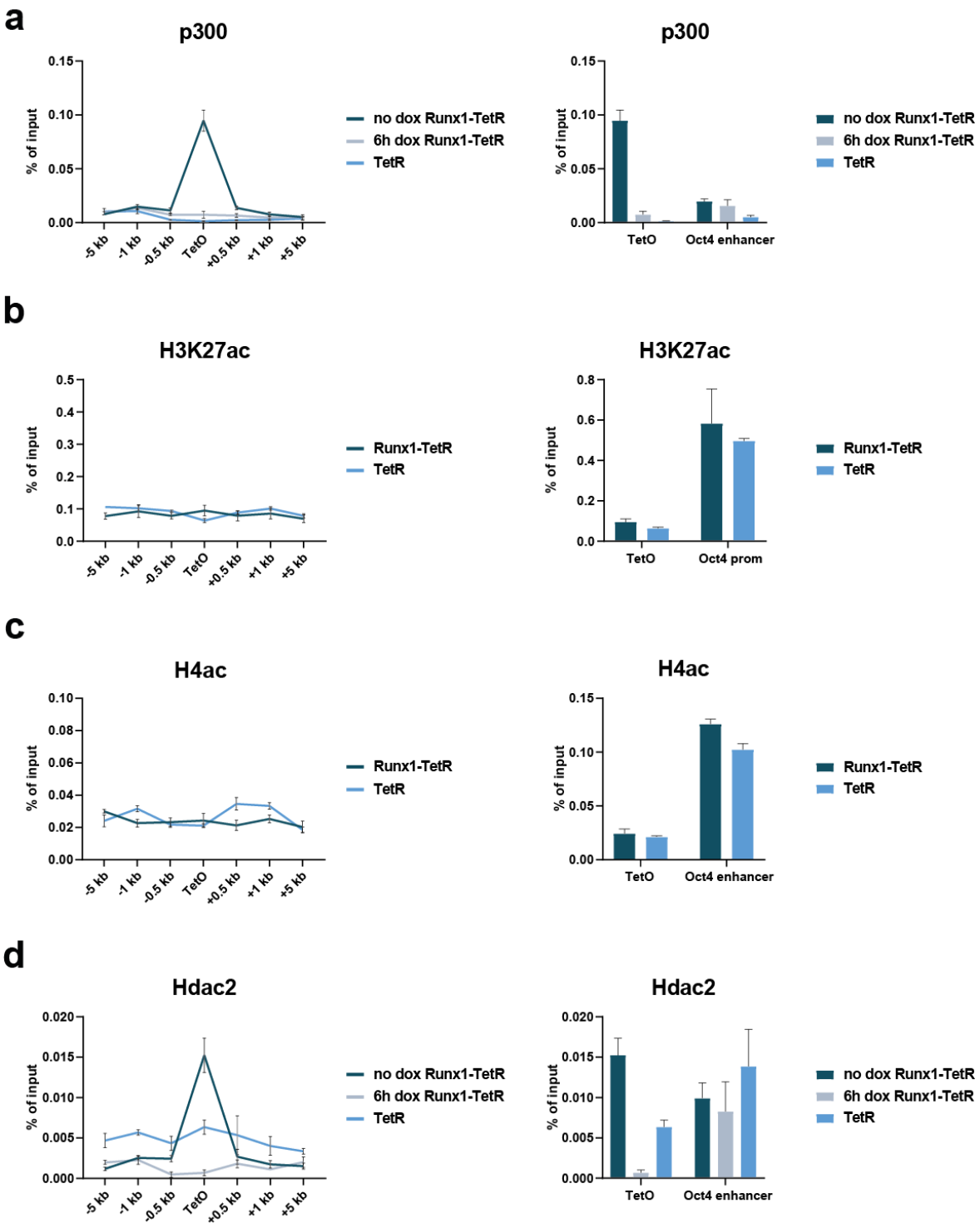


Figure 3.9: Runx1 is sufficient for the recruitment of p300 and Hdacs, but there is no enrichment in histone acetylation at the TetO locus. (a) ChIP-qPCR for p300 at the TetO locus of Runx1-TetR cells with/without dox treatment and in TetR-only cells shows that Runx1 is sufficient to recruit p300. (b) ChIP-qPCR for H3K27ac and (c) H4ac (with an antibody that recognises acetylated H4K5, K8, K12 and K16 residues) in Runx1-TetR and TetR-only mESCs does not show enrichment in histone acetylation at the TetO locus of Runx1-TetR mESCs. (d) ChIP-qPCR for Hdac2 at the TetO locus of Runx1-TetR cells with/without dox treatment and in TetR-only cells shows that Runx1 is sufficient to recruit Hdacs. ChIP experiments were performed in 3 biologicals replicates and error bars show

the standard error of the mean (S.E.M.). ChIP-qPCR at endogenous loci is also shown for comparison of ChIP signal.

To investigate whether the activity of Hdacs causes the lack of acetylation, despite the recruitment of p300 at significant levels, Runx1-TetR mESCs were treated for 24h with 100 ng/mL of Trichostatin A (TSA), which inhibits the activity of Hdacs (Vanhaecke et al., 2004). If treatment with TSA resulted in an increase in the levels of acetylation at the TetO locus, this would suggest that it is the Runx1-mediated recruitment of Hdacs that leads to the removal of the acetyl marks deposited by p300, and that the balance between these two complexes with counteracting activities leans towards removal of acetyl marks by Hdacs. TSA treatment of Runx1-TetR mESCs caused a significant increase in acetylation genome-wide, as revealed by Western blot for H3K27ac and H4ac (Figure 3.10 c), and ChIP-qPCR for these histone marks showed their increase at endogenous loci after TSA treatment (Figure 3.10 a,b). However, no increase in the acetylation levels at the TetO locus was seen (Figure 3.10 a,b) suggesting that other mechanisms account for the lack of acetylation at the TetO locus, rather than the function of Hdacs.

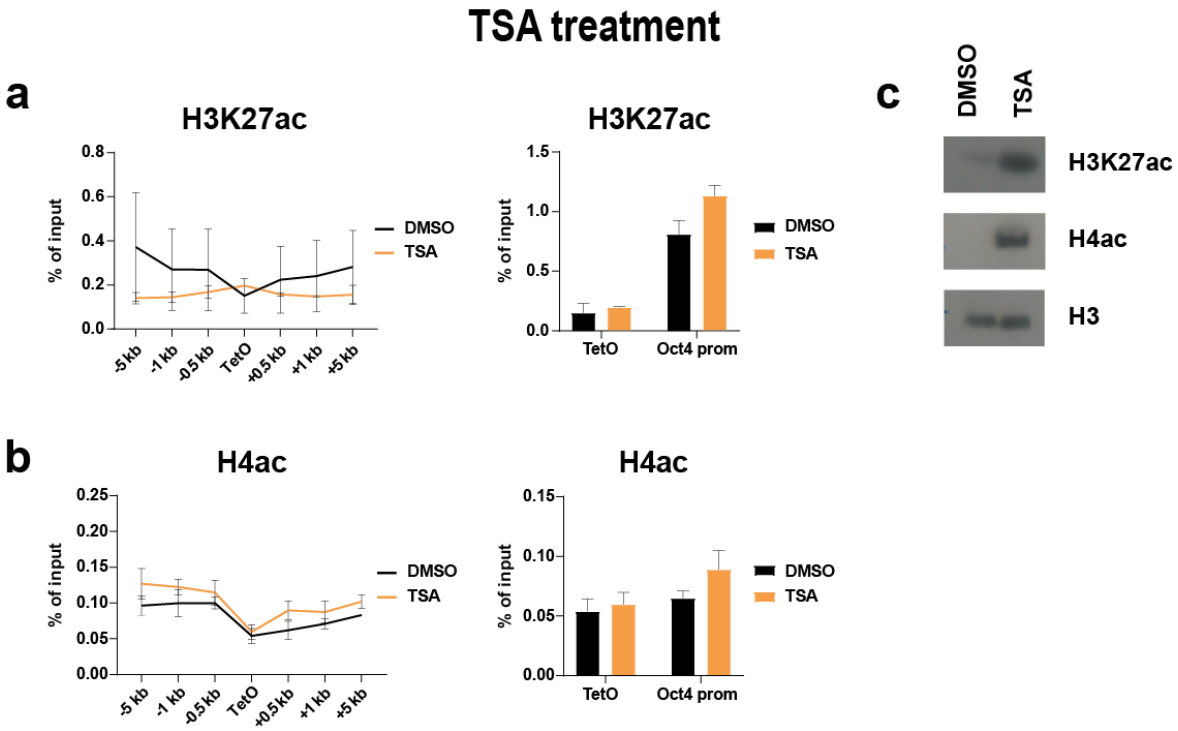


Figure 3.10: Inhibition of Hdac activity in Runx1-TetR cells does not result in increased acetylation levels at the TetO locus. ChIP-qPCR for H3K27ac (a) and H4ac (b) at the TetO locus of Runx1-TetR

mESCs treated with 100 ng/mL TSA for 24h or DMSO shows that inhibition of the Hdac activity does not result in increased acetylation levels at the TetO locus. ChIP experiments were performed in 2 biologicals replicates (for H3K27ac) and 3 biological replicates (H4ac). ChIP-qPCR at endogenous loci is also shown for comparison of ChIP signal. (c) Western blot for H3K27ac, H4ac and H3 in Runx1-TetR mESCs after TSA treatment or DMSO control.

To further investigate the lack of histone acetylation, we aimed to test whether the presence of PcG proteins inhibits the deposition of acetyl marks by acetyltransferases. It has been shown that Polycomb proteins inhibit the catalytic activity of acetyltransferases (Tie et al., 2016), and recently it has been demonstrated that the widespread accumulation of H2AK119ub results in the reduction of the transcription-associated histone modifications H3K27ac and H3K4me3 (Fursova et al., 2021). To examine whether the catalytic activity of Polycomb proteins causes the lack of acetylation, I used the proteasome inhibitor MG132 which functions by reducing the degradation of ubiquitin-conjugated proteins (Dantuma et al., 2006), and it has been suggested that it has no adverse secondary effects with short-term exposure (Szutorisz et al., 2006). Treatment with MG132 caused genome-wide reduction of H2AK119ub as previously described (Dantuma et al., 2006; Tavares et al., 2012). In our Runx1-TetR cells, 6h treatment with 10 μM of the inhibitor led to a significant reduction in H2AK119ub at the TetO locus (Figure 3.11 a), accompanied by an increase in H3K27ac (Figure 3.11 b). This suggests that the lack of acetylation at the TetO locus of Runx1-TetR mESCs is, at least in part, due to the catalytic activity of PcG proteins. In addition, reduction of H2AK119ub after treatment with the proteasome inhibitor MG132 also resulted in an increase in H3K4me3 (Figure 3.11 c), in agreement with previous observations showing that reduction in H2AK119ub is followed by an increase in this histone mark (Dobrinic et al., 2021; Nakagawa et al., 2008).

Despite the lack of acetylation, ChIP for Brd4 showed that anchoring Runx1 is sufficient for its recruitment (Figure 3.12). More specifically, Brd4 ChIP-qPCR showed that Brd4 is bound at the TetO locus of Runx1-TetR cells, at levels comparable to its levels at endogenous active control loci, while it is not bound at the TetO locus of TetR-only cells. Treatment of Runx1-TetR cells with doxycycline for 6h, which abolishes the binding of Runx1-TetR to the TetO locus, also reduced Brd4 binding to

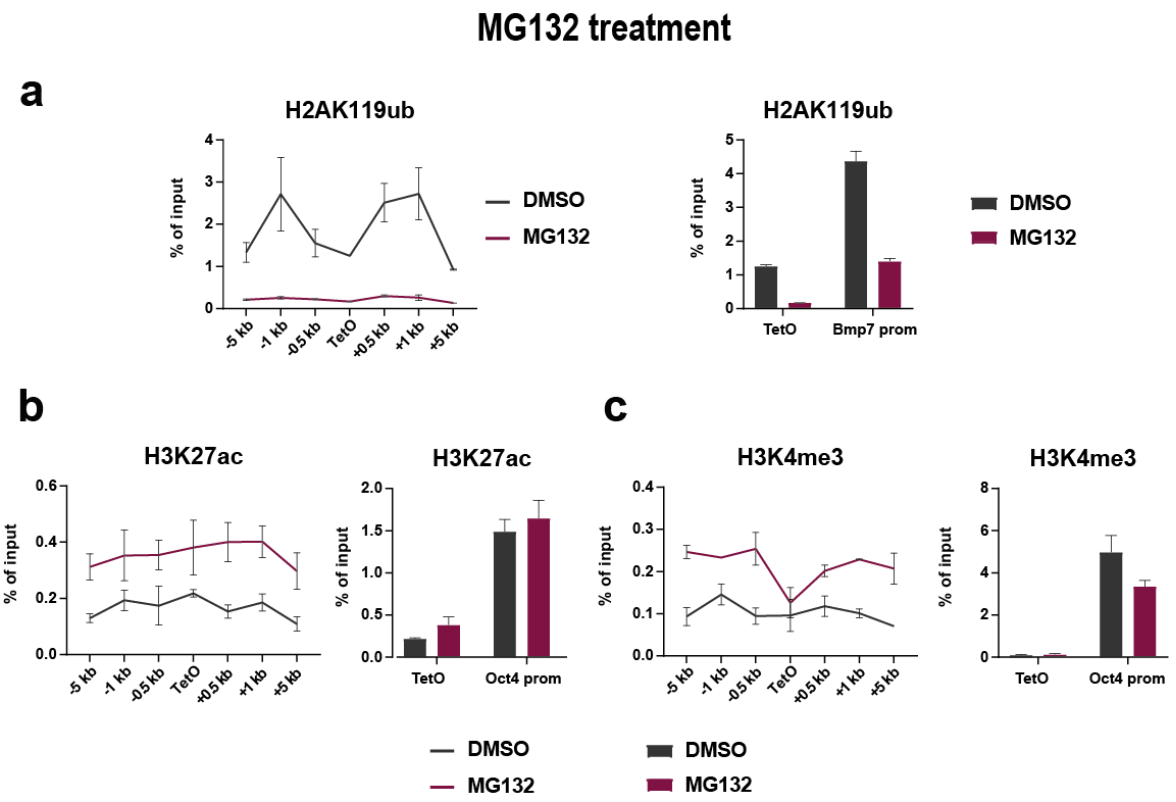


Figure 3.11: Reduction in H2AK119ub results in increase in active modifications at the TetO locus of Runx1-TetR cells. 6h treatment of Runx1-TetR mESCs with 10 µM MG132 inhibitor results in a reduction in H2AK119ub (a), and increases in H3K27ac (b) and H3K4me3 levels (c) at the TetO locus in MG132-treated Runx1-TetR mESCs (dark grey) compared to DMSO-treated cells (purple). ChIP-qPCR at endogenous loci is also shown for comparison of ChIP signal. Experiments were performed in at least 2 biological replicates and error bars represent the standard error of the mean (S.E.M.).

the TetO locus to background levels (Figure 3.12). Although the main mechanism of Brd4 recruitment is considered to be its association with acetylation marks, these data suggest that Runx1 is sufficient for its recruitment independent of acetylation. Potentially, the interaction between p300 and Brd4 (Wu et al., 2018), despite the lack of acetylation, accounts for Runx1-mediated recruitment of Brd4.

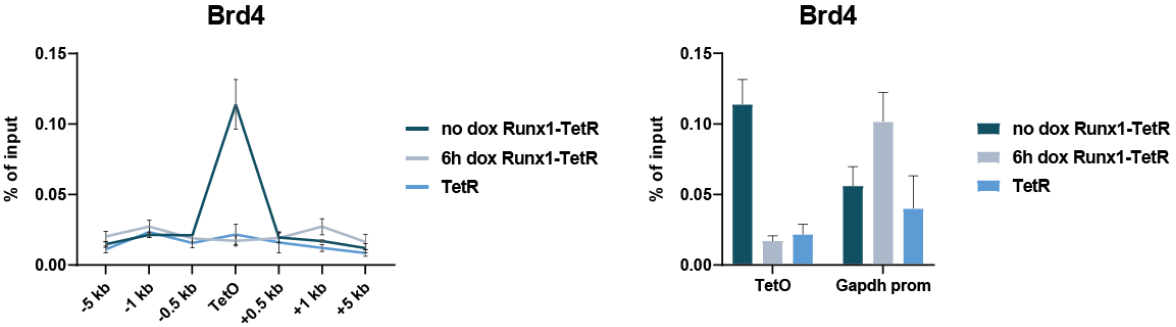


Figure 3.12: Runx1 is sufficient to recruit Brd4. ChIP-qPCR for Brd4 at the TetO locus of Runx1-TetR cells with/without dox treatment (grey and blue lines respectively) and TetR-only cells (light blue line), showing that Runx1 is sufficient to recruit Brd4. ChIP-qPCR at *Gapdh* promoter is also shown for comparison of ChIP signal. ChIP experiments were performed in 3 biological replicates and error bars represent the standard error of the mean (S.E.M.).

To test whether Runx1-mediated recruitment of p300 and Brd4 is also sufficient for the recruitment of the transcriptional machinery at the TetO locus, ChIP for the transcriptional proteins Tbp, Taf1c and RNA polymerase (RNAPII) was performed in Runx1-TetR cells. ChIP-qPCR at the TetO locus revealed the binding of these proteins in Runx1-TetR cells, and in fact the levels of binding of these transcription-associated proteins at the TetO locus are comparable to or higher than their levels of binding at highly expressed endogenous loci (Figure 3.13 a,b,c). Treatment of Runx1-TetR cells with doxycycline for 6h resulted in a reduction in the binding of Tbp, Taf1c and RNAPII at the TetO locus, and ChIP-qPCR for these proteins in TetR-only cells showed that they are not bound at the TetO locus. This further validates that their recruitment is due to the presence of Runx1.

To examine whether the presence of RNAPII leads to transcription at the TetO locus, I performed qRT-PCR in Runx1-TetR and TetR-only mESCs. Surprisingly, this showed no evidence of transcription at the TetO locus of Runx1-TetR cells when compared to TetR-only cells (Figure 3.13 d). The presence of RNAPII in combination with the lack of transcription suggests that the RNAPII is in a conformation and phosphorylation status that does not allow it to initiate transcription, possibly due to the recruitment of and establishment of a Polycomb repressive domain.

3.3.3 Reduction in H2AK119ub levels converts the TetO locus of Runx1-TetR mESCs into a regulatory element with promoter-like features.

Bivalent domains in mESCs and differentiated tissues have previously been described, that are bound by PcG proteins and enriched both in H3K27me3 and H3K4me3 (Bernstein et al., 2006a; Mohn et al., 2008; Pan et al., 2007) and are also bound by RNAPII, but no transcription is observed or the genes are transcribed at very low levels. The TetO locus of the Runx1-TetR system seems to

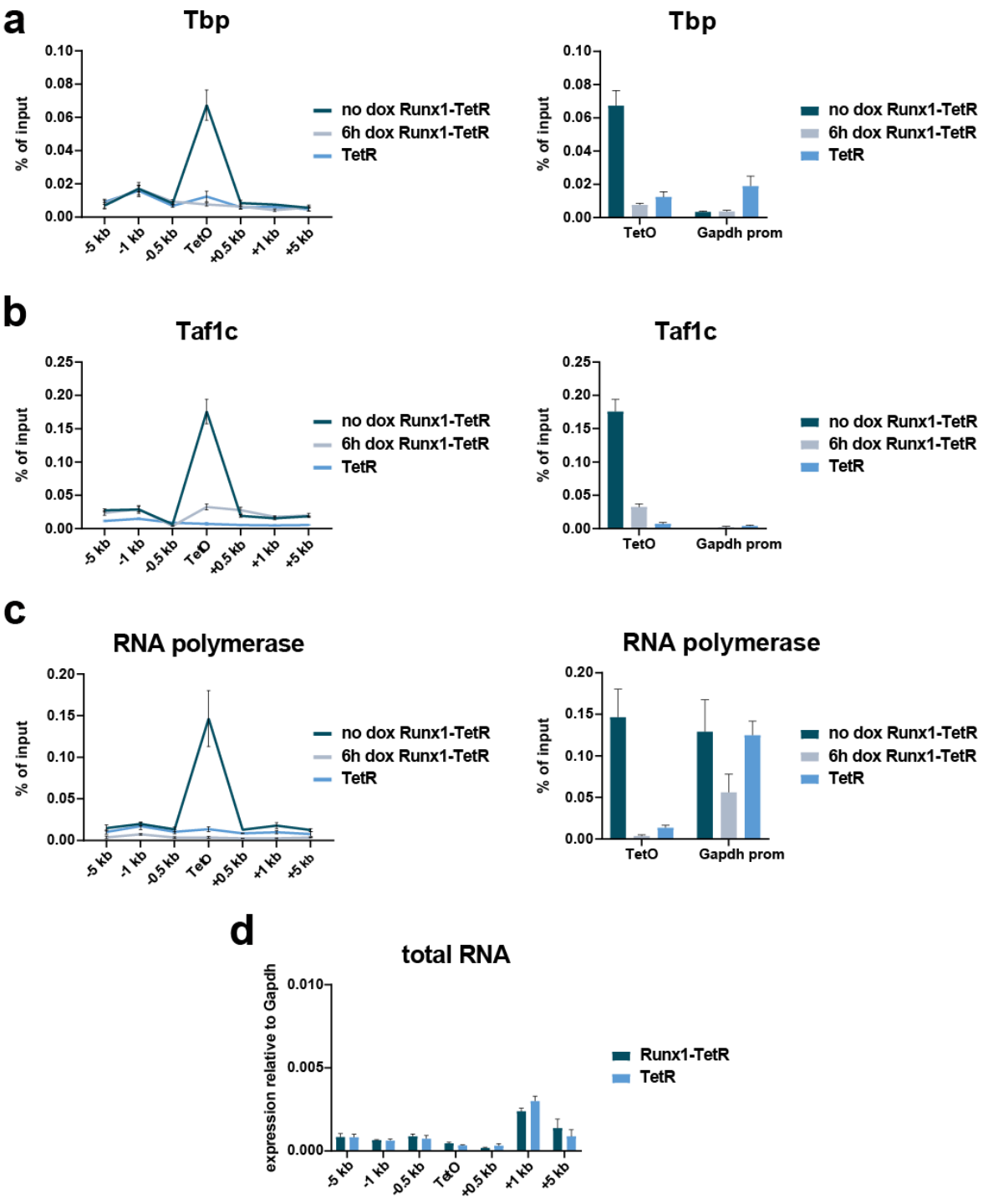


Figure 3.13: Runx1 is sufficient for the recruitment of the transcriptional machinery, but no transcription is observed at the TetO locus. ChIP-qPCR for Tbp (a), Taf1c (b) and RNA polymerase II (c) at the TetO locus of Runx1-TetR cells with/without dox treatment (grey and dark blue lines respectively) and TetR-only cells (light blue line), showing that Runx1 is sufficient to recruit the transcriptional machinery. ChIP experiments were performed in 3 biologicals replicates and error bars show the standard error of the mean (S.E.M.). (d) qRT-PCR for total RNA across the TetO locus in Runx1-TetR and TetR-only mESCs (dark blue and light blue lines respectively), showing that anchoring Runx1 at the TetO locus is not sufficient to promote transcriptional initiation. Experiments were performed in 3 biological replicates and error bars show standard error of the mean (SEM). Values are normalised to *Gapdh* mRNA levels.

share the same characteristics as bivalent domains (i.e. binding of a TF, activators and PcG repressors, and enrichment in H3K4me3 and H3K27me3, in addition to RNAPII binding without active transcription).

Previous studies have proposed that RNAPII at bivalent domains has a different conformation, i.e. is in a poised state, and can be recognised with antibodies detecting Ser5 phosphorylation (Brookes et al., 2012; Stock et al., 2007; Tee et al., 2014). Additionally, it has been reported that loss of H2AK119ub following *Ring1a/b* KO in mESCs results in a change in the conformation status of RNAPII, and more specifically in a reduction in the levels of Ser5 phosphorylated RNAPII and an increase in the total levels of RNAPII, and ultimately gene derepression (Brookes et al., 2012; Stock et al., 2007). Furthermore, it has been reported that reduction in H2AK119ub leads to transcriptional initiation (Dobrinic et al., 2021; Nakagawa et al., 2008). Therefore, to determine whether reduction in H2AK119ub in the Runx1-TetR system has the same effect, i.e. a change in the conformation status of RNAPII and transcriptional initiation, ChIP for Ser5 phosphorylated RNAPII and for total RNAPII after treatment with the MG132 inhibitor for 6h was performed (Figure 3.14). In addition to increased H3K27ac and H3K4me3 levels at the TetO locus of MG132-treated Runx1-TetR mESCs compared to DMSO-treated cells (Figure 3.11), I found that the reduction in H2AK119ub also caused a decrease in the levels of Ser5 phosphorylated RNAPII (Figure 3.14 a) and an increase in the levels of total RNAPII (Figure 3.14 b). The change in the conformation status of RNAPII at the TetO locus led me to check whether reduction in H2AK119ub resulted in transcriptional initiation at the TetO locus. qRT-PCR of MG132-treated Runx1-TetR mESCs compared to TetR-only cells indeed showed transcriptional initiation (albeit at low levels), and more specifically, the production of RNA transcripts from the TetO locus up to 5 kb upstream (Figure 3.14 c). This unidirectional transcription from the TetO locus in combination with the increase in the levels of H3K27ac and H3K4me3, suggest that the reduction in H2AK119ub converts the TetO locus from a bivalent domain to a promoter-like element (albeit with weak transcriptional activity).

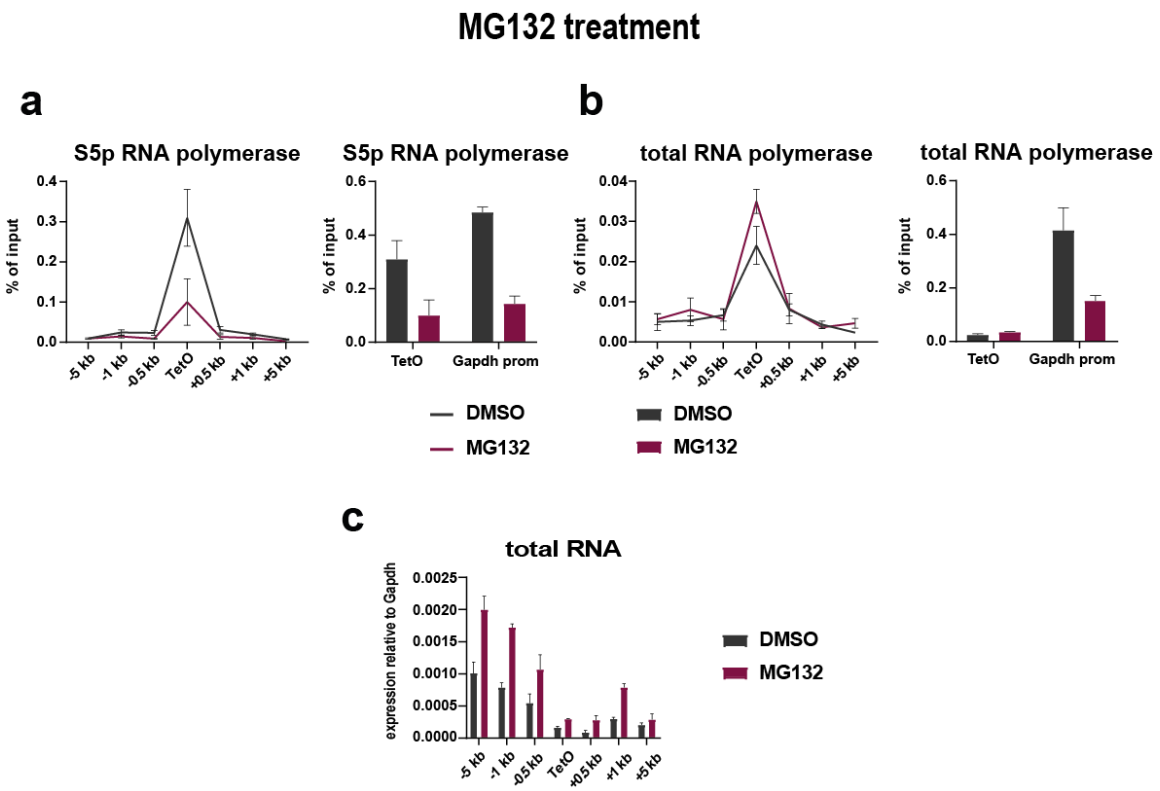


Figure 3.14: Reduction in H2AK119ub results in change of the conformational status of RNAPII and transcriptional initiation from the TetO locus of Runx1-TetR cells. 6h treatment of Runx1-TetR mESCs with 10 µM MG132 inhibitor resulted in (a) reduction in the levels of the S5 phosphorylated form of RNA polymerase II (Millipore/05-623 antibody) and (b) increase in the total RNA polymerase II levels (Santa Cruz sc-899 antibody) in MG132-treated Runx1-TetR mESCs (dark grey) compared to DMSO-treated cells (purple). ChIP-qPCR at endogenous loci is also shown for comparison of ChIP signal. Experiments were performed in at least 2 biological replicates and error bars represent the standard error of the mean (S.E.M.). (c) Expression analysis by qRT-PCR across the TetO in MG132- and DMSO-treated mESCs shows unidirectional transcription from the TetO locus following reduction in H2AK119ub. Experiments were performed in 3 biological replicates and values are normalised to *Gapdh* mRNA levels. Error bars show the standard error of the mean (S.E.M.).

The mechanism by which PcG proteins negatively impact transcription is a challenging question. Previous data support that PcG proteins could negatively affect accessibility (Fitzgerald and Bender, 2001; Zink and Paro, 1995) leading to reduced binding of RNAPII; that they may affect the step of transcriptional initiation (Dobrinic et al., 2021; Nakagawa et al., 2008); or that they may cause the formation of a poised form of RNAPII after initiation (Brookes et al., 2012; Stock et al., 2007). In the Runx1-TetR system, my data show that the binding of PcG proteins does not inhibit the binding of co-activators and our data supports the hypothesis that H2AK119ub catalysed by PRC1 results in the inhibition of transcriptional initiation by RNAPII. This is in accordance with recent data,

according to which H2AK119ub is central to PcG-mediated repression in mESCs (Blackledge et al., 2020; Tamburri et al., 2020) and seems to limit transcriptional burst frequency (Dobrinic et al., 2021).

3.3.4 Runx1-mediated recruitment of PcG proteins results in transcriptional repression at endogenous loci.

In the Runx1-TetR mESC lines, the Runx1 protein that is fused to TetR is full-length and retains its DNA binding ability. We were able to take advantage of the fact that endogenous Runx1 is not expressed in mESCs, in order to determine if TetR-Runx1 had any impact on the regulation of endogenous loci. ChIP-seq for TetR (ChIP with the FS2 antibody recognising the FLAG STREPx2 tag on the C-terminus of TetR) and for Runx1 co-factor Cbfb, revealed that the Runx1-TetR fusion protein binds to over 50,000 loci genome-wide (Figure 3.15). Taking advantage of that, I wanted to investigate how Runx1 binding impacts the recruitment of gene-regulatory proteins at endogenous loci, rather than at the gene-poor TetO locus. The fact that TFs that interact with Runx1 in haematopoietic tissues are not expressed or are expressed at low levels (Figure 3.4) offers the advantage of investigating the effects of Runx1 binding at endogenous sites alone, rather than the potential co-operative effects of Runx1 with additional haematopoietic TFs.

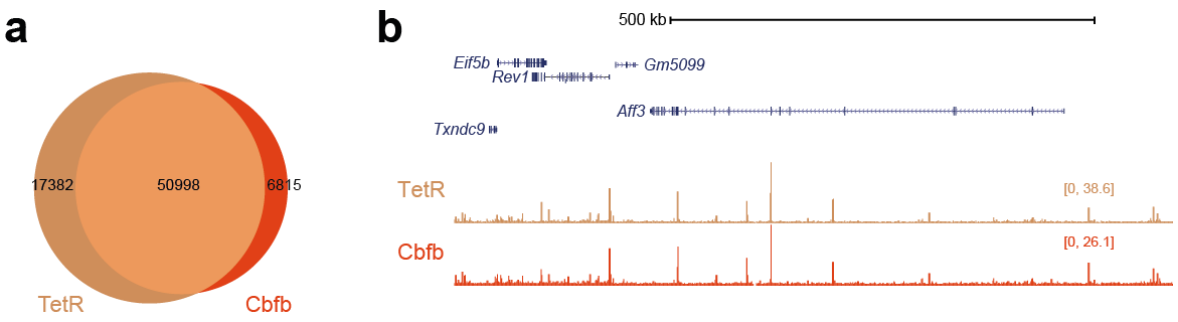


Figure 3.15: Runx1-TetR binding to thousands of loci genome-wide in mESCs. (a) Venn diagram showing the overlap of Cbfb and TetR (antibody for the FS2 tag) ChIP-seq peaks in Runx1-TetR mESCs. (b) UCSC ChIP-seq tracks for Cbfb and TetR at the *Aff3* locus.

Firstly, I investigated the effect of Runx1 binding at endogenous PcG loci. I noticed that among the Runx1 binding sites across the genome there are many PcG sites, in fact about 30% of PcG sites in mESCs (defined as sites bound by Ring1b in TetR-only mESCs) are bound by the Runx1-TetR fusion protein (Figure 3.16 a). By performing reference-normalised ChIP-seq for Ring1b in Runx1-TetR cells and TetR-only cells, I was able to show that Runx1 binding at PcG sites results in increased Ring1b binding (Figure 3.16 b,c). I also validated my ChIP-seq results (with ChIP-qPCR at a subset of loci) that Runx1 binding results in increased Ring1b binding and increased levels of the repressive PcG histone modifications (Figure 3.17 a,b). At these genes, which are expressed at low levels, nascent RNA-seq in TetR and Runx1-TetR mESCs further shows that Runx1-mediated recruitment of PcG proteins results in transcriptional repression. In addition, at these loci (*Crim1, Satb1)*, and at other Runx1-bound PcG-repressed sites (*Irak3, Htra1, Amfr*), I found that Runx1 binding resulted in increased binding of the PcG subunit Phc1 (Figure 3.17 c). The increased binding of Phc1 could be explained both through its association with Runx1, as the results at the TetO locus show that Runx1 is sufficient to recruit this subunit, and most likely the increased levels of H3K27me3 resulted in increased Phc1 binding through the association of Cbx proteins of the cPRC1 complexes with H3K27me3 (Bernstein et al., 2006b; Fischle et al., 2003; Min et al., 2003). In addition, I have found other examples of Runx1 binding at non-transcribed regions, where, similarly to the TetO locus, Runx1 binding leads to Ring1b binding and to the *de novo* formation of a Polycomb domain (Figure 3.17 d).

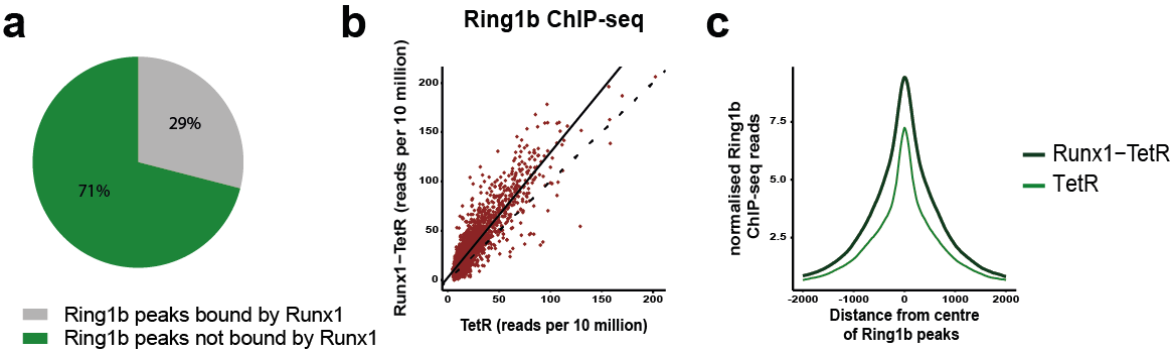


Figure 3.16: Runx1 binding results in increased PcG binding at PcG sites. (a) Pie chart showing the proportion of Ring1b ChIP-seq peaks in TetR-only cells that are bound by Runx1 (using TetR ChIP-

seq peaks in Runx1-TetR cells), indicating that 30% of Runx1 sites overlap with PcG sites. (b) Correlation of normalised Ring1b ChIP-seq reads at Ring1b peaks bound by Runx1-TetR in Runx1-TetR and TetR-only mESCs, showing that upon Runx1 binding Ring1b ChIP-seq signal increases at the majority of Ring1b sites. (c) Metaplot of mean normalised Ring1b ChIP-seq reads at Ring1b peaks bound by Runx1-TetR in Runx1-TetR and TetR-only mESCs (dark green and light green lines respectively), indicating that Runx1 binding results in increased Ring1b binding at PcG loci.

In contrast, at "classical" PcG-repressed sites, such as *Hoxb5* and *Pax7*, whose promoters are bound at background levels by Runx1 in Runx1-TetR cells, my reference-normalised Ring1b ChIP-seq results do not show an increase in Ring1b binding in Runx1-TetR cells compared to TetR-only cells (Figure 3.17 e,f). Together, this suggests that increases in Ring1b binding at a subset of loci are the direct result of Runx1 recruitment.

Next, I evaluated whether the increase in PcG binding at PcG-repressed sites after Runx1-TetR binding contributes to transcriptional repression. To do that, I focused on Runx1-bound PcG-repressed genes. From this analysis, I identified 2993 Runx1-TetR peaks at PcG sites (TetR peaks in Runx1-TetR cells overlapping with Ring1b sites in TetR cells), corresponding to 2779 genes. From these genes 1510 are expressed, while the rest are not transcribed. By focusing on the PcG-repressed genes that are lowly transcribed (expression less than 10 CPM in TetR cells, which corresponds to 754 genes), I found that Runx1-TetR binding further contributes to gene repression at most of these sites (Figure 3.18 c). This suggests that at lowly expressed PcG-repressed genes that are bound by Runx1, Runx1 binding further represses transcription.

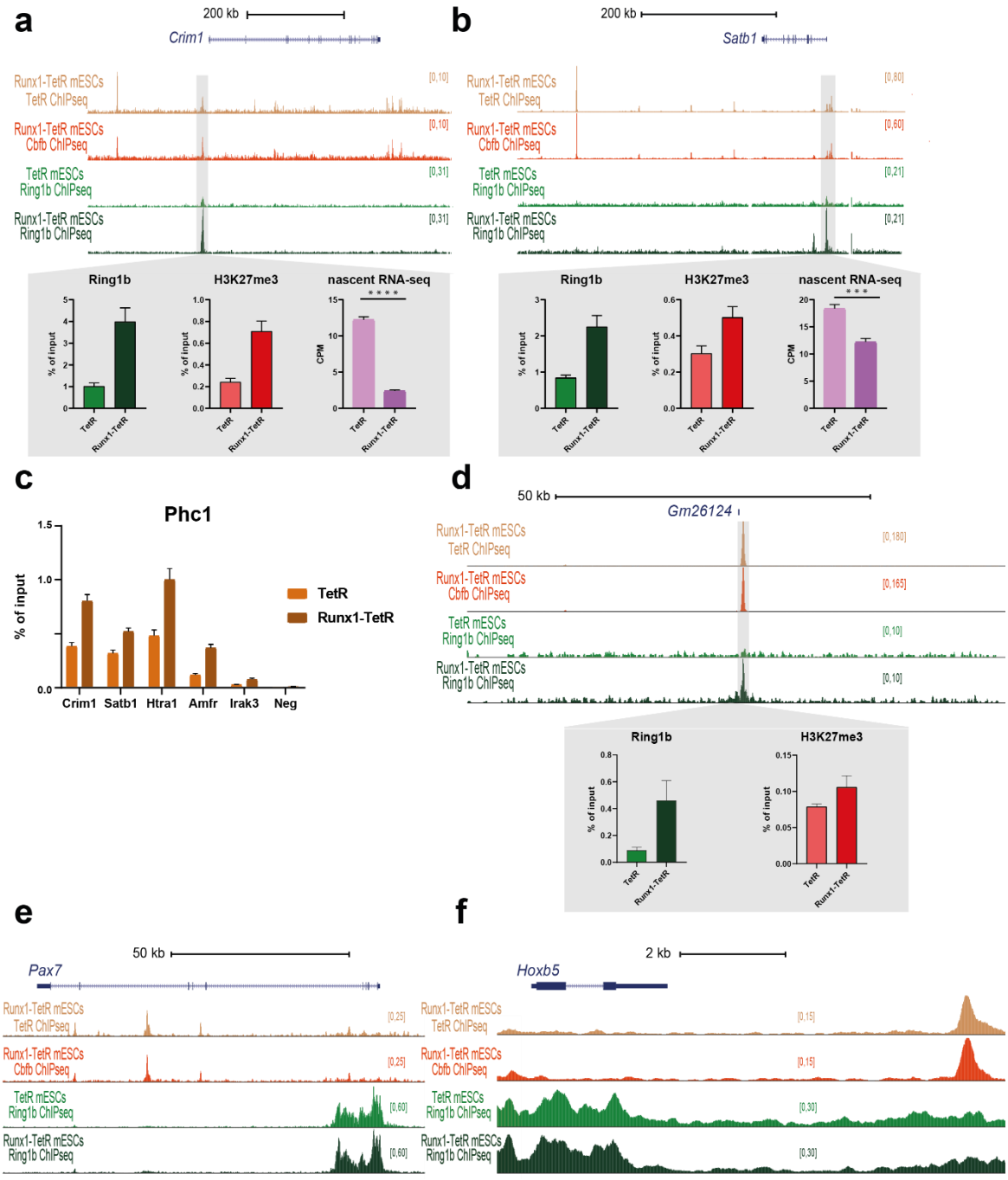


Figure 3.17: Runx1 binding at endogenous Polycomb sites results in increased binding of PcG proteins and transcriptional repression. *Top:* UCSC ChIP-seq tracks for TetR (FS2 tag) and Cbfb in Runx1-TetR mESCs and reference-normalised ChIP-seq for Ring1b in TetR-only and Runx1-TetR mESCs at *Crim1* promoter (a) and *Satb1* promoter (b), showing increased Ring1b binding in Runx1-TetR cells compared to TetR-only cells. *Bottom:* ChIP-qPCR for Ring1b and H3K27me3, and nascent RNA-seq results in TetR-only and Runx1-TetR mESCs at the same loci. *Crim1* and *Satb1* are downregulated in Runx1-TetR cells compared to TetR-only cells (****FDR < 0.0001, ***FDR < 0.001). (c) ChIP-qPCR for Phc1 in TetR-only and Runx1-TetR mESCs at *Crim1, Satb1, Htra1, Amfr* and *Irek3* loci and a negative control region downstream the TetO locus. (d) *Top:* ChIP-seq tracks for TetR (FS2 tag) and Cbfb in Runx1-TetR mESCs and for Ring1b in TetR-only and Runx1-TetR mESCs at *Gm26124* locus. *Bottom:* ChIP-qPCR for Ring1b and H3K27me3 at the same locus, indicating that Runx1 binding at an inactive locus results in the *de novo* formation of a Polycomb domain. (e) ChIP-seq tracks for TetR and Cbfb in Runx1-TetR mESCs and Ring1b in TetR-only and

Runx1-TetR mESCs at *Pax7* and (f) *Hoxb5* loci, showing examples of PcG-repressed genes, whose promoters are not bound by Runx1-TetR. At these sites, reference-normalised Ring1b ChIP-seq does not show increase in Ring1b binding in Runx1-TetR cells compared to TetR-only cells. ChIP-qPCR and nascent RNA-seq experiments were performed in 3 biological replicates per condition and error bars represent the standard error of the mean (S.E.M.).

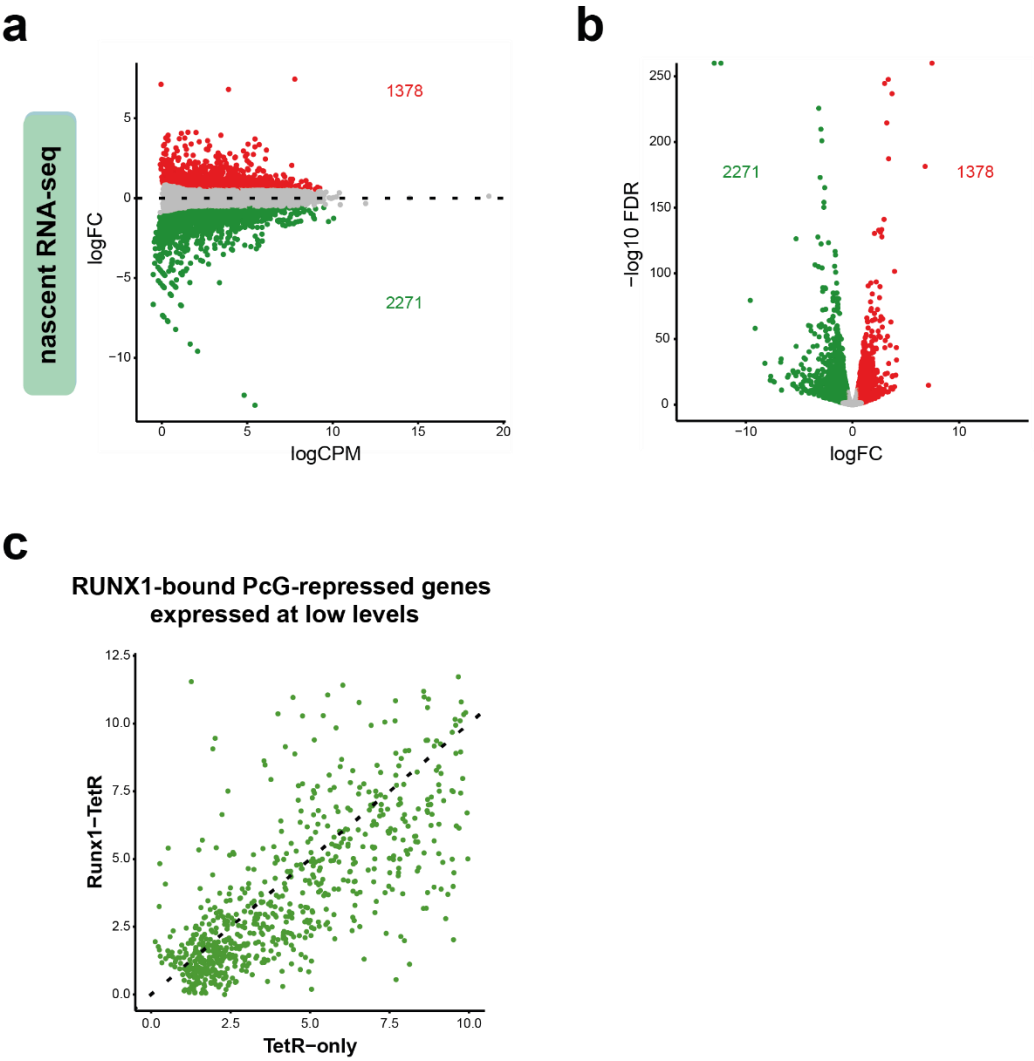


Figure 3.18: Runx1-TetR binding results in transcriptional repression of lowly expressed Runx1-bound PcG-repressed genes. (a) MA plot and (b) volcano plot showing genes differentially expressed in Runx1-TetR cells compared to TetR-only mESCs, after performing nascent RNA-seq. Upregulated genes (n= 1378) are shown in red, downregulated genes (n=2271) are shown in green, and unchanged genes in grey. Nascent RNA-seq was performed in 3 biological replicates per condition. (c) Dotplot showing in TetR-only cells and Runx1-TetR cells the expression levels of PcG-repressed genes that are lowly expressed (CPM in TetR-only cells <=10, n=754 genes) and are bound by Runx1-TetR, demonstrating that most of Runx1-bound PcG-bound genes are downregulated in Runx1-TetR cells.

3.3.5 Runx1-mediated recruitment of co-factors is context-dependent.

Finally, I investigated how Runx1 binding impacts gene regulation at active loci. In contrast to my observations at PcG-repressed sites, I noticed that Runx1 binding at active loci (for example at active genes *Fkbp3, Ppp2ca, Nanog, Triml2)* resulted in increased binding of p300 and increased levels of histone acetylation (Figure 3.19 a,b,c,d). In contrast, there was no stable binding of Ring1b at active Runx1-bound sites in Runx1-TetR cells. I also found that at a subset of active loci, binding of Runx1 and increased deposition of active modifications led to transcriptional upregulation (*Fkbp3* and *Ppp2ca* genes) (Figure 3.19 a,b), although increased binding of co-activators and enrichment in active histone marks did not always translate to increase in transcription (*Nanog, Triml2* genes) (Figure3.19 c,d). To determine genome-wide how Runx1-TetR binding at active loci affects transcription, I divided expressed genes that are bound by Runx1 at their promoters and are not bound by PcG proteins into 4 quartiles based on their expression levels in TetR-only cells (with the 1st quartile representing the genes expressed at the lowest levels, and the 4th quartile the most highly expressed genes), then calculated the logFC of gene expression in Runx1-TetR cells compared to TetR-only cells in each of these quartiles. My results show that for the lowly expressed genes, Runx1 binding on average results in gene repression, while Runx1 binding has the opposite effect on highly expressed genes, i.e. on average, Runx1 binding further leads to gene activation (Figure 3.20).

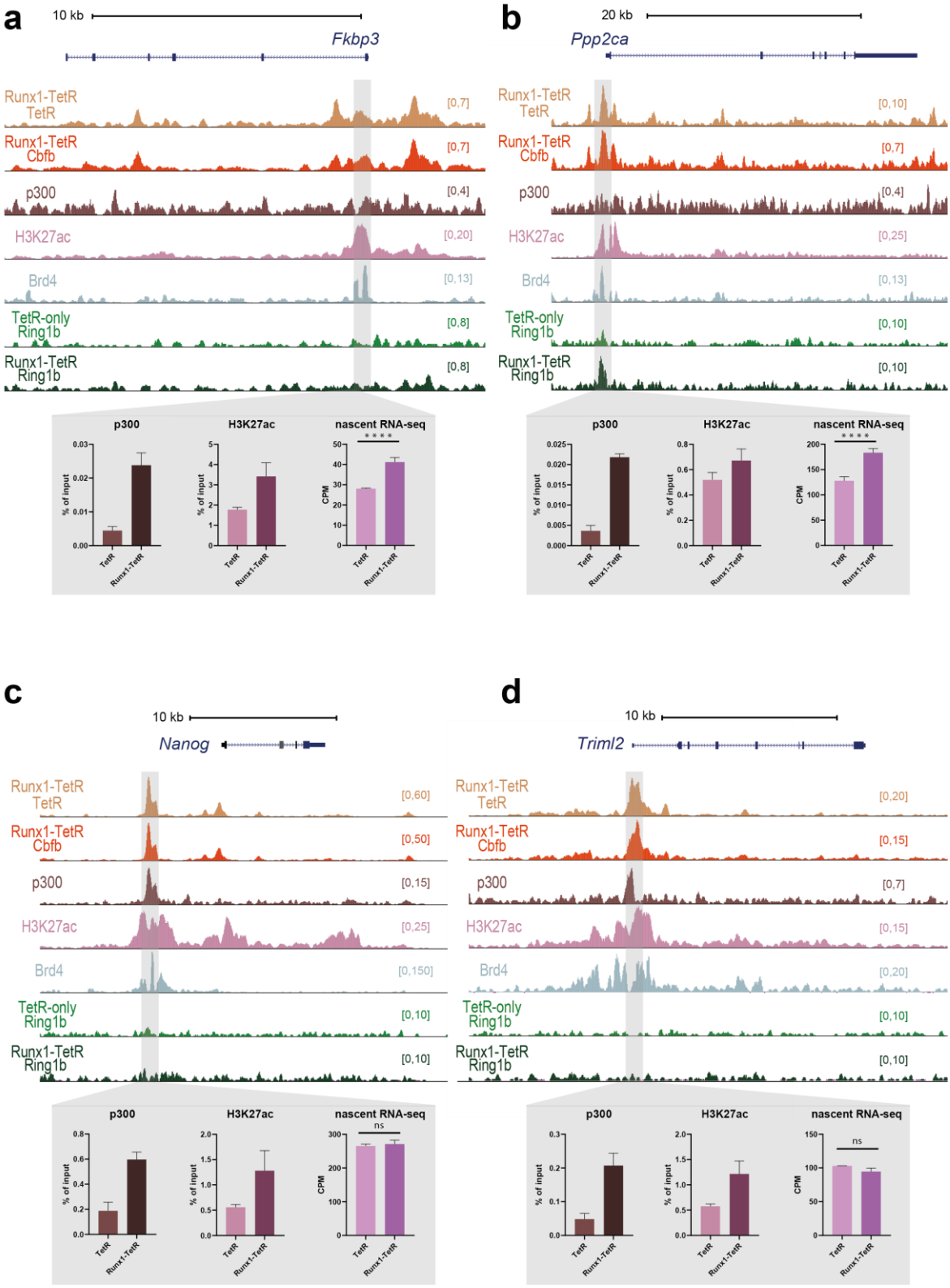


Figure 3.19: Runx1 binding at active loci results in increased binding of co-activators. ChIP-seq tracks for TetR and CbFb in Runx1-TetR mESCs, for p300, H3K27ac and Brd4 in mESCs (from publically available datasets (Ferrari et al., 2014; Gonzales-Cope et al., 2016; Martire et al., 2019)) and for Ring1b in TetR-only mESCs and Runx1-TetR mESCs at the *Fkbp3* (a), *Ppp2ca* (b), *Nanog* (c) and *Triml2* loci (d). Below the ChIP-seq tracks, ChIP-qPCR results for p300 and H3K27ac in TetR-only and Runx1-TetR mESCs at the same loci, and nascent RNA-seq results for the corresponding genes are shown. ChIP-qPCR and nascent RNA-seq experiments were performed in 3 biological replicates,

and error bars show standard error of the mean (S.E.M.) *Fkbp3* and *Triml2* are upregulated in Runx1-TetR cells compared to TetR-only cells (****FDR < 0.0001), while the transcriptional changes for *Nanog* and *Triml2* are not statistically significant (ns).

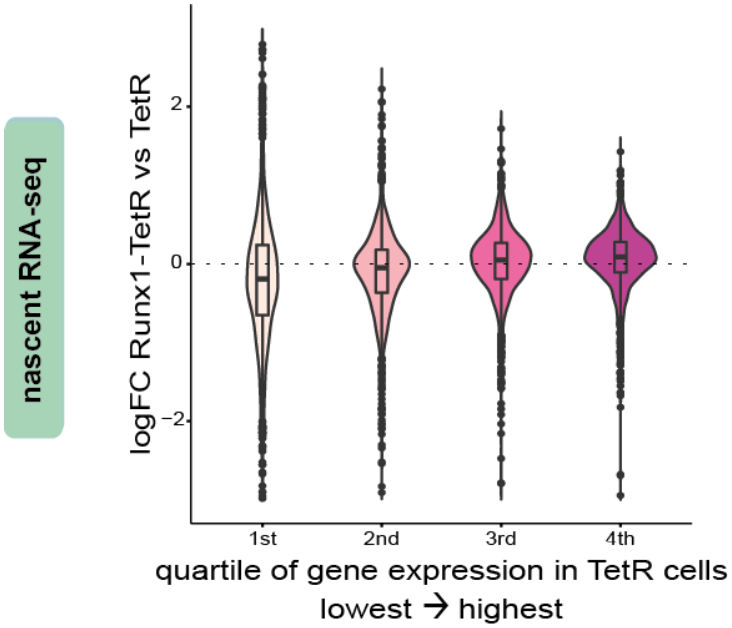


Figure 3.20: Runx1 binding at highly expressed active loci results in transcriptional upregulation. Violin plots showing logFC of nascent RNA levels in Runx1-TetR compared to TetR-only cells, for active genes bound by Runx1-TetR at their TSS. Genes were divided in 4 quartiles based on their expression levels in TetR cells. LogFC was calculated from the average of 3 biological replicates per condition.

Thus, using this model system, my data suggest that Runx1 modifies the chromatin landscape by recruiting co-activators or repressors. However, Runx1-mediated recruitment of chromatin-modifying proteins is context-dependent: Runx1 stabilises the binding of PcG repressors at PcG-repressed regions, while it stabilises the binding of activators at active regions. This suggests that Runx1 “amplifies” active or repressive signals depending on the existing chromatin landscape, and can “switch” between co-factors based on the spatiotemporal conditions.

3.4 Discussion

In these experiments, by employing the TetR system in order to determine protein recruitment, I have shown that RUNX1 is sufficient to recruit both co-activators (p300, Brd4, TrxG proteins Mll and Brg1) and the pre-initiation complex and also co-repressors (PRC1 complexes, Hdacs). Interestingly, among the co-factors that are recruited by Runx1 are pairs of gene-regulatory proteins with opposite function (p300-Hdacs and the PRC1-Bap1 complexes).

Especially for PRC1 recruitment, these results clearly show that RUNX1, a DNA binding sequence-specific TF, is sufficient to recruit PRC1 in mammalian cells. So far, robust evidence for TF-mediated recruitment of PRC1 complexes comes from the studies focusing on PRC1.6 complex, which is recruited by the E2F6-DP1/MGA-MAX TFs (Endoh et al., 2017; Scelfo et al., 2019; Stielow et al., 2018; Trojer et al., 2011). My results show that RUNX1 is sufficient to recruit both canonical PRC1 complexes independent of H3K27me3, and variant complexes, including the PRC1.1 and PRC1.6 complexes. Unlike in *Drosophila,* where the binding of PcG proteins at Polycomb Response Elements (PREs) is mediated by TFs (Kassis and Brown, 2013), in mammalian cells PcG proteins seem to function mainly by sampling the chromatin and binding more stably to unmethylated CpG islands at transcriptionally permissive sites that lack activating signals and transcription (Klose et al., 2013). My results support, complementary to this mechanism of binding, the role of RUNX1 and potentially other TFs that can recruit PRC1.

The *de novo* recruitment of these factors at the TetO locus due to Runx1 binding results in the establishment of a regulatory element that shares the same characteristics with bivalent domains (binding of PcG and TrxG proteins and RNAPII, and enrichment with the active mark H3K4me3 and the repressive marks H2AK119ub and H3K27me3). Bivalent domains have been identified in mESCs and in more differentiated tissues, and their formation is believed to be a mechanism for establishing a poised state for the expression of key lineage-specific genes (Bernstein et al., 2006a; Mohn et al., 2008; Pan et al., 2007). The expression of the appropriate TFs and the signalling cues

during the subsequent steps of differentiation result in either activation of the corresponding genes, repression or maintenance of the bivalent state. In the Runx1-TetR system used here, despite the recruitment of co-activators, Runx1-mediated recruitment of PcG proteins is sufficient to establish a repressive Polycomb domain. I hypothesise that the lack of transcription and the lack of activating marks at this neutral region results in the prevalence of PcG repressive marks. It would be of interest to see whether Runx1-mediated recruitment of PcG proteins would still result in the establishment of a Polycomb domain at the TetO locus, if it was enriched in active marks, for example by introducing Runx1-TetR in mESCs after having targeted dCas9-p300 to the TetO locus.

In the Runx1-TetR system used here, despite the recruitment of p300 and RNAPII respectively at the TetO locus, it is not enriched in histone acetylation and is not transcribed. By further dissecting the lack of acetylation and the lack of transcription, my experiments revealed that reduction in H2AK119ub that is catalysed by PRC1 resulted in an increase in the active histone modifications H3K27ac and H3K4me3 and to transcriptional initiation, albeit the transcript levels at the TetO locus are rather low. This suggests that the reduction in H2AK119ub renders the TetO locus into a promoter-like regulatory element with weak transcriptional activity. In agreement with this, others have shown that reduction in H2AK119ub in mESCs results in an increase in transcription-associated active histone modification H3K4me3 (Dobrinic et al., 2021; Nakagawa et al., 2008), while accumulation of H2AK119ub leads to decrease in H3K4me3 and H3K27ac levels (Fursova et al., 2021).

