Genome-wide Analyses Reveal New Insights into MLL Fusion Protein Recruitment Mechanisms

JONATHAN KERRY

Green Templeton College



A thesis submitted to the Medical Sciences Division of the University of Oxford in partial fulfilment of the requirements for the Degree of Doctor of Philosophy

Michaelmas Term 2015

Molecular Haematology Unit
The Weatherall Institute of Molecular Medicine
University of Oxford

Abstract

Genome-wide Analyses Reveal New Insights into MLL Fusion Protein Recruitment Mechanisms

Jonathan Kerry D.Phil. Clinical Laboratory Sciences

Green Templeton College Michaelmas Term 2015

Leukaemias arising from rearrangements of the *MLL1* gene are responsible for the majority of infant leukaemias as well as a proportion of childhood and adult leukaemias. The aggressive nature and resistance to therapy of these leukaemias has made them the focus of many therapeutic studies. Targeted therapies for MLL-rearranged (MLL-r) leukaemias are starting to enter clinical trials in an attempt to improve on outcomes from more standard treatments such as chemotherapy.

MLL fusion proteins (MLL-FPs) are recruited to gene targets where they cause aberrant upregulation of gene expression, leading to leukaemia development. Therefore, targeting the recruitment of MLL-FPs to gene targets presents one possibility for a therapeutic strategy. Before this can be achieved a thorough understanding of MLL-FP recruitment is required.

Data presented in this thesis tests some of the current models of MLL-FP recruitment on a genome-wide level. This study shows that the Menin/LEDGF complex is a good candidate for global MLL-FP recruitment to gene targets. However, data here shows that MLL-FPs can also recruit Menin, suggesting that models up to this point have been incomplete with regards to the Menin-MLL-FP interaction. It is shown here that the PAF elongation complex is unable to recruit MLL-FPs unlike previously thought. Investigation into genes marked by broad, or “spreading”, domains of MLL-FP recruitment shows that these genes have increased expression and correlate with poor survival outcomes in MLL-r leukaemia patients. Analysis of the mechanism behind MLL-FP spreading recruitment implicates the Menin/LEDGF and super elongation complex as important for this process.

Acknowledgments

I would like to thank my supervisor, Tom Milne, for his teaching and support throughout the four years of this D.Phil. project. His enthusiasm for science has been a real encouragement throughout my time in his lab. Special thanks go to other members of the lab: Erica Ballabio, Phill North, Marta Tapia, Ross Thorne, Laura Godfrey and I-Jun Lau for their advice on many scientific techniques and ideas relating to this project.

My gratitude goes to inspirational scientists in Rob Klose's lab, both past and present, who have helped with techniques used in the project, particularly Neil Blackledge and Hannah Long. Thanks also go to Huimin Geng for her collaboration with patient studies.
I would like to thank the Medical Research Council (MRC), the Nuffield Department of Clinical Laboratory Sciences (NDCLS) and the Weatherall Institute of Molecular Medicine (WIMM) for providing funding that has made this D.Phil. possible.

I am especially grateful to friends I have made during my time at the WIMM and in Oxford for many fun and enjoyable times outside of the lab. In particular I would like to thank Erica Bello for her critical reading of this thesis. Finally I would like to thank my family, particularly my parents, for their continued support and attempts to understand the work I have been doing during these four years.

Contents

Abstract 2
Acknowledgments 3
Contents 4
List of Figures 8
List of Tables 9
1. Chapter One - Introduction 10
1.1. MLL1 in Normal Development 11
1.2. MLL-rearranged Leukaemias 12
1.2.1. Acute Lymphoid Leukaemia 12
1.2.2. Acute Myeloid Leukaemia 14
1.3. Transcription Proteins 14
1.3.1. Transcription Initiation 14
1.3.2. Transcription Elongation 15
1.3.3. Interference by MLL Fusion Proteins 16
1.4. Chromatin Marks 18
1.4.1. H3K79me2 18
1.4.2. H3K4me3 20
1.4.3. H3K27me3 23
1.4.4. Other Chromatin Marks 24
1.4.4.1. H3K36me2/3 24
1.4.4.2. H3K4me1 / H3K27ac 25
1.4.4.3. DNA Methylation 25
1.4.5. CpG Islands 26
1.4.6. Super Enhancers 27
1.5. Gene Targets of MLL Fusion Proteins 27
1.6. Recruitment of Chromatin Proteins 29
1.6.1. COMPASS Proteins 31
1.6.2. PcG Proteins 32
1.6.3. Competition Between COMPASS and PcG Proteins 33
1.6.4. Interacting Proteins 34
1.7. Recruitment of MLL Fusion Proteins 34
1.7.1. Menin/LEDGF 35
1.7.2. CXXC Domain 39
1.7.3. PAFc 43
1.8. Genome-wide approaches 44
1.9. Aim of this D.Phil 46
2. Chapter Two – Materials and Methods 48
2.1. Mammalian Cell Culture 48

2.1.1. Media 48
2.2. Cell Lines and Maintenance 48
2.2.1. Cell Lines 48
Table 2.1 Cell Lines 48
2.2.2. Mycoplasma Testing 49
2.2.3. Storage and Recovery of Cell Lines 50
2.3. Transfection of Mammalian Cells 50
2.3.1. DNA Transfections 50
2.3.1.1. Lipofectamine 2000 50
2.3.1.2. PEI 51
2.3.1.3. Transit-LT1 51
2.3.1.4. MEF2 Nucleofector 51
2.3.2. Retroviral Transductions 52
2.4. Generation of Stable Cell Lines 52
2.5. Protein Techniques 53
2.5.1. Whole Cell Extracts 53
2.5.1.1. BC300 Buffer 53
2.5.2. Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE) 53
2.5.2.1. Transfer Buffer 54
2.5.3. Western Blotting (WB) 54
Table 2.2 Antibodies 55
2.5.4. Immunoprecipitation (IP) 56
2.5.5. Chromatin Immunoprecipitation (ChIP) 56
2.5.5.1. SDS Lysis Buffer / Elution Buffer 58
2.5.5.2. ChIP Dilution Buffer (CDB) 58
2.5.5.3. RIPA Wash Buffer 58
2.5.6. Biotinylated CXXC Affinity Purification (Bio-CAP) 58
2.5.6.1. CAP Buffer 59
2.6. Nucleic Acid Techniques 59
2.6.1. RNA Extraction 59
2.6.2. Generation of cDNA 59
2.6.2.1. Reverse Transcription Reaction Master Mix 59
2.6.3. RT-qPCR 60
Table 2.3 RT-qPCR Run Conditions 60
2.6.3.1. SYBR Green Reaction Mix 60
2.6.3.2. Taqman Reaction Mix 60
2.6.3.3. Calculations 60
2.6.3.4. Primer Efficiency 61
Table 2.4 Primers 62

2.7. Cloning 64
2.7.1. Polymerase Chain Reaction (PCR) 64
Table 2.5 KAPA HiFi HotStart PCR Conditions 64
2.7.2. Restriction Digests 64
2.7.3. InFusion Cloning 65
2.7.4. Ligation-Independent Cloning (LIC) 65
Table 2.6 T4 DNA Polymerase Reaction 66
Table 2.7 Vectors 66
2.7.5. Bacterial Transformations 66
2.7.6. Plasmid Amplification and Sequencing 67
2.8. ChIP-seq Analysis 68
2.8.1. Handling Raw Data 68
2.8.1.1. Mapping *fastq* files 68
2.8.1.2. Peak-calling 68
2.8.1.3. Visualisation 69
2.8.2. Genome-wide Analyses 69
2.8.2.1. Heat Maps 69
2.8.2.2. Peak/Gene Set Comparisons 69
2.8.2.3. Composite Read Plots 70
2.8.2.4. Expression analysis 70
Table 2.8 ChIP/RNA-seq 71
Table 2.9 Self-written *Perl* scripts for computational analyses 73
2.9. Patient Data 73
2.10. Statistics 74
2.10.1. ChIP-seq Read Correlations (Figure 3.5F) 74
2.10.2. Gene expression of spreading MLL-AF4 (Figure 5.2A) 74
2.10.3. Survival Analysis, Kaplan-Meier Curves (Figure 5.2C) 75
3. Chapter Three – MLL Fusion Proteins Exhibit Unique Genome-wide Binding Profiles That Co-localise with Interacting Factors 76
3.1. Introduction 76
3.2. MLL Fusion Proteins Bind at a Limited Subset of Unmethylated CpG Islands 78
3.3. MLL Fusion Proteins Bind at Active Epigenetic Landscapes 86
3.4. Binding of the Menin/LEDGF Complex Correlates with MLL-FPs 91
3.5. The PAF Elongation Complex Shows No Specific Correlation with MLL-FPs 97
3.6. Discussion 101
4. Chapter Four – The CXXC Domain is the Primary Recruitment Module of MLL Fusion Proteins and Can Function in the Absence of Other Complexes 103
4.1. Introduction 103
4.2. The CXXC Domain is Sufficient to Bind at all Active Promoter uCGIs 104

4.3. Interaction with the Menin/LEDGF Complex Confers Bi-Directional Recruitment 108
4.4. The Menin/LEDGF Complex is not Necessary for MLL-AF4 Recruitment..... 112
4.5. PAFc is Unable to Recruit MLL-AF4..... 115
4.6. Discussion 115
5. Chapter Five – Spreading MLL Fusion Proteins Mark a Distinct Subset of Genes that are Important in Leukaemia 120
5.1. Introduction..... 120
5.2. Spreading MLL-FPs Define a Unique Class of Gene Targets..... 120
5.3. Spreading MLL-FPs Overlap with Members of the SEC and Menin/LEDGF Complex..... 125
5.4. MLL-FP Spreading does not Occur Across Large Non-CGI Regions 127
5.5. Domains Surrounding the CXXC Domain are Necessary for Spreading Across uCGIs 129
5.6. Spreading MLL-AF4 Overlaps with Other Known Broad Domains..... 133
5.7. Common Targets of Spreading MLL-FPs are Associated with Poor Prognosis in MLL-r Leukaemia Patients..... 137
5.8. Discussion 138
6. Chapter Six - General Discussion..... 140
References..... 148

List of Figures

Figure 1.1	Simplified structures of MLL1 and MLL-FPs	13
Figure 1.2	Simplified model of transcription	17
Figure 1.3	Recruitment of chromatin proteins	30
Figure 3.1	Bio-CAP and ChIP optimisation	80
Figure 3.2	Genome-wide recruitment of MLL-FPs occurs at uCGIs	83
Figure 3.3	KDM2B and CFP1 are recruited to the majority of uCGIs	86
Figure 3.4	MLL-FPs are recruited to active loci	89
Figure 3.5	The Menin/LEDGF complex co-localises with MLL-FPs	95
Figure 3.6	MLL-FPs show non-specific overlap with PAFc	99
Figure 3.7	KDM2B and CFP1 overlap analyses in SEM cells	100
Figure 3.8	A model for MLL-AF4 recruitment to uCGIs	103
Figure 4.1	The DNMT region cannot recruit Menin or PAF1	106
Figure 4.2	The DNMT region is recruited to all H3K4me3-marked uCGIs	108
Figure 4.3	Menin and MLL-AF4 show bi-directional recruitment	112
Figure 4.4	The Menin/LEDGF complex is not necessary for MLL-AF4 recruitment	115
Figure 4.5	PAF1 and MLL-AF4 cannot recruit each other	117
Figure 4.6	Schematic of MLL-AF4, Menin and PAF1 recruitment potentials	118
Figure 4.7	A model for MLL-FP recruitment	120
Figure 5.1	MLL-FPs show spreading recruitment	123
Figure 5.2	Gene targets of spreading MLL-AF4 have increased expression	125
Figure 5.3	MLL-FP spreading recruitment correlates with members of the Menin/LEDGF complex and SEC	127
Figure 5.4	MLL-AF4 does not spread between distal uCGIs	129
Figure 5.5	The isolated CXXC domain of MLL-FPs does not show spreading recruitment	132
Figure 5.6	Spreading MLL-FP targets are a subset of genes marked by broad H3K4me3	136
Figure 5.7	Spreading recruitment is common amongst MLL-FPs	140
Figure 6.1	Proposed mechanisms of spreading recruitment	147

List of Tables

Table 2.1	Cell Lines	50
Table 2.2	Antibodies	56
Table 2.3	RT-qPCR Run Conditions	61
Table 2.4	Primers	63
Table 2.5	KAPA HiFi HotStart PCR Conditions	65
Table 2.6	T4 DNA Polymerase Reaction	67
Table 2.7	Vectors	67
Table 2.8	ChIP/RNA-seq	72
Table 2.9	Self-written *Perl* scripts for computational analyses	74

1. Chapter One - Introduction

Rearrangements of the *Mixed Lineage Leukaemia 1* (*MLL1*) gene cause the majority of infant leukaemias and a proportion of childhood and adult leukaemias. The poor survival rate and resistance to therapy of these leukaemias has brought them into sharp focus, requiring a thorough understanding of the disease mechanism for more effective ways to treat them. Recent studies on small molecule inhibitors targeting proteins that interact with MLL fusion proteins (MLL-FPs) have shown promising results and are now candidates for clinical trials. MLL-FPs cause misregulation of key genes in haematopoiesis, resulting in the leukaemia. An attractive therapeutic strategy is to target recruitment of MLL-FPs to these genes, preventing the subsequent misregulation. Many studies have investigated the recruitment of MLL-FPs at a localised level with one or two starting to assess the mechanism genome-wide. Furthermore studies have not been definitive towards determining the recruitment mechanism, often making several assumptions.

The aim of this DPhil was to take a comprehensive genome-wide approach of MLL-FP recruitment and build on models proposed from recent studies. Using ChIP-sequencing, it was shown that all MLL-FP recruitment events occur at unmethylated CpG islands (uCGIs), with the Menin/Lens Epithelium-Derived Growth Factor (LEDGF) complex also binding at these sites. An isolated recruitment system was used to show that Menin can recruit MLL-FPs as previously proposed but the reverse is also true, showing that the system is not necessarily hierarchical as originally thought. Importantly, whereas previous studies have implicated the Polymerase Associated Factor elongation complex (PAFc) in MLL-FP recruitment, data in this thesis argues against a direct role for PAFc. At certain MLL-FP target sites the fusion protein showed broad binding events, or “spreading”, into the gene body. These genes appear to have an important role in MLL-

rearranged (MLL-r) disease progression and potentially represent a minimal subset of targets that should be the focus for therapeutic targeting. Studies showed that the N- and C-termini of MLL-FPs, facilitates this spreading. These portions of the fusion protein can interact with the Menin/LEDGF and Super Elongation Complex, respectively, potentially implicating these complexes in spreading recruitment.

1.1. MLL1 in Normal Development

MLL1 (Mll1 in mice) is one of two human homologues of *Drosophila* Trithorax (Trx) and has a predominant role in methylating lysine 4 of histone H3 (H3K4) (see section 1.4.2) specifically at developmental *Hox* genes, with loss of Mll1 resulting in no change to bulk H3K4 methylation levels (Milne et al., 2002; Nakamura et al., 2002; Wang et al., 2009). *Mll1* mutant mice showed specific developmental axial defects and retarded growth (Yu et al., 1995). In this context, the activity of Mll1 maintained the expression of already active *Hox* genes (Yu et al., 1998). Conditional Mll1 knockout mouse models showed that Mll1 is necessary for the maintenance of HSCs and early lineage progenitors in adult haematopoiesis (Jude et al., 2007; McMahon et al., 2007). Loss of Mll1 resulted in exhaustion of the HSC compartment and an inability to compete with normal HSCs even in a normal bone marrow environment. Specifically, loss of Mll1 induced cell cycling of HSCs, depleting the quiescent pool of cells, with a similar result observed in fetal haematopoiesis. Early lineage progenitors exhibited reduced proliferation whereas Mll1 was not required for maintenance or differentiation of mature blood cell types. Interestingly however, deleting the SET domain of Mll1, and thus its H3K4 methyltransferase activity, had only a minor effect on embryogenesis (Terranova et al., 2006) and almost no effect on haematopoiesis (Mishra et al., 2014). See section 1.4.2 for a more thorough discussion of the molecular role of MLL1.

1.2. MLL-rearranged Leukaemias

Balanced translocations of the *MLL1* gene can occur with over 100 different partner genes. However, the majority of MLL-r leukaemias are caused by one of six common MLL-FPs: MLL-AF4, MLL-AF6, MLL-AF9, MLL-AF10, MLL-ENL and MLL-ELL (Meyer et al., 2013). These six MLL-FPs retain the N-terminal half of MLL1, including the DNA-binding CXXC domain and AT hooks of MLL1 (Figure 1.1). The interface of MLL1 that interacts with the proteins Menin and LEDGF is also retained in MLL-FPs. Menin is a tumour suppressor protein that has roles in cell cycle regulation and DNA damage repair (Gallo et al., 2010; Kim et al., 2003; Kim et al., 1999; Milne et al., 2005a). Menin binds at the N-terminus of MLL-FPs and serves as a bridge between the fusion protein and LEDGF (Yokoyama and Cleary, 2008). LEDGF binds directly to chromatin via an N-terminal PWWP domain and has been shown to direct sites of HIV-1 DNA integration (Shun et al., 2008). The roles of Menin and LEDGF in MLL-FP-mediated leukaemogenesis are discussed at length in section 1.7. Translocations that retain the N-terminal half of MLL1 lose the C-terminal half which contains the Plant Homeodomain (PHD) fingers, a bromodomain and the SET domain responsible for the HMT activity of MLL1.

1.2.1. Acute Lymphoid Leukaemia

Infant (<1 year old) acute lymphoid leukaemias (ALL) constitute 5% of childhood (0-18 years) ALL, with 70% of infant ALL cases resulting from an MLL rearrangement (Muntean and Hess, 2012). Whereas event-free (EFS) and overall survival (OS) are around 85% and 92%, respectively, for childhood ALL patients (Conter et al., 2010; Hunger et al., 2012; Pui et al., 2009; Vora et al., 2013), EFS and OS remain significantly lower for infant ALL patients, at around 45% and 53% (Dreyer et al., 2015; Hilden et al., 2006; Pieters et al., 2007; Tomizawa et al., 2007).

Figure 1.1

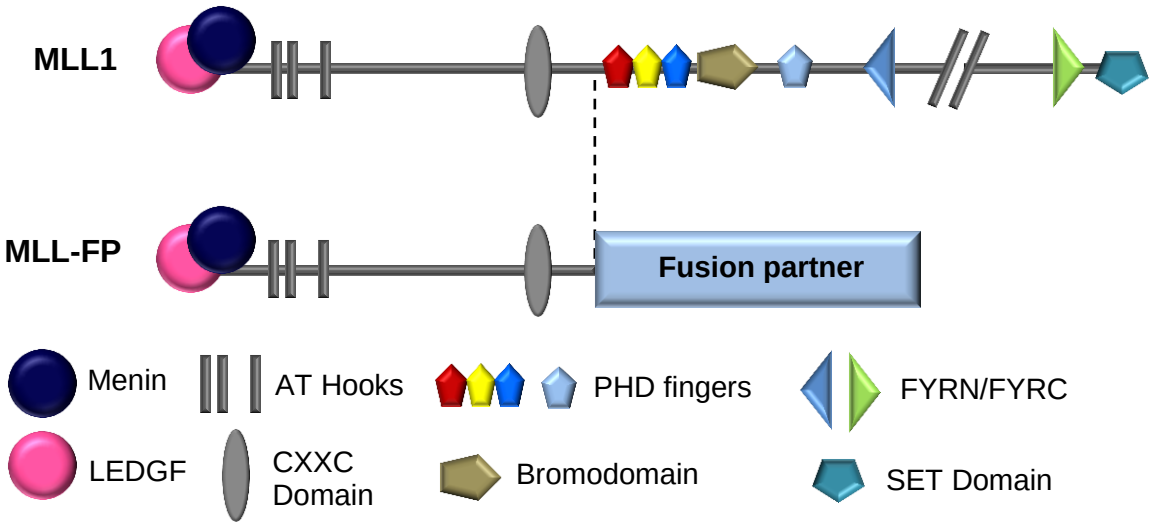


Figure 1.1 Simplified structures of MLL1 and MLL-FPs (adapted from Milne et al. 2010): Translocations of MLL1 result in MLL fusion proteins that retain the DNA-binding CXXC domain and AT hooks of MLL1. MLL-FPs retain the interaction with the proteins Menin and LEDGF. The C-terminal half of MLL1 is replaced by a fusion partner, commonly AF4, AF6, AF9, AF10, ENL or ELL.

The low survival rate of infant ALL patients is likely due to the high incidence of MLL-r leukaemias amongst these cases, which are aggressive and have an increased resistance to chemotherapy. Infant ALL patients with an MLL rearrangement have an EFS of around 20% (Pui et al., 2003). MLL rearrangements also cause ALL in children above 1 year old and in adults, accounting for approximately 10% of these cases. Among the six most common MLL1 fusions, MLL-AF4 is the major cause of ALL in both children and adults (Meyer et al., 2013). This fusion protein in particular is associated with a poor prognosis (Pui et al., 2003). MLL-AF9 and MLL-ENL have also been observed in a number of infant ALL cases (Meyer et al., 2013).

1.2.2. Acute Myeloid Leukaemia

MLL-FPs are responsible for approximately 35% to 50% of infant acute myeloid leukaemias (AML) (Muntean and Hess, 2012). In childhood MLL-r AML patients, MLL-AF9 is the most common cause of the leukaemia, followed by MLL-AF10. MLL-AF9 is the main cause of MLL-r AML in adults, followed by MLL-AF6 (Meyer et al., 2013). Patients with the MLL-AF9 translocation have the most favourable prognosis amongst MLL-r cases (Rubnitz et al., 2002). MLL-AF6 leukaemia patients often have the poorest clinical outcome, with EFS between 11% and 23.3% and an OS of 22% in cases of childhood AML (Balgobind et al., 2009; Pigazzi et al., 2011).

The aggressive nature of MLL-r leukaemias and their high incidence among childhood leukaemias demands an understanding of the underlying molecular mechanisms of oncogenesis in order to develop effective treatments.

1.3. Transcription Proteins

All cellular physiology is driven by transcription of genes and the subsequent RNA or proteins that arise. Cancers arise as a result of gene mutations that ultimately disturb the carefully regulated transcriptional machinery. MLL-FPs cause aberrant gene regulation by interacting with proteins involved in transcription initiation and elongation, resulting in leukaemia development. Therefore, an understanding of the functions of proteins involved in transcription is important to understanding the molecular mechanisms of MLL-r leukaemogenesis.

1.3.1. Transcription Initiation

For transcription to occur, RNA Polymerase II (RNAPII) is recruited to gene promoters along with auxiliary factors that form the pre-initiation complex (PIC) and regulate RNAPII activity. Initially, DNA sequence-specific transcription factors (TF) bind at promoters and make contact with the Mediator complex, a member of the PIC (Figure

1.2) (Holstege et al., 1998; Myers et al., 1999). Mediator can recruit RNAPII and other members of the PIC, including the general transcription factors (GTFs), to the promoter through direct interactions (Baek et al., 2006; Esnault et al., 2008; Jishage et al., 2012; Johnson and Carey, 2003; Seizl et al., 2011). Transcription from promoters is initiated when RNAPII is phosphorylated on Serine-5 (Ser5) of its C-Terminal Domain (CTD) by the GTF TFIIH, disrupting its interaction with Mediator (Svejstrup et al., 1997).

1.3.2. Transcription Elongation

RNAPII pauses after transcribing 30-60 nucleotides downstream from initiation (Core et al., 2008a; Seila et al., 2008). Recruitment of Negative Elongation Factor (NELF) and DRB-Sensitivity Inducing Factor (DSIF) to the site of RNAPII sets up this pausing which acts as the major rate-limiting step in transcription (Wada et al., 1998; Yamaguchi et al., 1999). The Positive Transcription Elongation Factor B (P-TEFb) complex is recruited to sites of transcription by the chromatin-binding Bromodomain Protein 4 (BRD4), and mediates the transition from paused to productively elongating RNAPII (Jang et al., 2005; Jonkers et al., 2014; Rahl et al., 2010; Yang et al., 2005). P-TEFb phosphorylates Ser2 on RNAPII CTD as well as phosphorylating NELF and DSIF, thereby relieving the inhibition on RNAPII (Fujinaga et al., 2004; Garriga and Graña, 2004; Kim and Sharp, 2001; Price, 2000; Yamada et al., 2006). The Polymerase Associated Factor complex (PAFc) is also required for full phosphorylation of Ser2 on the RNAPII CTD (Mueller et al., 2004). PAFc is composed of six subunits in mammals and associates with BRD4 and RNAPII at actively transcribed genes (Dawson et al., 2011; Kim et al., 2010). P-TEFb is part of the larger Super Elongation Complex (SEC) or AF4 family/ENL family/P-TEFb (AEP) complex that contains the elongation factors ELL1, ELL2 and ELL3, and the common MLL fusion partners AF9, ENL, AF4 and AFF4 (Lin et al., 2010; Yokoyama et al., 2010). The SEC is enriched at actively transcribed genes and is necessary for the induction of transcription (Lin et al., 2011; Lin et al., 2010). The positive effect on transcription is mediated by the ELL proteins and P-TEFb, with other proteins, such as

AFF4 stabilising the complex (Lin et al., 2010). Subunit 26 of the Mediator complex (MED26) interacts with the SEC and aids RNAPII transcription past the +1 nucleosome, highlighting a role for Mediator in elongation as well as initiation (Nock et al., 2012; Takahashi et al., 2011).

1.3.3. Interference by MLL Fusion Proteins

The MLL fusion partners AF4, AF9, ENL and ELL have all been purified as part of the SEC (He et al., 2010; Lin et al., 2010). ELL has been shown to stimulate transcription elongation (Shilatifard, 1998; Shilatifard et al., 1996) and may provide direct activation of RNAPII at gene targets of the fusion protein. Furthermore, inclusion of p-TEFb in the SEC provides a direct link between the fusion proteins and this positive regulator of transcriptional elongation (Lin et al., 2010). The association between MLL-FPs and elongation proteins may result in circumvention of the rate-limiting initiation-to-elongation checkpoint at gene targets of the fusion protein, causing overexpression of these genes ultimately leading to leukaemia progression.

The interaction of BRD4 with the SEC and PAFc potentially facilitates recruitment of MLL-FPs to a variety of key oncogenic targets leading to increased expression of these genes and leukaemia development (Dawson et al., 2011). BRD4 has been shown to be important for acute myeloid leukaemia (AML) progression due to the fact that its gene targets are prominent oncogenes, for example *MYC* (Dawson et al., 2014; Dawson et al., 2011; Zuber et al., 2011). Inhibitors targeting BRD4 were shown to induce cell cycle arrest and apoptosis in MLL-r cell lines and increased survival in MLL-r leukaemia mouse models, possibly by disrupting BRD4-PAFc/SEC-mediated recruitment of the fusion protein to its gene targets (Dawson et al., 2011; Zuber et al., 2011). In this instance BRD4 does not have a direct oncogenic function but instead links the fusion protein to genes where it has detrimental effects.

Figure 1.2

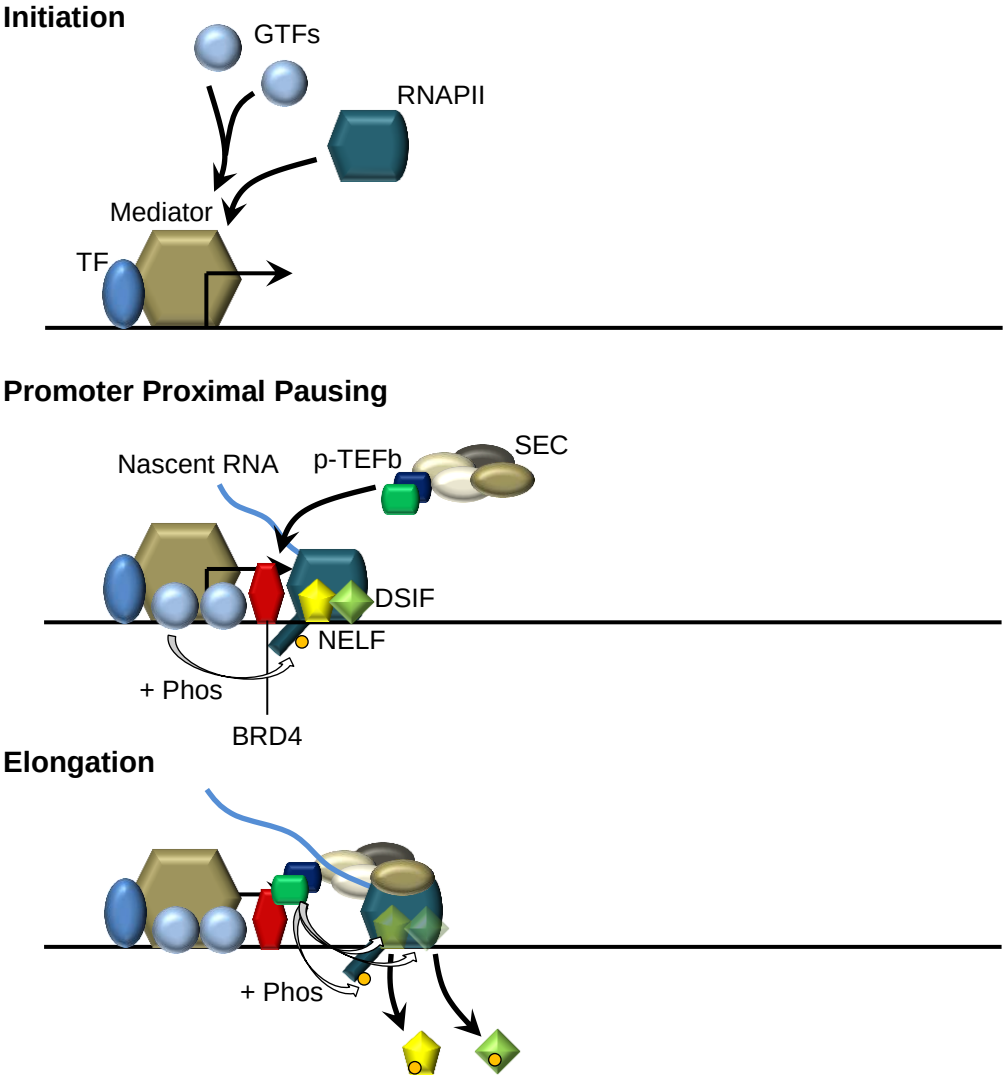


Figure 1.2 Simplified model of transcription. Initiation: Transcription factors (TF) bind to DNA and recruit the Mediator complex which in turn recruits RNA Pol II (RNAPII) and general transcription factors (GTFs). **Promoter Proximal Pausing:** RNAPII pauses after transcribing 30-60 nucleotides. Pausing is brought about by the negative regulators NELF and DSIF, and phosphorylation of RNAPII Ser-5 by the GTF TFIIH. BRD4 binds at promoters and recruits p-TEFb/SEC in the transition to elongation. **Elongation:** Phosphorylation of RNAPII Ser-2, NELF and DSIF by p-TEFb relieves pausing and allows RNAPII to transition to elongation, acting in combination with the SEC.

1.4. Chromatin Marks

Methylation and acetylation of specific histone residues are associated with different states of transcriptional activity. MLL-FP binding has been shown to correlate with certain histone marks, suggesting that these marks may play a role in recruitment of the fusion protein or may be a direct consequence of the fusion protein. Direct interactions between MLL-FPs and histone methyltransferases, enzymes that catalyse the formation of methyl marks, are responsible for some of these observations. The association between histone marks and transcription represents another example of how MLL-FPs can affect regulation of the transcription machinery at genes, leading to leukaemia development.

1.4.1. H3K79me2

Lysine 79 resides in the globular domain of histone H3 but is still subject to methylation. H3K79 di- and tri-methylation (H3K79me2/3) occur at the 5' end of gene bodies and have been linked to active transcription (Okada et al., 2005; Okada et al., 2006; Steger et al., 2008; Vakoc et al., 2006), specifically correlating with transcription elongation (Bitoun et al., 2007; Mueller et al., 2007).

Disruptor of Telomeric Silencing 1-Like protein (DOT1L) is the only known H3K79 methyltransferase and has been biochemically purified in a complex (DotCom) with the common MLL fusion partners ENL, AF9, AF10 and AF17 (Jones et al., 2008; Mohan et al., 2010). DOT1L interacts with actively-transcribing RNAPII and through this mechanism is associated with active genes (Kim et al., 2012).

Many studies have shown that MLL-r leukaemia cell lines or mouse BM cells transformed by an MLL-FP have increased levels of H3K79me2 at MLL-FP targets or genes that are upregulated in clinical MLL-r leukaemias (Bernt et al., 2011; Deshpande et al., 2013; Deshpande et al., 2014; Krivtsov et al., 2008; Milne et al., 2005b; Nguyen et

al., 2011; Okada et al., 2005; Wang et al., 2011). Direct interactions between DOT1L and AF9, AF10 or ENL have all been shown to be maintained in the corresponding MLL fusions (Kuntimaddi et al., 2015; Monroe et al., 2011; Mueller et al., 2007; Okada et al., 2005; Yokoyama et al., 2010). These interactions correlated with mislocalisation of DOT1L to MLL-FP gene targets important in haematopoietic transformation and a subsequent increase in H3K79me2 and transcription (Chang et al., 2010; Deshpande et al., 2014; Milne et al., 2005b; Monroe et al., 2011; Mueller et al., 2007; Nguyen et al., 2011). Importantly, knockout of DOT1L or disruption of its interaction with the fusion protein prevented MLL-mediated transformation of BM cells and inhibited colony-forming potential of MLL-r leukaemia cell lines (Chang et al., 2010; Deshpande et al., 2014; Kuntimaddi et al., 2015; Nguyen et al., 2011; Okada et al., 2005).

Despite no direct interaction between AF4 and DOT1L, an increase in associated H3K79me2 was observed at gene targets of the fusion protein in an Mll-AF4 mouse model (Krivtsov et al., 2008; Yokoyama et al., 2010). Furthermore, the importance of DOT1L was confirmed by shRNA-mediated knockdown of DOT1L in MLL-AF4 cell lines which led to loss of H3K79me2, gene expression and clonogenic capacity of the cells (Krivtsov et al., 2008). The mechanism of aberrant DOT1L recruitment to sites of MLL-AF4 is unknown but the inclusion of AF4 in the SEC might play an indirect role in this process.

More surprising is the reliance of MLL-AF6-mediated leukaemogenesis on DOT1L despite the cytoplasmic AF6 fusion partner showing no interaction with DOT1L or the SEC (Yokoyama et al., 2010). Nevertheless, an increase in H3K79me2 was seen at MLL-AF6 gene targets, and knockout or catalytic inhibition of DOT1L suppressed the leukaemogenic activity of MLL-AF6 cell lines (Deshpande et al., 2013; Wang et al., 2011). AF6 exhibits homodimerisation and this might be important for its leukaemogenic

activity as has previously been shown for other MLL-FPs (Dobson et al., 2000; Martin et al., 2003; So et al., 2004).

The association between MLL-FPs and DOT1L has been exploited to treat MLL-r leukaemias which exhibit reduced H3K79 methylation and MLL-FP target gene expression with DOT1L inhibitors, resulting in anti-proliferative effects on the leukaemia (Daigle et al., 2013; Daigle et al., 2011). These inhibitors are now in clinical trials (Stein and Tallman, 2015).

1.4.2. H3K4me3

Tri-methylation of lysine 4 on histone H3 (H3K4me3) is predominantly associated with the promoters of active genes, particularly those with a high CpG content, and has shown a strong correlation with MLL-FP binding (Bernstein et al., 2005; Guenther et al., 2008; Heintzman et al., 2007; Kim et al., 2005; Mikkelsen et al., 2007). So-called bivalent promoters are marked by H3K4me3 and tri-methylation of lysine 27 on histone H3 (H3K27me3), maintaining them in a poised state. Poised promoters function to silence developmental genes in embryonic stem cells (ESCs) before their necessary activation during differentiation (Bernstein et al., 2006; Jiang et al., 2011b; Ku et al., 2008). H3K4me3 has been observed across broad intergenic regions at the *Hox* clusters and is possibly indicative of transcription across the clusters (Bernstein et al., 2005; Guenther et al., 2005). Recently, broad domains of H3K4me3 were characterised at genes other than *Hox* clusters and were found to correlate with increased transcriptional consistency at key cell identity and function genes (Benayoun et al., 2014). It has been difficult to establish whether H3K4me3 is a cause or a consequence of transcription, however PHD finger-mediated binding of Transcription Associated Factor 3 (TAF3) to H3K4me3 has been shown to promote PIC formation and facilitate transcription (Lauberth et al., 2013). On the other hand, H3K4me3 is not required to maintain steady-state transcriptional activation (Clouaire et al., 2014).

The COMplex of Proteins ASsociated with Set1 (COMPASS) family consists of six proteins identified in mammals, including MLL1, which together are responsible for H3K4 methylation in cells (Shilatifard, 2012). These proteins are found in independent complexes with the common subunits WDR5, Ash2L, Rbbp5 and DPY30 (WRAD) as well as some unique subunits. The specific roles of the six SET domain-containing members (MLL1-4, Set1A/B) are described below.

The SET domain of MLL1 is capable of H3K4 mono- and di-methylation but relies on members of the WRAD complex for full efficiency and H3K4 trimethylation activity (Dou et al., 2006; Jiang et al., 2011a; Milne et al., 2002; Nakamura et al., 2002; Wysocka et al., 2005). Since deletion of the SET domain has negligible effect on the physiological function of MLL1, it is possible that the main role of MLL1 is mediated via the proteins it interacts with, such as the histone H4 lysine 16 (H4K16) acetyltransferase MOF (Mishra et al., 2014; Terranova et al., 2006). MLL1 was shown to interact with members of PAFc, providing a direct link between MLL1 and H3K4me3 with the transcription elongation machinery (Milne et al., 2010; Muntean et al., 2010).

MLL-FPs are unable to transform cells in the absence of wild-type MLL1 (Milne et al., 2010; Thiel et al., 2010) highlighting a role for MLL1 in MLL-r leukaemia development. Interaction of MLL1 with the COMPASS member WDR5 was shown to be essential for its methyltransferase activity (Dou et al., 2006; Dou et al., 2005). Targeting this interaction with a small molecule inhibitor blocked proliferation of MLL-r leukaemias and caused myeloid differentiation (Cao et al., 2014). The active chromatin conformation provided by MLL1-mediated H3K4 methylation might be necessary for MLL-FP recruitment, making MLL1 essential in MLL-r leukaemogenesis (Thiel et al., 2012).

MLL2, the other human homologue of *Drosophila* Trx, also causes H3K4 methylation at *Hox* genes (Hu et al., 2013b; Wang et al., 2009), with *Mll2* mutant mice showing slowed development and retarded growth similar to *Mll1* mutant mice (Glaser et al., 2006; Yu et al., 1995). However, MLL1 and MLL2 appear to have distinct gene targets, with MLL2-mediated H3K4 methylation predominantly at bivalent promoters (Bach et al., 2008; Denissov et al., 2014; Hu et al., 2013b). Unique amongst the COMPASS family, MLL1 and MLL2 interact with Menin and LEDGF at their N-termini which might be important for the recruitment of MLL1 and MLL2 (see sections 1.6 and 1.7) (Hughes et al., 2004; van Nuland et al., 2013; Yokoyama et al., 2005; Yokoyama et al., 2004).

Despite interacting with both wild-type MLL1 and MLL-FPs, Menin interaction with MLL-FPs seems to be more important, with Menin binding showing minimal overlap with MLL1 gene targets and being seemingly unimportant for normal B-cell development (Li et al., 2013; Yokoyama et al., 2005). Interaction between Menin and MLL-FPs has been shown to be necessary for MLL-FP-mediated leukaemogenesis (Yokoyama et al., 2005). Treatment of MLL-r leukaemias with inhibitors that disrupt this interaction has shown promising results, highlighting the MLL-FP-Menin interaction as a promising target for therapeutic intervention in MLL-r leukaemia patients (Borkin et al., 2015; Grembecka et al., 2010; Grembecka et al., 2012; He et al., 2015).

MLL3 and MLL4, the human homologues of *Drosophila* Trithorax-related (Trr) are predominantly found at enhancer regions where they catalyse H3K4 mono- and di-methylation in a redundant manner (Herz et al., 2012; Hu et al., 2013a; Lee et al., 2013). Despite a common H3K4 methyltransferase activity, the differing roles between MLL1, MLL2 and MLL3 help explain their non-redundant functions (Glaser et al., 2006; Lee et al., 2006; Yu et al., 1995).

Set1A and Set1B are responsible for the majority of H3K4me2 and H3K4me3 in mammals (Wu et al., 2008), similar to dSet1 in *Drosophila* (Herz et al., 2012; Mohan et al., 2011). Set1/COMPASS complexes uniquely contain Wdr82 and CXXC-Finger Protein 1 (CFP1), as well as HCF1 which is found in MLL1/2/COMPASS complexes (Lee et al., 2007; Lee and Skalnik, 2005; Wu et al., 2008; Wysocka et al., 2003). CFP1 is important for targeting Set1 complexes and H3K4 methylation to promoters throughout the genome (Clouaire et al., 2014; Clouaire et al., 2012; Thomson et al., 2010).

1.4.3. H3K27me3

As well as marking the promoters of poised genes, H3K27 tri-methylation (H3K27me3) is also found at the promoters of repressed genes (Mikkelsen et al., 2007). This histone mark can act as a binding site for polycomb complexes involved in the deposition of repressive monoubiquitination on lysine 119 of histone H2B (H2AK119ub1), thereby reinforcing the repressed state (Cao et al., 2005; Fischle et al., 2003; Gao et al., 2012; Tavares et al., 2012). H3K27me3 correlates with CpG-rich DNA (Ku et al., 2008) and may simply be a consequence of gene repression as opposed to a cause. The role of H3K27me3 in transcriptional repression has remained enigmatic but it might function to inhibit the deposition of the activation-associated acetylation on lysine 27 of histone H3 (H3K27ac) (Simon and Kingston, 2013). Furthermore, H3K27me3 might facilitate repression by recruiting chromatin remodellers or impeding the recruitment of transcriptional activators.

Polycomb Repressive Complex 2 (PRC2) consists of the subunits EED, SUZ12 and RbAp48, with an interchangeable catalytic subunit of EZH1 or EZH2 (Cao et al., 2002; Margueron et al., 2008). EZH2 catalyses global H3K27 methylation, whereas EZH1 is predominantly involved in catalysing H3K27 mono-methylation (H3K27me1). Both gain-of-function and loss-of-function mutations of EZH2 have been reported in patients with

haematopoietic malignancies (Lund et al., 2014). Small molecule inhibitors blocking the enzymatic activity of EZH2 have shown promising results in treating EZH2-mutated or overexpressed lymphomas (Bradley et al., 2014; Knutson et al., 2014; Knutson et al., 2012; McCabe et al., 2012; Zhang et al., 2012b). Furthermore, the apparent reliance of MLL-r leukaemias on EZH2 (Neff et al., 2012; Shi et al., 2013; Tanaka et al., 2012) have made inhibition of this enzyme an effective treatment for patients with these leukaemias (Xu et al., 2015).

1.4.4. Other Chromatin Marks

1.4.4.1. H3K36me2/3

H3K36 di- and tri-methylation (H3K36me2/3) are predominantly associated with the bodies of actively transcribed genes, increasing towards the 3' end (Bannister et al., 2005; Mikkelsen et al., 2007). An increase in H3K36me3 was shown at MLL-FP targets genes after MLL-FP-mediated transformation of mBM cells, possibly as a consequence of increased gene expression (Bernt et al., 2011). Deposition of H3K36 methylation throughout the gene body is mediated via interaction between the H3K36 methyltransferase SETD2 and elongating RNAPII (Li et al., 2003; Li et al., 2002; Yuan et al., 2009). The link between H3K36 methylation and transcription is possibly due to its antagonism with repressive complexes and mutual exclusivity with H3K27me3 (Yuan et al., 2011). H3K36 methylation can also mediate its effects by acting as a docking site for proteins containing a PWWP domain. The histone deacetylase Rpd3 recognises H3K36me3 and removes acetyl marks in co-ordination with transcribing RNAPII, preventing inappropriate transcription initiation within the coding region of gene bodies (Lee and Shilatifard, 2007). H3K36 methylation has also been implicated in various mechanisms of gene repression, one of which is mediated by DNA methylation caused by the PWWP-containing DNMT3A (Dhayalan et al., 2010).

1.4.4.2. H3K4me1 / H3K27ac

Mono-methylation of lysine 4 on histone H3 (H3K4me1) is associated with enhancers (Heintzman et al., 2009; Heintzman et al., 2007; Visel et al., 2009). Whilst the presence of H3K4me1 at enhancers generally remains constant, enhancers active early in development contain H3K27ac, which is absent from H3K27me3-marked poised enhancers proximal to developmental genes (Rada-Iglesias et al., 2011). An effector function of H3K4me1 might be mediated by binding of HIV-1 Tat1-Interactive Protein 60 (TIP60), which facilitates acetylation and incorporation of the histone variant H2A.Z (Altaf et al., 2010; Jeong et al., 2011). H3K4me1 may also enhance gene activity by preventing repressive DNA methyltransferases binding to their substrate of unmodified H3K4 (Ooi et al., 2007). As well as marking enhancers, H3K27ac is a marker of active promoters (Karlić et al., 2010; Liu et al., 2015) and may act as a docking site for transcriptional activators, similar to H3K4me3.

1.4.4.3. DNA Methylation

De novo methylation of DNA is carried out by the DNA methyltransferases DNMT3A and DNMT3B, whose activities are largely influenced by different histone marks (Ooi et al., 2007). DNMT1 works in conjunction with these enzymes to maintain methylation during processes such as DNA replication (Jones and Liang, 2009). Methylation is associated with the promoters of long-term repressed genes and can occur at CpG island promoters (see 1.4.5), which are predominantly unmethylated. In line with this, transcription start sites marked by the active H3K4me2, H3K4me3 and H2A.Z show a negative correlation with DNA methylation (Conerly et al., 2010; Ooi et al., 2007; Zilberman et al., 2008). Conversely, in gene bodies the presence of DNA methylation is a feature of active transcription (Hellman and Chess, 2007; Wolf et al., 1984). Specifically, methylation within the gene body might play a role in splicing since an increase in methylation was observed within exons at the intron/exon boundaries (Laurent et al., 2010).

1.4.5. CpG Islands

CpG Islands (CGIs) are sequences that are approximately 1kb in length with higher than expected GC and CpG content (Deaton and Bird, 2010). The mutagenic properties of methylated CpG (meCpG) mean that the majority of the mammalian genome is lacking in CpG dinucleotides. However, CGIs are refractory to methylation and have persisted throughout evolution, with the number and positions of CGIs being well-conserved between different species indicative of a functionally important feature (Long et al., 2013b). The unmethylated state is thought to be maintained by DNA-binding proteins that protect the CGIs from methylation (Erfurth et al., 2008), by Tet1-mediated demethylation via a hydroxymethyl-Cytosine intermediate (Guo et al., 2011; He et al., 2011; Ito et al., 2011), or by histone marks at CGIs blocking DNMT recruitment (Ooi et al., 2007). CGIs are associated with 70% of annotated gene promoters and correlate with active transcription (Carninci et al., 2006; Core et al., 2008b; Seila et al., 2008). In accordance with their association with transcription, most CGIs have a typically active chromatin signature, particularly H3K4me3 as well as H3 and H4 acetylation (Tazi and Bird, 1990; Thomson et al., 2010). Furthermore, there is a direct correlation between CpG density and amount of H3K4me3 (Illingworth et al., 2010). It has also been shown that recruitment of KDM2A to CGIs results in depletion of H3K36me2 (Blackledge et al., 2010). Importantly, these epigenetic modifications are mediated by proteins with CXXC domains that bind specifically to unmethylated CGIs (see 1.6) identifying the CGI as the platform between protein and histone mark. CGIs are also binding sites for PRC2 and are often seen underlying bivalent promoters (Lynch et al., 2012; Mendenhall et al., 2010). Although CGIs are generally unmethylated, it is possible for DNA methylation to occur at these sites, particularly during developmental processes such as gene imprinting and X-inactivation (Mohn et al., 2008; Payer and Lee, 2008; Stein et al., 1982). In these instances the CGIs are often already marked by H3K27me3, with the DNA methylation acting to “lock-in” the repressed state. It is also possible for CGI methylation

to occur at the promoters of genes in somatic cells, particularly during the early stages of differentiation (Schilling and Rehli, 2007; Shen et al., 2007; Weber et al., 2007). Approximately 50% of CGIs occur away from annotated promoters. These so-called “orphan” CGIs are located either inter- or intra-genically and are also capable of producing transcripts (Illingworth et al., 2008; Illingworth et al., 2010).

1.4.6. Super Enhancers

Super enhancers were recently identified as clusters of enhancers marked by different proteins and histone marks, including H3K4me1, H3K27ac, BRD4 and subunits 1 and 12 of the Mediator complex (MED1 and MED12, respectively) (Lovén et al., 2013; Sengupta et al., 2015; Whyte et al., 2013). Super enhancers were shown to associate with highly active genes that determine cell identity and are critical for the development of disease. This correlation with disease-associated genes explains recent findings that super enhancers are good therapeutic targets. In multiple myeloma the *MYC* oncogene can be inhibited by treatment with a BET-bromodomain inhibitor that causes loss of BRD4 from super enhancers and disrupts transcription elongation (Dawson et al., 2011; Lovén et al., 2013; Zuber et al., 2011).

1.5. Gene Targets of MLL Fusion Proteins

Understanding the gene targets that are misregulated as a result of MLL-FPs is important in determining disease pathology. Analysis of gene expression profiles of MLL-r leukaemia patients revealed that MLL1 rearrangements represent a unique class of leukaemias in terms of downstream abnormalities (Armstrong et al., 2002; Ferrando et al., 2003; Ross et al., 2004; Stam et al., 2010). In the majority of cases, MLL-r leukaemias show upregulation of *HoxA* cluster genes and *Meis1* compared to non-MLL-r leukaemias. MLL-r leukaemia patients can be segregated by gene expression profiles

when the leukaemias are analysed in a fusion protein- or lineage-specific manner (Stam et al., 2010).

Importantly, certain genes have been shown to act as prognostic markers when expressed in MLL-r leukaemias, and can predict response to treatment. Despite overexpression of *HoxA9* being associated with poor prognosis in leukaemia (Golub et al., 1999), a subset of MLL-r ALLs displaying a lack of *HoxA* gene expression have a higher risk of disease relapse (Stam et al., 2010). 35 MLL-r AMLs separated based on expression of *EVI1*, a recruiter of various epigenetic and transcriptional regulators, showed that *EVI1* expression was more common in MLL-AF6 and MLL-AF9 leukaemias and that it was essential for the pathogenesis of the AML, correlating with a poor prognosis (Bindels et al., 2012; Matsuo et al., 2014). MLL-r leukaemias displaying *KRAS* or *NRAS* mutations showed increased resistance to chemotherapy and a significant decrease in patient EFS and OS (Driessen et al., 2013). Interestingly, all RAS-mutant samples in this study lacked *HoxA* gene expression and it was found that the RAS mutation was the more likely cause of the increased risk of relapse in *HOXA*-negative patients (Driessen et al., 2013; Stam et al., 2010). MLL-r leukaemia patients could also be distinguished based on CpG methylation status, having increased methylation compared to non-MLL-r leukaemia patients (Stumpel et al., 2009; Stumpel et al., 2011). This increased methylation correlated with decreased gene expression, an increased incidence of relapse and a lower OS. However, *in vitro* work suggested that these leukaemias were more responsive to treatment with the de-methylating agent zebularine (Stumpel et al., 2009).

Gene expression profiles of MLL-r leukaemia patient cells lines and MLL-FP-transformed primary cells have been analysed to determine how closely they resemble gene expression profiles of MLL-r leukaemia patients. In many instances, the expression of the posterior *HoxA* genes and *Meis1* was observed in cell lines, similar to as in

patients (Ayton and Cleary, 2003; Chen et al., 2008b; Krivtsov et al., 2006; Metzler et al., 2006; Zeisig et al., 2004). Furthermore, gene set enrichment analyses showed that the gene expression profiles of MLL-r cell lines mimic that of clinical MLL-r leukaemias (Krivtsov et al., 2013). These models have allowed testing of the importance of certain genes in disease initiation and maintenance. Unlike wild-type mBM cells, mBM cells from *HoxA9*$^{-/-}$ mice could not be transformed by MLL-ENL (Ayton and Cleary, 2003). Furthermore, knockdown of *HoxA9* reduced the cell viability of mouse leukaemic-Granulocyte Macrophage Progenitors (GMPs) and their colony-forming ability (Faber et al., 2009). However, consistent with lack of *HoxA* expression in a subset of MLL-r leukaemia patients, *HoxA7* and *HoxA9*-deficient mice were susceptible to MLL-FP-mediated transformation, albeit at a lower efficiency (So et al., 2004).

The recruitment of MLL-FPs to targets causes aberrant gene regulation by consistently triggering DOT1L-mediated H3K79me2 upregulation, and affecting SEC recruitment to the gene. This mechanism underlies the importance of recruitment of the fusion protein to these genes in driving leukaemogenesis.

1.6. Recruitment of Chromatin Proteins

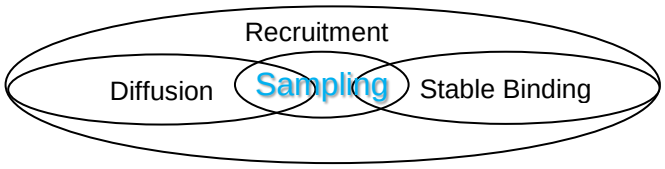
The concept of recruitment is central to a protein's ability to regulate gene activity. The most thorough studies on protein recruitment have been done on transcription factors (TF) in bacterial cells (Berg et al., 1981; Elf et al., 2007). In these systems the proteins undergo 3-dimensional (3D) diffusion within the cell but due to the scarcity of the sequence-specific TF-binding sites, this process alone might be too inefficient for TF recruitment and function. Instead, diffusion to the DNA template at non-specific sites is followed by 1-dimensional (1D) diffusion to the binding site. 1D diffusion greatly increases the efficiency of the search for the specific TF-binding sequence, where the protein can stably bind, preventing further diffusion. By this means the protein is

recruited to its target site. Whilst 1D diffusion of proteins has not been observed in mammals, these concepts can still be used in terms of helping us to understand recruitment. Mechanisms building on these ideas are discussed throughout the results of this chapter.

An idea linked to recruitment that is gaining momentum is the concept of sampling. In the bacterial system above, the TF samples binding sites as it diffuses along the DNA until it finds its specific binding sequence. In mammals, sampling has been used to describe a responsive model for the antagonistic recruitment of repressive polycomb group (PcG) and activating trithorax group (TrxG) proteins to CGIs, where stable binding is only permitted in response to an absence of active or repressive histone marks, respectively (Klose et al., 2013). Building on ideas discussed above, sampling describes a scenario where the protein diffuses towards its binding site (a CGI) where stable binding is prevented or facilitated depending on interactions with histone marks, DNA methylation or the presence or absence of other occupying proteins (Figure 1.3A).

Figure 1.3

A)



B)

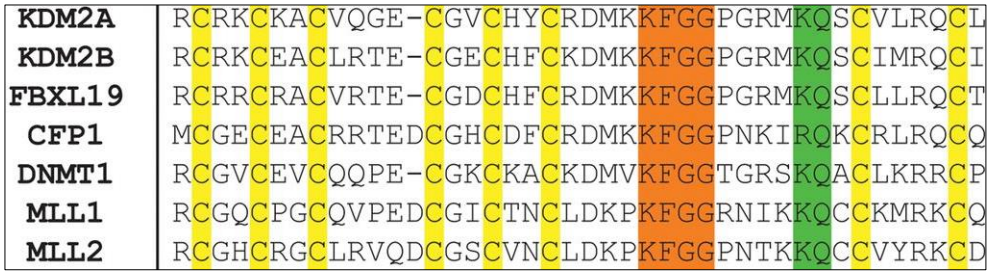


Figure 1.3 Recruitment of chromatin proteins. A) Schematic of how protein diffusion, sampling and binding link in recruitment. **B)** Primary sequences of CXXC domains found in different mammalian chomatin proteins, with conserved residues highlighted (adapted from (Long et al., 2013a)).

As highlighted by the success of treatment with some of the small molecule inhibitors, targeting the recruitment of chromatin proteins is a potential effective treatment strategy and important mechanism to understand. Various domains and interactions have been implicated in the recruitment of chromatin proteins to gene targets. Since recruitment is an essential first step preceding functional output, an appreciation of the different recruitment mechanisms is important. Before a detailed discussion about MLL-FP recruitment is given, mechanisms common to recruitment of chromatin proteins are described.

1.6.1. COMPASS Proteins

MLL1 recruitment has been shown to involve its zinc-finger CXXC domain and third Plant Homeodomain finger (PHD3) (Milne et al., 2010). CXXC domains bind specifically and with high-affinity to uCGIs (Allen et al., 2006; Birke et al., 2002; Erfurth et al., 2008; Shin Voo et al., 2000). PHD fingers in different proteins have been shown to bind methylated and unmethylated forms of H3K4, K9 and K36, as well as acetylated H3K9 and K14 (Musselman and Kutateladze, 2011). Specifically PHD3 of MLL1 binds to H3K4me3 (Milne et al., 2010; Wang et al., 2010). A point mutation in MLL1 PHD3 that abrogated H3K4me3 binding, in conjunction with two point mutations in the CXXC domain, abolished binding of MLL1 at *HoxA9* (Milne et al., 2010). One of the point mutations in the CXXC domain disrupted its DNA-binding capability and the other reduced its interaction with PAFc (Allen et al., 2006; Milne et al., 2010). This implicates a role for PAFc in MLL1 recruitment to gene targets possibly via interaction with RNAPII or BRD4 (Dawson et al., 2011; Kim et al., 2010). These data show that the CXXC domain and PHD3 are necessary for MLL1 recruitment to *HoxA9*. Both domains were also found

to be sufficient for MLL1 recruitment to *HoxA9* since a construct containing just these domains recapitulated binding of full-length MLL1 (Milne et al., 2010). MLL2 has a CXXC domain alongside PHD fingers but lacks the equivalent PHD3 of MLL1. MLL3 and MLL4 do not contain CXXC domains but do contain PHD fingers. Whilst no work has been done to assess the contribution of these domains to the recruitment of MLL2-4 it is possible that they allow for binding at gene targets in a similar manner as in MLL1.

The CFP1 subunit of Set1/COMPASS complexes also contains a CXXC domain and a PHD finger. However, whilst CFP1 is crucial for Set1-mediated H3K4me3, this does not appear to be dependent on the DNA-binding ability of its CXXC domain (Clouaire et al., 2012). CFP1 with a DNA-binding deficient CXXC domain was able to rescue H3K4me3 levels in CFP1-knockout ES cells. Recruitment of CFP1 to active promoters may instead be directed by its PHD finger, since the yeast orthologue of this PHD finger binds H3K4me3 (Shi et al., 2007). Instead it appears that the CXXC domain of CFP1 restricts recruitment of Set1 to CGIs since mutation of the CXXC domain led to aberrant H3K4 methylation at non-CGI sites (Clouaire et al., 2012). These data show that whilst the CXXC domain is not necessary for CFP1 recruitment, it does play a role in limiting its target sites.

1.6.2. PcG Proteins

Polycomb Repressive Complex 1 (PRC1) is recruited to a subset of uCGIs and this was shown to be mediated by KDM2B (Farcas et al., 2012; He et al., 2013; Wu et al., 2013). KDM2B, and its closely related KDM2A, contain CXXC domains that are very similar in structure to the CXXC domains of both MLL1 and CFP1 (Figure 1.3B). KDM2B binds to CGIs genome-wide but only recruits PRC1 to a subset of CGI promoters (Farcas et al., 2012; He et al., 2013; Wu et al., 2013). This mechanism is dependent on the CXXC domain of KDM2B but not its PHD finger (Blackledge et al., 2014; Farcas et al., 2012;

He et al., 2013). PRC1 places H2AK119ub1 at these sites leading to repression of the associated genes (de Napoles et al., 2004; Tavares et al., 2012; Wang et al., 2004). H2AK119ub1 has recently been shown to recruit PRC2, resulting in H3K27 methylation at sites of H2AK119ub1, establishing a polycomb domain (Blackledge et al., 2014).

Interaction with non-coding RNAs (ncRNA) has been shown to recruit PcG proteins to sites that are subsequently repressed. This mechanism has been observed during X inactivation (Mak et al., 2002; Plath et al., 2003; Silva et al., 2003; Zhao et al., 2008), gene imprinting (Fitzpatrick et al., 2002; Pandey et al., 2008) and at *Hox* clusters (Rinn et al., 2007).

1.6.3. Competition Between COMPASS and PcG Proteins

CGIs are binding sites for both COMPASS and PcG proteins (Farcas et al., 2012; Ku et al., 2008; Lynch et al., 2012; Mendenhall et al., 2010; Thomson et al., 2010), however the antagonism between these groups of proteins suggests a degree of competition for binding. CFP1 and KDM2A have been shown to bind the majority of CGIs genome-wide but, unlike KDM2B, are not found at polycomb-repressed CGIs (Blackledge et al., 2014; Farcas et al., 2012; Thomson et al., 2010). The mechanism behind whether a CGI establishes an active or a repressed state is still unclear but histone marks appear to play a key role in this process, with regions marked by active histones being refractory to polycomb-mediated silencing. Specifically, TrxG-mediated H3K4me3 inhibits binding of the polycomb subunit Suz12, and acetylation of H3K27 prevents H3K27 methylation (Schmitges et al., 2011; Tie et al., 2014). H3K36me2/3 also serves a role in inhibiting PRC2 HMT activity (Yuan et al., 2011). It was shown that a repressed state arises by default in the absence of activating signals (Lynch et al., 2012; Mendenhall et al., 2010). For the proposed sampling recruitment mechanism, in the absence of active transcription at a locus, PcG proteins can bind and set up a long-term repressed state (Klose et al., 2013). Conversely, at an actively transcribed region, RNAPII-interacting

TrxG proteins initiate a positive-feedback mechanism that reinforces the active chromatin signature and inhibits polycomb activity. The existence of bivalent promoters suggests that co-occupancy of loci by these proteins can also occur (Voigt et al., 2013). Indeed both TrxG and PcG proteins have been identified at the same loci (Beisel et al., 2007; Enderle et al., 2011).

1.6.4. Interacting Proteins

Specific subset of CGIs that are targeted by chromatin proteins could be determined by interacting partners of the chromatin proteins. LEDGF interacts with MLL1 and MLL-FPs and contains a PWWP domain that binds to chromatin, specifically thought to be H3K36me2/3 and directly to DNA (Daugaard et al., 2012; Eidahl et al., 2013; Okuda et al., 2014; Pradeepa et al., 2012). The chromatin-binding ability of the LEDGF PWWP domain is thought to be involved in recruitment of MLL-FPs (Okuda et al., 2014; Yokoyama and Cleary, 2008) and might define target sites of the wild-type or fusion protein as those marked by H3K36me2/3. BRD4 contains bromodomains that bind to acetylated lysines (Zhang et al., 2012a; Zippo et al., 2009). BRD4 interacts with members of PAFc and the SEC and through this mechanism might interact with MLL1 and MLL-FPs (Dawson et al., 2011). It is possible that this interaction restricts MLL1 and MLL-FP binding sites to only those CGIs marked by acetylated lysines.

1.7. Recruitment of MLL Fusion Proteins

The study of the protein domains and interactions involved in MLL-FP recruitment is essential to understand MLL-r leukaemia pathogenesis since the fusion protein must be recruited to gene targets before it can exert it aberrant function. Therefore, an understanding of MLL-FP recruitment provides a potential therapeutic avenue. Since the focus of this thesis is a genome-wide assessment of MLL-FP recruitment, a detailed description of the current model is given below.

1.7.1. Menin/LEDGF

Truncation constructs of MLL1 first identified the interaction with Menin as occurring in the first 1400 amino acids of MLL1 (Yokoyama et al., 2004). Since this region is maintained in MLL-FPs this allowed immunoprecipitation experiments involving Menin and either MLL-P300 or MLL-ENL (ME), confirming an interaction between Menin and MLL-FPs for the first time (Yokoyama et al., 2004). The domain comprising amino acids (AAs) 6 to 10 in the N-terminus of MLL, termed the high-affinity Menin-Binding Motif (hMBM), was found to necessary for the interaction with Menin (Yokoyama 2005). This 5 AA motif was shown to be important for interaction with Menin regardless of the fusion partner. Whilst necessary, the hMBM did not seem to be sufficient for interaction with Menin, since a reverse approach showed that a minimal domain of AAs 5 to 44 was required to pull-down Menin (Caslini et al., 2007). These studies suggest that there is more than one Menin-interaction motif, with a second one occurring downstream of the hMBM.

The first evidence that the Menin interaction plays an important role in MLL-FP leukaemia was found in a study showing that, in contrast with normal MLL-ENL, MLL-ENL with an hMBM deletion (MEd) did not cause immortalisation of c-kit^{+} mouse bone marrow (mBM) cells (Yokoyama et al., 2005). When transplanted into mice, MEd cells did not result in lethality by day 200, whereas 60% of mice transplanted with ME-transduced cells died by the same time point. Induced knockout of Menin also prevented MLL-ENL from maintaining colony-forming potential of ME-transduced cells. In a similar study, removal of N-terminal motifs from MLL-AF9 impeded maintenance of serial replating ability and led to differentiation of MLL-AF9-transduced BM cells (Caslini et al., 2007). Menin can be sequestered away from MLL-FP by using dominant negative constructs such as the AA 5 to 44 fragment. This causes a reduction in cell growth and colony-forming potential of MLL-FP-transduced mBM cells (Caslini et al., 2007).

Menin interaction with MLL-FPs was shown to facilitate activation of fusion protein target genes. In the presence of MEd or *MEN1* knockout, decrease in *HoxA7* and *HoxA9* expression, but not of *p27* expression (possibly cell-type specific or just a target of wild-type MLL1), was found compared to the control counterparts (Yokoyama et al., 2005). In a similar study, *MEN1* knockout in MLL-AF9-transformed mBM cells caused a significant reduction in growth and expression of *HoxA9* (although activation by wild-type MLL1 cannot be ruled out for the latter) (Chen et al., 2006). Furthermore, transfection of dominant-negative Menin-binding fragments into MLL-AF9 immortalised mBM cells resulted in a decrease in *HoxA9* and *Meis1* expression three days post-transfection (Caslini et al., 2007).

The hypothesis that Menin could play a role in recruiting MLL-FPs to gene targets came from chromatin immunoprecipitation (ChIP) experiments that showed co-localisation of MLL-AF6 and Menin binding at *HoxA7*, *HoxA9* and *HoxA10* in ML-2 cells (Yokoyama et al., 2005). In MLL-AF9-transformed mBM cells with an induced *MEN1* knockout MLL-AF9 binding reportedly decreased at *HoxA9* (Thiel et al., 2010). However, the AF9(C) antibody used in ChIP experiments in this study would detect wild-type AF9 as well as MLL-AF9 making it impossible to distinguish between the two. Importantly, the MLL-FP associated histone mark H3K79me2 showed a 2-fold reduction in signal at the *HoxA9* locus, suggesting loss of the fusion protein. Interestingly, in a myeloblastic cell line transformed with inducible MLL-ENL-ER, induction of MLL-ENL led to increased binding of Menin as well as MLL-ENL binding across *HoxA9*, suggesting that the fusion protein could be recruiting Menin to genes (Caslini et al., 2007). This hypothesis is further supported by ChIP experiments done in the presence of a Menin binding-deficient fusion protein that showed a reduction in Menin binding at *HoxA9* compared to when ChIP of Menin was done in the presence of the normal fusion protein (Yokoyama and Cleary, 2008).

The interaction with Menin was shown to be part of a larger, trimeric interaction with LEDGF (Yokoyama and Cleary, 2008). Whilst Menin is able to interact with the fusion protein in the absence of LEDGF, the interaction between the fusion protein and LEDGF is dependent on the presence of Menin. Menin is also unable to interact with LEDGF in the absence of an MLL intermediate (van Nuland et al., 2013). A point mutation in the MLL N-terminus (F129A) that disrupted interaction between MLL-ENL and LEDGF prevented serial replating of MLL-ENL-transformed mBM cells and caused a decrease in *HoxA9* expression (Yokoyama and Cleary, 2008). It was identified that the disruption to the MLL-ENL transformation potential in the absence of Menin was likely an indirect effect due to the role of Menin in linking LEDGF to the fusion protein, with LEDGF being the key contributor to the disease. In MLL-ENL-transformed BM cells, mutant Menin proteins (P12L or L22R) that no longer interact with LEDGF could not rescue the loss of colony forming potential or *HoxA9* expression resulting from Menin knockout. Furthermore, when the chromatin-binding PWWP domain of LEDGF was used to replace the Menin- and LEDGF-interacting motifs in the N-terminus of MLL-ENL (pME), the fusion protein could still transform cells and maintain the expression of *HoxA9*, even in the absence of Menin or LEDGF.

The importance of the LEDGF chromatin-binding PWWP domain as well as co-localisation of Menin, LEDGF and MLL-AF6 (ML-2 cells) or MLL1 (U937 cells) across the *HoxA7* and *HoxA9* loci gave rise to the hypothesis that LEDGF might target fusion proteins to genes (Yokoyama and Cleary, 2008). ChIP experiments showed that pME still bound to *HoxA9,* but other domains of the fusion protein could also have been responsible for this recruitment. Since binding of pME with a chromatin binding-deficient PWWP domain was not tested it is not possible to determine a role for the PWWP domain in targeting the fusion protein from these ChIP experiments.

The LEDGF PWWP domain was found to be one of three components necessary for MLL-FP driven leukaemogenesis (Okuda et al., 2014). A minimal construct containing just the CXXC domain of MLL1 and the transactivation domain of MLL-ENL fused to the PWWP domain of LEDGF (PCE), immortalised c-kit^{+} BM cells, maintaining their colony-forming potential and increasing expression of the fusion protein targets *HoxA7*, *HoxA9*, *Myb*, *Hmgb3* and *Cbx5*. Removal of either the LEDGF PWWP domain or the MLL1 CXXC domain from PCE inhibited its transforming ability. It has previously been reported that the PWWP domain localises to chromatin by binding to both DNA and H3K36me3 (Daugaard et al., 2012; Eidahl et al., 2013; Pradeepa et al., 2012). In a pull-down experiment histones with H3K36me2 and me3 co-precipitated with the LEDGF PWWP domain but the specific interaction was not tested (Okuda et al., 2014). The co-precipitated histones also contained K27me3, K27ac and K4me2. Importantly, a W21A point mutation in the PWWP domain that supposedly disrupts histone binding was inserted to investigate the LEDGF targeting mechanism. Whilst mutation of the PWWP domain prevented nucleosome localisation and abolished binding of the construct at *HoxA7*, *HoxA9* and *ARP1* as detected by ChIP, this result might be an artefact of the native ChIP technique. Removal of the Menin/LEDGF interaction domain of MLL-FPs from the CXXC domain did not alter binding at *HoxA9* if the material is double-fixed (Milne et al., 2010). These data together suggest that the CXXC domain alone can bind to DNA, but it binds more strongly in the presence of the LEDGF PWWP domain. Therefore, in terms of sampling, LEDGF could represent a protein that pre-occupies the binding site and facilitates stable binding of the CXXC domain. ChIP across *HoxA7* and *HoxA9* in ML-2 cells reveals an overlap between MLL-AF6, Menin, LEDGF binding and H3K36me3 at the promoter of *HoxA9* (Okuda et al., 2014). Although H3K36me2 shows a broader pattern across this range, MLL-AF6 and LEDGF binding does not seem to correlate. However, binding of these two proteins correlated more with H3K36me2 than H3K36me3 at the TSS of *CDKN1B* and *CDKN2C,* suggesting that LEDGF preference for either histone mark is locus-specific. Interestingly ChIP-seq of H3K36me3 shows

higher enrichment in the body of MLL-AF6 target genes compared to the genomic average.

1.7.2. CXXC Domain

The importance of the CXXC domain in MLL-FP-mediated transformation was first determined in studies where MLL-ENL harbouring a deletion of this domain (termed BCB at this point, for “basic-cysteine-basic”) was transfected into mBM cells and CXXC deletions were found to inhibit the colony-forming potential of the fusion protein (Slany et al., 1998). A similar deletion in MLL-ENL also prevented colony formation when transfected into Cos-7 cells (Ayton et al., 2004). In this study, the decrease in transformation ability correlated with a decrease in gene activation as measured by a luciferase assay.

An ability of the MLL-FP CXXC domain to bind DNA was first shown by probing GST-tagged BCB with different types of DNA (Slany et al., 1998). CXXC binding to DNA was later discovered to be dependent on unmethylated CpG dinucleotides (Birke et al., 2002) (Birke 2002). 52bp oligonucleotides with six regularly-spaced CpGs were used in the assay and binding was assessed by gel-shift. Importantly, when the CpGs were replaced with either GpC or methylated CpG (meCpG), no binding was observed. Binding was observed in the presence of hemi-methylated CpG although noticeably less than unmethylated. A dissociation constant (K_D) of 32.8nM was measured for the MLL-FP CXXC in the presence of the 52bp CpG oligo, compared to >1mM (i.e. no binding) in the presence of meCpG. This specificity for unmethylated CpG was confirmed in other studies using a similar approach (Allen et al., 2006; Bach et al., 2008).

The binding affinity of the MLL-FP CXXC domain has been tested in a variety of contexts, using slightly different sizes of the domain depending on the flanking residues included, and with different CpG-based oligos. A 12-mer with a central CpG recorded a

K_D of 4.3µM (Allen et al., 2006). Similarly another study found that a 12-mer with a central CGCG resulted in a K_D of 3.4µM whereas CGCGCG resulted in a K_D of 7.4µM (Erfurth et al., 2008). This study suggested that the DNA-contact region of the CXXC domain extends beyond the CpG dinucleotide and hence the increase in CpG content does not affect the K_D much due to steric hindrance inhibiting extra CXXC domains from binding. A further study measured K_D values between 16µM and 30µM for a single CpG (approximately eight times higher than the study by Allen et al.), possibly influenced by neighbouring DNA residues reducing binding efficiency (Cierpicki et al., 2010). The authors suggest that higher binding at CGCG and CGCGCG over CG may be indicative of 1D diffusion. An oligo with three CpGs equally spaced apart by two nucleotides resulted in a K_D of 24uM with the MLL-FP CXXC domain, closer to that of a single CpG. Based on their suggestions, the spacing prevents diffusion but is still sufficiently short that steric hindrance would prevent multiple CXXC binding events (Erfurth et al., 2008). An analysis of different CXXC domains found that the MLL-FP CXXC domain had the lowest K_D, at 0.64µM, and thus the strongest binding, with the CXXC domains of CFP1 and DNMT1 at 4.44µM and 21.9µM, respectively (Risner et al., 2013).

Chemical shift perturbation analysis upon DNA binding, revealed that the majority of the DNA-contacting residues present on one face of the MLL-FP CXXC domain, from AAs 1183 to 1188 (Allen et al., 2006). Point mutations in several of these residues inhibited binding of the CXXC domain to the DNA template in gel shift assays, particularly the K1186A mutation. A CXXC mutation not located on the main contact surface of the domain, K1176A, was also shown to inhibit CXXC DNA binding activity. Importantly, this study identified which mutations of the CXXC domain could be made without altering its structure. Only mutations that maintained structural integrity are mentioned here.

The importance of the CXXC domain of MLL-FPs in leukaemic transformation made it a good candidate in the mechanism of targeting MLL-FPs to gene targets. Interestingly,

the isolated CXXC domain and its flanking residues can recapitulate MLL-AF9 binding at *HoxA9*, suggesting that the CXXC domain is sufficient for targeting the fusion protein to this locus (Milne et al., 2010). This binding at *HoxA9* does however require the presence of wildtype Mll1. Conversely, an R1153A, K1176A double point mutation in the CXXC domain of full-length MLL-AF9 abolishes binding of the fusion protein at *HoxA9*. This evidence shows that the CXXC domain is necessary for MLL-AF9 recruitment to *HoxA9* and suggests that the Menin/LEDGF complex cannot target the fusion protein to *HoxA9* in the absence of a functional CXXC domain. The double point mutant CXXC domain in MLL-AF9 resulted in loss of *HoxA9* expression and transformation potential of the fusion protein, showing a direct link between gene targeting and disease. Similarly a study using an MLL-FP harbouring a K1186A point mutation in the CXXC domain, showed loss of MLL-FP binding at the transcription start sites of *EVI1*, *PITX2*, *HoxA7* and *HoxA9*, all of which were marked by unmethylated CpGs (Okuda et al., 2014). This correlated with a loss of gene expression and colony-forming ability.

A role beyond recruitment was found for the CXXC domain of MLL1 and MLL-FPs which implicated this domain as important for protecting CGIs from methylation. A CGI upstream of the foetal exon of *HoxA9* was analysed for methylation status in wild-type versus *Mll1*$^{-/-}$ Mouse Embryonic Fibroblasts (MEFs) (Erfurth et al., 2008). Methylation levels were higher in *Mll1*$^{-/-}$ MEFs compared to wild-type fibroblasts. Increased methylation of CpGs at other *Hox* loci has also been noticed in *Mll1*$^{-/-}$ MEFs (Milne et al., 2002). Restoration of Mll1 expression in *Mll1*$^{-/-}$ MEFs reduced methylation, whereas shRNA-mediated knockdown of *Mll1* in wild-type MEFs increased methylation (Erfurth et al., 2008). Interestingly, expression of MLL-AF4 or MLL-AF9 into *Mll*$^{-/-}$ MEFs did not protect as many CpGs in the CGI from methylation as wild-type Mll1. Moreover, the fusion proteins took longer (>1 week) to reverse the already present methylation compared to the wild-type Mll1. This may be correlated with a decrease in MLL-FP binding efficiency in the absence of wild-type MLL1 (Milne et al., 2010). A more recent

study (Cierpicki et al., 2010) showed that *Mll1*$^{-/-}$ MEFs transfected with MLL-AF9 harbouring a DNA-binding deficient CXXC domain had increased methylation at CGIs compared to *Mll1*$^{-/-}$ MEFs transfected with MLL-AF9 without the mutation. This study specifically implicated the CXXC domain in protecting uCGIs from methylation. Despite containing a DNA binding-deficient CXXC domain, MLL-AF9 still bound at CGIs in this study. The authors suggested that the interaction of MLL-AF9 with the Menin/LEDGF complex can maintain binding of the fusion protein despite contact between the CXXC domain and CpG dinucleotides being lost, however they did not test this. In accordance with evidence for MLL-FP-mediated protection from methylation, CGIs in MLL-r AML patient samples are predominantly hypomethylated, whereas in IDH-mutant patient samples CGIs are mainly hypermethylated (Akalin et al., 2012).

A study showed that when MLL-AF9 containing the CXXC domain of either CFP1 or MBD1 in place of its own CXXC domain was used to transform mBM cells *HoxA9* expression was abolished and the fusion protein lost its ability to transform cells, despite still binding at *HoxA9* (Risner et al., 2013). It should be noted, however, that a very low signal was detected in some of these ChIP experiments and no negative control binding region was tested. Similar to previous studies (Cierpicki et al., 2010), the authors claimed that the MLL-FP CXXC domain was not necessary for targeting the fusion protein to genes. This finding is in direct contrast to studies that showed the necessity of the CXXC domain in MLL-FP recruitment (Milne et al., 2010; Okuda et al., 2014). The common finding of these studies is the importance of the CXXC domain in MLL-FP leukaemic transformation. A separate study showed that Mll2 fused to ENL was unable to transform primary haematopoietic cells, unlike MLL-ENL which was able to transform these cells (Bach et al., 2008). Importantly, swapping the MLL-ENL CXXC domain and its surrounding basic residues into Mll2-ENL conferred colony-forming potential to Mll2-ENL. This was despite both MLL-ENL and Mll2-ENL CXXC domains showing the same binding specificity for unmethylated CpGs. Furthermore, the CXXC-swapped Mll2-ENL

showed increased activation of *HoxA7*, *HoxA9* and *Meis1*. A difference in nuclear loci occupancy between MLL-ENL and Mll2-ENL suggests that the specific CXXC domains may have different targeting abilities, possibly due to differences in their ability to interact with PAFc (Milne et al., 2010). However when the CXXC domain of MLL2 was swapped in place of the CXXC domain of MLL-ENL, it did not inhibit the transformation potential of the fusion protein (Okuda et al., 2014). The CXXC domains of MLL2, DNMT1 and KDM2A could all maintain *HoxA9* expression and colony-forming potential when swapped in place of the MLL CXXC domain, with the KDM2A swap shown to maintain binding at *PITX2*, *HoxA7* and *HoxA9*.

Interestingly, inserting of the third PHD finger (PHD3), which binds H3K4me3, downstream of the MLL-FP CXXC domains was shown to inhibit the transforming ability of the fusion protein (Chen et al., 2008a; Muntean et al., 2008). This mechanism might play a role in retargeting the fusion protein, perhaps limiting or expanding the number of uCGIs that the CXXC domain is targeted to. Whilst MLL-PHD3-FP binding was not assessed, inclusion of PHD2 triggered a reduction in binding at the *HoxA9* promoter. However, this reduction is likely caused by structural obstructions, since PHD2 does not have histone binding ability (Muntean et al., 2008).

1.7.3. PAFc

An interaction between residues surrounding the MLL1 CXXC domain and PAFc is retained in the MLL-FP (Milne et al., 2010; Muntean et al., 2010). This interaction does not occur with MLL2 or the Set1 complexes (Milne et al., 2010). Importantly, the R1153A point mutation in the MLL-FP CXXC domain severely disrupts interaction with PAFc and reduces binding of MLL-AF9 at *HoxA9*. Knockdowns of PAFc components caused a similar reduction in binding of MLL-ENL at *HoxA9* (Muntean et al., 2010). Both mutation of PAFc interaction and knockdown of PAFc members reduced the leukaemogenic activity of the fusion protein, reducing the colony-forming ability of MLL-FP transformed

cells. Despite a reduction in binding, recruitment of MLL-AF9 was only abolished in the presence of both R1153A and the DNA-binding K1176A mutation suggesting that both of these properties of the CXXC domain are important (Milne et al., 2010).

1.8. Genome-wide approaches

This D.Phil. aims to analyse the validity of current MLL-FP recruitment models on a genome-wide level. This is important to address since many of the studies listed above are limited to an analysis of a few select genes. It is important to first summarise other genome-wide approaches that have looked at MLL-FP binding and its associated activity.