By further examining the effect of PcG proteins on the conformation status of RNAPII I have shown that upon reduction of H2AK119ub, the levels of Ser5 phosphorylated RNAPII decrease and the levels of total RNAPII increase. This is consistent with published studies that support that RNAPII bound at bivalent regions is found in a poised state, and can be recognised with antibodies detecting Ser5 phosphorylation of the carboxyterminal domain (CTD) of RNAPII, but it cannot be recognised by the 8WG16 antibodies which preferentially recognize non-phosphorylated Ser2

residues (Brookes and Pombo, 2009; Stock et al., 2007). In accordance with that, in the Runx1-TetR experimental system, RNAPII was detected at the TetO locus with an antibody that preferentially recognises Ser5 phosphorylation (Figures 3.13 c and 3.14), but not with an 8WG16 antibody (ChIP with 8WG16 antibody not shown). In addition, *Ring1a/b* knockout experiments in mESCS have shown that the reduction in PcG binding and H2AK119ub levels at bivalent domains is followed by, similarly to TetO bivalent domain in our experimental system, decrease in the binding of Ser5 phosphorylated RNAPII, increase in total RNAPII levels and gene derepression (Brookes et al., 2012; Stock et al., 2007). Interestingly, other groups have supported that the phosphorylation of Ser5 of the poised RNAPII at bivalent genes is not mediated by the TFIIH CTD kinase subunit Cdk7, as it would be expected, but instead by the Erk1/2 kinases (Ma et al., 2016; Tee et al., 2014). Further ChIP experiments for Erk1/2 and Cdk7 in the Runx1-TetR system could be used to validate whether Erk1/2 kinases are bound at the TetO locus of Runx1-TetR cells, and whether their inhibition would result in transcriptional initiation, or whether it is TFIIH that mediates Ser5 phosphorylation of RNAPII.

In my experiments, the reduction in H2AK119ub following treatment with MG132 should not disrupt the Runx1-mediated PRC1 recruitment to the TetO locus. Further *Ring1a/b* conditional knockout experiments in the Runx1-TetR system would help to dissect whether the removal of PcG proteins leads to a further increase in activating marks and increase in transcription, or whether the PcG-mediated inhibition of acetyltransferase activity and transcriptional initiation mainly takes place through the deposition of ubiquitin molecules. Although the exact mechanisms of how PcG proteins inhibit transcription are not defined and previous studies have reported that PcG proteins negatively affect accessibility (Fitzgerald and Bender, 2001; Zink and Paro, 1995), the assembly of the pre-initiation complex, transcriptional initiation (Dellino et al., 2004; Nakagawa et al., 2008) or result to a poised state of RNAPII post initiation (Brookes et al., 2012; Brookes and Pombo, 2009; Stock et al., 2007), recent data support that PcG repression is, at least in mESC, dependent on H2AK119ub (Blackledge et al., 2020; Tamburri et al., 2020), which seems to limit transcriptional

initiation by affecting transcriptional burst frequency (Dobrinic et al., 2021). However, in order to fully dissect the exact mechanistic details of how PcG proteins affect transcription, detailed biochemical studies are needed, and also it is vital that the function of PcG proteins is studied in other systems beyond mESCs, in which it seems that PcG repression is independent from H2AK119ub (Pengelly et al., 2015).

Finally, my experiments have shown that Runx1-mediated recruitment of co-factors is context-dependent. More specifically, at least in this experimental system, Runx1 binding at PcG-repressed loci further promotes the binding of PcG proteins, while at active loci Runx1 binding stabilises the binding of co-activators. The presence of intrinsically disordered domains in Runx1 could provide an explanation for its ability to interact with multiple gene-regulatory proteins, and my data support that Runx1 can "switch" between its co-factors, and exchange its co-factors with binding partners with the opposite function, depending on the existing chromatin context. I hypothesise that context-dependent stabilisation of co-activator or co-repressor binding relies on multivalent interactions (Rothbart et al., 2013; Ruthenburg et al., 2011; Ruthenburg et al., 2007b); the low-affinity transient interactions between Runx1 and co-activators or between Runx1 and co-repressors are stabilised by additional low affinity interactions between co-activators and active histone marks or with co-repressors and repressive PcG modifications respectively. The sum of low affinity interactions ultimately leads to the formation of higher affinity interactions, which is translated to stabilisation of binding of co-activators or co-repressors, and to transcriptional changes.

Chapter 4 - Understanding RUNX1-mediated PRC1 recruitment in leukaemia

4.1 Introduction

4.1.1 The role of RUNX1 in leukaemias with fusion proteins

RUNX1 is a key protein in myeloid neoplasms, and is often a target of mutations and translocations in AML and myelodysplastic syndromes (MDS) (Sood et al., 2017). In addition, wildtype RUNX1 promotes the survival of leukaemic cells in other forms of leukaemia including MLL-rearranged AML and ALL (Goyama et al., 2013; Harman et al., 2021; Wilkinson et al., 2013) and E2A-PBX1 ALL leukaemias (Pi et al., 2020). However, despite the established role of mutant and wildtype RUNX1 protein in the survival of leukaemic cells, the mechanistic details of how RUNX1 impacts gene regulation in leukaemia remain elusive.

In the previous chapter, using the TetR system I demonstrated that RUNX1 is sufficient to recruit many gene regulatory proteins including the TrxG and PcG complexes, and that depending on the existing chromatin landscape, RUNX1 binding results in the stabilisation of co-activator or co-repressor binding. More specifically, I showed that although RUNX1 binding to the gene-poor TetO locus caused recruitment of both co-activators and co-repressors, the lack of transcription and activating signals results in the prevalence of PcG proteins and PcG repressive marks. In addition, RUNX1 binding to PcG-repressed loci further contributes to the recruitment of PcG proteins and the deposition of repressive PcG histone marks.

The TetR system provides the opportunity to study RUNX1 function at a gene poor region without the complications of endogenous gene regulatory elements and without the presence of other haematopoietic TFs that associate with RUNX1 in haematopoietic tissues. However, given the importance of RUNX1 in human leukaemias, it is important that the information about the RUNX1 co-factors gained from the TetR system, is also used to study how RUNX1 mediates the recruitment of co-factors and impacts gene regulation in leukaemia models. Among the co-factors that are

recruited by RUNX1, I chose to focus on PcG proteins, as there are only a few clear examples of TFs that have been shown to recruit PcG proteins in mammalian cells (for example the E2F6/DP1 and MGA/MAX TFs, responsible for mediating the recruitment of the PRC1.6 complex (Endoh et al., 2017; Scelfo et al., 2019; Stielow et al., 2018; Trojer et al., 2011)).

I decided to investigate the function of RUNX1 in MLL-AF4 and RUNX1-RUNX1T1 leukaemias, both of which are dependent on wildtype RUNX1 for leukaemia cell survival (Ben-Ami et al., 2013; Goyama et al., 2013; Wilkinson et al., 2013). More specifically, I wanted to determine whether RUNX1 mediates the recruitment of PcG proteins in a leukaemic cell context, and whether it affects gene expression through their recruitment. In the case of MLL-AF4 leukaemias, *RUNX1* is a direct downstream target of the MLL-AF4 fusion protein (Wilkinson et al., 2013). In RUNX1-RUNX1T1 leukaemias, the *t(8;21)* translocation results in the fusion of one *RUNX1* allele to the *RUNX1T1* gene (also known as *ETO),* leading to the generation of the RUNX1-RUNX1T1 fusion protein (Miyoshi et al., 1993; Miyoshi et al., 1991). In the RUNX1-RUNX1T1 fusion protein only the N-terminal part of RUNX1, including the DNA binding domain, is retained, while the C-terminal part is lost (Figure 1.8 a,b). The other *RUNX1* allele encodes wildtype RUNX1 that regulates the expression of cell signalling genes and has an anti-apoptotic role in RUNX1-RUNX1T1 leukaemic cells (Ben-Ami et al., 2013).

4.1.2 Increased frequency of *RUNX1* and *PcG* mutations in leukaemia

In a significant proportion of RUNX1-RUNX1T1 patients, additional mutations in *PcG* genes have been identified. More specifically, this includes the *EZH2, ASXL1, ASXL2, BCOR* and *BCORL1* genes (Duployez et al., 2016; Hollein et al., 2018) (Figure 4.1). In a recent study with a large cohort of adult patients it was estimated that *ASXL2*, *ASXL1* and *EZH2* mutations are found in 25%, 11% and 7% of RUNX1-RUNX1T1 patients respectively (Hollein et al., 2018). In addition, *ASXL1* mutations have been associated with reduced relapse-free survival (Hollein et al., 2018; Krauth et al., 2014; Micol

et al., 2014), while for *ASXL2* mutations, there are conflicting data on the effect of the mutations on event-free survival of RUNX1-RUNX1T1 patients in different patient cohorts (Jahn et al., 2017; Micol et al., 2014). Similarly, *RUNX1* mutations in MDS, AML and early T-cell precursor acute lymphoblastic leukaemia (ETP ALL) have been shown to frequently co-occur with *PcG* mutations (Booth et al., 2018; Gaidzik et al., 2016; Lehnertz et al., 2017; Ochi et al., 2020; Sashida et al., 2014; Wu et al., 2019). In fact, pair-wise correlation of the genes that are most frequently mutated in AML and MDS cases, performed by Ochi et al., 2020, in order to identify the pairs of genes with mutually exclusive mutations or with mutations that frequently co-occur, shows a significant co-occurrence of *RUNX1* mutations and mutations in genes encoding PcG proteins ASXL1, BCOR and EZH2 in patients (Ochi et al., 2020) (Figure 4.2).

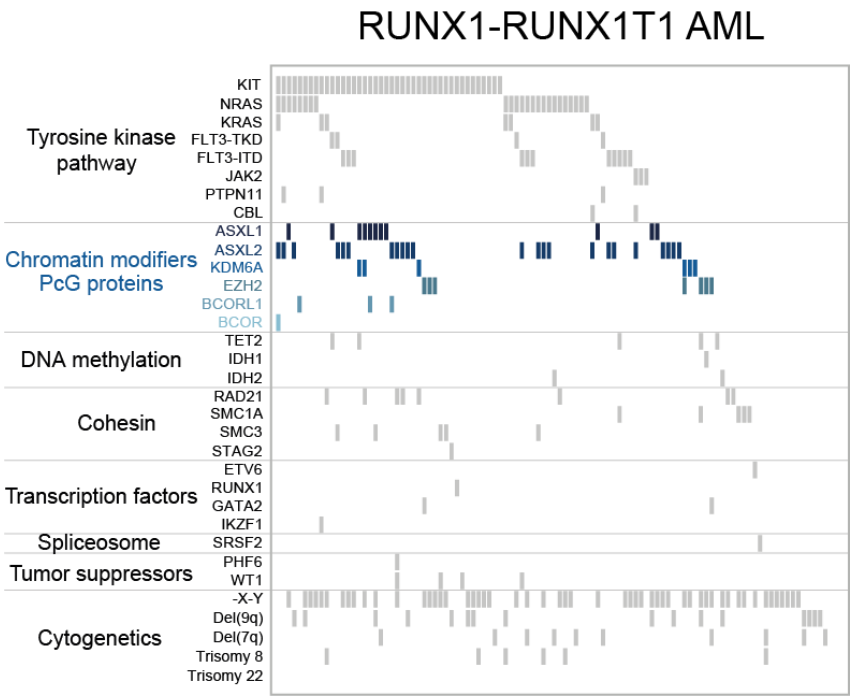


Figure 4.1: A subset of RUNX1-RUNX1T1 patients have additional mutations in PcG genes. Mutational status of RUNX1-RUNX1T1 AML patients. Each column represents a single patient. Highlighted are mutations in *PcG* genes. The figure was adapted from (Duployez et al., 2016).

RUNX1 mutations are usually localised at the sequence coding for its DNA-binding domain and the resulting mutant protein shows reduced DNA binding. Alternatively, nonsense and frameshift

mutations occur across the *RUNX1* gene. Frequently these mutations occur in the nucleotide sequence after the DNA-binding domain and result in the formation of a truncated RUNX1 mutant protein that, similarly to the RUNX1-RUNX1T1 fusion protein, retains only the N-terminal part of wildtype RUNX1 (Gaidzik et al., 2016; Tang et al., 2009) (Figure 1.8 c).

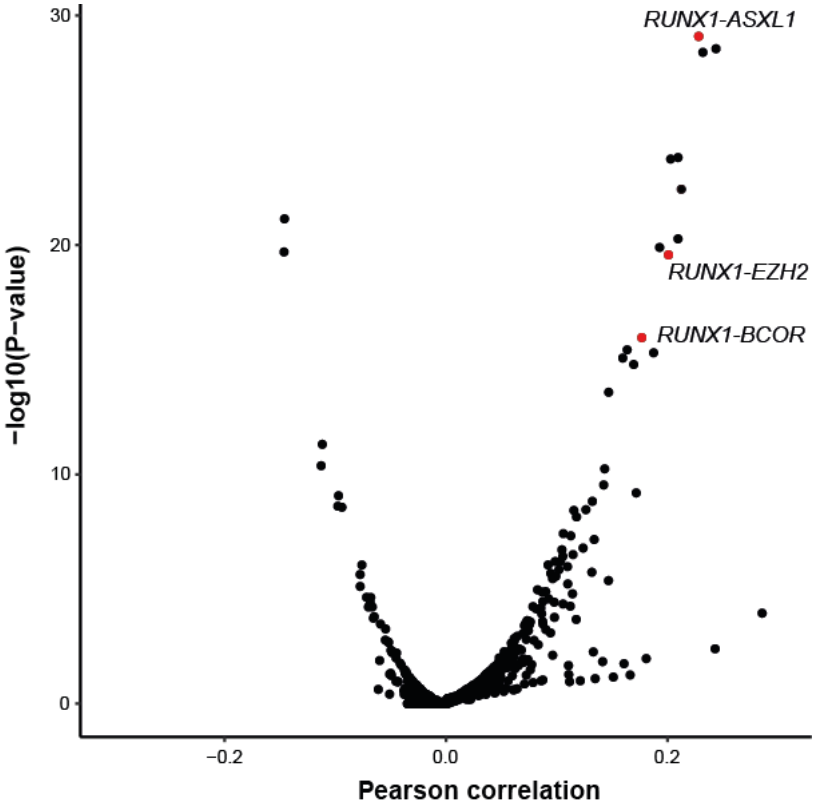


Figure 4.2: *RUNX1* and *PcG* mutations frequently co-occur in AML/MDS patients. Volcano plot showing the Pearson correlation values and –log10(P-values) of pairs of mutations in AML and MDS that significantly co-occur or are mutually exclusive. Positive Pearson correlation suggests co-occurrence of pairs of mutations. *RUNX1* mutations frequently co-occur with mutations on *PcG* genes *ASXL1, BCOR* and *EZH2.* Pearson correlation values and the corresponding P-values were calculated in (Ochi et al., 2020), and the volcano plot was reproduced from (Ochi et al., 2020) to emphasize co-occurrence of *RUNX1* and *PcG* mutations.

In turn, mutations in genes encoding PcG proteins result in inappropriate transcriptional upregulation of PcG-repressed genes (Asada et al., 2018; Balasubramani et al., 2015; Ernst et al., 2010; Score et al., 2012; Yang et al., 2018). More specifically, previous studies have demonstrated that *PcG* mutations function by either attenuating the deposition of the repressive Polycomb histone modifications (*EZH2* loss-of-function mutations and *BCOR* mutations) (Ernst et al., 2010;

Schaefer et al., 2021; Score et al., 2012), or by enhancing their removal (*ASXL1-ASXL2* mutations) (Asada et al., 2018; Balasubramani et al., 2015; Yang et al., 2018), ultimately leading to aberrant upregulation of key PcG-repressed genes. In particular, *EZH2* mutations seem to cluster around nucleotides encoding the catalytic SET domain and result in reduced methyltransferase activity of EZH2, or they occur in the post-SET domain, which contributes to the active conformation of the complex (Duployez et al., 2016; Wu et al., 2013a). These mutations in AML cause reduced H3K27me3 levels and upregulation of PcG-repressed genes. ASXL1/2 are components of the BAP1 deubiquitinase complex and *ASXL1/2* mutations in AML are gain-of-function and result in increased activity of the BAP1 complex. In turn, increased BAP1 activity leads to increased removal of the repressive mark H2AK119ub and upregulation of a subset of PcG-repressed genes, resulting in a Polycomb phenotype (Asada et al., 2018; Balasubramani et al., 2015). In addition, it has been recently shown that *BCOR* and *BCORL1* mutations in AML compromise the structural integrity of the PRC1.1 complex and result in aberrant expression of PcG-repressed genes (Schaefer et al., 2021). Overall, these studies suggest that *PRC* mutations result in the reduction of repressive histone modifications and the loss of PRC repression of key target genes.

The frequent co-occurrence of *RUNX1* and *PcG* mutations suggests a selective advantage for leukaemia clones bearing this combination of mutations (Booth et al., 2018; Lehnertz et al., 2017; Sashida et al., 2014), but the mechanistic details for the co-occurrence of *RUNX1-PcG* mutations have not been investigated. The RUNX1-mediated recruitment of PcG proteins that was described in the previous chapter could provide an explanation for the relative enrichment of *PcG* mutations in leukaemias with *RUNX1* mutations and translocations. Thus, in this chapter I aimed to investigate whether RUNX1 recruits Polycomb proteins and contributes to PcG-mediated gene repression in leukaemia models, and whether *RUNX1* mutations and translocations affect the recruitment of PcG proteins.

4.2 Aims

In this chapter, MLL-AF4 ALL and RUNX1-RUNX1T1 AML leukaemic cell line models were used to determine whether RUNX1 affects the binding of Polycomb proteins and PcG-mediated gene repression in MLL-AF4 and RUNX1-RUNX1T1 leukaemias.

In addition, I aimed to determine the domain of RUNX1 that mediates the RUNX1-PRC1 interaction, and investigate whether truncating *RUNX1* mutations, including the *RUNX1-RUNX1T1* translocation, which result in loss of the C-terminal part of RUNX1, affect the ability of RUNX1 to associate with PRC1 proteins, and whether they impact the transcriptional regulation of PcG-repressed genes.

Finally, I also examined whether *PcG*-mutant RUNX1-RUNX1T1 and *RUNX1*-mutant AML patients have distinct transcriptional signatures compared to *PcG*-wildtype AML patients, and whether RUNX1 binding affects PcG recruitment and gene expression of genes deregulated in *PcG*-mutant RUNX1-RUNX1T1 and *RUNX1*-mutant patients.

4.3 Results

4.3.1 RUNX1 binding stabilises PRC1 binding at active and PcG-repressed genes in MLL-AF4 leukaemia, and contributes to PcG-mediated gene repression.

As mentioned in the previous chapter, ChIP-seq correlation analysis in MLL-AF4 ALL cells (Figure 3.5), revealed a significant chromatin co-occupancy between RUNX1 and RING1B, and more specifically approximately half of the RING1B sites are also bound by RUNX1 (Figure 4.3 a,b). Additional analysis revealed that RING1B peaks overlap not only with PcG-repressed PRC2-bound regions (Figure 4.3 a) that are enriched in H3K27me3 and H2AK119ub repressive histone marks, such as the *ASAP1* promoter, but also that RUNX1 and RING1B are bound together at active loci that are devoid of repressive PcG histone modifications as well, for example at the promoters of

active genes *FLT3* and *PAN3* (Figure 4.3 c). In addition, despite the fact that PRC1 is mostly bound at TSSs in mESCs, in MLL-AF4 cells RING1B is bound both at TSSs and at distal regulatory elements; in fact about half of RING1B peaks are found at distal elements rather than TSSs (Figure 4.3 d).

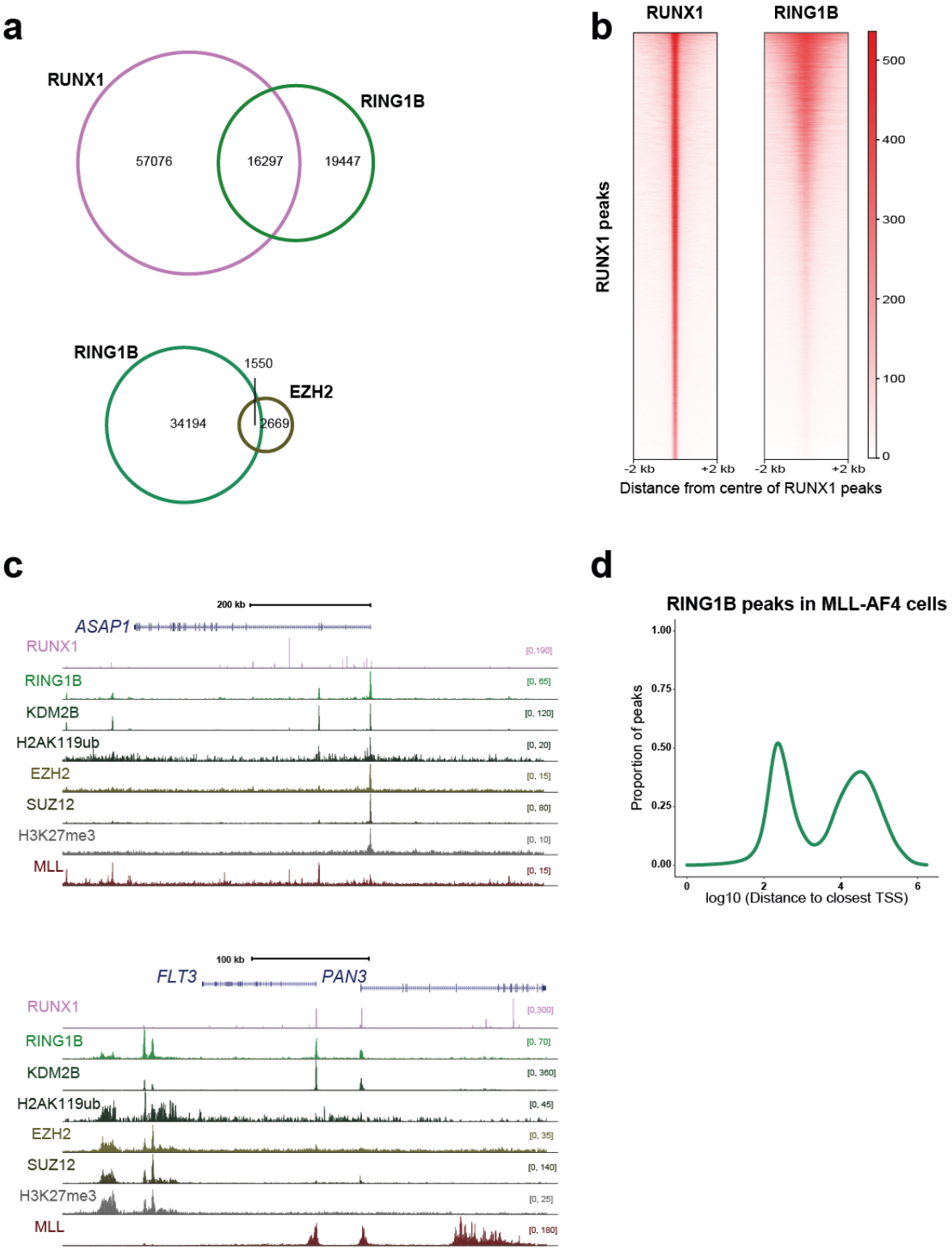


Figure 4.3: Significant chromatin co-occupancy between RUNX1 and RING1B is observed in MLL-AF4 cells. (a) *Top:* Venn diagram showing the overlap between RUNX1 and RING1B ChIP-seq peaks in the MLL-AF4 ALL cell line SEM. *Bottom:* Venn diagram showing the overlap between RING1B and EZH2 peaks in SEM cells. (b) Heatmap showing RUNX1 and RING1B ChIP-seq enrichment for a 4 kb region centred on the RUNX1 peaks. RUNX1 peaks are ranked based on the level of RUNX1 ChIP-seq signal. (c) UCSC ChIP-seq tracks for RUNX1, RING1B, KDM2B, H2AK119ub, EZH2, SUZ12, H3K27me3 and MLL showing RUNX1-RING1B co-occupancy at a PRC2-bound gene (*ASAP1* promoter) and at active genes (*FLT3* – *PAN3* promoters), not bound by PRC2. (d) Density plot showing the distribution of RING1B peaks in SEM cells as a distance from the closest TSS.

In order to examine whether RUNX1 binding affects the binding of PRC1 at PcG-repressed and at active sites in an ALL context, *RUNX1* siRNA KD was performed in SEM cells followed by ChIP-qPCR for RING1B. Electroporation with the *RUNX1* siRNA was repeated 48h after the 1st electroporation, and cells were harvested for downstream analysis 48h afterwards (96h post the 1st electroporation). Western blot for RING1B showed that *RUNX1* siRNA KD does not result in reduction in RING1B protein levels (Figure 4.4 a), and nascent RNA-seq after *RUNX1* siRNA KD showed that *RUNX1* siRNA KD does not lead to downregulation of key PRC1 components (Figure 4.4 b). ChIP-qPCR for RING1B revealed that *RUNX1* siRNA KD results in decreased RING1B binding both at PcG-repressed loci (*ASAP1, CTIF, ZIC2, CMTM7*) (Figure 4.4 c); at RUNX1-bound promoters of active genes that are devoid of repressive PcG modifications (i.e. H3K27me3 and H2AK119ub on *ELK3, CEBPG* and *FLT3*) (Figure 4.4 c,d); as well as at RUNX1-bound enhancers (for example at the *SPI1* upstream enhancer) (Figure 4.4 e). This suggests that RUNX1 contributes to PcG recruitment in MLL-AF4 cells.

To determine how RUNX1 binding affects the binding of RING1B genome-wide, I performed *RUNX1* siRNA KD followed by reference-normalised ChIP-seq for RUNX1 and RING1B in SEM cells. These results revealed that the reduction in RUNX1 binding is followed by a reduction in RING1B binding, and further analysis revealed reduction in RING1B binding both at TSSs and at distal regulatory elements (Figure 4.4 d,e,f,g). By overlapping the distal RING1B sites (~20,000 sites) with H3K27ac ChIP-seq peaks in SEM cells, I identified globally RING1B-bound enhancer elements (~6,000 sites), and meta-analysis of reference-normalised RING1B ChIP-seq showed significant reduction of

RING1B binding at these sites as well after *RUNX1* KD (Figure 4.4 g). This suggests that the RUNX1-RING1B association results in the stabilisation of PRC1 binding at the promoters of active and PcG-repressed genes, and also at enhancers.

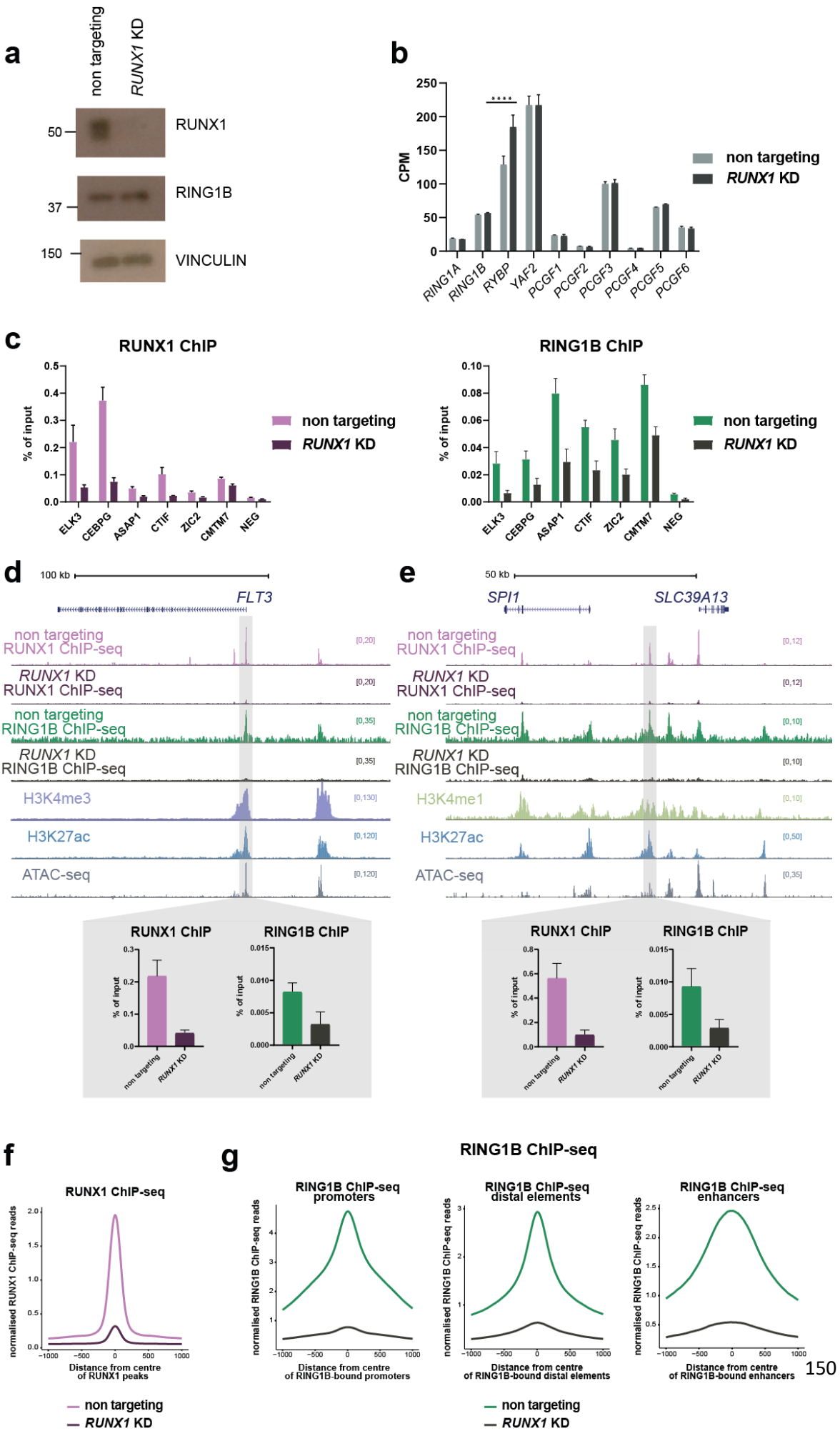

a
non targeting
RUNX1 KD
50
RUNX1
37
RING1B
150
VINCULIN
b
CPM
****
non targeting
RUNX1 KD
RING1A
RING1B
RYBP
YAF2
PCGF1
PCGF2
PCGF3
PCGF4
PCGF5
PCGF6
c
RUNX1 ChIP
% of input
non targeting
RUNX1 KD
ELK3
CEBPG
ASAP1
CTIF
ZIC2
CMTM7
NEG
RING1B ChIP
d
100 kb
FLT3
non targeting RUNX1 ChIP-seq
RUNX1 KD RUNX1 ChIP-seq
non targeting RING1B ChIP-seq
RUNX1 KD RING1B ChIP-seq
H3K4me3
H3K27ac
ATAC-seq
e
50 kb
SPI1
SLC39A13
H3K4me1
f
RUNX1 ChIP-seq
normalised RUNX1 ChIP-seq reads
Distance from centre of RUNX1 peaks
g
RING1B ChIP-seq
RING1B ChIP-seq promoters
RING1B ChIP-seq distal elements
RING1B ChIP-seq enhancers
normalised RING1B ChIP-seq reads
Distance from centre of RING1B-bound promoters
Distance from centre of RING1B-bound distal elements
Distance from centre of RING1B-bound enhancers

Figure 4.4: RUNX1 stabilises RING1B binding at promoters and distal regulatory elements in MLL-AF4 cells. (a) Western blot for RUNX1, RING1B and VINCULIN control after *RUNX1* siRNA KD or non targeting control in SEM cells. (b) Quantification of nascent RNA-seq levels for key PRC1 components in SEM cells after *RUNX1* siRNA knockdown or non targeting control. The mean of three biological replicates is shown and error bars represent the standard error of the mean (SEM). *RYBP* is upregulated after *RUNX1* siRNA KD (****FDR < 0.0001), and the transcriptional changes for the other PRC1 components are not statistically significant. (c) ChIP-qPCR for RUNX1 (*left)* and RING1B *(right)* in SEM cells after *RUNX1* siRNA KD or non targeting control at active genes (*ELK3, CEBPG)* and PcG-repressed genes (*ASAP1, CTIF, ZIC2, CMTM7)*. (d), (e) UCSC ChIP-seq tracks for reference-normalised RUNX1 and RING1B ChIP-seq in SEM cells after *RUNX1* siRNA KD and non targeting control at *FLT3* gene *(left)* and *SPI1* gene *(right).* H3K4me3, H3K4me1 and H3K27ac ChIP-seq and ATAC-seq tracks are also shown. ChIP-qPCR for RUNX1 and RING1B after *RUNX1* siRNA KD or non targeting control at *FLT3* promoter and *SPI1* upstream enhancer are shown below the ChIP-seq tracks. ChIP-qPCR experiments were performed in 3 biological replicates and error bars represent the standard error of the mean (SEM). (f) Metaplot of mean normalised RUNX1 ChIP-seq reads at RUNX1 peaks in SEM cells after *RUNX1* siRNA KD or non targeting control. (g) Metaplot of mean normalised RING1B ChIP-seq reads at RING1B-bound TSSs, RING1B-bound distal sites and RING1B-bound enhancers in SEM cells after *RUNX1* siRNA KD or non targeting control.

Interestingly, published ChIP-seq data for variant PRC1.1 component KDM2B from our lab (Kerry et al., 2017) show significant overlap between RUNX1 and KDM2B peaks (Figure 4.5 a) and also that a significant proportion of KDM2B peaks are found at distal sites (Figure 4.5 b). I also observed RUNX1-KDM2B co-occupancy at promoters of active genes (for example the promoters of *CEBPG* and *EZH2 gene*s) (Figure 4.5 c,e) and enhancers (for example the enhancer elements around *SPI1* gene and upstream of the *EZH2* locus (Figure 4.5 d,e). Similarly ChIP-qPCR results show enrichment for cPRC1 proteins CBX4 and PHC2 (the most highly expressed of CBX and PHC proteins in SEM cells respectively) (Figure 4.5 f) not only at PcG-repressed sites, but also at active promoters and enhancers bound by RUNX1 (Figure 4.5 g). The main mechanism for the recruitment of PRC1.1 is the association of KDM2B with unmethylated CpG islands mostly found at gene promoters (Farcas et al., 2012; He et al., 2013; Wu et al., 2013b) and in turn, cPRC1 proteins are mainly recruited through the association of the chromodomain of CBX proteins with H3K27me3 at repressed sites (Bernstein et al., 2006b; Cao et al., 2002; Fischle et al., 2003; Min et al., 2003). In addition, recently the Ph proteins (homologues of mammalian PHC proteins) have been identified to be bound at active sites in *Drosophila melanogaster* both at enhancers and promoters (Loubiere et al., 2020).

Furthermore, in breast cancer cells CBX4 has been demonstrated to be bound at active promoters of genes that become upregulated after oestrogen administration (Zhang et al., 2021). My results showing significant reduction of RING1B binding after *RUNX1* siRNA KD at active promoters and enhancer elements, in combination with the observations that canonical and variant PRC1 complexes are bound at these sites, and also the TetR results in the previous chapter, suggest a role for mammalian TFs in stabilising the binding of canonical and variant PRC1 complexes at active loci.

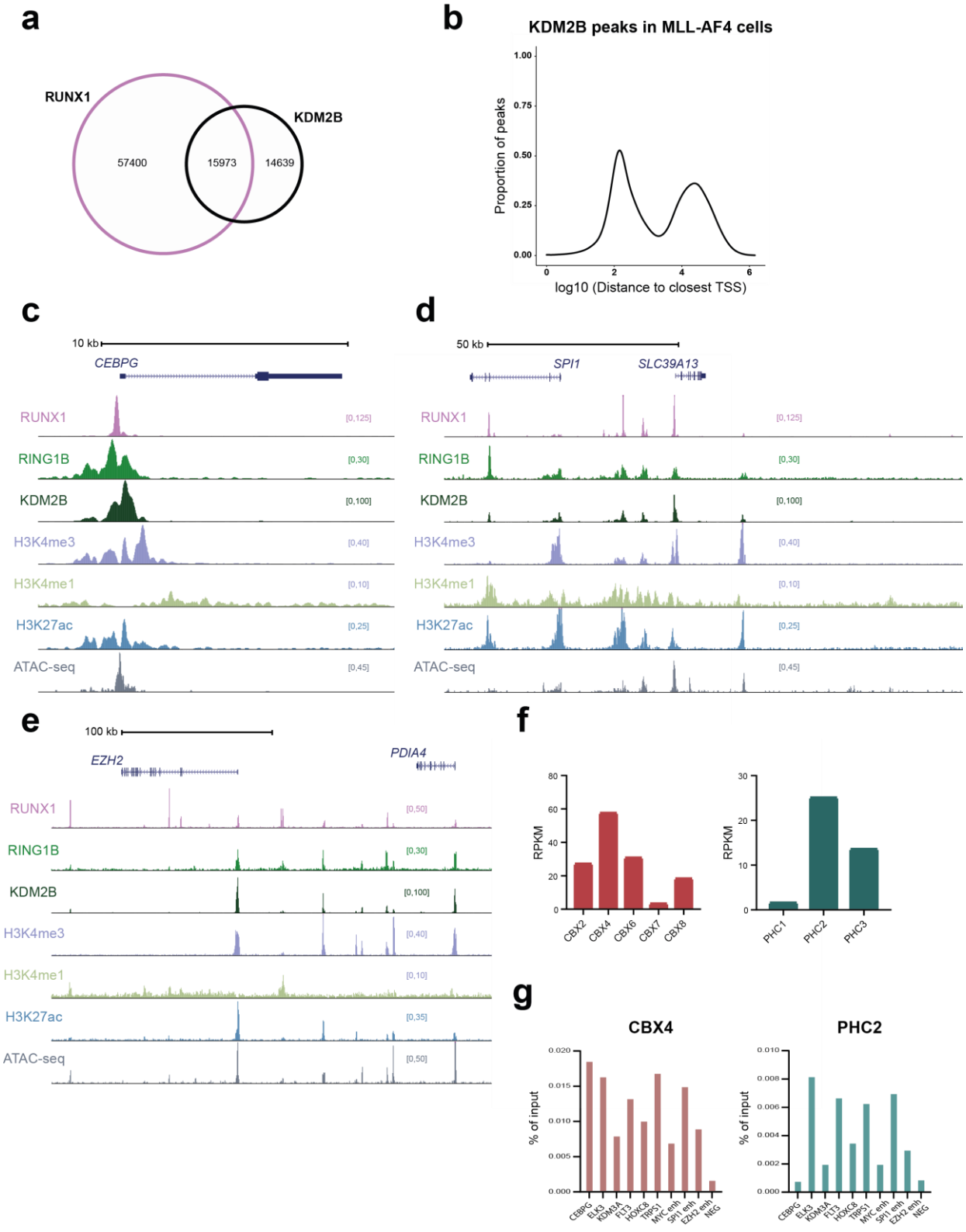


Figure 4.5: Canonical and variant PRC1 complexes are bound at RUNX1-bound active promoters and enhancers in MLL-AF4 cells. (a) Venn diagram showing the overlap of RUNX1 and KDM2B ChIP-seq peaks in SEM cells. (b) Density plot showing the distribution of KDM2B peaks in SEM cells as a distance from the closest TSS. (c), (d), (e) UCSC ChIP-seq tracks for RUNX1, RING1B, KDM2B, H3K4me3, H3K4me1, H3K27ac and ATAC-seq tracks at *CEBPG* gene (c) and *SPI1* (d) and *EZH2* genes (e) and upstream regulatory elements, showing KDM2B binding at RUNX1-bound active genes and enhancer elements. (f) RNA-seq data for expression of CBX and PHC proteins, components of cPRC1

complexes, in SEM cells. (g) ChIP-qPCR for cPRC1 proteins CBX4 and PHC2 at RUNX1-bound active genes (*CEBPG, ELK3, KDM3A, FLT3)*, PcG-repressed genes (*HOXC8, TRPS1)* and RUNX1-bound enhancers (*MYC, SPI1* and *EZH2* enhancers) in SEM cells. One representative ChIP-qPCR replicate is shown.

I next wanted to examine whether the stabilisation of RING1B binding at repressed loci by RUNX1 contributes to PcG-mediated gene repression. In order to determine how the reduction of RUNX1 and subsequently RING1B binding affects gene expression at PcG-repressed genes, I performed nascent RNA-seq after *RUNX1* siRNA KD. This revealed that ~350 genes get downregulated after *RUNX1* KD, and that ~400 genes get upregulated (Figure 4.6 a), including a subset of PcG-repressed genes that becomes upregulated after *RUNX1* siRNA KD. More specifically, about 10% of PcG-repressed genes (marked by H3K27me3) seem to become upregulated after *RUNX1* KD (Figure 4.6 b), including the RUNX1-bound PcG-repressed genes *PLCB1, MAP7, CMTM7, NCKAP1, FRMPD1* and *NR3C2* (Figure 4.6 c). These results confirm that a key function of RUNX1 in ALL is to contribute to PcG-mediated recruitment and gene repression, at least at a subset of loci, similarly to my results in mESCs in which RUNX1 was artificially expressed.

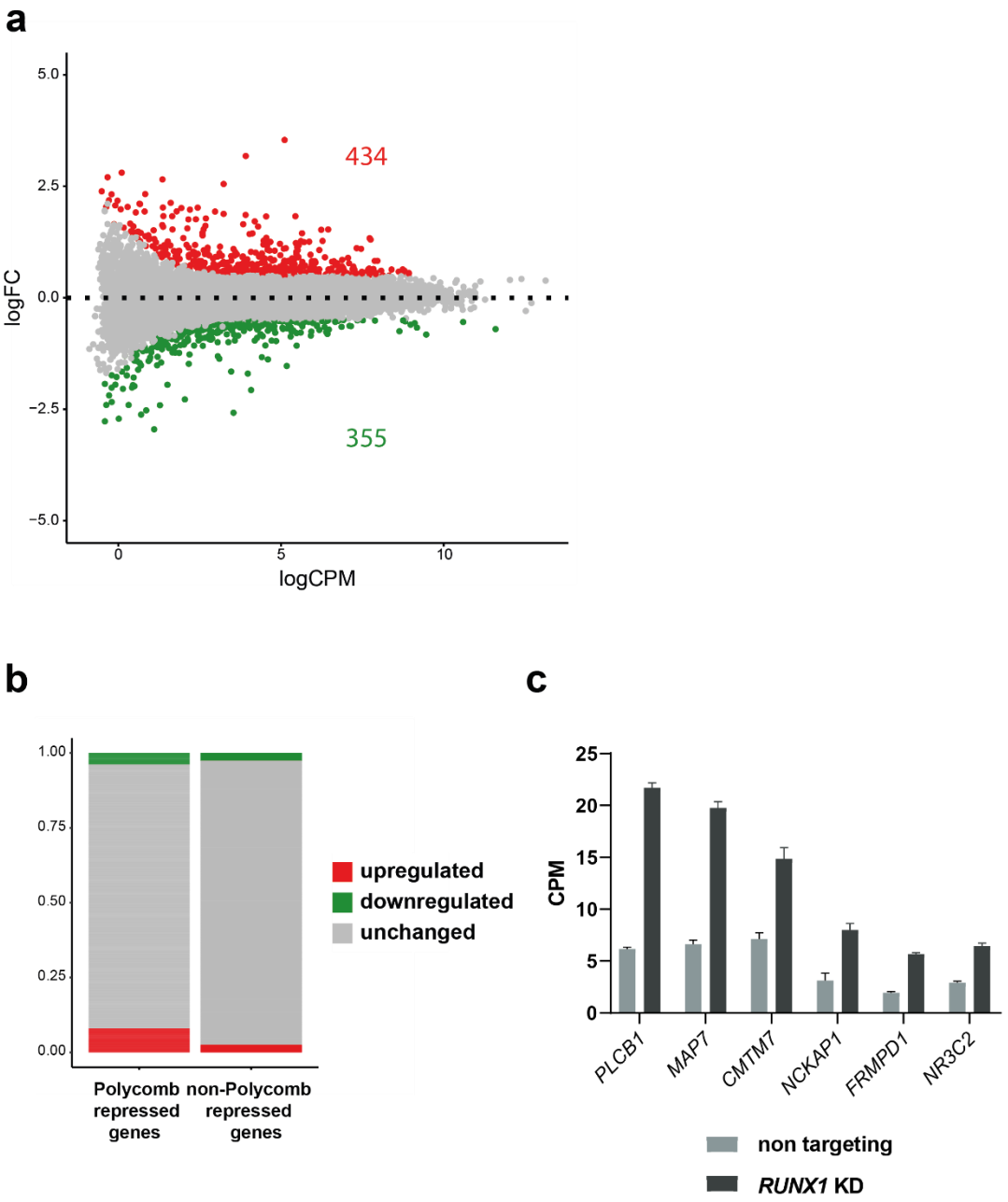


Figure 4.6: A subset of PcG-repressed genes become upregulated after *RUNX1* siRNA KD in MLL-AF4 cells. (a) MA plot showing genes differentially expressed in SEM cells after *RUNX1* siRNA KD compared to non targeting control, after performing nascent RNA-seq. Upregulated genes (n=434, logFC > 0.5, FDR <0.05) are shown in red, downregulated genes (n=355, logFC < -0.5, FDR <0.05) are shown in green, and unchanged genes in grey. Nascent RNA-seq was performed in 3 biological replicates per condition. (b) Bar plots showing the proportion of PcG-repressed genes (defined as marked by H3K27me3) and non-Polycomb repressed genes that get upregulated, downregulated or remain unchanged after *RUNX1* siRNA knockdown. (c) Quantification of nascent RNA-seq levels in SEM cells after *RUNX1* siRNA knockdown or non targeting control for RUNX1-bound genes *PLCB1, MAP7, CMTM7, NCKAP1, FRMPD1* and *NR3C2* that are bound by the PRC2 repressive complex. These genes become upregulated after *RUNX1* siRNA KD (FDR < 0.0001). The mean of three biological replicates is shown and error bars represent the standard error of the mean (SEM).

4.3.2 RUNX1 mediates recruitment of OGT at active PRC1-bound sites.

Recent observations in mESCs showed that PRC1 and BAP1 complexes, which have opposite activities, are found at distinct sites and do not overlap (Conway et al., 2021). In contrast, my TetR data showed RUNX1-mediated recruitment of BAP1 and OGT (Figure 3.8), in addition to RUNX1-mediated recruitment of PRC1. Thus, I wanted to examine whether RUNX1 contributes to the recruitment of the BAP1 complex in addition to PRC1 complex in MLL-AF4 cells. By overlapping the RUNX1 and RING1B ChIP-seq peaks in MLL-AF4 cells with the ChIP-seq peaks of the BAP1 component OGT, I observed significant chromatin co-occupancy of the three (Figure 4.7 a). I also observed significant overlap between RING1B and OGT at active loci. More specifically, by overlapping the RING1B peaks with EZH2 peaks in SEM cells, I went on to divide the RING1B peaks into those found at PcG-repressed loci (that overlap with EZH2), and those found at active sites (Figure 4.7 b). This showed that the RING1B-OGT co-binding mostly occurs at active sites (for example at the promoters of active genes *ELK3* and *CEBPG*) (Figure. 4.7 b,c). In fact, about half of the OGT peaks overlap with RING1B non-PRC2 peaks at active sites, while a very small proportion of OGT peaks are found at PcG-repressed sites (Figure 4.7 b).

Given that the results in the TetR system show that, apart from PRC1, RUNX1 is sufficient to recruit the BAP1 complex as well, and taking into consideration the significant overlap of RUNX1-RING1B-OGT in MLL-AF4 cells, I next wanted to determine how RUNX1 binding affects binding of OGT. After performing *RUNX1* siRNA KD followed by ChIP-qPCR for OGT, we observed that reduction in RUNX1 binding results in reduced binding of OGT at active RUNX1-bound sites (Figure 4.7 d). Nascent RNA-seq results after *RUNX1* siRNA KD did not show significant transcriptional changes in components of the BAP1 complex (Figure 4.7 e), suggesting that the reduced OGT binding is not due to reduced expression of the proteins of the complex. In addition to other mechanisms that have been proposed to inhibit the formation of a repressive Polycomb domain at active sites (the activity of AUTS2 (Gao et al., 2014) and Aurora B kinase (Frangini et al., 2013) and the presence of activating signals and ongoing transcription (Klose et al., 2013)) (described in Figure 4.7 f), RUNX1-mediated

recruitment of BAP1 could also result in the removal of RING1B-deposited H2AK119ub. The removal of this repressive histone mark could potentially function as an alternative mechanism that impedes the generation of a PcG-repressed locus at active sites (Figure 4.7 g).

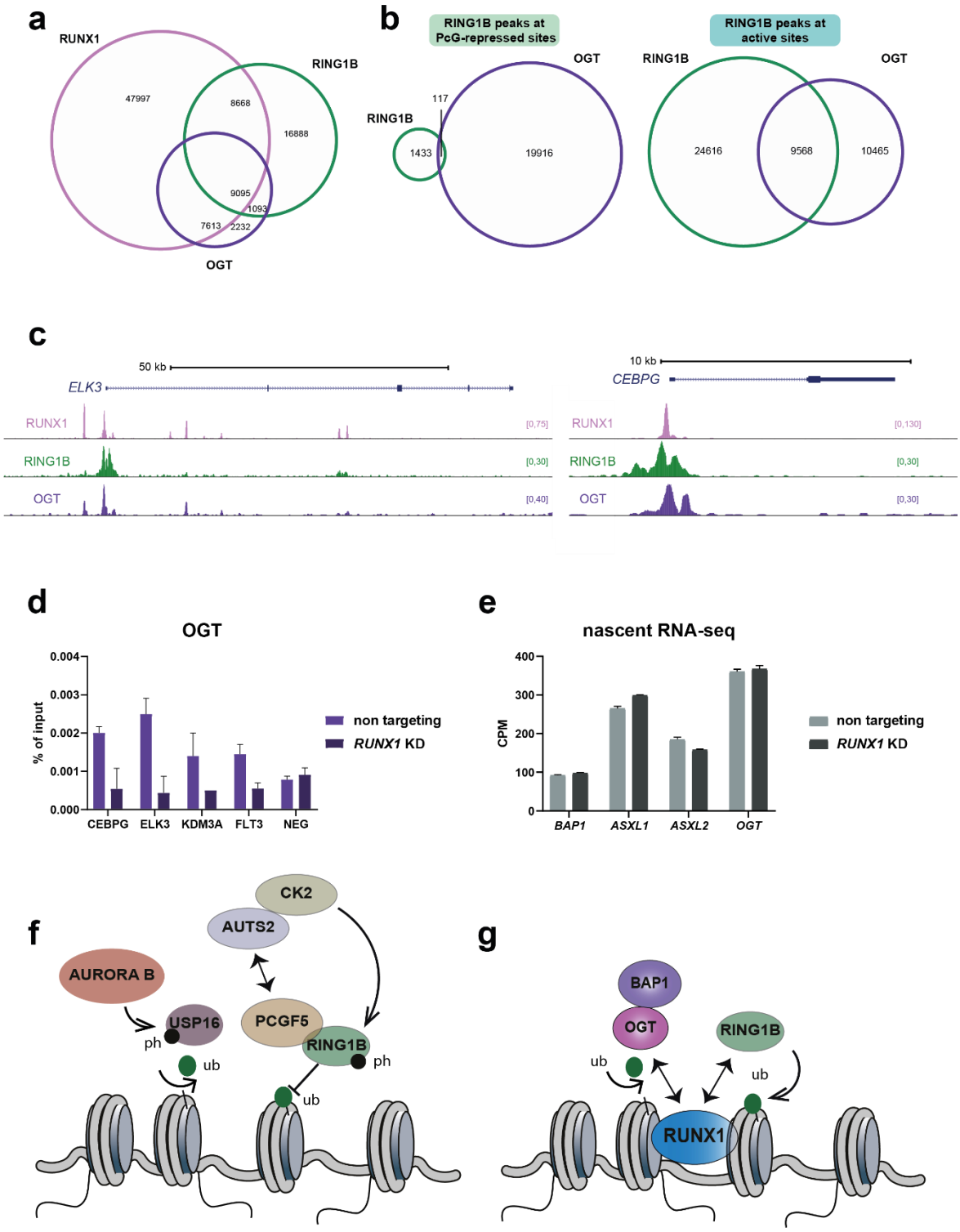


Figure 4.7: RUNX1 stabilises OGT binding in MLL-AF4 cells. (a) Venn diagram showing the overlap between RUNX1, RING1B and OGT ChIP-seq peaks in SEM cells. (b) Venn diagram showing the

overlap between OGT peaks and RING1B peaks found at active sites (not PRC2-bound) and PcG-repressed sites (PRC2-bound). (c) ChIP-seq tracks for RUNX1, RING1B and OGT at active sites *ELK3* and *CEBPG.* (d) ChIP-qPCR for OGT in SEM cells after *RUNX1* siRNA KD or non targeting control. ChIP-qPCR experiments were performed in 3 biological replicates, and error bars show the standard error of the mean (SEM). (e) Quantification of nascent RNA-seq levels in SEM cells after *RUNX1* siRNA knockdown or non targeting control for the components of the BAP1 deubiquitinase complex *BAP1, ASXL1, ASXL2* and *OGT*. The mean of three biological replicates is shown, and error bars represent the standard error of the mean (SEM). Changes in the expression levels of these genes are not statistically significant (defined as logFC > 0.5 or logFC <-0.5, FDR < 0.05). (f) Schematic of mechanisms leading to reduced H2AK119ub at active sites. TF-mediated recruitment of Aurora B kinase at active loci phosphorylates the deubiquitinase USP16, which removes the H2AK119ub mark. In addition, the association between PRC1.5 complex (comprising of PCGF5 subunit) with the AUTS2-CK2 complex leads to CK2-mediated phosphorylation of RING1B, which reduces its ability to monoubiquitinate H2AK119. The figure was adapted from (Gao et al., 2014) and (Frangini et al., 2013). (g) Schematic showing that RUNX1-mediated recruitment of OGT (component of the BAP1 deubiquitinase complex) could remove H2AK119ub, deposited by RING1B, at active loci.

4.3.3 The C-terminal part of Runx1, lost in RUNX1 truncating mutations and leukaemia-associated RUNX1 translocations, is sufficient for Polycomb recruitment.

4.3.3.1 The N-terminal part of Runx1 is not sufficient for PRC1 recruitment.

Given the increased frequency of *RUNX1* mutations and translocations in human myeloid leukaemias and MDS (Sood et al., 2017), I next decided to investigate whether RUNX1 contributes to Polycomb recruitment in myeloid leukaemia cells. Interestingly, as mentioned in the introduction of this chapter, it has been previously reported that there is a relative enrichment for *PcG* mutations in *RUNX1*-mutant leukaemias, such as *RUNX1*-mutant MDS, ETP-ALL and AML (Booth et al., 2018; Gaidzik et al., 2016; Lehnertz et al., 2017; Ochi et al., 2020; Sashida et al., 2014; Simon et al., 2017; Wu et al., 2019), including RUNX1-RUNX1T1 AML (Duployez et al., 2016; Faber et al., 2016; Micol et al., 2014).

Given the frequent co-occurrence of *PcG* mutations in *RUNX1*-mutant leukaemias, I hypothesised that the combination of *RUNX1* and *PcG* mutations would result in a selective advantage of these leukaemia clones. Taking into consideration the findings that a key function of RUNX1 is to recruit

PcG proteins and contribute to PcG-mediated gene repression, I hypothesised that in the case of *RUNX1*/PcG-mutant leukaemia, *RUNX1* mutations would result in decreased PcG binding and, thus, functional co-operation between *RUNX1* and *PcG* mutations (Figure 4.8). Specifically, I hypothesised firstly that *RUNX1* mutations in its DNA binding domain, that lead to reduced binding ability of the mutant proteins to RUNX1 targets, would also result in reduced PcG binding at RUNX1-bound PcG-repressed loci (Figure 4.8 a). In addition, given the increased frequency of *PcG* mutations together with *RUNX1* mutations, including mutations resulting in truncated RUNX1 mutant proteins and in RUNX1-RUNX1T1 leukaemias, I hypothesised the truncated RUNX1 mutant proteins fail to associate with PcG proteins, resulting in decreased PcG binding at key loci (Figure 4.8 b). The additional *PcG* mutations that negatively affect the repressive activity of PcG proteins may also result in upregulation of a subset of PcG-repressed genes.

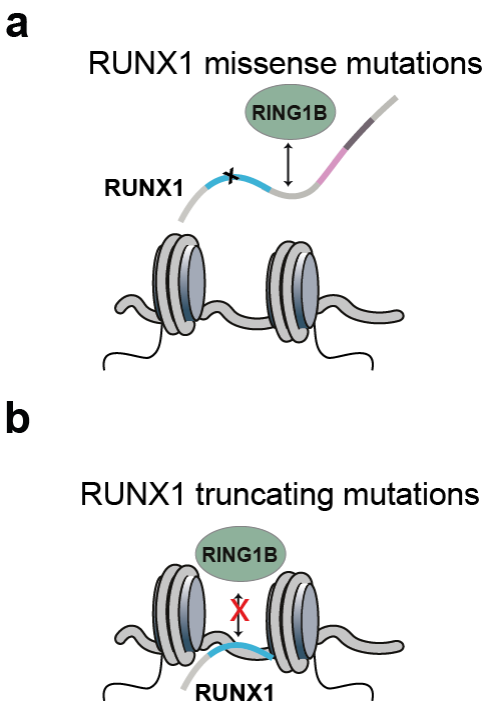


Figure 4.8: Schematic of interaction between RUNX1 mutant proteins and PRC1. Schematic of the hypothesis that RUNX1 missense mutations (a), which function by reducing the DNA binding ability of RUNX1, also reduce the binding of PRC1. (b) Schematic of the hypothesis that truncating RUNX1 mutations that result in the loss of the C-terminal part of RUNX1 result in loss of RUNX1-PRC1 interaction, leading to reduced PRC1 binding at RUNX1 sites.

To test the hypothesis that truncating *RUNX1* mutations including the *RUNX1-RUNX1T1* translocation disrupt RUNX1 mediated recruitment of PcG proteins, I went back to the TetR system. To identify domains of RUNX1 important for PRC1 recruitment, I created stable mESC lines expressing the N-terminal part of Runx1 (aa 1-177, including the DNA-binding Runt domain) fused to the TetR, from now on referred to as Runx1(1-177aa)-TetR cells (Figure 4.9 a). In order to test whether the N-terminal part of Runx1 fused to the TetR is sufficient to recruit known interaction partners, I performed ChIP-qPCR for Cbfb, the co-factor of Runx1, in Runx1(1-177aa)-TetR cells. Cbfb associates with the DNA binding domain of Runx1 that is included in Runx1(1-177aa)-TetR fusion protein, and consistent with this, ChIP-qPCR showed that Cbfb is recruited to the TetO locus in Runx1(1-177aa)-TetR cells, in contrast to TetR-only cells (Figure 4.9 b). In addition, after 6h treatment of the Runx1(1-177aa)-TetR cells with doxycycline, Cbfb binding at the TetO locus is abolished (Figure 4.9 b) confirming that the fusion protein is functional and sufficient to recruit Cbfb.