The first genome-wide assessment of MLL fusion protein binding utilised ChIP-seq of MLL(N) and AF4(C) to assess gene targets of MLL-AF4 in the pre-B ALL MLL-AF4 cell line SEM, identifying 2490 gene targets (Guenther et al., 2008). Since no antibody exists that can uniquely detect the MLL-AF4 fusion protein, this is the most efficient way to assess MLL-AF4 gene targets. However, binding of wild-type MLL1 and AF4 will also be detected by this ChIP-seq technique. The list of potential MLL-AF4 gene targets was refined by removing gene targets that were also co-occupied by MLL(N) and AF4(C) in the non MLL-r pre-B ALL cell line REH. Some of the resulting 226 genes were verified as MLL-AF4 gene targets by transfection of HA-tagged MLL-AF4 followed by ChIP-qPCR. However, a separate approach (Wilkinson et al., 2013) used MLL(N), AF4(C) and H3K79me2 overlap in RS4;11 cells and compared the gene target list to the 2490 targets from the study by Guenther et al. A set of 491 MLL-AF4 gene targets was identified as common between the cell lines, with verification of the targets that were not identified by Guenther et al. by knockdown of *MLL-AF4*. These ChIP-seq datasets have been used in other studies to identify specific target genes of MLL-AF4 whose role in the leukaemia has then been analysed in detail (Geng et al., 2012). The MLL-AF6 leukaemia cell line ML-2 lacks the wild-type *MLL1* allele and hence provides an ideal

model for ChIP-seq experiments with an MLL(N) antibody. ChIP-seq in this cell line identified 154 genes occupied by MLL-AF6 (Okuda et al., 2014). A specific approach in primary cells was achieved by transducing c-kit^{+} mBM cells with biotin-labelled human MLL-AF9 followed by streptavidin ChIP-seq. This identified 139 MLL-AF9 targets (Bernt et al., 2011).

The ChIP-seq experiments outlined in these studies provided a large-scale method for obtaining genes potentially important in MLL-r leukaemias. Microarray data from earlier studies was used to show that two thirds of the MLL-AF4 gene targets were overexpressed in clinical MLL-r leukaemias compared to non-MLL-r leukaemias (Guenther et al., 2008). Not only did this implicate gene targets of the fusion protein as important to leukaemia development but also showed that MLL-r cell lines are representative models of clinical MLL-r leukaemias. Studies that looked at specific gene examples, verified that these genes overexpressed in MLL-r leukaemia patients (*BCL6* and *RUNX1*) were direct targets of the fusion protein. The binding of MLL-AF4 was shown to be a direct influence on the gene expression since knockdown of MLL-AF4 led to their downregulation (Geng et al., 2012; Wilkinson et al., 2013).

A genome-wide study of MLL-AF6 in ML-2 cells showed that promoters occupied by the fusion protein were almost 2-fold enriched for unmethylated CpGs compared to non-fusion protein targets, indicative of CXXC domain targeting, despite a comparable CpG content between the two sets of genes (Okuda et al., 2014).

ChIP-seq of H3K79me2 in MLL-r leukaemia cells lines revealed a strong overlap between H3K79me2 and MLL and AF4 ChIP-seq at MLL-AF4 target genes in both SEM and RS4;11 cells (Guenther et al., 2008; Wilkinson et al., 2013). Moreover, higher H3K79me2 signal was observed at MLL-AF4 targets genes in SEM cells compared to the same genes in REH cells (Guenther et al., 2008). MLL-AF9 targets had higher

H3K79me2 signal compared to non-targets with a similar expression level In MLL-AF9-transformed mBM cells (Bernt et al., 2011). These MLL-AF9 targets also had higher levels of H3K79me2 in primary AML samples with MLL-AF9 compared to AML samples with a normal karyotype. Another study assessed these gene targets in the human MLL-AF9 cell line THP-1 and found they had a significantly higher enrichment of H3K79me2 even compared to highly expressed genes (Wang et al., 2013).

Whilst an increase in H3K4me3 signal was also observed at MLL-AF4 target genes in SEM cells, this may not be fusion protein specific since a similar increase was observed at the same genes in REH cells (Guenther et al., 2008). However, a separate study in mBM cells before and after transformation with MLL-AF9 showed an increase in both H3K4me3 and H3K79me2 levels at gene targets of the fusion protein post-transformation (Bernt et al., 2011). In many cases the H3K4me3 pattern matches that of the fusion protein and H3K79me2 but the correlation is less strong (Guenther et al., 2008).

1.9. Aim of this D.Phil.

Whilst there is complete agreement in the field about the importance in MLL-r leukaemia development of the Menin/LEDGF complex, the MLL-FP CXXC domain and, by extension, the PAF elongation complex, it is clear that the roles played by these in recruiting MLL-FPs to genes is still up for debate. If a role exists for the Menin/LEDGF complex in MLL-FP recruitment the importance of this role in relation to that of the CXXC domain is not clear. Different experimental outcomes regarding these domains could be loci- or cell-type specific, or simply a result of the technique used.

As previously described, the roles of Menin, LEDGF and PAFc in gene regulation and MLL-FP recruitment in the context of MLL-r leukaemias has been restricted to the 5'

HoxA9 cluster as well as a few other genes such as *Meis1*. Importantly, the hypothesis that the Menin/LEDGF complex recruits MLL-FPs to these loci has not been tested but only suggested based on binding correlations and knockout or induced systems, which suffer from the potential of indirect effects. Moreover, some on these studies indicate that the fusion protein might actually be targeting Menin and LEDGF to chromatin. In order to assess the recruitment potential of these proteins, an isolated system is required that is not influenced by other factors other than those in question. Furthermore, a bi-directional approach is required to see if the recruitment is restricted to a certain hierarchy.

Whilst genome-wide studies in MLL-r leukaemias have characterised MLL-FP binding, histone marks and proteins associated with transcription, none have analysed genome-wide binding of Menin, LEDGF or PAFc. The well-known importance of these proteins makes them crucial candidates for assessing genome-wide binding and function of MLL-FPs. Furthermore, when assessing gene expression, these studies relied on microarray data from patient samples. The easily-obtainable, large quantities of cell lines lend themselves to the more precise assessment of RNA- and nascent RNA-sequencing.

This DPhil analyses mechanisms by focusing on two MLL-FPs, MLL-AF4 and MLL-AF6, in the human cell lines SEM and ML-2, respectively. MLL-AF4 is important due to its prevalence among MLL-r leukaemias, particularly in cases of infant ALL. On the opposite end of the spectrum, MLL-AF6 is responsible for a large proportion of adult MLL-r AMLs. Furthermore, the particular aggressiveness of the leukaemias associated with these two fusion proteins makes them important to understand. Determining a mechanism for genome-wide MLL-FP recruitment will aid our understanding of the molecular mechanisms driving these leukaemias and provide a platform for effective targeted therapies.

2. Chapter Two – Materials and Methods

2.1. Mammalian Cell Culture

2.1.1. Media

Dulbecco's Modified Eagle's Medium (DMEM)

[+]4.5g/L D-Glucose, [+]L-Glutamine, [-]Pyruvate DMEM (#41965-039, Gibco, Life Technologies) supplemented with 10% Heat-Inactivated Fetal Bovine Serum (FBS, #10500-064, Gibco, Life Technologies) and 1% GlutaMAX™-I CTS™ (100X) (#A12860-01, Gibco, Life Technologies).

Iscove's Modified Dulbecco's Medium (IMDM)

[+]L-Glutamine, [+]25mM HEPES (#21980-032, Gibco, Life Technologies) supplemented with 10% Heat-Inactivated Fetal Bovine Serum (FBS, #10500-064, Gibco, Life Technologies) and 1% GlutaMAX™-I CTS™ (100X) (#A12860-01, Gibco, Life Technologies).

Opti-MEM® I + GlutaMAX™-I (Opti-MEM)

Reduced Serum Medium (#51985-026, Gibco, Life Technologies).

RPMI Medium 1640 (RPMI)

[+]L-Glutamine (#21875-034, Gibco, Life Technologies) supplemented with 10% Heat-Inactivated Fetal Bovine Serum (FBS, #10500-064, Gibco, Life Technologies) and 1% GlutaMAX™-I CTS™ (100X) (#A12860-01, Gibco, Life Technologies).

2.2. Cell Lines and Maintenance

2.2.1. Cell Lines

All cell lines were maintained at 37˚C, 5% CO_2.

Table 2.1 Cell Lines

Cell Line	Company, Cat. #	Media	Supplements	Notes
HEK293T	Dharmacon, #HCL4517	DMEM		
ML-2	DSMZ	RPMI		Carry t(6;11) translocation (MLL-AF6)
E13.5 Mouse Embryonic Fibroblasts (**MEF**)	Gift from Alan Engelman	DMEM	100Unit/mL Penicillin; 100µg/mL Streptomycin	*LEDGF*$^{f/+}$ (E1) and *LEDGF*$^{-/-}$ (E2) variants (Shun et al., 2008)
E14 Mouse Embryonic Stem Cells (**mESC**)	Gift from Rob Klose	DMEM	100Unit/mL Penicillin; 100µg/mL Streptomycin; LIF; NEAA; β-mercaptoethanol	TOT2N variant (Blackledge et al., 2014)
SEM	DSMZ, #ACC 546	IMDM		Carry t(4;11) translocation (MLL-AF4)

ML-2 and SEM cells grow in suspension. They were maintained at cell numbers between 5 x 10^5 and 2 x 10^6 cells/mL, being passaged every second or third day depending on cell density. Cell densities were determined by counting live cells using a haematocytometer.

293T, MEF and mESC are adherent. Cells were grown in flat-bottomed plates and passaged at 90% confluency to a 1:10 dilution. Media was aspirated and replaced with 1X Trypsin (0.5% Trypsin-EDTA (10X), #15400-054, Gibco, Life Technologies) in Phosphate Buffered Saline (PBS pH7.4, #10010-015, Gibco, Life Technologies). Cells were returned to 37˚C until they were detached. Trypsin was inactivated using FBS-supplemented DMEM before the cells were added to a fresh plate. For mESCs, the new plate was pre-emptively covered in 0.1% Gelatin (Gelatin Solution, Type B, 2% in H_2O, #G1393-100ML, Sigma, Life Science) diluted in PBS and left for 5 minutes before the solution was aspirated.

2.2.2. Mycoplasma Testing

Cells were spun at 90rcf for 10 minutes and the supernatant was used to test for *Mycoplasma* contamination using a luminometer (Lucetta Luminometer, Lonza). The

MycoAlert Mycoplasma Detection Assay (Lonza) was followed as per the manufacturer's instructions. Only cells free of *Mycoplasma* contamination were used for cell culture.

2.2.3. Storage and Recovery of Cell Lines

For storage, cells were collected and spun at 90rcf for 10 minutes to pellet them. They were then resuspended in 80% FBS, 20% Dimethyl-Sulphoxide (DMSO Hybri-MAX®, #D2650, Sigma, Life Science) and aliquoted in 1mL into freezing vials. Vials were then transferred to a Cryo 1˚C Freezing Container (#5100-0001, Nalgene) and put in a -80˚C freezer overnight. The next day cells were transferred to liquid nitrogen storage facilities.

For recovery, cells were put directly on to dry ice from the liquid nitrogen. Vials were then submerged in 37˚C water with gentle shaking until ice was no longer visible. The 1mL aliquot of cells was then immediately transferred to the appropriate media. In order to remove the DMSO, cells were spun at 90rcf for 10 minutes, the media was aspirated and cells were resuspended in fresh media. Cells were transferred to an appropriate flask or plate before being put in the incubator.

2.3. Transfection of Mammalian Cells

2.3.1. DNA Transfections

2.3.1.1. Lipofectamine 2000

Mouse ESCs were seeded 1:10 into a fresh 15cm plate and transfections were conducted when cells were at 60-70% confluency. 80µL Lipofectamine 2000 Reagent (#11668-019, Invitrogen, Life Technologies) was added to 2mL Opti-MEM and incubated at room temperature (RT) for 5 minutes. 40µg DNA construct was added to a separate 2mL Opti-MEM. After the 5-minute incubation, both mixtures were combined and incubated at RT for 20 minutes. The solution was then added drop-wise to the plate of mESCs before it was returned to the incubator.

2.3.1.2. PEI

Confluent 293T cells were seeded 1:10 into a fresh 15cm plate and transfections were conducted when cells were at 60-70% confluency. Polyethylenimine (PEI, 25kDa, branched, #408727-100ML, Sigma-Aldrich) stock was diluted in water to 1mg/mL, adjusted to pH7.0 and filter sterilised. 15µg DNA construct was added to 1mL Opti-MEM. 75µL PEI was added to the mixture which was then vortexed for 10 seconds and incubated at RT for 15 minutes. During this incubation, media on the 293T cells was replaced with DMEM supplement with reduced (2%) FBS to increase transfection efficiency. DNA solution was added drop-wise to the plate which was then returned to the incubator.

2.3.1.3. Transit-LT1

Confluent 293T cells were seeded 1:10 into a fresh 15cm plate and transfections were conducted 24 hours later. 45µL Transit-LT1 Reagent (#MIR2300, Mirus) was added to 1.4mL Opti-MEM and incubated at RT for 5 minutes. 15µg DNA plasmid was added to the solution and incubated at RT for 20 minutes. The mixture was added drop-wise to the plate of 293T cells which was then returned to incubator.

2.3.1.4. MEF2 Nucleofector

MEFs were trypsinised and counted. Appropriate cell numbers were spun at 90rcf for 10 minutes to pellet cells. Media was aspirated and aliquots of 2 x 10^6 cells were resuspended in 100µL MEF2 Nucleofection solution (#VAPD-1005, Lonza). 5µL DNA construct was added to the cell mixture which was then transferred to a cuvette and used in the AMAXA transfection machine, programme T-20. Cells were resuspended in 500µL warm DMEM and transferred to 10ml pre-warmed DMEM in a 10cm dish which was then returned to the incubator.

2.3.2. Retroviral Transductions

To package DNA into a virus, 6 x 10^6 293T cells were seeded in a 10cm plate. 24 hours later 80µL Transit-LT1 was added to 3mL OptiMem and incubated at RT for 5 minutes. 35µg DNA plasmid and 20µg EcoPak (containing Gag-Pol-Env) were added to the OptiMem mix and incubated at RT for 20 minutes. The mixture was then added drop-wise to the 10cm plate. Media was replaced 24 hours post-transfection. Virus-containing supernatant was harvested at 48-, 72- and 96-hours post-transfection, filtered through a 0.45µM filter, and frozen in aliquots at -80˚C. To generate stably expressing cell lines, $4x10^5$ fresh 293T cells were plated in a 12-well plate. 1.5mL was added to each well along with 2µL Lipofectamine Reagent (#18324-010, Invitrogen, Life Technologies). The plate was wrapped in cling film at spun at 1500g for 90 minutes.

2.4. Generation of Stable Cell Lines

24 hours post-transfection (2.3), puromycin was added to cell culture media at a concentration of 1µg/mL to kill untransfected cells. After 8-10 days colonies of resistant cells could be seen. Individual colonies were picked using a 200µL pipette and transferred to individual wells of a 96-well plate. Colonies were left in Trypsin at 37˚C for 30-45 minutes to break them up. Trypsin was inactivated with FBS-supplemented DMEM and cells were left to grow at 37˚C with the media being changed the next day. Cell lines were continuously grown in puromycin-supplemented media. Where applicable, GFP expression was also used to test for positive cell lines under a fluorescent microscope. Every 2-3 days cells needed to be expanded into larger wells. Cells were Trypsinised and sequentially moved to 24-well, to 12-well, to 6-well plates. Confluent wells of a 6-well plate were used for freezing down cell lines (2.2.3) and for whole cell extracts to test for protein expression (2.5.1/.2/.3).

2.5. Protein Techniques

2.5.1. Whole Cell Extracts

Cells were collected and spun at 90rcf for 10 minutes. Media was aspirated and the cell pellet was resuspended in 1mL PBS before being transferred to 1.5mL Eppendorf tube. Cells were then spun at 420rcf for 3' minutes before PBS was removed and the pellet was resuspended in 1mL cold BC300 + 0.5% NP-40 (IGEPAL® CA-630, #I8896-100ML, Sigma) + protease inhibitors (EDTA-free protease inhibitor cocktail, Roche). The cell lysis suspension was incubated end-over-end at 4˚C for one hour. After the incubation, the lysed material was spun at 17,900rcf for 10 minutes at 4˚C. The supernatant was transferred to a new Eppendorf tube and stored at -80˚C.

2.5.1.1. BC300 Buffer

20mM Tris-HCL pH7.5; 20% glycerol; 300mM KCl; 5mM EDTA

2.5.2. Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Whole cell extracts mixed with 1X LDS Sample Buffer (NuPAGE® LDS Sample Buffer (4X), #NP0008, Novex, Life Technologies) were incubated at 95˚C for 5 minutes to denature proteins before being loaded onto polyacrylamide gels. For the detection of proteins up to approximately 200kDa, pre-cast 4-12% Bis-Tris gels (Novex, Life Technologies) were used. For proteins larger than 200kDa, pre-cast 3-8% Tris-Acetate gels (Novex, Life Technologies) were used. 180V were supplied to the gels assembled in a tank filled with either 1X MOPS SDS Running Buffer (NuPAGE® MOPS SDS Running Buffer (20X), #NP0001, Novex, Life Technologies) (Bis-Tris gels) or 1X Tris-Acetate SDS Running Buffer (#LA0041, Novex, Life Technologies) (Tris-Acetate gels), diluted in water. Gels were run for at least an hour, ensuring appropriate migration and separation of molecular weights as determined by the protein ladder.

Proteins were transferred to a 60cm^2 PVDF membrane (Immobilon®-P Transfer Membrane, #IPVH00010, EMD Millipore) using a wet transfer setup. Gel and membrane were sandwiched between filter paper and submerged in a tank filled with transfer buffer. 100V was supplied to the tank for 1 hour.

2.5.2.1. Transfer Buffer

192mM glycine; 25mM Trizma® base; 10% methanol

2.5.3. Western Blotting (WB)

Membranes were incubated with the desired antibody in 10mL 5% milk in TBS, overnight at 4˚C. The following day the membrane was washed 3 times for 30 minutes in 0.05% TBS-Tween. The membrane was subsequently incubated in 10mL 5% milk with the appropriate secondary antibody, at a concentration of 1:10000. The membrane was then washed a further 3 times for 30 minutes each with 0.05% TBS-Tween. Detection of proteins was achieved using either “strong” or “weak” ECL. Strong ECL (ECL Prime Western Blotting Detection Reagent, #RPN2232, Amersham, GE Healthcare) solutions A & B were mixed in the ratio 1mL:1mL and added over the entire membrane and left on for 5 minutes. Excess solution was removed and the membrane was developed using X-ray film (Amersham Hyperfilm ECL, #28906837, GE Healthcare) and an X-ray developing machine (Xograph Imaging). Weak ECL (Pierce ECL Western Blotting Substrate, #32106, Thermo Scientific) solutions were also mixed 1mL:1mL but were only left on the membrane for 1 minute prior to development.

Table 2.2 Antibodies

Antibody	Company & catalogue #	Dilution and application	Species
AF4	Abcam, #ab31812	1:1000 WB; 1:500 IP; 1:500 ChIP	Rabbit
BRD4	Bethyl, #A301-985A	1:500 ChIP	Rabbit
CFP1	Bethyl, #A303-161A	1:500 ChIP	Rabbit
ENL	Bethyl, #A302-268A	1:500 ChIP	Rabbit
EZH2	Cell Signalling, #5246/4	1:500 ChIP	Rabbit
Flag-Strep-Strep (FS2)	Gift from Rob Klose	1:1000 WB; 1:333 ChIP	Mouse
GAPDH	Bethyl, #A300-641A	1:10000 WB	Rabbit
H3K27ac	Diagenode, #pAb-196-050	1:500 ChIP	Rabbit
H3K27me3	Millipore, #07-449	1:500 ChIP	Rabbit
H3K36me2	Diagenode, #CS-127-100	1:500 ChIP	Rabbit
H3K36me3	Diagenode, #pAb-192-050	1:500 ChIP	Rabbit
H3K4me1	Diagenode, #pAb-194-050	1:500 ChIP	Rabbit
H3K4me3	Diagenode, #pAb-003-050	1:500 ChIP	Rabbit
H3K79me2	Millipore, #04-835	1:500 ChIP	Rabbit
HA	Bethyl, #A190-108A	1:333 ChIP	Rabbit
KDM2B	Gift from Rob Klose	1:333 ChIP	Rabbit
LEDGF	Bethyl, #A300-848A	1:5000 WB; 1:500 ChIP	Rabbit
LEO1	Bethyl, #A300-175A	1:500 ChIP	Rabbit
MED1	Bethyl, #A300-793A	1:500 ChIP	Rabbit
Menin	Bethyl, #A300-105A	1:5000 WB; 1:500 ChIP	Rabbit
MLL1	Bethyl, #A300-086A	1:5000 WB; 1:500 IP; 1:500 ChIP	Rabbit
MLL1 (DNMT)	Bethyl, #A300-087A	1:500 ChIP	Rabbit
PAF1	Bethyl, #A300-172A	1:500 ChIP	Rabbit
Strep	Qiagen, #34850	1:2000 WB; 1:200 ChIP	Mouse
SUZ12	Cell Signalling, #3737/3	1:500 ChIP	Rabbit

2.5.4. Immunoprecipitation (IP)

Protein A and Protein G Dynabeads (#10001/4D, Novex, Life Technologies) were resuspended in BC300 and added to whole cell extracts for pre-clearing. Two 30-minute incubations were done with fresh Dynabeads end-over-end at 4˚C. Tubes were spun at 17,900rcf for 1 minute at 4˚C and the supernatant was transferred to a new Eppendorf tube. IP antibody was added to the extract and incubated end-over-end at 4˚C overnight. The following day 10µL Dynabeads (Protein A for rabbit IP antibody; Protein G for mouse IP antibody) were added to the tubes and incubated end-over-end at 4˚C for 5 hours. Beads were washed twice in BC300 using a magnet (DynaMag-2, #123.21D, Invitrogen) and then rinsed in PBS. PBS was removed and proteins were eluted from beads using 2X LDS Sample Buffer + β-mercaptoethanol (#63689, Sigma) at 95˚C for 5'. Using a magnet, solutions were loaded directly onto a gel for SDS-PAGE (2.5.2).

2.5.5. Chromatin Immunoprecipitation (ChIP)

Day 1

Each individual ChIP experiment was carried out on 1 x 10^7 cells. Cells were counted and rinsed in PBS before double- (protein) or single- (histone mark) fixation. For double fixation, cells were fixed in 1mL 2mM disuccinimidyl glutarate (DSG, #80424-50MG-F, Sigma) end-over-end at room temperature for 30 minutes, followed by 1mL 1% formaldehyde (FA) end-over-end at room temperature for 30 minutes. For single fixation, cells were fixed in 1mL 1% FA end-over-end for 10 minutes. Fixed material was spun at 13,000rpm for 1 minute and the pellet was rinsed in PBS before being frozen on dry ice.

Cell pellets were thawed on wet ice, resuspended in 100µL SDS lysis buffer + 1:200 protease inhibitor cocktail and syringe-passaged through a 27-gage (27G) needle to remove clumps. Lysed material was transferred to a small Covaris tube (microTUBE AFA Fiber Pre-Slit Snap-CAP 6x16mm, #520045, Covaris Inc.) for use in a sonicator

(S220, Covaris Inc.). Tubes were loaded into the degassed water at 4˚C and run using the Covaris SonoLite software. The specific parameters were:

Duty cycle = 10%

Intensity = 5

Cycles / burst = 200

Duration = 300s

For single-fixed material this setup was run once (i.e. a single 5-minute run) whereas for double-fixed material it was run twice consecutively (a 10-minute run).

After sonication, samples were transferred to Eppendorf tubes and spun at 13,000rpm for 10 minutes at 4˚C. Supernatant was retained at diluted in ChIP dilution buffer (CDB) at a factor of 1:11. Protein A/G Dynabeads resuspended in CDB were added to the ChIP solution at a 1:500 concentration from pre-clearing. Two 30-minute incubations were done with fresh Dynabeads, end-over-end at 4˚C. After pre-clearing, tubes were spun at 2200rcf for 3 minutes and ChIP solution was split into 1mL aliquots with 3 aliquots of 50µL being retained for input material and to test sonication efficiency. Antibodies were added to the 1mL aliquots and incubated end-over-end at 4˚C overnight.

Day 2

10µL of Protein A/G Dynabeads resuspended in CDB were added to each 1mL ChIP aliquot and incubated end-over-end at 4˚C for 5 hours. Beads were washed 3 times in RIPA wash buffer using a magnet, switching to fresh Eppendorf tubes after the first wash. Beads were rinsed in TE + 50mM NaCl, spun at 1090rcf for 3 minutes, resuspended in 210µL elution buffer and put at 65˚C for 30 minutes to elute material. Beads were spun at 17,900rcf for 1 minute and 200µL eluted material was transferred to new Eppendorf tubes before reverse crosslinking overnight at 65˚C. Input samples were diluted to 200µL and put at 65˚C overnight.

Day 3

1µL RNAse A (10mg/mL, #EN0531, Thermo Scientific) was added to each sample and placed at 37˚C for 1 hour. 1µL Proteinase K (20mg/mL, #EO0491, Thermo Scientific)

was added to each sample and put at 45˚C for 1 hour. DNA was extracted using a PCR Purification kit (#28106, Qiagen), and eluted in 60-140µL water depending on the volume required. Samples were stored at -20˚C until needed.

2.5.5.1. SDS Lysis Buffer / Elution Buffer

50mM Tris-HCl pH8.1; 10mM EDTA; 1% SDS

2.5.5.2. ChIP Dilution Buffer (CDB)

16.7mM Tris-HCl pH8.1; 167mM NaCl; 1.2mM EDTA; 0.01% SDS; 1.1% Triton X-100 (#X100-100ML, Sigma)

2.5.5.3. RIPA Wash Buffer

50mM HEPES-KOH pH7.6; 500mM LiCl; 1mM EDTA; 1% NP-40; 0.7% Na-Deoxycholate

2.5.6. Biotinylated CXXC Affinity Purification (Bio-CAP)

50µL biotinylated KDM2B CXXC domain (0.5µg/µL, a generous gift from Rob Klose) was diluted with 425µL BC150 and added to 25µL BC150-rinsed magnetic neutravidin beads (Sera-Mag SpeedBeads NeutraAvidin Microparticles, Thermo Scientific). As a control, 25µL beads were also added to 475µL of BC150 alone. Tubes were rotated end-over-end at 4˚C for 1 hour. Beads were washed 3 times with 1M CAP Buffer, using a magnet, and then 100mM CAP Buffer. 1M washes remove contaminating bacterial DNA and 0.1mM wash equilibrates beads. 100µL DNA (100ng/µL) was diluted with 900µL BC150 (10µL DNA (10%) was diluted with 90µL BC150 for input). 500µL of DNA in BC150 was added to KDM2B-hybridised beads and the other 500µL was added to the beads-only control. Tubes were incubated end-over-end at 4˚C for 1 hour. After incubation, supernatant was removed from beads using a magnet – this was the flow-through material. Beads were washed twice with 100mM CAP Buffer. Material was then eluted from beads using increasing concentrations of salt: 300mM, 500mM, 700mM and 1M. This was done in 50µL CAP buffer at room temperature for 10 minutes, with gentle

tapping, twice to give a total volume of 100µL for each elution. DNA was purified using a PCR purification kit and stored at -20˚C until needed.

2.5.6.1. CAP Buffer

100/300/500/700/1000mM NaCl; 20mM HEPES pH7.9; 0.1% Triton X-100; 12.5% glycerol

2.6. Nucleic Acid Techniques

2.6.1. RNA Extraction

To extract whole cell RNA a Qiagen RNeasy Mini Kit (#74106) was used as per the manufacturer's instructions. RNA was stored at -80˚C.

2.6.2. Generation of cDNA

8µL RNA was mixed with 1µL dNTP Mix (10mM, #18427-013, Invitrogen) and 1µL random hexamers (50µM, #S06405, Roche) and placed at 65˚C for 1 minute before being snap-cooled on ice for 1 minute. The mixture was added to 10µL reaction master mix and the entire 20µL was transferred to a PCR tube. The tube was run on the PCR machine using the following parameters:

25˚C for 10' (primer extension)

50˚C for 50' (cDNA synthesis)

85˚C for 5' (Termination)

The PCR reaction mixture was diluted with 200µL water before quantification using a NanoDrop (Thermo Scientific). cDNA was stored at -20˚C.

2.6.2.1. Reverse Transcription Reaction Master Mix

4µL First-Strand Buffer (5X, #y02321, Invitrogen); 2µL water; 2µL DTT (0.1M #y00147, Invitrogen); 1µL RNaseOUT (40units/µL, #51535, Invitrogen); 1µL SuperScript III Reverse Transcriptase (200units/µL, #56575, Invitrogen).

2.6.3. RT-qPCR

ChIP DNA or cDNA was used in Real Time Quantitative Polymerase Chain Reaction (RT-qPCR) for the detection of gene targets or gene expression, respectively.

For SYBR green primers, 5µL sample or input DNA was added to 20µL reaction mix in a single well of a 96-well PCR plate (MicroAmp Fast Optical 96-well Reaction Plate, #4346906, Applied Biosystems). For TaqMan primers, 5µL sample or input DNA was added to 15µL reaction mix in a single well. Plates were sealed with film (MicroAmp Adhesive Film, #4311971, Applied Biosystems) and spun at 1000rpm for 1 minutes before being run on a RT-qPCR machine (7500 Fast Real Time PCR System, Applied Biosystems).

Reactions were run using the following RT-qPCR program ($\Delta\Delta C_T$; Fast; DNA):

Table 2.3 RT-qPCR Run Conditions

Temperature (˚C)	Time (s)	Cycles	Reason
95	20	1	Initial denaturation
95	3	40	Denaturation
60	30		Annealing & Extension
95	15	Continuous	Melting Curve (SYBR Green primers only)
60	60		
95	15		
60	15		

2.6.3.1. SYBR Green Reaction Mix

12.5µL Fast SYBR® Green Master Mix (#4385612, Applied Biosystems); 0.25µL forward primer (10µM); 0.25µL reverse primer (10µM); 7µL water

2.6.3.2. Taqman Reaction Mix

10µL TaqMan® Fast Advanced Master Mix (#4444557, Applied Biosystems); 0.5µL forward primer (10µM) 0.5µL reverse primer (10µM); 0.5µL TaqMan probe (5µM); 3.5µL water

2.6.3.3. Calculations

ChIP (% Input)

To calculate the ChIP signal as a percentage of the input, the following formula was used:

i) $\Delta C_T = C_T(\text{input}) - C_T(\text{sample})$

ii) % Input = $2^{\Delta CT} \times 5$

The input sample represented 5% of the total IP material hence multiplication by 5.

Gene Expression

For gene expression, samples were always run in parallel with a *GAPDH* control for normalisation. A reference was also used for fold difference of gene expresison (e.g. untreated or wild-type cell line). The following calculation was used:

i) $\Delta C_T = C_T(\textit{Gene}) - C_T(\textit{GAPDH})$

ii) $\Delta\Delta C_T = \Delta C_T(\text{sample}) - \Delta C_T(\text{reference})$

iii) Fold difference in expression = $2^{-\Delta\Delta CT}$

2.6.3.4. Primer Efficiency

To test the efficiency of primers for material of varying concentration, they were run in reactions with a serial dilution of (100, 10, 1, 0.1ng) cDNA to generate a standard curve. Only primers with good PCR efficiency and yielding a single amplification product as assesed by melting curve, were used.

Table 2.4 Primers

Primer Name	Organism	Purpose	Region	Sequence (5' to 3')	TaqMan Probe (5' to 3')
HOXA9 Coding For	Human	ChIP	*HOXA9* exon1	atgcttgtggttctcctccagttg	SYBR green
HOXA9 Coding Rev				ccgccgctctcattctcagc	
18205 For			18.2kb upstream of *Myc* – negative control region	ggctcctgtaaccaaccactacc	
18205 Rev				cctctgggctggcttcattc	
HOXA9 P1 For	Mouse		*HOXA9* promoter	tgacccctcagcaagacaaac	cctcttgatgttgactggcgattttccc
HOXA9 P1 Rev				tcccgctccccagactg	
HOXA9 1461 For			1.4kb upstream of *HOXA9* promoter	tagactcacaaggacaatatctcctttt	tgaattttccccctttgggccac
HOXA9 1461 Rev				aggtactgagtattaagcagctgtttaca	
HOXA9 TATA2 For			*HOXA9* TSS	cgcgatccctttgcataaaa	attatgactgcaaaacaccgggccatt
HOXA9 TATA2 Rev				cgtaaatcactccgcacgct	
HOXA9 2207 For			*HOXA9* exon1	ggtgcgctctccttcgc	taccctccagccggccttatggc
HOXA9 2207 Rev				gcatagtcagtcagggacaaagtgt	
HOXA9 HD For			*HOXA9* exon 2	tctctctccctccgcagataac	ccaactggctacatgctcgctcca
HOXA9 HD Rev				gggcatcgcttcttccg	
HOXA9 FeExon For			*HOXA9* foetal exon	gccatcaaggcctaatcgtg	cctgcggtggcaacctcagatcc
HOXA9 FeExon Rev				aagacccgaagctcctcctg	
TetO -10 For			10kb upstream of TetO array	ggcaaggatacagggagtca	SYBR Green
TetO -10 Rev				aaggaccaccttagccaaca	
TetO -5 For			5kb upstream of TetO array	ttggacatagcactgagaaagaag	
TetO -5 Rev				tccagacagggactttgagtg	
TetO -1 For			1kb upstream of TetO array	ttttcctcatgatggtacatgg	
TetO -1 Rev				aatttgttgtaggtgccattca	
TetO -0.5 For			500bp upstream of TetO array	ctttaggaggggagctgcat	
TetO -0.5 Rev				ttgccagttttaagagaatttgc	
TetO For			Within TetO array	tgtgggcaggtcacaagtta	
TetO Rev				atctcatccccaatgctgtc	
TetO +0.5 For			500bp downstream of TetO array	aaaggaaaatgaggggccta	
TetO +0.5 Rev				caggcaggtaaacacgataagtt	
TetO +1 For			1kb downstream of TetO array	caaatccgtccttgatcttga	
TetO +1 Rev				gcagtaggatggggagacag	
TetO +5 For			5kb downstream of TetO array	tcagatgtacctagctgtgatg	

TetO +5 Rev	Mouse	ChIP		cactgagcatcaatcacattg	SYBR Green
TetO +10 For			10kb downstream of TetO array	ttcttccaaggccctaacct	
TetO +10 Rev				caaggcaacgtcaaacaatg	
BCL11A CGI1 For	Human		CGI in *BCL11A* intron 1	acacccagtgcccagaattg	
BCL11A CGI1 Rev				cgcgggtcctgagattcatt	
BCL11A CGI2 For			CGI in *BCL11A* intron 2	aggctgagtgttgtggagac	
BCL11A CGI2 Rev				cgttggaagctgcctttgtt	
BCL11A CGI3 For			CGI (in SEM not 293T) in *BCL11A* intron 2	gccaggctaacatagacctcc	
BCL11A CGI3 Rev				gaggccaactcttccactcc	
ARID1B CGI2 For			CGI in *ARID1B* intron 1	aagtgccgtgaccttcacat	
ARID1B CGI2 Rev				ccttgcattacctctgccca	
CPEB2 CGI For			CGI in *CPEB2* intron 1	attgctggagaaggtgcgtg	
CPEB2 CGI Rev				tccctcattggacagcgaga	
CPEB2 non-CGI For			Non-CGI region in *CPEB2* intron 3	catatgggccatcctattctctg	
CPEB2 non-CGI Rev				gctctggttctaactcctagaaa	
MLL-AF4 For	Human	cDNA	Spans exons either side of breakpoint	aggtccagagcagagcaaac	SYBR Green
MLL-AF4 Rev				cggccatgaatgggtcatttc	
MLL1 For	Mouse		Spans exons 22 and 23	ctgaatgacctctctgactgtgaaga	actctttcctattggataccagtgttctcggg
MLL1 Rev				ggcatctgtggtgctccagta	
GAPDH For				gtctcctgcgacttcagc	SYBR Green
GAPDH Rev				tcattgtcataccaggaaatgagc	

2.7. Cloning

2.7.1. Polymerase Chain Reaction (PCR)

All PCR for the purpose of cloning was done using the KAPA HiFi HotStart as this was the best compromise between efficiency and accuracy. 10ng plasmid DNA was mixed with 1.5µL 10mM forward/reverse primers and 12.5µL 2X KAPA HiFi HotStart Ready Mix. If the PCR product was particularly GC-rich (>70%), 5% DMSO was also added. The mixture was made up to 25µL with water and run on the PCR machine (T100 Thermal Cycler, Bio-RAD) using the following program:

Table 2.5 KAPA HiFi HotStart PCR Conditions

Temperature (˚C)	Time (s)	Cycles	Reason
95	300	1	Initial denaturation
98	20	1	Denaturation
65	15	1	Annealing
72	30*	29	Extension
72	60	1	Final extension
4	Infinite	1	Cooling

*Extension time is adjusted for length of PCR product (approximately 60s/kb)

Depending on the cloning reaction the product was to be subsequently used for, different flanking sequences were added to the primers (e.g. restriction digest sites). If the PCR was clean and the size of PCR product was >200bp it was purified using a PCR purification kit else a gel extraction kit was used.

2.7.2. Restriction Digests

For single digests, 1µg DNA contruct was digested with 1µL of the appropriate restriction enzyme (New England Biosciences) and 2µL buffer (10X CutSmart Buffer, #B72045, NEB) in a total volume of 20µL. The reaction was carried out at 37˚C for 1 hour. For double digests, 1µL of each enzyme was added to the DNA with 2.5µL buffer in a total of 25µL. This extra volume was to prevent inhibition of the reaction by the glycerol that the

enzymes are supplied in (can occur if glycerol >10% total volume). To assess efficiency of digestion, DNA was run on an agarose gel (concentration varied according to size of DNA) and visualised on a molecular imager (Gel Doc XR+, Bio-RAD). Depending on the cloning experiment, DNA was either purified using a PCR purification kit (#28106, Qiagen) or the DNA band was excised and purified using a gel extraction kit (#28704, Qiagen), following the manufacturer's instructions.

DNA fragments with matching sticky ends were ligated together using 1µL T4 DNA Ligase (400,000units/mL, #M0202S, NEB) and 2µL buffer (10X T4 DNA Ligase Buffer, #B0202S, NEB) in a total volume of 20µL, at room temperature for 2 hours. Different ratios of insert to vector were used to optimise the ligation. Specifically 5:1, 3:1 and 0.2:1 ratios of ends were used alongside a vector only control.

2.7.3. InFusion Cloning

DNA to be inserted into a vector was amplified by PCR (2.7.3.) flanked by 15bp complimentary sequences to the vector either side of a unique restriction digest site. 100ng vector digested at this site was combined with 100ng amplified PCR product and mixed with 2µL buffer (5X In-Fusion HD Enzyme Premix, Clontech Laboratories Inc.) in 10µL total volume. The reaction was put at 50˚C for 15 minutes and then placed on ice. During the 15 minute incubation, homology-directed repair between the complementary sequences results in insertion of the PCR product into the vector.

2.7.4. Ligation-Independent Cloning (LIC)

DNA to be inserted into a vector was amplified flanked 15bp LIC sites, complementary to identical sites flanking BaeI-digestion sites on a pre-designed LIC vector. The vector was digested with BaeI and the linearised vector was gel purified. Vector and PCR plasmid were incubated in the presence of T4 DNA polymerase at room temperature for 30 minutes as follows:

Table 2.6 T4 DNA Polymerase Reaction

	Cat. #, Company	Amount for vector (µL)	Amount for PCR product (µL)
DNA		15	15
T4 DNA Polymerase	3000units/mL, #M0203S, NEB	3	3
10X NEBuffer 2	#B7002S, NEB	5	5
10X BSA	#B9001S, NEB	5	5
2mM dGTP	#R0161, Thermo Sci.	2	-
2mM dCTP	#R0151, Thermo Sci.	-	2
Water		20	20

This reaction makes use of the 3' to 5' exonuclease activity of T4 DNA polymerase to chew back the LIC sites until continuous addition and removal of GTP (vector) or CTP (PCR product) inhibits further digestion. This generates 15bp overhangs complimentary between vector and PCR product. After the incubation the reaction was put at 75˚C for 20 minutes to heat-inactivate the enzyme. Insertion of the PCR product occurs when it is mixed with the vector since the overhangs are long enough to anneal at room temperature, with the DNA backbone being fully repaired when transformed into bacteria. For optimisation of the annealing, PCR product and vector were incubated in ratios 0.2:1, 1:1 and 2:1 alongside a vector only control. The reaction was done at room temperature for 30 minutes.

Table 2.7 Vectors

Vector Name	Company	Promoter	Markers	Antibiotic Resistance
LV067	Applied Biological Materials, Inc.	EF1a	GFP-2A-Puro^R	Kanamycin
MSCVpuro	Clontech Laboratories, Inc.	MSCV	Puro^R	Ampicillin
pCAGIpuro-LIC	Gift from Rob Klose	CAGpro	Puro^R	Ampicillin

2.7.5. Bacterial Transformations

100µL aliquots of frozen high transformation efficiency *E.coli* HST08 strains (Stellar Competent Cells, #636763, Clontech Laboratories Inc.) were thawed on ice and aliquoted. 5µL DNA from cloning reactions was added to 50µL cells and incubated on

ice for 30 minutes. For DNA from minipreps or maxipreps (2.7.6), 0.5µL DNA was added to 10µL cells. Bacterial cells were then heat-shocked at 42˚C for 45 seconds followed by immediate incubation on ice for 2 minutes. 450µL (or 990µL for mini/maxiprep DNA) pre-warmed SOC media was added to the bacteria and they were incubated in a 37˚C shaker for 30 minutes.100µL of the bacterial media was streaked on pre-warmed agar plates containing 50µg/mL of the appropriate antibiotic. 1:10 dilution was also done in case of overcrowding of the plate by the bacteria. Plates were incubated at 37˚C overnight. Plates were removed the following day and successful transformations, identified by the presence of significantly more colonies on the reaction plates compared to the vector-only control, were stored at 4˚C until required.

2.7.6. Plasmid Amplification and Sequencing

For minipreps, colonies were incubated overnight in 3ml LB supplemented with 50µg/mL of the appropriate antibiotic, at 37˚C, 220rpm. For maxipreps, colonies were incubated for 8 hours in 2mL LB supplemented with 50µg/mL of the appropriate antibiotic, in a 37˚C shaker. The 2mL bacterial solution was then transferred to 200mL LB plus 50µg/mL antibiotic and incubated in a 37˚C shaker overnight. The next day DNA was prepared using either a MiniPrep (Qiagen, #27104) or MaxiPrep (Endofree®, #12362, Qiagen) kit as per the manufacturer's instructions.

Purified DNA was sent for Sanger sequencing along with appropriate primers, using a pre-paid next-day service (Mix2Seq, EuroFins Genomics). Sequences were verified using SnapGene software (snapgene.com, GSL Biotech LLC).

2.8. ChIP-seq Analysis

2.8.1. Handling Raw Data

2.8.1.1. Mapping *fastq* files

Libraries of DNA from ChIP experiments were made by the Wellcome Trust and sequenced using a HiSeq 2500, 50bp paired-end (Wellcome Trust Centre for Human Genomics, Old Road Campus, Oxford). Subsequent computational analysis was performed by myself. Raw *fastq* files from sequencing runs were mapped to the *Homo sapiens* NCBI36/hg18 genome using *Bowtie* (http://bowtie-bio.sourceforge.net/index.shtml), with the following command:

bowtie -p 1 -m 2 --best --strata --sam --chunkmb 256 –sam /databank/indices/bowtie/hg18/hg18 -1 sequence_1.fastq -2 sequence_2.fastq output.sam

where *–p* specifies the number of parallel search threads, run on different processors
-m specifies the upper-limit of alignments that are accepted for a read
--best --strata reports only the best alignments, based on Phred quality
--chunkmb specifies the amount of megabytes of memory a thread is given to store data
--sam specifies that the output of the alignment be in *.sam* format

The outputted *sam* file was converted to a *bam* file using the *Samtools view* (http://samtools.sourceforge.net/) command. Directories of ChIP-seq tags (reads) were generated from the *sam* files using the *HOMER makeTagDirectory* (http://homer.salk.edu/homer/) command. The *makeUCSCfile* command was used to generate *bigwig* files from these tag directories; this normalised the tag counts per 10 million tags.

2.8.1.2. Peak-calling

SeqMonk (http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/) was used for peak-calling. *Bam* files of ChIP-seq tracks and their corresponding input tracks were imported to an hg18 genome, and normalised per million reads. Duplicate reads were removed at the import stage and peaks were generated using the “Contig Probe

Generator". Parameters were varied to give the best peak-set as determined by eye. Reads under peaks were normalised to reads in the input track and a cut-off was defined to remove those peaks that were less than 2-fold enriched over input. Peak sets were exported as *txt* files and converted to *bed* files using a self-written *Perl* script *ParseTxtToBed.pl* (Table 2.8).

2.8.1.3. Visualisation

All ChIP-seq tracks displayed in the results section (chapters 3, 4 and 5) are representations of *bigwig* files visualised using the UCSC Genome Brower (https://genome.ucsc.edu/). The displays represent tags per 10 million. Peak displays represent *bed* files generated from *SeqMonk* (2.8.1.2).

2.8.2. Genome-wide Analyses

2.8.2.1. Heat Maps

Heat maps in chapters 3 and 4 were generated using the *HOMER annotatePeaks* command with tag directories generated as described in 2.8.1.1, and peak sets generated from *SeqMonk* (2.8.1.2). These heat maps represent tracks as tags per 10 million per base-pair. Tag value matrices generated from this were visualised using *Java TreeView* (http://jtreeview.sourceforge.net/). Heat maps of spreading peaks in chapter 5 were generated by using the *Samtools view –c* command to get the reads under the peaks in 2.5kb bins. The output file was used in an R command (written by Ross Thorne) to generate the heat map.

2.8.2.2. Peak/Gene Set Comparisons

Peak set comparisons represented by Venn diagrams throughout the results section were determined using the self-written *Perl* script *OverlapPeakCall.pl*. This determines overlap by matching the genome co-ordinates of peaks. Exact overlap was always used such that even if two peaks had only 1bp gap between them they were not included in

the output. Gene targets of proteins were determined by matching peak sets with the hg18 *RefSeq* gene list in the self-written *Perl* script *PeakGeneTargets.pl*. This determined gene targets if the peak co-ordindates overlapped the gene at any point between the TSS or TES. The proximity could be altered to allow the peak to fall outside the gene by a certain distance (e.g. +/-2kb). Depending on the analysis, sometimes only gene targets where the peak overlapped the promoter were required. For this the self-written *PeakGeneTargetsTSS.pl* was used. This works in much the same way but the peaks have to overlap at the promoter which is user-defined as #kb either side of the TSS. For all examples in this thesis the promoter was defined as the TSS +/-2kb.

2.8.2.3. Composite Read Plots

Composite read plots were generated using the *HOMER annotatePeaks* command with tag directories generated as described in 2.8.1.1, and peak sets generated from *SeqMonk* (2.8.1.2). These plots represent tracks as tags per 10 million per base-pair. Tag values and co-ordinates were copied to *GraphPad Prism* (http://www.graphpad.com/scientific-software/prism/) and visualised as an XY graph.

2.8.2.4. Expression analysis

Nascent RNA-seq (generated by Erica Ballabio) was used to determine expression levels of genes. For a given gene, the raw reads over the exons were determined using the *Samtools view –c* command, supplied with the exon co-ordinates obtained from the hg18 *RefSeq* gene list. Reads over exons were normalised per million reads and per base-pair of exon using the self-written *Perl* script AllGenes_Expression_ExonCount_multRNA.pl. For gene expression of spreading MLL-AF4 targets, the mean average expression of the 149 genes was determined. For comparison, all MLL-AF4 targets were sampled in groups of 149 randomly selected genes. 1000 individual groups were sampled and the mean average expression of each group was plotted. The gene expression of the spreading MLL-AF4 targets was plotted relative to this standard curve.

Table 2.8 ChIP/RNA-seq

Cell Line	Antibody	ChIP done by	Library Prep.	Computational Analysis	Refers to figure
SEM	H3K4me3	Me	Wellcome Trust	Me	3.4; 3.7; 5.6
	H3K4me1				5.6
	H3K27ac				3.4; 5.6
	H3K79me2				3.4; 5.2
	H3K36me3				3.4
	Menin				3.5; 3.7; 5.1
	LEDGF				3.5
	KDM2B				3.3; 3.7; 5.1; 5.3
	(Bio-CAP)				3.2; 3.3; 5.1; 5.3; 5.5
	MLL(N)	Erica Ballabio			3.2; 3.4; 3.5; 3.6; 5.1; 5.3; 5.7
	AF4(C)				3.2; 3.4; 3.5; 3.6; 5.1; 5.3
	ATAC-seq				3.2; 3.3; 5.1
	ENL				5.3
	BRD4				5.6
	MED1	Tom Milne			5.4
	CFP1				3.3; 3.7; 5.1
	PAF1				3.6; 3.7; 5.3
	LEO1				3.6; 3.7
	EZH2				3.7; 5.3
	SUZ12				3.7
	RNAPII				5.3
	H3K27me3	Marta Tapia			3.4; 3.7
ML-2	H3K4me3	Me			3.4
	H3K27ac				3.4
	H3K79me2				3.4
	H3K36me3				3.4
	Menin				3.5; 5.1; 5.3
	LEDGF				3.5; 5.1
	KDM2B				3.3; 5.1

ML-2	(Bio-CAP)	Me	Wellcome Trust	Me	3.2; 3.3; 5.1
	MLL(N)	Erica Ballabio			3.2; 3.4; 3.5; 3.6; 5.1; 5.3; 5.7
	CFP1	Tom Milne			3.3
	PAF1				3.6
	LEO1				3.6
MV4;11	MLL(N)	Erica Ballabio			5.7
KOPN-8	MLL(N)				
THP-1	MLL(N)				
SHI-1	MLL(N)				
CCRF-CEM	MLL(N)				
RCH-ACV	MLL(N)	Tom Milne			
DNMT-SII 293T (Clone #7)	Strep	Me			4.2; 5.5
	(Bio-CAP)				4.2; 5.5
SEM	Nascent RNA	Erica Ballabio			3.4; 5.2

Table 2.9 Self-written *Perl* scripts for computational analyses

Script Name	Function	Used For
AllGenes_Expression_ ExonCount_multRNA.pl	Determine nascent RNA reads over exons of genes	Gene expression analysis
BedReadDensitySort.pl	Sorts a set of peaks based on number of reads for a given ChIP-seq track	Heat-maps
CompareGeneLists.pl	Determines overlap of gene targets in two separate lists	Venn diagrams
Menin_MA4_Read.pl	Determines reads of two ChIP-seq experiments under a supplied set of peaks	Menin/MLL-AF4 read correlation (XY) under MLL-AF4 peaks
OverlapPeakCall.pl	Determines the overlap between two ChIP-seq peak sets	Venn diagrams
PeakGeneTargets.pl	Determines the overlap between a ChIP-seq peak set and *RefSeq* hg18 genes	Venn diagrams; Gene expression analysis
PeakGeneTargetsTSS.pl	Determines the overlap between a ChIP-seq peak set and *RefSeq* hg18 gene TSS +/-2kb	Venn diagrams; Gene expression analysis
ParseTxtToBed.pl	Converts peak co-ordinate *txt* file output from SeqMonk to a 4-column *bed* file	UCSC visualisation; Peak overlaps; Heat-maps
SortBedByPeakLength.pl	Sorts peak bed files by length, with longest peak at the top	Spreading MLL-FP heat-maps
SpreadingMatrix.pl	Determines reads across the body of a spreading MLL-FP gene target and output them into a matrix of 2.5kb bins	Spreading MLL-FP heat-maps

2.9. Patient Data

Survival data from two studies (1. ECOG 2993, adult ALL, n=165; 2. COG P9906, childhood ALL, n=207) was analysed using a SuperPC analysis (Bair and Tibshirani, 2004), by Huimin Geng. Within a supplied group, genes that were above a given threshold of expression, based on microarray data, were used to determine prognostic score for each patient. Patients were divided into high-risk (prognostic score above the median) or low-risk (below the median) for survival analysis, plotted as Kaplan-Meier

curves. The R packages *superPC* and *survival* (version 2.35-8) were used for this analysis.

2.10. Statistics

2.10.1. ChIP-seq Read Correlations (Figure 3.5F)

Pearson correlation coefficient (r) for X- vs. Y- plotted ChIP-seq reads from two different datasets was calculated using GraphPad Prism, testing for a two-tailed significance. This was used to calculate R^2. For positive correlations, R^2>0.6 was deemed significant. The value for r is calculated using the following formula[1]:

$$r = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_i (x_i - \overline{x})^2} \sqrt{\sum_i (y_i - \overline{y})^2}}$$

where x_i/y_i = the sample value, and x-bar/y-bar = the sample mean.

2.10.2. Gene expression of spreading MLL-AF4 (Figure 5.2A)

Mean average gene expression of spreading MLL-AF4 was compared to the mean average gene expression of 1000 groups of 149 randomly sampled groups of MLL-AF4. The significance was determined by the number of groups that spreading MLL-AF4 showed higher expression than. In this instance the group was higher than all 1000 groups therefore **p = 1/1000 = 0.001**. Nascent RNA-seq data from this analysis represented 4 biological replicates in SEM cells.

[1] Taken from http://www.socscistatistics.com/tests/pearson/

2.10.3. Survival Analysis, Kaplan-Meier Curves (Figure 5.2C)

To determine the significance between survival curves, the log-rank test was used. The log-rank test statistic is calculated using the following formula (Goel et al., 2010):

$$\text{Log-rank test statistic} = \frac{(0_1 - E_1)^2}{E_1} + \frac{(0_2 - E_2)^2}{E_2}$$

where O = observed events, E= expected events, and 1/2 = group number.

3. Chapter Three – MLL Fusion Proteins Exhibit Unique Genome-wide Binding Profiles That Co-localise with Interacting Factors

3.1. Introduction

Since recruitment of MLL-FPs to gene targets is integral to the aberrant gene expression that causes leukaemia, a full understanding of this process is therapeutically important. Studies have shown an essential role for the CXXC domain in MLL-FP recruitment to unmethylated CpG Islands (uCGIs) (Allen et al., 2006; Ayton et al., 2004; Bach et al., 2008; Birke et al., 2002; Cierpicki et al., 2010; Milne et al., 2010; Okuda et al., 2014). Interactions between MLL-FPs and both the Menin/LEDGF complex and PAF elongation complex (PAFc) are also thought to be important for MLL-FP recruitment (Okuda et al., 2014; Thiel et al., 2010; Yokoyama and Cleary, 2008; Yokoyama et al., 2005) (Dawson et al., 2011; Milne et al., 2010; Muntean et al., 2010). These studies are limited in that they use ChIP to assess binding at a few select loci whilst attempting to make conclusions about MLL-FP recruitment in general. In addition, the techniques used do not definitively assess recruitment. Whilst the handful of gene targets tested in these studies are of importance in MLL-r leukaemia, there are other important targets that remain untested. Furthermore, MLL-r leukaemias have been observed in the clinic that show no overexpression of the genes regularly examined e.g. *HoxA9*. Since MLL-FP recruitment mechanisms might vary between gene targets it is important to analyse MLL-FP recruitment on a global level. Previous studies that have used ChIP-seq to assess global MLL-FP binding did not do ChIP-seq of MLL-FP interacting proteins that might be influencing recruitment of the fusion protein (Bernt et al., 2011; Guenther et al., 2008; Okuda et al., 2014). Therefore it was decided to analyse binding of MLL-FPs and possible recruiting factors, genome-wide.

Previously, CGIs have been defined using a computational algorithm (Gardiner-Garden and Frommer, 1987). This classes a CGI as a region of the genome greater than 200bp that has a GC% greater than 50% and an observed:expected CpG ratio greater than 0.6. However, using this stringency often misses CGIs and cannot reveal any information about methylation status. A Biotinylated CXXC Affinity Purification (Bio-CAP) assay has previously been used to detect uCGIs experimentally (Blackledge et al., 2012). This assay uses an isolated CXXC domain to enrich for unmethylated CpG dinucleotides throughout the genome. It has the advantage over the algorithm in that it has higher sensitivity, detecting more CGIs, specifically those that are well-conserved between species indicative of an important function (Long et al., 2013b). The role of CGIs in gene regulation and transcription is well-known, acting as a binding platform for chromatin proteins - a feature that can be used to further define their definition.

ChIP-seq can be used to detect where a protein binds throughout the genome. If two proteins bind at the same place it is possible that there is a role for one in the recruitment of the other. If protein A recruits protein B, they should correlate genome-wide. This does not prove recruitment but if they fail to correlate it falsifies the possibility that A recruits B. For an initial assessment of MLL-FP recruitment genome-wide, a description of the global binding profiles of MLL-AF4 and MLL-AF6 was done, alongside interacting factors. Specifically, ChIP-seq of Menin, LEDGF, members of PAFc (PAF1 and LEO1) and Bio-CAP-seq (uCGIs) were compared to MLL-FP binding. A high-degree of overlap between binding of the fusion protein and another protein suggests a potential role in recruitment, particularly if both the site and shape of binding are similar. Genome-wide approaches provide more than just a snapshot of a few genes. The handfuls of genes that have been tested to date are biased towards the more important genes in the leukaemia and might have unique recruitment criteria.

3.2. MLL Fusion Proteins Bind at a Limited Subset of Unmethylated CpG Islands

The DNA-binding ability of the CXXC domain has been shown to target MLL-FPs to genes marked by uCGIs (Milne et al., 2010; Okuda et al., 2014). The extent to which this is true genome-wide remains elusive but if true we would expect all MLL-FP binding events to occur at a uCGI. Although MLL-AF6 binding sites have been shown to be enriched for uCpG genome-wide (Okuda et al., 2014), the degree of overlap between MLL-AF6 and uCGIs was not assessed. The correlation between MLL-FP binding sites and global uCGIs was determined using ChIP-seq and Bio-CAP-seq, respectively, in MLL-r leukaemia cell lines. Bio-CAP was performed on SEM (MLL-AF4 cell line) and ML-2 (MLL-AF6 cell line) cells to enrich for regions of unmethylated CpG dinucleotides. Genomic DNA incubated in the presence of the CXXC domain was eluted using different concentrations of salt to optimise for the best conditions for subsequent high-throughput sequencing (Figure 3.1A). When a control incubation was performed in the absence of the CXXC domain, all material, regardless of uCpG content, was present in the flow-through (Figure 3.1B). The high affinity binding of CXXC domains (K_D = 0.64µM (Blackledge et al., 2010; Risner et al., 2013)) meant that loci with a high uCpG content, as tested using the *HoxA9* and *α-globin* genes, could only be eluted with 700mM salt when incubated in the presence of the CXXC domain (Figure 3.1C). This was sufficient to elute almost all of the material, making the 1M elution somewhat redundant. Hence when samples were sent for sequencing, only a 700mM elution was performed. Successful sequencing of the material revealed an overlap with approximately half of the predicted CGIs determined by the computational algorithm but also identified many new regions of unmethylated CpG that were potential CGIs (Figure 3.1D). Computationally-derived CGIs that did not have a corresponding peak in the Bio-CAP were presumed to be methylated.

In order to increase the specificity of CGIs determined by Bio-CAP, data from Assay for Transposase Accessible Chromatin (ATAC)-seq in SEM cells (performed in the Milne lab by E. Ballabio) was analysed to detect regions accessible by proteins involved in regulation. The technique uses the hyperactive Tn5 transposase to insert adapters into accessible chromatin which are subsequently used for high-throughput sequencing. ATAC-seq peaks were overlapped with Bio-CAP peaks to identify 18,630 uCGIs in SEM cells. ATAC-seq data was unavailable in ML-2 cells so Bio-CAP-seq alone was used.

Figure 3.1

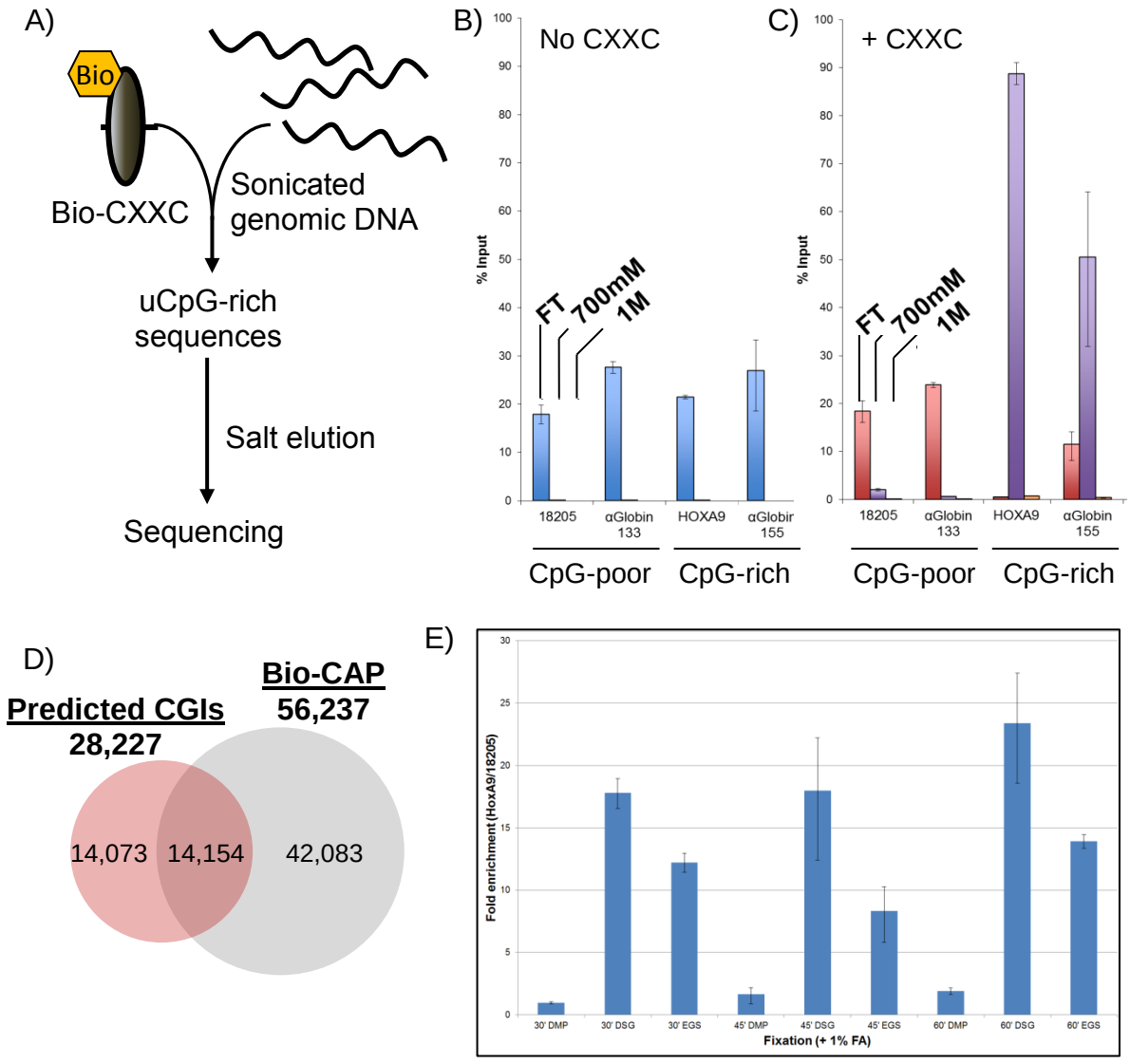


Figure 3.1 Bio-CAP and ChIP optimisation. A) Schematic of Bio-CAP technique. A biotinylated CXXC domain is incubated in the presence of sonicated genomic DNA. High salt is used to elute unmethylated CpG-rich sequences which are sent for sequencing. **B) & C)** Bio-CAP in ML-2 Cells. Enrichment of different CpG-poor and CpG-rich sequences was tested for in the absence (B) and presence (C) of the CXXC domain. FT = Flow-through, 700mM and 1M = salt concentration. **D)** Overlap between predicted CGIs and Bio-CAP peaks in SEM cells. **E)** Optimisation of ChIP fixation. ML-2 cells were fixed in the presence of DMP, DSG or EGS for 30', 45' or 60' followed by 1% FA of the same time length. ChIP was performed in the presence of a LEDGF antibody. Signal was determined as % input at a positive control locus (*HoxA9*) over % input at a negative control region (*18205*). Error bars represent standard deviations of 2 technical replicates.

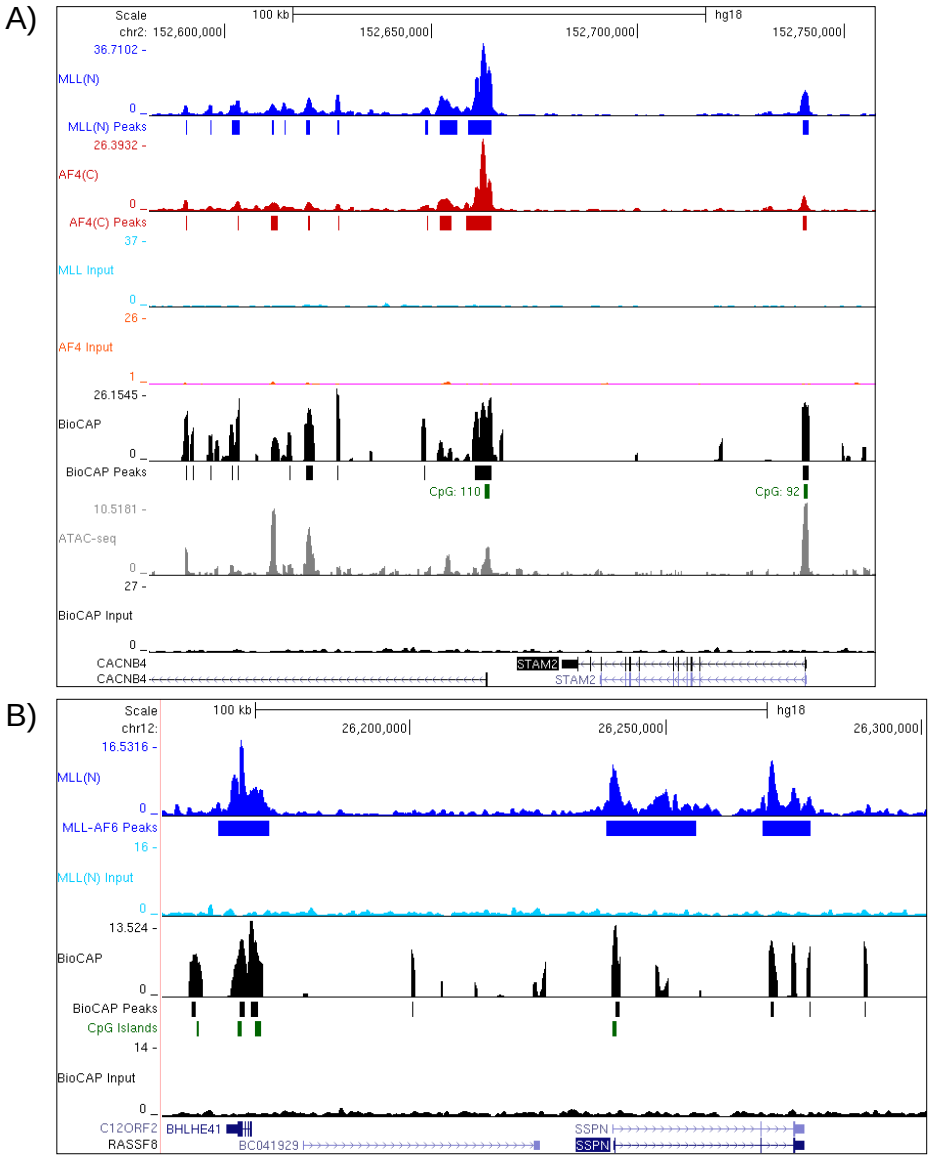
ChIP was used to enrich for binding sites of various proteins and histone marks in SEM and ML-2 cells. Optimisation of ChIP is necessary due to variations that can occur with antibody specificities for fixed material. Furthermore the degree of fixation, sonication and washing can drastically affect the signal detection. In order to ensure suitable fixation of protein to chromatin, double fixation was used. For optimisation, the chemical cross-linkers DSG, DMP or EGS were used followed by 1% FA, with individual fixation steps being carried out for 30, 45 or 60 minutes. 30 minutes of 2mM DSG, followed by 30 minutes of 1% FA was found to give a sufficiently high signal for the test antibody and was subsequently used as the standard condition for all subsequent protein ChIP (Figure 3.1E). Fixation conditions were only altered as a possible solution to the rare event of a ChIP yielding low signal. The efficiency of the ChIP was tested using qPCR with primers against a known positive (*HoxA9*) and negative (*18205*) control. The *18205* primers amplify a gene desert region void of histone marks that is approximately 18kb upstream of the *Myc* gene. The signal level, the degree of background, and the fold enrichment of positive over negative were all taken into consideration when assessing quality.

ChIP performed on material from 10^8 cells was sufficient for high throughput sequencing. Where ChIP was done by other members of the lab it is stated in the figure legends. All library preparation and sequencing was done by the Wellcome Trust. All computational analysis of ChIP-seq data was done by myself (for full details, see Chapter 2, Table 2.8).

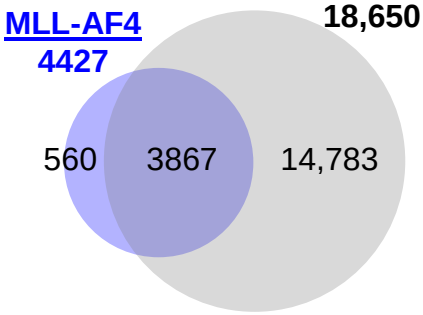
The lack of a wild-type *MLL1* allele in ML-2 cells means that MLL-AF6 could be unambiguously detected using an antibody against the N-terminus of MLL (MLL(N)). The presence of wild-type MLL1 and AF4 in SEM cells complicates MLL-AF4 detection, requiring an overlap between MLL(N) and C-terminal AF4 (AF4(C)) ChIP-seq to narrow down possible sites of MLL-AF4. This method has been applied previously (Guenther et al., 2008; Wilkinson et al., 2013). ChIP-seq data of MLL(N) in SEM and ML-2 cells, as well as AF4(C) in SEM cells, available from the Milne lab, was analysed to determine MLL-AF4 and MLL-AF6 binding sites throughout the genome (Figure 3.2A; B). To assess the importance of the CXXC domain in recruitment, these binding sites, or peaks, were overlapped with the uCGI peaks described earlier. Peaks were determined by a cut-off that assessed the enrichment of ChIP signal over the input at a given region (see Chapter 2, section 2.8). Approximately 90% of MLL-AF4 and MLL-AF6 binding events occurred at uCGIs, implicating a likely role for the CXXC domain in MLL-FP recruitment at all of these sites (Figure 3.2C; D). Since peak-calling relies on a set cut-off, it is inevitable that peaks of low enrichment will be omitted. A heat-map of the ChIP-seq data was generated to determine if the 10% of MLL-AF4 peaks that did not occur at uCGIs, as assessed by peak overlap, did actually occur at uCGIs (Figure 3.2E). This heat-map visualised the intensity of Bio-CAP and ATAC reads relative to the surrounding area under all 4427 MLL-AF4 peaks, and showed enrichment of Bio-CAP and ATAC (i.e. uCGIs) at every site. It is therefore likely that the CXXC domain is involved in MLL-FP binding at all sites. The low number of MLL-AF6 peaks meant that a heat-map was not a good visual representation of the data. Instead, individual observation of the remaining 15 MLL-AF6 peaks also revealed them to be at uCGIs (example shown in Figure 3.2F).

A composite plot of Bio-CAP and ATAC reads at all MLL-AF4 peaks showed strongest uCGI enrichment at the centre of the binding sites (Figure 3.2G). This was also true for MLL-AF6 but a more wide-spread Bio-CAP profile was observed possibly due to a larger proportion of broad binding events (discussed in Chapter 5) (Figure 3.2H). These data show that, independent of the fusion partner, MLL-FPs are recruited to sites of CXXC domain binding genome-wide.

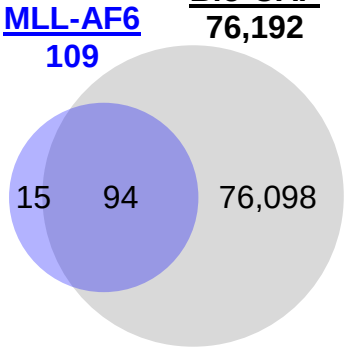
Figure 3.2



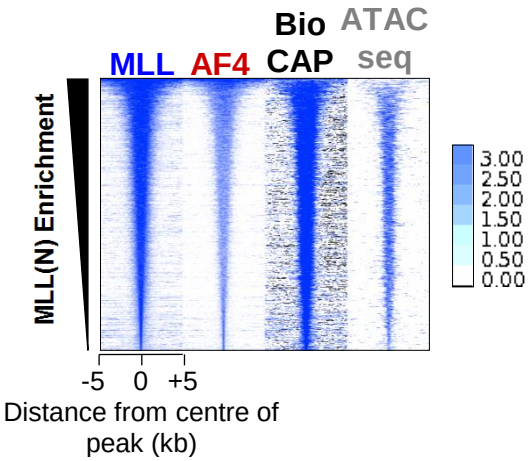
C)


MLL-AF4
4427
uCGI
18,650
560
3867
14,783

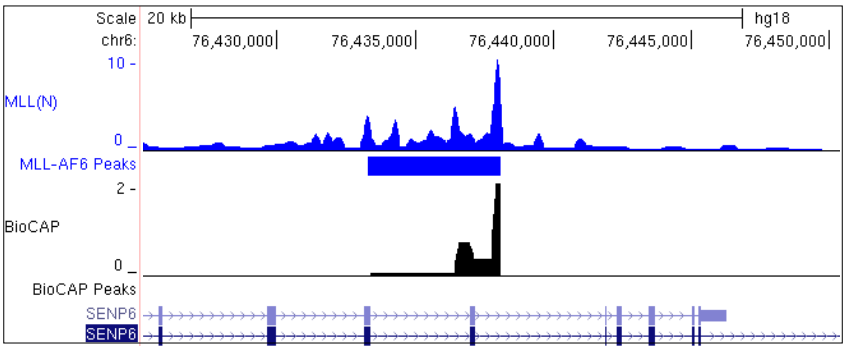
D)


MLL-AF6
109
Bio-CAP
76,192
15
94
76,098

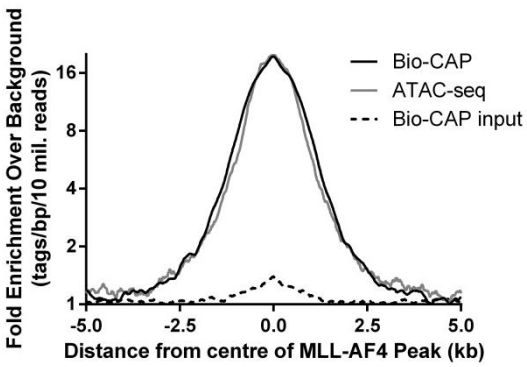
E)


Bio
CAP
ATAC
seq
MLL
AF4
MLL(N) Enrichment
3.00
2.50
2.00
1.50
1.00
0.50
0.00
-5
0
+5
Distance from centre of peak (kb)

F)


Scale
20 kb
hg18
chr6:
76,430,000
76,435,000
76,440,000
76,445,000
76,450,000
MLL(N)
MLL-AF6 Peaks
BioCAP
BioCAP Peaks
SENP6
SENP6

G)


Bio-CAP
ATAC-seq
Bio-CAP input
Fold Enrichment Over Background (tags/bp/10 mil. reads)
Distance from centre of MLL-AF4 Peak (kb)

H)

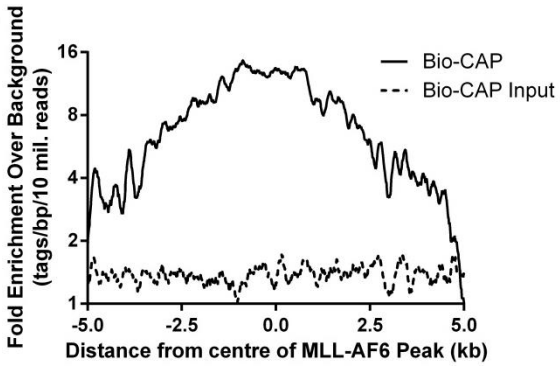
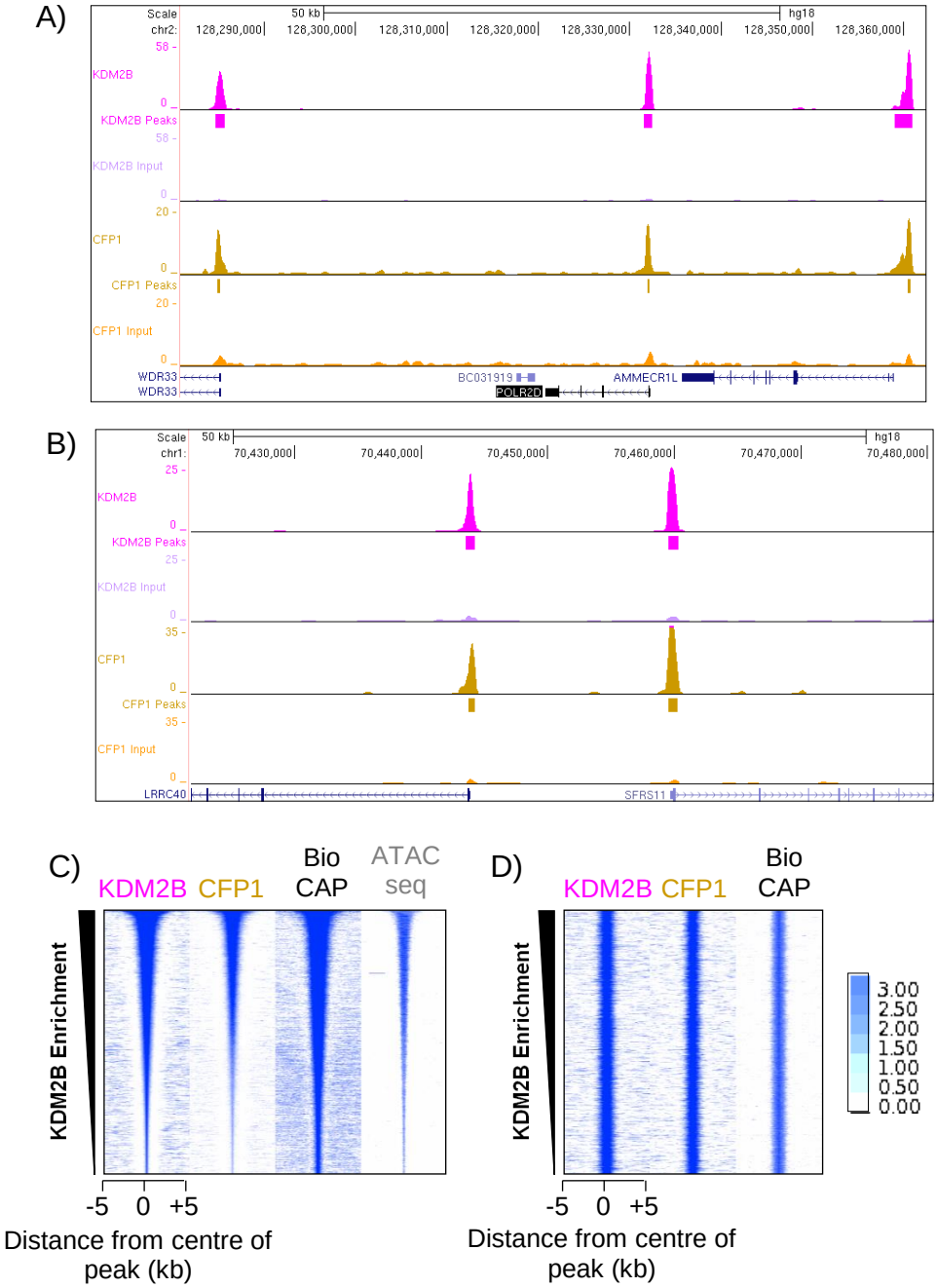

Bio-CAP
Bio-CAP Input
Fold Enrichment Over Background (tags/bp/10 mil. reads)
Distance from centre of MLL-AF6 Peak (kb)

Figure 3.2 Genome-wide recruitment of MLL-FPs occurs at uCGIs. A) Example ChIP-seq track of MLL-AF4 (MLL(N) + AF4(C), ChIP done by E. Ballabio) binding at uCGIs (Bio-CAP + ATAC-seq, ATAC done by E. Ballabio) in SEM cells. **B)** Example ChIP-seq track of MLL-AF6 (MLL(N), ChIP done by E. Ballabio) binding at uCGIs in ML-2 cells. Scale bars represent tags per 10 million. **C)** Heat-map showing uCGI enrichment at all 4427 MLL-AF4 peaks in SEM cells. Scale bar represents tags per base-pair per 10 million. **D)** Degree of overlap between uCGIs and MLL-AF4 in SEM cells. **E)** Degree of overlap between Bio-CAP and MLL-AF6 in ML-2 cells. **F)** Example ChIP-seq track of an MLL-AF6 peak at a Bio-CAP site that fell short of the cut-off. **G)** Composite plot showing the distribution of Bio-CAP and ATAC-seq reads relative to the centre of all MLL-AF4 peaks in SEM cells. **H)** Composite plot showing the distribution of Bio-CAP reads relative to the centre of all MLL-AF6 peaks in ML-2 cells.

Surprisingly, both MLL-AF4 and MLL-AF6 only bound at a small proportion of the total number of uCGIs (20.7% and 0.1%, respectively) (Figure 3.2C; D), showing that a uCGI is not sufficient for MLL-FP recruitment. To determine if this is true for other CXXC domain proteins, ChIP-seq of KDM2B and CFP1 was done in SEM and ML-2 cells (Figure 3.3A; B). Importantly both of these proteins contain CXXC domains that are highly related to that of MLL1 and bind unmethylated CpG dinucleotides in the same manner (Long et al., 2013a) (Figure 1.3B). Interestingly, when the CXXC domain of KDM2A, a member of the same family as KDM2B, is swapped in place of the MLL-FP CXXC domain, the fusion protein maintains its transformation ability (Okuda et al., 2014). However, the same is not possible when the CXXC domain of CFP1 is used despite the fusion protein still being able to bind at gene targets (Risner et al., 2013). Similar to MLL-FP recruitment, all KDM2B and CFP1 binding sites occurred at uCGIs in both cell lines (Figure 3.3C; D). However, unlike the fusion proteins, these two proteins occupied a much larger proportion of the total number of uCGIs, with KDM2B found at approximately 90% and CFP1 at approximately 50% (Figure 3.3E; F). This argues that in different contexts the CXXC domain has varied propensities for stable binding at uCGIs or that the proteins are expressed at different levels. The idea of sampling is particularly important here since it is possible that all proteins containing a CXXC

domain are capable of diffusion to all uCGIs. What then determines how they respond to the environment at the uCGI that is potentially facilitating or preventing the stable binding? This could be due to different chromatin-binding domains in the proteins or due to differences in interactions with other proteins. For example MLL-FP binding might be limited to uCGIs that are marked by a particular chromatin signature or are preoccupied by certain proteins. Following on from this, whilst a particular active state may be required, the absence of repressive factors might also be necessary.