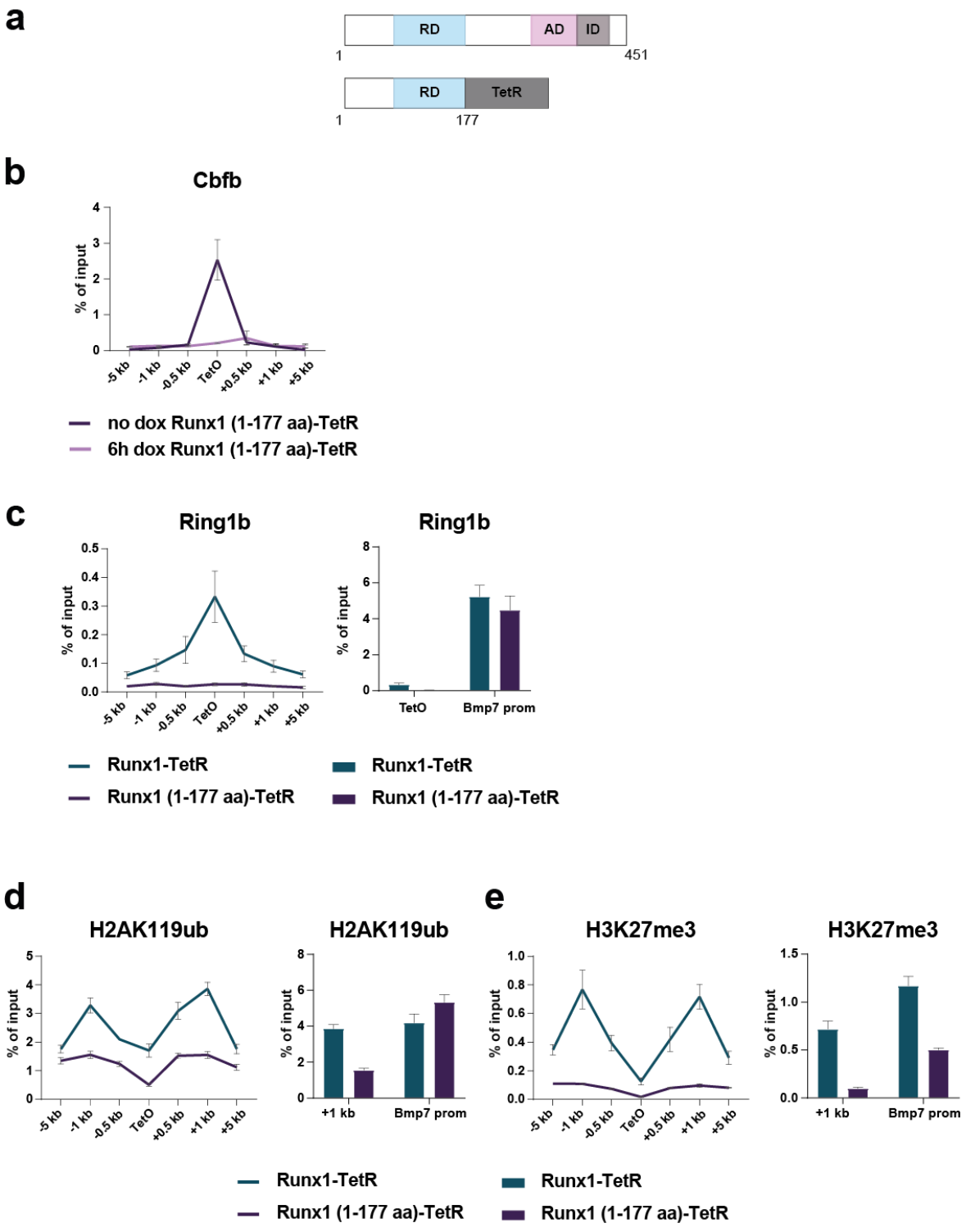


Figure 4.9: The N-terminal part of RUNX1 is not sufficient to recruit PRC1. (a) Schematic of the full-length Runx1 protein and the N-terminal of Runx1 (1-177 aa) fused to TetR (RD=Runt domain (DNA-binding domain), AD=Activation domain, ID=Inhibitory domain). (b) ChIP-qPCR for Cbfb in Runx1 (1-177 aa)-TetR mESCs with/without dox treatment showing that the N-terminal part of Runx1 is sufficient to recruit Cbfb. (c)-(e) ChIP-qPCR for Ring1b (c), H2AK119ub (d) and H3K27me3 (e) at the TetO locus and at an endogenous control (*Bmp7* promoter) in mESCs expressing Runx1-TetR as control (*blue*) or Runx1 (1-177 aa)-TetR (*purple)*, revealing that the N-terminal part of Runx1, unlike full-length Runx1, is not sufficient to establish a Polycomb domain. ChIP-qPCR

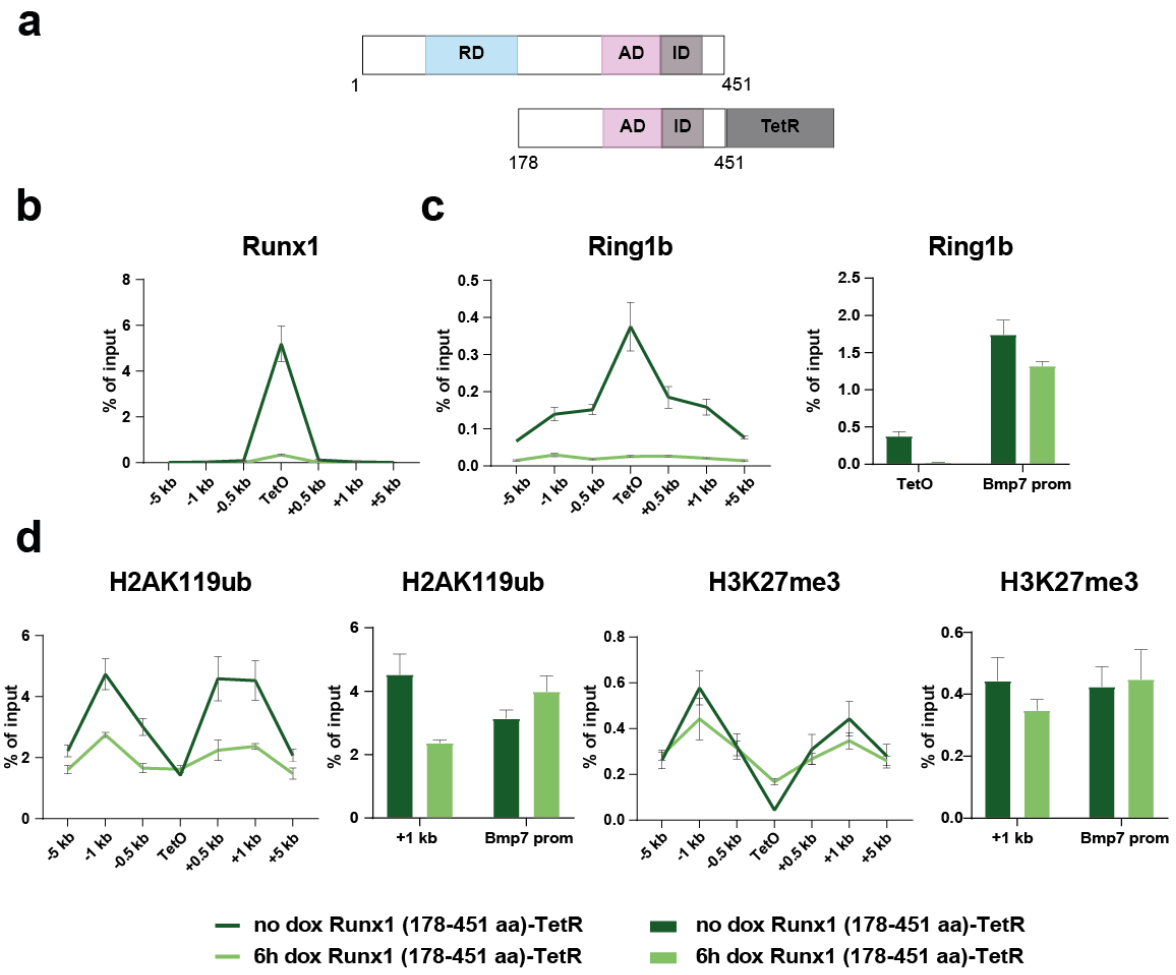
experiments were performed in 3 biological replicates and error bars represent the standard error of the mean (SEM).

To test whether the N-terminal part of Runx1 is sufficient to recruit PRC1, I performed ChIP-qPCR for Ring1b at the TetO locus of Runx1(1-177aa)-TetR cells. In contrast to the Runx1-TetR cells, in which the full length protein is fused to the TetR, in the Runx1(1-177aa)-TetR cells, Ring1b binding was not detected at the TetO locus (Figure 4.9 c). Similarly, in contrast to Runx1-TetR cells, ChIP-qPCR for the repressive PcG modifications H2AK119ub and H3K27me3 did not show enrichment for these marks at the TetO locus of Runx1(1-177aa)-TetR cells (Figure 4.9 d,e). These data suggest that the N-terminal part of RUNX1, which is retained in mutations that result in truncated RUNX1 proteins and in the RUNX1-RUNX1T1 fusion protein, is not sufficient to recruit PRC1.

4.3.3.2 The C-terminal part of Runx1 mediates the recruitment of PRC1.

To further validate the hypothesis that it is the C-terminal part of Runx1 that is actually mediating the recruitment of PRC1, I created stable mESC lines expressing the C-terminal part of Runx1 (aa 178-451, i.e. the part after the DNA binding domain) fused to TetR, from now on referred to as Runx1(178-451aa)-TetR cells (Figure 4.10 a). This is the part of RUNX1 that is not included in the RUNX1-RUNX1T1 fusion protein and is often lost in *RUNX1* truncating mutations. The C-terminal part of Runx1 is recognised by the Runx1 antibody, and ChIP-qPCR for Runx1 in Runx1(178-451aa)-TetR cells validated the binding of the fusion protein at the TetO locus, which is abolished after 6h treatment with doxycycline (Figure 4.10 b). ChIP-qPCR for Ring1b at the TetO locus of Runx1(178-451aa)-TetR cells showed that, consistent with our hypothesis, Ring1b was recruited to the TetO locus. Treatment with doxycycline for 6h resulted in reduction in Ring1b binding following loss of Runx1 binding (Figure 4.10 c). Similarly to the results with the full length Runx1 protein fused to the TetR, tethering the C-terminal part of Runx1 to TetR resulted in significant enrichment in H2AK119ub and H3K27me3 at the nucleosomes flanking the TetO locus (Figure 4.10 d). This

demonstrates that the C-terminal part of Runx1 is sufficient to recruit PRC1 and result in the formation of a Polycomb domain that spreads a few kbs, when anchored to a gene-poor region (Figure 4.10 e).



e

RUNX1 fragment	PcG recruitment
RD AD ID (1–451)	✓
RD (1–177) N-terminal	✗
C-terminal AD ID (178–451)	✓

Figure 4.10: The C-terminal part of RUNX1, lost in the RUNX1-RUNX1T1 fusion protein and in truncated RUNX1 mutant proteins, is sufficient to recruit PRC1 and establish a Polycomb domain. (a) Schematic of the full-length Runx1 protein and the C-terminal of Runx1 (178-451 aa) fused to TetR. (b) – (d) ChIP-qPCR for Runx1 (b), Ring1b (c), H2AK119ub and H3K27me3 (d) at the TetO locus

and at an endogenous locus (*Bmp7* promoter) in mESCs expressing Runx1 (178-451 aa)-TetR with/without doxycycline treatment, revealing that the C-terminal part of Runx1 is sufficient to establish a Polycomb domain. ChIP-qPCR experiments were performed in 3 biological replicates and error bars show the standard error of the mean (SEM). (e) Schematic table illustrating that the RUNX1-PRC1 interaction is mediated by the C-terminal part of RUNX1 (aa 178-451), while the N-terminal part of RUNX1 (aa 1-177) does not associate with RING1B.

Taken together, my results show that RUNX1 is sufficient for the recruitment of PRC1 and this interaction is mediated by the C terminal "half" of RUNX1. These results provide a potential mechanistic explanation for the co-occurrence and functional cooperation between *PcG* mutations and truncating *RUNX1* mutations and translocations, as the latter seem to function, at least in part, by abolishing the RUNX1-PRC1 interaction.

4.3.4 RUNX1 binding stabilises PRC1 binding at PcG-repressed sites in RUNX1-RUNX1T1 leukaemia.

The C-terminal part of RUNX1 is not included in the RUNX1-RUNX1T1 fusion proteins that arise due to the *t(8;21)* translocation. This further suggests that the fusion protein, unlike wildtype RUNX1, is not sufficient for the recruitment of PcG proteins. To test whether wildtype RUNX1, contributes to PcG recruitment and gene repression in RUNX1-RUNX1T1 leukaemias, I set out to investigate how *RUNX1* siRNA KD affects the binding of PcG proteins. Firstly, I identified PcG-repressed TSSs that are bound by wildtype RUNX1 (and the co-factor of RUNX1 CBFb) and not the fusion protein in SKNO-1 cells (RUNX1-RUNX1T1 leukaemia cell line) from publicly available datasets (Martens et al., 2012; Micol et al., 2017). There is not an extensive overlap between RUNX1 binding sites and loci enriched in H3K27me3, but there are approximately 1,000 sites of RUNX1-H3K27me3 overlap (Figure 4.11 a, b). In contrast, the overlap between RUNX1-RUNX1T1 fusion protein and H3K27me3 was rather weak and limited to ~ 200 sites (Figure 4.11 a, b).

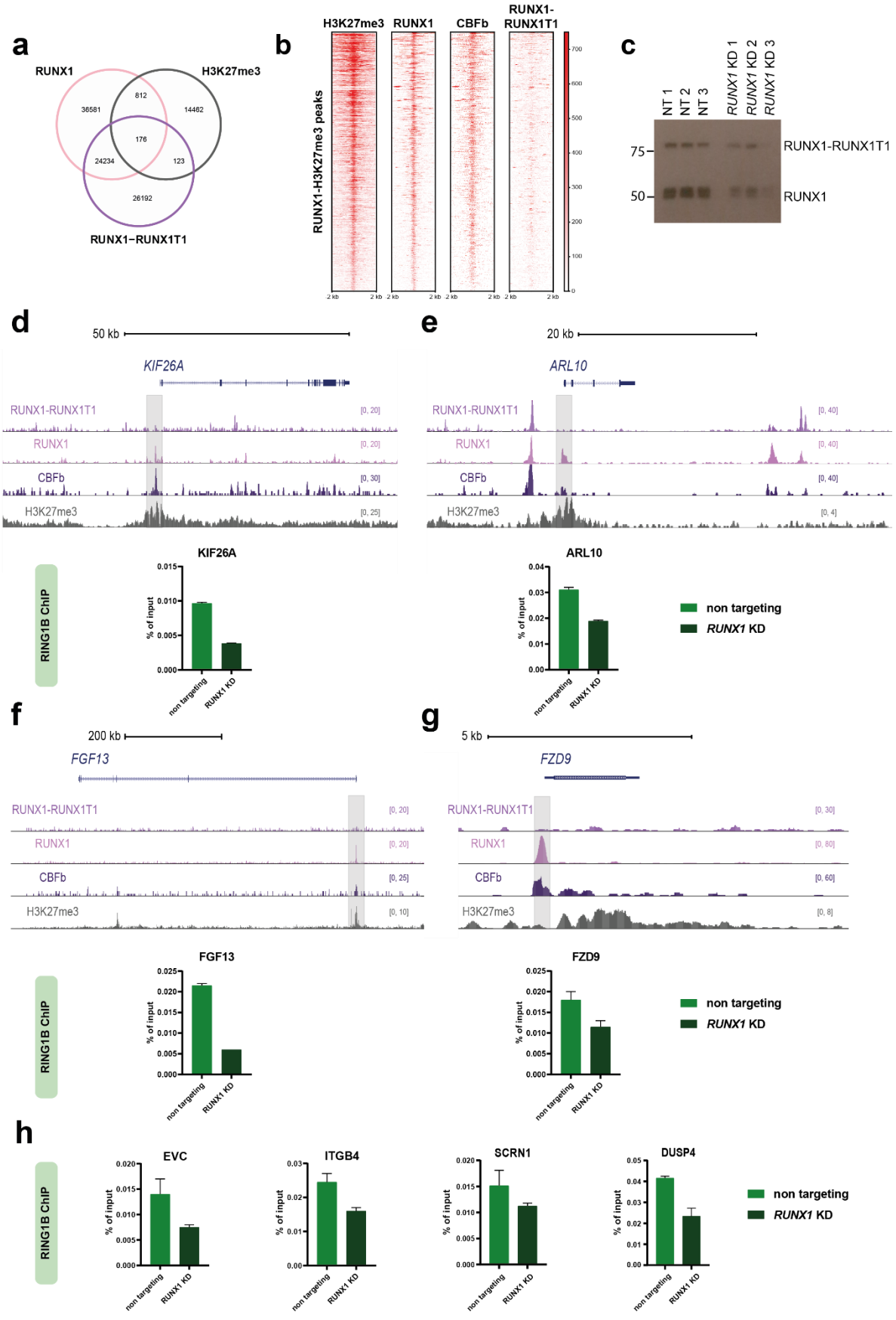

a
RUNX1
H3K27me3
RUNX1–RUNX1T1
36581
812
14462
176
24234
123
26192
b
H3K27me3
RUNX1
CBFb
RUNX1-RUNX1T1
RUNX1-H3K27me3 peaks
c
NT 1
NT 2
NT 3
RUNX1 KD 1
RUNX1 KD 2
RUNX1 KD 3
75
50
RUNX1-RUNX1T1
RUNX1
d
50 kb
KIF26A
RUNX1-RUNX1T1
RUNX1
CBFb
H3K27me3
e
20 kb
ARL10
RING1B ChIP
KIF26A
ARL10
% of input
non targeting
RUNX1 KD
f
200 kb
FGF13
g
5 kb
FZD9
h
EVC
ITGB4
SCRN1
DUSP4

Figure 4.11: RUNX1 recruits RING1B at PcG-repressed loci in RUNX1-RUNX1T1 AML cells. (a) Venn diagram showing the overlap between RUNX1, RUNX1T1 and H3K27me3 ChIP-seq peaks in SKNO-1 RUNX1-RUNX1T1 cell line. (b) Heatmaps showing H3K27me3, RUNX1, CBFb and RUNX1-RUNX1T1 ChIP-seq enrichment for a 4 kb region centred on the H3K27me3 peaks in SKNO-1 cells that overlap with RUNX1 peaks. (c) Western blot for RUNX1 (and RUNX1-RUNX1T1) after *RUNX1* siRNA KD or non targeting control (NT) in SKNO-1 cells. UCSC ChIP-seq tracks for RUNX1-RUNX1T1, RUNX1, CBFb and H3K27me3 in SKNO-1 cells at (d) *KIF26A*, (e) *ARL10,* (f) *FGF13* and (g) *FZD9* loci, showing that these PcG-repressed sites are bound by RUNX1, but not the fusion protein RUNX1-RUNX1T1. Below the ChIP-seq tracks ChIP-qPCR results for RING1B are shown after *RUNX1* siRNA knockdown or non targeting control, demonstrating that RUNX1 stabilises RING1B binding. (h) ChIP-qPCR for RING1B in SKNO-1 cells after *RUNX1* siRNA knockdown or non targeting control at PcG-repressed loci *EVC, ITGB4, SCRN1* and *DUSP4,* which are bound by RUNX1 but not RUNX1-RUNX1T1, showing reduced RING1B binding after reduction in RUNX1 binding. ChIP-qPCR experiments were performed in 2 biological replicates. RUNX1-RUNX1T1, RUNX1 and CBFb ChIP-seq in SKNO-1 cells were generated in (Martens et al., 2012) and H3K27me3 ChIP-seq in (Micol et al., 2017).

Next, I performed *RUNX1* siRNA knockdown in SKNO-1 cells, using an siRNA that targets the C-terminal part of RUNX1 that is not shared with the fusion protein (although a slight reduction in RUNX1-RUNX1T1 protein levels is also observed after *RUNX1* siRNA KD) (Figure 4.11 c), and found that *RUNX1* KD resulted in a decrease in RING1B binding at the TSSs of a subset of PcG-repressed RUNX1-bound genes (Figure 4.11 d-h). For example, the TSSs of *KIF26A, ARL10, FGF13, FZD9*, *EVC, ITGB4, SCRN1* and *DUSP4* genes are bound by RUNX1 and CBFb, but not the RUNX1-RUNX1T1 fusion protein, and are PcG-repressed as denoted by the enrichment of H3K27me3 at the same sites. At these loci, reduction in RUNX1 binding after *RUNX1* siRNA KD resulted in reduction in RING1B binding, and at some of these loci (*KIF26A, FGF13, EVC)* the reduction was more than 50% (Figure 4.11 d,f,h). These results support the conclusion that a key function of wildtype RUNX1 in RUNX1-RUNX1T1 AML is to recruit PcG proteins and contribute to PcG repression (Figure 4.12).

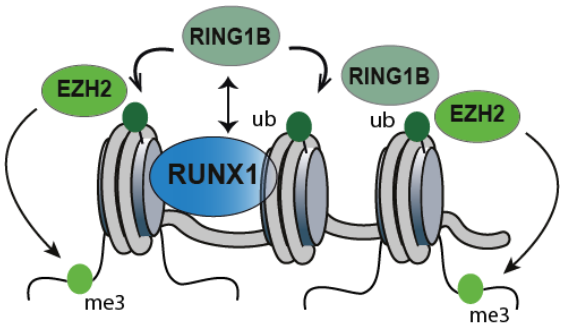


Figure 4.12: Schematic of RUNX1-mediated recruitment of PRC1 at PcG-repressed loci in RUNX1-RUNX1T1 leukaemia. At PcG-repressed loci in RUNX1-RUNX1T1 AML cells (marked by the binding of RING1B and EZH2, and the presence of the repressive histone marks H2AK119ub and H3K27me3), RUNX1 binding further contributes to the recruitment of PRC1 component RING1B.

4.3.5 Differential gene expression analysis reveals genes upregulated in *PcG*-mutant RUNX1-RUNX1T1 patients, including RUNX1-bound PcG-repressed genes.

In the RUNX1-RUNX1T1 leukaemias, additional cooperating mutations are needed for the development of leukaemia, as the fusion protein alone is not sufficient for leukaemogenesis (Higuchi et al., 2002; Yuan et al., 2001). Interestingly, similarly to other forms of leukaemias with *RUNX1* mutations, as discussed in the introduction, among the recurrent mutations in RUNX1-RUNXT1 patients are mutations in genes encoding PcG proteins (*ASXL1, ASXL2, EZH2, and BCOR*) (Christen et al., 2019; Duployez et al., 2016; Hollein et al., 2018).

Given the functional cooperation of RUNX1-RING1B in RUNX1-RUNX1T1 leukaemias, I next wanted to investigate the transcriptional signatures in *PcG*-mutant RUNX1-RUNX1T1 patients compared to *PcG*-wildtype ones, and determine if i) there are PcG-repressed genes that are upregulated in *PcG*-mutant patients and ii) whether RUNX1 contributes to recruitment of PRC1 at a subset of those genes.

To do this, I used publicly available RNA-seq data from RUNX1-RUNX1T1 patients with or without additional *PcG* mutations (Lavallee et al., 2016). These data comprised 9 patients with mutations in *ASXL1* and *ASXL2* genes and 4 patients who were *PcG*-wildtype ones. The mutations in *ASXL1/2* genes are gain-of-function mutations that lead to increased activity of the deubiquitinase BAP1 complex, which is responsible for removing H2AK119ub. As a result, these mutations lead to increased removal of the repressive H2AK119ub mark, and subsequently result to derepression of PcG targets (Asada et al., 2018; Balasubramani et al., 2015; Yang et al., 2018). In accordance with the expected number of mutations in RUNX1-RUNX1T1 patient samples (on average 2.2 additional

mutations (Christen et al., 2019)), the *PcG*-mutant and *PcG*-wildtype patients had few additional mutations, mainly *KIT* mutations and mutations in the RAS pathway (Appendix table 7.2).

By performing differential gene expression analysis in *PcG*-mutant vs *PcG*-wildtype RUNX1-RUNX1T1 patients, I identified a subset of PcG-repressed genes that become upregulated in *PcG*-mutant (*ASXL1/2*-mutant) patients (Figure 4.13 a,b). Principal component analysis (PCA) based on these differentially expressed genes showed that this ~100 gene signature separated the *PcG*-mutant and *PcG*-wildtype patient samples into two distinct clusters (Figure 4.13 c). Among the PcG-bound genes that became upregulated in *PcG*-mutant compared to *PcG*-wildtype RUNX1-RUNX1T1 patients, was a subset of RUNX1-bound PcG-repressed genes (including the *CCND1*, *RBPMS*, *KIF26B, LAMC3, MTA3, PARD3* genes) (Figure 4.13 a,b). These represent RUNX1-bound PcG-bound genes that become derepressed in patients with *PcG* mutations.

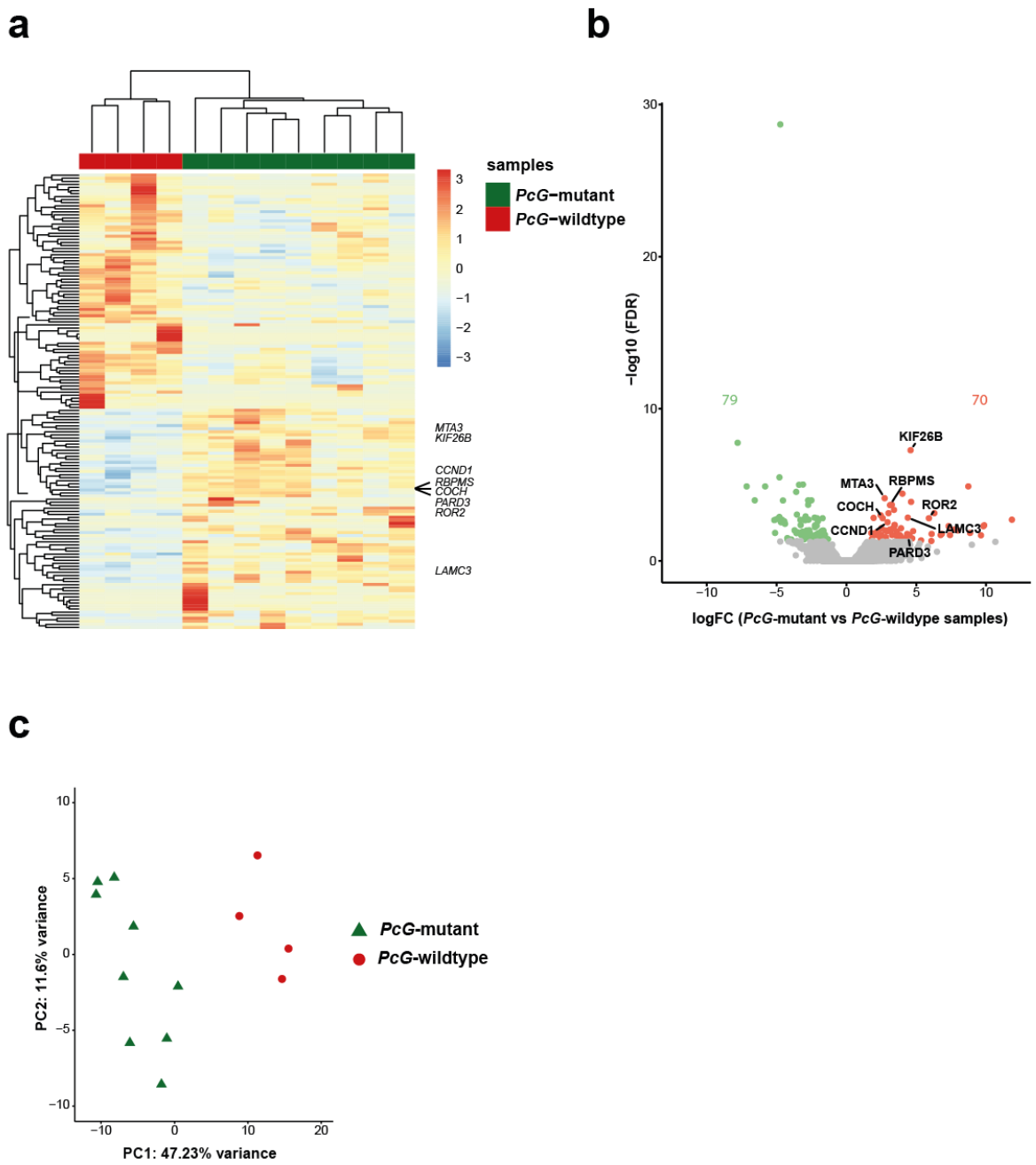


Figure 4.13: Identification of genes differentially expressed in *PcG*-mutant RUNX1-RUNX1T1 AML patient samples compared to *PcG*-wildtype ones. (a) Heatmap showing clustering of *PcG*-mutant (green, n=9) and *PcG*-wildtype (red, n=4) RUNX1-RUNX1T1 patient samples based on top differentially expressed genes (FDR < 0.05). Scale is log2 (normalised counts per million (CPM) +1). Examples of RUNX1-bound PcG-repressed genes upregulated in *PcG*-mutant RUNX1-RUNX1T1 patients compared to *PcG*-wildtype ones are annotated. RNA-seq data from Lavallée et al, *Blood*, 2016. (b) Volcano plot showing differentially expressed genes between *PcG*-mutant and *PcG*-wildtype RUNX1-RUNX1T1 patient samples. Genes upregulated in *PcG*-mutant compared to *PcG*-wildtype RUNX1-RUNX1T1 patients shown in red (n=70, logFC > 0, FDR <0.05). Genes downregulated in *PcG*-mutant compared to *PcG*-wildtype RUNX1-RUNX1T1 patients shown in light green (n=79, logFC < 0, FDR <0.05). Genes not significantly differentially expressed shown in grey. Examples of genes upregulated in *PcG*-mutant RUNX1-RUNX1T1 patients compared to *PcG*-wildtype ones are annotated. (c) PCA plot of log2(normalised counts per million (CPM) +1) of

differentially expressed genes showing clustering of *PcG*-mutant (n=9, green) and *PcG*-wildtype (n=4, red) RUNX1-RUNX1T1 patients.

To investigate the hypothesis that RUNX1 binding at those PcG-repressed loci further contributes to binding of PcG proteins, I checked how the reduction in RUNX1 binding in SKNO-1 cells affected RING1B binding at the same sites. To do that, I performed *RUNX1* siRNA KD, and cells were harvested 48h after electroporation. Consistent with my hypothesis, *RUNX1* siRNA KD resulted in reduction in RING1B binding at the loci tested (Figure 4.14). For example, the promoters of *CCDN1* and *RBPMS* genes, that are upregulated in *PcG*-mutant RUNX1-RUNX1T1 patients compared to *PcG*-wildtype ones, are bound by RUNX1 and CBFb, but not the fusion protein RUNX1-RUNX1T1, and their TSSs are enriched in the repressive H3K27me3 histone mark. At these sites, *RUNX1* KD followed by ChIP for RING1B showed reduction in RING1B binding (Figure 4.14 a). Similarly, I found that at other sites that are RUNX1-bound and PcG-repressed (*KIF26B* and *LAMC3), RUNX1* siRNA KD resulted in reduction of RING1B binding (Figure 4.14 b).

To determine whether the reduction in RING1B binding after *RUNX1* siRNA KD also caused upregulation of RUNX1-bound PcG-repressed genes, I performed nascent RNA-seq after *RUNX1* siRNA KD for 48h. This demonstrated downregulation of genes associated with replication and mitosis (*MCM3, MCM4, MCM7, MIS12, CDC6)* and downregulation of heat shock protein genes (*HSP90AA1*, *HSP90B1*, *HSPA13*, *HSPA5, HSPA8, HSPH1, HSPD1),* in accordance with previous observations (Ben-Ami et al., 2013) (Figure 4.15 a,b). Interestingly, genes encoding histone variants were also downregulated (Figure 4.15 b), suggesting reduction in replication-dependent histone biosynthesis (Singh et al., 2018). However, in contrast to the MLL-AF4 ALL results, there was no upregulation of RUNX1-bound PcG-repressed genes after *RUNX1* siRNA KD, despite the reduction in RING1B binding (Figure 4.15 c). More specifically, the common RUNX1-H3K27me3 peaks in SKNO-1 cells (989 peaks) corresponded to 933 RUNX1-bound PcG-repressed genes, of which 305

genes are expressed in SKNO-1 cells (in non targeting control cells or after *RUNX1* siRNA KD).

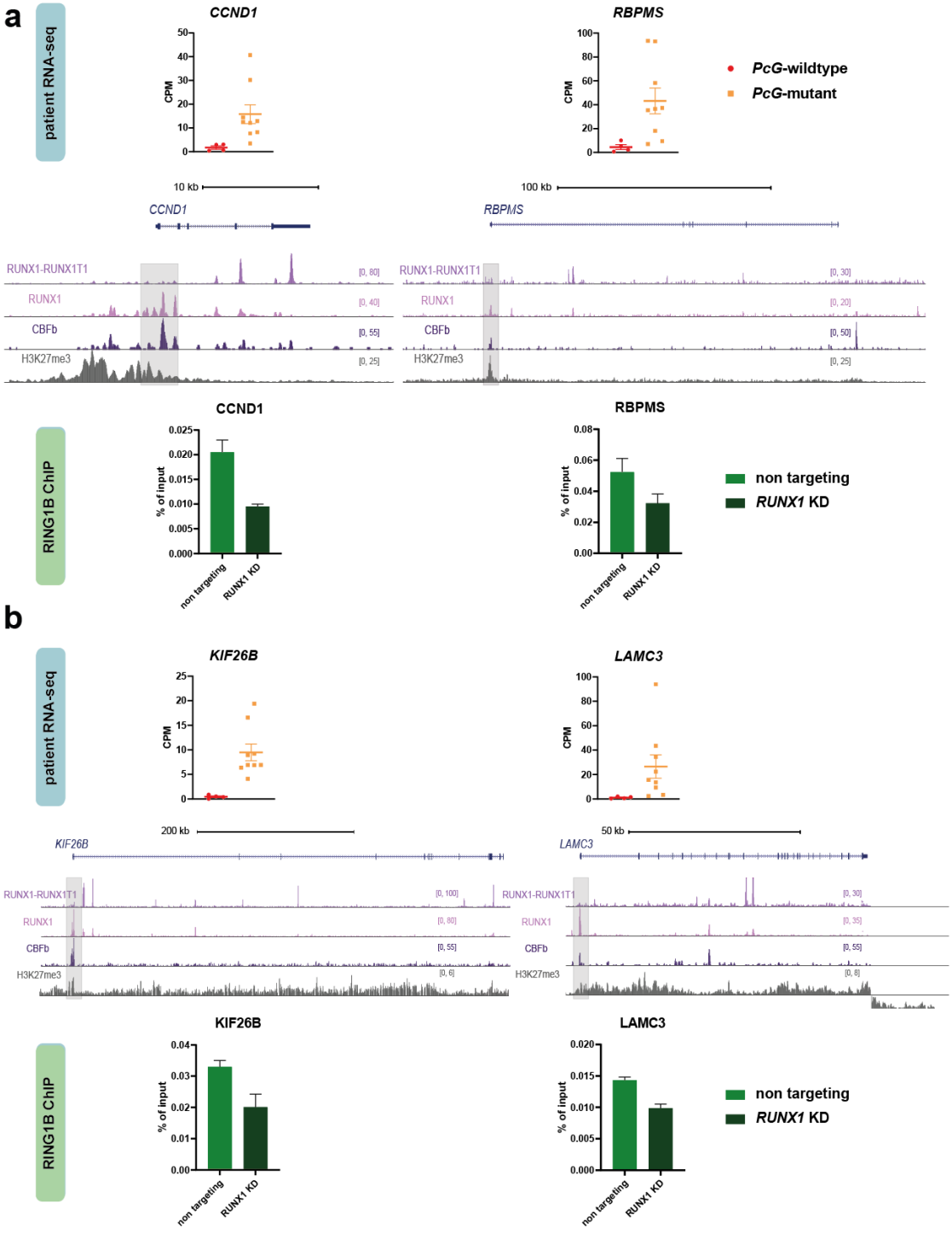
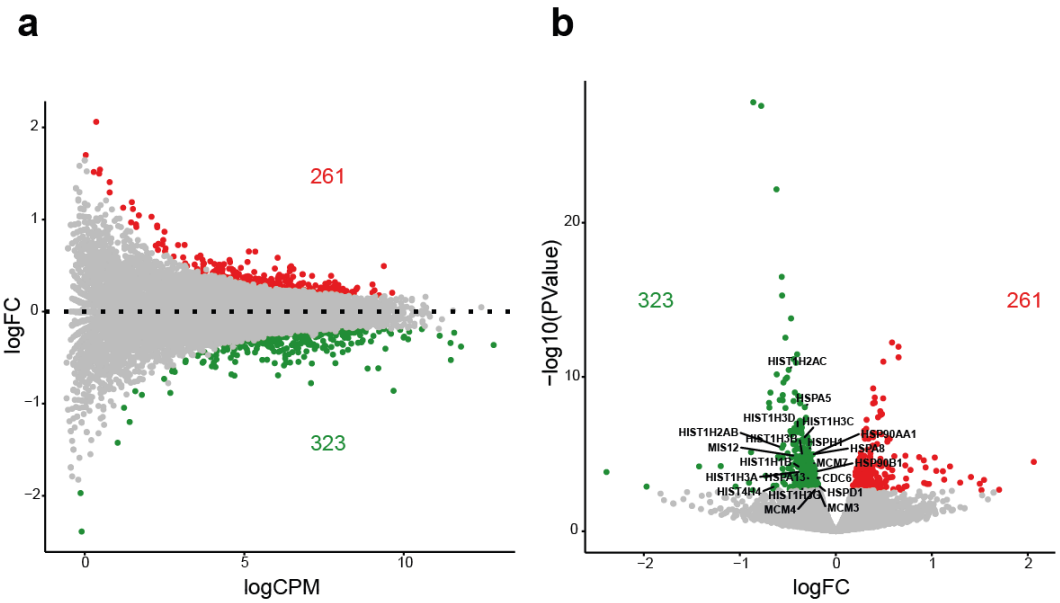


Figure 4.14: RUNX1 contributes to the stabilisation of RING1B binding at a subset of PcG-repressed genes upregulated in *PcG*-mutant RUNX1-RUNX1T1 AML patients. (a) *Top*: Dotplots showing expression of *CCND1 (left)* and *RBPMS (right)* in *PcG*-wildtype (red, n=4) and *PcG*-mutant (yellow, n=9) RUNX1-RUNX1T1 patient samples. *Middle*: UCSC ChIP-seq tracks for RUNX1-RUNX1T1, RUNX1, CBFb and H3K27me3 at *CCND1* and *RBPMS* loci in SKNO-1 cells (ChIP-seq from (Martens et al., 2012) and (Micol et al., 2017)). *Bottom*: ChIP-qPCR for RING1B in SKNO-1 cells at the promoters of *CCND1* and *RBPMS* after *RUNX1* siRNA knockdown or non targeting control. (b)

Like in (a) for *KIF26B* and *LAMC3* genes. ChIP-qPCR experiments were performed in 2 biological replicates.

However, 48h after transfection with *RUNX1* siRNA there was no upregulation of PcG-repressed genes that are expressed at low levels, or emergence of transcription from PcG-repressed genes that are not expressed in SKNO-1 cells (Figure 4.15 c). This suggests that, while RUNX1 binding contributes to the stabilisation of PcG proteins, the residual binding of PcG proteins after *RUNX1* siRNA KD is sufficient for maintaining gene repression.



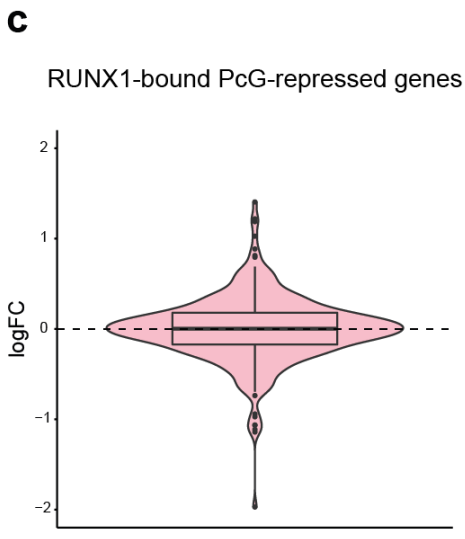


Figure 4.15: *RUNX1* siRNA KD in SKNO-1 cells results in downregulation of replication-associated genes, but does not lead to upregulation of PcG-repressed genes. (a)-(b) MA plot (a) and volcano plot (b) showing changes in nascent RNA levels in SKNO-1 cells after *RUNX1* siRNA KD compared to non targeting control. Upregulated genes (logFC> 0, FDR< 0.05) are shown in red, downregulated genes (logFC< 0, FDR< 0.05) are shown in green, and unchanged genes in grey. Statistically significant changes were determined from 4 biological replicates per condition. In the volcano plot, downregulated genes after *RUNX1* siRNA KD are annotated. (c) Violin plot showing logFC of nascent RNA levels in SKNO-1 cells after *RUNX1* siRNA KD compared to non-targeting control, for PcG-repressed genes (enriched in H3K27me3) that are bound by RUNX1.

4.3.6 Effects of CCND1 upregulation on the cell cycle in RUNX1-RUNX1T1 leukaemia cells

One of the PcG-repressed genes, where RUNX1 binding further contributes to PcG recruitment, is the *CCND1* gene, as mentioned in the previous section (Figure 4.13 a,b, Figure 4.14 a). *CCND1* is a cell cycle gene that controls G1-S phase transition by forming a complex with cyclin-dependent kinase 4 and 6 (CDK4, CDK6). *CCND1* is frequently mutated and overexpressed in many forms of cancer and can be therapeutically targeted (Montalto and De Amicis, 2020). Mutations in *CCND1* and (mostly) *CCND2* genes have been previously identified in a significant proportion of adult RUNX1-RUNX1T1 patients (Christen et al., 2019; Eisfeld et al., 2017; Faber et al., 2016), but, in contrast to other forms of cancer, are rare events in non-RUNX1-RUNX1T1 AML patients (Eisfeld et al., 2017). *CCND1/CCND2* mutations in RUNX1-RUNX1T1 leukaemias cluster around the conserved residue Thr280 (Eisfeld et al., 2017; Faber et al., 2016), similarly to *CCND2* mutations in developmental syndromes (Mirzaa et al., 2014). These mutations seem to promote protein stabilisation by rendering the CCND1/2 proteins resistant to proteosomal degradation (Faber et al., 2016; Kida et al., 2007; Mirzaa et al., 2014). I therefore hypothesised that loss of Polycomb repression in *PcG*-mutant RUNX1-RUNX1T1 patients leads to *CCND1* upregulation and could be an alternative mechanism to *CCND1/2* mutations for increased expression levels of these genes.

To "mimic" the effect of *PcG*-mutations in RUNX1-RUNX1T1 patients, I performed *EZH2* siRNA knockdown in SKNO-1 cells which harbour the *t(8;21)* translocation. *EZH2* siRNA KD resulted in significant upregulation of *CCND1* (Figure 4.16 a). Similarly, treatment of cells with PRC2 inhibitors

also resulted in *CCND1* upregulation. More specifically, treatment with 5 μM of the EZH2/1 inhibitor UNC1999 (Konze et al., 2013) and 2 μM of the EZH2-specific inhibitor GSK126 (McCabe et al., 2012b) for 120h resulted in a significant upregulation of *CCND1,* following reduction of H3K27me3 (Figure 4.16 b). In order to test whether the upregulation of *CCND1* that is observed in *PcG*-mutant RUNX1-RUNX1T1 patients would affect the cycling of cells, i.e. lead to increased proportion of cells in the S phase, I performed a cell cycle analysis in cells treated with PRC2 inhibitors. A higher percentage of cells treated with PRC2 inhibitors were found in S phase compared to untreated cells, and a smaller percentage of cells were in G1 phase (Figure 4.16 c), suggesting increased cell proliferation. The more potent EZH2/1 inhibitor UNC1999 caused a more significant derepression of *CCND1* compared to the EZH2-specific inhibitor GSK126, and also resulted in a more pronounced increase in the proportion of cells found in the S phase of the cell cycle (Figure 4.16 b,c). Together, these results show that EZH2 inhibition can increase CCND1 expression and are also associated with an increase in the S phase, suggesting that the cells are undergoing increased proliferation. In this way, upregulation of PcG-repressed gene *CCND1* may provide a selective advantage to leukaemia clones bearing *PcG* mutations (similarly to leukaemic clones bearing *CCND1/CCND2* mutations in other RUNX1-RUNX1T1 patients).

As mentioned above, *CCND1 is* frequently mutated or overexpressed in many forms of cancer and is a therapeutic target. Palbociclib is a selective inhibitor of the cyclin-dependent kinases CDK4 and CDK6 and is prescribed together with other drugs for metastatic breast cancer with *CCND1* upregulation, as CCND1 forms a complex with CDK4 and promotes the G1-S transition and cell proliferation (Taylor-Stokes et al., 2019). Based on our previous results, I hypothesised that RUNX1-RUNX1T1 patients with additional *PcG* mutations that lead to upregulation of *CCND1* (as well as patients with *CCND1/2* mutations) would benefit more from treatments targeting CDK4/6 compared to the other subsets of RUNX1-RUNX1T1 patients. It has been recently proposed that RUNX1-RUNX1T1 patients would benefit from combination treatments including palbociclib, due to the fact that the fusion protein RUNX1-RUNX1T1 binds to and activates *CCND2* (Martinez-Soria

et al., 2018). However, given that palbociclb functions by inhibiting cell cycle progression, i.e. it is cytostatic rather than cytotoxic, it does not eradicate the leukaemia clones, and is used as part of combination treatments. By treating SKNO-1 cells with either the EZH2/1 inhibitor UNC1999 or with DMSO control (for 72h), followed by 72h treatment with the EZH1/2 inhibitor plus different

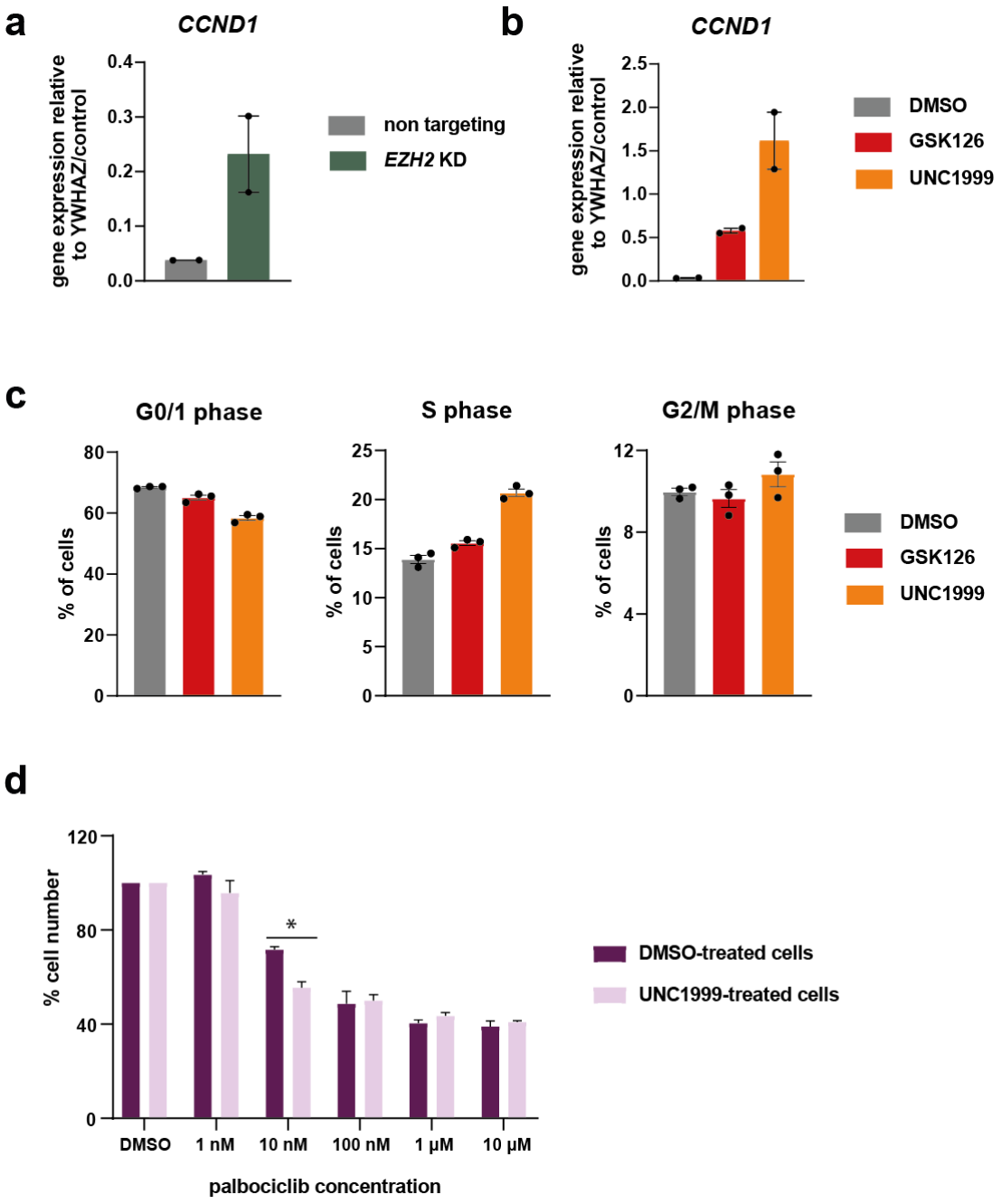


Figure 4.16: *CCND1* upregulation results in changes in cell cycling. qRT-PCR analysis of *CCND1* expression in SKNO-1 cells after (a) *EZH2* siRNA knockdown and (b) after treatment with 2 µM EZH2 inhibitor GSK126 or 5 µM EZH2/1 inhibitor UNC1999 or for 5 days (right), normalized to YWHAZ and relative to cells treated with non targeting control and DMSO-treated cells respectively. The mean of 2 biological replicates per condition is shown. (c) Cell cycle analysis in SKNO-1 cells treated

with 2 µM EZH2 inhibitor GSK126 for 5 days (red), 5 µM EZH2/1 inhibitor UNC1999 for 5 days (orange) or DMSO control (grey). Barplots show from left to right the percentage of cells in G0/1, S and G2/M phases. Error bars represent the SEM of 3 biological replicates. (d) Cell proliferation of DMSO- and UNC1999-treated SKNO-1 cells after treatment with different concentrations of palbociclib. Cells were treated with 5 µM UNC1999 (or DMSO control) for 72h, followed by 72h treatment with UNC1999 (or DMSO) plus palbociclib in concentration ranging from 1 nM to 10 µM. % of cells for each palbociclib concentration was calculated in comparison to DMSO-treated cells. In UNC1999-treated cells treated with 10 nM of palbociclb we observe a more potent cytostatic effect compared to DMSO-treated cells that have been treated with the same concentration of the inhibitor (p-value = 0.0286, two-tailed Mann-Whitney test). Error bars represent the standard error of the mean (SEM) of 4 biological replicates.

concentrations of palbociclib (ranging from 1 nM to 1 µM) I found that in both UNC1999- and DMSO-treated cells, there is a reduction in cell proliferation with increased concentrations of the inhibitor. In addition, in the case of UNC1999-treated cells, there is a more significant cytostatic effect with smaller concentrations of the inhibitor (10 nM) compared to DMSO-treated cells that do not have *CCND1* upregulation (Figure 4.16 d).

Finally, although differential expression analysis did not identify the PcG-repressed gene *EPCAM* as upregulated in *PcG*-mutant RUNX1-RUNX1T1 patients, I found that it is overexpressed in *ASXL1*-mutant and a subset of *ASXL2*-mutant RUNX1-RUNX1T1 patients (Figure 4.17). *EPCAM* encodes a cell surface protein that is involved in cell adhesion and its overexpression has been associated with chemoresistance in AML (Zheng et al., 2017). As EPCAM is a promising target for adoptive immune effector cell therapies (Zhang et al., 2019; Zhou et al., 2019), its targeting could be beneficial for this subset of *PcG*-mutant RUNX1-RUNX1T1 patients that show *EPCAM* overexpression.

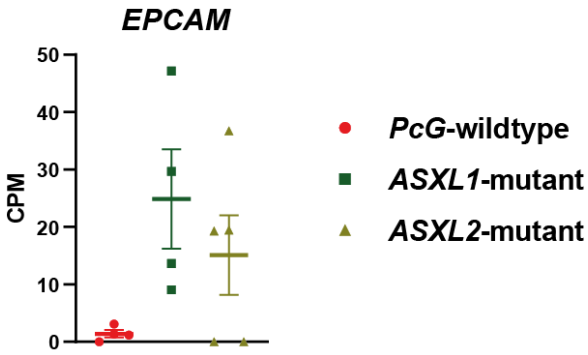


Figure 4.17: *EPCAM* is upregulated in a subset of *PcG*-mutant RUNX1-RUNX1T1 patients. Dotplot showing expression of *EPCAM* in *PcG*-wildtype (red, n=4), *ASXL1*-mutant (dark green, n=4) and

ASXL2-mutant (light green, n=5) RUNX1-RUNX1T1 patient samples, demonstrating that it is upregulated in *ASXL1*-mutant and a subset of *ASXL2*-mutant RUNX1-RUNX1T1 patients.

4.3.7 Differential gene expression analysis reveals genes upregulated in *PcG*-mutant compared to *PcG*-wildtype RUNX1-mutant patients.

As discussed before, similarly to RUNX1-RUNX1T1 patients, in *RUNX1*-mutant AML patients an increased frequency in *PcG* mutations has been observed. *RUNX1*-mutant patients have a distinct transcriptional signature compared to *RUNX1*-wildtype ones (Simon et al., 2017), and more specifically it has been shown that *RUNX1*-mutant patients have upregulation of genes expressed in primitive haematopoietic cells and B-cell progenitors (*DNTT, BAALC, BLNK, CD109, RBPMS, FLT3*) and downregulation of genes that promote myelopoiesis such as *CEBPA* (Greif et al., 2012; Mendler et al., 2012; Simon et al., 2017).

To determine whether, similarly to RUNX1-RUNX1T1 patients, *RUNX1*-mutant AML patients with additional *PcG* mutations have a distinct transcriptional signature compared to *RUNX1*-mutant AML patients without *PcG* mutations, I analysed publicly available RNA-seq and mutational data from Simon et al., 2017 (Appendix table 7.3). I compared 17 *PcG*-wildtype to 26 *PcG*-mutant *RUNX1*-mutant AML samples (with *ASXL1, BCOR, EZH2* mutations and one sample with *HIST1H3H* K27M in addition to *BCOR* mutation (Lehnertz et al., 2017; Simon et al., 2017)). This revealed ~100 genes differentially expressed between the two subsets of patients (Figure 4.18 a). PCA analysis based on these differentially expressed genes showed that this ~100 gene signature separates the *PcG*-mutant and *PcG*-wildtype *RUNX1*-mutant patient samples into two clusters (Figure 4.18 b). Despite the significant heterogeneity in the mutational status of the patients (Appendix table 7.3), I found upregulation of PcG-repressed genes in a subset of *PcG*-mutant patients, including genes *CCND1* and *KIF26B* (like in *PcG*-mutant RUNX1-RUNX1T1 patients) and *SCN1B*. Furthermore, in a subset of *PcG*-mutant *RUNX1*-mutant AML patients, the *BLNK* gene is upregulated, which, as mentioned

above is a lymphoid regulator and is also upregulated in *RUNX1*-mutant compared to *RUNX1*-wildtype AML cases (Greif et al., 2012; Mendler et al., 2012) (Figure 4.18 c). In addition, in *PcG*-mutant patients there was upregulation of the oncofoetal gene *PLAG1,* which has been previously shown to be upregulated in *in vivo* AML mouse models with *Ezh2* loss (Figure 4.18 c) (Basheer et al., 2019). In these models, it has been shown that loss of Ezh2 repression results in increased expression of *Plag1,* which activates the IGF2 signalling pathway (Van Dyck et al., 2007) and has been previously described as an oncogene in CBFb-SMMHC leukaemias (Landrette et al., 2005). It is possible that upregulation of the cell cycle gene *CCND1* together with the oncofoetal gene *PLAG1* would result in increased proliferation and confer a selective advantage of *PcG*-mutant clones in *RUNX1*-mutant AML patients.

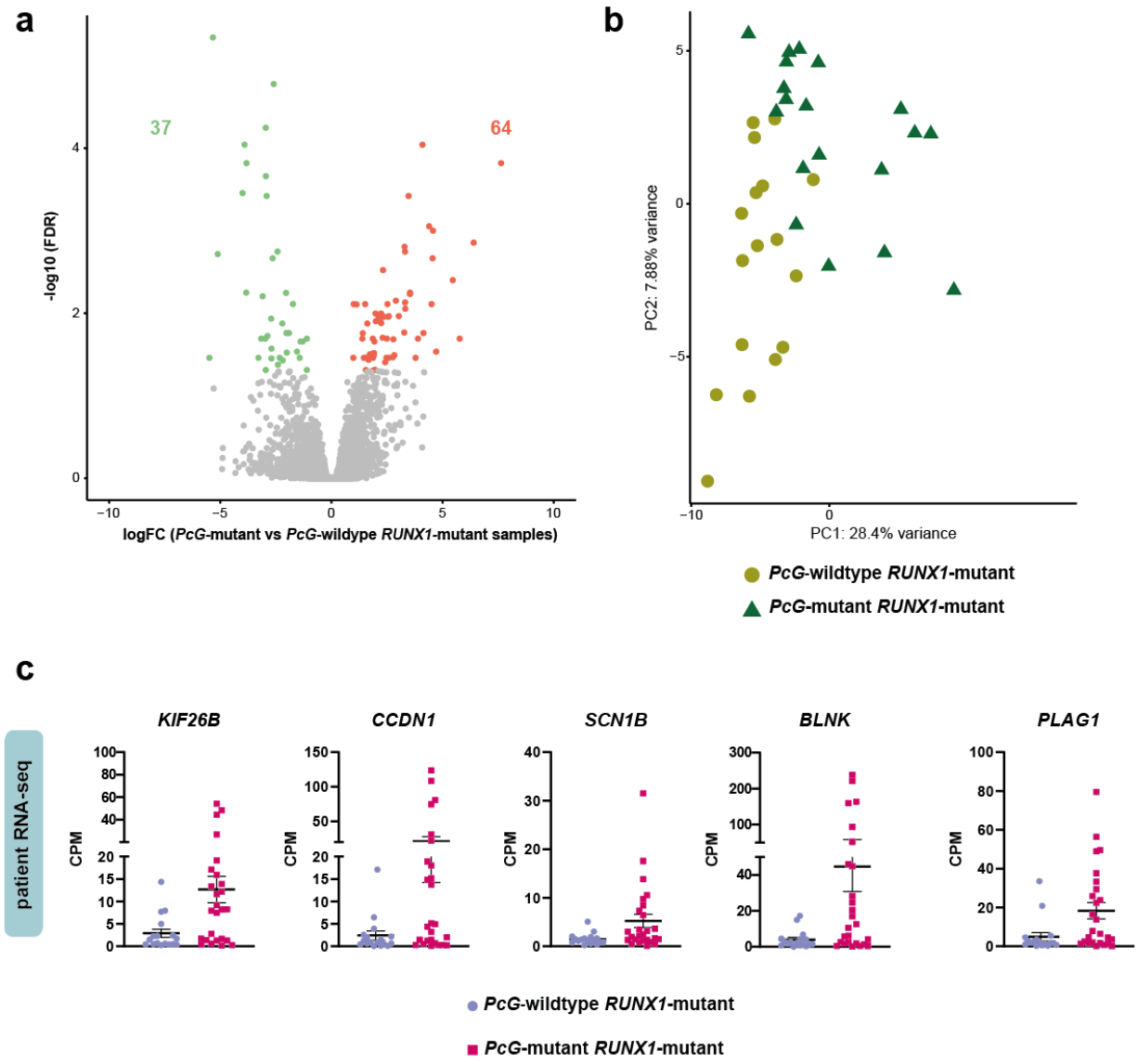


Figure 4.18: Identification of genes differentially expressed in *PcG*-mutant *RUNX1*-mutant AML patient samples compared to *PcG*-wildtype *RUNX1*-mutant samples. (a) Volcano plot showing differentially expressed genes between *PcG*-mutant (n=26) and *PcG*-wildtype (n=17) *RUNX1*-mutant patient samples. Genes upregulated in *PcG*-mutant compared to *PcG*-wildtype *RUNX1*-mutant patients shown in red (n=64, logFC > 0.5, FDR <0.05). Genes downregulated in *PcG*-mutant compared to *PcG*-wildtype *RUNX1*-mutant patients shown in light green (n=37, logFC < -0.5, FDR <0.05). Genes not significantly differentially expressed shown in grey. Patient RNA-seq from (Simon et al., 2017). (b) PCA plot showing clustering of *PcG*-mutant (n=26, green) and *PcG*-wildtype (n=17, lime) *RUNX1*-mutant AML patients based on log2(normalised counts per million (CPM) +1) of differentially expressed genes. (c) Dotplots showing expression of *KIF26B*, *CCND1*, *SCN1B, BLNK* and *PLAG1* genes in *PcG*-wildtype (light blue, n=17) and *PcG*-mutant (magenta, n=26) *RUNX1*-mutant AML patient samples.