Figure 3.3



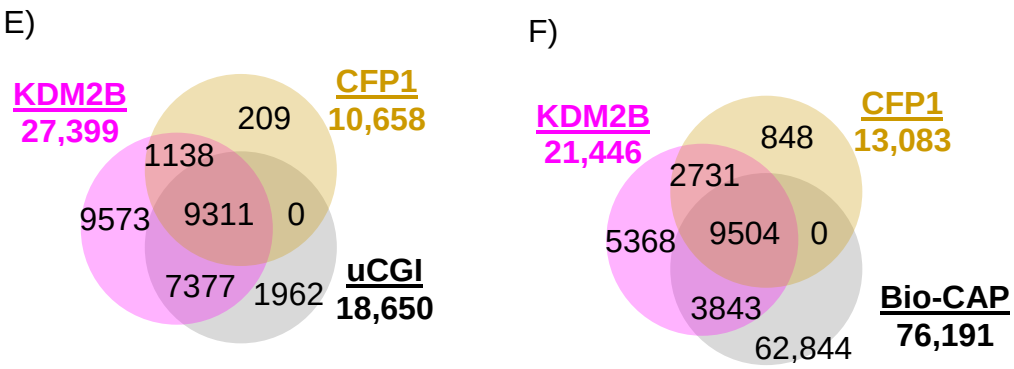


Figure 3.3 KDM2B and CFP1 are recruited to the majority of uCGIs. A) Example ChIP-seq track of KDM2B and CFP1 in SEM and **B)** ML-2 cells (CFP1 ChIP done by T. Milne). Scale bars represent tags per 10 million. **C)** Heat-map showing CFP1 and uCGI enrichment at all 27,399 KDM2B in SEM cells. **D)** Heat-map showing CFP1 and Bio-CAP enrichment at all 21,446 KDM2B in ML-2 cells. Scale bar represent tags per base-pair per 10 million. **E)** Overlap between KDM2B and CFP1 with uCGI peaks in SEM cells and **F)** Bio-CAP peaks in ML-2 cells.

3.3. MLL Fusion Proteins Bind at Active Epigenetic Landscapes

In order to investigate whether other proteins or chromatin marks are required for MLL-FP recruitment, ChIP-sequencing was performed on various histone marks and proteins that interact with MLL-FPs. For ChIP-sequencing of histone marks, $4x10^7$ cells fixed with 1% FA for 10 minutes was found to be sufficient (Figure 3.4A; B). MLL-FP-bound uCGIs were largely marked by the active-associated histone marks H3K4me3 and H3K27ac (Figure 3.4C). This is in accordance with the fusion proteins overlapping with CFP1 and MLL1, two members of H3K4 methyltransferase complexes, and an association that would suggest the pre-requisite of an active chromatin state for binding. Notably, MLL-FP binding sites lack the repressive H3K27me3 mark (Figure 3.4C). This is in agreement with the proposed sampling recruitment mechanism introduced in Chapter 1.

A requirement for active marks and an absence of repressive marks is not mutually exclusive for a protein's binding criteria since whilst KDM2B binds at "active" uCGIs, its role in recruiting polycomb repressive complexes means that it is also found at repressed uCGIs (Figure 3.7A; B). This may help to explain why KDM2B is found at a larger number of uCGIs than MLL-FPs since KDM2B binds at uCGIs that the fusion protein is excluded from due to its apparent antagonism with repressive chromatin (Figure 3.2C; D; Figure 3.3E).

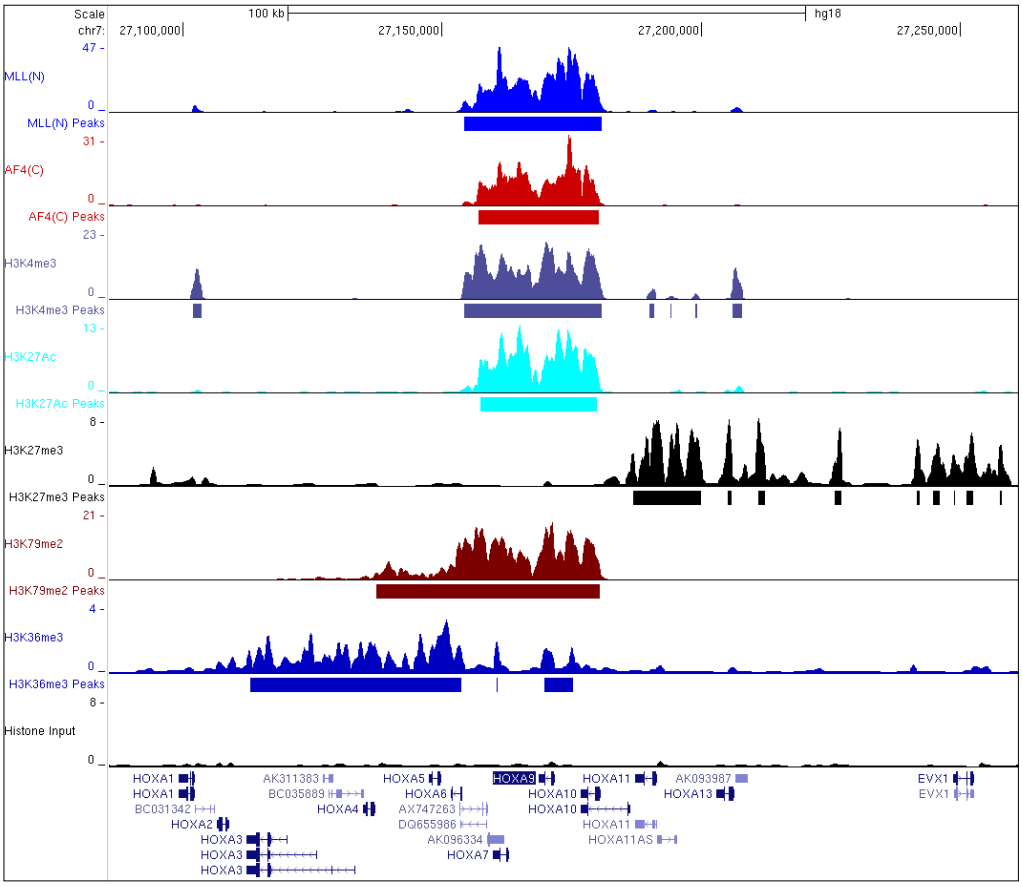
H3K79me2 is a mark associated with actively transcribed genes, located towards the 5' end of the gene body. The correlation of MLL-FPs with H3K79me2 is well known due to the association between the fusion proteins and DOT1L. H3K79me2 was found to perfectly correlate with MLL-FPs in both SEM and ML-2 cells both in terms of number and breadth of binding events (Figure 3.4A-D). Whether this mark is present before, and required for, MLL-FP recruitment or whether it is a direct consequence of MLL-FP recruitment remains unknown. However, previous studies have shown an increase in H3K79me2 at MLL-FP target genes once the fusion protein has been transduced into bone marrow (Bernt et al., 2011). Since DOT1L is found at actively transcribed genes in the normal context it is likely that a basal level of H3K79me2 is present but that this is enhanced by MLL-FP-mediated DOT1L recruitment. When all MLL-AF4 target genes in SEM cells were compared to non-target genes marked by H3K79me2, the former group showed an increase in H3K79me2 signal in the gene body (Figure 3.4F).

H3K36me3 is also associated with the gene body of actively transcribed genes but is localised towards the 3' of the gene. A strong anti-correlation was seen between H3K79me2 and H3K36me3 in SEM and ML-2 cells at MLL-FP binding sites (Figure 3.4D). This is not abnormal, since an analysis of all gene targets marked by H3K36me3 in ML-2 cells showed the well-characterised exclusivity between H3K79me2, near the promoter, and H3K36me3, away from the promoter (Figure 3.4E). However, at certain

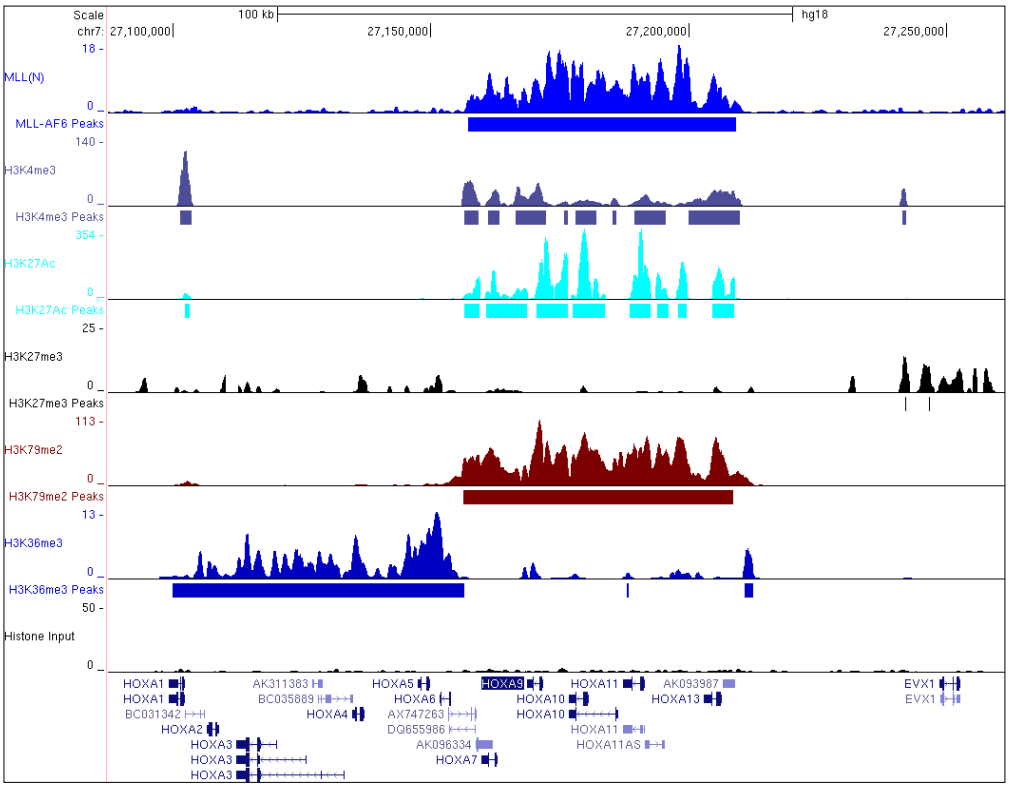
gene loci the association of the fusion proteins with abnormally broad domains of H3K79me2 seems to restrict the appearance of H3K36me3 where it might otherwise be found. The presence of H3K79me2 but absence of H3K36me3 represents a unique chromatin profile not seen at any other genes (Figure 3.4G, top). Instead, at all other genes H3K36me3 increases as H3K79me2 decreases (Figure 3.4E; Figure 3.4G, bottom). Nascent RNA-seq data from the Milne lab was analysed, and verified that these genes with an abnormal K79/K36 relationship were still actively transcribed potentially indicating a novel transcription mechanism as a result of the fusion protein (Figure 3.4G, top). Together, genome-wide assessment of histone marks shows that recruitment of MLL-FPs occurs exclusively at sites of active chromatin and is limited by not being able to bind at sites of repression.

Figure 3.4

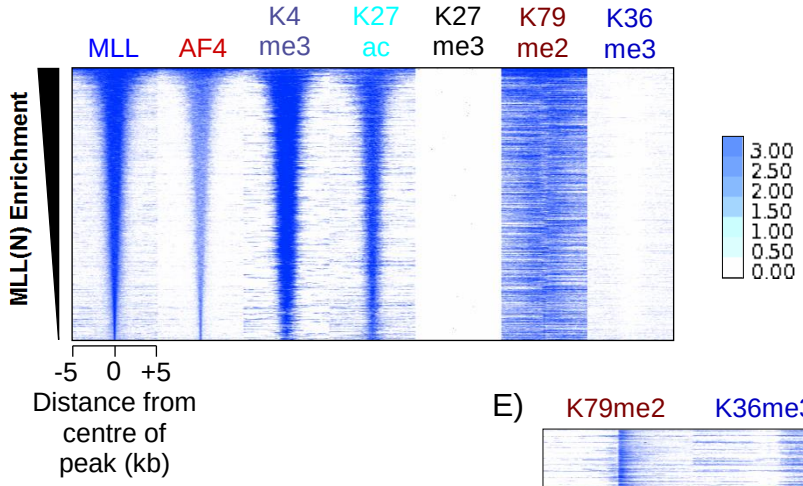
A)



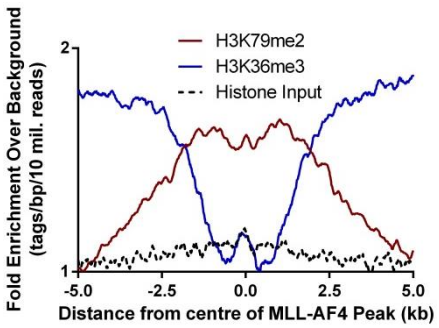
B)


Scale
chr7:
100 kb
hg18
27,100,000
27,150,000
27,200,000
27,250,000
MLL(N)
MLL-AF6 Peaks
H3K4me3
H3K4me3 Peaks
H3K27Ac
H3K27Ac Peaks
H3K27me3
H3K27me3 Peaks
H3K79me2
H3K79me2 Peaks
H3K36me3
H3K36me3 Peaks
Histone Input
HOXA1
HOXA1
BC031342
HOXA2
HOXA3
HOXA3
HOXA3
AK311383
BC035889
HOXA4
HOXA5
HOXA6
AX747263
DQ655986
AK096334
HOXA7
HOXA9
HOXA10
HOXA10
HOXA11
HOXA11AS
AK093987
HOXA13
EVX1
EVX1

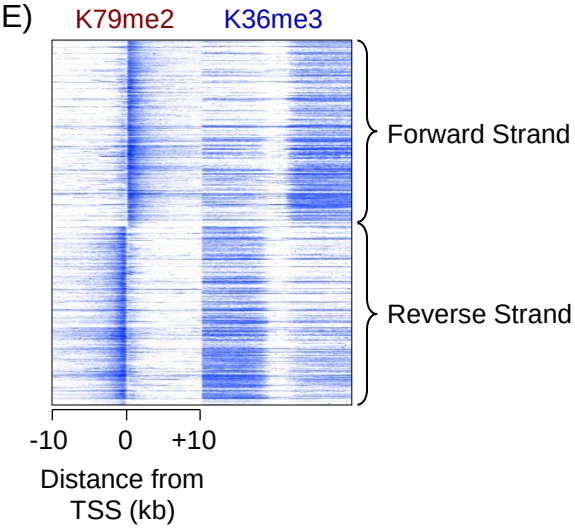
C)


MLL
AF4
K4 me3
K27 ac
K27 me3
K79 me2
K36 me3
MLL(N) Enrichment
3.00
2.50
2.00
1.50
1.00
0.50
0.00
-5
0
+5
Distance from centre of peak (kb)

D)


H3K79me2
H3K36me3
Histone Input
Fold Enrichment Over Background (tags/bp/10 mil. reads)
2
1
-5.0
-2.5
0.0
2.5
5.0
Distance from centre of MLL-AF4 Peak (kb)

E)


K79me2
K36me3
Forward Strand
Reverse Strand
-10
0
+10
Distance from TSS (kb)

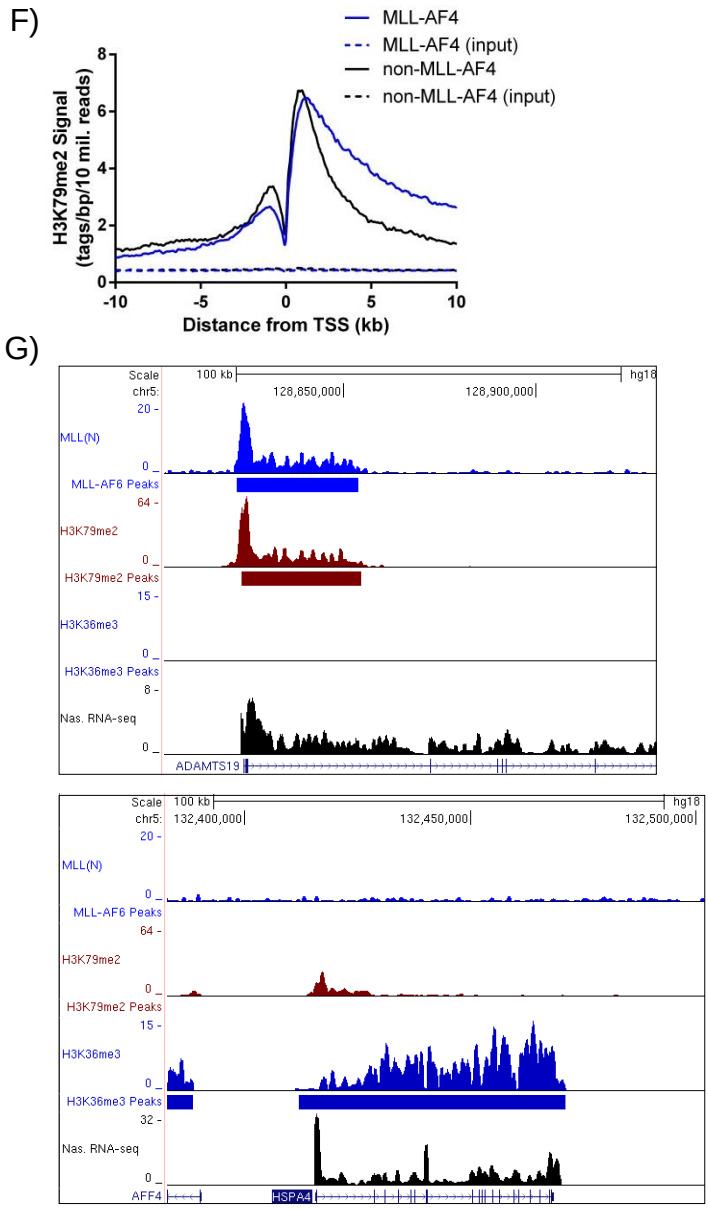


Figure 3.4 MLL-FPs are recruited to active loci. A) Example ChIP-seq tracks of histone marks at the *HOXA* cluster in SEM (H3K27me3 ChIP done by M. Tapia) and **B)** ML-2 cells. **C)** Heat-map showing enrichment of active- and repressive- associated histone marks at all 4427 MLL-AF4 peaks in SEM cells. Scale bar represents tags per bp per 10 million. **D)** Composite binding plot of H3K79me2 and H3K36me3 reads at all MLL-AF4 peaks in SEM cells. **E)** Heat-map showing H3K79me2 and H3K36me3 enrichment around the TSS of genes marked by H3K36me3 in ML-2 cells. Scale as in (C) **F)** Composite binding plot of H3K79me2 reads at MLL-AF4 gene targets (blue) and non-MLL-AF4 genes targets (black) in SEM cells. **G)** Example ChIP-seq track of unique H3K79me2 and H3K36me3 relationship at MLL-AF6 target gene (top) compared to a non-MLL-AF6 target (bottom), in ML-2 cells. Scale bars represent tags per 10 million.

3.4. Binding of the Menin/LEDGF Complex Correlates with MLL-FPs

The above data suggests that MLL-FP recruitment is restricted to actively marked uCGIs but the absence of MLL-FPs at thousands of these sites indicates that there are other limiting factors imposed on its binding. MLL-FPs were shown to interact with other proteins that can bind to chromatin and that this binding is functionally important, for example with the Menin/LEDGF complex (Yokoyama and Cleary, 2008; Yokoyama et al., 2005). The interaction with these proteins could therefore also be a limiting factor on MLL-FP recruitment. ChIP-seq was performed on the Menin/LEDGF complex to see if it correlates with MLL-FP binding sites. Menin ChIP was of sufficient quality for sequencing in both SEM and ML-2 cells; however LEDGF ChIP only produced good enrichment in ML-2 cells (Figure 3.5A; B). The reasons for this are unclear but attempts at ChIP with different LEDGF antibodies failed to improve the issue. As such, in SEM cells Menin alone was used to represent the Menin/LEDGF complex.

A currently published model proposes that the Menin/LEDGF complex is necessary for MLL-FP recruitment (Okuda et al., 2014; Yokoyama and Cleary, 2008). However, as described in detail in Chapter 1, there is not enough evidence to make this claim, and it has never been tested genome-wide. Peak-calling revealed numerous MLL-AF4 binding sites (87.2%) that were absent of Menin in SEM cells (Figure 3.5C). This either implies that there is only a subset of MLL-AF4 binding sites where the Menin/LEDGF complex is required, or this is due to the limitations of peak-calling low ChIP signal. A heat-map was used to determine if two distinct categories existed (Menin overlap vs non-Menin overlap) or if a continuum of Menin enrichment was seen across all MLL-AF4 target sites (Figure 3.5E). A certain level of Menin signal was seen at all MLL-AF4 binding sites. Furthermore, there was a strong correlation between the number of MLL(N) and Menin reads under MLL-AF4 peaks (r^2 = 0.96) (Figure 3.5F). This same correlation was not

observed between Menin and CFP1 (r^2 = 0.27) (data not shown). These data show that the Menin/LEDGF complex is a good candidate for directing genome-wide recruitment of MLL-FPs but since this is just a correlation further work is required for such a conclusion. The strong correlation could also be due to MLL-FPs recruiting Menin. Since the heat-map revealed a small number of Menin binding sites absent of MLL-AF4 it is possible to conclude that Menin is not sufficient for MLL-FP recruitment (Figure 3.5E). Peak-calling in ML-2 cells revealed a 92% overlap between MLL-AF6 and the Menin/LEDGF complex (Figure 3.5D). This stronger degree of overlap compared to in SEM cells likely reflects the better ChIP signal for these proteins in this cell line. Interestingly, the total number of Menin and LEDGF binding sites are both very similar to that of MLL-AF6, suggesting few roles independent from each other.

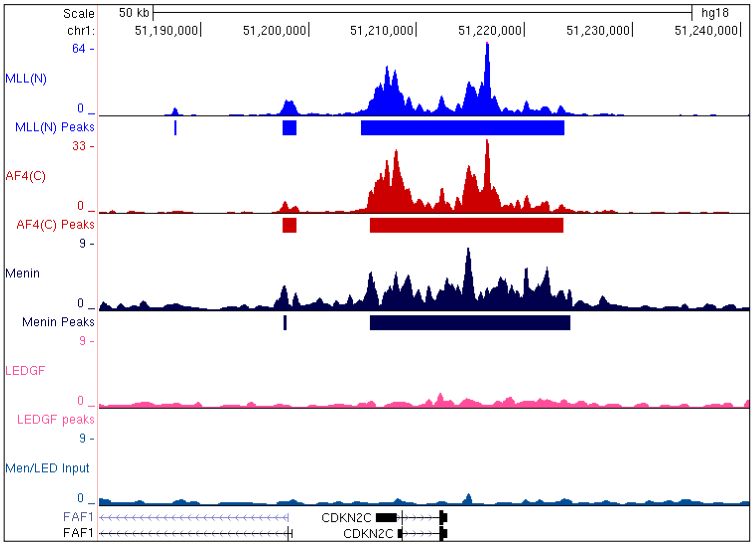
The 568 MLL-AF4 peaks that were determined as overlapping with Menin by peak-calling were used to represent MLL-AF4 binding sites where Menin enrichment was the highest (Figure 3.5C). Since a strong correlation was already shown in terms of number of reads (Figure 3.5F), the height of MLL enrichment was observed at the Menin vs non-Menin peaks. Where the Menin/LEDGF complex overlapped with MLL-FPs there was higher enrichment of the fusion protein compared to where there was no overlap with the complex (Figure 3.5G). These may represent the most important set of genes, possibly essential for maintaining the MLL leukaemia. Returning to the recruitment model, whilst an active uCGI might be sufficient for sampling to facilitate binding of the MLL-FP, the fusion protein might instantaneously diffuse away from the site in the absence of the Menin/LEDGF complex, which acts to stabilise the interaction at the uCGI (Figure 3.8). In other words, in the absence of Menin/LEDGF the binding on-off rate is shifted to the off state to such a degree that binding is never measurable by ChIP-seq i.e. not stable. Comparing to the bacterial recruitment system, here the uCGI acts as the DNA binding template and Menin/LEDGF would be playing a similar role to the sequence-specific TF binding site.

Since ChIP-seq has been performed on a heterogeneous population, genes that are bound by Menin/LEDGF in a larger proportion of the population are also more likely to have MLL-FPs bound (i.e. the fusion protein has a slower off-rate), as observed (Figure 3.5G). The majority of Menin binding sites overlapped with MLL-FP binding suggesting that this correlation is important. As well as this, the breadth of the Menin/LEDGF complex peaks closely matched that of the fusion protein (Figure 3.5H; I). A role for the Menin/LEDGF complex in MLL-FP recruitment was further supported by findings that there was less overlap between the Menin/LEDGF complex and KDM2B or CFP1 despite a significantly larger number of KDM2B and CFP1 binding sites compared to the MLL-FPs (Figure 3.5C-D; 3.7C-D). These data together strongly implicate the Menin/LEDGF complex as important for MLL-FP recruitment genome-wide.

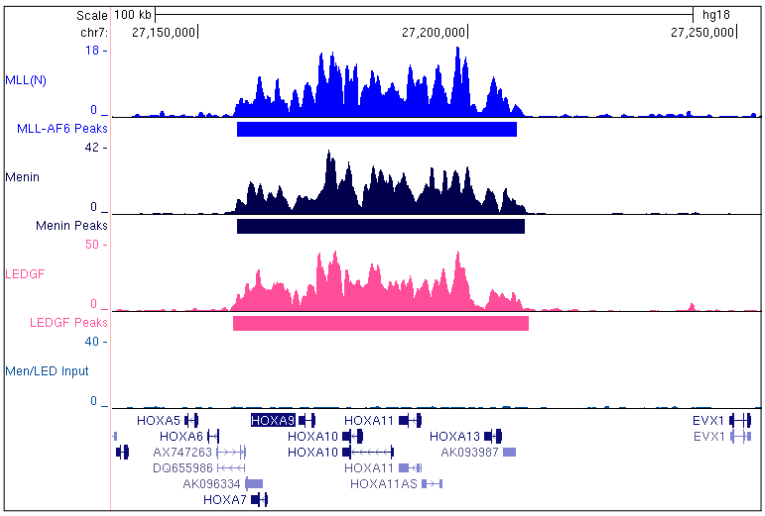
The Menin/LEDGF complex has been shown to mediate MLL-FP binding via the PWWP domain of LEDGF (Yokoyama and Cleary, 2008). Mutations of amino acids in the PWWP domain that are important for chromatin binding led to reduced binding of MLL-FPs at certain loci. Currently it is thought that this mechanism involves an interaction of the PWWP domain with H3K36me2/3. Conversely, ChIP-sequencing of H3K36me3 revealed a lack of this mark at fusion protein binding sites, in both SEM and ML-2 cells (Figure 3.4A-D; F). This implies that H3K36me3 is not required for recruiting MLL-FPs but that the Menin/LEDGF complex is associating with chromatin in another way. ChIP of H3K36me2 was attempted but a high enough quality could not be reached for successful sequencing. Since both Menin and LEDGF are able to bind DNA directly, this represents another possibility of how they are interacting with gene loci.

Figure 3.5

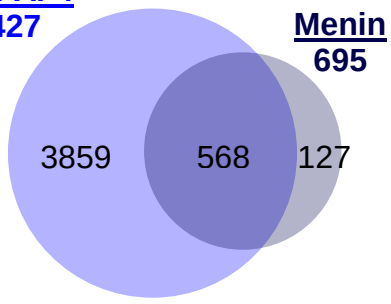
A)



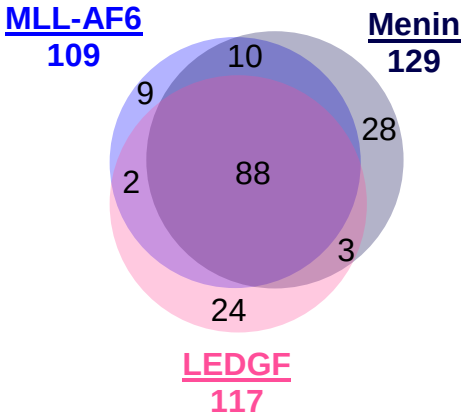
B)



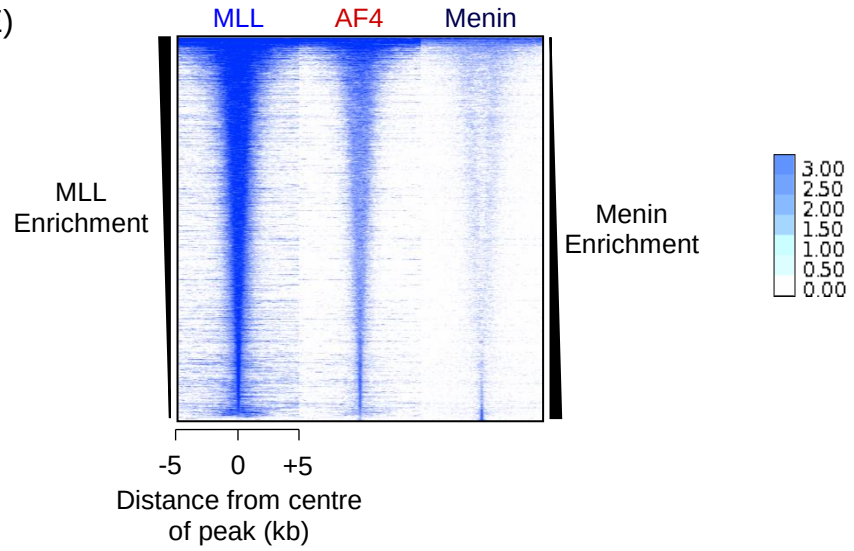
C)



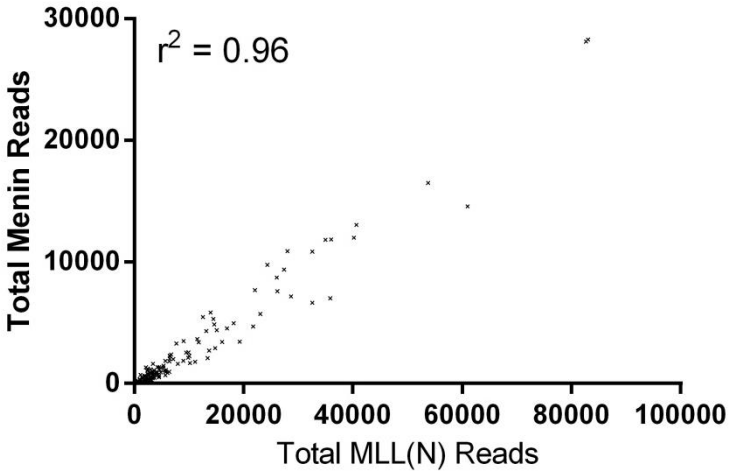
D)



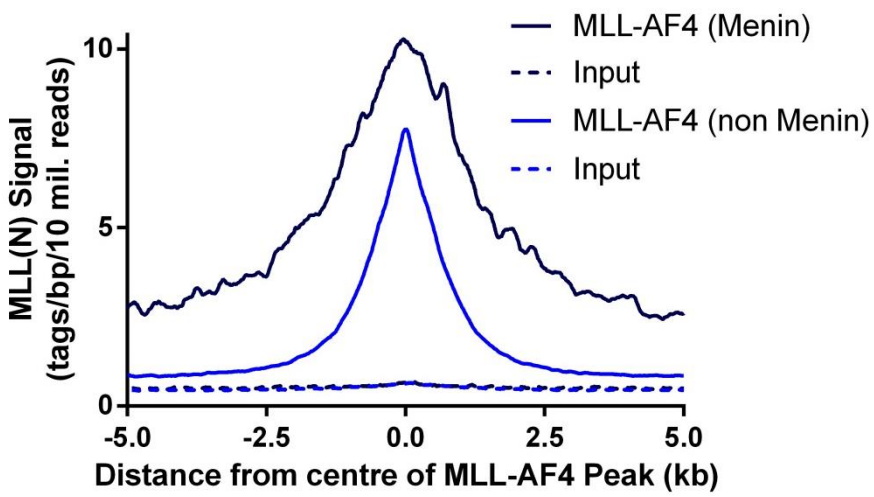
E)


MLL
AF4
Menin
MLL Enrichment
Menin Enrichment
3.00
2.50
2.00
1.50
1.00
0.50
0.00
-5
0
+5
Distance from centre of peak (kb)

F)


r^2 = 0.96
Total Menin Reads
30000
20000
10000
0
0
20000
40000
60000
80000
100000
Total MLL(N) Reads

G)


MLL-AF4 (Menin)
Input
MLL-AF4 (non Menin)
Input
MLL(N) Signal (tags/bp/10 mil. reads)
10
5
0
-5.0
-2.5
0.0
2.5
5.0
Distance from centre of MLL-AF4 Peak (kb)

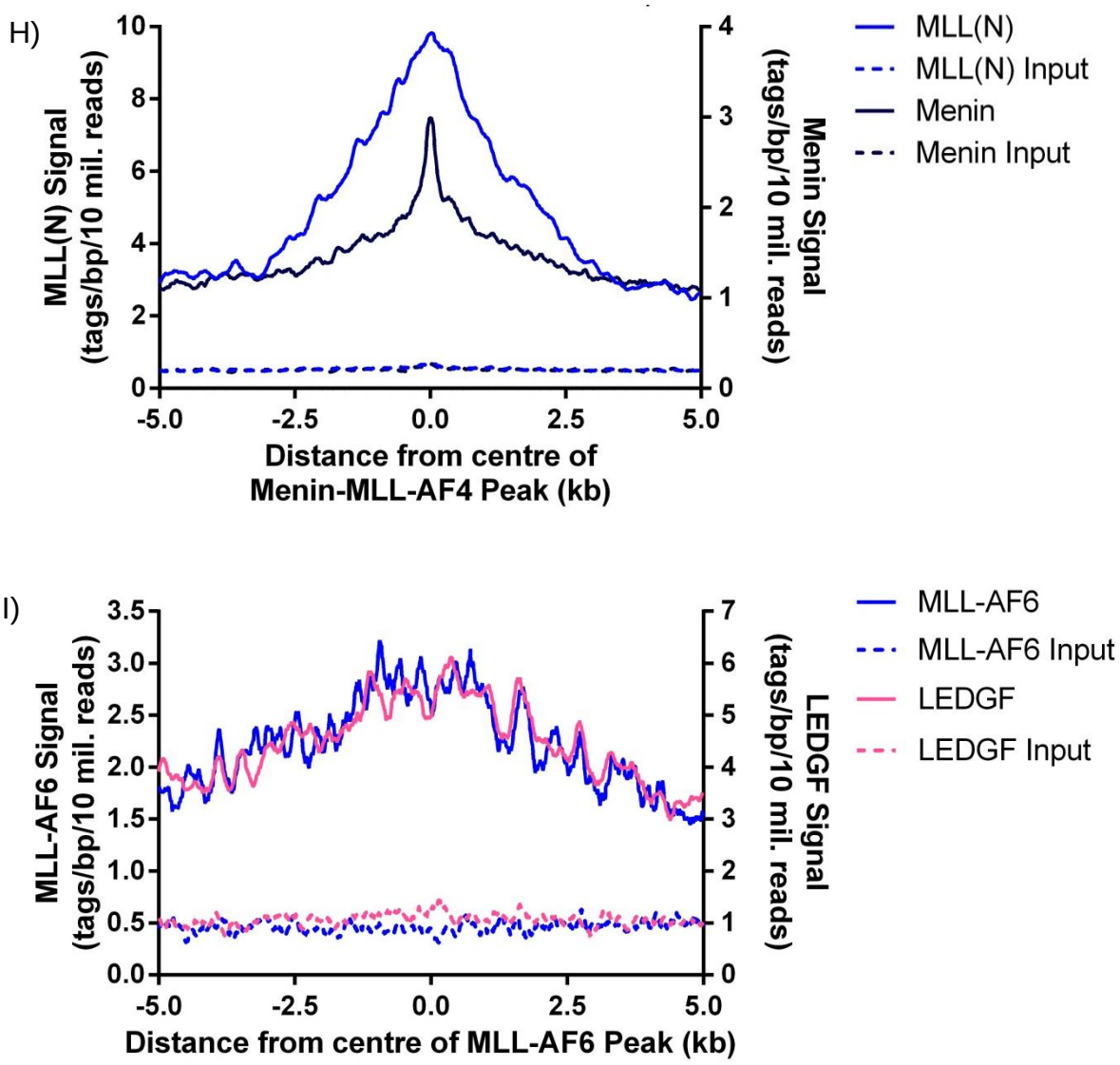


Figure 3.5 The Menin/LEDGF complex co-localises with MLL-FPs. A) Example ChIP-seq track of Menin/LEDGF complex at MLL fusion protein target site in SEM and **B)** ML-2 cells. Scale bars represent tags per 10 million. **C)** Overlap between MLL-AF4 and Menin binding sites in SEM cells. **D)** Overlap between MLL-AF6 and Menin/LEDGF complex in ML-2 cells. **E)** Heat-map showing Menin enrichment at all 4427 MLL-AF4 sites in SEM cells. Scale bar represents tags per base-pair per 10 million. **F)** Plot of MLL(N) and Menin reads at individual MLL-AF4 peaks **G)** Composite binding plot of MLL(N) reads at MLL-AF4 peaks that overlap with Menin (purple) and MLL-AF4 peaks that don't overlap with Menin (blue) in SEM cells. **H)** Composite binding plot of MLL (N) and Menin reads at MLL-AF4 peaks that overlap with Menin in SEM cells. **I)** Composite binding plot of MLL(N) and LEDGF reads at MLL-AF6 peaks in ML-

3.5. The PAF Elongation Complex Shows No Specific Correlation with MLL-FPs

It was previously shown that PAFc is involved in MLL-FP binding at certain gene targets (Milne et al., 2010; Muntean et al., 2010). The mechanism is not fully understood but may be mediated by the association of PAFc with the transcription machinery, particularly RNAPII, or BRD4 (Dawson et al., 2011; Kim et al., 2010). If the PAFc-MLL interaction is necessary for MLL-FP recruitment, it is possible that MLL-FP recruitment requires the presence of transcriptional elongation. ChIP-seq data from the Milne lab of two members of PAFc, PAF1 and LEO1, in both SEM and ML-2 cells was analysed (Figure 3.6A-B). Both proteins were found in high abundance due to their roles in transcription (Figure 3.6C-D). As such there was an approximately 50% overlap with between member of PAFc and MLL-FPs. However, similar degrees of overlap with PAFc were seen for both KDM2B and CFP1, suggesting that the correlation is not specific to MLL-FPs and is more likely a result of the large number of PAFc binding sites (Figure 3.7E). Furthermore, there were many sites of MLL-FP binding where PAFc was not present. These either represented intergenic regions or genes with low level transcription. Overall this data shows that PAFc is not necessary for MLL-FP recruitment genome-wide and that associations seen here are not necessarily specific. Certainly for wild-type MLL1, a previous study has shown that its interaction with PAF1 only confers overlap at a few loci, as assessed by immunofluorescence (Milne et al., 2010). Whereas the shape of the Menin/LEDGF complex binding events matched those of MLL-FPs (Figure 3.5A-B; F-G), the overlap between MLL-FP and PAFc binding was seen at the promoters but not in the gene body where PAFc spread independently of the fusion protein (Figure 3.6A-B).