4.4 Discussion

The aim of the work in this chapter was to determine whether the RUNX1-mediated recruitment of PRC1 observed in the TetR system in Chapter 3 also takes place in leukaemic cells, and whether RUNX1 affects gene regulation through the recruitment of PcG proteins. To do this I used a number of leukaemic cell lines and also analysed publicly available RNA sequencing data from primary human leukaemia samples.

4.4.1 RUNX1 binding stabilises PRC1 binding in MLL-AF4 cells.

First, I confirmed that in the MLL-AF4 ALL cell line SEM, in which *RUNX1* is not mutated, there is extensive overlap between RUNX1 and RING1B binding. Furthermore, reference-normalised ChIP-seq for RING1B after *RUNX1* siRNA KD, showed significant reduction in RING1B binding after reduction in RUNX1 binding at the same sites. This reduction in RING1B binding occurred not only at RUNX1-bound gene promoters, but also at distal RUNX1-bound RING1B sites, including enhancer elements. The additional observations that the variant PRC1 component KDM2B and the canonical PRC1 components CBX4 and PHC2 are bound at the promoters of active genes and at enhancer elements, in addition to the TetR data from the previous chapter that RUNX1 can mediate recruitment of variant PRC1 complexes and of canonical PRC1 complexes independent of H3K27me3, provide a plausible mechanism for the localisation of cPRC1 proteins at active promoters and enhancers.

Given these observations and the findings of others that in differentiated cells PcG proteins are bound not only at repressed sites but also at active loci (Cohen et al., 2018; Cohen et al., 2019; Frangini et al., 2013; Gao et al., 2014; Zhang et al., 2021), in addition to the emerging evidence that PcG proteins can mediate interactions between active sites (including interactions between promoters and enhancers) (Loubiere et al., 2020; Zhang et al., 2021), it is possible that RUNX1 could promote the formation of enhancer-promoter interactions between active sites via the recruitment

of PcG proteins. More specifically, I hypothesise that the recruitment of cPRC1 proteins at active sites by RUNX1 (and potentially other TFs) could lead to chromatin interactions, mediated by the polymerisation of the SAM domains of PHC proteins (Isono et al., 2013; Kim et al., 2002; Kundu et al., 2017).

Based on these observations, future experiments could aim to dissect whether RUNX1-mediated recruitment of PRC1 contributes to chromatin interactions in MLL-AF4 leukaemic cells. More specifically, it would be of interest to perform Capture-C at a subset of downregulated genes after *RUNX1* siRNA KD, in order to investigate whether loss of RUNX1 binding leads to reduction in enhancer-promoter interactions. It would also be useful to perform Capture-C after reduction in RING1A/RING1B binding, after *siRNA* KD of both of these proteins (or after rapid depletion of these proteins with an inducible degron protein degradation system), to determine whether the reduction in binding of PcG proteins affects chromatin interactions and whether it has the same effect on enhancer-promoter interactions as reduction in RUNX1 binding.

4.4.2 RUNX1 binding stabilises PRC1 binding in RUNX1-RUNX1T1 AML

In considering the role of RUNX1 in leukaemias of other lineages, particularly AML and ETP ALL, I noticed that several clinical studies and case reports have documented the co-occurrence of *RUNX1* and *PcG* mutations (Booth et al., 2018; Gaidzik et al., 2016; Lehnertz et al., 2017; Ochi et al., 2020; Sashida et al., 2014; Simon et al., 2017; Wu et al., 2019). This is true for patients with *RUNX1* point mutations, nonsense and frameshift mutations and also in patients with the *t(8;21)* translocation that gives rise to the RUNX1-RUNX1T1 fusion protein. The frequent co-occurrence of *RUNX1* and *PcG* mutations (not only in AML but also in MDS and ETP-ALL) suggests that the neoplastic clones carrying this combination of mutations have a selective advantage. In addition, *in vivo* studies have validated that *PcG* mutations accelerate the progression of leukaemia in mouse models of *RUNX1*-mutant and RUNX1-RUNX1T1 AML (Asada et al., 2018; Lehnertz et al., 2017; Micol et al., 2017;

Sashida et al., 2014). However, the mechanistic details of the functional co-operation of *RUNX1* and *PcG* mutations have not been investigated.

The relative enrichment of *PcG* mutations in *RUNX1*-mutant leukaemias, especially the ones with truncating *RUNX1* mutations, and in RUNX1-RUNX1T1 AML cases, in combination with my data that RUNX1 is sufficient to recruit RING1B, prompted me to investigate the hypothesis that the C-terminal part of RUNX1 is mediating the RUNX1-RING1B interaction. Interestingly, despite previous reports showing that the DNA-binding domain of RUNX1 mediates the RUNX1-BMI1 interaction (Yu et al., 2012), I was able, using the TetR system, to show that the C-terminal part of RUNX1, which is not included in the RUNX1-RUNX1T1 fusion protein and is frequently lost in truncating *RUNX1* mutations, interacts (directly or indirectly) with RING1B, and is responsible for the RUNX1-mediated recruitment of RING1B. The loss of the C-terminal part of RUNX1 via nonsense and frameshift mutations could result in decreased stabilisation of RING1B binding at common RING1B-RUNX1 binding sites. Similarly, the RUNX1-RUNX1T1 fusion protein that lacks the C-terminal part of RUNX1 does not retain the ability to interact with PRC1, in contrast to wildtype RUNX1. Additionally, these data suggest that the missense *RUNX1* mutations that abrogate the ability of RUNX1 to bind to its cognate DNA sequence could negatively affect its ability to stabilise the binding of PcG proteins at a subset of target genes.

In the case of RUNX1-RUNX1T1 leukaemias, I used publicly available ChIP-seq data from RUNX1-RUNX1T1 cell lines and identified ~ 1,000 RUNX1 binding sites that are enriched in the repressive H3K27me3 mark, and have shown that at a subset of loci reduction in the binding of RUNX1 results in reduced RING1B binding. In contrast, the co-occupancy between the fusion protein and the H3K27me3 mark is rather limited. Interestingly, nascent RNA-seq results after *RUNX1* siRNA KD in RUNX1-RUNX1T1 cells did not show emergence of transcription from these non-expressed RUNX1-bound PcG-bound genes. This may indicate that, while RUNX1 contributes to PcG recruitment at those sites, the residual binding of PcG proteins after *RUNX1* siRNA KD is sufficient for maintaining

gene repression. One of the caveats of the way in which differential gene expression analysis is performed, is the assumption that the majority of genes do not change expression across conditions, which may lead to masking of subtle changes in gene expression. In this way for example, a subset of upregulated or downregulated genes could be classified as unchanged. Spike-in controls for normalisation of gene expression across samples would be necessary to determine this. In addition, since nascent RNA-seq was performed 48h after introducing the *RUNX1* siRNA in RUNX1-RUNX1T1 cells, it should be noted that RUNX1 protein levels were not completely abolished. Therefore to test whether abolishing the levels of RUNX1 would result in upregulation of RUNX1-bound PcG-repressed genes, a repeat of a *RUNX1* siRNA transfection or RUNX1 depletion with an inducible degron technology could be employed.

Given the frequency of *PcG* mutations in AML patients with *RUNX1* mutations and translocations, I compared the transcriptional status of *PcG*-mutant and *PcG*-wildtype RUNX1-RUNX1T1 patients. In this way, I was able to find a transcriptional signature of *PcG*-mutant RUNX1-RUNX1T1 patients compared to *PcG*-wildtype ones, including PcG-repressed genes that become upregulated in *PcG*-mutant patients that are bound by RUNX1. Among these genes is the cell cycle gene *CCND1.* This is interesting because *CCND1,* and more frequently *CCND2,* are mutated in a subset of RUNX1-RUNX1T1 patients (Christen et al., 2019; Eisfeld et al., 2017; Faber et al., 2016), and these mutations lead to stabilisation of the mutant proteins (Faber et al., 2016; Kida et al., 2007; Mirzaa et al., 2014). This suggested that *CCND1* upregulation due to loss of PcG repression as a result of mutations in *PcG* genes might promote cell proliferation by functioning similarly to *CCND1/CCND2* mutations. In addition, *RUNX1* siRNA KD showed that RUNX1 stabilises PcG binding at a subset of RUNX1-bound PcG-repressed loci that become upregulated in *PcG*-mutant RUNX1-RUNX1T1 patients, including *CCND1*. The upregulation of cell cycle genes that promote cell proliferation like *CCND1* may be an important part of the mechanism underlying the selective advantage of *PcG*-mutant RUNX1-RUNX1T1 leukaemia cells.

In addition to *PcG*-mutant RUNX1-RUNX1T1 leukaemia, my gene expression analyses have shown that *RUNX1*-mutant AML cases with *PcG* mutations also have a distinct transcriptional signature compared to *PcG*-wildtype *RUNX1*-mutant AML patients. Despite heterogeneity between patients with mutations in different *PcG* genes (*EZH2, BCOR, ASXL1)*, I was able to identify PcG-repressed genes in AML that become upregulated in *PcG*-mutant *RUNX1*-mutant AML patients. Among them are the RUNX1-bound PcG-repressed genes *CCND1* and *KIF26B.* The upregulation of *CCND1,* a cell cycle gene, in a subset of *RUNX1*-mutant patients with additional *PcG* mutations may result in increased cell cycling and proliferation, expansion of the leukaemic clones and accumulation of additional mutations that confer a selective advantage – as noted for *PcG*-mutant RUNX1-RUNX1T1 AML. In addition, *EPCAM* is upregulated in *ASXL1*-mutant and a subset of *ASXL2*-mutant RUNX1-RUNX1T1 patients. Although *ASXL1* and *ASXL2* mutations are mutually exclusive in RUNX1-RUNX1T1 patients (Micol et al., 2014), suggesting that these mutations result in common deregulated pathways, it has been shown in haematopoietic cells that ASXL1 deficiency causes a more significant reduction in H3K27me3 compared to ASXL2 deficiency (Micol et al., 2017). This differential effect on H3K27me3 levels could potentially explain the difference in *EPCAM* expression in ASXL1-mutant compared to *ASXL2*-mutant RUNX1-RUNX1T1 patients, as only a subset of *ASXL2*-muant patients show *EPCAM* upregulation. Similarly to the case of CCND1, *PcG*-mutant RUNX1-RUNX1T1 patients that show EPCAM overexpression could benefit from treatments targeting this protein.

Due to the significant heterogeneity of the leukaemic clones, in order to better understand the transcriptional outcome of *PcG* mutations, future experiments using single cell techniques, such as Target-seq, would allow the transcriptional and mutational status of *PcG*-mutant AML cells to be identified in the same cell (Rodriguez-Meira et al., 2019). In this way, the transcriptional status of leukaemic subclones with *PcG* mutations could be compared to the transcriptional status of subclones of the same patient that do not carry *PcG* mutations. Furthermore, by performing these experiments in *PcG*-mutant patient samples at diagnosis, post treatment and at relapse it should

be possible to understand how the treatment affects the mutational and transcriptional status of leukaemic subclones with and without *PcG* mutations. This could be of clinical significance, as the combination of *RUNX1* mutations with mutations in *ASXL1* are associated with worse overall survival (Gaidzik et al., 2016).

4.4.3 Summary and conclusion

In summary, the results in this chapter, build on the data in Chapter 3 identifying PRC1 as a RUNX1 interaction partner, to show that RUNX1 stabilises PcG binding in ALL and AML leukaemic cells. In addition, I was able to demonstrate that the C-terminal part of RUNX1 is mediating the recruitment of RING1B, which suggests that truncating *RUNX1* mutations and translocations lead to loss of the RUNX1-PcG interaction. Finally, I have identified a distinct transcriptional signature for *PcG*-mutant RUNX1-RUNX1T1 and *PcG*-mutant *RUNX1*-mutant AML patients compared to *PcG*-wildtype ones.

Chapter 5 - Effects of PRC2 inhibition on the chromatin landscape and RUNX1 binding in MLL-AF4 leukaemia

5.1 Introduction

TFs affect the chromatin environment through the recruitment of gene-regulatory complexes, and in turn, the chromatin environment shaped from the activity of these proteins further affects the binding of TFs. In Chapter 3, it was demonstrated that RUNX1 affects the chromatin environment through the recruitment of both co-activators and co-repressors, and it was also shown that the RUNX1-mediated recruitment of co-factors is highly context-dependent; at active sites, RUNX1 stabilises the binding of co-activators, while at PcG-repressed loci, RUNX1 further recruits PcG repressors and contributes to the stabilisation of Polycomb domains. It is clear from the literature that there are also RUNX1 independent ways of recruiting PcG complexes to gene targets, especially the binding to unmethylated CpG islands (Farcas et al., 2012; He et al., 2013; Li et al., 2017; Wu et al., 2013b). It is quite likely that the recognition of CpG islands represents the main way of recruiting and establishing Polycomb domains at repressed genes. Given my findings that a key function of RUNX1 is to increase the stabilisation of PcG binding at repressed gene targets, I wanted to investigate how, or if, the initial binding of PcG proteins and their activity impacts RUNX1 binding.

It has been suggested that the reduction of the repressive H3K27me3 mark catalysed by the PRC2 complex offers a “window” of opportunity for the binding of lineage-specific TFs; more specifically, the lag in the deposition of H3K27me3 on nascent DNA after replication in ESCs and haematopoietic progenitors allows the binding of TFs (Petruk et al., 2017a; Petruk et al., 2017b). In addition, it has been suggested that H3K27 methylation outside the Polycomb domains that are bound by PcG proteins, and in particular the H3K27me2 and H3K27me3 marks, antagonise transcription, potentially by antagonising the deposition of H3K27ac and preventing the activation of non-cell-type-specific enhancers (Ferrari et al., 2014; Lee et al., 2015). EZH2 inhibitors have recently been

used as a way of sensitizing AML cells to standard induction therapy, and the proposed mechanism is that a reduction in H3K27me3 globally opens up the chromatin (Porazzi et al., 2022). However, the mechanistic details of how H3K27 methylation affects chromatin accessibility genome-wide and how it specifically impacts TF binding are still incompletely understood.

5.2 Aims

In this chapter, I aimed to investigate the effects of PRC2 activity (and more specifically the effects of changes in H3K27 methylation) on accessibility and RUNX1 binding using PRC2 small molecule inhibitors in MLL-AF4 cells. MLL-AF4 cells have been previously shown to be dependent on PRC2 activity, as PRC2 inhibition results in reduced survival of MLL-AF4 cells in *in vitro* and *in vivo* models (Xu et al., 2015). I specifically aimed to determine how the loss of H3K27me3 after the inhibition of the PRC2 activity in MLL-AF4 cells affects accessibility and RUNX1 binding.

5.3 Results

(In this chapter, I have performed bioinformatics analysis of next generation sequencing experiments after treatment with the PRC2 inhibitor UNC1999 in MLL-AF4 cells, that had been performed by Dr M. Tapia, as indicated where appropriate).

5.3.1 Reduction in H3K27me3 in MLL-AF4 cells results in increased chromatin accessibility and increased RUNX1 binding.

Firstly, to understand the effects of H3K27me3 levels on RUNX1 binding, I have treated the MLL-AF4 cell line SEM with GSK126 (2 μM for 120h). GSK126 functions by inhibiting the activity of EZH2, the catalytic component of the PRC2 complex, and more specifically it competes with the methyl donor, S-adenosyl-methionine (SAM), for binding to EZH2, resulting in a global reduction in H3K27me3 (McCabe et al., 2012b) (Figure 5.1 a). ChIP-qPCR showed reduction in H3K27me3 levels

(Figure 5.1 b), and reference-normalised ChIP-seq for H3K27me3 after treatment with the GSK126 inhibitor showed a significant reduction in the number of H3K27me3 peaks identified (~ 24,500 H3K27me3 peaks identified in the untreated cells compared to ~ 2,000 H3K27me3 peaks in the GSK126-treated cells), and reduction in H3K27me3 levels at H3K27me3-enriched regions (Figure 5.1 c,d and e). To investigate how the reduction in H3K27me3 affected the binding of RUNX1, RUNX1 ChIP-qPCR was performed after using the GSK126 inhibitor, showing an increase in RUNX1 binding (Figure 5.1 f). This suggests that the reduction of this histone mark facilitates the binding of RUNX1, although more replicates need to be performed and more loci need to be tested to confirm these findings.

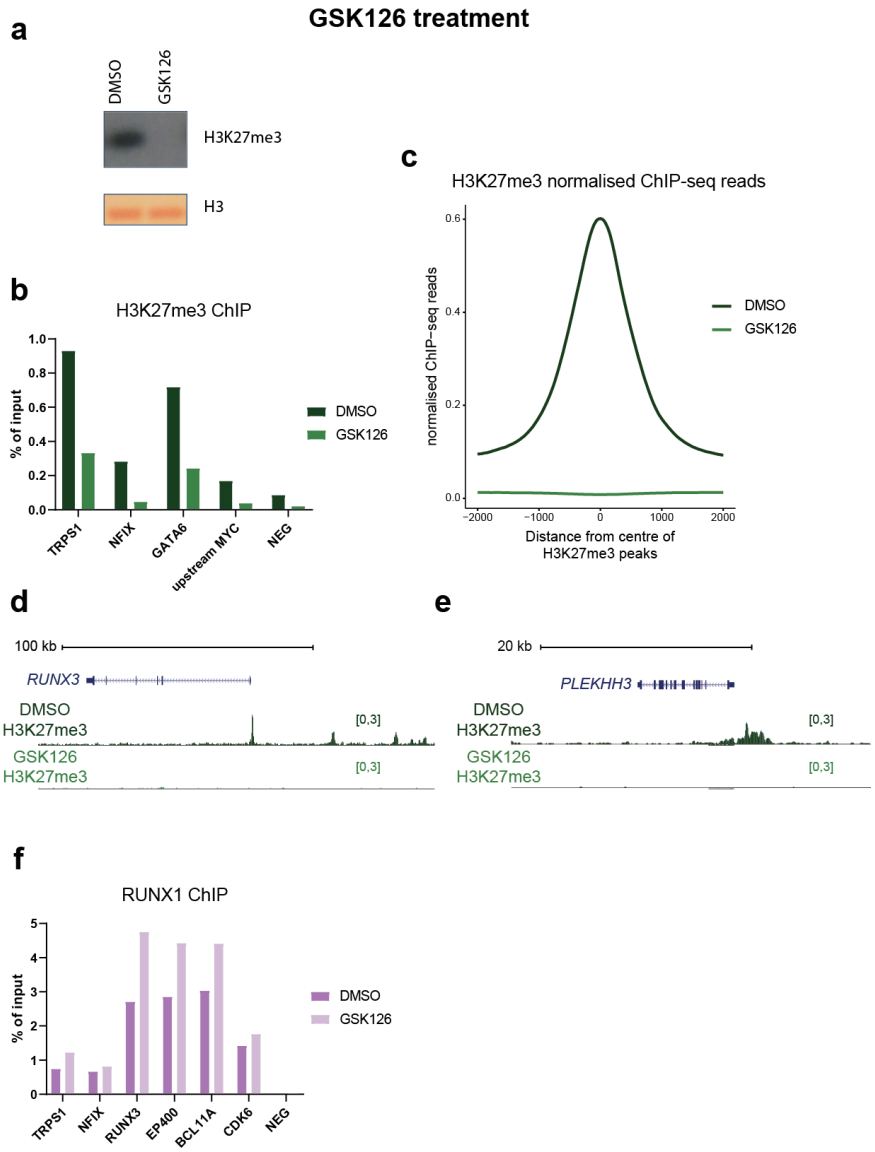


Figure 5.1: GSK126 treatment in SEM cells results in increased RUNX1 binding. (a) Western blot for H3K27me3 (and Ponceau stain showing H3 levels) in SEM cells after GSK126 treatment (2 μM for 120h) and DMSO control. (b) ChIP-qPCR for H3K27me3 in SEM cells after GSK126 treatment and DMSO control, showing reduced H3K27me3 levels following GSK126 treatment (n=1). (c) Metaplot of mean reference-normalised H3K27me3 ChIP-seq reads at H3K27me3 peaks after GSK126 treatment or DMSO control. (d), (e) UCSC ChIP-seq tracks for reference-normalised H3K27me3 ChIP-seq at *RUNX3* (b) and *PLEKHH3* loci (e) after GSK126 treatment and DMSO control, showing reduction in H3K27me3 levels after GSK126 treatment. (f) ChIP-qPCR for RUNX1 after GSK126 treatment and DMSO control in SEM cells, showing increased RUNX1 binding after GSK126 treatment (n=1).

To test whether the increased TF binding after reduction in H3K27 methylation following treatment with the GSK126 inhibitor is accompanied by increased chromatin accessibility, I performed ATAC-seq after GSK126 treatment (2 μM for 120h) or treatment with DMSO. Indeed, treatment with the EZH2-specific inhibitor GSK126 resulted in a genome-wide increase in accessibility. In addition, overlap between ATAC-seq peaks in GSK126-treated and DMSO-treated cells showed the emergence of novel ATAC-seq peaks (Figure 5.2 a and Figure 5.3 a,b). In addition, the majority of ATAC-seq peaks identified in untreated cells are retained after GSK126 treatment, and are marked with increased ATAC-seq reads compared to DMSO-treated cells (Figure 5.2 a,b and Figure 5.3 a,b). The increase in accessibility is evident mainly in H3K27me3-marked TSSs, but an increased accessibility in non-PcG repressed TSSs that lack the H3K27me3 mark is also observed (Figure 5.2 c). This suggests that the genome-wide reduction in H3K27me3, and also the reduction in H3K27me2, which is spread as a "blanket" at intergenic regions (Ferrari et al., 2014), results in increased chromatin accessibility.

Furthermore, the new ATAC-seq peaks that appear after the GSK126 treatment are mainly localised at distal sites, rather than at gene promoters (Figure 5.3 c), suggesting the appearance of novel enhancers. This is consistent with previous reports demonstrating that the genome-wide deposition of H3K27me3, and mainly H3K27me2, antagonises the inappropriate emergence of enhancers, and that the reduction of these histone marks leads to the activation of non-cell-type-specific enhancer elements (Ferrari et al., 2014). Overall, these results suggest that the reduction

in H3K27 methylation results in increased accessibility, and potentially leads to the emergence of novel enhancer elements genome-wide. Whether the novel ATAC-seq peaks in GSK126-treated cells are also enriched in the H3K27ac and H3K4me1 marks compared to untreated cells, and whether their appearance correlates with transcriptional changes has not been investigated further in this thesis, but could be addressed in future work.

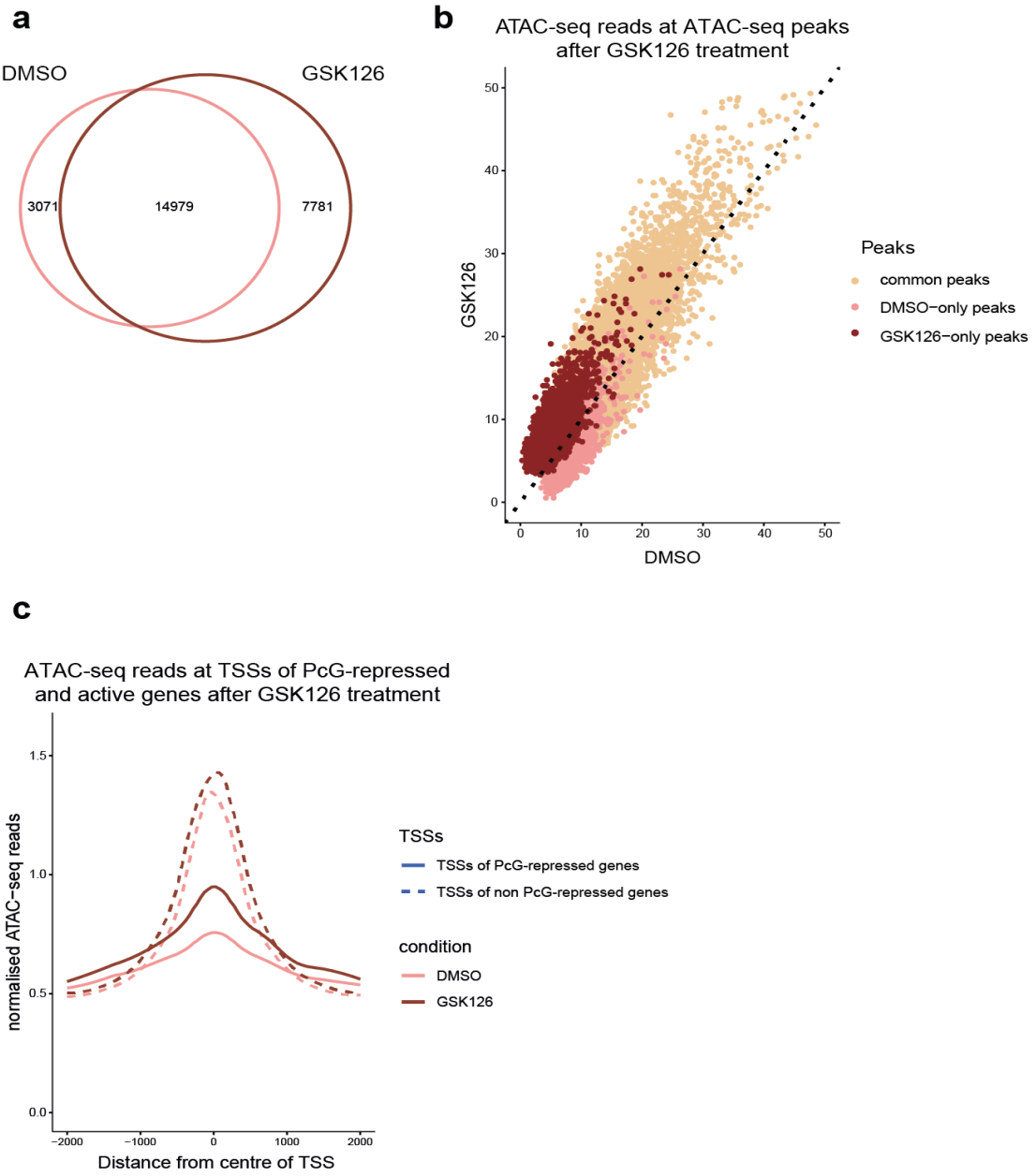


Figure 5.2: GSK126 treatment in MLL-AF4 cells results in increased chromatin accessibility. (a) Venn diagram showing the overlap of ATAC-seq peaks in SEM cells after GSK126 treatment and DMSO control. ATAC-seq was performed in one replicate per condition. (b) Correlation of normalised ATAC-seq reads after GSK126 treatment and DMSO control in SEM cells, after

annotating the ATAC-seq peaks as common in GSK126- and DMSO-treated cells, and DMSO-only and GSK126-only peaks, showing increased ATAC-seq reads at the common and GSK126-only peaks after GSK126 treatment. (c) Metaplot of ATAC-seq reads at the TSSs of PcG-repressed and non PcG-repressed genes (classified based on the presence of H3K27me3 peaks) after GSK126 treatment and DMSO control.

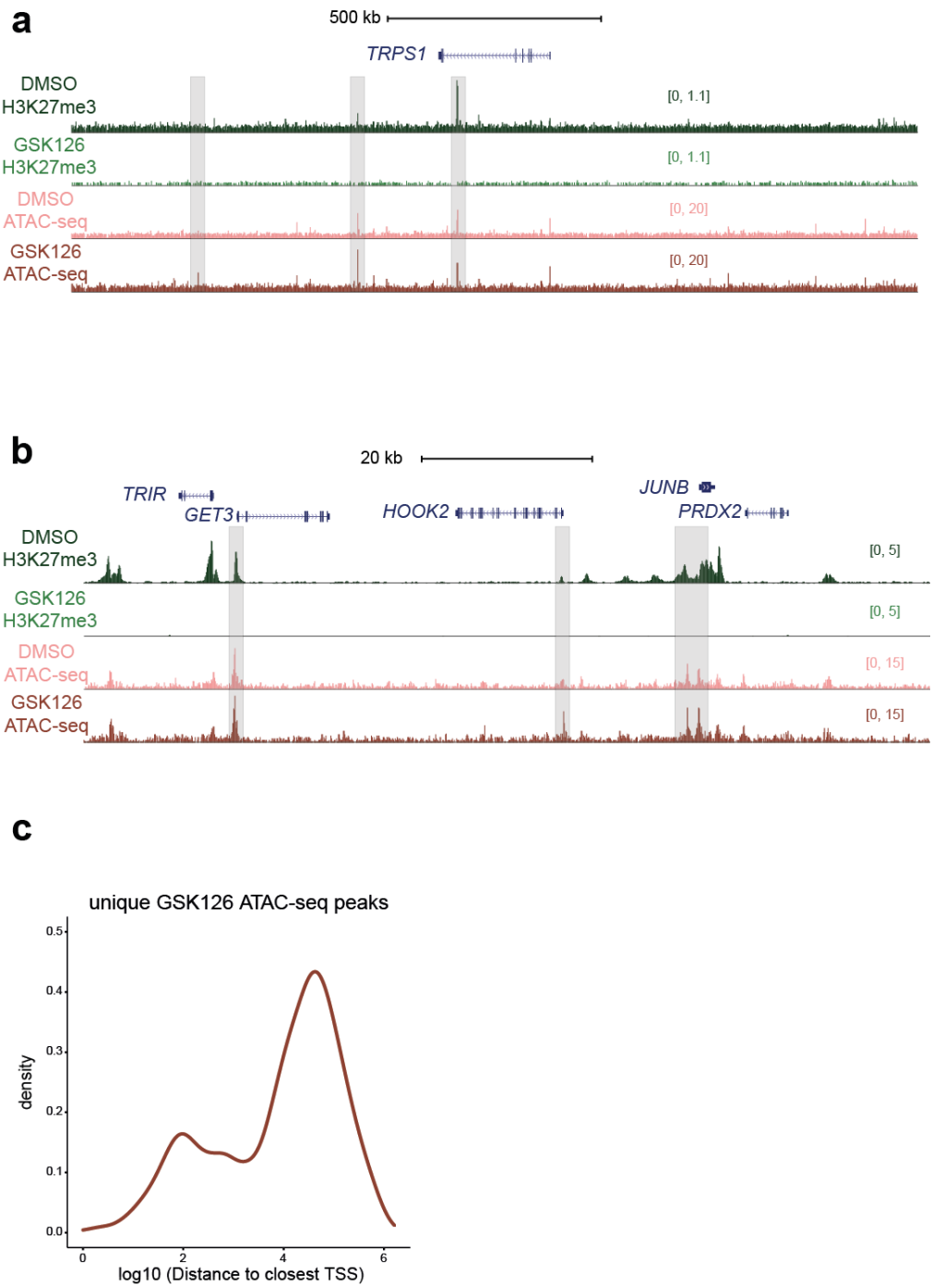


Figure 5.3: GSK126 treatment results in increased accessibility and the emergence of novel distal ATAC-seq peaks. (a), (b) UCSC tracks for reference-normalised H3K27me3 ChIP-seq and ATAC-seq tracks at *TRPS1* (a) and *GET3* loci (b) after GSK126 treatment and DMSO control. Vertical grey bars highlight examples of novel ATAC-seq peaks, and peaks with increased ATAC-seq reads after

GSK126 treatment. (c) Density plot showing the distribution of the novel ATAC-seq peaks that appear after GSK126 treatment as a distance from the closest TSS. ATAC-seq was performed in one replicate per condition.

5.3.2 Reduction in H3K27me3 levels in MLL-AF4 cells following EZH2/1 inhibition results in reduced TF binding and reduced chromatin accessibility.

Despite the genome-wide reduction in H3K27me3 after using the GSK126 inhibitor in MLL-AF4 cells, I observed that at a subset of loci, H3K27me3 levels were retained (Figure 5.4 a). Increasing the concentration of the inhibitor (6 μM instead of 2 μM) did not reduce H3K27me3 further (data not shown), suggesting that the residual methyltransferase activity of EZH1, which is not inhibited by the GSK126 inhibitor, is sufficient for the maintenance of H3K27me3 levels at those sites. In order to determine the effects of the complete erasure of H3K27me3 on accessibility and TF binding, I treated the cells with UNC1999, a dual inhibitor of EZH2 and EZH1 (Konze et al., 2013). Treatment of MLL-AF4 cells with this inhibitor (5 μM for 120h) resulted in global reduction in H3K27me3, as determined by reference-normalised ChIP-seq for H3K27me3 in cells treated with the inhibitor compared to DMSO-treated cells (ChIP-seq performed by Dr Marta Tapia) (Figure 5.4 b,d,e), and I further validated with ChIP-qPCR the reduction in H3K27me3 (Figure 5.4 c). The reduction in H3K27 methylation was also accompanied by an increase in H3K27ac (Figure 5.4 f), as previously described (Xu et al., 2015).

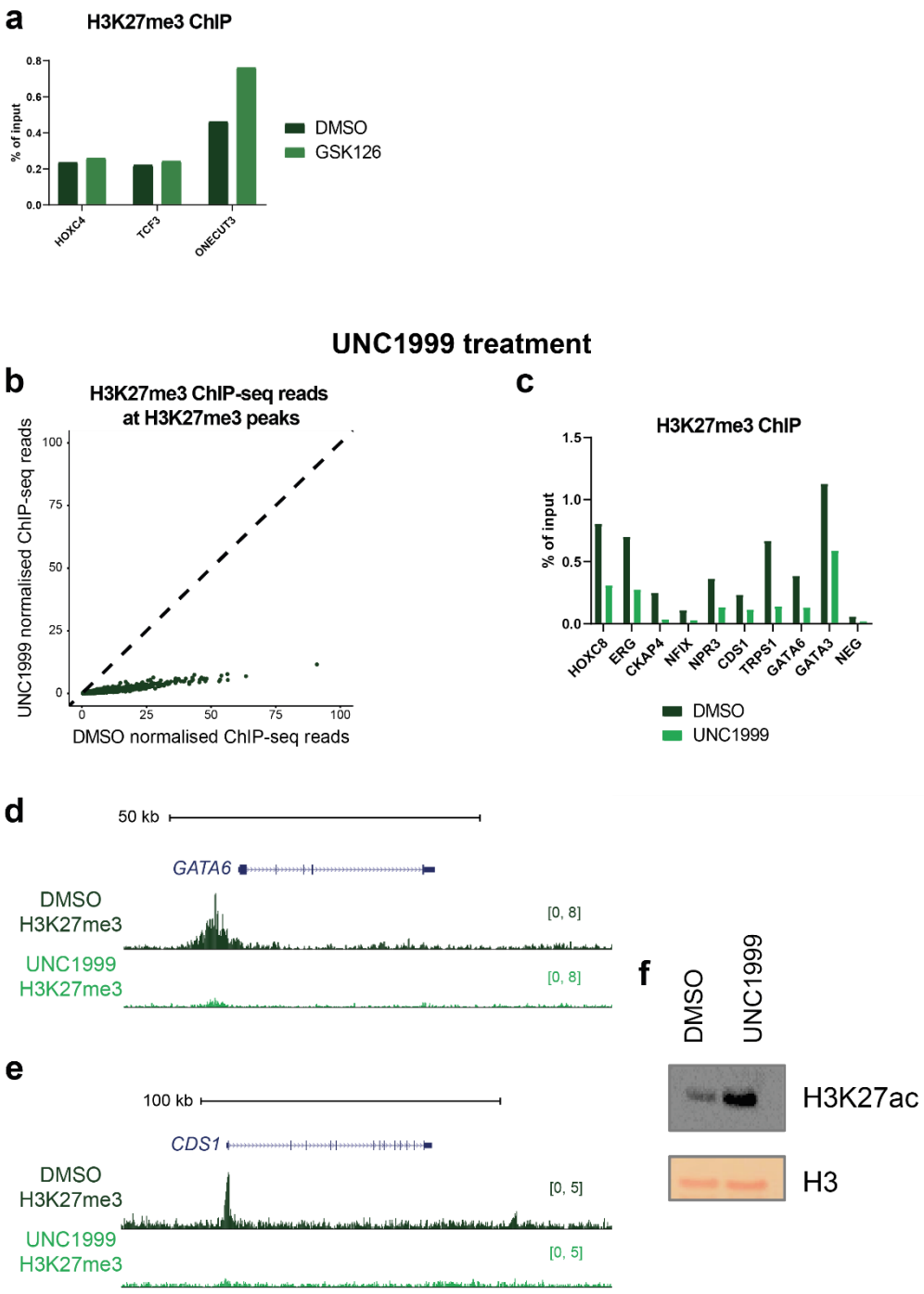


Figure 5.4: UNC1999 treatment in MLL-AF4 cells results in global decrease in H3K27me3. (a) ChIP-qPCR for H3K27me3 in SEM cells after GSK126 treatment (2 µM for 120h) and DMSO control, showing that at a subset of PcG-repressed loci, H3K27me3 levels are retained after GSK126 treatment (n=1). (b) Correlation of reference-normalised H3K27me3 ChIP-seq reads at H3K27me3 peaks after UNC1999 treatment and DMSO control, showing reduction in H3K27me3 reads after UNC1999 treatment. (c) ChIP-qPCR for H3K27me3 in SEM cells after UNC1999 treatment or DMSO control. One representative ChIP-qPCR replicate is shown. (d), (e) UCSC ChIP-seq tracks for reference-normalised H3K27me3 ChIP-seq at *GATA6* (d) and *CDS1* genes (e), showing reduction in H3K27me3 levels after UNC1999 treatment. Reference normalised H3K27me3 ChIP-seq after

UNC1999 treatment was performed by Dr M. Tapia. (f) Western blot for H3K27ac (and Ponceau stain showing H3 levels) in SEM cells after UNC1999 treatment and DMSO control.

To investigate the effect of the loss of H3K27me3 following the use of UNC1999 on RUNX1 binding, I performed reference-normalised ChIP-seq for RUNX1 in UNC1999-treated cells compared to untreated ones, and, in contrast to the effects of GSK126 on RUNX1 binding, I observed a significant reduction in RUNX1 binding genome-wide (Figure 5.5 a,b,c). In addition, ChIP-qPCR for RUNX1 at a subset of loci also showed reduction in RUNX1 binding after UNC1999 treatment, validating the ChIP-seq results (Figure 5.5 d).

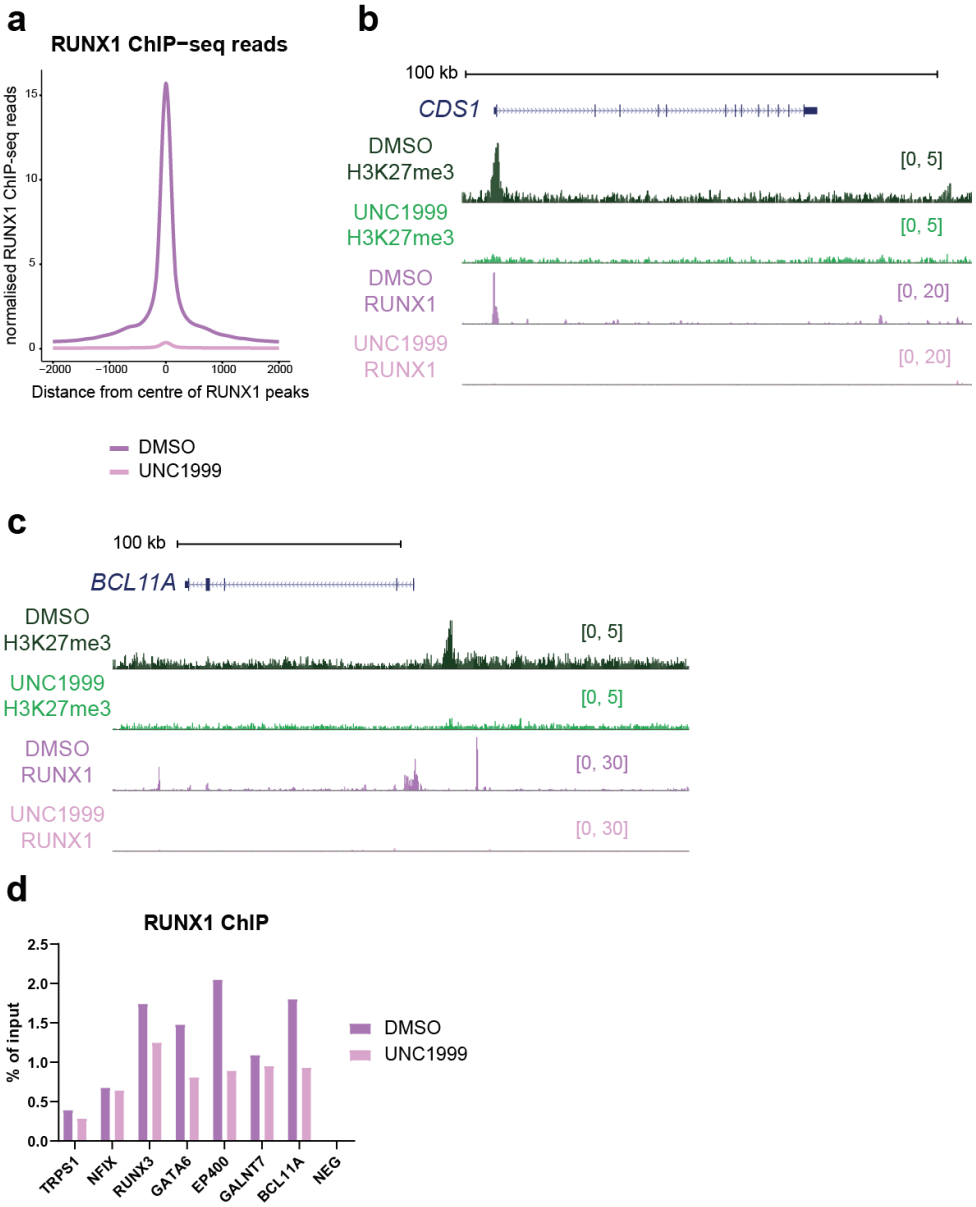


Figure 5.5: UNC1999 treatment in MLL-AF4 cells results in genome-wide reduction in RUNX1 binding. (a) Metaplot of mean reference-normalised RUNX1 ChIP-seq reads at RUNX1 peaks after UNC1999 treatment (5 μM for 120h) or DMSO control in SEM cells. (b) UCSC ChIP-seq tracks for reference-normalised H3K27me3 and RUNX1 ChIP-seq at *CDS1* (b) and *BCL11A* genes (c), showing that the reduction in H3K27me3 after UNC1999 treatment leads to reduction in RUNX1 binding. Reference normalised H3K27me3 ChIP-seq after UNC1999 treatment was performed by Dr M. Tapia. (d) ChIP-qPCR for RUNX1 in SEM cells after UNC1999 treatment or DMSO control. One representative ChIP-qPCR replicate from three biological replicates is shown.

The global reduction in RUNX1 binding after reduction in the repressive mark H3K27me3 following treatment with UNC1999, was rather unexpected. To check whether this is the case for other TFs as well, I examined if UNC1999 inhibitor treatment also affects binding of the ligand-induced TF Glucocorticoid receptor (GR). GR is found at the cytoplasm and upon treatment with dexamethasone, it gets rapidly relocated to the nucleus where it binds to its target genes (Weikum et al., 2017). Glucocorticoids are also commonly used to treat ALL, and are thought to induce differentiation of leukaemia cells. ChIP-qPCR for GR in DMSO-treated and UNC1999-treated cells followed by 1h treatment with dexamethasone showed the same pattern as RUNX1 binding, i.e. reduced GR binding in UNC1999-treated cells (Figure 5.6 a). This suggests that the chromatin landscape after the treatment of UNC1999 impedes the binding of TFs. Furthermore, the binding of the chromatin-modifying enzyme MLL was also reduced after UNC1999 treatment, as seen in MLL ChIP-seq (performed by Dr M. Tapia) (Figure 5.6 b), and this finding was further validated with ChIP-qPCR for MLL (Figure 5.6 c).

In addition, to understand how the treatment with UNC1999 inhibitor affects chromatin accessibility, ATAC-seq was performed. This showed that the inhibitor results in a reduced number of ATAC-seq peaks (Figure 5.7 a), and reduced accessibility in the peaks common to DMSO-treated and UNC1999-treated cells (ATAC-seq after UNC1999 treatment was performed by Dr Marta Tapia) (Figure 5.7 b). This is in contrast to the ATAC-seq results after treatment of the MLL-AF4 cells with the EZH2-specific inhibitor GSK126, which showed increased chromatin accessibility (Figure 5.2).

Overall, these results suggest that UNC1999 treatment in MLL-AF4 cells results in reduced chromatin accessibility and in reduced binding of TFs and chromatin complexes.

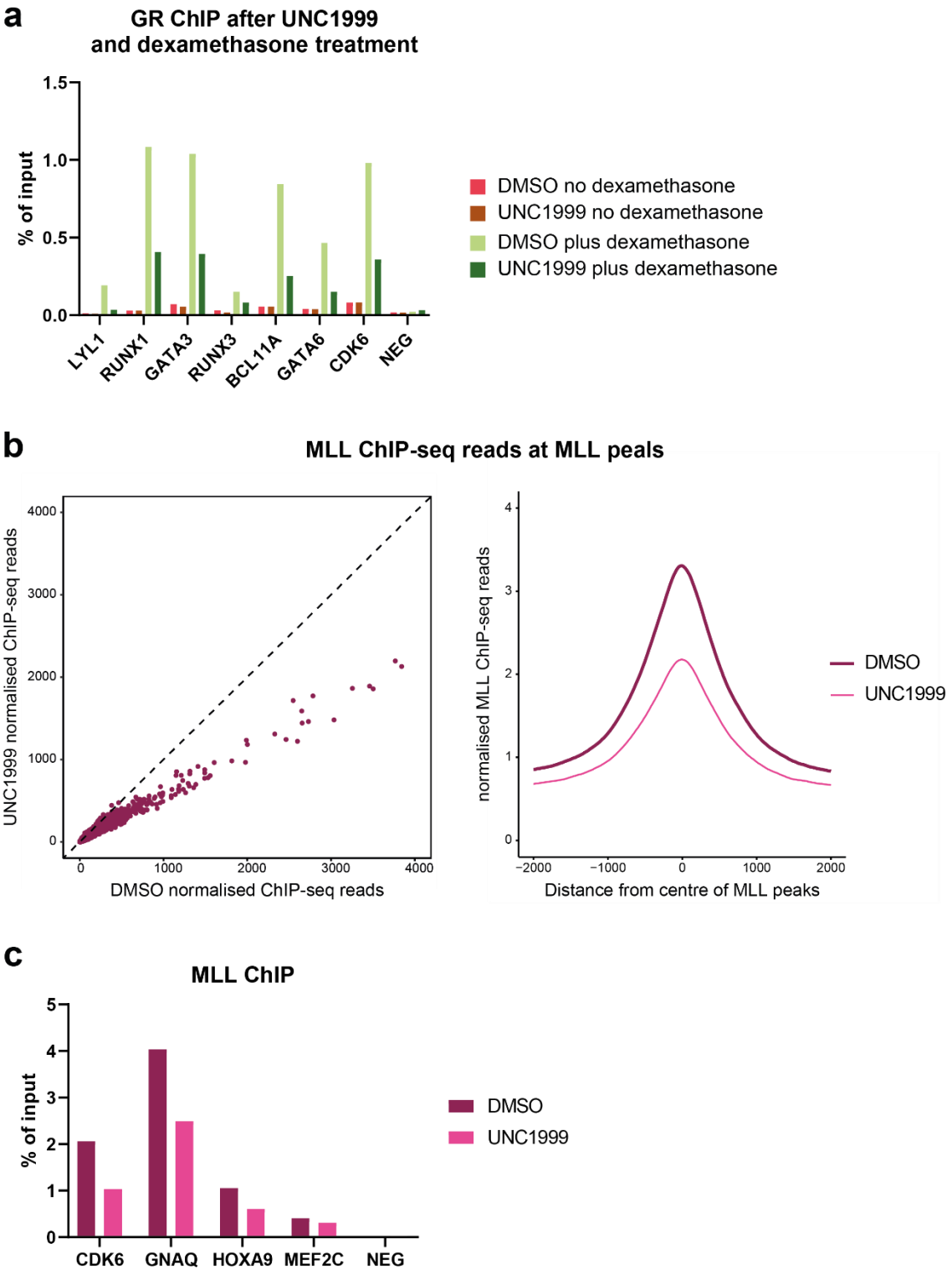


Figure 5.6: EZH2/1 inhibition after UNC1999 treatment in SEM cells results in reduced binding of GR and MLL. (a) ChIP-qPCR for glucocorticoid receptor (GR) in SEM cells after UNC1999 treatment and DMSO control, in unstimulated cells (no dexamethasone) and after 1h stimulation with

dexamethasone, showing that the treatment with dexamethasone results in GR binding to target genes in the nucleus, but its binding is reduced in cells that have been treated with UNC1999 (n=1). (b – *left panel*) Correlation of normalised MLL ChIP-seq reads at MLL peaks after UNC1999 treatment and DMSO control, showing reduced MLL binding after UNC1999 treatment. (*right panel*) Metaplot of mean normalised MLL ChIP-seq reads at MLL peaks after UNC1999 treatment or DMSO control in SEM cells. (c) ChIP-qPCR for MLL in SEM cells after UNC1999 treatment and DMSO control (n=1).

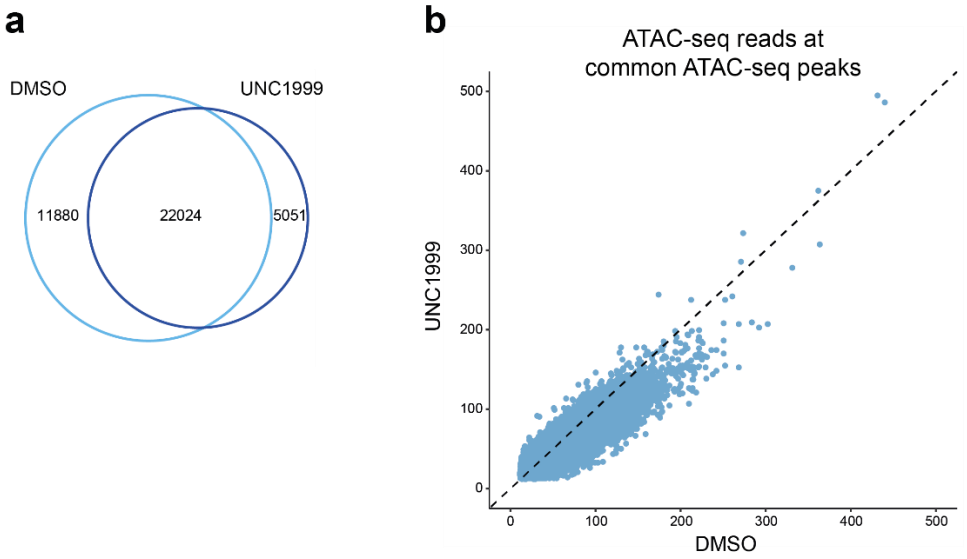


Figure 5.7: The dual EZH2/1 inhibitor UNC1999 results in reduced chromatin accessibility in MLL-AF4 cells. (a) Venn diagram showing the overlap of ATAC-seq peaks in SEM cells after UNC1999 treatment and DMSO control. (b) Correlation of ATAC-seq reads at peaks commonly identified in DMSO-treated and UNC1999-treated cells. ATAC-seq after UNC1999 treatment was performed by Dr M. Tapia.

Furthermore, consistent with the observed decrease in accessibility, reference-normalised ChIP-seq for H3K27ac in UNC1999-treated cells showed reduction in H3K27ac at H3K27ac ChIP-seq peaks compared to DMSO-treated cells (Figure 5.8 a), despite the global increase in H3K27ac (Figure 5.4 f). In contrast, at the TSSs of PcG-repressed genes that are marked by H3K27me3, I observed an increase in acetylation, while the opposite is the case for the TSSs of non PcG-repressed genes (Figure 5.8 b). This suggests that the increase in acetylation takes place at PcG-repressed regions and at low levels across the genome, i.e. at regions which were previously marked with H3K27me3 and H3K27me2. Indeed, after dividing the genome in 100 kb windows and calculating the H3K27ac ChIP-seq reads per window, I showed that, on average, H3K27ac ChIP-seq reads increase genome-wide in UNC1999-treated cells (Figure 5.8 c). The increase in acetylation at low levels across the

genome (and at regions previously marked with H3K27me3), rather than at H3K27ac ChIP-seq peaks, suggests that the deposition of acetyl marks by histone acetyltransferases occurs in a 'hit & run' fashion to the nucleosomes previously marked with H3K27me2 and H3K27me3, rather than after stable binding of the enzymes to TF binding sites.

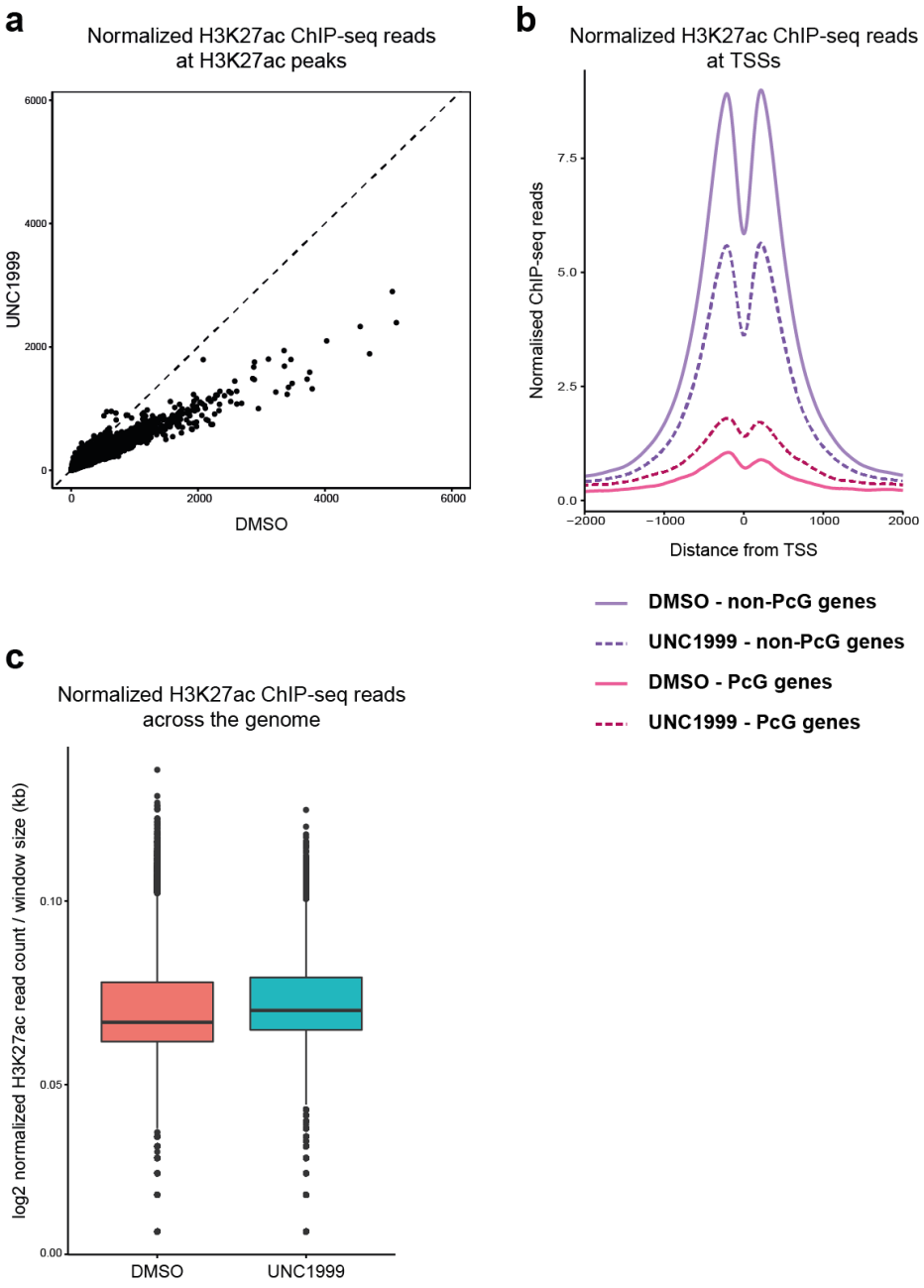


Figure 5.8: UNC1999 treatment results in increased H3K27ac levels across the genome and at PcG-repressed genes. (a) Correlation of reference-normalised H3K27ac ChIP-seq reads at H3K27ac peaks after UNC1999 treatment and DMSO control. (b) Metaplot of mean reference-normalised H3K27ac ChIP-seq reads at the TSSs of PcG-repressed genes (marked by H3K27me3) and non PcG-repressed genes, after UNC1999 treatment and DMSO control. (c) Boxplots of reference-

normalised H3K27ac reads at 100 kb genomic windows in DMSO-treated and UNC1999-treated cells. Reference-normalised H3K27ac ChIP-seq was performed by Dr M. Tapia.

To understand the transcriptional changes in MLL-AF4 cells after treatment with the EZH2/1 inhibitor UNC1999, I performed analysis of nascent RNA-seq results after UNC1999 treatment (nascent RNA-seq performed by Dr M.Tapia). This showed that ~1,300 genes were upregulated and ~700 were downregulated after UNC1999 treatment (Figure 5.9 a,b). By classifying the genes as PcG-repressed (associated with an H3K27me3 peak) or non-PcG-repressed, I have further shown that, on average, PcG-repressed genes tend to be upregulated, whereas the genes that are not PcG-repressed tend to be downregulated (Figure 5.9 c), suggesting that the reduction in H3K27me3 (accompanied by an increase in H3K27ac) at PcG-repressed genes is sufficient for their activation.