Figure 3.6

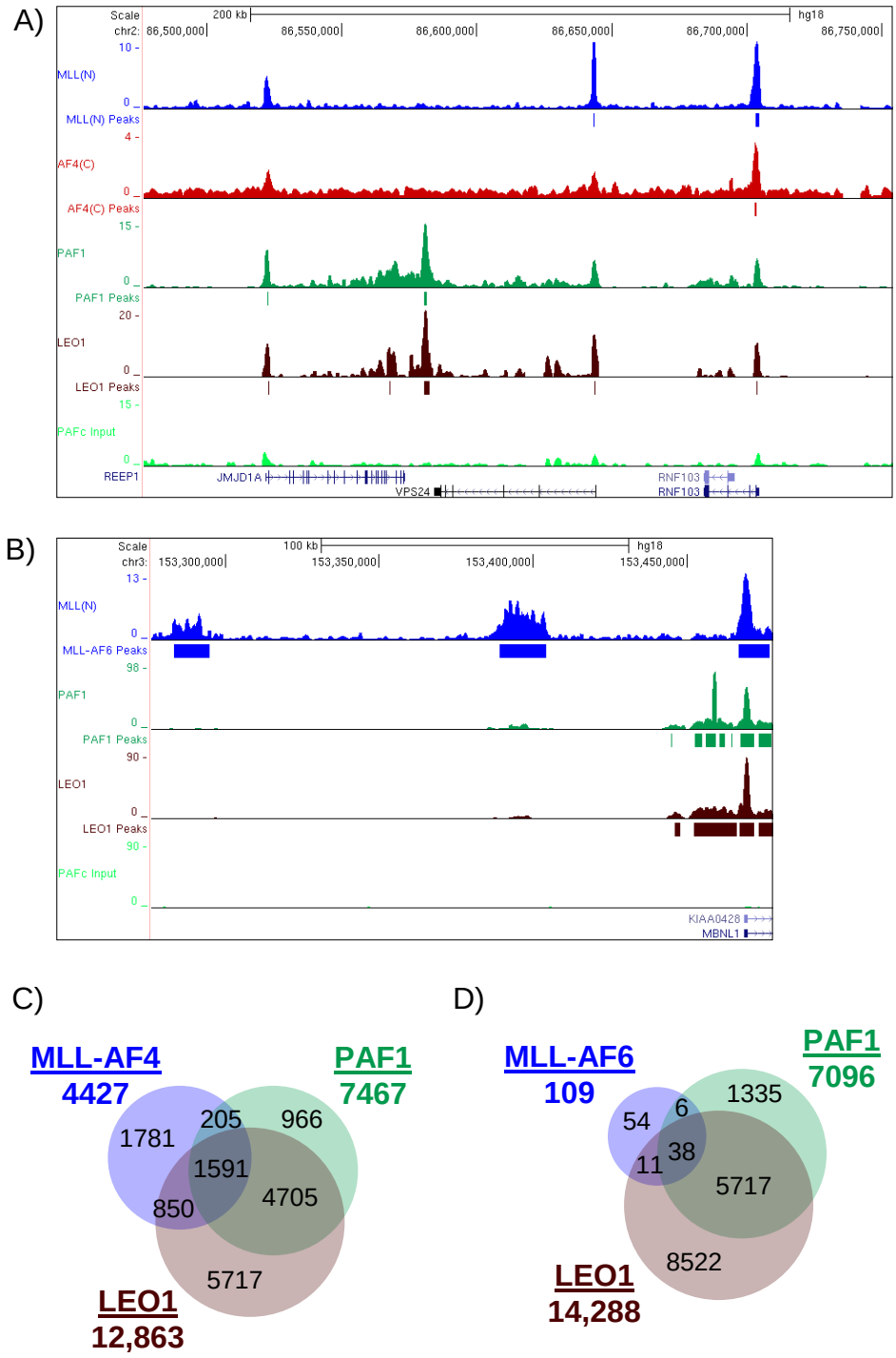
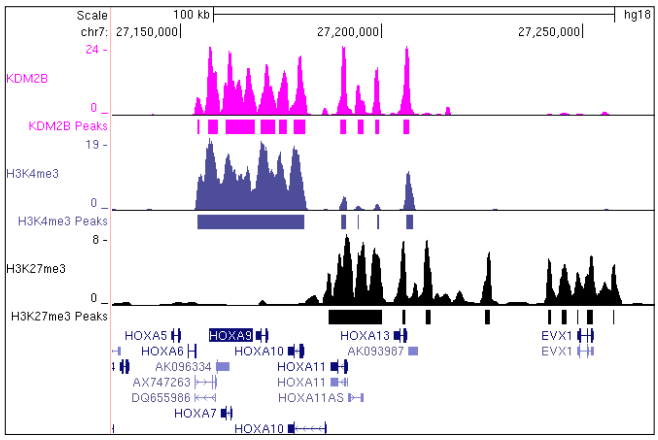


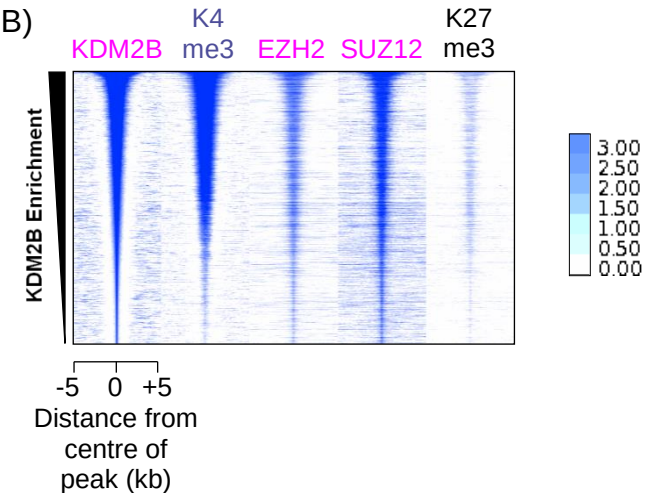
Figure 3.6 MLL-FPs show non-specific overlap with PAFc. A) Example ChIP-seq tracks of PAFc members PAF1 and LEO1 at MLL fusion protein binding sites in SEM and **B)** ML-2 cells (PAF1 and LEO1 ChIP done by T. Milne). Scale bars represent tags per 10 million. **C)** Overlap between MLL-AF4, PAF1 and LEO1 binding sites in SEM cells and **D)** ML-2 cells.

Figure 3.7

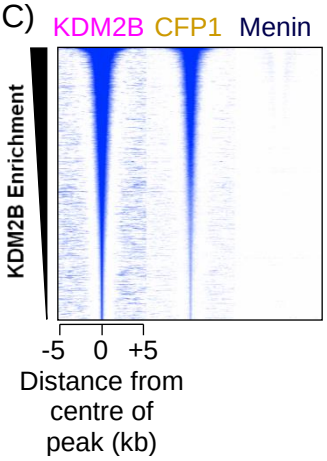
A)



B)



C)



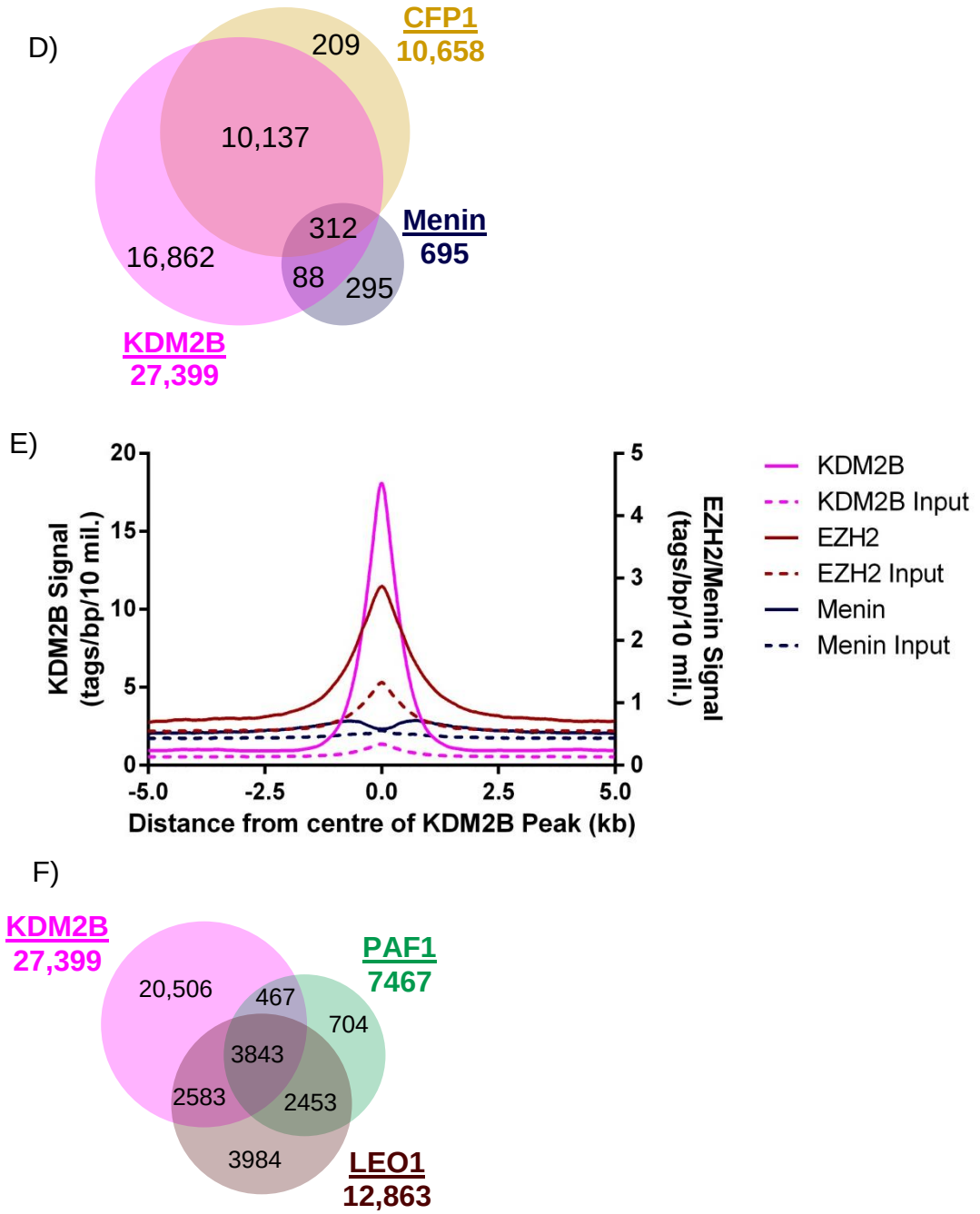


Figure 3.7 KDM2B and CFP1 overlap analyses in SEM cells. A) Example ChIP-seq track of KDM2B with H3K4me3 and H3K27me3. Scale bar represents tags per 10 million. **B)** Heat-map showing enrichment of H3K4me3, members of PRC2 and its associated H3K27me3 at all 27,399 KDM2B sites. (EZH2 and SUZ12 ChIP done by T. Milne) **C)** Heat-map showing CFP1 and Menin enrichment at all KDM2B sites. Scale bars represent tags per base-pair per 10 million. **D)** Overlap between KDM2B, CFP1 and Menin binding sites. **E)** Composite binding plot of KDM2B, Menin and EZH2 reads at KDM2B binding sites. **F)** Overlap between KDM2B, PAF1 and LEO1 binding sites.

3.6. Discussion

Genome-wide correlations are a powerful tool to assess the association of various proteins and chromatin features on a large scale. Using ChIP-seq, this chapter has shown that MLL-FPs are limited in their ability to bind uCGIs compared to other CXXC domain proteins. This shows that despite appearing to be necessary, uCGIs are not sufficient for recruitment. This chapter has, for the first time, compared the binding profiles of MLL-FPs with other proteins believed to be necessary for their recruitment. Strong overlap of uCGIs and the Menin/LEDGF complex with MLL-FPs reinforces, but does not confirm, previous studies proposing a role for these in recruitment. Importantly, data here extrapolates these studies to a global scale. Results with PAFc suggest that data from previous studies implicating its role in recruitment are either an indirect effect of the assay used or that it is only involved in recruitment at certain loci. Whilst ChIP-seq is useful for global analysis it is limited in its ability to provide functional conclusion since correlation does not mean causation. Furthermore, as highlighted above, it is limited by the ability of the antibody to enrich for fixed material and by the quality of the sequencing. Therefore what might be considered as an area of no protein binding may simply be a result of lack of signal due to the limitations of the technique.

Recruitment of MLL-FPs appears to require an actively-marked uCGI as a result of CXXC-mediated binding. Discrimination over which uCGIs are bound by the fusion protein may in part be directed by the interaction with, and hence presence of, other protein complexes which in turn may also be dependent on the epigenetic or transcriptional context at a given gene. Alternatively, this model would mean that MLL-FP recruitment is limited by expression of interacting proteins, particularly relevant for the Menin/LEDGF complex which shows relatively few binding sites. The expression level of the fusion protein itself may also be the reason for its low binding frequency. Since a high expression level of the fusion protein could be toxic to the cells, the limited

amount of MLL-FP may be directed to those genes where it has the most beneficial effect for maintenance of the leukaemia.

MLL-FPs might be capable of binding to “active” uCGIs in the absence of other factors but the strong correlation with Menin/LEDGF implicates a role for this complex. Studies thus far have assumed that the Menin/LEDGF and PAF complexes are recruiting the fusion protein. However, whether the complexes allows the fusion protein to bind at a site, i.e. play a direct role in sampling, or whether they simply stabilise the fusion protein after binding by another means remains unclear. Data in the next chapter assesses the ability of both complexes to recruit the fusion protein. Furthermore, it examines the recruitment of the isolated CXXC domain to gene targets in the absence of an interaction with the Menin/LEDGF or PAF complexes.

Figure 3.8

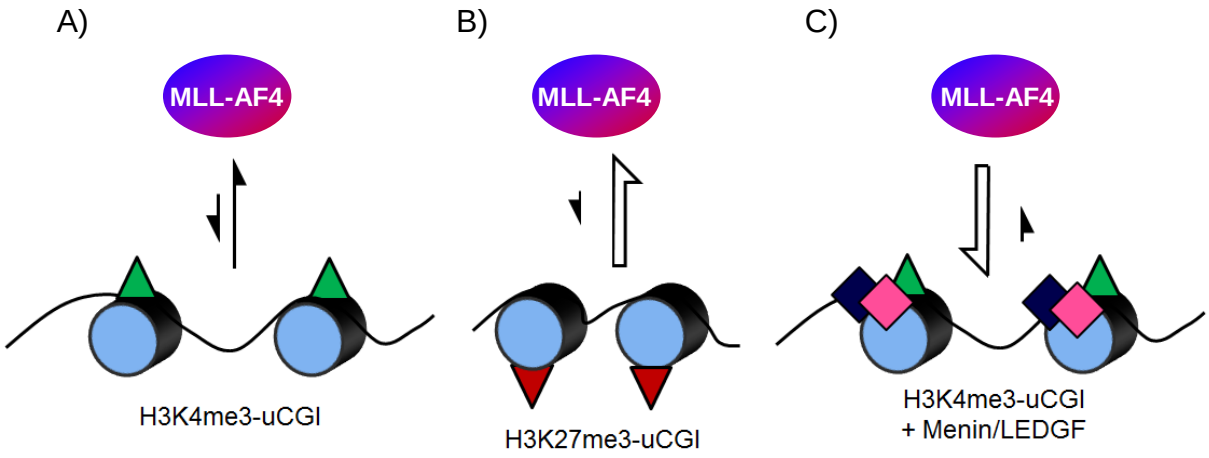


Figure 3.8 A model for MLL-AF4 recruitment to uCGIs. In a sampling based recruitment model, MLL-FPs are capable of diffusion to all uCGIs however recruitment to the site is influenced by the surrounding region. **A)** MLL-FPs are able to bind at H3K4me3-marked uCGIs but the interaction is not stable, eliciting a higher off-rate. **B)** H3K27me3-marked uCGIs are refractory to any kind of MLL-AF4 binding. **C)** At H3K4me3-marked uCGIs occupied by the Menin/LEDGF complex, binding occurs as in (A) but Menin/LEDGF stabilise binding through direct interaction with the fusion protein resulting in a higher on-rate.

4. Chapter Four – The CXXC Domain is the Primary Recruitment Module of MLL Fusion Proteins and Can Function in the Absence of Other Complexes

4.1. Introduction

Recruitment to a gene target is difficult to assess due to the presence of multiple protein complexes and activities that are taking place at a given gene at any one time. For example, at transcriptionally active loci, formation of the transcriptional machinery, elongation complexes, transcription factors and chromatin remodelling complexes can all be present, altering the accessibility and epigenetic landscape. Removing any of these proteins can have knock-on effects to various other members at the locus, either directly or indirectly. Thus knock-down experiments at endogenous loci are limited since the immediate effects are obscured, making it difficult to draw conclusions. Furthermore, it is not possible to determine from this approach which protein arrived first, which proteins were recruited as a result, or which proteins simply have preferential binding due to the presence of other factors already there. Although artificial, systems that utilise isolation from the normal activity of the genome can remove these limitations. Where available, knockout systems are advantageous over knockdowns since they allow for a fully functioning system to be established in the absence of the protein of interest and ensure no residual protein is present.

Results from Chapter 3 provided a correlation between MLL-FPs and factors thought to be involved in their recruitment. Correlations are useful because they can potentially eliminate proteins not involved in recruitment. For example, if protein A causes recruitment of protein B there will be a correlation between the two. However, if there is no correlation it is possible to conclude that there is no causative connection. Correlations are limited by the fact that they cannot prove causation. For example,

despite a correlation, it cannot be determined if protein A can recruit protein B or vice versa. To build on experiments from Chapter 3, more functional approaches were taken to assess the roles of the Menin/LEDGF and PAF complexes as well as the CXXC domain.

4.2. The CXXC Domain is Sufficient to Bind at all Active Promoter uCGIs

Data from Chapter 3 showed that MLL-FPs are recruited to a restricted subset of uCGIs, potentially due to a limitation imposed on the CXXC domain by surrounding domains, and proteins that interact with the full-length fusion protein. To investigate the binding capability of the CXXC domain in the absence of these, an isolated part of the fusion protein containing just the CXXC domain and its flanking regions (DNMT region from now on) was generated (Figure 4.1A). Since previous studies have shown an interaction between this region and members of PAFc and Menin (Milne et al., 2010; Muntean et al., 2010), it was important to determine if these interactions translated into functional recruitment that could complicate recruitment of the isolated CXXC domain. To test this, an artificial Tet recruitment system designed and created by the Klose lab (Blackledge 2013) was used. Briefly, a Tet-Operon (TetO) array was centrally inserted into a BAC that was void of CGIs, promoter and enhancer features. Using Tol2 sites, this BAC was inserted into chromosome 8 of mouse embryonic stem cells (mESCs). By fusing Tet-Repressor (TetR) to a protein of interest, that protein will be the only thing that actively binds to the TetO array. If other factors can be detected using ChIP across the TetO array, it can be assumed that they were recruited there by the TetR-protein (Figure 4.1B). Stable cell lines were generated by transfecting the TetO-Tol2 (TOT2N) mESCs with TetR-DNMT or DNMT-TetR, both with a DNA-binding deficient CXXC domain, in a plasmid expressing puromycin resistance (see Materials and Methods for further details).

Figure 4.1

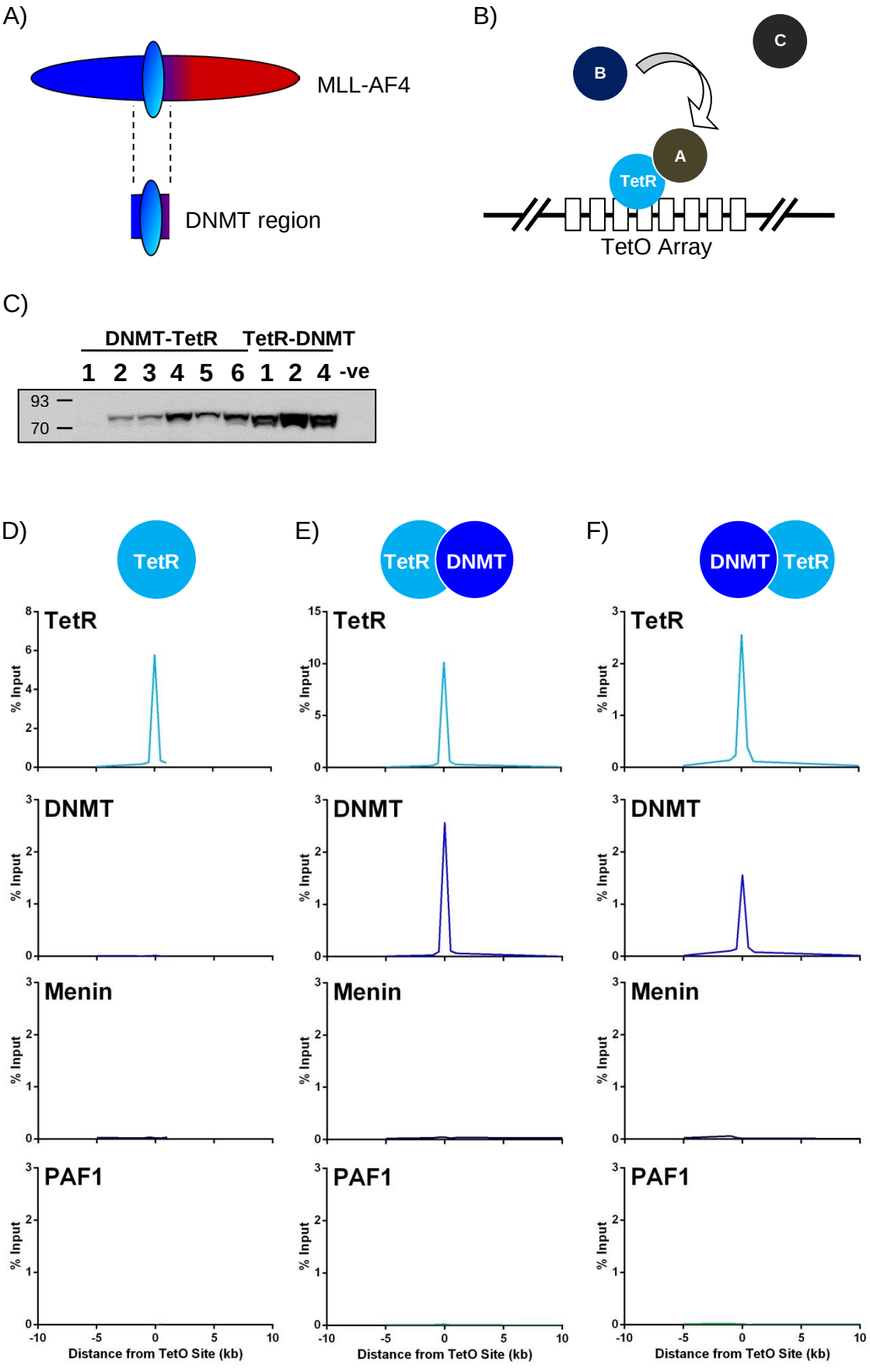

A)
MLL-AF4
DNMT region
B)
C
B
A
TetR
TetO Array
C)
DNMT-TetR
TetR-DNMT
1 2 3 4 5 6 1 2 4 -ve
93
70
D)
E)
F)
TetR
DNMT
TetR
DNMT
TetR
DNMT
Menin
PAF1
% Input
Distance from TetO Site (kb)

Figure 4.1 The DNMT region cannot recruit Menin or PAF1. A) Schematic of the cloned DNMT construct in the context of full-length MLL-AF4. **B)** Schematic of the Tet recruitment system: Protein A is fused to the Tet repressor (TetR) causing it to bind at the Tet operon (TetO) array. Through interaction with Protein A, Protein B is recruited to the array. Protein C is not recruited by Protein A thus it is undetectable. **C)** Western blot of stable TOT2N mESCs containing either DNMT-TetR or TetR-DNMT. A high-affinity FS2 antibody was used. **D)** ChIP-qPCR of various proteins in a TOT2N mES line expressing TetR alone, **E)** TetR-DNMT or **F)** DNMT-TetR.

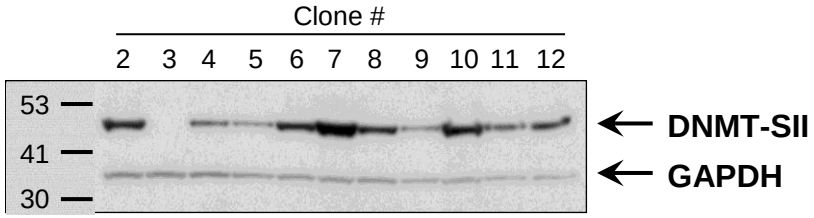
Clonal cell lines were selected by growing in puromycin-supplemented media. The presence of either construct was confirmed by western blot, with clones showing the highest expression selected for further experiments (Figure 4.1C). A flag-strep-strep (FS2) tag at the N-terminus of TetR allowed for strong detection of the construct in ChIP. ChIP-qPCR confirmed that both TetR-DNMT and DNMT-TetR bound at the TetO array (Figure 4.1D-F, top row). A N-terminal MLL(N) antibody that binds in the DNMT region was used to confirm that both constructs bind at the TetO in the stable cell lines but not in the TetR-only control (Figure 4.1D-F, second row). Contrary to immunoprecipitation data (Milne et al., 2010; Muntean et al., 2010), neither Menin nor PAFc members were detected at the TetO array when the DNMT region was present (Figure 4.1D-F, bottom two rows). Therefore, although there may be a weak interaction between these proteins, the interaction is not sufficient for recruitment in this system. This data showed that the binding of the CXXC domain could be assessed without confounding issues relating to an interaction with the Menin/LEDGF or PAFc complexes.

In order to test the genome-wide binding capability of the isolated DNMT region, it was cloned with a C-terminal streptavidin tag (DNMT-SII) to allow for high affinity ChIP detection, and transduced into HEK293T cells (293T cells from here on). Clonal cell lines that stably expressed DNMT-SII were selected for using puromycin and the presence of the construct was confirmed by western blot (Figure 4.2A). The cell line with

the highest expression, clone #7, was selected to allow for the strongest detection by ChIP. ChIP-sequencing was performed on the DNMT-SII clone using an antibody against the strep tag. Bio-CAP-seq was also performed in this cell line. Publicly available data from ENCODE was accessed to analyse sites of H3K4me3 in wild-type 293T cells as a close comparison. DNMT-SII was found to bind distinctly and with high affinity at H3K4me3-marked Bio-CAP CGIs (Figure 4.2B). A composite read plot showed that at DNMT binding sites, the binding profile of the protein almost perfectly matched that of Bio-CAP (Figure 4.2C). Furthermore, genome-wide analysis showed that of the 25,902 Bio-CAP sites, the DNMT region bound at all 15,500 that were marked by H3K4me3 (Figure 4.2D). This is significantly greater than the number of uCGIs bound by MLL-AF4 in SEM cells suggesting that the CXXC domain of the fusion protein has a larger binding capability when in isolation compared to in the context of the larger protein. However, cell-type specific differences cannot be ruled out. This follows on from ideas in Chapter 3 suggesting that a combination of size and regions of the fusion protein surrounding the CXXC domain could cause steric hindrance and might explain why it has a lower binding capacity than KDM2B or CFP1. Interestingly, although some binding of the DNMT region was seen at Bio-CAP sites not marked by H3K4me3, it was noticeably less than those that do contain the histone mark (Figure 4.2D). This in line with the epigenetic preference of full-length MLL-AF4 and reinforces the idea that an actively marked region is required for stable binding. Based on results from the TetO experiments this also confirms that the MLL-FP CXXC domain is sufficient for recruitment to uCGIs without the need for the Menin/LEDGF or PAF complexes.

Figure 4.2

A)



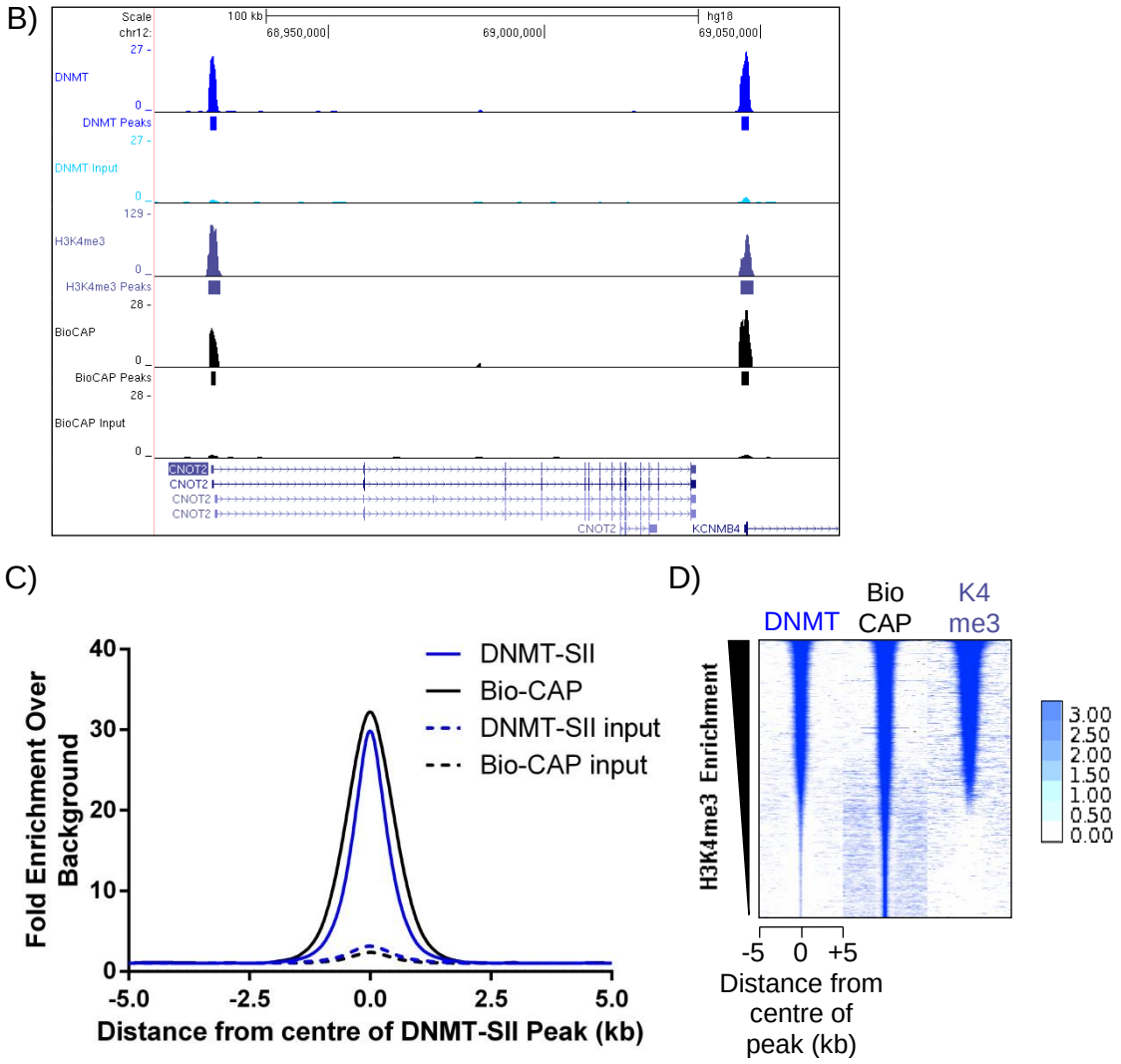


Figure 4.2 The DNMT region is recruited to all H3K4me3-marked uCGIs. A) Western blot using streptavidin and GAPDH antibodies to detect stable expression of DNMT-SII in 293T clonal cell lines. **B)** Example ChIP-seq tracks in a clonal 293T cell line stably expressing the DNMT-SII construct**.** Scale bars represent tags per 10 million. **C)** Composite binding plot of DNMT-SII and Bio-CAP at DNMT-SII binding sites. **D)** Heat-map showing enrichment of DNMT-SII and H3K4me3 at all 25,902 Bio-CAP sites in the clonal 293T cell line. Scale bar represents tags per base-pair per 10 million.

4.3. Interaction with the Menin/LEDGF Complex Confers Bi-Directional Recruitment

In a system where one protein, protein A, recruits another, protein B, there are certain assumptions that must hold true. If protein A is sufficient for recruitment of protein B then

the interaction between the proteins must confer the appearance of protein B anywhere protein A goes. Correlations from Chapter 3 showed a strong overlap between MLL-FPs and the Menin/LEDGF complex however, as explained above, this cannot show if one is capable of recruiting the other. The Tet artificial recruitment system was used to assess the ability of Menin to recruit MLL-AF4 in this way. This system is powerful because it allows for testing of recruitment in the absence of any other factors at the binding site. At endogenous loci there are many factors and events taking place that could be influencing recruitment of proteins. In this isolated system these factors can be ignored. For example, it is possible to assess the potential recruitment of MLL-AF4 by Menin in the absence of PAFc or RNAPII complexes. The weakness of the system is that it is artificial and any hierarchy of recruitment that occurs at natural loci cannot be determined. It does however still give an indication of whether or not a specific interaction is sufficient for recruitment of a protein.

TetR was independently cloned onto the N- and C- termini of full-length MLL-AF4 with a DNA-binding deficient CXXC domain, and the generation of stably expressing cell lines was attempted in the same way as with the DNMT region. However, either due to the large size of the construct or due to toxicity of the fusion protein to the mESCs, no clones were recovered. This may be as a result of growth inhibition from increased activation of cell cycle regulators by overexpressed MLL-AF4, as has been previously reported (Xia et al., 2005). To get around this issue, TetR-MLL-AF4 was transiently transfected into the original TOT2N mES cell line and cells were collected and fixed for ChIP 24 hours later to reduce the impact of the fusion protein on the survival of the cells. TetR-Menin and TetR-PAF1 were cloned at a later date and in the interest of time, transient transfections of these constructs were also chosen over the generation of stable cell lines.

TetR-Menin and TetR-PAF1 were co-transfected with a C-terminally triple-HA tagged MLL-AF4 (MLL-AF4-3HA) to assess their ability to recruit the fusion protein. To test the efficiency of the 3HA tag for ChIP, the construct was first transfected into 293T cells. Strong enrichment at *HoxA9* over background was seen for both MLL(N) and AF4(C) antibodies, but most importantly for an HA antibody (Figure 4.3A). Expression of the MLL-AF4-3HA construct in the TOT2N mES cells proved too difficult to detect by western blot so RNA levels were determined and compared to untransfected cells. Primers that amplify specifically across the fusion breakpoint showed strong expression of MLL-AF4 in the transfected cells (Figure 4.3B). Binding of TetR-Menin at the centre of the TetO array was confirmed using ChIP with antibodies against FS2 and Menin (Figure 4.3D). This detection was specific to the TetR-Menin protein, since TetR alone resulted in no detection of Menin (Figure 4.3C). Importantly, when co-transfected with MLL-AF4-3HA, Menin was able to recruit the fusion protein to the TetO array (Figure 4.3D). To date, characterisation of the interaction between Menin and the fusion protein has been done *in vitro*. Here, recruitment of MLL-AF4 by Menin in a cell shows that the interaction is functional for recruitment *in vivo*. In a separate experiment TetR was fused to the N-terminus of MLL-AF4. Interestingly, when transfected into TOT2N mESCs the fusion protein was able to recruit Menin showing that the recruitment ability conferred by the interaction works in both directions (Figure 4.3E). This could also hold true at endogenous loci since the CXXC domain of MLL-AF4 is sufficient for recruitment to genes (Figure 4.2). These results reinforce data from Chapter 3 identifying Menin as a potential recruiter of MLL-FPs but show that the current model of a hierarchical recruitment system is too limited. Instead this shows that either protein has the ability to recruit the other to gene targets.

Figure 4.3

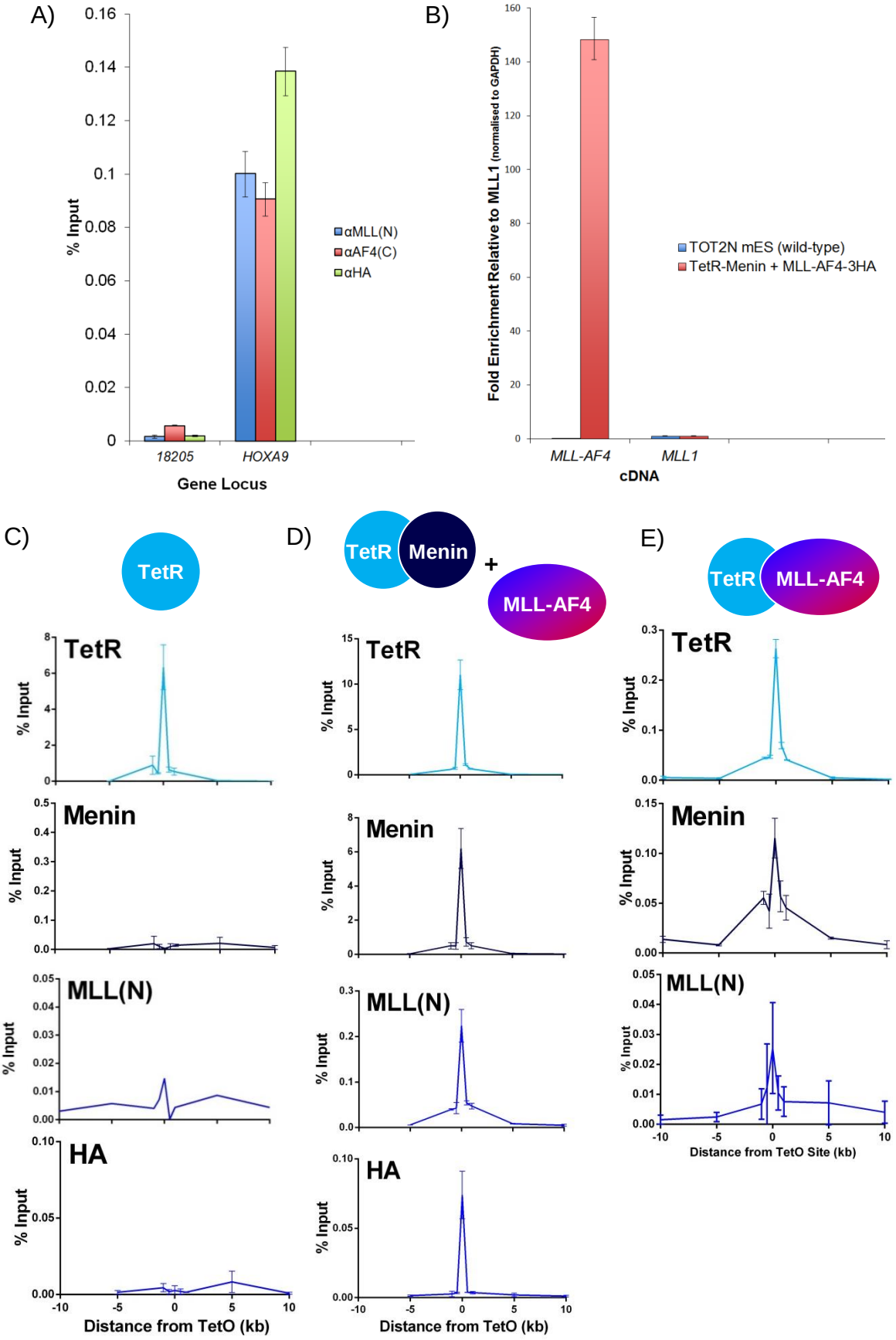

A)
% Input
αMLL(N)
αAF4(C)
αHA
18205
HOXA9
Gene Locus
B)
Fold Enrichment Relative to MLL1 (normalised to GAPDH)
TOT2N mES (wild-type)
TetR-Menin + MLL-AF4-3HA
MLL-AF4
MLL1
cDNA
C)
TetR
TetR
Menin
MLL(N)
HA
Distance from TetO (kb)
D)
TetR
Menin
MLL-AF4
Distance from TetO (kb)
E)
TetR
MLL-AF4
Distance from TetO Site (kb)

Figure 4.3 Menin and MLL-AF4 show bi-directional recruitment. A) ChIP-qPCR in 293T cells transfected with MLL-AF4-3HA at the negative and positive binding regions *18205* and *HoxA9*. Error bars represent standard deviations of 2 technical replicates. **B)** Expression of MLL-AF4 in TOT2N mES cells transfected with TetR-Menin and MLL-AF4-3HA. Error bars represent standard deviations of 3 technical replicates. **C)** ChIP-qPCR in a TOT2N mESC line stably expressing TetR alone. **D)** ChIP-qPCR in a TOT2N mESC line transiently transfected with TetR-Menin and MLL-AF4-3HA, and **E)** TetR-MLL-AF4. Error bars represent standard deviations of 2 biological replicates.

4.4. The Menin/LEDGF Complex is not Necessary for MLL-AF4 Recruitment

Another assumption that must hold true if protein A is necessary for the recruitment of protein B is that protein A must be present in the cell for recruitment to occur. So far it has been shown that there is a large variety in the amount of Menin/LEDGF complex between sites of MLL-FP recruitment (Figure 3.5). Furthermore, at sites of overlap the interaction is not restricted to uni-directional recruitment (Figure 4.3). To test if the complex is necessary to recruit the fusion protein, a streptavidin-tagged MLL-AF4 was transfected into *LEDGF* knockout (*LEDGF*$^{-/-}$) Mouse Embryonic Fibroblasts (MEFs). The lack of LEDGF was confirmed by western blot and compared to the littermate *LEDGF*$^{f/+}$ MEF line (Figure 4.4A). Loss of LEDGF binding in the knockout cell line was confirmed by ChIP-qPCR at multiple loci in both cell lines (Figure 4.4B). Despite the absence of LEDGF, MLL-AF4 was still able to bind at *HOXA9*, with minimal detection at a negative control region (Figure 4.4C). Since *HOXA9* is a site of overlap between MLL-AF4 and the Menin/LEDGF complex, this shows that the complex is not necessary for recruitment to this site. Over-expression of the fusion protein as a result of transfection might explain why MLL fusion proteins were seen in the absence of the Menin/LEDGF complex in this system but not in SEM or ML-2 cells. Nonetheless, this experiment still shows a

molecularly important concept with regards to binding of the fusion protein. A point mutation was made in the CXXC domain of MLL-AF4 at a lysine residue that is involved in direct contact with the DNA (K1176A). This drastically reduced binding of the fusion protein at *HOXA9* in the *LEDGF*$^{-/-}$ cells but did not abolish it (Figure 4.4C). This could be due to other residues in the CXXC domain that also make contact with the DNA (Allen et al., 2006) or it could be due to stabilisation through AF4 interaction with the SEC/BRD4 complex (Dawson et al., 2011). Unfortunately transfection of *LEDGF*$^{f/+}$ MEFs with the fusion protein led to severe cell death and further experiments were not possible. This is perhaps not surprising since MLL-AF4 can have full activity in the presence of the Menin/LEDGF complex and this likely led to toxicity, possibly via upregulation of cell cycle inhibitors as mentioned above (Xia et al., 2005).