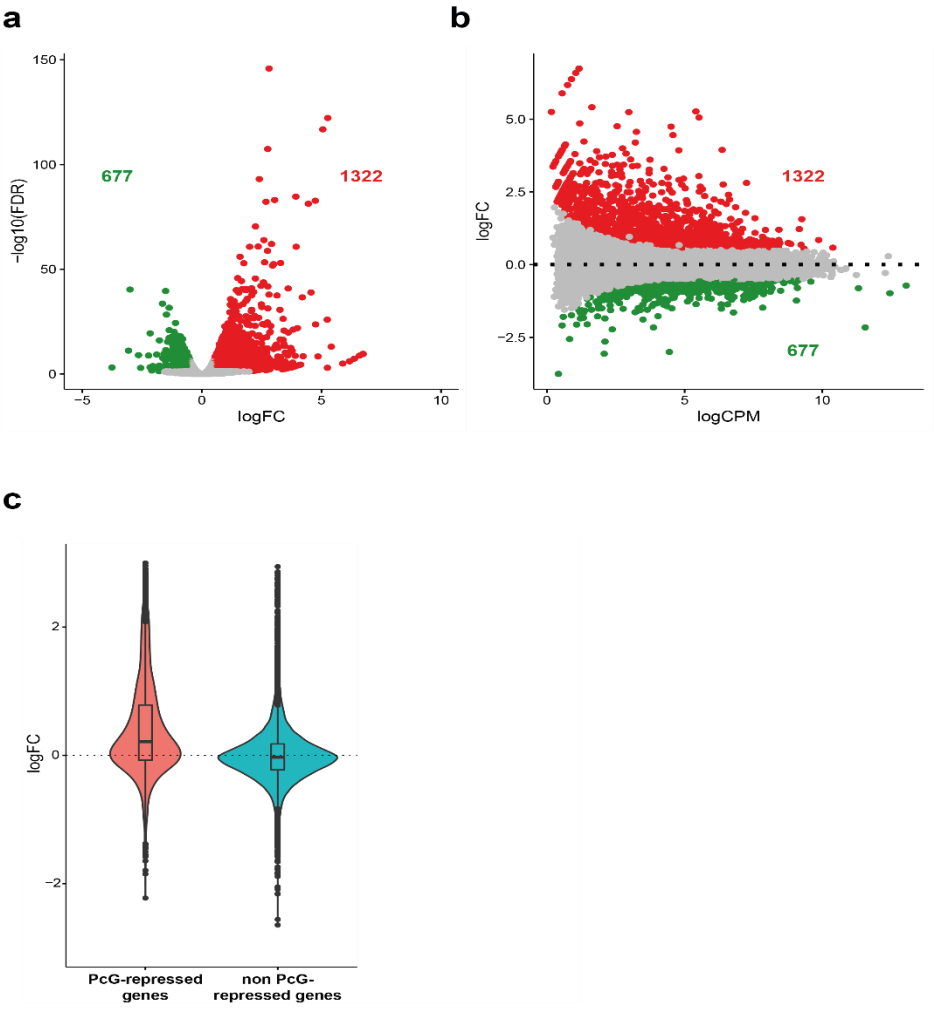


Figure 5.9: UNC1999 treatment in MLL-AF4 cells results in upregulation of PcG-repressed genes. (a) Volcano plot and (b) MA plot showing genes differentially expressed in UNC1999-treated cells compared to DMSO-treated cells, after performing nascent RNA-seq. Upregulated genes (n= 1322) are shown in red, downregulated genes (n=677) are shown in green, and unchanged genes in grey. Nascent RNA-seq was performed in 3 biological replicates per condition by Dr M. Tapia. (c) Violin plots showing logFC of PcG-repressed genes (marked by H3K27me3), and non PcG-repressed genes after UNC1999 treatment.

GO-term analysis for differentially expressed genes after UNC1999 treatment showed enrichment for the terms 'tissue morphogenesis' and 'tyrosine kinase signalling' for the upregulated genes (Figure 5.10 a), and enrichment for the terms 'cell cycle', 'cell cycle phase transition' and for 'DNA repair', for the downregulated genes (Figure 5.10 b). Consistent with this, cell cycle experiments have shown a G1-S arrest in a significant proportion of UNC1999-treated cells (cell cycle analysis performed by Dr M.Tapia – data not shown).

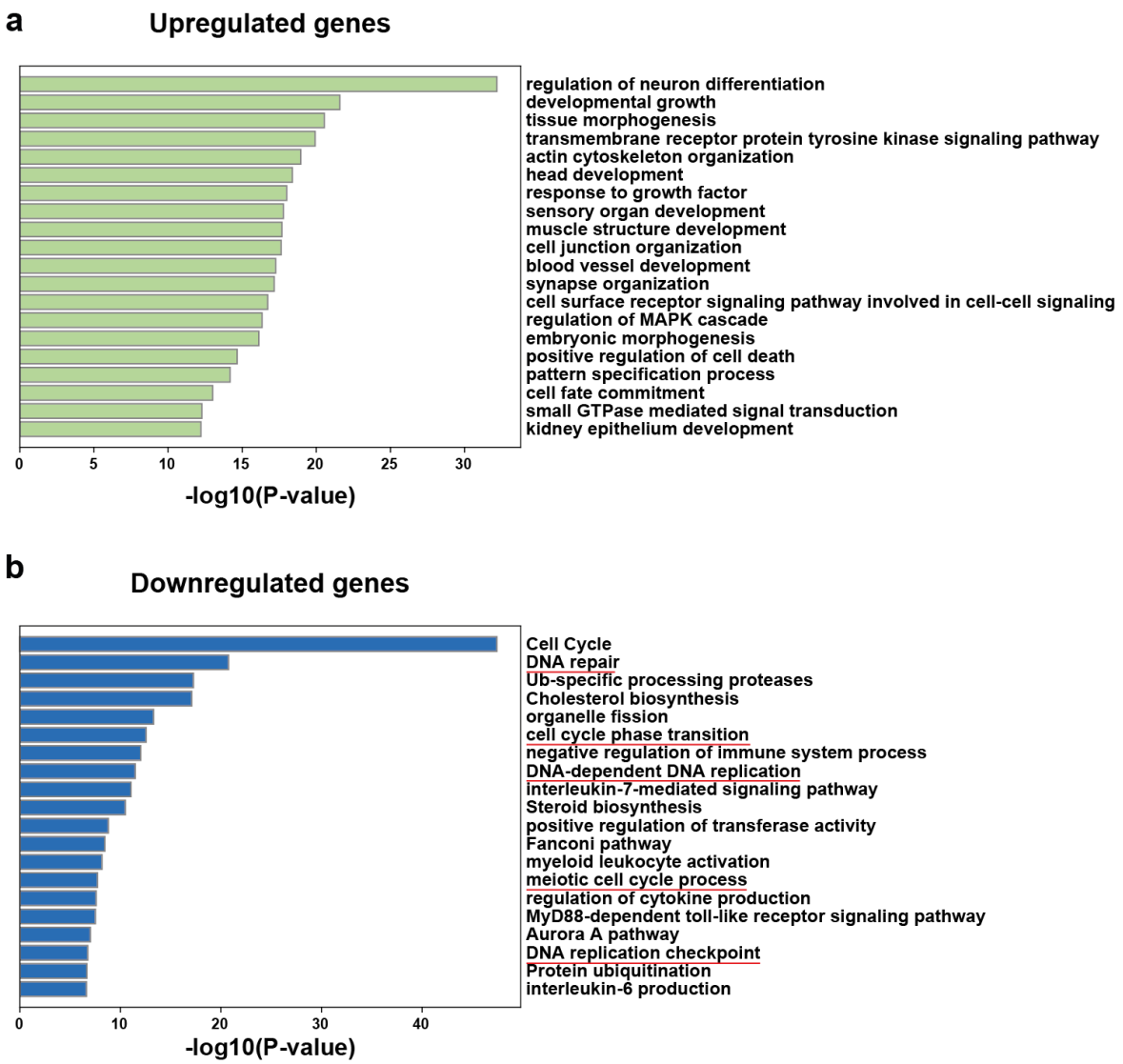


Figure 5.10: GO term analysis after UNC1999 treatment in MLL-AF4 cells demonstrates downregulation of cell cycle genes. GO term analysis for upregulated (a) and downregulated (b) genes after UNC1999 treatment.

5.3.3 EZH2/1 inhibition in MLL-AF4 cells results in accumulation of H3K9me3.

To determine whether the reduction in accessibility and TF binding is due to the inhibition of the EZH2/1 catalytic activity or a non-specific side-effect of the inhibitor, another dual inhibitor of EZH2 and EZH1 was used, DS-3201b, which has been used in preclinical models of *EZH2*-mutant lymphomas (Yamagishi et al., 2019) and in clinical trials for relapsed T-cell leukaemia/lymphoma. Similarly to the treatment with UNC1999, treatment with the DS-3201b inhibitor (2uM for 120h)

resulted in a significant reduction in H3K27me3 (Figure 5.11 a). ChIP-qPCR for RUNX1 after treatment with the DS-3201b inhibitor resulted in reduced RUNX1 binding (Figure 5.11 b), validating the effects of the UNC1999 inhibitor. These results further suggest that the inhibition of the catalytic activity of both H3K27 methyltransferases results in a reduction in TF binding.

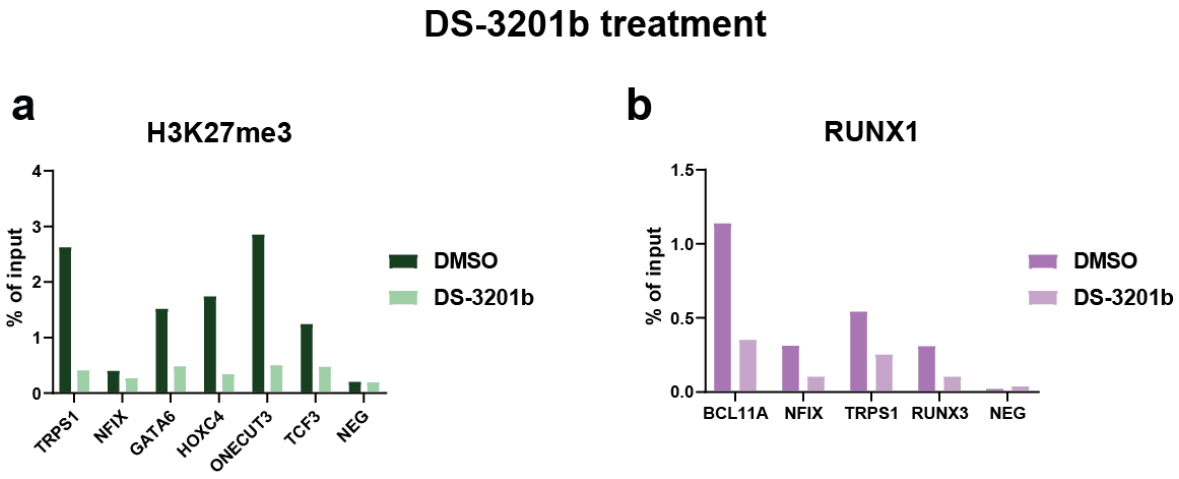


Figure 5.11: The dual EZH2/1 inhibitor DS-3201b results in reduced RUNX1 binding in MLL-AF4 cells. ChIP-qPCR for H3K27me3 (a) and RUNX1 (b) in SEM cells after DS-3201b treatment (2 μM for 120h) or DMSO control, showing that the reduction in H3K27me3 is accompanied by reduction in RUNX1 binding. One representative ChIP-qPCR replicate from 2 biological replicates is shown.

These observations, i.e. the reduction of accessibility and TF binding genome-wide after the reduction in the repressive histone marks H3K27me2/3 in MLL-AF4 cells are rather counterintuitive and unexpected. In order to understand the aetiology for the global reduction in TF binding, I then examined whether the treatment with the EZH2/1 inhibitors causes a genome-wide increase in DNA methylation or a gain in a repressive histone modification globally, and more specifically gain in the repressive mark H3K9me3, that could impede TF binding (Soufi et al., 2012). Analysis in UNC1999-treated cells compared to untreated ones did not show any changes in the global levels of DNA methylation or hydroxymethylation (mass spectrometry for DNA methylation performed by Dr Paolo Spingardi) (Figure 5.12 a). In contrast, ChIP-qPCR for H3K9me3 showed a significant increase in this mark after treatment of the cells with the UNC1999 inhibitor, both at PcG-repressed loci (for example *TRPPS1, NFIX, RUNX3*) and at active sites (*CEBPG, ELK3, KDM3A, CDK6, BCL2)*

(Figure 5.12 b). Western blotting for H3K9me3 also validated the global increase in this modification after treatment with UNC1999 (Figure 5.12 c). Similarly to UNC1999 treatment, treatment with the other dual EZH2/1 inhibitor DS-3201b also resulted in an increase in H3K9me3 levels (Figure 5.9 d). In contrast, treatment of the cells with the EZH2-specific inhibitor GSK126 did not lead to increased levels of H3K9me3 (Figure 5.12 e).

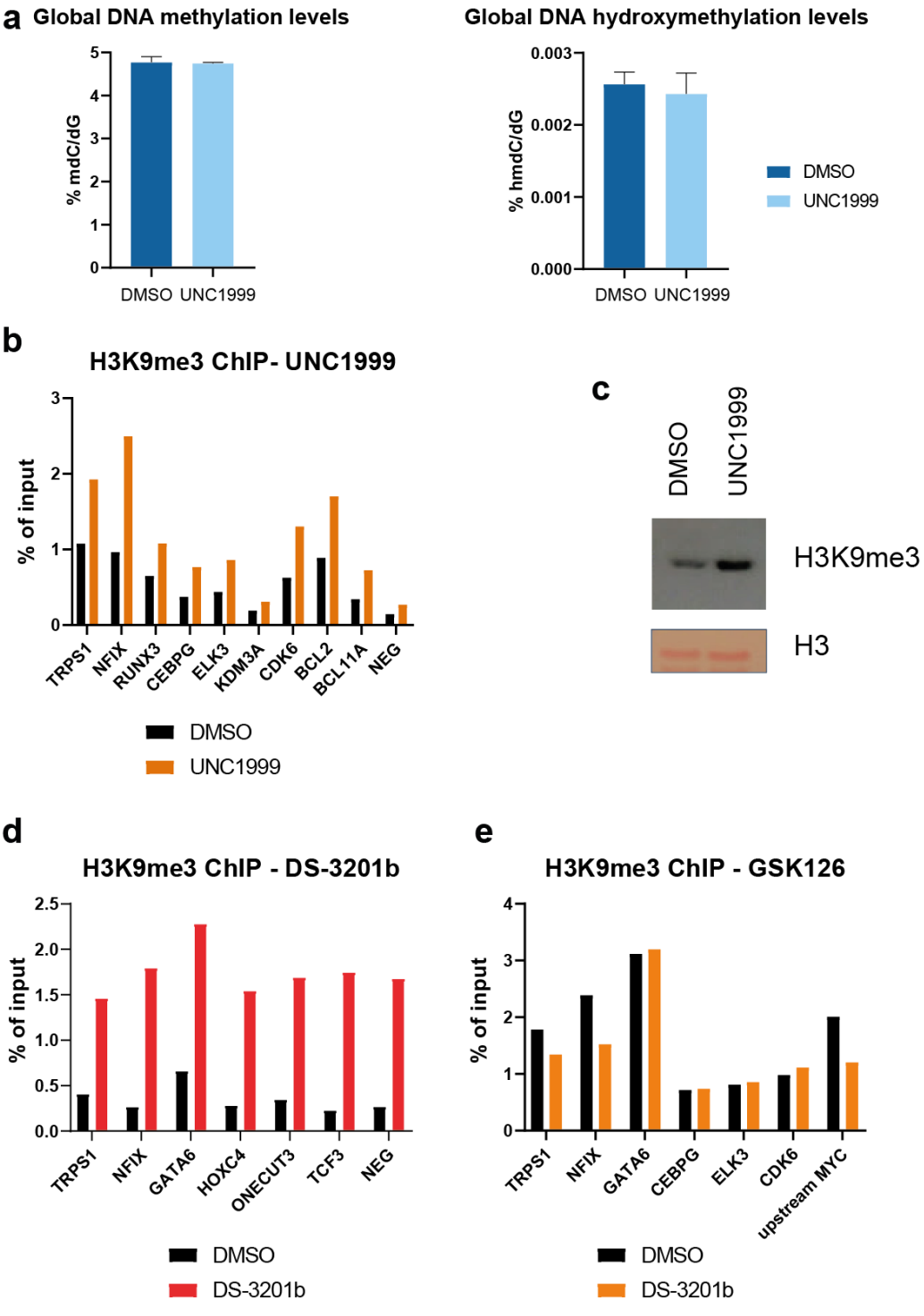


Figure 5.12: Dual EZH2/1 inhibitors, but not EZH2-specific inhibitors, result in increased H3K9me3 levels in MLL-AF4 cells. (a) Mass spectrometry for global DNA methylation and DNA hydroxymethylation in DMSO- and UNC1999-treated cells, showing that UNC1999 treatment does not result in changes in DNA methylation and hydroxymethylation levels. Mass spectrometry experiments were performed by Dr Paolo Spingardi in 3 biological replicates per condition. (b) ChIP-qPCR for H3K9me3 in SEM cells after UNC1999 treatment and DMSO control (n=1). (c) Western blot for H3K9me3 (and Ponceau stain showing H3 levels) in SEM cells after UNC1999 treatment and DMSO control, showing increased H3K9me3 levels after UNC1999 treatment. (d), (e) ChIP-qPCR for H3K9me3 in SEM cells after DS-3201b (d) and GSK126 (e) treatments compared to DMSO-treated cells (n=1), showing that the dual EZH2/1 inhibitor DS-3201b results in increased H3K9me3 levels, in contrast to the EZH2-specific inhibitor GSK126.

H3K9me3 has previously been shown to impede TF binding (Soufi et al., 2012), and its accumulation in MLL-AF4 cells after treatment with the dual EZH2/1 inhibitors provides an explanation for the significant reduction in chromatin accessibility and the reduction in TF binding. The reduction in TF binding after an increase in H3K9me3, despite the loss of H3K27me2/3 following the use of these inhibitors, also suggests that H3K9me3 (associated with constitutive heterochromatin) has a more dramatic effect in reducing accessibility and TF binding compared to the repressive PcG modification H3K27me3 (associated with facultative heterochromatin).

Analysis of the nascent RNA-seq data after UNC1999 treatment did not show transcriprional upregulation of H3K9 methyltransferases, or upregulation of the HP1 proteins (also known as CBX1, CBX3 and CBX5) that bind to H3K9me3 and promote transcriptional repression (Figure 5.13 a), nor downregulation of H3K9me2/3 demethylases (Figure 5.13 b). This suggests that other mechanisms are implicated in the accumulation of H3K9me3, rather than the upregulation of H3K9 methyltransferases or the downregulation of H3K9 demethylases.

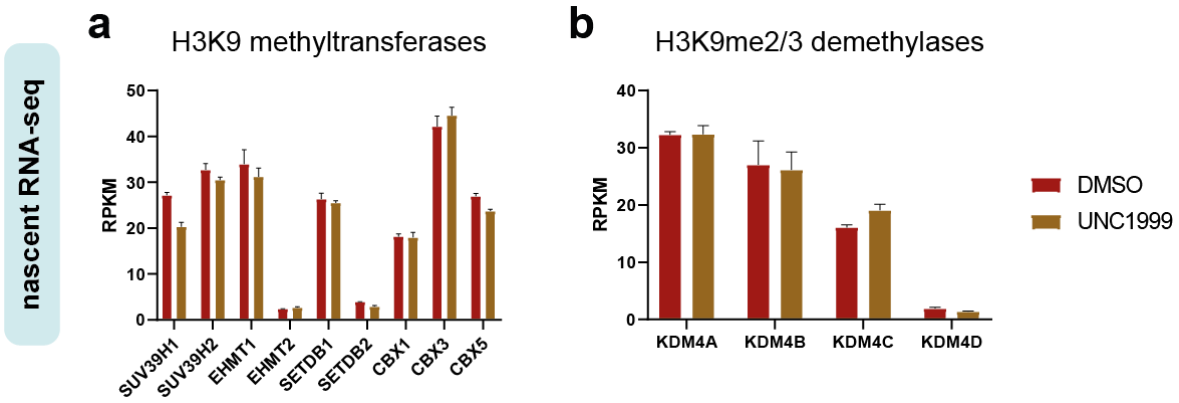


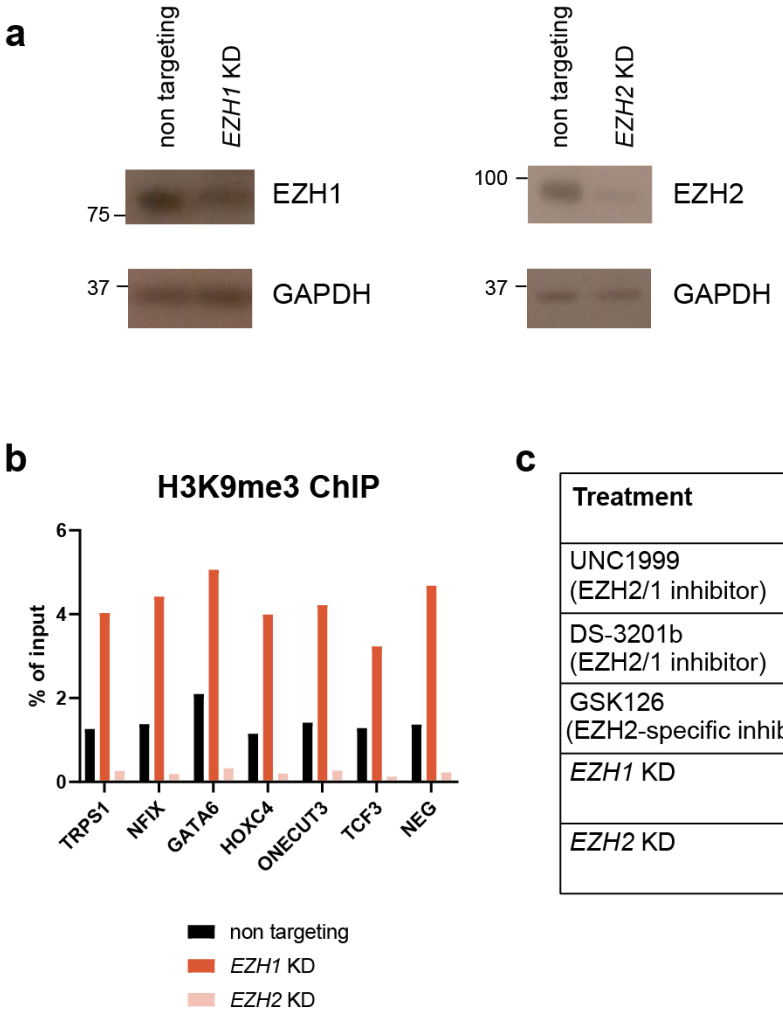
Figure 5.13: UNC1999 treatment does not lead to transcriptional changes in genes encoding H3K9 methyltransferases and demethylases. Nascent RNA-seq levels for genes encoding H3K9 methyltransferases (SUV39H1/2, EHMT1/2, SETDB1/2), HP1 proteins (also known as CBX1/3/5) (a) and H3K9me2/3 demethylases (KDM4A, KDM4B, KDM4C, KDM4D) (b) in DMSO-treated and UNC1999-treated cells. Nascent RNA-seq experiments were performed in 3 biological replicates and error bars show standard error of the mean (SEM). For all the genes included, differential gene expression analysis showed that their transcriptional changes after UNC1999 treatment were not statistically significant.

5.3.4 *EZH1* knockdown in MLL-AF4 cells results in accumulation of H3K9me3.

Given that the EZH1/2 inhibitors UNC1999 and DS-3201b result in increases in H3K9me3, while the EZH2-specific inhibitor GSK126 does not, I decided to check whether this was the effect of EZH1 inhibition. Thus, I performed siRNA knockdown of *EZH1* in MLL-AF4 cells; and indeed ChIP-qPCR validated the increase in H3K9me3 after *EZH1* KD, but this was not observed after *EZH2* KD (Figure 5.14 a,b,c). This suggests that EZH1 has an additional function compared to EZH2 in MLL-AF4 cells, in addition to its function as an H3K27 methyltransferase.

In addition, in MLL-AF4 cells, I observed that a subset of active promoters (not enriched in the H3K27me3 and H2AK119ub repressive marks) are bound by the PRC2 component SUZ12 but not by EZH2, based on ChIP-seq data from our lab. Examples are the promoters of the cell cycle genes *CDC25A* and *SKP2* genes (Figure 5.15 a,b), which are downregulated after treatment with the dual EZH2/1 inhibitor UNC1999 (Figure 5.15 c). This suggests that at these sites, the bound PRC2 complex consists of EZH1 rather than EZH2. However, validation that the bound protein at these loci is EZH1 was not feasible with ChIP-qPCR or ChIP-seq, as ChIP-qPCR experiments for EZH1 with commercially available antibodies were not successful (data not shown). The function of EZH1 in MLL-AF4 cells has not been further investigated in this thesis, but given the different effects of EZH1 and EZH2 inhibition, it could be addressed in future work. This could help us understand whether the binding of EZH1 at active genes (including cell cycle genes) has an activating role, as has been described in other cell types (Mousavi et al., 2012; Stojic et al., 2011), and also identify the

molecular mechanisms linking the inhibition of the EZH1 catalytic activity in MLL-AF4 cells to the accumulation of the repressive H3K9me3 mark.



c

Treatment	Increase in H3K9me3 ?
UNC1999 (EZH2/1 inhibitor)	yes
DS-3201b (EZH2/1 inhibitor)	yes
GSK126 (EZH2-specific inhibitor)	no
EZH1 KD	yes
EZH2 KD	no

Figure 5.14: *EZH1* siRNA knockdown in MLL-AF4 cells results in increase in H3K9me3. (a) Western blot for EZH1 (and GAPDH control) after *EZH1* KD and non targeting control (*left*) and for EZH2 (and GAPDH control) after *EZH2* KD and non targeting control (*right*). (b) ChIP-qPCR for H3K9me3 in SEM cells after *EZH1* and *EZH2* siRNA KD, and after non targeting control, showing that *EZH1* siRNA KD, but not *EZH2* siRNA KD, results in an increase in H3K9me3. One representative ChIP-qPCR replicate from 2 biological replicates is shown. (c) Table summarising the effects of PRC2 inhibitors and *EZH1/2* siRNA KD experiments on H3K9me3 levels, showing that *EZH1* siRNA KD and inhibitors that also target the catalytic activity of EZH1 result in increase in H3K9me3.

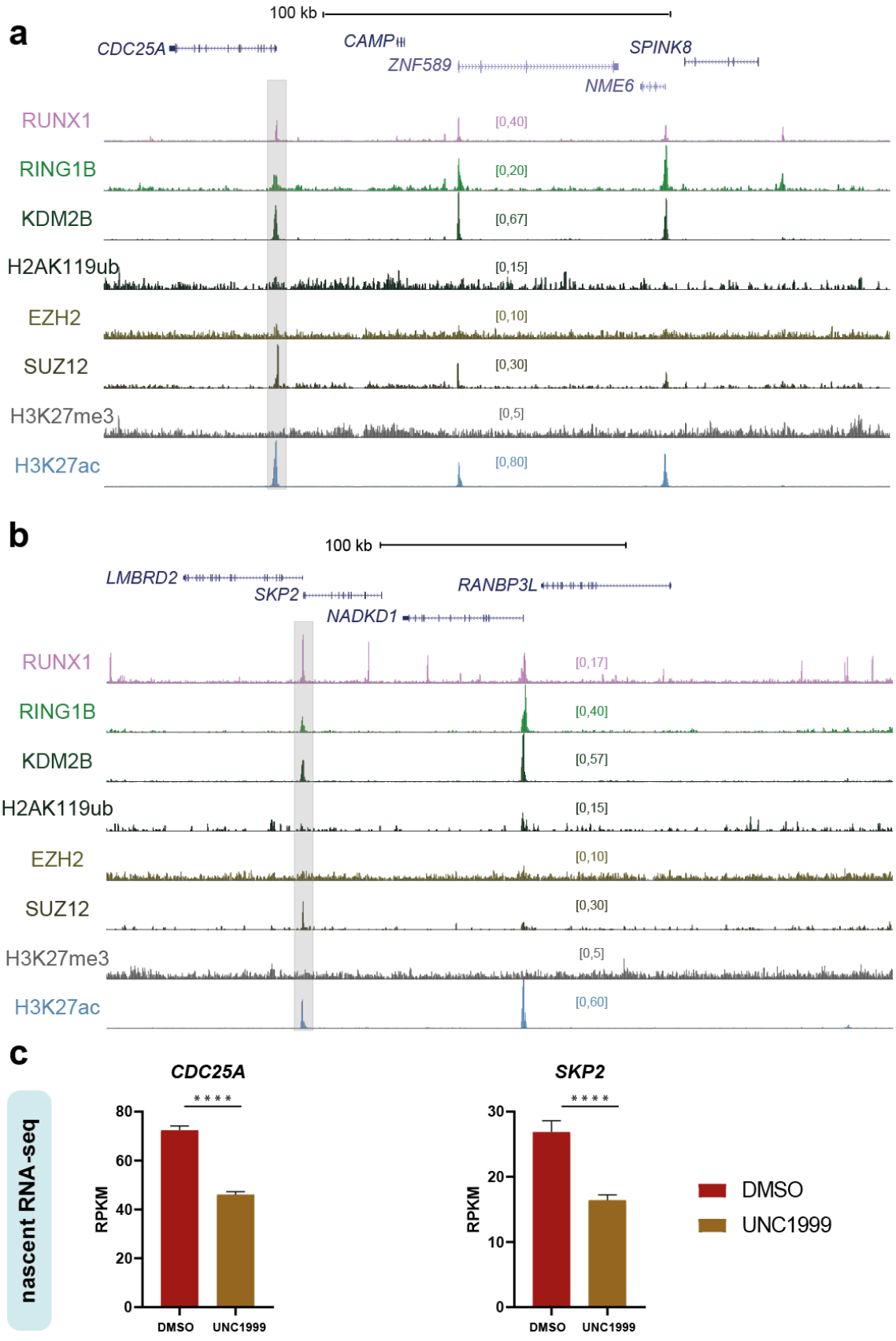


Figure 5.15: The PRC2-EZH1 complex is bound at the promoters of active genes, including cell cycle genes *CDC25A* and *SKP2*. (a), (b) UCSC ChIP-seq tracks for RUNX1, RING1B, KDM2B, H2AK119ub, EZH2, SUZ12, H3K27me3 and H3K27ac at the *CDC25A* (a) and *SKP2* (b) loci, showing

examples of active genes (enriched in H3K27ac, but not in the repressive PcG marks H3K27me3 and H2AK119ub) bound by the PRC2 (and PRC1) complexes. The binding of SUZ12, but not EZH2, at the promoters of *CDC25A* and *SKP2* genes suggests that the bound PRC2 complex contains EZH1. (c) Nascent RNA-seq levels for *CDC25A* and *SKP2* genes after UNC1999 treatment and DMSO control, showing downregulation of these genes after UNC1999 treatment. Nascent RNA-seq was performed in 3 biological replicates by Dr M. Tapia. ****FDR < 0.0001

5.4 Discussion

In this chapter, I aimed to investigate how alterations in the chromatin environment affect the binding of TFs, and more specifically the binding of RUNX1. I have shown in the previous chapters that a key function of RUNX1 in different contexts is the recruitment of PcG proteins and the stabilisation of Polycomb domains. For this chapter, I examined how, in turn, the chromatin environment shaped by the activity of PcG proteins affects the binding of RUNX1. For this, MLL-AF4 cells, which are dependent on PRC2 activity for their survival (Xu et al., 2015), were treated with PRC2 inhibitors.

I observed that the reduction in H3K27me3 levels after inhibition of the catalytic activity of EZH2 results in a significant increase in chromatin accessibility and increase in RUNX1 binding at a subset of RUNX1 binding sites, although more replicates are needed to robustly validate these observations. Future experiments could aim to dissect how the reduction in H3K27me3 affects RUNX1 binding genome-wide, by performing reference-normalised ChIP-seq for RUNX1. In addition, it is worth investigating whether the novel ATAC-seq peaks that emerge following the reduction in H3K27me3 and H3K27me2 levels are also enriched in H3K27ac, suggesting the emergence of novel enhancers, by performing reference-normalised ChIP-seq for H3K27ac. If this is the case, it would also be of interest to examine whether, novel enhancer-promoter interactions are formed, by performing Capture-C in GSK126-treated cells compared to DMSO-treated ones. Furthermore, nascent RNA-seq after GSK126 treatment in MLL-AF4 cells would provide information on how changes in accessibility correlate with transcriptional changes.

In contrast to the inhibition of the EZH2 catalytic activity, the dual inhibition of EZH2 and EZH1 results, unexpectedly, in significant reduction in chromatin accessibility and reduction in TF binding, including global reduction in the binding of RUNX1. I have shown that this is due to the accumulation of the repressive mark H3K9me3, and my data further demonstrate that *EZH1* KD recapitulates the increase in H3K9me3 that is observed after the use of the dual EZH2/1 inhibitors, suggesting an additional role of EZH1 compared to EZH2 in MLL-AF4 cells. Although EZH1 and EZH2 are the mutually exclusive catalytic components of the PRC2 complex, it has been shown in previous studies that the two proteins can have different molecular functions. More specifically, the activity of EZH1 as an H3K27me3 methyltransferase is not as potent as that of EZH2 (Lavarone et al., 2019; Shen et al., 2008), probably due to inefficient allosteric stimulation of the PRC2-EZH1 complex compared to the PRC2-EZH2 complex (Lee et al., 2018); in addition, *in vitro* experiments have shown that, in contrast to EZH2, EZH1 can cause compaction of nucleosome arrays (Margueron et al., 2008). On the other hand, it has also been shown in myoblasts and myocytes that, although the PRC2 complex is traditionally thought to be localised at PcG-repressed loci, EZH1 can also be found at the promoters of active genes, and that it contributes to their activation (Mousavi et al., 2012; Stojic et al., 2011). Despite the fact that the mechanisms by which EZH1 promotes the activation of target genes have not fully elucidated, it is possible that post-translational modifications abrogate its ability to deposit H3K27me3, similar to what has been suggested for EZH2 in prostate cancer (Xu et al., 2012). In addition, given that the SET domain of EZH1 has been reported to be indispensable for gene activation (Mousavi et al., 2012), it is possible that EZH1 is contributing to transcriptional activation through the deposition of H3K27me1 rather than H3K27me3. As discussed in the introduction (section 1.6.4), there is recent evidence that H3K27me1, which is found at the gene bodies of active genes, contributes to gene activation (Wang et al., 2018), in contrast to the repressive role of H3K27me2/3 marks.

Although the molecular mechanisms explaining how the inhibition of EZH1 results in the accumulation of H3K9me3, and in the downregulation of cell cycle genes and cell cycle arrest, have

not been fully addressed in this thesis, I have hypothesised that EZH1 binds to, and activates a subset of genes, including the cell cycle genes *CDC25A* and *SKP2*. Potentially this occurs through the deposition of H3K27me1. Inhibition of the catalytic activity of EZH1 could result in the downregulation of these cell cycle genes, resulting in cell cycle arrest. The inhibition of cell cycle progression due to cellular stress has been previously shown to result in the formation of senescence-associated heterochromatic foci (SAHF), enriched in H3K9me3 (Lanigan et al., 2011; Narita et al., 2003), linking the observed cell cycle arrest with the increase in H3K9me3.

More specifically, I have hypothesised that EZH1 inhibition leads to the downregulation of *CDC25A* and *SKP2* genes (which promote cell cycle progression), resulting in reduced activation of cyclin-dependent kinases (CDKs) and senescence-like cell cycle arrest (Hoffmann et al., 1994; Shen and Huang, 2012). The reduction in CDK activity leads to hypophosphorylation of the Retinoblastoma protein (Rb), which in turn, remains bound to the E2F TFs, and recruits repressors including HDACs and the SUV39H1 H3K9 methyltransferase to E2F binding sites. This leads to the accumulation of heterochromatic marks at E2F binding sites in senescent cells (Lanigan et al., 2011; Narita et al., 2003). Consistent with this hypothesis, it has been shown the activity of EZH1 is required for the deposition of H3K27me1 and H3K27me2 in adult HSCs, and for the prevention of senescence-like cell cycle arrest (Hidalgo et al., 2012). In the case of adult HSCs, the activity of EZH1 prevents senescence through the repression of *CDKN2A*, which is not seen in UNC1999-treated MLL-AF4 cells (i.e. upregulation of *CDKN2A* is not observed after UNC1999 treatment), suggesting that, if the inhibition of the EZH1 catalytic activity results in senescence in MLL-AF4 cells, this takes place through the transcriptional regulation of other target genes.

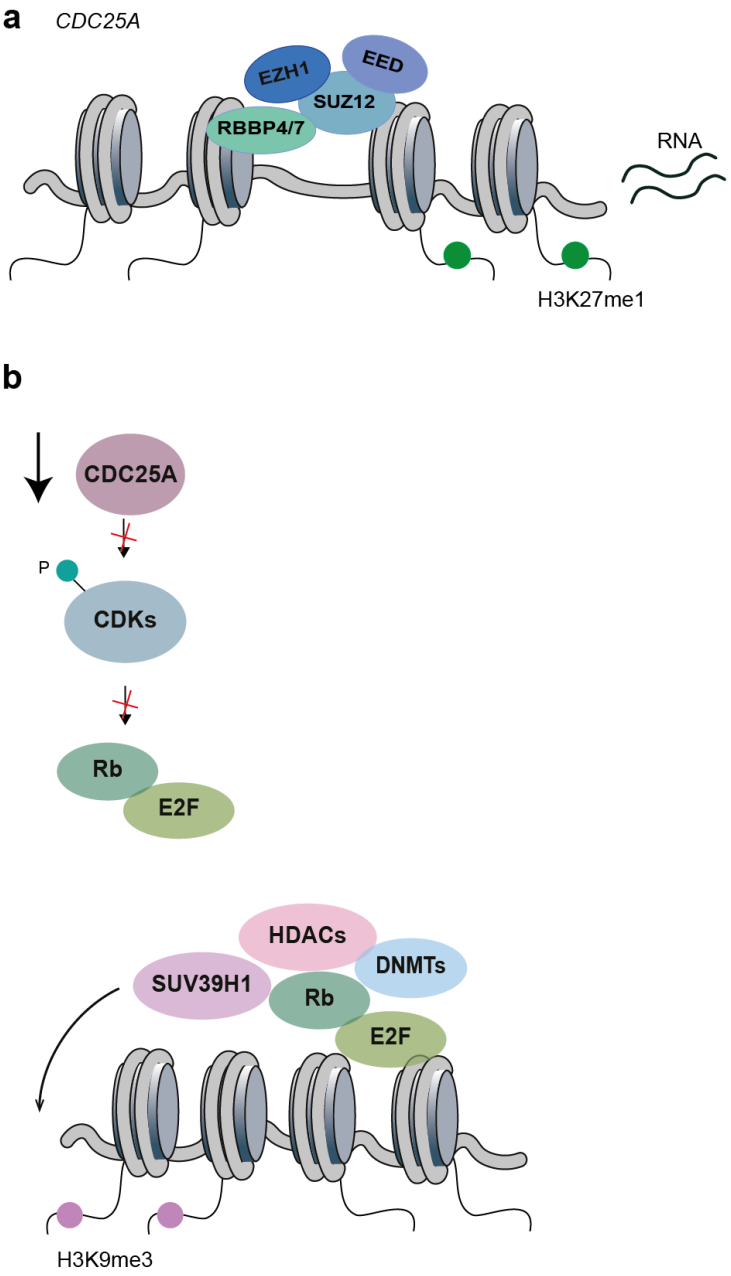


Figure 5.16: Schematic of the working hypothesis linking the inhibition of EZH1 activity to cell cycle arrest and accumulation of H3K9me3. (a) At the *CDC25A* locus and at other active loci, the PRC2-EZH1 complex activates genes through the deposition of H3K27me1. Upon reduction of the EZH1 catalytic activity, *CDC25A* gets downregulated. (b) (*Top*) CDC25A is a phosphatase that removes the inhibitory phosphorylation marks of CDKs, resulting in their activation. Downregulation of *CDC25A*, upon EZH1 inhibition, results in increased CDK phosphorylation and inhibition of their activity. In turn, this leads to reduced phosphorylation of Rb, which remains in complex with E2F TFs. At E2F binding sites in the nucleus, Rb recruits repressors including HDACs, DNA methyltransferases and the H3K9 methyltransferase SUV39H1, leading to the deposition of H3K9me3 and the formation of senescence-associated heterochromatic foci (SAHF).

To test if this is indeed the case, i.e. that EZH1 activates cell cycle genes and that the inhibition of its activity results in cell cycle arrest and the formation of senescence-associated heterochromatic foci, future experiments will aim to test whether *EZH1* KD results in the same transcriptional changes as UNC1999 treatment. More specifically, future experiments should aim to determine whether *EZH1* KD leads to the downregulation of cell cycle genes including *CDC25A*, and whether this is accompanied by a reduction in H3K27me1. If this is observed, additional experiments including *CDC25A* KD or the inhibition of the CDC25A catalytic activity will be performed, to test if these experiments also result in increase in H3K9me3, recapitulating the effect of the EZH2/1 inhibitors.

Chapter 6 - General discussion and future work

RUNX1 is a master regulator of the development of the haematopoietic system. Thus, it is not surprising that mutations and translocations affecting *RUNX1* are very frequently implicated in haematological neoplasms, including AML and MDS (Sood et al., 2017). In addition, the wildtype RUNX1 protein has been shown to be required for disease maintenance in leukaemias with fusion proteins, including MLL-AF4 and RUNX1-RUNX1T1 leukaemias (Ben-Ami et al., 2013; Goyama et al., 2013; Harman et al., 2021; Wilkinson et al., 2013). Despite the extensive studies on RUNX1 during normal haematopoiesis and in haematological malignancies, the molecular mechanisms of how RUNX1 impacts gene regulation remain incompletely understood and need to be investigated in detail. This will allow us to understand how RUNX1 controls the expression of downstream target genes, which will potentially help us identify pathways that can be therapeutically targeted.

6.1 RUNX1 is sufficient to recruit both co-activators and repressors.

In order to validate previously published RUNX1 gene-regulatory interaction partners and determine whether RUNX1 preferentially recruits complexes that function as activators or as repressors, the TetR system, an *in vivo* intracellular tethering system was employed (Blackledge et al., 2014). Using this system, I have shown that RUNX1 is sufficient to recruit co-activators (including the p300 acetyltransferase, the SWI/SNF nucleosome remodelling complex and the BAP1 deubiquitinase complex) and the transcriptional machinery. In addition, RUNX1 is sufficient to recruit repressors including histone deacetylases and the PRC1 complexes (Figure 6.1). This suggests that RUNX1 is sufficient to recruit multiple gene-regulatory complexes, including complexes with opposite function.

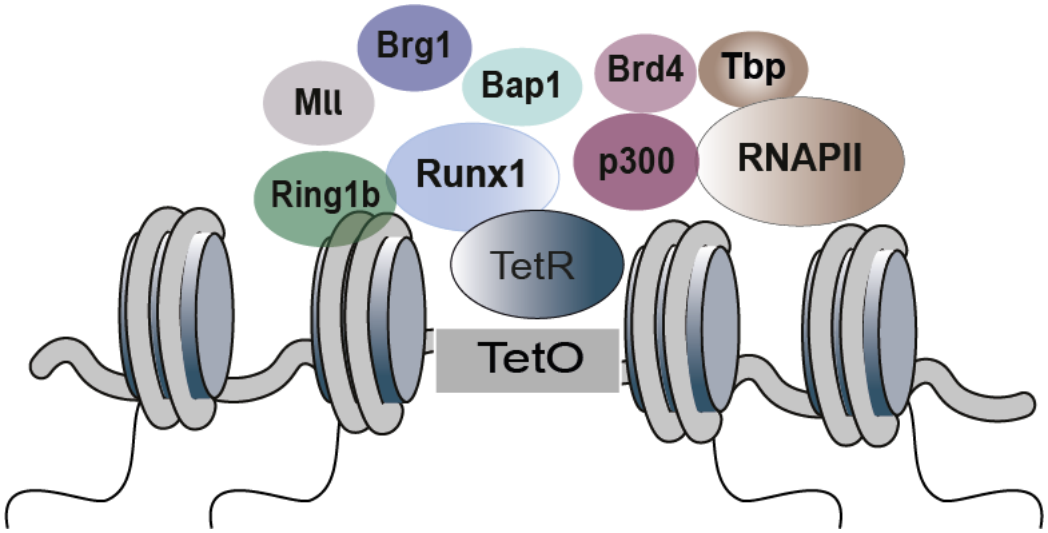


Figure 6.1: Schematic of the Runx1 interaction partners validated using the TetR system.

To further understand how the binding of RUNX1 impacts the recruitment of complexes with opposite activities, I have focused on endogenous sites bound by Runx1 in mESCs that express the Runx1-TetR fusion protein, and I have shown that RUNX1-mediated recruitment of co-factors is context-dependent. In particular, my results demonstrate that RUNX1 binding to active loci results in the stabilisation of co-activator binding and deposition of active histone marks, that correlates with transcriptional changes. In contrast, RUNX1 binding to PcG-repressed loci further results in the recruitment of PcG complexes, contributing to gene repression (Figure 6.2). The differential recruitment of co-activators and co-repressors by RUNX1 at active and repressed loci respectively is potentially due to the stabilisation of multivalent interactions (Rothbart et al., 2013; Ruthenburg et al., 2011; Ruthenburg et al., 2007b). More specifically, at active loci, the low-affinity interactions between RUNX1 and co-activators are further stabilised through additional low-affinity interactions between the RUNX1-recruited co-activators and the active histone modifications and the co-factors at these sites. Similarly, RUNX1-mediated PRC1 recruitment to PcG-repressed sites is potentially stabilised through the additional low-affinity interactions between PcG proteins and the repressive histone modifications at these loci. These results suggest that RUNX1 binding affects the chromatin

landscape through the recruitment of gene-regulatory and chromatin-modifying complexes, and in turn the existing chromatin environment affects the RUNX1-mediated stabilisation of co-factors.

Future experiments could examine how the differentiation of Runx1-expressing mESCs towards embryoid bodies (EBs), and the transcriptional changes that this transition entails, affect the preferential interaction between Runx1 and its co-factors at specific sites. Furthermore, it would be of interest to investigate in future experiments whether Runx1 impacts chromatin interactions through the recruitment of its co-factors. More specifically, Capture-C experiments could show whether RUNX1 binding to active sites contributes to the stabilisation of enhancer-promoter interactions, and also whether it results in increased interactions between PcG-occupied regions, when it is bound to PcG-repressed loci. In addition, it is important to examine whether tethering Runx1 to the TetO locus, in which the PcG proteins prevail, results in the emergence of novel chromatin interactions by performing Capture-C from the TetO locus.

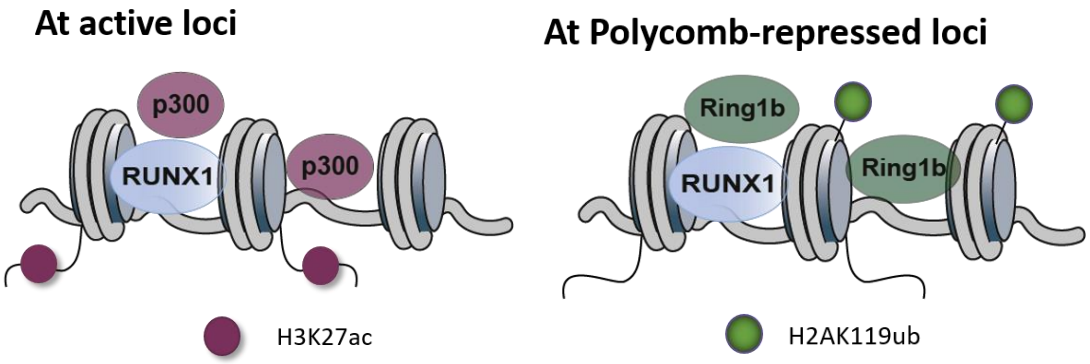


Figure 6.2: RUNX1-mediated recruitment of co-factors is context dependent. RUNX1 stabilises the binding of co-activators, including p300, at active sites through multivalent interactions, while it stabilises the binding of PcG repressive complexes at PcG-repressed sites.

6.2 RUNX1 contributes to the stabilisation of PRC1 binding in leukaemic contexts.

By focusing on the RUNX1-PRC1 interaction in a leukaemic context, I have observed significant RUNX1-PRC1 chromatin co-occupancy in MLL-AF4 ALL cells, both at PcG-repressed sites and also at active loci, including at the promoters of active genes and at enhancer elements. Furthermore, reduction in RUNX1 binding is followed by reduction in RING1B binding genome-wide, suggesting

that RUNX1 stabilises the recruitment of PRC1. This is translated as transcriptional upregulation of a subset of PcG-repressed loci. In addition, I have shown that the wildtype RUNX1 protein stabilises the binding of PRC1 at key RUNX1-bound PcG-repressed sites in RUNX1-RUNX1T1 leukaemias.

In MLL-AF4 cells I have also noticed the co-localisation of RUNX1 with both variant and canonical PRC1 complexes at active sites. In accordance with these data, recent studies have shown that PcG complexes are bound at active loci in different cell types (Cohen et al., 2018; Cohen et al., 2019; Frangini et al., 2013; Gao et al., 2014; Loubiere et al., 2020; Schaaf et al., 2013; Zhang et al., 2021), and in some instances they can promote enhancer-promoter interactions through the polymerisation of the canonical PRC1 components (Loubiere et al., 2020; Zhang et al., 2021). Thus, it would be relevant to investigate whether the RUNX1-mediated recruitment of PRC1 at active sites in MLL-AF4 cells results in chromatin interactions, including enhancer-promoter interactions between PcG-occupied sites by performing Capture-C after *RUNX1* and after *RING1A/B* siRNA KD (Figure 6.2).

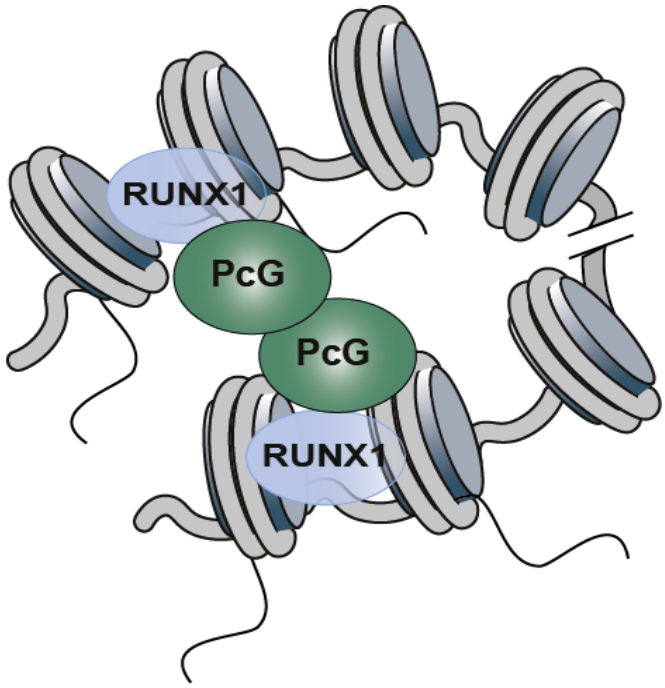


Figure 6.3: Schematic of chromatin interactions between RUNX1-bound loci, mediated through the recruitment of PcG proteins.

Additionally, taking this a step further, out of the context of leukaemia, given these findings that a key function of RUNX1 is the recruitment of PRC1 complexes, it is worth studying during normal haematopoiesis, for example during lymphocyte differentiation or megakaryocytic maturation, whether RUNX1 stabilises PRC1 binding. If this is the case, it would be interesting to investigate the functional role of this interaction; more specifically, future experiments using *in vitro* differentiation systems and *in vivo Runx1*-deficient mouse models could test whether the RUNX1-mediated PRC1 recruitment leads to chromatin interactions between repressed PcG-bound regions, or between active PcG-occupied key target genes, resulting in a regulatory topology that is required for the induction of gene expression during subsequent steps of the differentiation process (Cruz-Molina et al., 2017; Gentile et al., 2019; Kondo et al., 2014).

6.3 RUNX1-mediated recruitment of PcG proteins provides a mechanistic explanation for the co-occurrence of *RUNX1* and *PcG* mutations.

The relative enrichment of *PcG* mutations in leukaemias with *RUNX1* mutations, including mutations that give rise to truncated RUNX1 mutant proteins, and in RUNX1-RUNX1T1 AML cases (Booth et al., 2018; Duployez et al., 2016; Faber et al., 2016; Gaidzik et al., 2016; Lehnertz et al., 2017; Sashida et al., 2014; Simon et al., 2017), suggests that this combination of mutations offers a selective advantage to the leukaemic clones. However, the molecular mechanisms of the functional co-operation of *RUNX1* and *PcG* mutations have not been investigated in detail. This co-occurrence of mutations, in combination with the observation that RUNX1 stabilises the recruitment of PRC1, prompted me to test the hypothesis that *RUNX1* mutations function, at least partially, by reducing the RUNX1-mediated PRC1 recruitment.

To investigate whether this is the case for truncating *RUNX1* mutations that lead to the loss of the C-terminal part of RUNX1, I have shown, using the TetR system, that the C-terminal part of RUNX1

is mediating the RUNX1-PRC1 interaction. This suggests that truncating *RUNX1* mutations, common in AML, FPD/AML and ETP ALL, and also the *RUNX1-RUNX1T1* translocation, give rise to mutant proteins that lack the ability to associate with PRC1. Future experiments, using mouse models, could further investigate *in vivo* whether the loss of RUNX1 expression or the expression of truncated RUNX1 mutant proteins, affects RUNX1-mediated PRC1 recruitment at key RUNX1-bound PcG-repressed genes in haematopoietic tissues, and whether the acquisition of additional *PcG* mutations is required for aberrant activation of PcG-repressed genes.

Furthermore, by focusing on *PcG*-mutant RUNX1-RUNX1T1 leukaemias, I have identified a transcriptional signature of *PcG*-mutant RUNX1-RUNX1T1 patients compared to *PcG*-wildtype ones, including upregulated PcG-repressed genes; additionally, at a subset of these, RUNX1 binding stabilises the binding of RING1B. Similarly, *PcG*-mutant *RUNX1*-mutant AML patients have a distinct transcriptional profile compared to *RUNX1*-mutant AML patients without *PcG* mutations. In both cases, PcG-repressed genes including the cell cycle gene *CCND1* become upregulated in *PcG*-mutant patients. *CCND1* overexpression potentially results in changes in cell cycling and in increased proliferation, conferring a selective advantage to the leukaemic cells. This further suggests that patients could benefit from combination therapies including cytostatic regimens targeting the CDK4/6 proteins that associate with CCND1. As these analyses were performed on publicly available bulk RNA-seq data from heterogeneous leukaemic subclones, future experiments could aim to dissect at a single-cell level, using techniques like Target-seq, the effect of *PcG* mutations on gene expression (Rodriguez-Meira et al., 2019), by comparing the transcriptional status of *PcG*-mutant patient subclones compared to *PcG*-wildtype ones.

6.4 Effects of PRC2 inhibition in the chromatin landscape in MLL-AF4 leukaemia

The interaction between RUNX1 and PcG proteins, in combination with the fact that the existing chromatin landscape affects the binding of RUNX1 and the RUNX1-mediated recruitment of co-factors, prompted me to investigate how changes in Polycomb activity affect the binding of RUNX1. To address this, I have treated MLL-AF4 ALL cells with PRC2 inhibitors, which have been suggested as a therapeutic approach in *MLL*-rearranged leukaemias (Xu et al., 2015). I observed that the inhibition of EZH2 catalytic activity results in the emergence of novel ATAC-seq peaks, suggesting the emergence of novel enhancers. In addition, I observed an increase in RUNX1 binding at a subset of RUNX1 binding sites. It would be of interest to dissect with future experiments how RUNX1 binding changes genome-wide after use of the EZH2 inhibitors, and whether the changes in accessibility correlate with transcriptional changes.

In contrast, the use of compounds that inhibit the catalytic activity of both EZH2 and EZH1, has very different results on accessibility and TF binding. More specifically, I have observed a reduction in accessibility and reduced RUNX1 binding genome-wide using the EZH2/1 inhibitors. Further experiments showed that this is attributed to a global increase in the repressive mark H3K9me3, which is mediated by the inhibition of the catalytic activity of EZH1. As it has been shown that heterochromatic regions enriched in H3K9me3 are compacted and impede the binding of TFs (Hiragami-Hamada et al., 2016; Soufi et al., 2012), the accumulation of H3K9me3 provides a mechanistic explanation for the decrease in chromatin accessibility and the reduction in TF binding in my experiments. In addition, transcriptomic studies have shown that the increase in H3K9me3 following the use of EZH2/1 inhibitors in MLL-AF4 cells is accompanied by cell cycle arrest.

In order to link the effects of EZH1 inhibition on cell cycle progression and the increase in H3K9me3, future experiments should aim to validate whether the binding of EZH1 to the promoters of cell cycle genes results in their activation, and whether this takes place through the deposition of H3K27me1. It would also be of interest to examine whether the reduction in the expression levels

of EZH1-bound cell cycle genes results in cell cycle arrest and increased deposition of H3K9me3 through the activation of the E2F-pRb pathway and the formation of senescence-associated heterochromatic foci (Lanigan et al., 2011) (Figure 5.16), or whether other mechanisms link the inhibition of EZH1 activity with the observed cell cycle arrest and the accumulation of H3K9me3.

6.5 Summary

In this thesis I investigated how RUNX1 binding shapes the chromatin landscape and impacts gene regulation through the recruitment of gene-regulatory co-factors, and how, in turn, the chromatin environment affects the binding of RUNX1 and RUNX1-mediated recruitment of co-factors. Overall, I have shown that RUNX1 is sufficient to recruit co-activators and co-repressors, including the PcG proteins, in a context-dependent fashion. By focusing on the RUNX1-PcG interaction, I have shown that a key role of RUNX1 in leukaemias with fusion proteins is to stabilise PcG binding; and that truncated RUNX1 mutant proteins cannot associate with PRC1, providing a mechanistic understanding for the enrichment of *PcG* mutations in *RUNX1*-mutant leukaemias.

7. Appendix

Table 7.1: Accession numbers of ChIP-seq/BioCap datasets in SEM cells

Data	GEO Accession numbers
BioCap	GSE83671: GSM2212233
BRD4 ChIP-seq	GSE83671: GSM2212237
EZH2 ChIP-seq	Unpublished (performed by Dr T. Milne)
KDM2B ChIP-seq	GSE83671: GSM2212235
MED1 ChIP-seq	GSE83671: GSM2212239
MLL ChIP-seq	GSE74812: GSM1934095
MYB ChIP-seq	GSE117865: GSM3312815
H2AK119ub ChIP-seq	Unpublished (performed by D. Dimou)
H3K4me1 ChIP-seq	GSE74812: GSM1934091
H3K4me3 ChIP-seq	GSE74812: GSM1934090
H3K27ac ChIP-seq	GSE74812: GSM1934089
H3K79me3 ChIP-seq	GSE74812: GSM1934093
OGT ChIP-seq	Unpublished (performed by Dr T. Milne)
p300 ChIP-seq	GSE85988: GSM2290169
RING1B ChIP-seq	Unpublished (performed by Dr T. Milne)
RUNX1 ChIP-seq	GSE42075: GSM1032075
RUNX2 ChIP-seq	GSE117865: GSM3312806
RUNX3 ChIP-seq	Unpublished (performed by Dr T. Milne)

Table 7.2: Mutations and GEO accession numbers for RUNX1-RUNX1T1 AML patient RNA-seq datasets analysed in this thesis

sample ID	Mutations	Accession number
05H042	*ASXL1, KIT, NRAS, SMC1A*	GSE62190: GSM1521608
06H035	*ASXL1, NRAS*	GSE62190: GSM1521611
08H042	*ASXL1, KIT, SMC1A*	GSE62190: GSM1521614
09H040	*ASXL1*	GSE62190: GSM1521616
05H184	*ASXL2, KDM6A, ZBTB7A*	GSE62190: GSM1521610
12H098	*ASXL2, KIT*	GSE62190: GSM1521621
12H166	*ASXL2, KIT*	GSE62190: GSM1521622
12H183	*ASXL2*	GSE62190: GSM1521624
08H034	*ASXL2, FLT3, KIT, WT1*	GSE62190: GSM1521613
03H065	*TET2, CSF3R*	GSE62190: GSM1521606
10H119	*DNMT3A, KIT*	GSE62190: GSM1521618
10H030	*SMC1A, NRAS*	GSE62190: GSM1521617
13H169	*SMC3*	GSE67039: GSM1636915

Table 7.3: Mutations and GEO accession numbers for *RUNX1*-mutant AML patient RNA-seq datasets analysed in this thesis

sample ID	mutations	Accession number
03H024	*RUNX1, ASXL1, IDH2, SRSF2, STAG2*	GSE67040: GSM1636675
03H119	*RUNX1, ASXL1, BCOR, SF3B1*	GSE67040: GSM1203309
04H001	*RUNX1, ASXL1, SETBP1, SRSF2*	GSE67040: GSM1636684
04H055	*RUNX1, ASXL1, FLT3, SRSF2*	GSE67040: GSM1636690
04H103	*RUNX1, ASXL1, EZH2, KIT, TET2, ZRSR2*	GSE67040: GSM1636694
05H013	*RUNX1, ASXL1*	GSE67040: GSM1636705
06H133	*RUNX1, ASXL1, FLT3, PHF6, STAG2, U2AF1, WT1*	GSE67040: GSM1634297
07H124	*RUNX1, ASXL1, IDH1, SRFS2, TET2*	GSE67040: GSM1636750
07H155	*RUNX1, ASXL1, IDH2, KMT2A, KRAS, NRAS, SRSF2, STAG2*	GSE67040: GSM1636758
08H137	*RUNX1, ASXL1, DNMT3A, EZH2, TET2*	GSE67040: GSM1636782
09H079	*RUNX1, ASXL1, EP300, ETV6, SRFS2*	GSE67040: GSM1636801
10H063	*RUNX1, ASXL1, KRAS*	GSE67040: GSM1521589
10H136	*RUNX1, ASXL1, CEBPA, CSF3R, EZH2, WT1*	GSE67040: GSM1636830
12H030	*RUNX1, ASXL1, IDH2*	GSE67040: GSM1273641
12H151	*RUNX1, ASXL1, JAK2, NF1, PHF6, TET2*	GSE67040: GSM1636881
13H039	*RUNX1, ASXL1, IDH1, SRFS2*	GSE67040: GSM1636893
13H048	*RUNX1, ASXL1, BCOR, CBL, KMT2A, PTPN11, RAD21*	GSE67040: GSM1636895
14H027	*RUNX1, ASXL1, SRSF2*	GSE67040: GSM1636927
04H138	*RUNX1, BCOR, HIST1H3H, TET2, USP7*	GSE67040: GSM1636704
06H019	*RUNX1, BCOR, NRAS, U2AF1*	GSE67040: GSM1636719
09H015	*RUNX1, BCOR, FLT3, TET2*	GSE67040: GSM1636788
09H046	*RUNX1, BCOR, CEBPA, CUX1, DNMT3A, KMT2C, KMT2D, NF1, SMC5, STAG2, TET2, TP53, USP7*	GSE67040: GSM1273629
09H092	*RUNX1, BCOR, NRAS*	GSE67040: GSM1636807
11H232	*RUNX1, BCOR, KRAS, NF1*	GSE67040: GSM1636858
03H081	*RUNX1, DNMT3A, EZH2, GATA2, IDH1, IKZF1, U2AF1*	GSE67040: GSM1636681
08H088	*RUNX1, CEBPA, EZH2, NRAS, TET2*	GSE67040: GSM1636777
06H048	*RUNX1, DNMT3A, ETV6, KMT2A, NRAS, U2AF1*	GSE67040: GSM1636724
06H146	*RUNX1, FLT3, KMT2A*	GSE67040: GSM1634298
08H065	*RUNX1, CEBPA*	GSE67040: GSM1636774
13H139	*RUNX1, FLT3, IDH1, NPM1, SRSF2*	GSE67040: GSM1636910
12H106	*RUNX1, CSF3R, TP53*	GSE67040: GSM1636874
09H106	*RUNX1, DNMT3A, FLT3, KMT2A*	GSE67040: GSM1636810
11H138	*RUNX1, FLT3, SF3B1*	GSE67040: GSM1636846

12H175	*RUNX1, CEBPA, GATA2*	GSE67040: GSM1636886
11H046	*RUNX1, IDH1, JAK2, SRSF2*	GSE67040: GSM1203340
05H149	*RUNX1, DNMT3A, IDH2, JAK2, SRSF2*	GSE67040: GSM1203314
08H062	*RUNX1, DNMT3A, IDH2, SRSF2, TET2*	GSE67040: GSM1636772
08H087	*RUNX1, IKZF1, NRAS, PTPN11*	GSE67040: GSM1636776
03H030	*RUNX1, JAK2, STAG2*	GSE67040: GSM1636677
07H069	*RUNX1, KIT, KMT2C*	GSE67040: GSM1273621
10H007	*RUNX1, FLT3, KMT2A, KMT2D, TP53*	GSE67040: GSM1636814
05H180	*RUNX1, NRAS*	GSE67040: GSM1636714
11H140	*RUNX1, NRAS, SRSF2, STAG2, TET2*	GSE67040: GSM1636847

References

Agger, K., P.A. Cloos, J. Christensen, D. Pasini, S. Rose, J. Rappsilber, I. Issaeva, E. Canaani, A.E. Salcini, and K. Helin. 2007. UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature*. 449:731-734.

Alexander, J.M., J. Guan, B. Li, L. Maliskova, M. Song, Y. Shen, B. Huang, S. Lomvardas, and O.D. Weiner. 2019. Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *Elife*. 8.

Allen, M.D., C.G. Grummitt, C. Hilcenko, S.Y. Min, L.M. Tonkin, C.M. Johnson, S.M. Freund, M. Bycroft, and A.J. Warren. 2006. Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase. *EMBO J*. 25:4503-4512.

Allende, M.L., M. Manzanares, J.J. Tena, C.G. Feijoo, and J.L. Gomez-Skarmeta. 2006. Cracking the genome's second code: enhancer detection by combined phylogenetic footprinting and transgenic fish and frog embryos. *Methods*. 39:212-219.

Almeida, M., G. Pintacuda, O. Masui, Y. Koseki, M. Gdula, A. Cerase, D. Brown, A. Mould, C. Innocent, M. Nakayama, L. Schermelleh, T.B. Nesterova, H. Koseki, and N. Brockdorff. 2017. PCGF3/5-PRC1 initiates Polycomb recruitment in X chromosome inactivation. *Science*. 356:1081-1084.

Andersson, A.K., J. Ma, J. Wang, X. Chen, A.L. Gedman, J. Dang, J. Nakitandwe, L. Holmfeldt, M. Parker, J. Easton, R. Huether, R. Kriwacki, M. Rusch, G. Wu, Y. Li, H. Mulder, S. Raimondi, S. Pounds, G. Kang, L. Shi, J. Becksfort, P. Gupta, D. Payne-Turner, B. Vadodaria, K. Boggs, D. Yergeau, J. Manne, G. Song, M. Edmonson, P. Nagahawatte, L. Wei, C. Cheng, D. Pei, R. Sutton, N.C. Venn, A. Chetcuti, A. Rush, D. Catchpoole, J. Heldrup, T. Fioretos, C. Lu, L. Ding, C.H. Pui, S. Shurtleff, C.G. Mullighan, E.R. Mardis, R.K. Wilson, T.A. Gruber, J. Zhang, J.R. Downing, and P. St. Jude Children's Research Hospital-Washington University Pediatric Cancer Genome. 2015. The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias. *Nat Genet*. 47:330-337.

Andersson, R. 2015. Promoter or enhancer, what's the difference? Deconstruction of established distinctions and presentation of a unifying model. *Bioessays*. 37:314-323.

Andreu-Vieyra, C.V., R. Chen, J.E. Agno, S. Glaser, K. Anastassiadis, A.F. Stewart, and M.M. Matzuk. 2010. MLL2 is required in oocytes for bulk histone 3 lysine 4 trimethylation and transcriptional silencing. *PLoS Biol*. 8.

Arber, D.A., A. Orazi, R. Hasserjian, J. Thiele, M.J. Borowitz, M.M. Le Beau, C.D. Bloomfield, M. Cazzola, and J.W. Vardiman. 2016. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*. 127:2391-2405.

Ardehali, M.B., A. Mei, K.L. Zobeck, M. Caron, J.T. Lis, and T. Kusch. 2011. Drosophila Set1 is the major histone H3 lysine 4 trimethyltransferase with role in transcription. *EMBO J*. 30:2817-2828.

Arents, G., R.W. Burlingame, B.C. Wang, W.E. Love, and E.N. Moudrianakis. 1991. The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. *Proc Natl Acad Sci U S A*. 88:10148-10152.

Asada, S., S. Goyama, D. Inoue, S. Shikata, R. Takeda, T. Fukushima, T. Yonezawa, T. Fujino, Y. Hayashi, K.C. Kawabata, T. Fukuyama, Y. Tanaka, A. Yokoyama, S. Yamazaki, H. Kozuka-Hata, M. Oyama, S. Kojima, M. Kawazu, H. Mano, and T. Kitamura. 2018. Mutant ASXL1 cooperates with BAP1 to promote myeloid leukaemogenesis. *Nat Commun*. 9:2733.

Ayton, P.M., E.H. Chen, and M.L. Cleary. 2004. Binding to nonmethylated CpG DNA is essential for target recognition, transactivation, and myeloid transformation by an MLL oncoprotein. *Mol Cell Biol*. 24:10470-10478.

Bachmann, I.M., O.J. Halvorsen, K. Collett, I.M. Stefansson, O. Straume, S.A. Haukaas, H.B. Salvesen, A.P. Otte, and L.A. Akslen. 2006. EZH2 expression is associated with high proliferation rate and aggressive tumor subgroups in cutaneous melanoma and cancers of the endometrium, prostate, and breast. *J Clin Oncol*. 24:268-273.

Bajusz, I., L. Sipos, Z. Gyorgypal, E.A. Carrington, R.S. Jones, J. Gausz, and H. Gyurkovics. 2001. The Trithorax-mimic allele of Enhancer of zeste renders active domains of target genes accessible to polycomb-group-dependent silencing in Drosophila melanogaster. *Genetics*. 159:1135-1150.

Baker, L.A., C.D. Allis, and G.G. Wang. 2008. PHD fingers in human diseases: disorders arising from misinterpreting epigenetic marks. *Mutat Res*. 647:3-12.

Bakshi, R., M.Q. Hassan, J. Pratap, J.B. Lian, M.A. Montecino, A.J. van Wijnen, J.L. Stein, A.N. Imbalzano, and G.S. Stein. 2010. The human SWI/SNF complex associates with RUNX1 to control transcription of hematopoietic target genes. *J Cell Physiol*. 225:569-576.

Balasubramani, A., A. Larjo, J.A. Bassein, X. Chang, R.B. Hastie, S.M. Togher, H. Lahdesmaki, and A. Rao. 2015. Cancer-associated ASXL1 mutations may act as gain-of-function mutations of the ASXL1-BAP1 complex. *Nat Commun*. 6:7307.

Banerji, J., S. Rusconi, and W. Schaffner. 1981. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell*. 27:299-308.

Bannister, A.J., and T. Kouzarides. 1996. The CBP co-activator is a histone acetyltransferase. *Nature*. 384:641-643.

Bannister, A.J., and T. Kouzarides. 2011. Regulation of chromatin by histone modifications. *Cell Res*. 21:381-395.

Barrera, L.A., A. Vedenko, J.V. Kurland, J.M. Rogers, S.S. Gisselbrecht, E.J. Rossin, J. Woodard, L. Mariani, K.H. Kock, S. Inukai, T. Siggers, L. Shokri, R. Gordan, N. Sahni, C. Cotsapas, T. Hao, S. Yi, M. Kellis, M.J. Daly, M. Vidal, D.E. Hill, and M.L. Bulyk. 2016. Survey of variation in human transcription factors reveals prevalent DNA binding changes. *Science*. 351:1450-1454.

Barski, A., S. Cuddapah, K. Cui, T.Y. Roh, D.E. Schones, Z. Wang, G. Wei, I. Chepelev, and K. Zhao. 2007. High-resolution profiling of histone methylations in the human genome. *Cell*. 129:823-837.

Bartman, C.R., S.C. Hsu, C.C. Hsiung, A. Raj, and G.A. Blobel. 2016. Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. *Mol Cell*. 62:237-247.

Basheer, F., G. Giotopoulos, E. Meduri, H. Yun, M. Mazan, D. Sasca, P. Gallipoli, L. Marando, M. Gozdecka, R. Asby, O. Sheppard, M. Dudek, L. Bullinger, H. Dohner, R. Dillon, S. Freeman, O. Ottmann, A. Burnett, N. Russell, E. Papaemmanuil, R. Hills, P. Campbell, G.S. Vassiliou, and B.J.P. Huntly. 2019. Contrasting requirements during disease evolution identify EZH2 as a therapeutic target in AML. *J Exp Med*. 216:966-981.

Bauer, M., J. Trupke, and L. Ringrose. 2016. The quest for mammalian Polycomb response elements: are we there yet? *Chromosoma*. 125:471-496.

Beguelin, W., R. Popovic, M. Teater, Y. Jiang, K.L. Bunting, M. Rosen, H. Shen, S.N. Yang, L. Wang, T. Ezponda, E. Martinez-Garcia, H. Zhang, Y. Zheng, S.K. Verma, M.T. McCabe, H.M. Ott, G.S. Van Aller, R.G. Kruger, Y. Liu, C.F. McHugh, D.W. Scott, Y.R. Chung, N. Kelleher, R. Shaknovich, C.L. Creasy, R.D. Gascoyne, K.K. Wong, L. Cerchietti, R.L. Levine, O. Abdel-Wahab, J.D. Licht, O. Elemento, and A.M. Melnick. 2013. EZH2 is required for germinal center formation and somatic EZH2 mutations promote lymphoid transformation. *Cancer Cell*. 23:677-692.

Beltran, M., C.M. Yates, L. Skalska, M. Dawson, F.P. Reis, K. Viiri, C.L. Fisher, C.R. Sibley, B.M. Foster, T. Bartke, J. Ule, and R.G. Jenner. 2016. The interaction of PRC2 with RNA or chromatin is mutually antagonistic. *Genome Res*. 26:896-907.

Ben-Ami, O., D. Friedman, D. Leshkowitz, D. Goldenberg, K. Orlovsky, N. Pencovich, J. Lotem, A. Tanay, and Y. Groner. 2013. Addiction of t(8;21) and inv(16) acute myeloid leukemia to native RUNX1. *Cell Rep*. 4:1131-1143.

Benabdallah, N.S., I. Williamson, R.S. Illingworth, L. Kane, S. Boyle, D. Sengupta, G.R. Grimes, P. Therizols, and W.A. Bickmore. 2019. Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. *Mol Cell*. 76:473-484 e477.

Benedikt, A., S. Baltruschat, B. Scholz, A. Bursen, T.N. Arrey, B. Meyer, L. Varagnolo, A.M. Muller, M. Karas, T. Dingermann, and R. Marschalek. 2011. The leukemogenic AF4-MLL fusion protein causes P-TEFb kinase activation and altered epigenetic signatures. *Leukemia*. 25:135-144.

Benoist, C., and P. Chambon. 1981. In vivo sequence requirements of the SV40 early promotor region. *Nature*. 290:304-310.

Berg, T., S. Thoene, D. Yap, T. Wee, N. Schoeler, P. Rosten, E. Lim, M. Bilenky, A.J. Mungall, T. Oellerich, S. Lee, C.K. Lai, P. Umlandt, A. Salmi, H. Chang, L. Yue, D. Lai, S.W. Cheng, R.D. Morin, M. Hirst, H. Serve, M.A. Marra, G.B. Morin, R.D. Gascoyne, S.A. Aparicio, and R.K. Humphries. 2014. A transgenic mouse model demonstrating the oncogenic role of mutations in the polycomb-group gene EZH2 in lymphomagenesis. *Blood*. 123:3914-3924.

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

Bernstein, E., E.M. Duncan, O. Masui, J. Gil, E. Heard, and C.D. Allis. 2006b. Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Mol Cell Biol*. 26:2560-2569.

Bernt, K.M., N. Zhu, A.U. Sinha, S. Vempati, J. Faber, A.V. Krivtsov, Z. Feng, N. Punt, A. Daigle, L. Bullinger, R.M. Pollock, V.M. Richon, A.L. Kung, and S.A. Armstrong. 2011. MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L. *Cancer Cell*. 20:66-78.

Bird, A.P. 1980. DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res*. 8:1499-1504.

Birke, M., S. Schreiner, M.P. Garcia-Cuellar, K. Mahr, F. Titgemeyer, and R.K. Slany. 2002. The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. *Nucleic Acids Res*. 30:958-965.

Biswas, D., T.A. Milne, V. Basrur, J. Kim, K.S. Elenitoba-Johnson, C.D. Allis, and R.G. Roeder. 2011. Function of leukemogenic mixed lineage leukemia 1 (MLL) fusion proteins through distinct partner protein complexes. *Proc Natl Acad Sci U S A*. 108:15751-15756.

Blackledge, N.P., A.M. Farcas, T. Kondo, H.W. King, J.F. McGouran, L.L. Hanssen, S. Ito, S. Cooper, K. Kondo, Y. Koseki, T. Ishikura, H.K. Long, T.W. Sheahan, N. Brockdorff, B.M. Kessler, H. Koseki, and R.J. Klose. 2014. Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell*. 157:1445-1459.

Blackledge, N.P., N.A. Fursova, J.R. Kelley, M.K. Huseyin, A. Feldmann, and R.J. Klose. 2020. PRC1 Catalytic Activity Is Central to Polycomb System Function. *Mol Cell*. 77:857-874 e859.

Bledau, A.S., K. Schmidt, K. Neumann, U. Hill, G. Ciotta, A. Gupta, D.C. Torres, J. Fu, A. Kranz, A.F. Stewart, and K. Anastassiadis. 2014. The H3K4 methyltransferase Setd1a is first required at the epiblast stage, whereas Setd1b becomes essential after gastrulation. *Development*. 141:1022-1035.

Blow, M.J., D.J. McCulley, Z. Li, T. Zhang, J.A. Akiyama, A. Holt, I. Plajzer-Frick, M. Shoukry, C. Wright, F. Chen, V. Afzal, J. Bristow, B. Ren, B.L. Black, E.M. Rubin, A. Visel, and L.A. Pennacchio. 2010. ChIP-Seq identification of weakly conserved heart enhancers. *Nat Genet*. 42:806-810.

Blumenthal, E., S. Greenblatt, G. Huang, K. Ando, Y. Xu, and S.D. Nimer. 2017. Covalent Modifications of RUNX Proteins: Structure Affects Function. *Adv Exp Med Biol*. 962:33-44.

Boija, A., I.A. Klein, B.R. Sabari, A. Dall'Agnese, E.L. Coffey, A.V. Zamudio, C.H. Li, K. Shrinivas, J.C. Manteiga, N.M. Hannett, B.J. Abraham, L.K. Afeyan, Y.E. Guo, J.K. Rimel, C.B. Fant, J. Schuijers, T.I. Lee, D.J. Taatjes, and R.A. Young. 2018. Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell*. 175:1842-1855 e1816.

Bolouri, H., J.E. Farrar, T. Triche, Jr., R.E. Ries, E.L. Lim, T.A. Alonzo, Y. Ma, R. Moore, A.J. Mungall, M.A. Marra, J. Zhang, X. Ma, Y. Liu, Y. Liu, J.M.G. Auvil, T.M. Davidsen, P. Gesuwan, L.C. Hermida, B. Salhia, S. Capone, G. Ramsingh, C.M. Zwaan, S. Noort, S.R. Piccolo, E.A. Kolb, A.S. Gamis, M.A. Smith, D.S. Gerhard, and S. Meshinchi. 2018. The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nat Med*. 24:103-112.

Bonifer, C., E. Levantini, V. Kouskoff, and G. Lacaud. 2017. Runx1 Structure and Function in Blood Cell Development. *Adv Exp Med Biol*. 962:65-81.

Booth, C.A.G., N. Barkas, W.H. Neo, H. Boukarabila, E.J. Soilleux, G. Giotopoulos, N. Farnoud, A. Giustacchini, N. Ashley, J. Carrelha, L. Jamieson, D. Atkinson, T. Bouriez-Jones, R.K. Prinjha, T.A. Milne, D.T. Teachey, E. Papaemmanuil, B.J.P. Huntly, S.E.W. Jacobsen, and A.J. Mead. 2018. Ezh2 and Runx1 Mutations Collaborate to Initiate Lympho-Myeloid Leukemia in Early Thymic Progenitors. *Cancer Cell*. 33:274-291 e278.

Boyer, L.A., K. Plath, J. Zeitlinger, T. Brambrink, L.A. Medeiros, T.I. Lee, S.S. Levine, M. Wernig, A. Tajonar, M.K. Ray, G.W. Bell, A.P. Otte, M. Vidal, D.K. Gifford, R.A. Young, and R. Jaenisch. 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature*. 441:349-353.

Brien, G.L., R.B. Bressan, C. Monger, D. Gannon, E. Lagan, A.M. Doherty, E. Healy, H. Neikes, D.J. Fitzpatrick, O. Deevy, V. Grant, M.A. Marques-Torrejon, N. Alfazema, S.M. Pollard, and A.P. Bracken. 2021. Simultaneous disruption of PRC2 and enhancer function underlies histone H3.3-K27M oncogenic activity in human hindbrain neural stem cells. *Nat Genet*. 53:1221-1232.

Brookes, E., I. de Santiago, D. Hebenstreit, K.J. Morris, T. Carroll, S.Q. Xie, J.K. Stock, M. Heidemann, D. Eick, N. Nozaki, H. Kimura, J. Ragoussis, S.A. Teichmann, and A. Pombo. 2012. Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. *Cell Stem Cell*. 10:157-170.

Brookes, E., and A. Pombo. 2009. Modifications of RNA polymerase II are pivotal in regulating gene expression states. *EMBO Rep*. 10:1213-1219.

Brown, J.L., C. Fritsch, J. Mueller, and J.A. Kassis. 2003. The Drosophila pho-like gene encodes a YY1-related DNA binding protein that is redundant with pleiohomeotic in homeotic gene silencing. *Development*. 130:285-294.

Brown, J.L., D. Mucci, M. Whiteley, M.L. Dirksen, and J.A. Kassis. 1998. The Drosophila Polycomb group gene pleiohomeotic encodes a DNA binding protein with homology to the transcription factor YY1. *Mol Cell*. 1:1057-1064.

Brownell, J.E., and C.D. Allis. 1996. Special HATs for special occasions: linking histone acetylation to chromatin assembly and gene activation. *Curr Opin Genet Dev*. 6:176-184.

Buenrostro, J.D., P.G. Giresi, L.C. Zaba, H.Y. Chang, and W.J. Greenleaf. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*. 10:1213-1218.

Burel, S.A., N. Harakawa, L. Zhou, T. Pabst, D.G. Tenen, and D.E. Zhang. 2001. Dichotomy of AML1-ETO functions: growth arrest versus block of differentiation. *Mol Cell Biol*. 21:5577-5590.

Caganova, M., C. Carrisi, G. Varano, F. Mainoldi, F. Zanardi, P.L. Germain, L. George, F. Alberghini, L. Ferrarini, A.K. Talukder, M. Ponzoni, G. Testa, T. Nojima, C. Doglioni, D. Kitamura, K.M. Toellner, I.H. Su, and S. Casola. 2013. Germinal center dysregulation by histone methyltransferase EZH2 promotes lymphomagenesis. *J Clin Invest*. 123:5009-5022.

Calo, E., and J. Wysocka. 2013. Modification of enhancer chromatin: what, how, and why? *Mol Cell*. 49:825-837.

Cancer Genome Atlas Research, N., T.J. Ley, C. Miller, L. Ding, B.J. Raphael, A.J. Mungall, A. Robertson, K. Hoadley, T.J. Triche, Jr., P.W. Laird, J.D. Baty, L.L. Fulton, R. Fulton, S.E. Heath, J. Kalicki-Veizer, C. Kandoth, J.M. Klco, D.C. Koboldt, K.L. Kanchi, S. Kulkarni, T.L. Lamprecht, D.E. Larson, L. Lin, C. Lu, M.D. McLellan, J.F. McMichael, J. Payton, H. Schmidt, D.H. Spencer, M.H. Tomasson, J.W. Wallis, L.D. Wartman, M.A. Watson, J. Welch, M.C. Wendl, A. Ally, M. Balasundaram, I. Birol, Y. Butterfield, R. Chiu, A. Chu, E. Chuah, H.J. Chun, R. Corbett, N. Dhalla, R. Guin, A. He, C. Hirst, M. Hirst, R.A. Holt, S. Jones, A. Karsan, D. Lee, H.I. Li, M.A. Marra, M. Mayo, R.A. Moore, K. Mungall, J. Parker, E. Pleasance, P. Plettner, J. Schein, D. Stoll, L. Swanson, A. Tam, N. Thiessen, R. Varhol, N. Wye, Y. Zhao, S. Gabriel, G. Getz, C. Sougnez, L. Zou, M.D. Leiserson, F. Vandin, H.T. Wu, F. Applebaum, S.B. Baylin, R. Akbani, B.M. Broom, K. Chen, T.C. Motter, K. Nguyen, J.N. Weinstein, N. Zhang, M.L. Ferguson, C. Adams, A. Black, J. Bowen, J. Gastier-Foster, T. Grossman, T. Lichtenberg, L. Wise, T. Davidsen, J.A. Demchok, K.R. Shaw, M. Sheth, H.J. Sofia, L. Yang, J.R. Downing, et al. 2013. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med*. 368:2059-2074.

Cao, R., L. Wang, H. Wang, L. Xia, H. Erdjument-Bromage, P. Tempst, R.S. Jones, and Y. Zhang. 2002. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science*. 298:1039-1043.

Carrozza, M.J., B. Li, L. Florens, T. Suganuma, S.K. Swanson, K.K. Lee, W.J. Shia, S. Anderson, J. Yates, M.P. Washburn, and J.L. Workman. 2005. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell*. 123:581-592.

Casares, F., and E. Sanchez-Herrero. 1995. Regulation of the infraabdominal regions of the bithorax complex of Drosophila by gap genes. *Development*. 121:1855-1866.

Cenik, B.K., and A. Shilatifard. 2021. COMPASS and SWI/SNF complexes in development and disease. *Nat Rev Genet*. 22:38-58.

Cerase, A., D. Smeets, Y.A. Tang, M. Gdula, F. Kraus, M. Spivakov, B. Moindrot, M. Leleu, A. Tattermusch, J. Demmerle, T.B. Nesterova, C. Green, A.P. Otte, L. Schermelleh, and N. Brockdorff. 2014. Spatial separation of Xist RNA and polycomb proteins revealed by superresolution microscopy. *Proc Natl Acad Sci U S A*. 111:2235-2240.

Chamberlain, S.J., D. Yee, and T. Magnuson. 2008. Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. *Stem Cells*. 26:1496-1505.

Chan, C.S., L. Rastelli, and V. Pirrotta. 1994. A Polycomb response element in the Ubx gene that determines an epigenetically inherited state of repression. *EMBO J*. 13:2553-2564.

Chen, B., J. Teng, H. Liu, X. Pan, Y. Zhou, S. Huang, M. Lai, G. Bian, B. Mao, W. Sun, Q. Zhou, S. Yang, T. Nakahata, and F. Ma. 2017. Inducible overexpression of RUNX1b/c in human embryonic stem cells blocks early hematopoiesis from mesoderm. *J Mol Cell Biol*. 9:262-273.

Chen, J., Z. Zhang, L. Li, B.C. Chen, A. Revyakin, B. Hajj, W. Legant, M. Dahan, T. Lionnet, E. Betzig, R. Tjian, and Z. Liu. 2014a. Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell*. 156:1274-1285.

Chen, M.J., T. Yokomizo, B.M. Zeigler, E. Dzierzak, and N.A. Speck. 2009. Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter. *Nature*. 457:887-891.

Chen, S., J. Chen, Q. Zhan, Y. Zhu, H. Chen, X. Deng, Z. Hou, B. Shen, Y. Chen, and C. Peng. 2014b. H2AK119Ub1 and H3K27Me3 in molecular staging for survival prediction of patients with pancreatic ductal adenocarcinoma. *Oncotarget*. 5:10421-10433.

Chen, S., L. Jiao, M. Shubbar, X. Yang, and X. Liu. 2018. Unique Structural Platforms of Suz12 Dictate Distinct Classes of PRC2 for Chromatin Binding. *Mol Cell*. 69:840-852 e845.

Chiang, A., M.B. O'Connor, R. Paro, J. Simon, and W. Bender. 1995. Discrete Polycomb-binding sites in each parasegmental domain of the bithorax complex. *Development*. 121:1681-1689.

Cho, W.K., J.H. Spille, M. Hecht, C. Lee, C. Li, V. Grube, and Cisse, II. 2018. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science*. 361:412-415.

Cho, Y.W., T. Hong, S. Hong, H. Guo, H. Yu, D. Kim, T. Guszczynski, G.R. Dressler, T.D. Copeland, M. Kalkum, and K. Ge. 2007. PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. *J Biol Chem*. 282:20395-20406.

Chory, E.J., J.P. Calarco, N.A. Hathaway, O. Bell, D.S. Neel, and G.R. Crabtree. 2019. Nucleosome Turnover Regulates Histone Methylation Patterns over the Genome. *Mol Cell*. 73:61-72 e63.

Chou, F.S., A. Griesinger, M. Wunderlich, S. Lin, K.A. Link, M. Shrestha, S. Goyama, B. Mizukawa, S. Shen, G. Marcucci, and J.C. Mulloy. 2012. The thrombopoietin/MPL/Bcl-xL pathway is essential for survival and self-renewal in human preleukemia induced by AML1-ETO. *Blood*. 120:709-719.

Christen, F., K. Hoyer, K. Yoshida, H.A. Hou, N. Waldhueter, M. Heuser, R.K. Hills, W. Chan, R. Hablesreiter, O. Blau, Y. Ochi, P. Klement, W.C. Chou, I.W. Blau, J.L. Tang, T. Zemojtel, Y. Shiraishi, Y. Shiozawa, F. Thol, A. Ganser, B. Lowenberg, D.C. Linch, L. Bullinger, P.J.M. Valk, H.F. Tien, R.E. Gale, S. Ogawa, and F. Damm. 2019. Genomic landscape and clonal evolution of acute myeloid leukemia with t(8;21): an international study on 331 patients. *Blood*. 133:1140-1151.

Christiansen, D.H., M.K. Andersen, and J. Pedersen-Bjergaard. 2004. Mutations of AML1 are common in therapy-related myelodysplasia following therapy with alkylating agents and

are significantly associated with deletion or loss of chromosome arm 7q and with subsequent leukemic transformation. *Blood*. 104:1474-1481.

Ciferri, C., G.C. Lander, A. Maiolica, F. Herzog, R. Aebersold, and E. Nogales. 2012. Molecular architecture of human polycomb repressive complex 2. *Elife*. 1:e00005.

Cirillo, L.A., F.R. Lin, I. Cuesta, D. Friedman, M. Jarnik, and K.S. Zaret. 2002. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell*. 9:279-289.

Cirillo, L.A., C.E. McPherson, P. Bossard, K. Stevens, S. Cherian, E.Y. Shim, K.L. Clark, S.K. Burley, and K.S. Zaret. 1998. Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. *EMBO J*. 17:244-254.

Clauss, K., A.P. Popp, L. Schulze, J. Hettich, M. Reisser, L. Escoter Torres, N.H. Uhlenhaut, and J.C.M. Gebhardt. 2017. DNA residence time is a regulatory factor of transcription repression. *Nucleic Acids Res*. 45:11121-11130.

Claussnitzer, M., S.N. Dankel, K.H. Kim, G. Quon, W. Meuleman, C. Haugen, V. Glunk, I.S. Sousa, J.L. Beaudry, V. Puviindran, N.A. Abdennur, J. Liu, P.A. Svensson, Y.H. Hsu, D.J. Drucker, G. Mellgren, C.C. Hui, H. Hauner, and M. Kellis. 2015. FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *N Engl J Med*. 373:895-907.

Cohen, I., D. Zhao, C. Bar, V.J. Valdes, K.L. Dauber-Decker, M.B. Nguyen, M. Nakayama, M. Rendl, W.A. Bickmore, H. Koseki, D. Zheng, and E. Ezhkova. 2018. PRC1 Fine-tunes Gene Repression and Activation to Safeguard Skin Development and Stem Cell Specification. *Cell Stem Cell*. 22:726-739 e727.

Cohen, I., D. Zhao, G. Menon, M. Nakayama, H. Koseki, D. Zheng, and E. Ezhkova. 2019. PRC1 preserves epidermal tissue integrity independently of PRC2. *Genes Dev*. 33:55-60.

Collett, K., G.E. Eide, J. Arnes, I.M. Stefansson, J. Eide, A. Braaten, T. Aas, A.P. Otte, and L.A. Akslen. 2006. Expression of enhancer of zeste homologue 2 is significantly associated with increased tumor cell proliferation and is a marker of aggressive breast cancer. *Clin Cancer Res*. 12:1168-1174.

Conway, E., F. Rossi, D. Fernandez-Perez, E. Ponzo, K.J. Ferrari, M. Zanotti, D. Manganaro, S. Rodighiero, S. Tamburri, and D. Pasini. 2021. BAP1 enhances Polycomb repression by counteracting widespread H2AK119ub1 deposition and chromatin condensation. *Mol Cell*. 81:3526-3541 e3528.

Cooper, S., M. Dienstbier, R. Hassan, L. Schermelleh, J. Sharif, N.P. Blackledge, V. De Marco, S. Elderkin, H. Koseki, R. Klose, A. Heger, and N. Brockdorff. 2014. Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment. *Cell Rep*. 7:1456-1470.

Cooper, S., A. Grijzenhout, E. Underwood, K. Ancelin, T. Zhang, T.B. Nesterova, B. Anil-Kirmizitas, A. Bassett, S.M. Kooistra, K. Agger, K. Helin, E. Heard, and N. Brockdorff. 2016. Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2. *Nat Commun*. 7:13661.

Cordonnier, G., A. Mandoli, N. Cagnard, G. Hypolite, L. Lhermitte, E. Verhoeyen, V. Asnafi, N. Dillon, E. Macintyre, J.H.A. Martens, and J. Bond. 2020. CBFbeta-SMMHC Affects Genome-wide Polycomb Repressive Complex 1 Activity in Acute Myeloid Leukemia. *Cell Rep*. 30:299-307 e293.

Core, N., S. Bel, S.J. Gaunt, M. Aurrand-Lions, J. Pearce, A. Fisher, and M. Djabali. 1997. Altered cellular proliferation and mesoderm patterning in Polycomb-M33-deficient mice. *Development*. 124:721-729.

Cortini, R., and G.J. Filion. 2018. Theoretical principles of transcription factor traffic on folded chromatin. *Nat Commun*. 9:1740.

Cremer, T., and M. Cremer. 2010. Chromosome territories. *Cold Spring Harb Perspect Biol*. 2:a003889.

Creyghton, M.P., A.W. Cheng, G.G. Welstead, T. Kooistra, B.W. Carey, E.J. Steine, J. Hanna, M.A. Lodato, G.M. Frampton, P.A. Sharp, L.A. Boyer, R.A. Young, and R. Jaenisch. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A*. 107:21931-21936.

Croft, J.A., J.M. Bridger, S. Boyle, P. Perry, P. Teague, and W.A. Bickmore. 1999. Differences in the localization and morphology of chromosomes in the human nucleus. *J Cell Biol*. 145:1119-1131.

Crump, N.T., E. Ballabio, L. Godfrey, R. Thorne, E. Repapi, J. Kerry, M. Tapia, P. Hua, C. Lagerholm, P. Filippakopoulos, J.O.J. Davies, and T.A. Milne. 2021. BET inhibition disrupts transcription but retains enhancer-promoter contact. *Nat Commun*. 12:223.

Crump, N.T., and T.A. Milne. 2019. Why are so many MLL lysine methyltransferases required for normal mammalian development? *Cell Mol Life Sci*. 76:2885-2898.

Cruz-Molina, S., P. Respuela, C. Tebartz, P. Kolovos, M. Nikolic, R. Fueyo, W.F.J. van Ijcken, F. Grosveld, P. Frommolt, H. Bazzi, and A. Rada-Iglesias. 2017. PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation. *Cell Stem Cell*. 20:689-705 e689.

Cusack, M., H.W. King, P. Spingardi, B.M. Kessler, R.J. Klose, and S. Kriaucionis. 2020. Distinct contributions of DNA methylation and histone acetylation to the genomic occupancy of transcription factors. *Genome Res*. 30:1393-1406.

Cutter, A.R., and J.J. Hayes. 2015. A brief review of nucleosome structure. *FEBS Lett*. 589:2914-2922.

Danis, E., T. Yamauchi, K. Echanique, X. Zhang, J.N. Haladyna, S.S. Riedel, N. Zhu, H. Xie, S.H. Orkin, S.A. Armstrong, K.M. Bernt, and T. Neff. 2016. Ezh2 Controls an Early Hematopoietic Program and Growth and Survival Signaling in Early T Cell Precursor Acute Lymphoblastic Leukemia. *Cell Rep*. 14:1953-1965.

Dantuma, N.P., T.A. Groothuis, F.A. Salomons, and J. Neefjes. 2006. A dynamic ubiquitin equilibrium couples proteasomal activity to chromatin remodeling. *J Cell Biol*. 173:19-26.

Dao, L.T.M., A.O. Galindo-Albarran, J.A. Castro-Mondragon, C. Andrieu-Soler, A. Medina-Rivera, C. Souaid, G. Charbonnier, A. Griffon, L. Vanhille, T. Stephen, J. Alomairi, D. Martin, M. Torres, N. Fernandez, E. Soler, J. van Helden, D. Puthier, and S. Spicuglia. 2017. Genome-wide characterization of mammalian promoters with distal enhancer functions. *Nat Genet*. 49:1073-1081.

Dawson, M.A., R.K. Prinjha, A. Dittmann, G. Giotopoulos, M. Bantscheff, W.I. Chan, S.C. Robson, C.W. Chung, C. Hopf, M.M. Savitski, C. Huthmacher, E. Gudgin, D. Lugo, S. Beinke, T.D. Chapman, E.J. Roberts, P.E. Soden, K.R. Auger, O. Mirguet, K. Doehner, R. Delwel, A.K. Burnett, P. Jeffrey, G. Drewes, K. Lee, B.J. Huntly, and T. Kouzarides. 2011. Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. *Nature*. 478:529-533.

De Santa, F., M.G. Totaro, E. Prosperini, S. Notarbartolo, G. Testa, and G. Natoli. 2007. The histone H3 lysine-27 demethylase Jmjd3 links inflammation to inhibition of polycomb-mediated gene silencing. *Cell*. 130:1083-1094.

Deaton, A.M., and A. Bird. 2011. CpG islands and the regulation of transcription. *Genes Dev*. 25:1010-1022.

Dellino, G.I., Y.B. Schwartz, G. Farkas, D. McCabe, S.C. Elgin, and V. Pirrotta. 2004. Polycomb silencing blocks transcription initiation. *Mol Cell*. 13:887-893.

Delvecchio, M., J. Gaucher, C. Aguilar-Gurrieri, E. Ortega, and D. Panne. 2013. Structure of the p300 catalytic core and implications for chromatin targeting and HAT regulation. *Nat Struct Mol Biol*. 20:1040-1046.

Deng, W., J. Lee, H. Wang, J. Miller, A. Reik, P.D. Gregory, A. Dean, and G.A. Blobel. 2012. Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell*. 149:1233-1244.

Deng, W., J.W. Rupon, I. Krivega, L. Breda, I. Motta, K.S. Jahn, A. Reik, P.D. Gregory, S. Rivella, A. Dean, and G.A. Blobel. 2014. Reactivation of developmentally silenced globin genes by forced chromatin looping. *Cell*. 158:849-860.

Denholtz, M., G. Bonora, C. Chronis, E. Splinter, W. de Laat, J. Ernst, M. Pellegrini, and K. Plath. 2013. Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. *Cell Stem Cell*. 13:602-616.

Denissov, S., H. Hofemeister, H. Marks, A. Kranz, G. Ciotta, S. Singh, K. Anastassiadis, H.G. Stunnenberg, and A.F. Stewart. 2014. Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. *Development*. 141:526-537.

Dhalluin, C., J.E. Carlson, L. Zeng, C. He, A.K. Aggarwal, and M.M. Zhou. 1999. Structure and ligand of a histone acetyltransferase bromodomain. *Nature*. 399:491-496.

Dixon, J.R., S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J.S. Liu, and B. Ren. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 485:376-380.

Dobin, A., C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T.R. Gingeras. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 29:15-21.

Dobrinic, P., A.T. Szczurek, and R.J. Klose. 2021. PRC1 drives Polycomb-mediated gene repression by controlling transcription initiation and burst frequency. *Nat Struct Mol Biol*.

Dorighi, K.M., T. Swigut, T. Henriques, N.V. Bhanu, B.S. Scruggs, N. Nady, C.D. Still, 2nd, B.A. Garcia, K. Adelman, and J. Wysocka. 2017. Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. *Mol Cell*. 66:568-576 e564.

Drissen, R., R.J. Palstra, N. Gillemans, E. Splinter, F. Grosveld, S. Philipsen, and W. de Laat. 2004. The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes Dev*. 18:2485-2490.

Duncan, I., and G. Montgomery. 2002. E. B. Lewis and the bithorax complex: part II. From cis-trans test to the genetic control of development. *Genetics*. 161:1-10.

Duployez, N., A. Marceau-Renaut, N. Boissel, A. Petit, M. Bucci, S. Geffroy, H. Lapillonne, A. Renneville, C. Ragu, M. Figeac, K. Celli-Lebras, C. Lacombe, J.B. Micol, O. Abdel-Wahab, P. Cornillet, N. Ifrah, H. Dombret, G. Leverger, E. Jourdan, and C. Preudhomme. 2016. Comprehensive mutational profiling of core binding factor acute myeloid leukemia. *Blood*. 127:2451-2459.

Eagen, K.P., E.L. Aiden, and R.D. Kornberg. 2017. Polycomb-mediated chromatin loops revealed by a subkilobase-resolution chromatin interaction map. *Proc Natl Acad Sci U S A*. 114:8764-8769.

Eisfeld, A.K., J. Kohlschmidt, S. Schwind, D. Nicolet, J.S. Blachly, S. Orwick, C. Shah, M. Bainazar, K.W. Kroll, C.J. Walker, A.J. Carroll, B.L. Powell, R.M. Stone, J.E. Kolitz, M.R. Baer, A. de la Chapelle, K. Mrozek, J.C. Byrd, and C.D. Bloomfield. 2017. Mutations in the CCND1 and CCND2 genes are frequent events in adult patients with t(8;21)(q22;q22) acute myeloid leukemia. *Leukemia*. 31:1278-1285.

El Khattabi, L., H. Zhao, J. Kalchschmidt, N. Young, S. Jung, P. Van Blerkom, P. Kieffer-Kwon, K.R. Kieffer-Kwon, S. Park, X. Wang, J. Krebs, S. Tripathi, N. Sakabe, D.R. Sobreira, S.C. Huang, S.S.P. Rao, N. Pruett, D. Chauss, E. Sadler, A. Lopez, M.A. Nobrega, E.L. Aiden, F.J. Asturias, and R. Casellas. 2019. A Pliable Mediator Acts as a Functional Rather Than an Architectural Bridge between Promoters and Enhancers. *Cell*. 178:1145-1158 e1120.

Endoh, M., T.A. Endo, J. Shinga, K. Hayashi, A. Farcas, K.W. Ma, S. Ito, J. Sharif, T. Endoh, N. Onaga, M. Nakayama, T. Ishikura, O. Masui, B.M. Kessler, T. Suda, O. Ohara, A. Okuda, R. Klose, and H. Koseki. 2017. PCGF6-PRC1 suppresses premature differentiation of mouse embryonic stem cells by regulating germ cell-related genes. *Elife*. 6.

Erhardt, S., I.H. Su, R. Schneider, S. Barton, A.J. Bannister, L. Perez-Burgos, T. Jenuwein, T. Kouzarides, A. Tarakhovsky, and M.A. Surani. 2003. Consequences of the depletion of zygotic and embryonic enhancer of zeste 2 during preimplantation mouse development. *Development*. 130:4235-4248.

Ernst, J., and M. Kellis. 2010. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol*. 28:817-825.

Ernst, T., A.J. Chase, J. Score, C.E. Hidalgo-Curtis, C. Bryant, A.V. Jones, K. Waghorn, K. Zoi, F.M. Ross, A. Reiter, A. Hochhaus, H.G. Drexler, A. Duncombe, F. Cervantes, D. Oscier, J. Boultwood, F.H. Grand, and N.C. Cross. 2010. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat Genet*. 42:722-726.

Eskeland, R., M. Leeb, G.R. Grimes, C. Kress, S. Boyle, D. Sproul, N. Gilbert, Y. Fan, A.I. Skoultchi, A. Wutz, and W.A. Bickmore. 2010. Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol Cell*. 38:452-464.

Faber, Z.J., X. Chen, A.L. Gedman, K. Boggs, J. Cheng, J. Ma, I. Radtke, J.R. Chao, M.P. Walsh, G. Song, A.K. Andersson, J. Dang, L. Dong, Y. Liu, R. Huether, Z. Cai, H. Mulder, G. Wu, M. Edmonson, M. Rusch, C. Qu, Y. Li, B. Vadodaria, J. Wang, E. Hedlund, X. Cao, D. Yergeau, J. Nakitandwe, S.B. Pounds, S. Shurtleff, R.S. Fulton, L.L. Fulton, J. Easton, E. Parganas, C.H. Pui, J.E. Rubnitz, L. Ding, E.R. Mardis, R.K. Wilson, T.A. Gruber, C.G. Mullighan, R.F. Schlenk, P. Paschka, K. Dohner, H. Dohner, L. Bullinger, J. Zhang, J.M. Klco, and J.R. Downing. 2016. The genomic landscape of core-binding factor acute myeloid leukemias. *Nat Genet*. 48:1551-1556.

Fan, H., Y. Guo, Y.H. Tsai, A.J. Storey, A. Kim, W. Gong, R.D. Edmondson, S.G. Mackintosh, H. Li, S.D. Byrum, A.J. Tackett, L. Cai, and G.G. Wang. 2021. A conserved BAH module within mammalian BAHD1 connects H3K27me3 to Polycomb gene silencing. *Nucleic Acids Res*. 49:4441-4455.

Fan, H., J. Lu, Y. Guo, D. Li, Z.M. Zhang, Y.H. Tsai, W.C. Pi, J.H. Ahn, W. Gong, Y. Xiang, D.F. Allison, H. Geng, S. He, Y. Diao, W.Y. Chen, B.D. Strahl, L. Cai, J. Song, and G.G. Wang. 2020. BAHCC1 binds H3K27me3 via a conserved BAH module to mediate gene silencing and oncogenesis. *Nat Genet*. 52:1384-1396.

Farcas, A.M., N.P. Blackledge, I. Sudbery, H.K. Long, J.F. McGouran, N.R. Rose, S. Lee, D. Sims, A. Cerase, T.W. Sheahan, H. Koseki, N. Brockdorff, C.P. Ponting, B.M. Kessler, and R.J. Klose. 2012. KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *Elife*. 1:e00205.

Faust, C., A. Schumacher, B. Holdener, and T. Magnuson. 1995. The eed mutation disrupts anterior mesoderm production in mice. *Development*. 121:273-285.

Ferrari, K.J., A. Scelfo, S. Jammula, A. Cuomo, I. Barozzi, A. Stutzer, W. Fischle, T. Bonaldi, and D. Pasini. 2014. Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. *Mol Cell*. 53:49-62.

Ficz, G., R. Heintzmann, and D.J. Arndt-Jovin. 2005. Polycomb group protein complexes exchange rapidly in living Drosophila. *Development*. 132:3963-3976.

Finogenova, K., J. Bonnet, S. Poepsel, I.B. Schafer, K. Finkl, K. Schmid, C. Litz, M. Strauss, C. Benda, and J. Muller. 2020. Structural basis for PRC2 decoding of active histone methylation marks H3K36me2/3. *Elife*. 9.

Fischle, W., Y. Wang, S.A. Jacobs, Y. Kim, C.D. Allis, and S. Khorasanizadeh. 2003. Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev*. 17:1870-1881.

Fitzgerald, D.P., and W. Bender. 2001. Polycomb group repression reduces DNA accessibility. *Mol Cell Biol*. 21:6585-6597.

Flavahan, W.A., Y. Drier, B.B. Liau, S.M. Gillespie, A.S. Venteicher, A.O. Stemmer-Rachamimov, M.L. Suva, and B.E. Bernstein. 2016. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature*. 529:110-114.

Forster, V.J., M.H. Nahari, N. Martinez-Soria, A.K. Bradburn, A. Ptasinska, S.A. Assi, S.E. Fordham, H. McNeil, C. Bonifer, O. Heidenreich, and J.M. Allan. 2016. The leukemia-associated RUNX1/ETO oncoprotein confers a mutator phenotype. *Leukemia*. 30:250-253.

Fournier, M., G. Bourriquen, F.C. Lamaze, M.C. Cote, E. Fournier, C. Joly-Beauparlant, V. Caron, S. Gobeil, A. Droit, and S. Bilodeau. 2016. FOXA and master transcription factors recruit Mediator and Cohesin to the core transcriptional regulatory circuitry of cancer cells. *Sci Rep*. 6:34962.

Francis, N.J., R.E. Kingston, and C.L. Woodcock. 2004. Chromatin compaction by a polycomb group protein complex. *Science*. 306:1574-1577.

Frangini, A., M. Sjoberg, M. Roman-Trufero, G. Dharmalingam, V. Haberle, T. Bartke, B. Lenhard, M. Malumbres, M. Vidal, and N. Dillon. 2013. The aurora B kinase and the polycomb protein ring1B combine to regulate active promoters in quiescent lymphocytes. *Mol Cell*. 51:647-661.

Frankel, A.D., and P.S. Kim. 1991. Modular structure of transcription factors: implications for gene regulation. *Cell*. 65:717-719.

Fritsch, C., J.L. Brown, J.A. Kassis, and J. Muller. 1999. The DNA-binding polycomb group protein pleiohomeotic mediates silencing of a Drosophila homeotic gene. *Development*. 126:3905-3913.

Fujisawa, T., and P. Filippakopoulos. 2017. Functions of bromodomain-containing proteins and their roles in homeostasis and cancer. *Nat Rev Mol Cell Biol*. 18:246-262.

Fursova, N.A., N.P. Blackledge, M. Nakayama, S. Ito, Y. Koseki, A.M. Farcas, H.W. King, H. Koseki, and R.J. Klose. 2019. Synergy between Variant PRC1 Complexes Defines Polycomb-Mediated Gene Repression. *Mol Cell*. 74:1020-1036 e1028.

Fursova, N.A., A.H. Turberfield, N.P. Blackledge, E.L. Findlater, A. Lastuvkova, M.K. Huseyin, P. Dobrinic, and R.J. Klose. 2021. BAP1 constrains pervasive H2AK119ub1 to control the transcriptional potential of the genome. *Genes Dev*. 35:749-770.

Gaidzik, V.I., L. Bullinger, R.F. Schlenk, A.S. Zimmermann, J. Rock, P. Paschka, A. Corbacioglu, J. Krauter, B. Schlegelberger, A. Ganser, D. Spath, A. Kundgen, I.G. Schmidt-Wolf, K. Gotze, D. Nachbaur, M. Pfreundschuh, H.A. Horst, H. Dohner, and K. Dohner. 2011. RUNX1 mutations in acute myeloid leukemia: results from a comprehensive genetic and clinical analysis from the AML study group. *J Clin Oncol*. 29:1364-1372.

Gaidzik, V.I., V. Teleanu, E. Papaemmanuil, D. Weber, P. Paschka, J. Hahn, T. Wallrabenstein, B. Kolbinger, C.H. Kohne, H.A. Horst, P. Brossart, G. Held, A. Kundgen, M. Ringhoffer, K. Gotze, M. Rummel, M. Gerstung, P. Campbell, J.M. Kraus, H.A. Kestler, F. Thol, M. Heuser, B. Schlegelberger, A. Ganser, L. Bullinger, R.F. Schlenk, K. Dohner, and H. Dohner. 2016. RUNX1 mutations in acute myeloid leukemia are associated with distinct clinico-pathologic and genetic features. *Leukemia*. 30:2160-2168.

Gao, Z., P. Lee, J.M. Stafford, M. von Schimmelmann, A. Schaefer, and D. Reinberg. 2014. An AUTS2-Polycomb complex activates gene expression in the CNS. *Nature*. 516:349-354.

Gao, Z., J. Zhang, R. Bonasio, F. Strino, A. Sawai, F. Parisi, Y. Kluger, and D. Reinberg. 2012. PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Mol Cell*. 45:344-356.

Garcia, E., C. Marcos-Gutierrez, M. del Mar Lorente, J.C. Moreno, and M. Vidal. 1999. RYBP, a new repressor protein that interacts with components of the mammalian Polycomb complex, and with the transcription factor YY1. *EMBO J*. 18:3404-3418.

Gaucher, J., F. Boussouar, E. Montellier, S. Curtet, T. Buchou, S. Bertrand, P. Hery, S. Jounier, A. Depaux, A.L. Vitte, P. Guardiola, K. Pernet, A. Debernardi, F. Lopez, H. Holota, J. Imbert, D.J. Wolgemuth, M. Gerard, S. Rousseaux, and S. Khochbin. 2012. Bromodomain-dependent stage-specific male genome programming by Brdt. *EMBO J*. 31:3809-3820.

Gelmetti, V., J. Zhang, M. Fanelli, S. Minucci, P.G. Pelicci, and M.A. Lazar. 1998. Aberrant recruitment of the nuclear receptor corepressor-histone deacetylase complex by the acute myeloid leukemia fusion partner ETO. *Mol Cell Biol*. 18:7185-7191.

Gentile, C., S. Berlivet, A. Mayran, D. Paquette, F. Guerard-Millet, E. Bajon, J. Dostie, and M. Kmita. 2019. PRC2-Associated Chromatin Contacts in the Developing Limb Reveal a Possible Mechanism for the Atypical Role of PRC2 in HoxA Gene Expression. *Dev Cell*. 50:184-196 e184.

Gergen, J.P., and E.F. Wieschaus. 1985. The localized requirements for a gene affecting segmentation in Drosophila: analysis of larvae mosaic for runt. *Dev Biol*. 109:321-335.

Gilmour, J., S.A. Assi, L. Noailles, M. Lichtinger, N. Obier, and C. Bonifer. 2018. The Co-operation of RUNX1 with LDB1, CDK9 and BRD4 Drives Transcription Factor Complex Relocation During Haematopoietic Specification. *Sci Rep*. 8:10410.

Glancy, E., C. Ciferri, and A.P. Bracken. 2021. Structural basis for PRC2 engagement with chromatin. *Curr Opin Struct Biol*. 67:135-144.

Glaser, S., J. Schaft, S. Lubitz, K. Vintersten, F. van der Hoeven, K.R. Tufteland, R. Aasland, K. Anastassiadis, S.L. Ang, and A.F. Stewart. 2006. Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. *Development*. 133:1423-1432.

Gonzales-Cope, M., S. Sidoli, N.V. Bhanu, K.J. Won, and B.A. Garcia. 2016. Histone H4 acetylation and the epigenetic reader Brd4 are critical regulators of pluripotency in embryonic stem cells. *BMC Genomics*. 17:95.

Goyama, S., and J.C. Mulloy. 2011. Molecular pathogenesis of core binding factor leukemia: current knowledge and future prospects. *Int J Hematol*. 94:126-133.

Goyama, S., J. Schibler, L. Cunningham, Y. Zhang, Y. Rao, N. Nishimoto, M. Nakagawa, A. Olsson, M. Wunderlich, K.A. Link, B. Mizukawa, H.L. Grimes, M. Kurokawa, P.P. Liu, G. Huang, and J.C. Mulloy. 2013. Transcription factor RUNX1 promotes survival of acute myeloid leukemia cells. *J Clin Invest*. 123:3876-3888.

Grau, D.J., B.A. Chapman, J.D. Garlick, M. Borowsky, N.J. Francis, and R.E. Kingston. 2011. Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. *Genes Dev*. 25:2210-2221.

Greif, P.A., N.P. Konstandin, K.H. Metzeler, T. Herold, Z. Pasalic, B. Ksienzyk, A. Dufour, F. Schneider, S. Schneider, P.M. Kakadia, J. Braess, M.C. Sauerland, W.E. Berdel, T. Buchner, B.J. Woermann, W. Hiddemann, K. Spiekermann, and S.K. Bohlander. 2012. RUNX1 mutations in cytogenetically normal acute myeloid leukemia are associated with a poor prognosis and up-regulation of lymphoid genes. *Haematologica*. 97:1909-1915.

Grembecka, J., A.M. Belcher, T. Hartley, and T. Cierpicki. 2010. Molecular basis of the mixed lineage leukemia-menin interaction: implications for targeting mixed lineage leukemias. *J Biol Chem*. 285:40690-40698.

Gross, D.S., and W.T. Garrard. 1988. Nuclease hypersensitive sites in chromatin. *Annu Rev Biochem*. 57:159-197.

Growney, J.D., H. Shigematsu, Z. Li, B.H. Lee, J. Adelsperger, R. Rowan, D.P. Curley, J.L. Kutok, K. Akashi, I.R. Williams, N.A. Speck, and D.G. Gilliland. 2005. Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype. *Blood*. 106:494-504.

Gu, Y., T. Nakamura, H. Alder, R. Prasad, O. Canaani, G. Cimino, C.M. Croce, and E. Canaani. 1992. The t(4;11) chromosome translocation of human acute leukemias fuses the ALL-1 gene, related to Drosophila trithorax, to the AF-4 gene. *Cell*. 71:701-708.

Guertin, M.J., and J.T. Lis. 2010. Chromatin landscape dictates HSF binding to target DNA elements. *PLoS Genet*. 6:e1001114.

Guertin, M.J., and J.T. Lis. 2013. Mechanisms by which transcription factors gain access to target sequence elements in chromatin. *Curr Opin Genet Dev*. 23:116-123.

Hager, G.L., J.G. McNally, and T. Misteli. 2009. Transcription dynamics. *Mol Cell*. 35:741-753.

Hallson, G., R.E. Hollebakken, T. Li, M. Syrzycka, I. Kim, S. Cotsworth, K.A. Fitzpatrick, D.A. Sinclair, and B.M. Honda. 2012. dSet1 is the main H3K4 di- and tri-methyltransferase throughout Drosophila development. *Genetics*. 190:91-100.

Hanson, R.D., J.L. Hess, B.D. Yu, P. Ernst, M. van Lohuizen, A. Berns, N.M. van der Lugt, C.S. Shashikant, F.H. Ruddle, M. Seto, and S.J. Korsmeyer. 1999. Mammalian Trithorax and polycomb-group homologues are antagonistic regulators of homeotic development. *Proc Natl Acad Sci U S A*. 96:14372-14377.

Harada, H., Y. Harada, H. Niimi, T. Kyo, A. Kimura, and T. Inaba. 2004. High incidence of somatic mutations in the AML1/RUNX1 gene in myelodysplastic syndrome and low blast percentage myeloid leukemia with myelodysplasia. *Blood*. 103:2316-2324.

Harding, K., and M. Levine. 1988. Gap genes define the limits of antennapedia and bithorax gene expression during early development in Drosophila. *EMBO J*. 7:205-214.

Harman, J.R., R. Thorne, M. Jamilly, M. Tapia, N.T. Crump, S. Rice, R. Beveridge, E. Morrissey, M. de Bruijn, I. Roberts, A. Roy, T.A. Fulga, and T.A. Milne. 2021. A KMT2A-AFF1 gene regulatory network highlights the role of core transcription factors and reveals the regulatory logic of key downstream target genes. *Genome Res*.

He, J., L. Shen, M. Wan, O. Taranova, H. Wu, and Y. Zhang. 2013. Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. *Nat Cell Biol*. 15:373-384.

Healy, E., M. Mucha, E. Glancy, D.J. Fitzpatrick, E. Conway, H.K. Neikes, C. Monger, G. Van Mierlo, M.P. Baltissen, Y. Koseki, M. Vermeulen, H. Koseki, and A.P. Bracken. 2019. PRC2.1 and PRC2.2 Synergize to Coordinate H3K27 Trimethylation. *Mol Cell*. 76:437-452 e436.

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

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

Herglotz, J., O.N. Kuvardina, S. Kolodziej, A. Kumar, H. Hussong, M. Grez, and J. Lausen. 2013. Histone arginine methylation keeps RUNX1 target genes in an intermediate state. *Oncogene*. 32:2565-2575.