To further test the necessity of the Menin/LEDGF complex, a small molecule inhibitor (MI-2-2) that targets the MLL-binding pocket of Menin was used to treat SEM cells. This had the advantage over the experiments in MEF cells as it looked at MLL-AF4 in an MLL-r leukaemic state. Treatment of SEM cells with a concentration of 12.5μM MI-2-2 was sufficient to abolish the interaction between MLL-AF4 and Menin after 24 hours (Figure 4.4D). Interestingly, treatment with MI-2-2 reduced whole cell levels of Menin suggesting that the protein relies on being in complex with the fusion protein for stable expression, possibly explaining the rigid correlation between the two at gene targets. At *HOXA9* there was no change in binding of MLL-AF4 after 48 hours of MI-2-2 treatment, however at a uCGI in the gene body of *SPT3*, binding of MLL-AF4 was reduced but not abolished (Figure 4.4D). The failure to abolish binding of MLL-AF4 shows that it is still recruited to the genes in the presence of the inhibitor. Menin is still present at the gene loci after the 48-hour treatment but the Co-IP experiments suggest there is no interaction with the fusion protein (Figure 4.4D). The difference in effect at the *HOXA9* and *SPT3* loci argues for varying importance of the Menin/LEDGF complex at different regions. As already alluded to, the *SPT3* region is in the gene body whereas *HOXA9* is near the

promoter at a uCGI-dominant landscape. The latter could allow for greater involvement of the CXXC domain and require a reduced role for the Menin/LEDGF complex. These data together show that MLL-FPs can bind at genes despite absence of components of, or loss of interaction with, the Menin/LEDGF complex. This suggests that the Menin/LEDGF complex is not necessary for recruitment of the fusion protein.

Figure 4.4

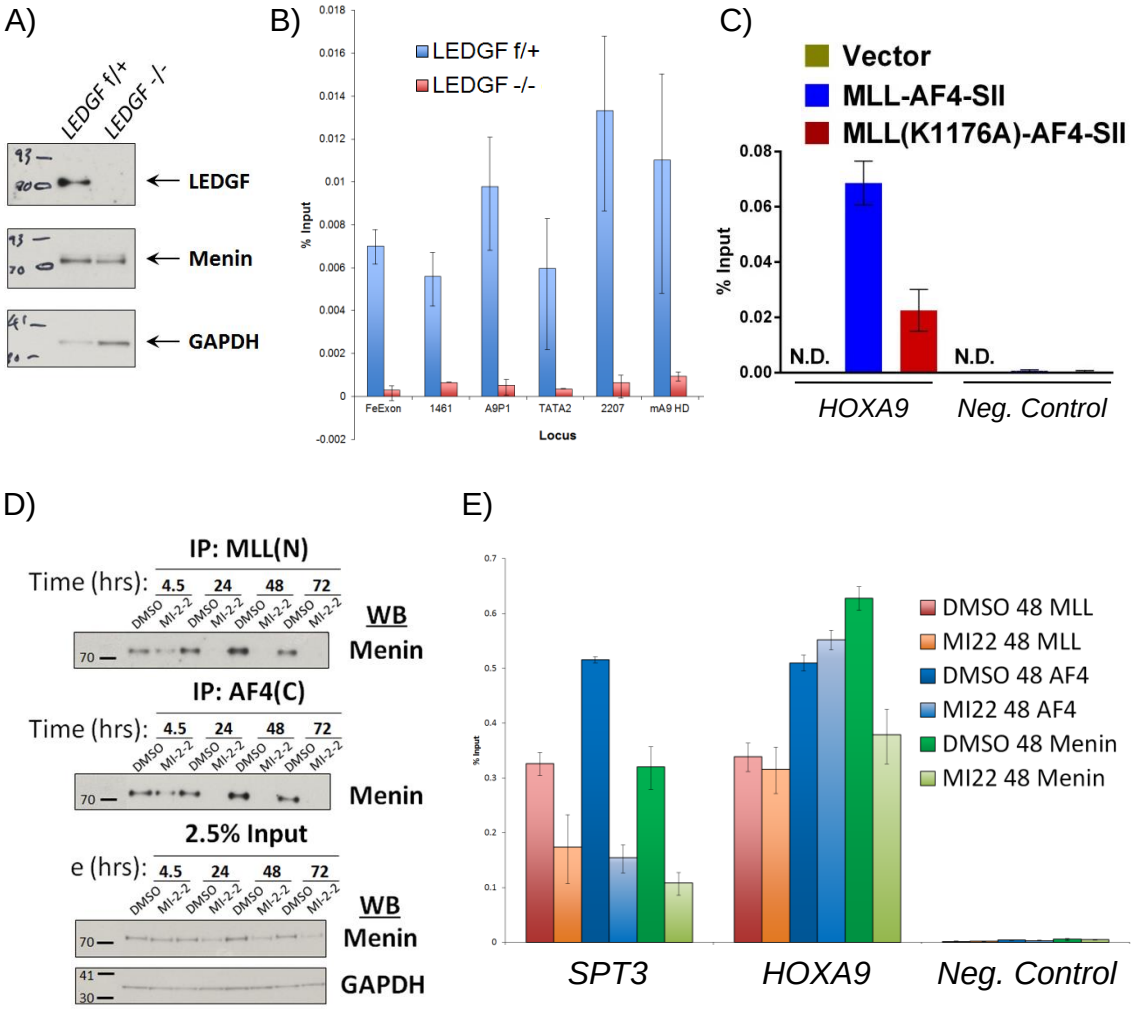


Figure 4.4 The Menin/LEDGF complex is not necessary for MLL-AF4 recruitment. A) Western blot showing expression levels of LEDGF and Menin in LEDGF f/+ and -/- MEF cells. **B)** ChIP-qPCR of LEDGF across the *HoxA9* locus in *LEDGF* f/+ and -/- MEF cells **C)** Streptavidin ChIP-qPCR in LEDGF -/- MEF cells transiently transfected with vector (green), MLL-AF4-SII (blue) or MLL-AF4-SII with a DNA-binding point mutation in the CXXC domain (K1176A, red). **D)** IP/IB experiments in SEM cells treated with 12.5μM MI-2-2 at the time-points indicated. **E)** MLL(N), AF4(C) and Menin ChIP-qPCR in SEM cells treated with 12.5μM MI-2-2 for 48 hours. Error bars represent standard deviations of 2 technical replicates.

4.5. PAFc is Unable to Recruit MLL-AF4

To investigate the recruitment properties of the PAFc interaction, TetR-PAF1 was cloned and expressed in the TOT2N mESCs. The presence of PAF1 was confirmed with FS2 and PAF1 antibodies and again shown to be a direct consequence of TetR-PAF1 and not TetR alone (Figure 4.5A; B). TetR-PAF1 was able to recruit LEO1, another member of PAFc, to the TetO array, showing that it was able to recruit other members of the complex as expected (Figure 4.5B). When co-transfected alongside MLL-AF4-3HA, the fusion protein could not be detected by HA ChIP (Figure 4.5B). The reverse experiment, with TetR-MLL-AF4, did not detect PAF1 or LEO1 at the TetO array (Figure 4.5C) showing that the interaction with PAFc does not confer recruitment in either direction. In contrast to previously published work (Milne et al., 2010; Muntean et al., 2010) we can rule out PAFc as a candidate for recruiting MLL-AF4. Data from previous studies (Milne et al., 2010; Muntean et al., 2010) that showed that loss of PAFc leads to reduced MLL-FP binding may argue that PAFc indirectly affects MLL-FP binding.

4.6. Discussion

The use of functional assays to assess the recruitment role of different proteins has allowed for a deeper understanding of their involvement. This chapter has shown that the CXXC domain of MLL-FPs is sufficient for recruitment to uCGIs even in the absence of an interaction with the Menin/LEDGF or PAF complexes. Furthermore, knockout and inhibitor studies were used to show that MLL-AF4 is still recruited to genes despite an absence or lack of interaction with the complex. Interestingly, the interaction between the fusion protein and the complex confers recruitment in both directions suggesting that unlike previously thought, MLL-AF4 may be recruiting Menin/LEDGF to genes (summarised in Figure 4.6).

Figure 4.5

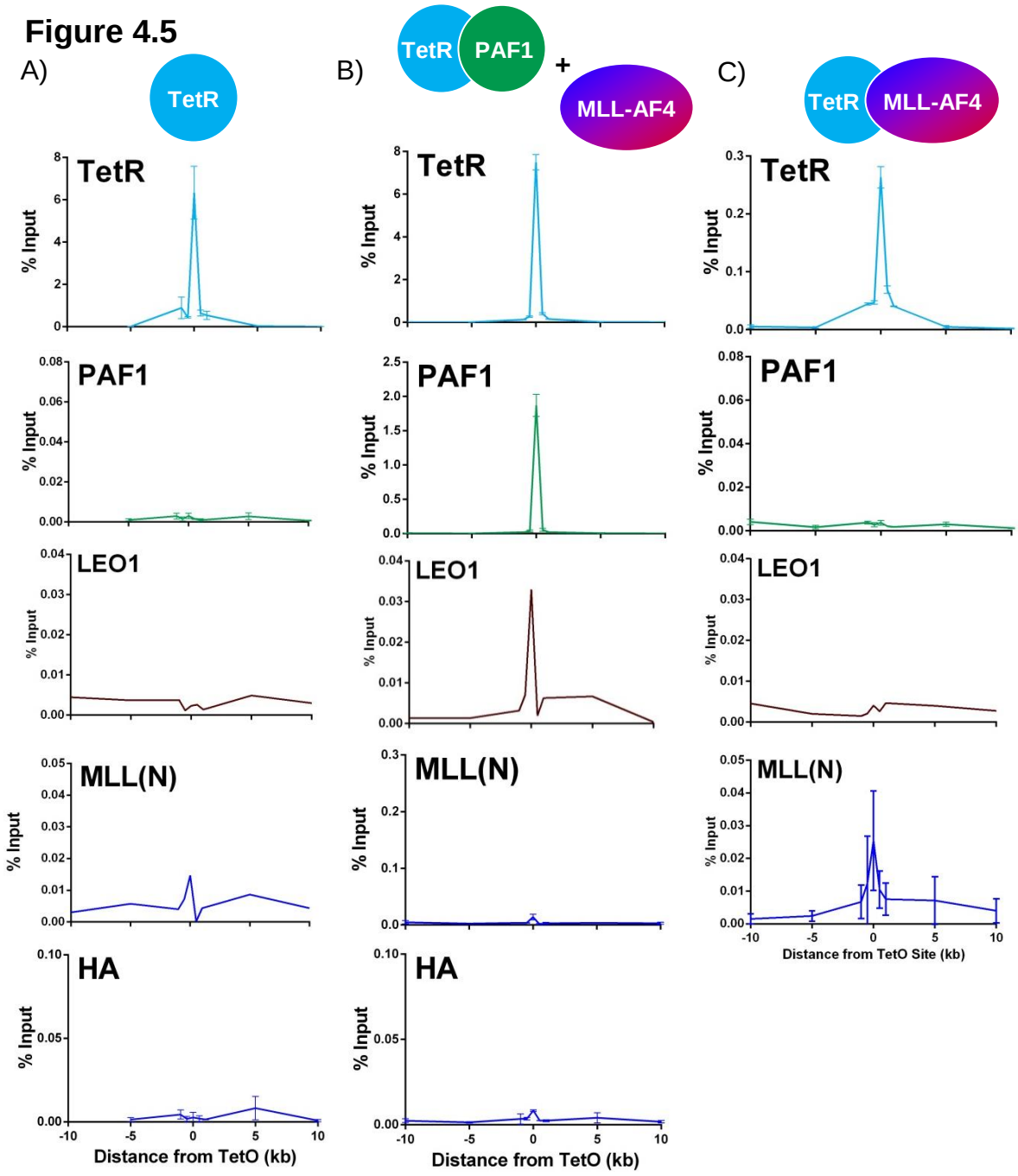


Figure 4.5 PAF1 and MLL-AF4 cannot recruit each other. A) ChIP-qPCR of various proteins in a TOT2N mESC line stably expressing TetR alone. **B)** ChIP-qPCR in a TOT2N mESC line transiently transfected with TetR-PAF1 and MLL-AF4-3HA, and **C)** TetR-MLL-AF4. Error bars represent standard deviations of 2 biological replicates.

Figure 4.6

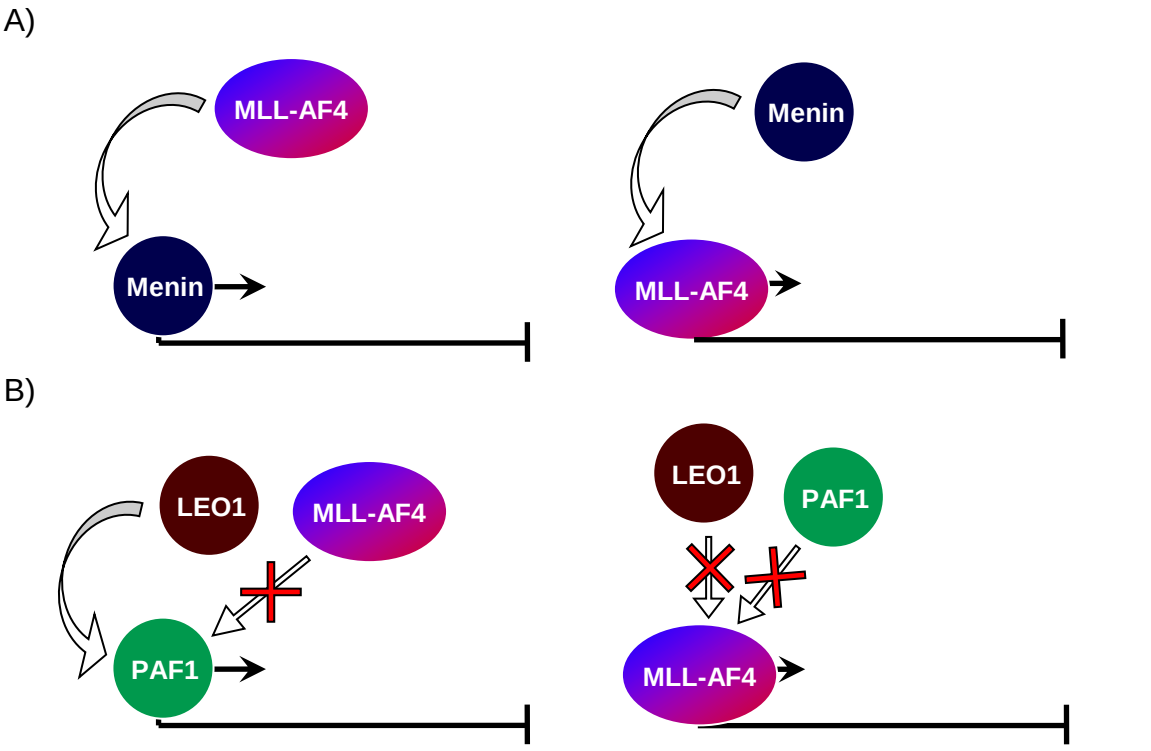


Figure 4.6 Schematic of MLL-AF4, Menin and PAF1 recruitment potentials. A, left) Menin is sufficient for the recruitment of MLL-AF4. **A, right)** MLL-AF4 is sufficient to recruit Menin. **B, left)** PAF1 can recruit other members of PAFc (LEO1) but not MLL-AF4. **B, right)** MLL-AF4 is not able to recruit PAFc members.

Correlations with PAFc in Chapter 3 were not significant enough to draw a meaningful conclusion. The use of an artificial recruitment system confirms that this complex is unable to recruit MLL-AF4 putting it aside as unimportant in the process. PAFc has previously been shown to mediate an interaction via BRD3/4 and MLL-AF4 (Dawson et al., 2011). It is possible that disruption of the MLL-FP-PAFc interaction in previous studies (Milne et al., 2010; Muntean et al., 2010) affected MLL-FP binding indirectly by interfering with the BRD3/4 recruitment pathway as opposed to having a direct effect on recruitment. Furthermore, the effects of PAFc in these studies might be very specific to the tested *HoxA9* since despite a PAF1-MLL interaction there are still many sites where no overlap is observed (chapter 3 of this thesis; (Milne et al., 2010)).

Ultimately this chapter shows that the current model of MLL fusion protein recruitment needs readdressing. Until now equal importance has been placed on the Menin/LEDGF complex and the CXXC domain with a less well-characterised role for PAFc. Results from Chapter 3 and 4 suggest a model whereby MLL-AF4 can be recruited to gene targets in the absence of the Menin/LEDGF and PAF complexes. However, a strong correlation with the Menin/LEDGF complex and a reduction in MLL-FP binding in the presence of an MLL-Menin inhibitor argues for an important role in stabilising binding of MLL-FPs. Since MLL-AF4 was still bound at gene targets despite a loss of interaction in the presence of the Menin inhibitor, the most important role of the Menin/LEDGF complex might be at the initial recruitment step after which it is less important. Interaction with MLL-FPs appears important for the stability of Menin, with the former possibly recruiting the latter (see summary Figure 4.7). This could explain why the Menin/LEDGF complex is found at sites absent of H3K36 methylation. Although not tested here, another possibility for stable binding of MLL-FPs in the absence of Menin/LEDGF could be via interaction with BRD3/4. Since MLL-FPs are recruited to H3-acetylated uCGIs, this provides a platform for BRD3/4 binding. Direct interaction between BRD3/4 and the AF4/SEC may facilitate the initial binding event or act to stabilise the fusion protein after initial CXXC-mediated binding.

Figure 4.7

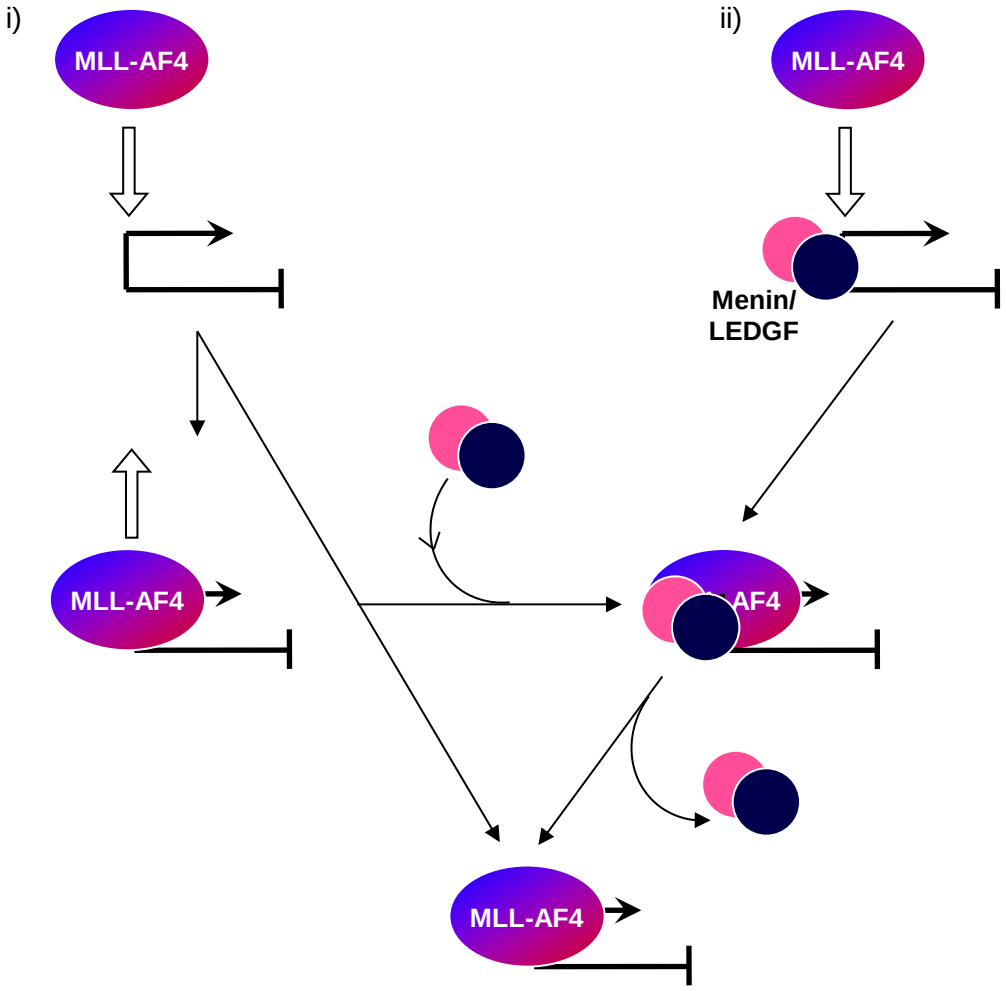


Figure 4.7 A model for MLL-FP recruitment. In all 3 scenarios, MLL-AF4 is able to bind independently at the gene promoter via its CXXC domain. The propensity for stable binding is assessed in different scenarios. **i)** The absence of stabilising factors (Menin/LEDGF in this example) means MLL-AF4 cannot stably bind and is lost from the promoter (solid line, left; refers to Figure 3.5). Alternatively, LEDGF KO studies argue that it is possible for a direct transition to stable binding in the absence of these factors (solid line, right; refers to Figure 4.4C). Instead, MLL-AF4 possibly recruits Menin/LEDGF as it binds in order to achieve stable binding (horizontal line, refers to Figure 4.3C). **ii)** CXXC-mediated binding of MLL-AF4 occurs at genes already occupied by Menin/LEDGF, which stabilises binding (refers to Figure 3.5). Once stable binding has been achieved, interaction with the Menin/LEDGF complex may be of minimal importance for its maintenance (lower third; refers to Figure 4.4D; E).

5. Chapter Five – Spreading MLL Fusion Proteins Mark a Distinct Subset of Genes that are Important in Leukaemia

5.1. Introduction

Studies have shown that specific gene expression signatures correlate with distinct subtypes of leukaemia (Armstrong et al., 2002; Ferrando et al., 2003; Ross et al., 2004; Stam et al., 2010). Certain genes have often been found to be upregulated and subsequent knockdown studies have shown these to be particularly important for disease maintenance, for example *HoxA9* in MLL-r leukaemias (Faber et al., 2009; Zeisig et al., 2004). Furthermore, where the leukaemia is driven by a particular protein, e.g. a fusion protein, these detrimental genes are often found to be direct targets of the driver protein. In MLL-r leukaemias, key upregulated genes have been found to be directly bound by the fusion protein (Milne et al., 2010; Milne et al., 2005b; Okuda et al., 2014). Building on this, the availability of ChIP-seq data for different proteins and histone marks in this thesis provided the means to identify subclasses of gene groups based on occupancy of different factors or unique binding profiles.

5.2. Spreading MLL-FPs Define a Unique Class of Gene Targets

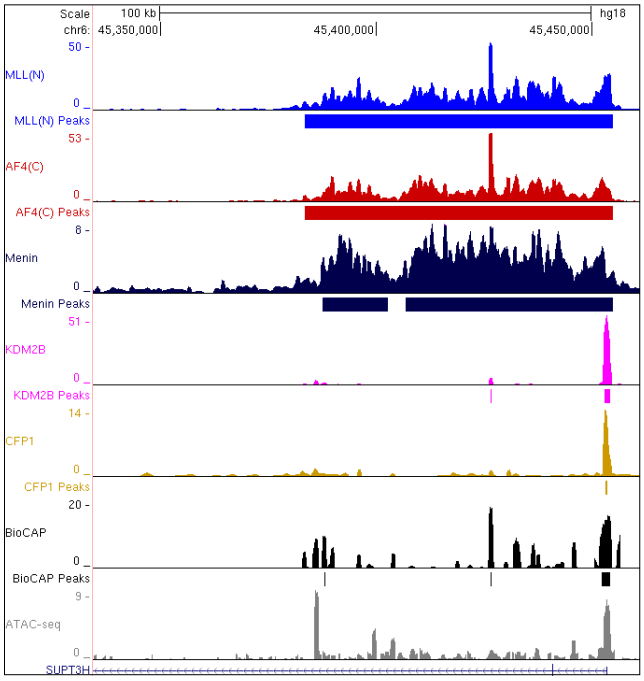
Broad regions of MLL-AF4 have been identified previously in SEM cells and were shown to mark genes important for maintenance of a stem cell-like, undifferentiated state (Guenther et al., 2008). These broad regions of MLL-AF4 were also identified in ChIP-seq presented here. To investigate their potential role in leukaemia further and to see if this was a general feature of MLL-FPs, in-depth analysis was done on these gene targets. A total of 149 genes were identified where the fusion protein spread from the promoter greater than 4kb into the gene body without extending beyond the end of the gene (Figure 5.1A-B). Due to this appearance this feature is referred to as “spreading”

from here on. These criteria were chosen because wild-type MLL1 peaks were not seen to spread greater than 4kb (Figure 5.1C). Furthermore, limiting the spreading to the end of the gene kept the peaks gene specific but also avoided regions where there is a large binding domain over lots of small genes, such as the *HoxA* cluster. A handful of these genes were verified as legitimate MLL-AF4 targets by using an siRNA specific to the fusion protein followed by ChIP for MLL(N) and AF4(C) (Figure 5.1D). Importantly, these knockdowns were tested at regions in the gene body as well as at the promoter (see Chapter 2, Table 2.4 for a full list of primer locations). Interestingly, MLL-AF6 also showed this spreading pattern, marking 47 genes in ML-2 cells (Figure 5.1E; F).

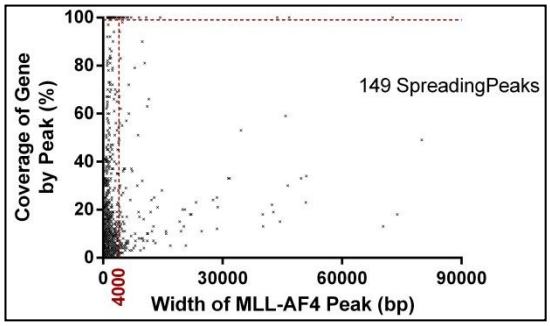
To assess if spreading is functionally important, nascent RNA-seq data from SEM cells, available in the Milne lab, was analysed to determine the relative transcription level of the 149 spreading targets. Initially, all MLL-AF4 targets were compared to targets of CFP1 which is associated with the promoters of active genes (Clouaire et al., 2014; Thomson et al., 2010). MLL-AF4 targets were more highly expressed, showing that targets of the fusion protein are in general overexpressed compared to what would be considered active genes. The gene expression of the 149 spreading targets was then compared against all MLL-AF4 targets which were sampled in groups of 149 randomly selected genes, one thousand times. The group of spreading targets was approximately 2-3-fold higher expressed compared to any other randomly sampled group (Figure 5.2A) showing a correlation between the presence of spreading and gene regulation. Whether increased expression is a cause or consequence of spreading is beyond the scope of this thesis, however the correlation revealed a link that warranted further investigation into this feature. Furthermore, spreading MLL-AF4 targets showed a significant enrichment of H3K79me2 signal compared to non-spreading targets and non-MLL-AF4 targets (Figure 5.2B). The increased abundance of this histone mark associated with gene transcription further supports a link between MLL-FP spreading and gene activation.

Figure 5.1

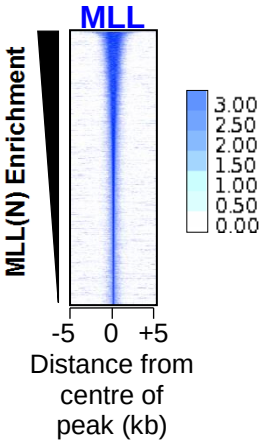
A)



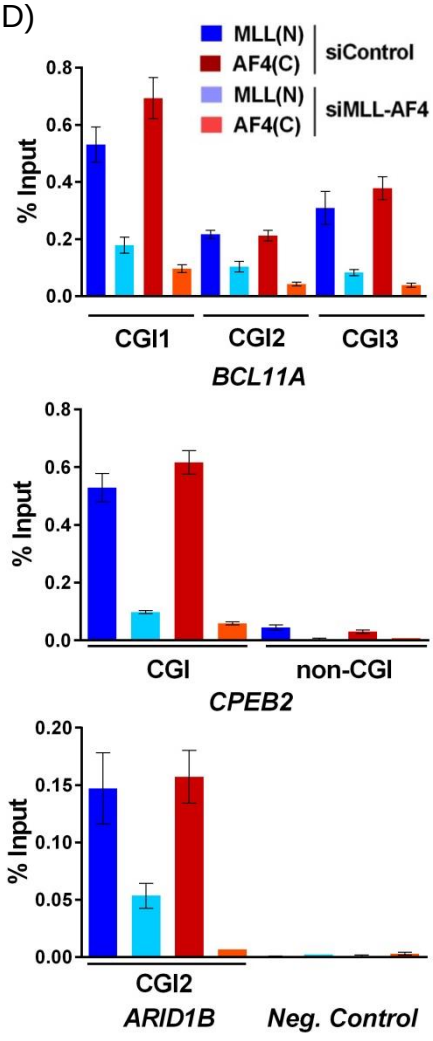
B)



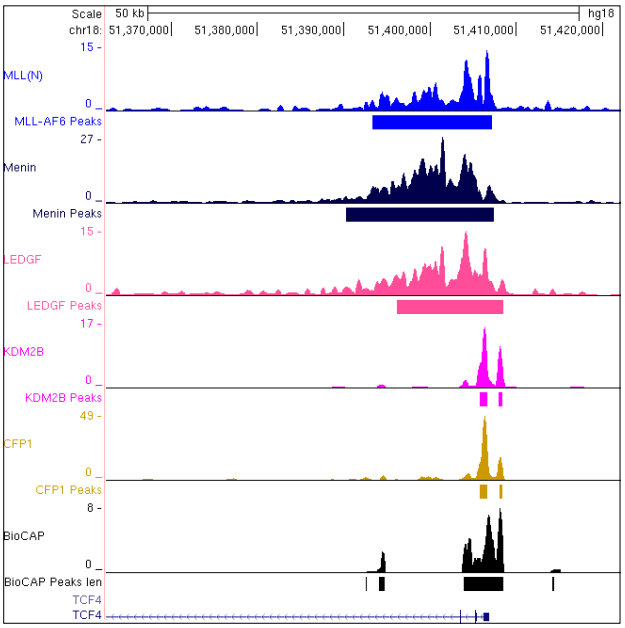
C)



D)



E)



F)

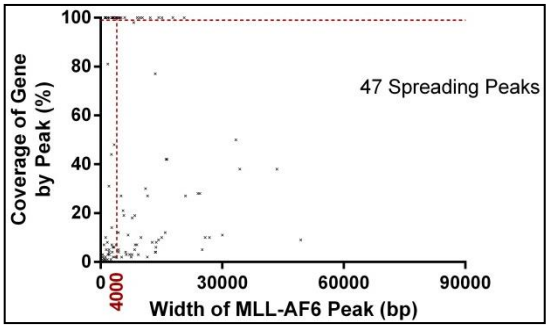


Figure 5.1 MLL-FPs show spreading recruitment. A) Example track of MLL-AF4 spreading into the gene body of *SUPT3H*, in SEM cells. Scale bars represent tags per 10 million **B)** Width of MLL-AF4 peaks against the percent coverage of the gene body by the peak. Peaks greater than 4kb that had less than 100% coverage were determined as spreading. **C)** Heat-map of wild-type MLL1 enrichment showing that it never spreads greater than 4kb. Scale bar represents tags per base-pair per 10 million **D)** *MLL-AF4* knockdown at spreading loci to confirm peaks as specific to the fusion protein. **E)** Example ChIP-seq track of MLL-AF6 spreading into the gene body of *TCF4*, in ML-2 cells. **F)** Selection of spreading MLL-AF6 peaks using the same criteria as in (B).

Figure 5.2

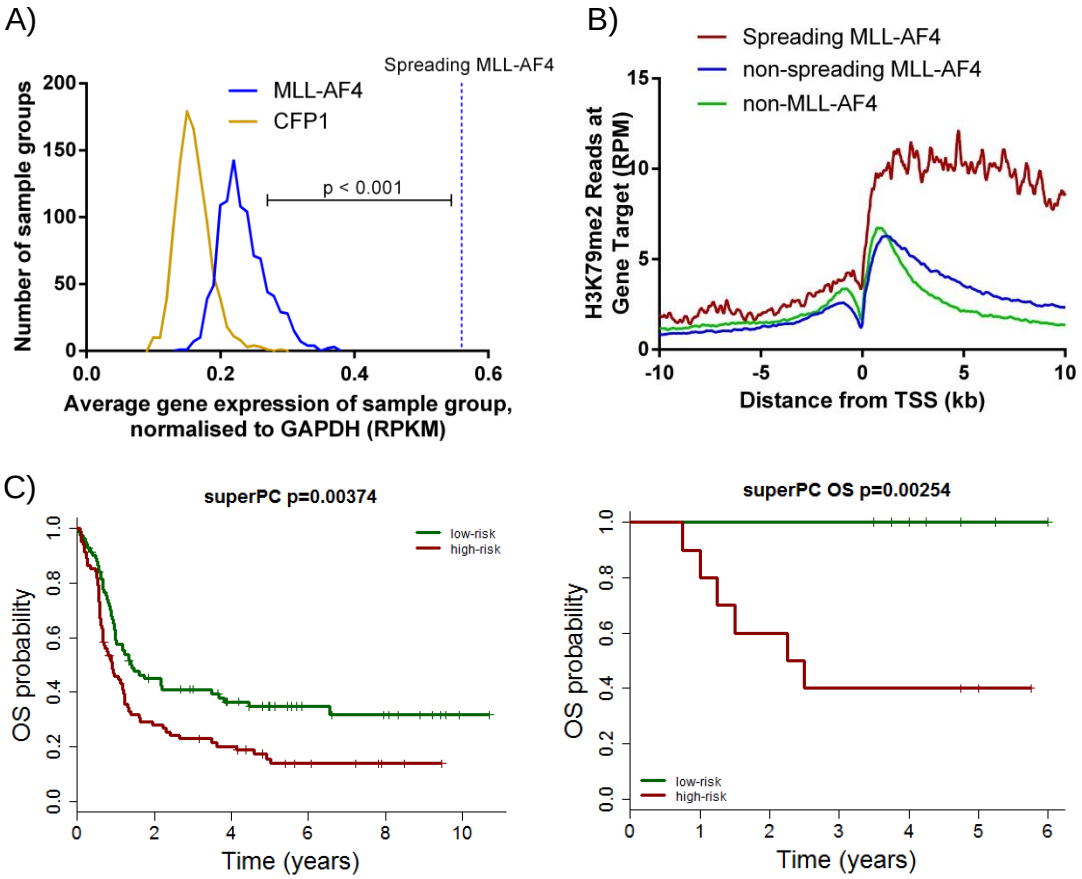


Figure 5.2 Gene targets of spreading MLL-AF4 have increased expression. A) Mean average expression of the 149 spreading MLL-AF4 gene targets compared to all MLL-AF4 (blue) and CFP1 (orange) targets. 1000 groups of 149 randomly selected genes were used to sample each set of gene targets. (Nascent RNA-seq was performed by E. Ballabio, n= 4 biological replicates) **B)** Composite read plot of H3K79me2 signal at spreading MLL-AF4 genes targets (red), non-Spreading MLL-AF4 targets (blue) and non-MLL-AF4 targets (green) **C)** Kaplan-Meier curves (analysis done by Huimin Geng) showing overall survival of patients using spreading MLL-AF4 targets as a prognostic score. **Left**, adult ALL patients, n=165. **Right,** childhood, MLL-r ALL patients, n=21.

Building on associations between spreading MLL-AF4 and increased gene expression, correlations with clinical outcomes were investigated. A SuperPC analysis was conducted by Huimin Geng, on two different ALL patient cohorts: 1. Adult ALL (ECOG 2993) and 2. Childhood ALL (COG P9906). When gene expression in patients was used as a prognostic marker, the group of spreading MLL-AF4 targets correlated with a

significant decrease in overall survival (OS) in adult ALL (n=165, p=0.00374) (Figure 5.2C, left). When just the MLL-r cases of childhood ALL were examined, this group of gene targets again correlated with reduced OS (n=21, p=0.00254) (Figure 5.2C, right). The link between spreading MLL-AF4 and reduced OS suggests that spreading recruitment of MLL-FPs might be important in leukaemia development and maintenance.

5.3. Spreading MLL-FPs Overlap with Members of the SEC and Menin/LEDGF Complex

Due to the potential importance of MLL-FP spreading in leukaemia, the mechanism behind it was investigated. To determine why spreading occurs at some gene targets but not others, ChIP-seq tracks at spreading loci were initially visualised. A heat-map of MLL(N) and AF4(C) signal at the genes was sorted by length, producing a distinct cone-shaped enrichment of both of these marks (Figure 5.3A). Spreading peaks were perfectly delimited by uCGIs, i.e. the beginning and end of the spread were always marked by this feature often with many occurring within the spreading peak (as exemplified in Figure 5.1A & E). Furthermore, the heat-map revealed that an increased enrichment of Bio-CAP followed the same cone-shaped pattern as the fusion protein (Figure 5.3A). This strongly suggests a role for the CXXC domain in facilitating this feature. However, the spreading peaks were also punctuated by non-uCGI regions suggesting that the CXXC domain alone cannot account for this feature. Furthermore, neither KDM2B nor CFP1 exhibit this spreading behaviour, despite containing a CXXC domain, instead being restricted to the promoter uCGIs (Figure 5.1A; E).

The identical spread of the Menin/LEDGF complex at spreading sites (Figure 5.1A; E) and the potentially more important role for the complex in the gene body over the promoters (Figure 4.4E) implicated it in the spreading process. Menin exhibited the same cone-shape of increased enrichment as MLL and AF4, at the spreading loci,

revealing its binding across exactly the same region as the fusion protein (Figure 5.3A). In ML-2 cells, a similar correlation was seen between spreading MLL-AF6 and Menin (Figure 5.3B). Interestingly, ENL, a member of the SEC that interacts directly with AF4 also correlated with the spreading pattern.

Figure 5.3

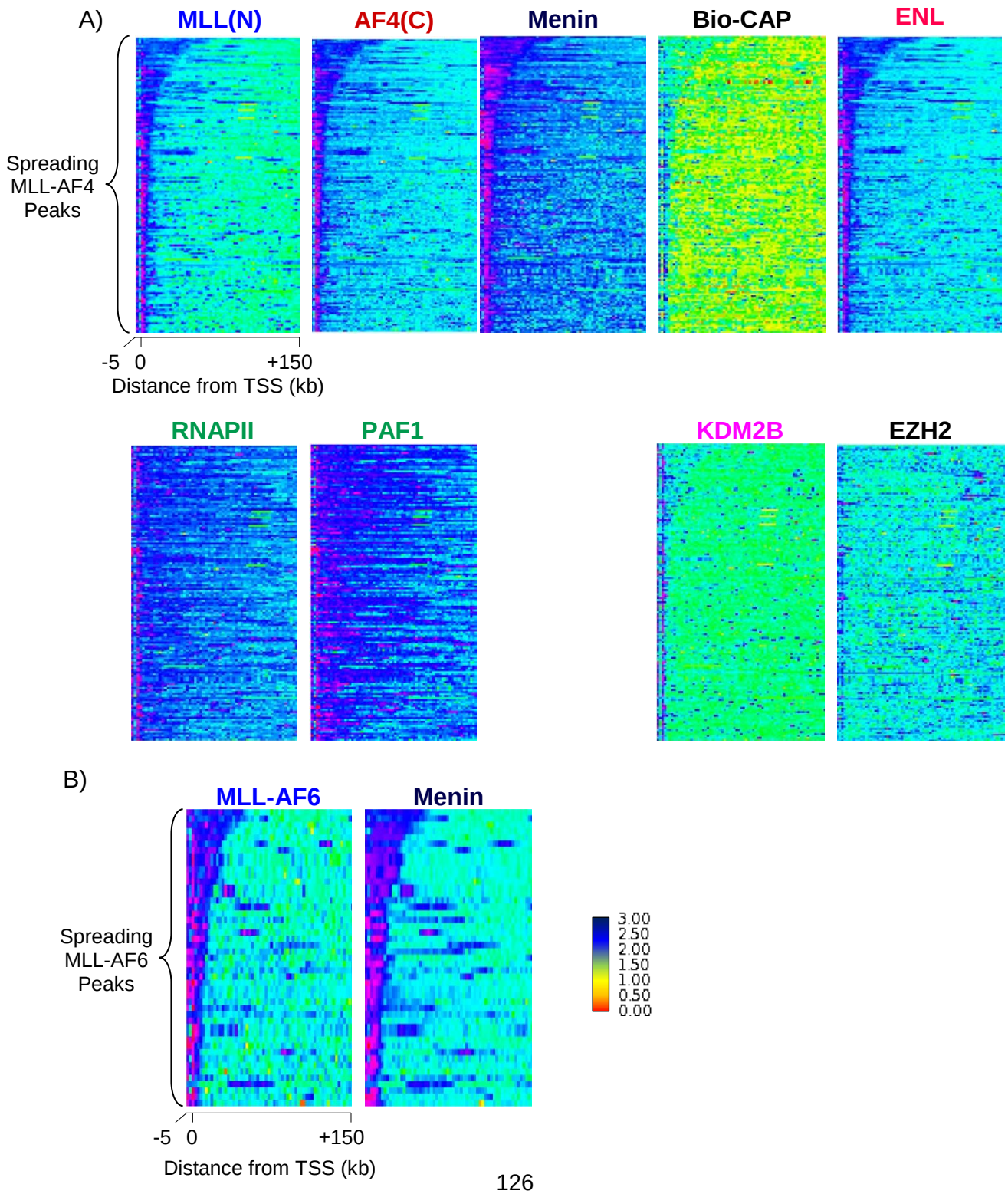


Figure 5.3 MLL-FP spreading recruitment correlates with members of the Menin/LEDGF complex and SEC. A) Heat map showing the enrichment of various proteins at all 149 spreading MLL-AF4 peaks in SEM cells, sorted by length (ENL ChIP done by E. Ballabio; RNAPII ChIP done by T. Milne). **B)** Heat-map showing the enrichment of MLL-AF6 and Menin at all 47 spreading MLL-AF6 peaks in ML-2 cells, sorted by length. Scale bar represents tags per base-pair per 10 million.