Hernandez-Munoz, I., P. Taghavi, C. Kuijl, J. Neefjes, and M. van Lohuizen. 2005. Association of BMI1 with polycomb bodies is dynamic and requires PRC2/EZH2 and the maintenance DNA methyltransferase DNMT1. *Mol Cell Biol*. 25:11047-11058.

Herz, H.M., M. Mohan, A.S. Garruss, K. Liang, Y.H. Takahashi, K. Mickey, O. Voets, C.P. Verrijzer, and A. Shilatifard. 2012. Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. *Genes Dev*. 26:2604-2620.

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

Higuchi, M., D. O'Brien, P. Kumaravelu, N. Lenny, E.J. Yeoh, and J.R. Downing. 2002. Expression of a conditional AML1-ETO oncogene bypasses embryonic lethality and establishes a murine model of human t(8;21) acute myeloid leukemia. *Cancer Cell*. 1:63-74.

Hilton, I.B., A.M. D'Ippolito, C.M. Vockley, P.I. Thakore, G.E. Crawford, T.E. Reddy, and C.A. Gersbach. 2015. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol*. 33:510-517.

Hinai, A., M. Pratcorona, T. Grob, F.G. Kavelaars, E. Bussaglia, M.A. Sanders, J. Nomdedeu, and P.J.M. Valk. 2019. The Landscape of KMT2A-PTD AML: Concurrent Mutations, Gene Expression Signatures, and Clinical Outcome. *Hemasphere*. 3:e181.

Hiragami-Hamada, K., S. Soeroes, M. Nikolov, B. Wilkins, S. Kreuz, C. Chen, I.A. De La Rosa-Velazquez, H.M. Zenn, N. Kost, W. Pohl, A. Chernev, D. Schwarzer, T. Jenuwein, M. Lorincz, B. Zimmermann, P.J. Walla, H. Neumann, T. Baubec, H. Urlaub, and W. Fischle. 2016. Dynamic and flexible H3K9me3 bridging via HP1beta dimerization establishes a plastic state of condensed chromatin. *Nat Commun*. 7:11310.

Hnisz, D., A.S. Weintraub, D.S. Day, A.L. Valton, R.O. Bak, C.H. Li, J. Goldmann, B.R. Lajoie, Z.P. Fan, A.A. Sigova, J. Reddy, D. Borges-Rivera, T.I. Lee, R. Jaenisch, M.H. Porteus, J. Dekker, and R.A. Young. 2016. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science*. 351:1454-1458.

Hodgson, J.W., B. Argiropoulos, and H.W. Brock. 2001. Site-specific recognition of a 70-base-pair element containing d(GA)(n) repeats mediates bithoraxoid polycomb group response element-dependent silencing. *Mol Cell Biol*. 21:4528-4543.

Hodis, E., I.R. Watson, G.V. Kryukov, S.T. Arold, M. Imielinski, J.P. Theurillat, E. Nickerson, D. Auclair, L. Li, C. Place, D. Dicara, A.H. Ramos, M.S. Lawrence, K. Cibulskis, A. Sivachenko, D. Voet, G. Saksena, N. Stransky, R.C. Onofrio, W. Winckler, K. Ardlie, N. Wagle, J. Wargo, K. Chong, D.L. Morton, K. Stemke-Hale, G. Chen, M. Noble, M. Meyerson, J.E. Ladbury, M.A. Davies, J.E. Gershenwald, S.N. Wagner, D.S. Hoon, D. Schadendorf, E.S. Lander, S.B. Gabriel, G. Getz, L.A. Garraway, and L. Chin. 2012. A landscape of driver mutations in melanoma. *Cell*. 150:251-263.

Hoffmann, I., G. Draetta, and E. Karsenti. 1994. Activation of the phosphatase activity of human cdc25A by a cdk2-cyclin E dependent phosphorylation at the G1/S transition. *EMBO J*. 13:4302-4310.

Hojfeldt, J.W., L. Hedehus, A. Laugesen, T. Tatar, L. Wiehle, and K. Helin. 2019. Non-core Subunits of the PRC2 Complex Are Collectively Required for Its Target-Site Specificity. *Mol Cell*. 76:423-436 e423.

Hollein, A., S. Jeromin, M. Meggendorfer, A. Fasan, N. Nadarajah, W. Kern, C. Haferlach, and T. Haferlach. 2018. Minimal residual disease (MRD) monitoring and mutational landscape in AML with RUNX1-RUNX1T1: a study on 134 patients. *Leukemia*. 32:2270-2274.

Hosogane, M., R. Funayama, M. Shirota, and K. Nakayama. 2016. Lack of Transcription Triggers H3K27me3 Accumulation in the Gene Body. *Cell Rep*. 16:696-706.

Hou, C., L. Li, Z.S. Qin, and V.G. Corces. 2012. Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. *Mol Cell*. 48:471-484.

Hu, D., X. Gao, M.A. Morgan, H.M. Herz, E.R. Smith, and A. Shilatifard. 2013a. The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. *Mol Cell Biol*. 33:4745-4754.

Hu, D., A.S. Garruss, X. Gao, M.A. Morgan, M. Cook, E.R. Smith, and A. Shilatifard. 2013b. The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. *Nat Struct Mol Biol*. 20:1093-1097.

Hu, G., D.E. Schones, K. Cui, R. Ybarra, D. Northrup, Q. Tang, L. Gattinoni, N.P. Restifo, S. Huang, and K. Zhao. 2011. Regulation of nucleosome landscape and transcription factor targeting at tissue-specific enhancers by BRG1. *Genome Res*. 21:1650-1658.

Huang, G., K. Shigesada, K. Ito, H.J. Wee, T. Yokomizo, and Y. Ito. 2001. Dimerization with PEBP2beta protects RUNX1/AML1 from ubiquitin-proteasome-mediated degradation. *EMBO J*. 20:723-733.

Huang, G., X. Zhao, L. Wang, S. Elf, H. Xu, X. Zhao, G. Sashida, Y. Zhang, Y. Liu, J. Lee, S. Menendez, Y. Yang, X. Yan, P. Zhang, D.G. Tenen, M. Osato, J.J. Hsieh, and S.D. Nimer. 2011. The ability of MLL to bind RUNX1 and methylate H3K4 at PU.1 regulatory regions is impaired by MDS/AML-associated RUNX1/AML1 mutations. *Blood*. 118:6544-6552.

Hur, M.W., J.D. Laney, S.H. Jeon, J. Ali, and M.D. Biggin. 2002. Zeste maintains repression of Ubx transgenes: support for a new model of Polycomb repression. *Development*. 129:1339-1343.

Huseyin, M.K., and R.J. Klose. 2021. Live-cell single particle tracking of PRC1 reveals a highly dynamic system with low target site occupancy. *Nat Commun*. 12:887.

Ichikawa, M., T. Asai, T. Saito, S. Seo, I. Yamazaki, T. Yamagata, K. Mitani, S. Chiba, S. Ogawa, M. Kurokawa, and H. Hirai. 2004. AML-1 is required for megakaryocytic maturation and lymphocytic differentiation, but not for maintenance of hematopoietic stem cells in adult hematopoiesis. *Nat Med*. 10:299-304.

Illingworth, R.S., U. Gruenewald-Schneider, S. Webb, A.R. Kerr, K.D. James, D.J. Turner, C. Smith, D.J. Harrison, R. Andrews, and A.P. Bird. 2010. Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet*. 6:e1001134.

Ingham, P.W., and A. Martinez-Arias. 1986. The correct activation of Antennapedia and bithorax complex genes requires the fushi tarazu gene. *Nature*. 324:592-597.

Isono, K., T.A. Endo, M. Ku, D. Yamada, R. Suzuki, J. Sharif, T. Ishikura, T. Toyoda, B.E. Bernstein, and H. Koseki. 2013. SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. *Dev Cell*. 26:565-577.

Issaeva, I., Y. Zonis, T. Rozovskaia, K. Orlovsky, C.M. Croce, T. Nakamura, A. Mazo, L. Eisenbach, and E. Canaani. 2007. Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth. *Mol Cell Biol*. 27:1889-1903.

Jack, T., and W. McGinnis. 1990. Establishment of the Deformed expression stripe requires the combinatorial action of coordinate, gap and pair-rule proteins. *EMBO J*. 9:1187-1198.

Jaensch, E.S., J. Zhu, J.C. Cochrane, S.K. Marr, T.A. Oei, M. Damle, E.Z. McCaslin, and R.E. Kingston. 2021. A Polycomb domain found in committed cells impairs differentiation when introduced into PRC1 in pluripotent cells. *Mol Cell*. 81:4677-4691 e4678.

Jaffredo, T., R. Gautier, A. Eichmann, and F. Dieterlen-Lievre. 1998. Intraaortic hemopoietic cells are derived from endothelial cells during ontogeny. *Development*. 125:4575-4583.

Jahn, N., M. Agrawal, L. Bullinger, D. Weber, A. Corbacioglu, V.I. Gaidzik, L. Schmalbrock, F. Thol, M. Heuser, J. Krauter, G. Gohring, A. Kundgen, W. Fiedler, M. Wattad, G. Held, C.H. Kohne, H.A. Horst, M. Lubbert, A. Ganser, R.F. Schlenk, H. Dohner, K. Dohner, and P. Paschka. 2017. Incidence and prognostic impact of ASXL2 mutations in adult acute myeloid leukemia patients with t(8;21)(q22;q22): a study of the German-Austrian AML Study Group. *Leukemia*. 31:1012-1015.

Jeong, K.W., K. Kim, A.J. Situ, T.S. Ulmer, W. An, and M.R. Stallcup. 2011. Recognition of enhancer element-specific histone methylation by TIP60 in transcriptional activation. *Nat Struct Mol Biol*. 18:1358-1365.

Jin, F., Y. Li, J.R. Dixon, S. Selvaraj, Z. Ye, A.Y. Lee, C.A. Yen, A.D. Schmitt, C.A. Espinoza, and B. Ren. 2013. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature*. 503:290-294.

John, S., P.J. Sabo, R.E. Thurman, M.H. Sung, S.C. Biddie, T.A. Johnson, G.L. Hager, and J.A. Stamatoyannopoulos. 2011. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat Genet*. 43:264-268.

Johnson, J.L., G. Georgakilas, J. Petrovic, M. Kurachi, S. Cai, C. Harly, W.S. Pear, A. Bhandoola, E.J. Wherry, and G. Vahedi. 2018. Lineage-Determining Transcription Factor TCF-1 Initiates the Epigenetic Identity of T Cells. *Immunity*. 48:243-257 e210.

Jones, R.S., and W.M. Gelbart. 1990. Genetic analysis of the enhancer of zeste locus and its role in gene regulation in Drosophila melanogaster. *Genetics*. 126:185-199.

Jourdan, E., N. Boissel, S. Chevret, E. Delabesse, A. Renneville, P. Cornillet, O. Blanchet, J.M. Cayuela, C. Recher, E. Raffoux, J. Delaunay, A. Pigneux, C.E. Bulabois, C. Berthon, C. Pautas, N. Vey, B. Lioure, X. Thomas, I. Luquet, C. Terre, P. Guardiola, M.C. Bene, C. Preudhomme, N. Ifrah, H. Dombret, and A.M.L.I. French. 2013. Prospective evaluation of gene mutations and minimal residual disease in patients with core binding factor acute myeloid leukemia. *Blood*. 121:2213-2223.

Juan, A.H., S. Wang, K.D. Ko, H. Zare, P.F. Tsai, X. Feng, K.O. Vivanco, A.M. Ascoli, G. Gutierrez-Cruz, J. Krebs, S. Sidoli, A.L. Knight, R.A. Pedersen, B.A. Garcia, R. Casellas, J. Zou, and V. Sartorelli. 2016. Roles of H3K27me2 and H3K27me3 Examined during Fate Specification of Embryonic Stem Cells. *Cell Rep*. 17:1369-1382.

Justin, N., Y. Zhang, C. Tarricone, S.R. Martin, S. Chen, E. Underwood, V. De Marco, L.F. Haire, P.A. Walker, D. Reinberg, J.R. Wilson, and S.J. Gamblin. 2016. Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2. *Nat Commun*. 7:11316.

Kagey, M.H., J.J. Newman, S. Bilodeau, Y. Zhan, D.A. Orlando, N.L. van Berkum, C.C. Ebmeier, J. Goossens, P.B. Rahl, S.S. Levine, D.J. Taatjes, J. Dekker, and R.A. Young. 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature*. 467:430-435.

Kalb, R., S. Latwiel, H.I. Baymaz, P.W. Jansen, C.W. Muller, M. Vermeulen, and J. Muller. 2014. Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nat Struct Mol Biol*. 21:569-571.

Kalenik, J.L., D. Chen, M.E. Bradley, S.J. Chen, and T.C. Lee. 1997. Yeast two-hybrid cloning of a novel zinc finger protein that interacts with the multifunctional transcription factor YY1. *Nucleic Acids Res*. 25:843-849.

Kasinath, V., C. Beck, P. Sauer, S. Poepsel, J. Kosmatka, M. Faini, D. Toso, R. Aebersold, and E. Nogales. 2021. JARID2 and AEBP2 regulate PRC2 in the presence of H2AK119ub1 and other histone modifications. *Science*. 371.

Kassis, J.A., and J.L. Brown. 2013. Polycomb group response elements in Drosophila and vertebrates. *Adv Genet*. 81:83-118.

Kassis, J.A., J.A. Kennison, and J.W. Tamkun. 2017. Polycomb and Trithorax Group Genes in Drosophila. *Genetics*. 206:1699-1725.

Kaufman, T.C., R. Lewis, and B. Wakimoto. 1980. Cytogenetic Analysis of Chromosome 3 in DROSOPHILA MELANOGASTER: The Homoeotic Gene Complex in Polytene Chromosome Interval 84a-B. *Genetics*. 94:115-133.

Kearns, N.A., H. Pham, B. Tabak, R.M. Genga, N.J. Silverstein, M. Garber, and R. Maehr. 2015. Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat Methods*. 12:401-403.

Keegan, L., G. Gill, and M. Ptashne. 1986. Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. *Science*. 231:699-704.

Kerry, J., L. Godfrey, E. Repapi, M. Tapia, N.P. Blackledge, H. Ma, E. Ballabio, S. O'Byrne, F. Ponthan, O. Heidenreich, A. Roy, I. Roberts, M. Konopleva, R.J. Klose, H. Geng, and T.A. Milne. 2017. MLL-AF4 Spreading Identifies Binding Sites that Are Distinct from Super-Enhancers and that Govern Sensitivity to DOT1L Inhibition in Leukemia. *Cell Rep*. 18:482-495.

Kida, A., K. Kakihana, S. Kotani, T. Kurosu, and O. Miura. 2007. Glycogen synthase kinase-3beta and p38 phosphorylate cyclin D2 on Thr280 to trigger its ubiquitin/proteasome-dependent degradation in hematopoietic cells. *Oncogene*. 26:6630-6640.

Kim, C.A., M. Gingery, R.M. Pilpa, and J.U. Bowie. 2002. The SAM domain of polyhomeotic forms a helical polymer. *Nat Struct Biol*. 9:453-457.

Kim, J., and R.E. Kingston. 2020. The CBX family of proteins in transcriptional repression and memory. *J Biosci*. 45.

Kim, T.K., M. Hemberg, J.M. Gray, A.M. Costa, D.M. Bear, J. Wu, D.A. Harmin, M. Laptewicz, K. Barbara-Haley, S. Kuersten, E. Markenscoff-Papadimitriou, D. Kuhl, H. Bito, P.F. Worley, G.

Kreiman, and M.E. Greenberg. 2010. Widespread transcription at neuronal activity-regulated enhancers. *Nature*. 465:182-187.

Kimura, H., and P.R. Cook. 2001. Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B. *J Cell Biol*. 153:1341-1353.

King, H.W., N.A. Fursova, N.P. Blackledge, and R.J. Klose. 2018. Polycomb repressive complex 1 shapes the nucleosome landscape but not accessibility at target genes. *Genome Res*. 28:1494-1507.

Kitabayashi, I., A. Yokoyama, K. Shimizu, and M. Ohki. 1998. Interaction and functional cooperation of the leukemia-associated factors AML1 and p300 in myeloid cell differentiation. *EMBO J*. 17:2994-3004.

Klauke, K., V. Radulovic, M. Broekhuis, E. Weersing, E. Zwart, S. Olthof, M. Ritsema, S. Bruggeman, X. Wu, K. Helin, L. Bystrykh, and G. de Haan. 2013. Polycomb Cbx family members mediate the balance between haematopoietic stem cell self-renewal and differentiation. *Nat Cell Biol*. 15:353-362.

Kleer, C.G., Q. Cao, S. Varambally, R. Shen, I. Ota, S.A. Tomlins, D. Ghosh, R.G. Sewalt, A.P. Otte, D.F. Hayes, M.S. Sabel, D. Livant, S.J. Weiss, M.A. Rubin, and A.M. Chinnaiyan. 2003. EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc Natl Acad Sci U S A*. 100:11606-11611.

Kloet, S.L., M.M. Makowski, H.I. Baymaz, L. van Voorthuijsen, I.D. Karemaker, A. Santanach, P. Jansen, L. Di Croce, and M. Vermeulen. 2016. The dynamic interactome and genomic targets of Polycomb complexes during stem-cell differentiation. *Nat Struct Mol Biol*. 23:682-690.

Klose, R.J., S. Cooper, A.M. Farcas, N.P. Blackledge, and N. Brockdorff. 2013. Chromatin sampling--an emerging perspective on targeting polycomb repressor proteins. *PLoS Genet*. 9:e1003717.

Knutson, S.K., T.J. Wigle, N.M. Warholic, C.J. Sneeringer, C.J. Allain, C.R. Klaus, J.D. Sacks, A. Raimondi, C.R. Majer, J. Song, M.P. Scott, L. Jin, J.J. Smith, E.J. Olhava, R. Chesworth, M.P. Moyer, V.M. Richon, R.A. Copeland, H. Keilhack, R.M. Pollock, and K.W. Kuntz. 2012. A selective inhibitor of EZH2 blocks H3K27 methylation and kills mutant lymphoma cells. *Nat Chem Biol*. 8:890-896.

Koch, F., R. Fenouil, M. Gut, P. Cauchy, T.K. Albert, J. Zacarias-Cabeza, S. Spicuglia, A.L. de la Chapelle, M. Heidemann, C. Hintermair, D. Eick, I. Gut, P. Ferrier, and J.C. Andrau. 2011. Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat Struct Mol Biol*. 18:956-963.

Koh, C.P., C.Q. Wang, C.E. Ng, Y. Ito, M. Araki, V. Tergaonkar, G. Huang, and M. Osato. 2013. RUNX1 meets MLL: epigenetic regulation of hematopoiesis by two leukemia genes. *Leukemia*. 27:1793-1802.

Kondo, T., K. Isono, K. Kondo, T.A. Endo, S. Itohara, M. Vidal, and H. Koseki. 2014. Polycomb potentiates meis2 activation in midbrain by mediating interaction of the promoter with a tissue-specific enhancer. *Dev Cell*. 28:94-101.

Konze, K.D., A. Ma, F. Li, D. Barsyte-Lovejoy, T. Parton, C.J. Macnevin, F. Liu, C. Gao, X.P. Huang, E. Kuznetsova, M. Rougie, A. Jiang, S.G. Pattenden, J.L. Norris, L.I. James, B.L. Roth, P.J. Brown, S.V. Frye, C.H. Arrowsmith, K.M. Hahn, G.G. Wang, M. Vedadi, and J. Jin. 2013. An orally bioavailable chemical probe of the Lysine Methyltransferases EZH2 and EZH1. *ACS Chem Biol*. 8:1324-1334.

Kornberg, R.D. 1977. Structure of chromatin. *Annu Rev Biochem*. 46:931-954.

Kowalczyk, M.S., J.R. Hughes, D. Garrick, M.D. Lynch, J.A. Sharpe, J.A. Sloane-Stanley, S.J. McGowan, M. De Gobbi, M. Hosseini, D. Vernimmen, J.M. Brown, N.E. Gray, L. Collavin, R.J. Gibbons, J. Flint, S. Taylor, V.J. Buckle, T.A. Milne, W.G. Wood, and D.R. Higgs. 2012. Intragenic enhancers act as alternative promoters. *Mol Cell*. 45:447-458.

Krauth, M.T., C. Eder, T. Alpermann, U. Bacher, N. Nadarajah, W. Kern, C. Haferlach, T. Haferlach, and S. Schnittger. 2014. High number of additional genetic lesions in acute myeloid leukemia with t(8;21)/RUNX1-RUNX1T1: frequency and impact on clinical outcome. *Leukemia*. 28:1449-1458.

Krejci, O., M. Wunderlich, H. Geiger, F.S. Chou, D. Schleimer, M. Jansen, P.R. Andreassen, and J.C. Mulloy. 2008. p53 signaling in response to increased DNA damage sensitizes AML1-ETO cells to stress-induced death. *Blood*. 111:2190-2199.

Krivega, I., R.K. Dale, and A. Dean. 2014. Role of LDB1 in the transition from chromatin looping to transcription activation. *Genes Dev*. 28:1278-1290.

Krivega, I., and A. Dean. 2017. LDB1-mediated enhancer looping can be established independent of mediator and cohesin. *Nucleic Acids Res*. 45:8255-8268.

Krivtsov, A.V., Z. Feng, M.E. Lemieux, J. Faber, S. Vempati, A.U. Sinha, X. Xia, J. Jesneck, A.P. Bracken, L.B. Silverman, J.L. Kutok, A.L. Kung, and S.A. Armstrong. 2008. H3K79 methylation profiles define murine and human MLL-AF4 leukemias. *Cancer Cell*. 14:355-368.

Krogan, N.J., J. Dover, S. Khorrami, J.F. Greenblatt, J. Schneider, M. Johnston, and A. Shilatifard. 2002. COMPASS, a histone H3 (Lysine 4) methyltransferase required for telomeric silencing of gene expression. *J Biol Chem*. 277:10753-10755.

Kundu, S., F. Ji, H. Sunwoo, G. Jain, J.T. Lee, R.I. Sadreyev, J. Dekker, and R.E. Kingston. 2017. Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation. *Mol Cell*. 65:432-446 e435.

Kurzhals, R.L., F. Tie, C.A. Stratton, and P.J. Harte. 2008. Drosophila ESC-like can substitute for ESC and becomes required for Polycomb silencing if ESC is absent. *Dev Biol*. 313:293-306.

Kuzmichev, A., K. Nishioka, H. Erdjument-Bromage, P. Tempst, and D. Reinberg. 2002. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev*. 16:2893-2905.

Lai, B., J.E. Lee, Y. Jang, L. Wang, W. Peng, and K. Ge. 2017. MLL3/MLL4 are required for CBP/p300 binding on enhancers and super-enhancer formation in brown adipogenesis. *Nucleic Acids Res*. 45:6388-6403.

Lambert, S.A., A. Jolma, L.F. Campitelli, P.K. Das, Y. Yin, M. Albu, X. Chen, J. Taipale, T.R. Hughes, and M.T. Weirauch. 2018. The Human Transcription Factors. *Cell*. 172:650-665.

Lan, F., P.E. Bayliss, J.L. Rinn, J.R. Whetstine, J.K. Wang, S. Chen, S. Iwase, R. Alpatov, I. Issaeva, E. Canaani, T.M. Roberts, H.Y. Chang, and Y. Shi. 2007. A histone H3 lysine 27 demethylase regulates animal posterior development. *Nature*. 449:689-694.

Lancrin, C., M. Mazan, M. Stefanska, R. Patel, M. Lichtinger, G. Costa, O. Vargel, N.K. Wilson, T. Moroy, C. Bonifer, B. Gottgens, V. Kouskoff, and G. Lacaud. 2012. GFI1 and GFI1B control the loss of endothelial identity of hemogenic endothelium during hematopoietic commitment. *Blood*. 120:314-322.

Lancrin, C., P. Sroczynska, C. Stephenson, T. Allen, V. Kouskoff, and G. Lacaud. 2009. The haemangioblast generates haematopoietic cells through a haemogenic endothelium stage. *Nature*. 457:892-895.

Landrette, S.F., Y.H. Kuo, K. Hensen, S. Barjesteh van Waalwijk van Doorn-Khosrovani, P.N. Perrat, W.J. Van de Ven, R. Delwel, and L.H. Castilla. 2005. Plag1 and Plagl2 are oncogenes that induce acute myeloid leukemia in cooperation with Cbfb-MYH11. *Blood*. 105:2900-2907.

Lanigan, F., J.G. Geraghty, and A.P. Bracken. 2011. Transcriptional regulation of cellular senescence. *Oncogene*. 30:2901-2911.

Lau, M.S., M.G. Schwartz, S. Kundu, A.J. Savol, P.I. Wang, S.K. Marr, D.J. Grau, P. Schorderet, R.I. Sadreyev, C.J. Tabin, and R.E. Kingston. 2017. Mutation of a nucleosome compaction region disrupts Polycomb-mediated axial patterning. *Science*. 355:1081-1084.

Lavallee, V.P., S. Lemieux, G. Boucher, P. Gendron, I. Boivin, R.N. Armstrong, G. Sauvageau, and J. Hebert. 2016. RNA-sequencing analysis of core binding factor AML identifies recurrent ZBTB7A mutations and defines RUNX1-CBFA2T3 fusion signature. *Blood*. 127:2498-2501.

Lavarone, E., C.M. Barbieri, and D. Pasini. 2019. Dissecting the role of H3K27 acetylation and methylation in PRC2 mediated control of cellular identity. *Nat Commun*. 10:1679.

Lee, C.H., M. Holder, D. Grau, R. Saldana-Meyer, J.R. Yu, R.A. Ganai, J. Zhang, M. Wang, G. LeRoy, M.W. Dobenecker, D. Reinberg, and K.J. Armache. 2018. Distinct Stimulatory Mechanisms Regulate the Catalytic Activity of Polycomb Repressive Complex 2. *Mol Cell*. 70:435-448 e435.

Lee, H.G., T.G. Kahn, A. Simcox, Y.B. Schwartz, and V. Pirrotta. 2015. Genome-wide activities of Polycomb complexes control pervasive transcription. *Genome Res*. 25:1170-1181.

Lee, J.E., C. Wang, S. Xu, Y.W. Cho, L. Wang, X. Feng, A. Baldridge, V. Sartorelli, L. Zhuang, W. Peng, and K. Ge. 2013. H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. *Elife*. 2:e01503.

Lee, J.H., and D.G. Skalnik. 2005. CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. *J Biol Chem*. 280:41725-41731.

Lee, J.H., C.M. Tate, J.S. You, and D.G. Skalnik. 2007a. Identification and characterization of the human Set1B histone H3-Lys4 methyltransferase complex. *J Biol Chem*. 282:13419-13428.

Lee, M.G., R. Villa, P. Trojer, J. Norman, K.P. Yan, D. Reinberg, L. Di Croce, and R. Shiekhattar. 2007b. Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. *Science*. 318:447-450.

Lehnertz, B., Y.W. Zhang, I. Boivin, N. Mayotte, E. Tomellini, J. Chagraoui, V.P. Lavallee, J. Hebert, and G. Sauvageau. 2017. H3(K27M/I) mutations promote context-dependent transformation in acute myeloid leukemia with RUNX1 alterations. *Blood*. 130:2204-2214.

Lettice, L.A., S.J. Heaney, L.A. Purdie, L. Li, P. de Beer, B.A. Oostra, D. Goode, G. Elgar, R.E. Hill, and E. de Graaff. 2003. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet*. 12:1725-1735.

Levantini, E., S. Lee, H.S. Radomska, C.J. Hetherington, M. Alberich-Jorda, G. Amabile, P. Zhang, D.A. Gonzalez, J. Zhang, D.S. Basseres, N.K. Wilson, S. Koschmieder, G. Huang, D.E. Zhang, A.K. Ebralidze, C. Bonifer, Y. Okuno, B. Gottgens, and D.G. Tenen. 2011. RUNX1 regulates the CD34 gene in haematopoietic stem cells by mediating interactions with a distal regulatory element. *EMBO J*. 30:4059-4070.

Lewis, E., and R. Mislove. 1947. New mutants report. *Drosoph. Inf. Serv*. 21:69.

Lewis, E.B. 1978. A gene complex controlling segmentation in Drosophila. *Nature*. 276:565-570.

Lewis, P.W., M.M. Muller, M.S. Koletsky, F. Cordero, S. Lin, L.A. Banaszynski, B.A. Garcia, T.W. Muir, O.J. Becher, and C.D. Allis. 2013. Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. *Science*. 340:857-861.

Li, B., M. Gogol, M. Carey, D. Lee, C. Seidel, and J.L. Workman. 2007. Combined action of PHD and chromo domains directs the Rpd3S HDAC to transcribed chromatin. *Science*. 316:1050-1054.

Li, G., X. Ruan, R.K. Auerbach, K.S. Sandhu, M. Zheng, P. Wang, H.M. Poh, Y. Goh, J. Lim, J. Zhang, H.S. Sim, S.Q. Peh, F.H. Mulawadi, C.T. Ong, Y.L. Orlov, S. Hong, Z. Zhang, S. Landt, D. Raha, G. Euskirchen, C.L. Wei, W. Ge, H. Wang, C. Davis, K.I. Fisher-Aylor, A. Mortazavi, M. Gerstein, T. Gingeras, B. Wold, Y. Sun, M.J. Fullwood, E. Cheung, E. Liu, W.K. Sung, M. Snyder, and Y. Ruan. 2012. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell*. 148:84-98.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and S. Genome Project Data Processing. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 25:2078-2079.

Li, H., R. Liefke, J. Jiang, J.V. Kurland, W. Tian, P. Deng, W. Zhang, Q. He, D.J. Patel, M.L. Bulyk, Y. Shi, and Z. Wang. 2017. Polycomb-like proteins link the PRC2 complex to CpG islands. *Nature*. 549:287-291.

Li, Y., L. Gao, X. Luo, L. Wang, X. Gao, W. Wang, J. Sun, L. Dou, J. Li, C. Xu, L. Wang, M. Zhou, M. Jiang, J. Zhou, M.A. Caligiuri, C. Nervi, C.D. Bloomfield, G. Marcucci, and L. Yu. 2013.

Epigenetic silencing of microRNA-193a contributes to leukemogenesis in t(8;21) acute myeloid leukemia by activating the PTEN/PI3K signal pathway. *Blood*. 121:499-509.

Li, Y., H. Wang, X. Wang, W. Jin, Y. Tan, H. Fang, S. Chen, Z. Chen, and K. Wang. 2016. Genome-wide studies identify a novel interplay between AML1 and AML1/ETO in t(8;21) acute myeloid leukemia. *Blood*. 127:233-242.

Li, Z., R. Cao, M. Wang, M.P. Myers, Y. Zhang, and R.M. Xu. 2006. Structure of a Bmi-1-Ring1B polycomb group ubiquitin ligase complex. *J Biol Chem*. 281:20643-20649.

Li, Z., J. Lu, M. Sun, S. Mi, H. Zhang, R.T. Luo, P. Chen, Y. Wang, M. Yan, Z. Qian, M.B. Neilly, J. Jin, Y. Zhang, S.K. Bohlander, D.E. Zhang, R.A. Larson, M.M. Le Beau, M.J. Thirman, T.R. Golub, J.D. Rowley, and J. Chen. 2008. Distinct microRNA expression profiles in acute myeloid leukemia with common translocations. *Proc Natl Acad Sci U S A*. 105:15535-15540.

Liao, Y., G.K. Smyth, and W. Shi. 2013. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res*. 41:e108.

Lichtinger, M., R. Ingram, R. Hannah, D. Muller, D. Clarke, S.A. Assi, A.L.M. Lie, L. Noailles, M.S. Vijayabaskar, M. Wu, D.G. Tenen, D.R. Westhead, V. Kouskoff, G. Lacaud, B. Gottgens, and C. Bonifer. 2012. RUNX1 reshapes the epigenetic landscape at the onset of haematopoiesis. *EMBO J*. 31:4318-4333.

Lickwar, C.R., F. Mueller, S.E. Hanlon, J.G. McNally, and J.D. Lieb. 2012. Genome-wide protein-DNA binding dynamics suggest a molecular clutch for transcription factor function. *Nature*. 484:251-255.

Lie, A.L.M., E. Marinopoulou, Y. Li, R. Patel, M. Stefanska, C. Bonifer, C. Miller, V. Kouskoff, and G. Lacaud. 2014. RUNX1 positively regulates a cell adhesion and migration program in murine hemogenic endothelium prior to blood emergence. *Blood*. 124:e11-20.

Lie, A.L.M., E. Marinopoulou, A.J. Lilly, M. Challinor, R. Patel, C. Lancrin, V. Kouskoff, and G. Lacaud. 2018. Regulation of RUNX1 dosage is crucial for efficient blood formation from hemogenic endothelium. *Development*. 145.

Lie, A.L.M., R. Mevel, R. Patel, K. Blyth, E. Baena, V. Kouskoff, and G. Lacaud. 2020. RUNX1 Dosage in Development and Cancer. *Mol Cells*. 43:126-138.

Lieberman-Aiden, E., N.L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B.R. Lajoie, P.J. Sabo, M.O. Dorschner, R. Sandstrom, B. Bernstein, M.A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L.A. Mirny, E.S. Lander, and J. Dekker. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 326:289-293.

Lin, C., E.R. Smith, H. Takahashi, K.C. Lai, S. Martin-Brown, L. Florens, M.P. Washburn, J.W. Conaway, R.C. Conaway, and A. Shilatifard. 2010. AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia. *Mol Cell*. 37:429-437.

Lister, R., M. Pelizzola, R.H. Dowen, R.D. Hawkins, G. Hon, J. Tonti-Filippini, J.R. Nery, L. Lee, Z. Ye, Q.M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A.H. Millar, J.A. Thomson,

B. Ren, and J.R. Ecker. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*. 462:315-322.

Liu, J., N.B. Perumal, C.J. Oldfield, E.W. Su, V.N. Uversky, and A.K. Dunker. 2006. Intrinsic disorder in transcription factors. *Biochemistry*. 45:6873-6888.

Liu, P., S.A. Tarle, A. Hajra, D.F. Claxton, P. Marlton, M. Freedman, M.J. Siciliano, and F.S. Collins. 1993. Fusion between transcription factor CBF beta/PEBP2 beta and a myosin heavy chain in acute myeloid leukemia. *Science*. 261:1041-1044.

Long, H.K., N.P. Blackledge, and R.J. Klose. 2013. ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. *Biochem Soc Trans*. 41:727-740.

Loubiere, V., G.L. Papadopoulos, Q. Szabo, A.M. Martinez, and G. Cavalli. 2020. Widespread activation of developmental gene expression characterized by PRC1-dependent chromatin looping. *Sci Adv*. 6:eaax4001.

Loven, J., H.A. Hoke, C.Y. Lin, A. Lau, D.A. Orlando, C.R. Vakoc, J.E. Bradner, T.I. Lee, and R.A. Young. 2013. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*. 153:320-334.

Lu, F., and T. Lionnet. 2021. Transcription Factor Dynamics. *Cold Spring Harb Perspect Biol*. 13.

Lu, Y., Y.B. Xu, T.T. Yuan, M.G. Song, M. Lubbert, M. Fliegauf, and G.Q. Chen. 2006. Inducible expression of AML1-ETO fusion protein endows leukemic cells with susceptibility to extrinsic and intrinsic apoptosis. *Leukemia*. 20:987-993.

Luger, K., A.W. Mader, R.K. Richmond, D.F. Sargent, and T.J. Richmond. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*. 389:251-260.

Lupianez, D.G., K. Kraft, V. Heinrich, P. Krawitz, F. Brancati, E. Klopocki, D. Horn, H. Kayserili, J.M. Opitz, R. Laxova, F. Santos-Simarro, B. Gilbert-Dussardier, L. Wittler, M. Borschiwer, S.A. Haas, M. Osterwalder, M. Franke, B. Timmermann, J. Hecht, M. Spielmann, A. Visel, and S. Mundlos. 2015. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*. 161:1012-1025.

Lupien, M., J. Eeckhoute, C.A. Meyer, Q. Wang, Y. Zhang, W. Li, J.S. Carroll, X.S. Liu, and M. Brown. 2008. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell*. 132:958-970.

Luscombe, N.M., S.E. Austin, H.M. Berman, and J.M. Thornton. 2000. An overview of the structures of protein-DNA complexes. *Genome Biol*. 1:REVIEWS001.

Lutterbach, B., J.J. Westendorf, B. Linggi, A. Patten, M. Moniwa, J.R. Davie, K.D. Huynh, V.J. Bardwell, R.M. Lavinsky, M.G. Rosenfeld, C. Glass, E. Seto, and S.W. Hiebert. 1998. ETO, a target of t(8;21) in acute leukemia, interacts with the N-CoR and mSin3 corepressors. *Mol Cell Biol*. 18:7176-7184.

Lyko, F. 2018. The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nat Rev Genet*. 19:81-92.

Ma, A., E. Stratikopoulos, K.S. Park, J. Wei, T.C. Martin, X. Yang, M. Schwarz, V. Leshchenko, A. Rialdi, B. Dale, A. Lagana, E. Guccione, S. Parekh, R. Parsons, and J. Jin. 2020. Discovery of a first-in-class EZH2 selective degrader. *Nat Chem Biol*. 16:214-222.

Ma, C., V. Karwacki-Neisius, H. Tang, W. Li, Z. Shi, H. Hu, W. Xu, Z. Wang, L. Kong, R. Lv, Z. Fan, W. Zhou, P. Yang, F. Wu, J. Diao, L. Tan, Y.G. Shi, F. Lan, and Y. Shi. 2016. Nono, a Bivalent Domain Factor, Regulates Erk Signaling and Mouse Embryonic Stem Cell Pluripotency. *Cell Rep*. 17:997-1007.

Magoc, T., and S.L. Salzberg. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*. 27:2957-2963.

Majer, C.R., L. Jin, M.P. Scott, S.K. Knutson, K.W. Kuntz, H. Keilhack, J.J. Smith, M.P. Moyer, V.M. Richon, R.A. Copeland, and T.J. Wigle. 2012. A687V EZH2 is a gain-of-function mutation found in lymphoma patients. *FEBS Lett*. 586:3448-3451.

Mak, W., J. Baxter, J. Silva, A.E. Newall, A.P. Otte, and N. Brockdorff. 2002. Mitotically stable association of polycomb group proteins eed and enx1 with the inactive x chromosome in trophoblast stem cells. *Curr Biol*. 12:1016-1020.

Mansour, M.R., B.J. Abraham, L. Anders, A. Berezovskaya, A. Gutierrez, A.D. Durbin, J. Etchin, L. Lawton, S.E. Sallan, L.B. Silverman, M.L. Loh, S.P. Hunger, T. Sanda, R.A. Young, and A.T. Look. 2014. Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science*. 346:1373-1377.

Margueron, R., N. Justin, K. Ohno, M.L. Sharpe, J. Son, W.J. Drury, 3rd, P. Voigt, S.R. Martin, W.R. Taylor, V. De Marco, V. Pirrotta, D. Reinberg, and S.J. Gamblin. 2009. Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature*. 461:762-767.

Margueron, R., G. Li, K. Sarma, A. Blais, J. Zavadil, C.L. Woodcock, B.D. Dynlacht, and D. Reinberg. 2008. Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Mol Cell*. 32:503-518.

Marshall, N.F., and D.H. Price. 1995. Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *J Biol Chem*. 270:12335-12338.

Martens, J.H., A. Mandoli, F. Simmer, B.J. Wierenga, S. Saeed, A.A. Singh, L. Altucci, E. Vellenga, and H.G. Stunnenberg. 2012. ERG and FLI1 binding sites demarcate targets for aberrant epigenetic regulation by AML1-ETO in acute myeloid leukemia. *Blood*. 120:4038-4048.

Martinez-Soria, N., L. McKenzie, J. Draper, A. Ptasinska, H. Issa, S. Potluri, H.J. Blair, A. Pickin, A. Isa, P.S. Chin, R. Tirtakusuma, D. Coleman, S. Nakjang, S. Assi, V. Forster, M. Reza, E. Law, P. Berry, D. Mueller, C. Osborne, A. Elder, S.N. Bomken, D. Pal, J.M. Allan, G.J. Veal, P.N. Cockerill, C. Wichmann, J. Vormoor, G. Lacaud, C. Bonifer, and O. Heidenreich. 2018. The Oncogenic Transcription Factor RUNX1/ETO Corrupts Cell Cycle Regulation to Drive Leukemic Transformation. *Cancer Cell*. 34:626-642 e628.

Martire, S., A.A. Gogate, A. Whitmill, A. Tafessu, J. Nguyen, Y.C. Teng, M. Tastemel, and L.A. Banaszynski. 2019. Phosphorylation of histone H3.3 at serine 31 promotes p300 activity and enhancer acetylation. *Nat Genet*. 51:941-946.

Matthews, J.M., and J.E. Visvader. 2003. LIM-domain-binding protein 1: a multifunctional cofactor that interacts with diverse proteins. *EMBO Rep*. 4:1132-1137.

Mayran, A., K. Khetchoumian, F. Hariri, T. Pastinen, Y. Gauthier, A. Balsalobre, and J. Drouin. 2018. Pioneer factor Pax7 deploys a stable enhancer repertoire for specification of cell fate. *Nat Genet*. 50:259-269.

McCabe, M.T., A.P. Graves, G. Ganji, E. Diaz, W.S. Halsey, Y. Jiang, K.N. Smitheman, H.M. Ott, M.B. Pappalardi, K.E. Allen, S.B. Chen, A. Della Pietra, 3rd, E. Dul, A.M. Hughes, S.A. Gilbert, S.H. Thrall, P.J. Tummino, R.G. Kruger, M. Brandt, B. Schwartz, and C.L. Creasy. 2012a. Mutation of A677 in histone methyltransferase EZH2 in human B-cell lymphoma promotes hypertrimethylation of histone H3 on lysine 27 (H3K27). *Proc Natl Acad Sci U S A*. 109:2989-2994.

McCabe, M.T., H.M. Ott, G. Ganji, S. Korenchuk, C. Thompson, G.S. Van Aller, Y. Liu, A.P. Graves, A. Della Pietra, 3rd, E. Diaz, L.V. LaFrance, M. Mellinger, C. Duquenne, X. Tian, R.G. Kruger, C.F. McHugh, M. Brandt, W.H. Miller, D. Dhanak, S.K. Verma, P.J. Tummino, and C.L. Creasy. 2012b. EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature*. 492:108-112.

McCall, K., and W. Bender. 1996. Probes of chromatin accessibility in the Drosophila bithorax complex respond differently to Polycomb-mediated repression. *EMBO J*. 15:569-580.

McKay, D.J., S. Klusza, T.J. Penke, M.P. Meers, K.P. Curry, S.L. McDaniel, P.Y. Malek, S.W. Cooper, D.C. Tatomer, J.D. Lieb, B.D. Strahl, R.J. Duronio, and A.G. Matera. 2015. Interrogating the function of metazoan histones using engineered gene clusters. *Dev Cell*. 32:373-386.

McKeon, J., and H.W. Brock. 1991. Interactions of thePolycomb group of genes with homeotic loci ofDrosophila. *Roux Arch Dev Biol*. 199:387-396.

McManus, K.J., and M.J. Hendzel. 2003. Quantitative analysis of CBP- and P300-induced histone acetylations in vivo using native chromatin. *Mol Cell Biol*. 23:7611-7627.

McManus, S., A. Ebert, G. Salvagiotto, J. Medvedovic, Q. Sun, I. Tamir, M. Jaritz, H. Tagoh, and M. Busslinger. 2011. The transcription factor Pax5 regulates its target genes by recruiting chromatin-modifying proteins in committed B cells. *EMBO J*. 30:2388-2404.

Medina-Rivera, A., D. Santiago-Algarra, D. Puthier, and S. Spicuglia. 2018. Widespread Enhancer Activity from Core Promoters. *Trends Biochem Sci*. 43:452-468.

Mendenhall, E.M., R.P. Koche, T. Truong, V.W. Zhou, B. Issac, A.S. Chi, M. Ku, and B.E. Bernstein. 2010. GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS Genet*. 6:e1001244.

Mendenhall, E.M., K.E. Williamson, D. Reyon, J.Y. Zou, O. Ram, J.K. Joung, and B.E. Bernstein. 2013. Locus-specific editing of histone modifications at endogenous enhancers. *Nat Biotechnol*. 31:1133-1136.

Mendler, J.H., K. Maharry, M.D. Radmacher, K. Mrozek, H. Becker, K.H. Metzeler, S. Schwind, S.P. Whitman, J. Khalife, J. Kohlschmidt, D. Nicolet, B.L. Powell, T.H. Carter, M. Wetzler, J.O. Moore, J.E. Kolitz, M.R. Baer, A.J. Carroll, R.A. Larson, M.A. Caligiuri, G. Marcucci, and C.D.

Bloomfield. 2012. RUNX1 mutations are associated with poor outcome in younger and older patients with cytogenetically normal acute myeloid leukemia and with distinct gene and MicroRNA expression signatures. *J Clin Oncol*. 30:3109-3118.

Meyer, C., T. Burmeister, D. Groger, G. Tsaur, L. Fechina, A. Renneville, R. Sutton, N.C. Venn, M. Emerenciano, M.S. Pombo-de-Oliveira, C. Barbieri Blunck, B. Almeida Lopes, J. Zuna, J. Trka, P. Ballerini, H. Lapillonne, M. De Braekeleer, G. Cazzaniga, L. Corral Abascal, V.H.J. van der Velden, E. Delabesse, T.S. Park, S.H. Oh, M.L.M. Silva, T. Lund-Aho, V. Juvonen, A.S. Moore, O. Heidenreich, J. Vormoor, E. Zerkalenkova, Y. Olshanskaya, C. Bueno, P. Menendez, A. Teigler-Schlegel, U. Zur Stadt, J. Lentes, G. Gohring, A. Kustanovich, O. Aleinikova, B.W. Schafer, S. Kubetzko, H.O. Madsen, B. Gruhn, X. Duarte, P. Gameiro, E. Lippert, A. Bidet, J.M. Cayuela, E. Clappier, C.N. Alonso, C.M. Zwaan, M.M. van den Heuvel-Eibrink, S. Izraeli, L. Trakhtenbrot, P. Archer, J. Hancock, A. Moricke, J. Alten, M. Schrappe, M. Stanulla, S. Strehl, A. Attarbaschi, M. Dworzak, O.A. Haas, R. Panzer-Grumayer, L. Sedek, T. Szczepanski, A. Caye, L. Suarez, H. Cave, and R. Marschalek. 2018. The MLL recombinome of acute leukemias in 2017. *Leukemia*. 32:273-284.

Meyer, C., E. Kowarz, J. Hofmann, A. Renneville, J. Zuna, J. Trka, R. Ben Abdelali, E. Macintyre, E. De Braekeleer, M. De Braekeleer, E. Delabesse, M.P. de Oliveira, H. Cave, E. Clappier, J.J. van Dongen, B.V. Balgobind, M.M. van den Heuvel-Eibrink, H.B. Beverloo, R. Panzer-Grumayer, A. Teigler-Schlegel, J. Harbott, E. Kjeldsen, S. Schnittger, U. Koehl, B. Gruhn, O. Heidenreich, L.C. Chan, S.F. Yip, M. Krzywinski, C. Eckert, A. Moricke, M. Schrappe, C.N. Alonso, B.W. Schafer, J. Krauter, D.A. Lee, U. Zur Stadt, G. Te Kronnie, R. Sutton, S. Izraeli, L. Trakhtenbrot, L. Lo Nigro, G. Tsaur, L. Fechina, T. Szczepanski, S. Strehl, D. Ilencikova, M. Molkentin, T. Burmeister, T. Dingermann, T. Klingebiel, and R. Marschalek. 2009. New insights to the MLL recombinome of acute leukemias. *Leukemia*. 23:1490-1499.

Michaud, J., F. Wu, M. Osato, G.M. Cottles, M. Yanagida, N. Asou, K. Shigesada, Y. Ito, K.F. Benson, W.H. Raskind, C. Rossier, S.E. Antonarakis, S. Israels, A. McNicol, H. Weiss, M. Horwitz, and H.S. Scott. 2002. In vitro analyses of known and novel RUNX1/AML1 mutations in dominant familial platelet disorder with predisposition to acute myelogenous leukemia: implications for mechanisms of pathogenesis. *Blood*. 99:1364-1372.

Micol, J.B., N. Duployez, N. Boissel, A. Petit, S. Geffroy, O. Nibourel, C. Lacombe, H. Lapillonne, P. Etancelin, M. Figeac, A. Renneville, S. Castaigne, G. Leverger, N. Ifrah, H. Dombret, C. Preudhomme, O. Abdel-Wahab, and E. Jourdan. 2014. Frequent ASXL2 mutations in acute myeloid leukemia patients with t(8;21)/RUNX1-RUNX1T1 chromosomal translocations. *Blood*. 124:1445-1449.

Micol, J.B., A. Pastore, D. Inoue, N. Duployez, E. Kim, S.C. Lee, B.H. Durham, Y.R. Chung, H. Cho, X.J. Zhang, A. Yoshimi, A. Krivtsov, R. Koche, E. Solary, A. Sinha, C. Preudhomme, and O. Abdel-Wahab. 2017. ASXL2 is essential for haematopoiesis and acts as a haploinsufficient tumour suppressor in leukemia. *Nat Commun*. 8:15429.

Miller, C.A., R.K. Wilson, and T.J. Ley. 2013. Genomic landscapes and clonality of de novo AML. *N Engl J Med*. 369:1473.

Miller, T., N.J. Krogan, J. Dover, H. Erdjument-Bromage, P. Tempst, M. Johnston, J.F. Greenblatt, and A. Shilatifard. 2001. COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proc Natl Acad Sci U S A*. 98:12902-12907.

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

Milne, T.A., J. Kim, G.G. Wang, S.C. Stadler, V. Basrur, S.J. Whitcomb, Z. Wang, A.J. Ruthenburg, K.S. Elenitoba-Johnson, R.G. Roeder, and C.D. Allis. 2010. Multiple interactions recruit MLL1 and MLL1 fusion proteins to the HOXA9 locus in leukemogenesis. *Mol Cell*. 38:853-863.

Min, J., Y. Zhang, and R.M. Xu. 2003. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev*. 17:1823-1828.

Mirzaa, G., D.A. Parry, A.E. Fry, K.A. Giamanco, J. Schwartzentruber, M. Vanstone, C.V. Logan, N. Roberts, C.A. Johnson, S. Singh, S.S. Kholmanskikh, C. Adams, R.D. Hodge, R.F. Hevner, D.T. Bonthron, K.P.J. Braun, L. Faivre, J.B. Riviere, J. St-Onge, K.W. Gripp, G.M. Mancini, K. Pang, E. Sweeney, H. van Esch, N. Verbeek, D. Wieczorek, M. Steinraths, J. Majewski, F.C. Consortium, K.M. Boycot, D.T. Pilz, M.E. Ross, W.B. Dobyns, and E.G. Sheridan. 2014. De novo CCND2 mutations leading to stabilization of cyclin D2 cause megalencephaly-polymicrogyria-polydactyly-hydrocephalus syndrome. *Nat Genet*. 46:510-515.

Misaghi, S., S. Ottosen, A. Izrael-Tomasevic, D. Arnott, M. Lamkanfi, J. Lee, J. Liu, K. O'Rourke, V.M. Dixit, and A.C. Wilson. 2009. Association of C-terminal ubiquitin hydrolase BRCA1-associated protein 1 with cell cycle regulator host cell factor 1. *Mol Cell Biol*. 29:2181-2192.

Mitchell, P.J., and R. Tjian. 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science*. 245:371-378.

Miyoshi, H., T. Kozu, K. Shimizu, K. Enomoto, N. Maseki, Y. Kaneko, N. Kamada, and M. Ohki. 1993. The t(8;21) translocation in acute myeloid leukemia results in production of an AML1-MTG8 fusion transcript. *EMBO J*. 12:2715-2721.

Miyoshi, H., K. Shimizu, T. Kozu, N. Maseki, Y. Kaneko, and M. Ohki. 1991. t(8;21) breakpoints on chromosome 21 in acute myeloid leukemia are clustered within a limited region of a single gene, AML1. *Proc Natl Acad Sci U S A*. 88:10431-10434.

Mohn, F., M. Weber, M. Rebhan, T.C. Roloff, J. Richter, M.B. Stadler, M. Bibel, and D. Schubeler. 2008. Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol Cell*. 30:755-766.

Monroe, S.C., S.Y. Jo, D.S. Sanders, V. Basrur, K.S. Elenitoba-Johnson, R.K. Slany, and J.L. Hess. 2011. MLL-AF9 and MLL-ENL alter the dynamic association of transcriptional regulators with genes critical for leukemia. *Exp Hematol*. 39:77-86 e71-75.

Montalto, F.I., and F. De Amicis. 2020. Cyclin D1 in Cancer: A Molecular Connection for Cell Cycle Control, Adhesion and Invasion in Tumor and Stroma. *Cells*. 9.

Morey, L., G. Pascual, L. Cozzuto, G. Roma, A. Wutz, S.A. Benitah, and L. Di Croce. 2012. Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells. *Cell Stem Cell*. 10:47-62.

Morin, R.D., N.A. Johnson, T.M. Severson, A.J. Mungall, J. An, R. Goya, J.E. Paul, M. Boyle, B.W. Woolcock, F. Kuchenbauer, D. Yap, R.K. Humphries, O.L. Griffith, S. Shah, H. Zhu, M. Kimbara, P. Shashkin, J.F. Charlot, M. Tcherpakov, R. Corbett, A. Tam, R. Varhol, D. Smailus,

M. Moksa, Y. Zhao, A. Delaney, H. Qian, I. Birol, J. Schein, R. Moore, R. Holt, D.E. Horsman, J.M. Connors, S. Jones, S. Aparicio, M. Hirst, R.D. Gascoyne, and M.A. Marra. 2010. Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat Genet*. 42:181-185.

Mousavi, K., H. Zare, A.H. Wang, and V. Sartorelli. 2012. Polycomb protein Ezh1 promotes RNA polymerase II elongation. *Mol Cell*. 45:255-262.

Mueller, D., M.P. Garcia-Cuellar, C. Bach, S. Buhl, E. Maethner, and R.K. Slany. 2009. Misguided transcriptional elongation causes mixed lineage leukemia. *PLoS Biol*. 7:e1000249.

Mukouyama, Y., N. Chiba, T. Hara, H. Okada, Y. Ito, R. Kanamaru, A. Miyajima, M. Satake, and T. Watanabe. 2000. The AML1 transcription factor functions to develop and maintain hematogenic precursor cells in the embryonic aorta-gonad-mesonephros region. *Dev Biol*. 220:27-36.

Muller-Tidow, C., B. Steffen, T. Cauvet, L. Tickenbrock, P. Ji, S. Diederichs, B. Sargin, G. Kohler, M. Stelljes, E. Puccetti, M. Ruthardt, S. deVos, S.W. Hiebert, H.P. Koeffler, W.E. Berdel, and H. Serve. 2004. Translocation products in acute myeloid leukemia activate the Wnt signaling pathway in hematopoietic cells. *Mol Cell Biol*. 24:2890-2904.

Muller, J., C.M. Hart, N.J. Francis, M.L. Vargas, A. Sengupta, B. Wild, E.L. Miller, M.B. O'Connor, R.E. Kingston, and J.A. Simon. 2002. Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell*. 111:197-208.

Muntean, A.G., J. Tan, K. Sitwala, Y. Huang, J. Bronstein, J.A. Connelly, V. Basrur, K.S. Elenitoba-Johnson, and J.L. Hess. 2010. The PAF complex synergizes with MLL fusion proteins at HOX loci to promote leukemogenesis. *Cancer Cell*. 17:609-621.

Muto, T., G. Sashida, M. Oshima, G.R. Wendt, M. Mochizuki-Kashio, Y. Nagata, M. Sanada, S. Miyagi, A. Saraya, A. Kamio, G. Nagae, C. Nakaseko, K. Yokote, K. Shimoda, H. Koseki, Y. Suzuki, S. Sugano, H. Aburatani, S. Ogawa, and A. Iwama. 2013. Concurrent loss of Ezh2 and Tet2 cooperates in the pathogenesis of myelodysplastic disorders. *J Exp Med*. 210:2627-2639.

Nakagawa, T., T. Kajitani, S. Togo, N. Masuko, H. Ohdan, Y. Hishikawa, T. Koji, T. Matsuyama, T. Ikura, M. Muramatsu, and T. Ito. 2008. Deubiquitylation of histone H2A activates transcriptional initiation via trans-histone cross-talk with H3K4 di- and trimethylation. *Genes Dev*. 22:37-49.

Narita, M., S. Nunez, E. Heard, M. Narita, A.W. Lin, S.A. Hearn, D.L. Spector, G.J. Hannon, and S.W. Lowe. 2003. Rb-mediated heterochromatin formation and silencing of E2F target genes during cellular senescence. *Cell*. 113:703-716.

Negishi, M., A. Saraya, S. Mochizuki, K. Helin, H. Koseki, and A. Iwama. 2010. A novel zinc finger protein Zfp277 mediates transcriptional repression of the Ink4a/arf locus through polycomb repressive complex 1. *PLoS One*. 5:e12373.

Nikoloski, G., S.M. Langemeijer, R.P. Kuiper, R. Knops, M. Massop, E.R. Tonnissen, A. van der Heijden, T.N. Scheele, P. Vandenberghe, T. de Witte, B.A. van der Reijden, and J.H. Jansen. 2010. Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. *Nat Genet*. 42:665-667.

Nolis, I.K., D.J. McKay, E. Mantouvalou, S. Lomvardas, M. Merika, and D. Thanos. 2009. Transcription factors mediate long-range enhancer-promoter interactions. *Proc Natl Acad Sci U S A*. 106:20222-20227.

North, T., T.L. Gu, T. Stacy, Q. Wang, L. Howard, M. Binder, M. Marin-Padilla, and N.A. Speck. 1999. Cbfa2 is required for the formation of intra-aortic hematopoietic clusters. *Development*. 126:2563-2575.

Ntziachristos, P., A. Tsirigos, P. Van Vlierberghe, J. Nedjic, T. Trimarchi, M.S. Flaherty, D. Ferres-Marco, V. da Ros, Z. Tang, J. Siegle, P. Asp, M. Hadler, I. Rigo, K. De Keersmaecker, J. Patel, T. Huynh, F. Utro, S. Poglio, J.B. Samon, E. Paietta, J. Racevskis, J.M. Rowe, R. Rabadan, R.L. Levine, S. Brown, F. Pflumio, M. Dominguez, A. Ferrando, and I. Aifantis. 2012. Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia. *Nat Med*. 18:298-301.

Nucifora, G., D.J. Birn, P. Erickson, J. Gao, M.M. LeBeau, H.A. Drabkin, and J.D. Rowley. 1993. Detection of DNA rearrangements in the AML1 and ETO loci and of an AML1/ETO fusion mRNA in patients with t(8;21) acute myeloid leukemia. *Blood*. 81:883-888.

Nusslein-Volhard, C., and E. Wieschaus. 1980. Mutations affecting segment number and polarity in Drosophila. *Nature*. 287:795-801.

O'Carroll, D., S. Erhardt, M. Pagani, S.C. Barton, M.A. Surani, and T. Jenuwein. 2001. The polycomb-group gene Ezh2 is required for early mouse development. *Mol Cell Biol*. 21:4330-4336.

O'Loghlen, A., A.M. Munoz-Cabello, A. Gaspar-Maia, H.A. Wu, A. Banito, N. Kunowska, T. Racek, H.N. Pemberton, P. Beolchi, F. Lavial, O. Masui, M. Vermeulen, T. Carroll, J. Graumann, E. Heard, N. Dillon, V. Azuara, A.P. Snijders, G. Peters, E. Bernstein, and J. Gil. 2012. MicroRNA regulation of Cbx7 mediates a switch of Polycomb orthologs during ESC differentiation. *Cell Stem Cell*. 10:33-46.

Ochi, Y., A. Kon, T. Sakata, M.M. Nakagawa, N. Nakazawa, M. Kakuta, K. Kataoka, H. Koseki, M. Nakayama, D. Morishita, T. Tsuruyama, R. Saiki, A. Yoda, R. Okuda, T. Yoshizato, K. Yoshida, Y. Shiozawa, Y. Nannya, S. Kotani, Y. Kogure, N. Kakiuchi, T. Nishimura, H. Makishima, L. Malcovati, A. Yokoyama, K. Takeuchi, E. Sugihara, T.A. Sato, M. Sanada, A. Takaori-Kondo, M. Cazzola, M. Kengaku, S. Miyano, K. Shirahige, H.I. Suzuki, and S. Ogawa. 2020. Combined Cohesin-RUNX1 Deficiency Synergistically Perturbs Chromatin Looping and Causes Myelodysplastic Syndromes. *Cancer Discov*. 10:836-853.

Ogryzko, V.V., R.L. Schiltz, V. Russanova, B.H. Howard, and Y. Nakatani. 1996. The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell*. 87:953-959.

Okada, Y., Q. Feng, Y. Lin, Q. Jiang, Y. Li, V.M. Coffield, L. Su, G. Xu, and Y. Zhang. 2005. hDOT1L links histone methylation to leukemogenesis. *Cell*. 121:167-178.

Okuda, H., M. Kawaguchi, A. Kanai, H. Matsui, T. Kawamura, T. Inaba, I. Kitabayashi, and A. Yokoyama. 2014. MLL fusion proteins link transcriptional coactivators to previously active CpG-rich promoters. *Nucleic Acids Res*. 42:4241-4256.

Okuda, T., Z. Cai, S. Yang, N. Lenny, C.J. Lyu, J.M. van Deursen, H. Harada, and J.R. Downing. 1998. Expression of a knocked-in AML1-ETO leukemia gene inhibits the establishment of normal

definitive hematopoiesis and directly generates dysplastic hematopoietic progenitors. *Blood*. 91:3134-3143.

Okuda, T., J. van Deursen, S.W. Hiebert, G. Grosveld, and J.R. Downing. 1996. AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. *Cell*. 84:321-330.

Oricchio, E., N. Katanayeva, M.C. Donaldson, S. Sungalee, J.P. Pasion, W. Beguelin, E. Battistello, V.R. Sanghvi, M. Jiang, Y. Jiang, M. Teater, A. Parmigiani, A.V. Budanov, F.C. Chan, S.P. Shah, R. Kridel, A.M. Melnick, G. Ciriello, and H.G. Wendel. 2017. Genetic and epigenetic inactivation of SESTRIN1 controls mTORC1 and response to EZH2 inhibition in follicular lymphoma. *Sci Transl Med*. 9.

Orlando, D.A., M.W. Chen, V.E. Brown, S. Solanki, Y.J. Choi, E.R. Olson, C.C. Fritz, J.E. Bradner, and M.G. Guenther. 2014. Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. *Cell Rep*. 9:1163-1170.

Orom, U.A., T. Derrien, M. Beringer, K. Gumireddy, A. Gardini, G. Bussotti, F. Lai, M. Zytnicki, C. Notredame, Q. Huang, R. Guigo, and R. Shiekhattar. 2010. Long noncoding RNAs with enhancer-like function in human cells. *Cell*. 143:46-58.

Osato, M. 2004. Point mutations in the RUNX1/AML1 gene: another actor in RUNX leukemia. *Oncogene*. 23:4284-4296.

Osato, M., N. Asou, E. Abdalla, K. Hoshino, H. Yamasaki, T. Okubo, H. Suzushima, K. Takatsuki, T. Kanno, K. Shigesada, and Y. Ito. 1999. Biallelic and heterozygous point mutations in the runt domain of the AML1/PEBP2alphaB gene associated with myeloblastic leukemias. *Blood*. 93:1817-1824.

Ott, H.M., A.P. Graves, M.B. Pappalardi, M. Huddleston, W.S. Halsey, A.M. Hughes, A. Groy, E. Dul, Y. Jiang, Y. Bai, R. Annan, S.K. Verma, S.D. Knight, R.G. Kruger, D. Dhanak, B. Schwartz, P.J. Tummino, C.L. Creasy, and M.T. McCabe. 2014. A687V EZH2 is a driver of histone H3 lysine 27 (H3K27) hypertrimethylation. *Mol Cancer Ther*. 13:3062-3073.

Ottersbach, K. 2019. Endothelial-to-haematopoietic transition: an update on the process of making blood. *Biochem Soc Trans*. 47:591-601.

Ougolkov, A.V., V.N. Bilim, and D.D. Billadeau. 2008. Regulation of pancreatic tumor cell proliferation and chemoresistance by the histone methyltransferase enhancer of zeste homologue 2. *Clin Cancer Res*. 14:6790-6796.

Pabst, T., B.U. Mueller, N. Harakawa, C. Schoch, T. Haferlach, G. Behre, W. Hiddemann, D.E. Zhang, and D.G. Tenen. 2001. AML1-ETO downregulates the granulocytic differentiation factor C/EBPalpha in t(8;21) myeloid leukemia. *Nat Med*. 7:444-451.

Pachano, T., G. Crispatzu, and A. Rada-Iglesias. 2019. Polycomb proteins as organizers of 3D genome architecture in embryonic stem cells. *Brief Funct Genomics*. 18:358-366.

Pan, G., S. Tian, J. Nie, C. Yang, V. Ruotti, H. Wei, G.A. Jonsdottir, R. Stewart, and J.A. Thomson. 2007. Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell*. 1:299-312.

Paro, R., and D.S. Hogness. 1991. The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. *Proc Natl Acad Sci U S A*. 88:263-267.

Pasini, D., A.P. Bracken, J.B. Hansen, M. Capillo, and K. Helin. 2007. The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Mol Cell Biol*. 27:3769-3779.

Pasini, D., A.P. Bracken, M.R. Jensen, E. Lazzerini Denchi, and K. Helin. 2004. Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO J*. 23:4061-4071.

Pasini, D., M. Malatesta, H.R. Jung, J. Walfridsson, A. Willer, L. Olsson, J. Skotte, A. Wutz, B. Porse, O.N. Jensen, and K. Helin. 2010. Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. *Nucleic Acids Res*. 38:4958-4969.

Pengelly, A.R., O. Copur, H. Jackle, A. Herzig, and J. Muller. 2013. A histone mutant reproduces the phenotype caused by loss of histone-modifying factor Polycomb. *Science*. 339:698-699.

Pengelly, A.R., R. Kalb, K. Finkl, and J. Muller. 2015. Transcriptional repression by PRC1 in the absence of H2A monoubiquitylation. *Genes Dev*. 29:1487-1492.

Pennacchio, L.A., N. Ahituv, A.M. Moses, S. Prabhakar, M.A. Nobrega, M. Shoukry, S. Minovitsky, I. Dubchak, A. Holt, K.D. Lewis, I. Plajzer-Frick, J. Akiyama, S. De Val, V. Afzal, B.L. Black, O. Couronne, M.B. Eisen, A. Visel, and E.M. Rubin. 2006. In vivo enhancer analysis of human conserved non-coding sequences. *Nature*. 444:499-502.

Petruk, S., J. Cai, R. Sussman, G. Sun, S.K. Kovermann, S.A. Mariani, B. Calabretta, S.B. McMahon, H.W. Brock, L. Iacovitti, and A. Mazo. 2017a. Delayed Accumulation of H3K27me3 on Nascent DNA Is Essential for Recruitment of Transcription Factors at Early Stages of Stem Cell Differentiation. *Mol Cell*. 66:247-257 e245.

Petruk, S., S.A. Mariani, M. De Dominici, P. Porazzi, V. Minieri, J. Cai, L. Iacovitti, N. Flomenberg, B. Calabretta, and A. Mazo. 2017b. Structure of Nascent Chromatin Is Essential for Hematopoietic Lineage Specification. *Cell Rep*. 19:295-306.

Pi, W.C., J. Wang, M. Shimada, J.W. Lin, H. Geng, Y.L. Lee, R. Lu, D. Li, G.G. Wang, R.G. Roeder, and W.Y. Chen. 2020. E2A-PBX1 functions as a coactivator for RUNX1 in acute lymphoblastic leukemia. *Blood*. 136:11-23.

Pintacuda, G., G. Wei, C. Roustan, B.A. Kirmizitas, N. Solcan, A. Cerase, A. Castello, S. Mohammed, B. Moindrot, T.B. Nesterova, and N. Brockdorff. 2017. hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. *Mol Cell*. 68:955-969 e910.

Platero, J.S., T. Hartnett, and J.C. Eissenberg. 1995. Functional analysis of the chromo domain of HP1. *EMBO J*. 14:3977-3986.

Plath, K., J. Fang, S.K. Mlynarczyk-Evans, R. Cao, K.A. Worringer, H. Wang, C.C. de la Cruz, A.P. Otte, B. Panning, and Y. Zhang. 2003. Role of histone H3 lysine 27 methylation in X inactivation. *Science*. 300:131-135.

Plys, A.J., C.P. Davis, J. Kim, G. Rizki, M.M. Keenen, S.K. Marr, and R.E. Kingston. 2019. Phase separation of Polycomb-repressive complex 1 is governed by a charged disordered region of CBX2. *Genes Dev*. 33:799-813.

Popp, A.P., J. Hettich, and J.C.M. Gebhardt. 2020. Transcription factor residence time dominates over concentration in transcription activation. *bioRxiv*:2020.2011.2026.400069.

Porazzi, P., S. Petruk, L. Pagliaroli, M. De Dominici, D. Deming, 2nd, M.V. Puccetti, S. Kushinsky, G. Kumar, V. Minieri, E. Barbieri, S. Deliard, A. Grande, M. Trizzino, A. Gardini, E. Canaani, N. Palmisiano, P. Porcu, A. Ertel, P. Fortina, C.M. Eischen, A. Mazo, and B. Calabretta. 2022. Targeting Chemotherapy to Decondensed H3K27me3-Marked Chromatin of AML Cells Enhances Leukemia Suppression. *Cancer Res*. 82:458-471.

Portoso, M., R. Ragazzini, Z. Brencic, A. Moiani, A. Michaud, I. Vassilev, M. Wassef, N. Servant, B. Sargueil, and R. Margueron. 2017. PRC2 is dispensable for HOTAIR-mediated transcriptional repression. *EMBO J*. 36:981-994.

Pradeepa, M.M., G.R. Grimes, Y. Kumar, G. Olley, G.C. Taylor, R. Schneider, and W.A. Bickmore. 2016. Histone H3 globular domain acetylation identifies a new class of enhancers. *Nat Genet*. 48:681-686.

Pradeepa, M.M., H.G. Sutherland, J. Ule, G.R. Grimes, and W.A. Bickmore. 2012. Psip1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing. *PLoS Genet*. 8:e1002717.

Prange, K.H.M., A. Mandoli, T. Kuznetsova, S.Y. Wang, A.M. Sotoca, A.E. Marneth, B.A. van der Reijden, H.G. Stunnenberg, and J.H.A. Martens. 2017. MLL-AF9 and MLL-AF4 oncofusion proteins bind a distinct enhancer repertoire and target the RUNX1 program in 11q23 acute myeloid leukemia. *Oncogene*. 36:3346-3356.

Ptasinska, A., S.A. Assi, D. Mannari, S.R. James, D. Williamson, J. Dunne, M. Hoogenkamp, M. Wu, M. Care, H. McNeill, P. Cauchy, M. Cullen, R.M. Tooze, D.G. Tenen, B.D. Young, P.N. Cockerill, D.R. Westhead, O. Heidenreich, and C. Bonifer. 2012. Depletion of RUNX1/ETO in t(8;21) AML cells leads to genome-wide changes in chromatin structure and transcription factor binding. *Leukemia*. 26:1829-1841.

Ptasinska, A., S.A. Assi, N. Martinez-Soria, M.R. Imperato, J. Piper, P. Cauchy, A. Pickin, S.R. James, M. Hoogenkamp, D. Williamson, M. Wu, D.G. Tenen, S. Ott, D.R. Westhead, P.N. Cockerill, O. Heidenreich, and C. Bonifer. 2014. Identification of a dynamic core transcriptional network in t(8;21) AML that regulates differentiation block and self-renewal. *Cell Rep*. 8:1974-1988.

Pulikkan, J.A., M. Hegde, H.M. Ahmad, H. Belaghzal, A. Illendula, J. Yu, K. O'Hagan, J. Ou, C. Muller-Tidow, S.A. Wolfe, L.J. Zhu, J. Dekker, J.H. Bushweller, and L.H. Castilla. 2018. CBFbeta-SMMHC Inhibition Triggers Apoptosis by Disrupting MYC Chromatin Dynamics in Acute Myeloid Leukemia. *Cell*. 174:172-186 e121.

Pulikkan, J.A., D. Madera, L. Xue, P. Bradley, S.F. Landrette, Y.H. Kuo, S. Abbas, L.J. Zhu, P. Valk, and L.H. Castilla. 2012. Thrombopoietin/MPL participates in initiating and maintaining RUNX1-ETO acute myeloid leukemia via PI3K/AKT signaling. *Blood*. 120:868-879.

Qi, W., H. Chan, L. Teng, L. Li, S. Chuai, R. Zhang, J. Zeng, M. Li, H. Fan, Y. Lin, J. Gu, O. Ardayfio, J.H. Zhang, X. Yan, J. Fang, Y. Mi, M. Zhang, T. Zhou, G. Feng, Z. Chen, G. Li, T. Yang, K. Zhao, X. Liu, Z. Yu, C.X. Lu, P. Atadja, and E. Li. 2012. Selective inhibition of Ezh2 by a small molecule inhibitor blocks tumor cells proliferation. *Proc Natl Acad Sci U S A*. 109:21360-21365.

Rada-Iglesias, A., R. Bajpai, T. Swigut, S.A. Brugmann, R.A. Flynn, and J. Wysocka. 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*. 470:279-283.

Ramirez, F., D.P. Ryan, B. Gruning, V. Bhardwaj, F. Kilpert, A.S. Richter, S. Heyne, F. Dundar, and T. Manke. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. 44:W160-165.

Reed-Inderbitzin, E., I. Moreno-Miralles, S.K. Vanden-Eynden, J. Xie, B. Lutterbach, K.L. Durst-Goodwin, K.S. Luce, B.J. Irvin, M.L. Cleary, S.J. Brandt, and S.W. Hiebert. 2006. RUNX1 associates with histone deacetylases and SUV39H1 to repress transcription. *Oncogene*. 25:5777-5786.

Reinitz, J., and M. Levine. 1990. Control of the initiation of homeotic gene expression by the gap genes giant and tailless in Drosophila. *Dev Biol*. 140:57-72.

Reisser, M., A. Palmer, A.P. Popp, C. Jahn, G. Weidinger, and J.C.M. Gebhardt. 2018. Single-molecule imaging correlates decreasing nuclear volume with increasing TF-chromatin associations during zebrafish development. *Nat Commun*. 9:5218.

Rennert, J., J.A. Coffman, A.R. Mushegian, and A.J. Robertson. 2003. The evolution of Runx genes I. A comparative study of sequences from phylogenetically diverse model organisms. *BMC Evol Biol*. 3:4.

Rennie, S., M. Dalby, M. Lloret-Llinares, S. Bakoulis, C. Dalager Vaagenso, T. Heick Jensen, and R. Andersson. 2018. Transcription start site analysis reveals widespread divergent transcription in D. melanogaster and core promoter-encoded enhancer activities. *Nucleic Acids Res*. 46:5455-5469.

Riising, E.M., I. Comet, B. Leblanc, X. Wu, J.V. Johansen, and K. Helin. 2014. Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. *Mol Cell*. 55:347-360.

Rinn, J.L., M. Kertesz, J.K. Wang, S.L. Squazzo, X. Xu, S.A. Brugmann, L.H. Goodnough, J.A. Helms, P.J. Farnham, E. Segal, and H.Y. Chang. 2007. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*. 129:1311-1323.

Risso, D., J. Ngai, T.P. Speed, and S. Dudoit. 2014. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat Biotechnol*. 32:896-902.

Roberts, G.A., B. Ozkan, I. Gachulincova, M.R. O'Dwyer, E. Hall-Ponsele, M. Saxena, P.J. Robinson, and A. Soufi. 2021. Dissecting OCT4 defines the role of nucleosome binding in pluripotency. *Nat Cell Biol*. 23:834-845.

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

Rodriguez-Meira, A., G. Buck, S.A. Clark, B.J. Povinelli, V. Alcolea, E. Louka, S. McGowan, A. Hamblin, N. Sousos, N. Barkas, A. Giustacchini, B. Psaila, S.E.W. Jacobsen, S. Thongjuea, and A.J. Mead. 2019. Unravelling Intratumoral Heterogeneity through High-Sensitivity Single-Cell Mutational Analysis and Parallel RNA Sequencing. *Mol Cell*. 73:1292-1305 e1298.

Rose, N.R., H.W. King, N.P. Blackledge, N.A. Fursova, K.J. Ember, R. Fischer, B.M. Kessler, and R.J. Klose. 2016. RYBP stimulates PRC1 to shape chromatin-based communication between Polycomb repressive complexes. *Elife*. 5.

Rothbart, S.B., B.M. Dickson, M.S. Ong, K. Krajewski, S. Houliston, D.B. Kireev, C.H. Arrowsmith, and B.D. Strahl. 2013. Multivalent histone engagement by the linked tandem Tudor and PHD domains of UHRF1 is required for the epigenetic inheritance of DNA methylation. *Genes Dev*. 27:1288-1298.

Ruan, H.B., X. Han, M.D. Li, J.P. Singh, K. Qian, S. Azarhoush, L. Zhao, A.M. Bennett, V.T. Samuel, J. Wu, J.R. Yates, 3rd, and X. Yang. 2012. O-GlcNAc transferase/host cell factor C1 complex regulates gluconeogenesis by modulating PGC-1alpha stability. *Cell Metab*. 16:226-237.

Ruthenburg, A.J., C.D. Allis, and J. Wysocka. 2007a. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol Cell*. 25:15-30.

Ruthenburg, A.J., H. Li, T.A. Milne, S. Dewell, R.K. McGinty, M. Yuen, B. Ueberheide, Y. Dou, T.W. Muir, D.J. Patel, and C.D. Allis. 2011. Recognition of a mononucleosomal histone modification pattern by BPTF via multivalent interactions. *Cell*. 145:692-706.

Ruthenburg, A.J., H. Li, D.J. Patel, and C.D. Allis. 2007b. Multivalent engagement of chromatin modifications by linked binding modules. *Nat Rev Mol Cell Biol*. 8:983-994.

Sabari, B.R., A. Dall'Agnese, A. Boija, I.A. Klein, E.L. Coffey, K. Shrinivas, B.J. Abraham, N.M. Hannett, A.V. Zamudio, J.C. Manteiga, C.H. Li, Y.E. Guo, D.S. Day, J. Schuijers, E. Vasile, S. Malik, D. Hnisz, T.I. Lee, Cisse, II, R.G. Roeder, P.A. Sharp, A.K. Chakraborty, and R.A. Young. 2018. Coactivator condensation at super-enhancers links phase separation and gene control. *Science*. 361.

Sanborn, A.L., S.S. Rao, S.C. Huang, N.C. Durand, M.H. Huntley, A.I. Jewett, I.D. Bochkov, D. Chinnappan, A. Cutkosky, J. Li, K.P. Geeting, A. Gnirke, A. Melnikov, D. McKenna, E.K. Stamenova, E.S. Lander, and E.L. Aiden. 2015. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci U S A*. 112:E6456-6465.

Sanjuan-Pla, A., C. Bueno, C. Prieto, P. Acha, R.W. Stam, R. Marschalek, and P. Menendez. 2015. Revisiting the biology of infant t(4;11)/MLL-AF4+ B-cell acute lymphoblastic leukemia. *Blood*. 126:2676-2685.

Sanulli, S., N. Justin, A. Teissandier, K. Ancelin, M. Portoso, M. Caron, A. Michaud, B. Lombard, S.T. da Rocha, J. Offer, D. Loew, N. Servant, M. Wassef, F. Burlina, S.J. Gamblin, E. Heard, and R. Margueron. 2015. Jarid2 Methylation via the PRC2 Complex Regulates H3K27me3 Deposition during Cell Differentiation. *Mol Cell*. 57:769-783.

Sasaki, K., H. Yagi, R.T. Bronson, K. Tominaga, T. Matsunashi, K. Deguchi, Y. Tani, T. Kishimoto, and T. Komori. 1996. Absence of fetal liver hematopoiesis in mice deficient in transcriptional coactivator core binding factor beta. *Proc Natl Acad Sci U S A*. 93:12359-12363.

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

Sato, T., S. Ohno, T. Hayashi, C. Sato, K. Kohu, M. Satake, and S. Habu. 2005. Dual functions of Runx proteins for reactivating CD8 and silencing CD4 at the commitment process into CD8 thymocytes. *Immunity*. 22:317-328.

Saurin, A.J., Z. Shao, H. Erdjument-Bromage, P. Tempst, and R.E. Kingston. 2001. A Drosophila Polycomb group complex includes Zeste and dTAFII proteins. *Nature*. 412:655-660.

Saurin, A.J., C. Shiels, J. Williamson, D.P. Satijn, A.P. Otte, D. Sheer, and P.S. Freemont. 1998. The human polycomb group complex associates with pericentromeric heterochromatin to form a novel nuclear domain. *J Cell Biol*. 142:887-898.

Scelfo, A., D. Fernandez-Perez, S. Tamburri, M. Zanotti, E. Lavarone, M. Soldi, T. Bonaldi, K.J. Ferrari, and D. Pasini. 2019. Functional Landscape of PCGF Proteins Reveals Both RING1A/B-Dependent-and RING1A/B-Independent-Specific Activities. *Mol Cell*. 74:1037-1052 e1037.

Schaaf, C.A., Z. Misulovin, M. Gause, A. Koenig, D.W. Gohara, A. Watson, and D. Dorsett. 2013. Cohesin and polycomb proteins functionally interact to control transcription at silenced and active genes. *PLoS Genet*. 9:e1003560.

Schaefer, E.J., H.C. Wang, C.A. Meyer, P. Cejas, M.D. Gearhart, E.R. Adelman, I. Fares, A. Apffel, K. Lim, Y. Xie, C.J. Gibson, M. Schenone, H.M. Murdock, E.S. Wang, L.P. Gondek, M.P. Carroll, R.S. Vedula, E.S. Winer, J.S. Garcia, R.M. Stone, M.R. Luskin, S.A. Carr, H.W. Long, V.J. Bardwell, M.E. Figueroa, and R.C. Lindsley. 2021. BCOR and BCORL1 mutations disrupt PRC1.1 repressive function in leukemia by unlinking the RING-PCGF1 enzymatic core from target genes. *bioRxiv*:2021.2003.2008.433705.

Scheuermann, J.C., A.G. de Ayala Alonso, K. Oktaba, N. Ly-Hartig, R.K. McGinty, S. Fraterman, M. Wilm, T.W. Muir, and J. Muller. 2010. Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. *Nature*. 465:243-247.

Schmidt, D., P.C. Schwalie, C.S. Ross-Innes, A. Hurtado, G.D. Brown, J.S. Carroll, P. Flicek, and D.T. Odom. 2010. A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res*. 20:578-588.

Schmitges, F.W., A.B. Prusty, M. Faty, A. Stutzer, G.M. Lingaraju, J. Aiwazian, R. Sack, D. Hess, L. Li, S. Zhou, R.D. Bunker, U. Wirth, T. Bouwmeester, A. Bauer, N. Ly-Hartig, K. Zhao, H. Chan, J. Gu, H. Gut, W. Fischle, J. Muller, and N.H. Thoma. 2011. Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol Cell*. 42:330-341.

Schmitt, S., M. Prestel, and R. Paro. 2005. Intergenic transcription through a polycomb group response element counteracts silencing. *Genes Dev*. 19:697-708.

Schnittger, S., F. Dicker, W. Kern, N. Wendland, J. Sundermann, T. Alpermann, C. Haferlach, and T. Haferlach. 2011. RUNX1 mutations are frequent in de novo AML with noncomplex karyotype and confer an unfavorable prognosis. *Blood*. 117:2348-2357.

Schnittger, S., T.M. Kohl, T. Haferlach, W. Kern, W. Hiddemann, K. Spiekermann, and C. Schoch. 2006. KIT-D816 mutations in AML1-ETO-positive AML are associated with impaired event-free and overall survival. *Blood*. 107:1791-1799.

Schoeftner, S., A.K. Sengupta, S. Kubicek, K. Mechtler, L. Spahn, H. Koseki, T. Jenuwein, and A. Wutz. 2006. Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *EMBO J*. 25:3110-3122.

Score, J., C. Hidalgo-Curtis, A.V. Jones, N. Winkelmann, A. Skinner, D. Ward, K. Zoi, T. Ernst, F. Stegelmann, K. Dohner, A. Chase, and N.C. Cross. 2012. Inactivation of polycomb repressive complex 2 components in myeloproliferative and myelodysplastic/myeloproliferative neoplasms. *Blood*. 119:1208-1213.

Serandour, A.A., S. Avner, F. Percevault, F. Demay, M. Bizot, C. Lucchetti-Miganeh, F. Barloy-Hubler, M. Brown, M. Lupien, R. Metivier, G. Salbert, and J. Eeckhoute. 2011. Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome Res*. 21:555-565.

Shahbazian, M.D., and M. Grunstein. 2007. Functions of site-specific histone acetylation and deacetylation. *Annu Rev Biochem*. 76:75-100.

Shao, Z., F. Raible, R. Mollaaghababa, J.R. Guyon, C.T. Wu, W. Bender, and R.E. Kingston. 1999. Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell*. 98:37-46.

Shen, T., and S. Huang. 2012. The role of Cdc25A in the regulation of cell proliferation and apoptosis. *Anticancer Agents Med Chem*. 12:631-639.

Shen, X., Y. Liu, Y.J. Hsu, Y. Fujiwara, J. Kim, X. Mao, G.C. Yuan, and S.H. Orkin. 2008. EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol Cell*. 32:491-502.

Shogren-Knaak, M., H. Ishii, J.M. Sun, M.J. Pazin, J.R. Davie, and C.L. Peterson. 2006. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science*. 311:844-847.

Silva, J., W. Mak, I. Zvetkova, R. Appanah, T.B. Nesterova, Z. Webster, A.H. Peters, T. Jenuwein, A.P. Otte, and N. Brockdorff. 2003. Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. *Dev Cell*. 4:481-495.

Simon, J., A. Chiang, and W. Bender. 1992. Ten different Polycomb group genes are required for spatial control of the abdA and AbdB homeotic products. *Development*. 114:493-505.

Simon, J., A. Chiang, W. Bender, M.J. Shimell, and M. O'Connor. 1993. Elements of the Drosophila bithorax complex that mediate repression by Polycomb group products. *Dev Biol*. 158:131-144.

Simon, L., V.P. Lavallee, M.E. Bordeleau, J. Krosl, I. Baccelli, G. Boucher, B. Lehnertz, J. Chagraoui, T. MacRae, R. Ruel, Y. Chantigny, S. Lemieux, A. Marinier, J. Hebert, and G. Sauvageau. 2017. Chemogenomic Landscape of RUNX1-mutated AML Reveals Importance of RUNX1 Allele Dosage in Genetics and Glucocorticoid Sensitivity. *Clin Cancer Res*. 23:6969-6981.

Singh, R., E. Bassett, A. Chakravarti, and M.R. Parthun. 2018. Replication-dependent histone isoforms: a new source of complexity in chromatin structure and function. *Nucleic Acids Res*. 46:8665-8678.

Slany, R.K. 2020. MLL fusion proteins and transcriptional control. *Biochim Biophys Acta Gene Regul Mech*. 1863:194503.

Slifer, E.H. 1942. A mutant stock of Drosophila with extra sex-combs. *Journal of Experimental Zoology*. 90:31-40.

Sneeringer, C.J., M.P. Scott, K.W. Kuntz, S.K. Knutson, R.M. Pollock, V.M. Richon, and R.A. Copeland. 2010. Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. *Proc Natl Acad Sci U S A*. 107:20980-20985.

Sofueva, S., E. Yaffe, W.C. Chan, D. Georgopoulou, M. Vietri Rudan, H. Mira-Bontenbal, S.M. Pollard, G.P. Schroth, A. Tanay, and S. Hadjur. 2013. Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J*. 32:3119-3129.

Song, W.J., M.G. Sullivan, R.D. Legare, S. Hutchings, X. Tan, D. Kufrin, J. Ratajczak, I.C. Resende, C. Haworth, R. Hock, M. Loh, C. Felix, D.C. Roy, L. Busque, D. Kurnit, C. Willman, A.M. Gewirtz, N.A. Speck, J.H. Bushweller, F.P. Li, K. Gardiner, M. Poncz, J.M. Maris, and D.G. Gilliland. 1999. Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nat Genet*. 23:166-175.

Sood, R., Y. Kamikubo, and P. Liu. 2017. Role of RUNX1 in hematological malignancies. *Blood*. 129:2070-2082.

Soufi, A., G. Donahue, and K.S. Zaret. 2012. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell*. 151:994-1004.

Souroullas, G.P., W.R. Jeck, J.S. Parker, J.M. Simon, J.Y. Liu, J. Paulk, J. Xiong, K.S. Clark, Y. Fedoriw, J. Qi, C.E. Burd, J.E. Bradner, and N.E. Sharpless. 2016. An oncogenic Ezh2 mutation induces tumors through global redistribution of histone 3 lysine 27 trimethylation. *Nat Med*. 22:632-640.

Spitz, F., and E.E. Furlong. 2012. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet*. 13:613-626.

Staber, P.B., P. Zhang, M. Ye, R.S. Welner, E. Levantini, A. Di Ruscio, A.K. Ebralidze, C. Bach, H. Zhang, J. Zhang, K. Vanura, R. Delwel, H. Yang, G. Huang, and D.G. Tenen. 2014. The Runx-PU.1 pathway preserves normal and AML/ETO9a leukemic stem cells. *Blood*. 124:2391-2399.

Staby, L., C. O'Shea, M. Willemoes, F. Theisen, B.B. Kragelund, and K. Skriver. 2017. Eukaryotic transcription factors: paradigms of protein intrinsic disorder. *Biochem J*. 474:2509-2532.

Stepanik, V.A., and P.J. Harte. 2012. A mutation in the E(Z) methyltransferase that increases trimethylation of histone H3 lysine 27 and causes inappropriate silencing of active Polycomb target genes. *Dev Biol*. 364:249-258.

Stern, M., R. Jensen, and I. Herskowitz. 1984. Five SWI genes are required for expression of the HO gene in yeast. *J Mol Biol*. 178:853-868.

Stielow, B., F. Finkernagel, T. Stiewe, A. Nist, and G. Suske. 2018. MGA, L3MBTL2 and E2F6 determine genomic binding of the non-canonical Polycomb repressive complex PRC1.6. *PLoS Genet*. 14:e1007193.

Stock, J.K., S. Giadrossi, M. Casanova, E. Brookes, M. Vidal, H. Koseki, N. Brockdorff, A.G. Fisher, and A. Pombo. 2007. Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat Cell Biol*. 9:1428-1435.

Stojic, L., Z. Jasencakova, C. Prezioso, A. Stutzer, B. Bodega, D. Pasini, R. Klingberg, C. Mozzetta, R. Margueron, P.L. Puri, D. Schwarzer, K. Helin, W. Fischle, and V. Orlando. 2011. Chromatin regulated interchange between polycomb repressive complex 2 (PRC2)-Ezh2 and PRC2-Ezh1 complexes controls myogenin activation in skeletal muscle cells. *Epigenetics Chromatin*. 4:16.

Struhl, G., and M. Akam. 1985. Altered distributions of Ultrabithorax transcripts in extra sex combs mutant embryos of Drosophila. *EMBO J*. 4:3259-3264.

Strutt, H., G. Cavalli, and R. Paro. 1997. Co-localization of Polycomb protein and GAGA factor on regulatory elements responsible for the maintenance of homeotic gene expression. *EMBO J*. 16:3621-3632.

Sun, C., L. Chang, and X. Zhu. 2017. Pathogenesis of ETV6/RUNX1-positive childhood acute lymphoblastic leukemia and mechanisms underlying its relapse. *Oncotarget*. 8:35445-35459.

Sun, F., C. Chronis, M. Kronenberg, X.F. Chen, T. Su, F.D. Lay, K. Plath, S.K. Kurdistani, and M.F. Carey. 2019. Promoter-Enhancer Communication Occurs Primarily within Insulated Neighborhoods. *Mol Cell*. 73:250-263 e255.

Sun, Y., B. Zhou, F. Mao, J. Xu, H. Miao, Z. Zou, L.T. Phuc Khoa, Y. Jang, S. Cai, M. Witkin, R. Koche, K. Ge, G.R. Dressler, R.L. Levine, S.A. Armstrong, Y. Dou, and J.L. Hess. 2018. HOXA9 Reprograms the Enhancer Landscape to Promote Leukemogenesis. *Cancer Cell*. 34:643-658 e645.

Suzuki, T., Y. Shimizu, E. Furuhata, S. Maeda, M. Kishima, H. Nishimura, S. Enomoto, Y. Hayashizaki, and H. Suzuki. 2017. RUNX1 regulates site specificity of DNA demethylation by recruitment of DNA demethylation machineries in hematopoietic cells. *Blood Adv*. 1:1699-1711.

Swiers, G., C. Baumann, J. O'Rourke, E. Giannoulatou, S. Taylor, A. Joshi, V. Moignard, C. Pina, T. Bee, K.D. Kokkaliaris, M. Yoshimoto, M.C. Yoder, J. Frampton, T. Schroeder, T. Enver, B. Gottgens, and M. de Bruijn. 2013. Early dynamic fate changes in haemogenic endothelium characterized at the single-cell level. *Nat Commun*. 4:2924.

Symmons, O., V.V. Uslu, T. Tsujimura, S. Ruf, S. Nassari, W. Schwarzer, L. Ettwiller, and F. Spitz. 2014. Functional and topological characteristics of mammalian regulatory domains. *Genome Res*. 24:390-400.

Szutorisz, H., A. Georgiou, L. Tora, and N. Dillon. 2006. The proteasome restricts permissive transcription at tissue-specific gene loci in embryonic stem cells. *Cell*. 127:1375-1388.

Takahashi, S., and A. Yokoyama. 2020. The molecular functions of common and atypical MLL fusion protein complexes. *Biochim Biophys Acta Gene Regul Mech*. 1863:194548.

Tamburri, S., E. Lavarone, D. Fernandez-Perez, E. Conway, M. Zanotti, D. Manganaro, and D. Pasini. 2020. Histone H2AK119 Mono-Ubiquitination Is Essential for Polycomb-Mediated Transcriptional Repression. *Mol Cell*. 77:840-856 e845.

Tang, J.L., H.A. Hou, C.Y. Chen, C.Y. Liu, W.C. Chou, M.H. Tseng, C.F. Huang, F.Y. Lee, M.C. Liu, M. Yao, S.Y. Huang, B.S. Ko, S.C. Hsu, S.J. Wu, W. Tsay, Y.C. Chen, L.I. Lin, and H.F. Tien. 2009. AML1/RUNX1 mutations in 470 adult patients with de novo acute myeloid leukemia: prognostic implication and interaction with other gene alterations. *Blood*. 114:5352-5361.

Taniuchi, I., M. Osato, T. Egawa, M.J. Sunshine, S.C. Bae, T. Komori, Y. Ito, and D.R. Littman. 2002. Differential requirements for Runx proteins in CD4 repression and epigenetic silencing during T lymphocyte development. *Cell*. 111:621-633.

Tavares, L., E. Dimitrova, D. Oxley, J. Webster, R. Poot, J. Demmers, K. Bezstarosti, S. Taylor, H. Ura, H. Koide, A. Wutz, M. Vidal, S. Elderkin, and N. Brockdorff. 2012. RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell*. 148:664-678.

Taylor-Stokes, G., D. Mitra, J. Waller, K. Gibson, G. Milligan, and S. Iyer. 2019. Treatment patterns and clinical outcomes among patients receiving palbociclib in combination with an aromatase inhibitor or fulvestrant for HR+/HER2-negative advanced/metastatic breast cancer in real-world settings in the US: Results from the IRIS study. *Breast*. 43:22-27.

Taylor, G.C., R. Eskeland, B. Hekimoglu-Balkan, M.M. Pradeepa, and W.A. Bickmore. 2013. H4K16 acetylation marks active genes and enhancers of embryonic stem cells, but does not alter chromatin compaction. *Genome Res*. 23:2053-2065.

Tee, W.W., S.S. Shen, O. Oksuz, V. Narendra, and D. Reinberg. 2014. Erk1/2 activity promotes chromatin features and RNAPII phosphorylation at developmental promoters in mouse ESCs. *Cell*. 156:678-690.

Telenius, J., T.W. Consortium, and J.R. Hughes. 2018. NGseqBasic - a single-command UNIX tool for ATAC-seq, DNaseI-seq, Cut-and-Run, and ChIP-seq data mapping, high-resolution visualisation, and quality control. *bioRxiv*:393413.

Thomson, J.P., P.J. Skene, J. Selfridge, T. Clouaire, J. Guy, S. Webb, A.R. Kerr, A. Deaton, R. Andrews, K.D. James, D.J. Turner, R. Illingworth, and A. Bird. 2010. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature*. 464:1082-1086.

Tie, F., R. Banerjee, C. Fu, C.A. Stratton, M. Fang, and P.J. Harte. 2016. Polycomb inhibits histone acetylation by CBP by binding directly to its catalytic domain. *Proc Natl Acad Sci U S A*. 113:E744-753.

Tiffen, J.C., D. Gunatilake, S.J. Gallagher, K. Gowrishankar, A. Heinemann, C. Cullinane, K. Dutton-Regester, G.M. Pupo, D. Strbenac, J.Y. Yang, J. Madore, G.J. Mann, N.K. Hayward, G.A. McArthur, F.V. Filipp, and P. Hersey. 2015. Targeting activating mutations of EZH2 leads to potent cell growth inhibition in human melanoma by derepression of tumor suppressor genes. *Oncotarget*. 6:27023-27036.

Tkachuk, D.C., S. Kohler, and M.L. Cleary. 1992. Involvement of a homolog of Drosophila trithorax by 11q23 chromosomal translocations in acute leukemias. *Cell*. 71:691-700.

Trojer, P., A.R. Cao, Z. Gao, Y. Li, J. Zhang, X. Xu, G. Li, R. Losson, H. Erdjument-Bromage, P. Tempst, P.J. Farnham, and D. Reinberg. 2011. L3MBTL2 protein acts in concert with PcG protein-mediated monoubiquitination of H2A to establish a repressive chromatin structure. *Mol Cell*. 42:438-450.

Tropberger, P., S. Pott, C. Keller, K. Kamieniarz-Gdula, M. Caron, F. Richter, G. Li, G. Mittler, E.T. Liu, M. Buhler, R. Margueron, and R. Schneider. 2013. Regulation of transcription through acetylation of H3K122 on the lateral surface of the histone octamer. *Cell*. 152:859-872.

Vakoc, C.R., D.L. Letting, N. Gheldof, T. Sawado, M.A. Bender, M. Groudine, M.J. Weiss, J. Dekker, and G.A. Blobel. 2005. Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol Cell*. 17:453-462.

Van Dyck, F., J. Declercq, C.V. Braem, and W.J. Van de Ven. 2007. PLAG1, the prototype of the PLAG gene family: versatility in tumour development (review). *Int J Oncol*. 30:765-774.

Vandenbunder, B., N. Fourre, A. Leray, F. Mueller, P. Volkel, P.O. Angrand, and L. Heliot. 2014. PRC1 components exhibit different binding kinetics in Polycomb bodies. *Biol Cell*. 106:111-125.

Vanhaecke, T., P. Papeleu, G. Elaut, and V. Rogiers. 2004. Trichostatin A-like hydroxamate histone deacetylase inhibitors as therapeutic agents: toxicological point of view. *Curr Med Chem*. 11:1629-1643.

Varambally, S., S.M. Dhanasekaran, M. Zhou, T.R. Barrette, C. Kumar-Sinha, M.G. Sanda, D. Ghosh, K.J. Pienta, R.G. Sewalt, A.P. Otte, M.A. Rubin, and A.M. Chinnaiyan. 2002. The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature*. 419:624-629.

Vella, P., I. Barozzi, A. Cuomo, T. Bonaldi, and D. Pasini. 2012. Yin Yang 1 extends the Myc-related transcription factors network in embryonic stem cells. *Nucleic Acids Res*. 40:3403-3418.

Vieux-Rochas, M., P.J. Fabre, M. Leleu, D. Duboule, and D. Noordermeer. 2015. Clustering of mammalian Hox genes with other H3K27me3 targets within an active nuclear domain. *Proc Natl Acad Sci U S A*. 112:4672-4677.

Visel, A., S. Prabhakar, J.A. Akiyama, M. Shoukry, K.D. Lewis, A. Holt, I. Plajzer-Frick, V. Afzal, E.M. Rubin, and L.A. Pennacchio. 2008. Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nat Genet*. 40:158-160.

von Hippel, P.H., and O.G. Berg. 1989. Facilitated target location in biological systems. *J Biol Chem*. 264:675-678.

Wang, C., J.E. Lee, B. Lai, T.S. Macfarlan, S. Xu, L. Zhuang, C. Liu, W. Peng, and K. Ge. 2016a. Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition. *Proc Natl Acad Sci U S A*. 113:11871-11876.

Wang, C.G., Y.J. Ye, J. Yuan, F.F. Liu, H. Zhang, and S. Wang. 2010a. EZH2 and STAT6 expression profiles are correlated with colorectal cancer stage and prognosis. *World J Gastroenterol*. 16:2421-2427.

Wang, H., L. Farnung, C. Dienemann, and P. Cramer. 2020. Structure of H3K36-methylated nucleosome-PWWP complex reveals multivalent cross-gyre binding. *Nat Struct Mol Biol*. 27:8-13.

Wang, H., L. Wang, H. Erdjument-Bromage, M. Vidal, P. Tempst, R.S. Jones, and Y. Zhang. 2004a. Role of histone H2A ubiquitination in Polycomb silencing. *Nature*. 431:873-878.

Wang, J., T. Hoshino, R.L. Redner, S. Kajigaya, and J.M. Liu. 1998. ETO, fusion partner in t(8;21) acute myeloid leukemia, represses transcription by interaction with the human N-CoR/mSin3/HDAC1 complex. *Proc Natl Acad Sci U S A*. 95:10860-10865.

Wang, L., J.L. Brown, R. Cao, Y. Zhang, J.A. Kassis, and R.S. Jones. 2004b. Hierarchical recruitment of polycomb group silencing complexes. *Mol Cell*. 14:637-646.

Wang, L., A. Gural, X.J. Sun, X. Zhao, F. Perna, G. Huang, M.A. Hatlen, L. Vu, F. Liu, H. Xu, T. Asai, H. Xu, T. Deblasio, S. Menendez, F. Voza, Y. Jiang, P.A. Cole, J. Zhang, A. Melnick, R.G. Roeder, and S.D. Nimer. 2011. The leukemogenicity of AML1-ETO is dependent on site-specific lysine acetylation. *Science*. 333:765-769.

Wang, L., P. Joshi, E.L. Miller, L. Higgins, M. Slattery, and J.A. Simon. 2018. A Role for Monomethylation of Histone H3-K27 in Gene Activity in Drosophila. *Genetics*. 208:1023-1036.

Wang, Q., T. Stacy, M. Binder, M. Marin-Padilla, A.H. Sharpe, and N.A. Speck. 1996a. Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. *Proc Natl Acad Sci U S A*. 93:3444-3449.

Wang, Q., T. Stacy, J.D. Miller, A.F. Lewis, T.L. Gu, X. Huang, J.H. Bushweller, J.C. Bories, F.W. Alt, G. Ryan, P.P. Liu, A. Wynshaw-Boris, M. Binder, M. Marin-Padilla, A.H. Sharpe, and N.A. Speck. 1996b. The CBFbeta subunit is essential for CBFalpha2 (AML1) function in vivo. *Cell*. 87:697-708.

Wang, R., A.B. Taylor, B.Z. Leal, L.V. Chadwell, U. Ilangovan, A.K. Robinson, V. Schirf, P.J. Hart, E.M. Lafer, B. Demeler, A.P. Hinck, D.G. McEwen, and C.A. Kim. 2010b. Polycomb group targeting through different binding partners of RING1B C-terminal domain. *Structure*. 18:966-975.

Wang, S., J.H. Su, B.J. Beliveau, B. Bintu, J.R. Moffitt, C.T. Wu, and X. Zhuang. 2016b. Spatial organization of chromatin domains and compartments in single chromosomes. *Science*. 353:598-602.

Wang, S., Q. Wang, B.E. Crute, I.N. Melnikova, S.R. Keller, and N.A. Speck. 1993. Cloning and characterization of subunits of the T-cell receptor and murine leukemia virus enhancer core-binding factor. *Mol Cell Biol*. 13:3324-3339.

Wang, S.W., and N.A. Speck. 1992. Purification of core-binding factor, a protein that binds the conserved core site in murine leukemia virus enhancers. *Mol Cell Biol*. 12:89-102.

Wang, W., J. Cote, Y. Xue, S. Zhou, P.A. Khavari, S.R. Biggar, C. Muchardt, G.V. Kalpana, S.P. Goff, M. Yaniv, J.L. Workman, and G.R. Crabtree. 1996c. Purification and biochemical heterogeneity of the mammalian SWI-SNF complex. *EMBO J*. 15:5370-5382.

Wang, Z., J. Song, T.A. Milne, G.G. Wang, H. Li, C.D. Allis, and D.J. Patel. 2010c. Pro isomerization in MLL1 PHD3-bromo cassette connects H3K4me readout to CyP33 and HDAC-mediated repression. *Cell*. 141:1183-1194.

Wang, Z., C. Zang, J.A. Rosenfeld, D.E. Schones, A. Barski, S. Cuddapah, K. Cui, T.Y. Roh, W. Peng, M.Q. Zhang, and K. Zhao. 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet*. 40:897-903.

Wani, A.H., A.N. Boettiger, P. Schorderet, A. Ergun, C. Munger, R.I. Sadreyev, X. Zhuang, R.E. Kingston, and N.J. Francis. 2016. Chromatin topology is coupled to Polycomb group protein subnuclear organization. *Nat Commun*. 7:10291.

Wedeen, C., K. Harding, and M. Levine. 1986. Spatial regulation of Antennapedia and bithorax gene expression by the Polycomb locus in Drosophila. *Cell*. 44:739-748.

Weikum, E.R., M.T. Knuesel, E.A. Ortlund, and K.R. Yamamoto. 2017. Glucocorticoid receptor control of transcription: precision and plasticity via allostery. *Nat Rev Mol Cell Biol*. 18:159-174.

Weintraub, A.S., C.H. Li, A.V. Zamudio, A.A. Sigova, N.M. Hannett, D.S. Day, B.J. Abraham, M.A. Cohen, B. Nabet, D.L. Buckley, Y.E. Guo, D. Hnisz, R. Jaenisch, J.E. Bradner, N.S. Gray, and R.A. Young. 2017. YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell*. 171:1573-1588 e1528.

Whyte, W.A., D.A. Orlando, D. Hnisz, B.J. Abraham, C.Y. Lin, M.H. Kagey, P.B. Rahl, T.I. Lee, and R.A. Young. 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*. 153:307-319.

Wijchers, P.J., P.H.L. Krijger, G. Geeven, Y. Zhu, A. Denker, M. Verstegen, C. Valdes-Quezada, C. Vermeulen, M. Janssen, H. Teunissen, L.C.M. Anink-Groenen, P.J. Verschure, and W. de Laat. 2016. Cause and Consequence of Tethering a SubTAD to Different Nuclear Compartments. *Mol Cell*. 61:461-473.

Wilkinson, A.C., E. Ballabio, H. Geng, P. North, M. Tapia, J. Kerry, D. Biswas, R.G. Roeder, C.D. Allis, A. Melnick, M.F. de Bruijn, and T.A. Milne. 2013. RUNX1 is a key target in t(4;11) leukemias that contributes to gene activation through an AF4-MLL complex interaction. *Cell Rep*. 3:116-127.

Wilson, N.K., S.D. Foster, X. Wang, K. Knezevic, J. Schutte, P. Kaimakis, P.M. Chilarska, S. Kinston, W.H. Ouwehand, E. Dzierzak, J.E. Pimanda, M.F. de Bruijn, and B. Gottgens. 2010.

Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell*. 7:532-544.

Worden, E.J., N.A. Hoffmann, C.W. Hicks, and C. Wolberger. 2019. Mechanism of Cross-talk between H2B Ubiquitination and H3 Methylation by Dot1L. *Cell*. 176:1490-1501 e1412.

Wu, H., H. Zeng, A. Dong, F. Li, H. He, G. Senisterra, A. Seitova, S. Duan, P.J. Brown, M. Vedadi, C.H. Arrowsmith, and M. Schapira. 2013a. Structure of the catalytic domain of EZH2 reveals conformational plasticity in cofactor and substrate binding sites and explains oncogenic mutations. *PLoS One*. 8:e83737.

Wu, P., J. Weng, M. Li, Z. Lu, C. Deng, Q. Sun, R. Xu, S. Geng, and X. Du. 2019. Co-occurrence of RUNX1 and ASXL1 mutations underlie poor response and outcome for MDS patients treated with HMAs. *Am J Transl Res*. 11:3651-3658.

Wu, T., Y.F. Kamikawa, and M.E. Donohoe. 2018. Brd4's Bromodomains Mediate Histone H3 Acetylation and Chromatin Remodeling in Pluripotent Cells through P300 and Brg1. *Cell Rep*. 25:1756-1771.

Wu, X., J.V. Johansen, and K. Helin. 2013b. Fbxl10/Kdm2b recruits polycomb repressive complex 1 to CpG islands and regulates H2A ubiquitylation. *Mol Cell*. 49:1134-1146.

Xie, L., S.E. Torigoe, J. Xiao, D.H. Mai, L. Li, F.P. Davis, P. Dong, H. Marie-Nelly, J. Grimm, L. Lavis, X. Darzacq, C. Cattoglio, Z. Liu, and R. Tjian. 2017. A dynamic interplay of enhancer elements regulates Klf4 expression in naive pluripotency. *Genes Dev*. 31:1795-1808.

Xu, B., D.M. On, A. Ma, T. Parton, K.D. Konze, S.G. Pattenden, D.F. Allison, L. Cai, S. Rockowitz, S. Liu, Y. Liu, F. Li, M. Vedadi, S.V. Frye, B.A. Garcia, D. Zheng, J. Jin, and G.G. Wang. 2015. Selective inhibition of EZH2 and EZH1 enzymatic activity by a small molecule suppresses MLL-rearranged leukemia. *Blood*. 125:346-357.

Xu, C., K. Liu, M. Lei, A. Yang, Y. Li, T.R. Hughes, and J. Min. 2018. DNA Sequence Recognition of Human CXXC Domains and Their Structural Determinants. *Structure*. 26:85-95 e83.

Xu, K., Z.J. Wu, A.C. Groner, H.H. He, C. Cai, R.T. Lis, X. Wu, E.C. Stack, M. Loda, T. Liu, H. Xu, L. Cato, J.E. Thornton, R.I. Gregory, C. Morrissey, R.L. Vessella, R. Montironi, C. Magi-Galluzzi, P.W. Kantoff, S.P. Balk, X.S. Liu, and M. Brown. 2012. EZH2 oncogenic activity in castration-resistant prostate cancer cells is Polycomb-independent. *Science*. 338:1465-1469.

Yamagishi, M., M. Hori, D. Fujikawa, T. Ohsugi, D. Honma, N. Adachi, H. Katano, T. Hishima, S. Kobayashi, K. Nakano, M. Nakashima, M. Iwanaga, A. Utsunomiya, Y. Tanaka, S. Okada, K. Tsukasaki, K. Tobinai, K. Araki, T. Watanabe, and K. Uchimaru. 2019. Targeting Excessive EZH1 and EZH2 Activities for Abnormal Histone Methylation and Transcription Network in Malignant Lymphomas. *Cell Rep*. 29:2321-2337 e2327.

Yang, H., S. Kurtenbach, Y. Guo, I. Lohse, M.A. Durante, J. Li, Z. Li, H. Al-Ali, L. Li, Z. Chen, M.G. Field, P. Zhang, S. Chen, S. Yamamoto, Z. Li, Y. Zhou, S.D. Nimer, J.W. Harbour, C. Wahlestedt, M. Xu, and F.C. Yang. 2018. Gain of function of ASXL1 truncating protein in the pathogenesis of myeloid malignancies. *Blood*. 131:328-341.

Yang, X., and K. Qian. 2017. Protein O-GlcNAcylation: emerging mechanisms and functions. *Nat Rev Mol Cell Biol*. 18:452-465.

Yang, Z., J.H. Yik, R. Chen, N. He, M.K. Jang, K. Ozato, and Q. Zhou. 2005. Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Mol Cell*. 19:535-545.

Yap, D.B., J. Chu, T. Berg, M. Schapira, S.W. Cheng, A. Moradian, R.D. Morin, A.J. Mungall, B. Meissner, M. Boyle, V.E. Marquez, M.A. Marra, R.D. Gascoyne, R.K. Humphries, C.H. Arrowsmith, G.B. Morin, and S.A. Aparicio. 2011. Somatic mutations at EZH2 Y641 act dominantly through a mechanism of selectively altered PRC2 catalytic activity, to increase H3K27 trimethylation. *Blood*. 117:2451-2459.

Yergeau, D.A., C.J. Hetherington, Q. Wang, P. Zhang, A.H. Sharpe, M. Binder, M. Marin-Padilla, D.G. Tenen, N.A. Speck, and D.E. Zhang. 1997. Embryonic lethality and impairment of haematopoiesis in mice heterozygous for an AML1-ETO fusion gene. *Nat Genet*. 15:303-306.

Yin, J.A., M.A. O'Brien, R.K. Hills, S.B. Daly, K. Wheatley, and A.K. Burnett. 2012. Minimal residual disease monitoring by quantitative RT-PCR in core binding factor AML allows risk stratification and predicts relapse: results of the United Kingdom MRC AML-15 trial. *Blood*. 120:2826-2835.

Yin, Y., E. Morgunova, A. Jolma, E. Kaasinen, B. Sahu, S. Khund-Sayeed, P.K. Das, T. Kivioja, K. Dave, F. Zhong, K.R. Nitta, M. Taipale, A. Popov, P.A. Ginno, S. Domcke, J. Yan, D. Schubeler, C. Vinson, and J. Taipale. 2017. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*. 356.

Yokoyama, A., and M.L. Cleary. 2008. Menin critically links MLL proteins with LEDGF on cancer-associated target genes. *Cancer Cell*. 14:36-46.

Yokoyama, A., M. Lin, A. Naresh, I. Kitabayashi, and M.L. Cleary. 2010. A higher-order complex containing AF4 and ENL family proteins with P-TEFb facilitates oncogenic and physiologic MLL-dependent transcription. *Cancer Cell*. 17:198-212.

Yokoyama, A., T.C. Somervaille, K.S. Smith, O. Rozenblatt-Rosen, M. Meyerson, and M.L. Cleary. 2005. The menin tumor suppressor protein is an essential oncogenic cofactor for MLL-associated leukemogenesis. *Cell*. 123:207-218.

Yu, B.D., J.L. Hess, S.E. Horning, G.A. Brown, and S.J. Korsmeyer. 1995. Altered Hox expression and segmental identity in Mll-mutant mice. *Nature*. 378:505-508.

Yu, M., T. Mazor, H. Huang, H.T. Huang, K.L. Kathrein, A.J. Woo, C.R. Chouinard, A. Labadorf, T.E. Akie, T.B. Moran, H. Xie, S. Zacharek, I. Taniuchi, R.G. Roeder, C.F. Kim, L.I. Zon, E. Fraenkel, and A.B. Cantor. 2012. Direct recruitment of polycomb repressive complex 1 to chromatin by core binding transcription factors. *Mol Cell*. 45:330-343.

Yuan, W., M. Xu, C. Huang, N. Liu, S. Chen, and B. Zhu. 2011. H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *J Biol Chem*. 286:7983-7989.

Yuan, Y., L. Zhou, T. Miyamoto, H. Iwasaki, N. Harakawa, C.J. Hetherington, S.A. Burel, E. Lagasse, I.L. Weissman, K. Akashi, and D.E. Zhang. 2001. AML1-ETO expression is directly involved in the development of acute myeloid leukemia in the presence of additional mutations. *Proc Natl Acad Sci U S A*. 98:10398-10403.

Yun, M., J. Wu, J.L. Workman, and B. Li. 2011. Readers of histone modifications. *Cell Res*. 21:564-578.

Zaret, K.S. 2020. Pioneer Transcription Factors Initiating Gene Network Changes. *Annu Rev Genet*. 54:367-385.

Zeleznik-Le, N.J., A.M. Harden, and J.D. Rowley. 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. *Proc Natl Acad Sci U S A*. 91:10610-10614.

Zhang, B.L., D. Li, Y.L. Gong, Y. Huang, D.Y. Qin, L. Jiang, X. Liang, X. Yang, H.F. Gou, Y.S. Wang, Y.Q. Wei, and W. Wang. 2019. Preclinical Evaluation of Chimeric Antigen Receptor-Modified T Cells Specific to Epithelial Cell Adhesion Molecule for Treating Colorectal Cancer. *Hum Gene Ther*. 30:402-412.

Zhang, J., L. Ding, L. Holmfeldt, G. Wu, S.L. Heatley, D. Payne-Turner, J. Easton, X. Chen, J. Wang, M. Rusch, C. Lu, S.C. Chen, L. Wei, J.R. Collins-Underwood, J. Ma, K.G. Roberts, S.B. Pounds, A. Ulyanov, J. Becksfort, P. Gupta, R. Huether, R.W. Kriwacki, M. Parker, D.J. McGoldrick, D. Zhao, D. Alford, S. Espy, K.C. Bobba, G. Song, D. Pei, C. Cheng, S. Roberts, M.I. Barbato, D. Campana, E. Coustan-Smith, S.A. Shurtleff, S.C. Raimondi, M. Kleppe, J. Cools, K.A. Shimano, M.L. Hermiston, S. Doulatov, K. Eppert, E. Laurenti, F. Notta, J.E. Dick, G. Basso, S.P. Hunger, M.L. Loh, M. Devidas, B. Wood, S. Winter, K.P. Dunsmore, R.S. Fulton, L.L. Fulton, X. Hong, C.C. Harris, D.J. Dooling, K. Ochoa, K.J. Johnson, J.C. Obenauer, W.E. Evans, C.H. Pui, C.W. Naeve, T.J. Ley, E.R. Mardis, R.K. Wilson, J.R. Downing, and C.G. Mullighan. 2012. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature*. 481:157-163.

Zhang, Y., T. Liu, F. Yuan, L. Garcia-Martinez, K.D. Lee, S. Stransky, S. Sidoli, R.E. Verdun, Y. Zhang, Z. Wang, and L. Morey. 2021. The Polycomb protein RING1B enables estrogen-mediated gene expression by promoting enhancer-promoter interaction and R-loop formation. *Nucleic Acids Res*.

Zhao, D., X. Zhang, H. Guan, X. Xiong, X. Shi, H. Deng, and H. Li. 2016. The BAH domain of BAHD1 is a histone H3K27me3 reader. *Protein Cell*. 7:222-226.

Zheng, X., X. Fan, B. Fu, M. Zheng, A. Zhang, K. Zhong, J. Yan, R. Sun, Z. Tian, and H. Wei. 2017. EpCAM Inhibition Sensitizes Chemoresistant Leukemia to Immune Surveillance. *Cancer Res*. 77:482-493.

Zhou, Y., P. Wen, M. Li, Y. Li, and X.A. Li. 2019. Construction of chimeric antigen receptormodified T cells targeting EpCAM and assessment of their antitumor effect on cancer cells. *Mol Med Rep*. 20:2355-2364.

Zink, D., and R. Paro. 1995. Drosophila Polycomb-group regulated chromatin inhibits the accessibility of a trans-activator to its target DNA. *EMBO J*. 14:5660-5671.