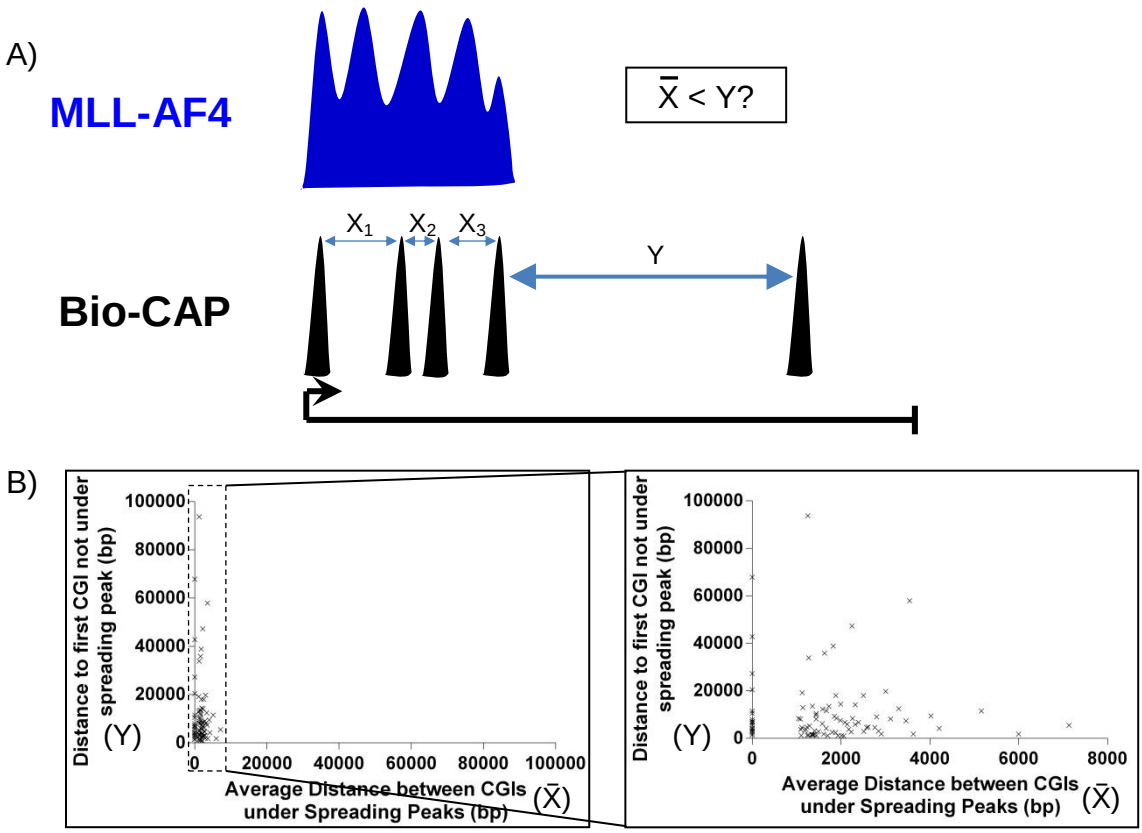
Since spreading occurs into the gene body, it may simply be a result of interaction with actively elongating RNAPII. However, both RNAPII and its interacting PAF1 showed no correlation with MLL-AF4 at spreading targets (Figure 5.3A, lower left). Instead they showed a disordered pattern, often binding significantly further into the gene bodies than spreading MLL-AF4. This reinforces data from chapter 3 and 4 which argues against a role for PAFc in MLL-FP recruitment. As controls, KDM2B, which so far has been shown to be restricted to the promoter, and EZH2 which is absent from MLL-AF4 binding sites, were also assessed for binding pattern at spreading loci (Figure 5.3A, lower right). As expected, neither showed an overlap with the spreading pattern.

5.4. MLL-FP Spreading does not Occur Across Large Non-CGI Regions

A closer observation of the relationship between spreading MLL-FPs and uCGIs revealed that uCGIs under the peaks were relatively close together. This led to the idea that MLL-FPs can spread across uCGIs in the gene body that are within a certain proximity but when the distance becomes too great spreading is inhibited. To this end, the mean average distance in base pairs between uCGIs under a spreading peak was calculated and compared to the distance to first gene uCGI that was not under the spreading peak (Figure 5.4A). Although close proximity of adjacent uCGIs was not sufficient for spreading to occur, it appeared to be necessary since spreading never occurred beyond a non-CGI region greater than 7kb, with the majority of spreading occurring across uCGIs within 1-2kb of each other (Figure 5.4B).

Next it was investigated whether differences in CpG density facilitate spreading or not. To this end the sequences under the uCGI peaks in genes where spreading of the fusion protein occurred were assessed for %GC and the observed to expected CpG ratio (Figure 5.4C). Plotting both sets of uCGIs – those under the spreading peaks versus those at a spreading gene target not under the peak – revealed no significant difference in CpG density of either group. This suggests that spreading of the fusion protein is not determined by CpG-density of the uCGIs in the gene body.

Figure 5.4



C)

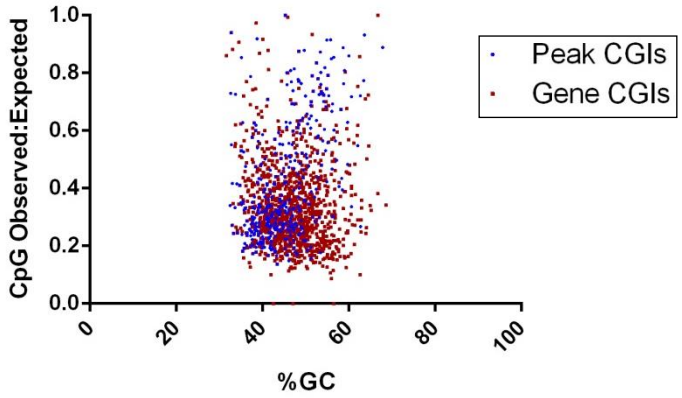


Figure 5.4 MLL-AF4 does not spread between distal uCGIs. A) Schematic of how proximity between Bio-CAP sites under spreading peaks was assessed to those outside. The mean average distance (bp) between Bio-CAP peaks under spreading peaks (X) was compared to the distance to the first Bio-CAP peak within the gene body that fell outside of the spreading peak (Y). **B)** Average distance between Bio-CAP peaks under spreading MLL-AF4 peaks was compared to distance to the first Bio-CAP peak outside the spreading peak but within the same gene. **C)** The percentage of GC dinucleotides and the Observed:Expected CpG ratio were compared for Bio-CAP peaks under the spreading MLL-AF4 peak (blue dots), against Bio-CAP peaks within the same gene but that fell outside the spreading peak (red dots).

5.5. Domains Surrounding the CXXC Domain are Necessary for Spreading Across uCGIs

In order to investigate the mechanism of spreading in more detail, various truncated constructs of full-length MLL-AF4 were made in an attempt to see if the presence or absence of any domains facilitated or inhibited spreading (Figure 5.5A). Haematopoietic cells proved difficult to transfect and on rare occasions where it was successful, the transfected constructs were toxic to the leukaemia cells, resulting in reduced proliferation or cell death. Thus, 293T cells were chosen for transfection of the constructs for subsequent ChIP-qPCR. A caveat of using a different cell type is the variation in uCGI landscape (Figure 5.5B; C). This therefore means that, if uCGI landscape is genuinely a prerequisite for spreading, loci where spreading occurs in SEM cells may not be susceptible to spreading in 293T cells. In order to select suitable

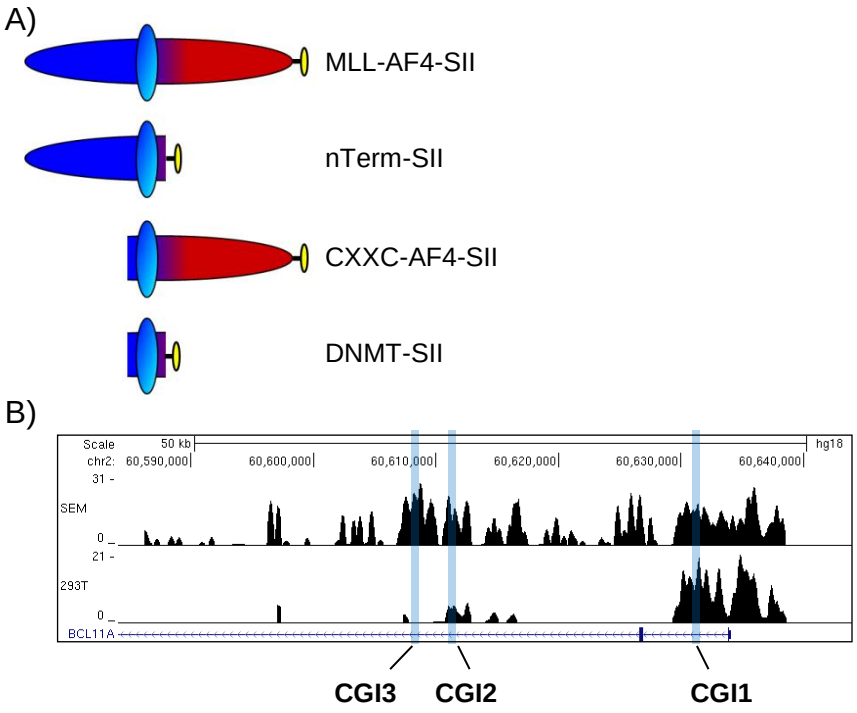
targets, Bio-CAP was performed in 293T cells and genes with uCGIs at the promoter and in close proximity within the gene body, in both cell lines, were chosen (Figure 5.5B; C). Importantly, primers were designed at these uCGIs to determine which constructs could bind at the different regions.

Constructs were transfected with 100ng/ul of PEI and cells were collected 48 hours later. The presence of a GFP cassette on the plasmids confirmed successful transfection. Expression of the protein constructs was verified by Western Blot (Figure 5.5D). The C-terminal strep tag following the AF4 moiety of the fusion protein was in some way inhibited for the purpose of western blotting. As such an AF4(C) antibody was used to detect both the CXXC-AF4 and full-length MLL-AF4 constructs. ChIP-qPCR using an antibody against the strep tag was successful for all constructs as shown by binding at the positive control *HoxA9* (Figure 5.5E). All 4 constructs bound with high affinity at the *BCL11A* promoter (BCL11A, CGI1) but did not bind at the gene body CGI that was absent in 293T cells (*BCL11A*, CGI3). Interestingly, at the two gene body uCGIs (*BCL11A* and *ARID1B*, CGI2), there was minimal binding of the DNMT region, with approximately ~2-fold higher binding of the N-terminal and C-terminal half constructs. Importantly, the N-terminal half maintains the interaction with the Menin/LEDGF complex whilst the C-terminal half maintains the interaction with factors such as ENL, mentioned above. Full-length MLL-AF4 bound with the highest affinity at all regions, including in the gene bodies. This suggests that the DNMT region is incapable of binding in the gene body and that domains of MLL-FPs in both halves of the protein can facilitate spreading, with an accumulative effect for the full-length protein. Whether this is by bridging non-CGI regions or by simply facilitating greater binding in the gene body remains unanswered. Importantly, whilst expression levels detected using anti-Strep could not be compared to those using anti-AF4, the western blots could be used to determine that DNMT-SII was higher expressed than nTerm-SII, and CXXC-AF4-SII was higher

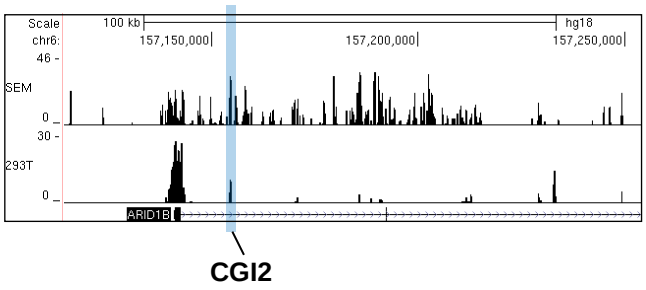
expressed than MLL-AF4-SII (Figure 5.5D). Therefore the observed higher binding is not simply the result of higher expression.

A more definitive approach to answer if constructs are in fact spreading was to do ChIP-sequencing of the constructs. Attempts to make 293T cell lines stably expressing the larger constructs repeatedly resulted in cell death. The only cell lines that remained viable were those that stably expressed DNMT-SII (Figure 4.2A). ChIP-sequencing using the strep antibody confirmed that the DNMT construct showed no spreading. In fact all DNMT peaks distinctly lined up with the promoters of genes, even at loci where Bio-CAP showed peaks downstream in the gene body (Figure 4.2B-C; 5.5F). It is possible that the chromatin binding ability of the Menin/LEDGF construct, or the elongation associated properties of the fusion partner, or a combination of both, facilitate the spreading observed for the full-length fusion protein. In fact, the binding profile of DNMT mirrors more that of KDM2B or CFP1 in the leukaemia cell lines (compare Figure 4.2C with 3.3), further suggesting that it is something unique about the surrounding domains in MLL-FP that allow for spreading to occur.

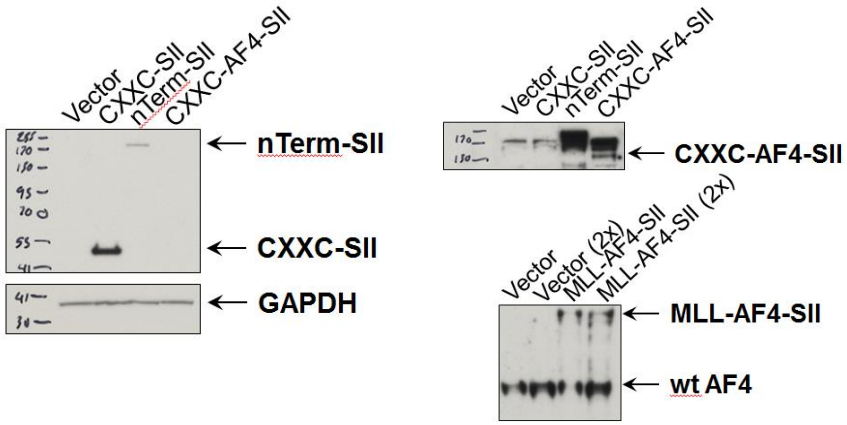
Figure 5.5



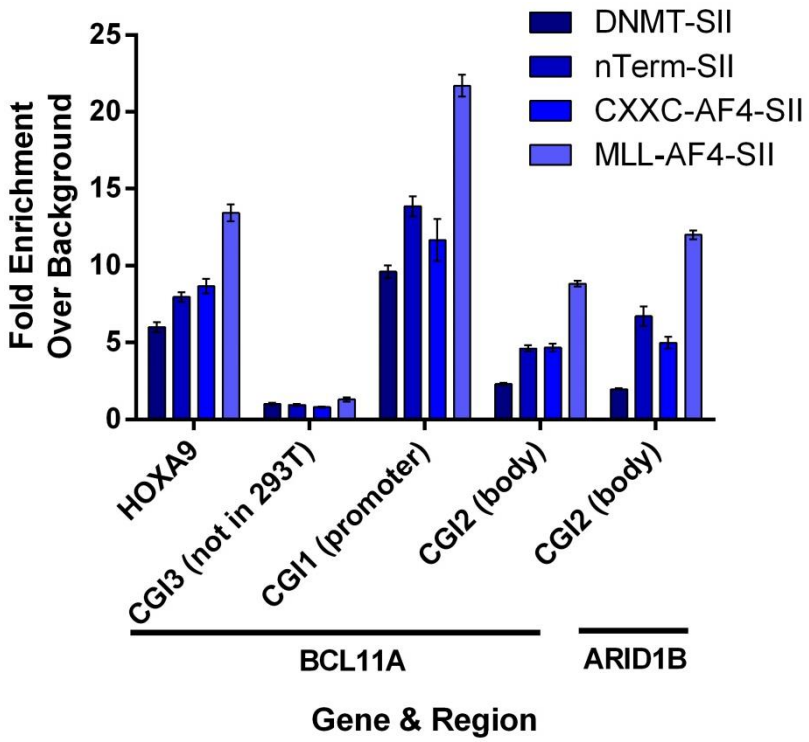
C)


Scale
100 kb
hg18
chr6:
157,150,000
157,200,000
157,250,000
46 -
SEM
0
30 -
293T
0
ARID1B
CGI2

D)


Vector
CXXC-SII
nTerm-SII
CXXC-AF4-SII
nTerm-SII
CXXC-SII
GAPDH
CXXC-AF4-SII
Vector (2x)
MLL-AF4-SII
MLL-AF4-SII (2x)
MLL-AF4-SII
wt AF4

E)


DNMT-SII
nTerm-SII
CXXC-AF4-SII
MLL-AF4-SII
Fold Enrichment Over Background
25
20
15
10
5
0
HOXA9
CGI3 (not in 293T)
CGI1 (promoter)
CGI2 (body)
CGI2 (body)
BCL11A
ARID1B
Gene & Region

F)

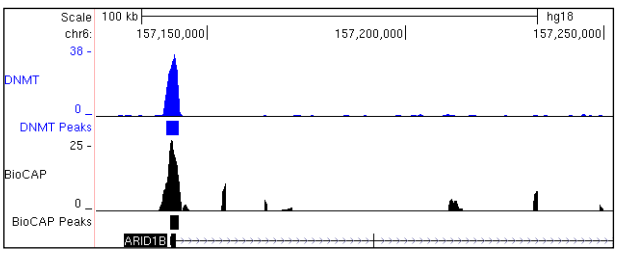

Scale
100 kb
hg18
chr6:
157,150,000
157,200,000
157,250,000
38 -
DNMT
0
DNMT Peaks
25 -
BioCAP
0
BioCAP Peaks
ARID1B

Figure 5.5 The isolated CXXC domain of MLL-FPs does not show spreading recruitment. **A)** Schematic of the 4 constructs used in 293T transfections. **B)** uCGI (Bio-CAP) landscape in SEM and 293T cells at *BCL11A* and **C)** *ARID1B*. **D)** Western blots showing expression of the 4 constructs transfected in to 293T cells. Upper left = αStrep antibody. Lower left = αGAPDH. Upper and lower right = αAF4(C). **E)** ChIP-qPCR showing binding of the 4 constructs at the uCGIs identified in (B) and (C), relative to a negative control. *HoxA9* serves as a positive binding control. **F)** Bio-CAP-seq and ChIP-seq of DNMT-SII in a stably expressing clone generated from 293T cells, showing the *ARID1B* locus as an example of the constructs inability to spread into the gene body.

5.6. Spreading MLL-AF4 Overlaps with Other Known Broad Domains

Other broad features have recently been described that correlate with gene regulatory pathways involved in cell identity and disease. Expansive regions of enhancer histone marks (H3K4me1 and H3K27Ac) along with BRD4 and members of the Mediator complex have been classified as super enhancers and mark genes believed to be important in both determining cell lineage and driving disease. Furthermore, genes associated with super enhancers showed increased expression compared to those associated with normal enhancers (Lovén et al., 2013; Whyte et al., 2013). Broad regions of H3K4me3 were shown to be a distinct group from super enhancers but similarly marked genes associated with cell identity, being better able to distinguish between closely associated cell types (Benayoun et al., 2014). It was not investigated whether this feature marks genes important in disease. Unlike super enhancers, genes associated with broad H3K4me3 showed an increase in transcriptional consistency between replicates but not an increase in expression.

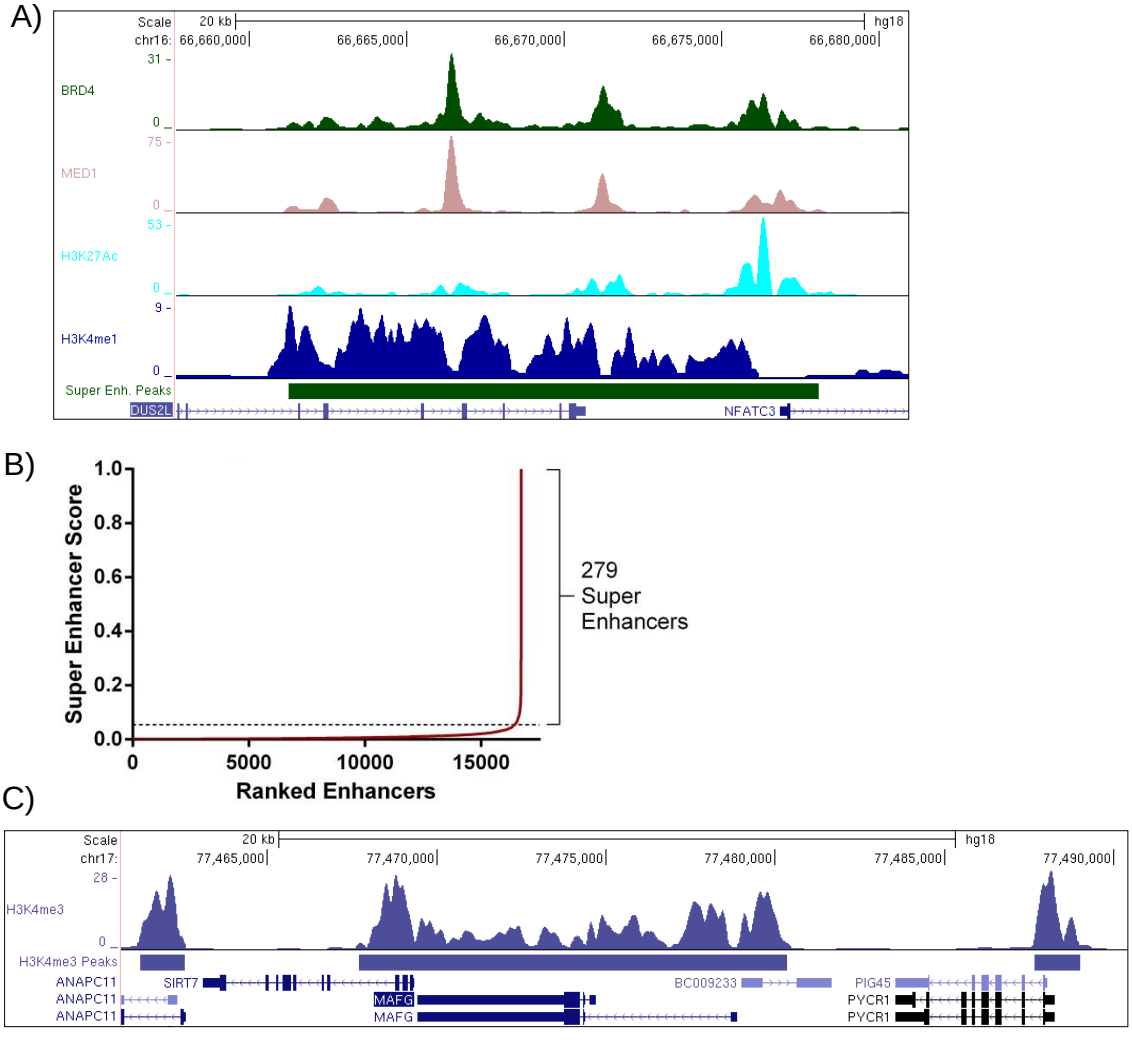
The study that initially identified broad regions of MLL-AF4 showed that these region mark genes associated with maintaining a stem cell-like, undifferentiated state but at this time super enhancers and broad H3K4me3 had not been characterised (Guenther et al.,

2008). However, a correlation with broad MLL-AF4 and H3K4me3 marks was commented on. To determine if spreading MLL-FPs correlated with either of these broad features, super enhancers and broad H3K4me3 domains were identified in SEM cells in the same manner as the original studies, making use of BRD4 and MED1 ChIP-seq available in the Milne lab. In total, 239 super enhancers (Figure 5.6A; B) and 527 broad (top 5%) H3K4me3 domains (Figure 5.6C; D) were found. An overlap with spreading MLL-AF4 revealed that this group marked a subset of the broad H3K4me3 group but was distinct from the super enhancer group (Figure 5.6E). However, the correlation between super enhancer marks and spreading MLL-AF4 is striking, with the only difference appearing to be H3K4me3 in place of H3K4me1 (Figure 5.6F; compare with Figure 5.3A).

It was previously shown that spreading MLL-AF4 targets showed increased expression compared to non-spreading targets (Figure 5.2A) but it was unknown if they showed increased transcriptional consistency, similar to broad H3K4me3-associated genes. The original study on broad H3K4me3 domains (Benayoun et al., 2014) determined transcriptional consistency by analysing the differences in read counts between biological replicates of RNA-seq experiments. Gene targets of the broadest H3K4me3 domains showed the least variation between replicates regardless of whether whole cell or nascent RNA was analysed. The authors showed that increased transcriptional consistency correlated with increased paused RNAPII and transcription elongation, two factors previously used to explain increased transcriptional consistency (Boettiger et al., 2011; Lagha et al., 2013). Similarly, for the purposes of this DPhil, the variation in read counts between biological replicates of nascent RNA-seq in SEM cells (performed by E. Ballabio) was determined for spreading MLL-AF4 targets. There was no increase in transcriptional consistency for spreading MLL-AF4 targets compared to non-spreading (Figure 5.6G). However, MLL-AF4 targets as a whole showed less variation in transcription level compared to the group of broad H3K4me3 genes. This suggests that

a more stringent level of control is required for the level of expression of MLL-AF4 targets. It may be that precise levels of downstream targets are crucial to the functioning and maintenance of the leukaemia. Thus, whilst spreading MLL-AF4 genes are distinct from super enhancers due to differences in H3K4me3 and H3K4me1 occupancy, their similarities in terms of increased expression and overlap with BRD4 and MED1 suggests that they are half-way between the two previously characterised broad features. The co-occurrence with BRD4 with spreading MLL-AF4 potentially links to the observation made previously about ENL at these loci. BRD4 is thought to recruit MLL-FPs to genes through interaction with the SEC and PAFc (Dawson et al., 2011). Since ENL is a member of the SEC, a potential explanation for spreading recruitment is an SEC-BRD4-mediated binding of MLL-FP to broad regions of H3K27ac (Figure 5.6F).

Figure 5.6



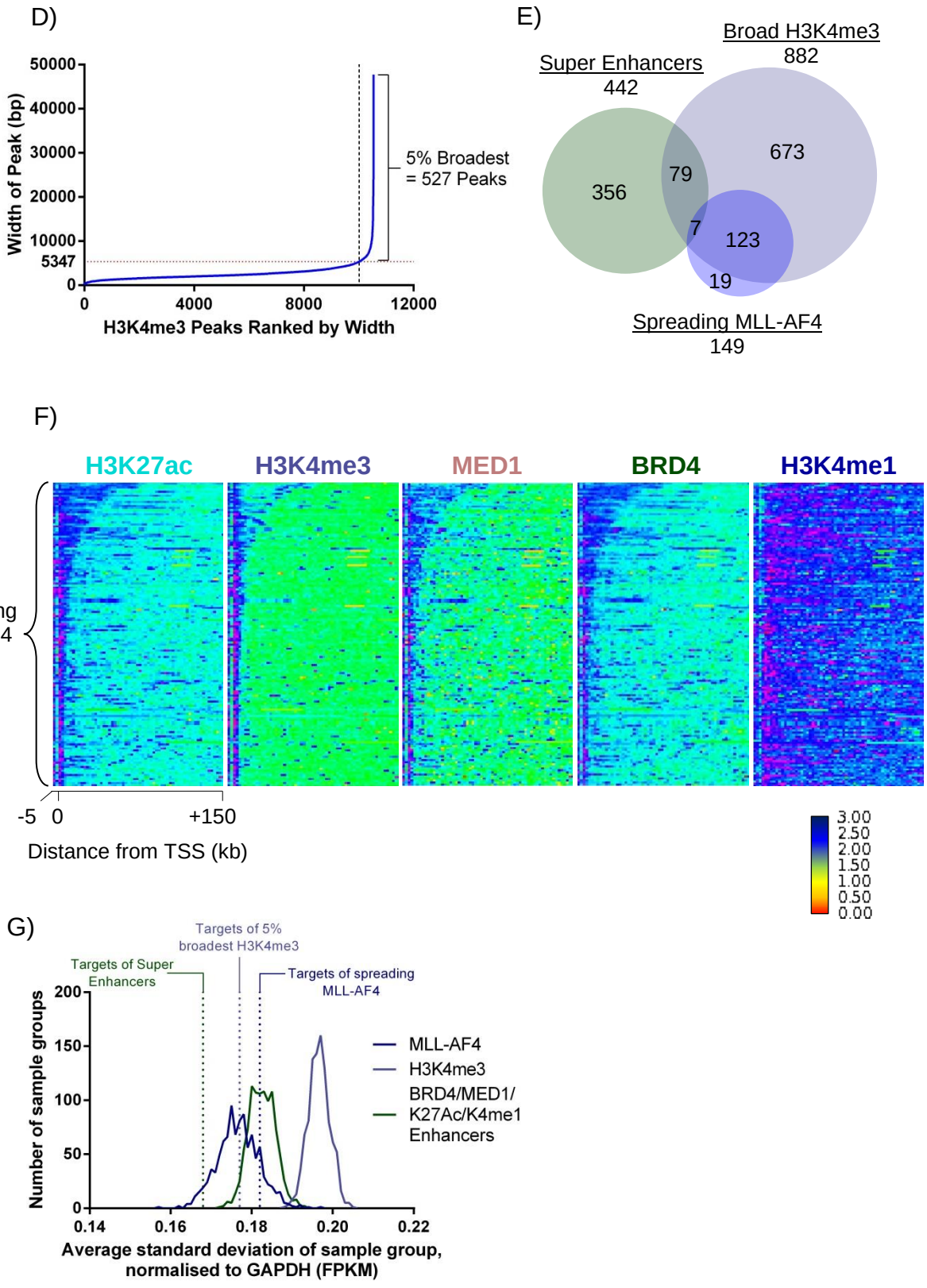

D)
Width of Peak (bp)
50000
40000
30000
20000
10000
5347
0
0
4000
8000
12000
H3K4me3 Peaks Ranked by Width
5% Broadest
= 527 Peaks
E)
Super Enhancers
442
Broad H3K4me3
882
356
79
673
7
123
19
Spreading MLL-AF4
149
F)
H3K27ac
H3K4me3
MED1
BRD4
H3K4me1
Spreading
MLL-AF4
Peaks
-5
0
+150
Distance from TSS (kb)
3.00
2.50
2.00
1.50
1.00
0.50
0.00
G)
Targets of 5%
broadest H3K4me3
Targets of Super
Enhancers
Targets of spreading
MLL-AF4
Number of sample groups
200
150
100
50
0
0.14
0.16
0.18
0.20
0.22
Average standard deviation of sample group,
normalised to GAPDH (FPKM)
MLL-AF4
H3K4me3
BRD4/MED1/
K27Ac/K4me1
Enhancers

Figure 5.6 Spreading MLL-FP targets are a subset of genes marked by broad H3K4me3. A) Example ChIP-seq track showing marks of a super enhancer spanning the *DUS2L* and *NFATC3* genes (BRD4 ChIP done by E. Ballabio; ENL ChIP done by T. Milne). **B)** Ranked enhancers in SEM cells showing the selection of super enhancers. **C)** Example ChIP-seq track showing a broad H3K4me3 domain spanning *SIRT7* and *MAFG* genes. The region is flanked by two normal promoter-bound H3K4me3 marks at *ANAPC11* and *PYCR1*. **D)** H3K4me3 peaks in SEM cells, ranked by length, showing the selection of the 5% broadest domains. **E)** Overlap of gene targets associated with super enhancers, broad H3K4me3 domains and spreading MLL-AF4. **F)** Heat-maps showing super enhancer and H3K4me3 marks at all 149 spreading MLL-AF4 peaks. Scale bar represents tags per base-pair per 10 million. **G)** Transcriptional consistency of genes associated with enhancers, H3K4me3 and MLL-AF4, with the broad domains associated with each group highlighted. Data is from 4 biological replicates of nascent RNA-seq, performed by E. Ballabio.

5.7. Common Targets of Spreading MLL-FPs are Associated with Poor Prognosis in MLL-r Leukaemia Patients

The presence and potential importance of spreading in MLL-AF4 leukaemias has been shown (Figure 5.1). To determine if spreading is important in other MLL-r leukaemias, ChIP-sequencing of MLL(N) in various MLL-FP cell lines was investigated. Spreading of MLL(N) ChIP-seq was a common feature observed in 4 other MLL-r cell lines – MV4;11 (MLL-AF4), THP-1 (MLL-AF9), KOPN-8 (MLL-ENL) and SHI-1 (MLL-AF6), but not in the non-MLL leukaemia cell lines RCH-ACV and CCRF-CEM (Figure 5.7A). Gene targets of spreading fusion proteins in these four MLL-r cell lines were characterised. As with ML-2 cells, MLL-AF6 showed very few sites of spreading and hence a minimal gene target list in SHI-1 cells. As a result an overlap was done between spreading gene targets of MLL-AF4 (SEM and MV4;11), MLL-AF9 and MLL-ENL. This revealed 14 gene targets that were common to all four spreading MLL-FPs (Figure 5.7B). Closer inspection of this list revealed that of the 9 genes that had been tested, 7 were upregulated in clinical cases of MLL leukaemias, with 5 targets correlating with a poor prognosis. This suggests that spreading is an important mechanism across the MLL-r spectrum and occurs at a subset

of genes that may be critical for leukaemia growth. By simply identifying targets of spreading MLL-FP it is likely we can select for a core set of driver genes.

5.8. Discussion

Broad domains of MLL-AF4 have previously been noted (Guenther et al. 2008) but the significance of this feature in disease has never been investigated. By characterising spreading MLL-AF4 in SEM cells, here it has been shown that its gene targets have increased expression compared to non-spreading MLL-AF4 and correlate with reduced OS in ALL patients. An important role for the Menin/LEDGF complex or factors interacting with the fusion partner, such as ENL (SEC) and BRD4, may be to facilitate spreading between regions of uCGIs. Indeed, the striking similarity between the uCGI landscape and the spreading peaks suggests that the CXXC domain is central to this process but the presence of non-CGI regions hints at other factors being involved. Importantly, wild-type MLL1 did not show a similar spreading mechanism in either MLL-r or non-MLL-r cell lines. In fitting with suggestions here, MLL1 does not interact directly with the SEC, and despite being able to interact with the Menin/LEDGF complex the two have different gene targets and this complex is seemingly unimportant for normal B-cell development (Li et al., 2013). Furthermore, two other CXXC proteins, KDM2B and CFP1, do not show spreading and do not interact with either the SEC or Menin/LEDGF complex. This further reinforces the idea that these complexes may play a role in spreading of the fusion protein.

By using truncated versions of the fusion protein it was shown that domains surrounding the CXXC domain can increase binding affinity for uCGIs in the gene body, potentially with an accumulative effect for the full-length protein. This approach is limited in that it does not confirm spreading per se but difficulties with generating stable cells lines make this hard to show definitively. Importantly, spreading was seen to be a common feature

between MLL-r leukaemia cell lines, marking genes that show overexpression and correlate with a poor prognosis in MLL-r leukaemia patients.

Figure 5.7

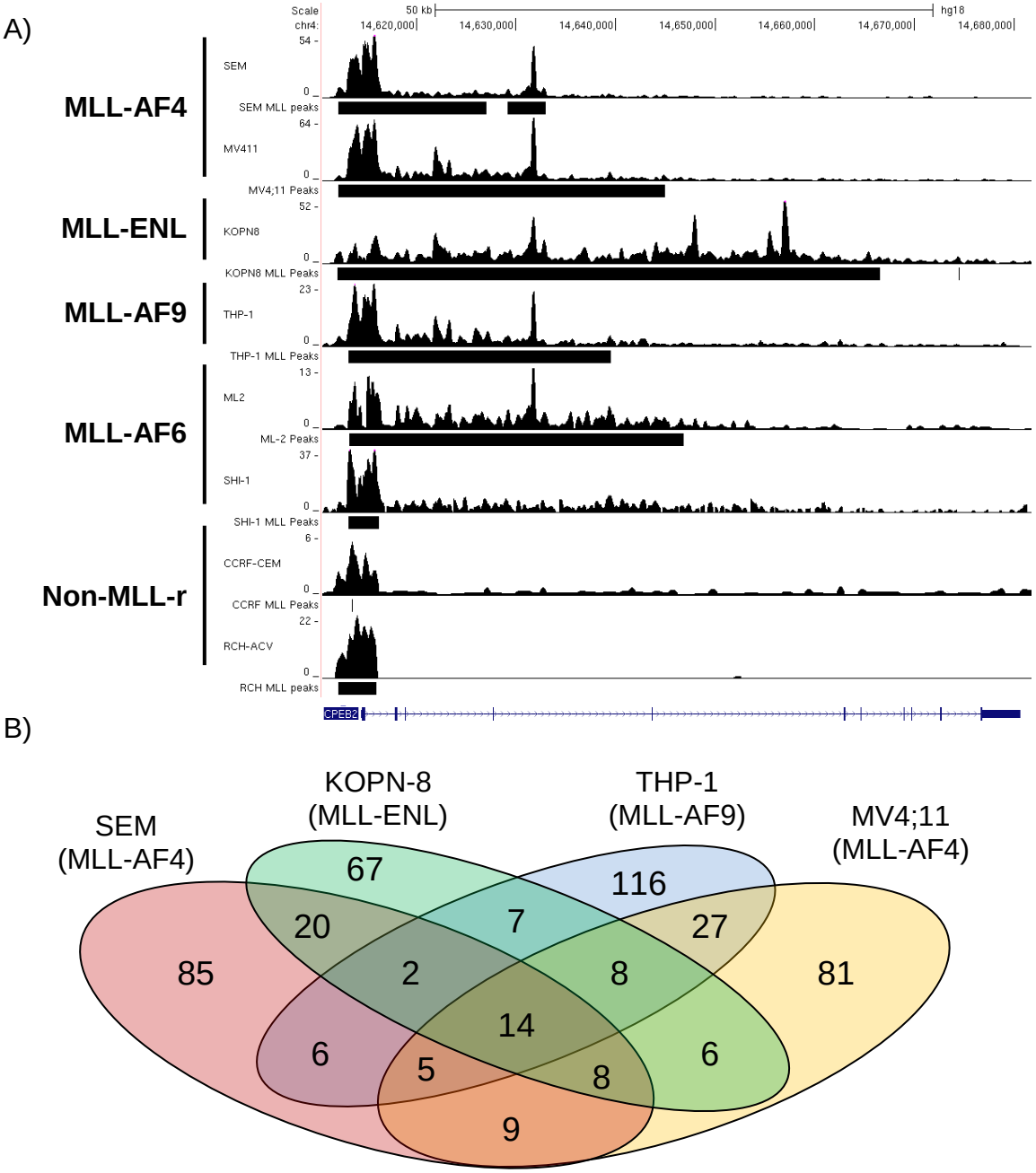


Figure 5.7 Spreading recruitment is common amongst MLL-FPs. A) Example ChIP-seq track of MLL(N) in various MLL-r and non-MLL-r leukaemia cell lines, showing spreading at *CPEB2* (MLL(N) ChIP in RCH-ACV done by T. Milne; all other ChIP done by E. Ballabio). **B)** Overlap of gene targets associated with spreading MLL-AF4, MLL-AF9 and MLL-ENL in 4 different MLL-r leukaemia cell lines.

6. Chapter Six - General Discussion

The predominance of MLL rearrangements among infant ALLs along with their aggressive nature has resulted in the development of many strategies in an attempt to treat these leukaemias (Chessells et al., 2002; Pieters et al., 2007). Stem cell transplantation, chemotherapy, immunotherapy and, more recently, targeted therapies have been used with varying degrees of success (Koh et al., 2015; Pieters et al., 1998; Stam et al., 2003). Despite this, there is still much progress to be made towards achieving higher overall survival in patients. An important factor towards this goal is an understanding of the molecular mechanisms behind MLL-FP-mediated leukaemogenesis. Since recruitment of MLL-FPs to gene targets links the fusion protein to aberrant gene expression responsible for leukaemia development, this D.Phil. investigated current hypotheses of MLL-FP recruitment, on a genome-wide level.

This study began by characterising the global binding profiles of two MLL-FPs in human MLL-r leukaemia cell lines and determined their binding sites alongside the binding sites of interacting proteins and chromatin marks. As has been previously reported for MLL-FPs (Guenther et al., 2008; Okuda et al., 2014), all MLL-AF4 and MLL-AF6 binding correlated with uCGIs marked by active histone marks (Figure 3.2; 3.4). This indicates a role for the CXXC domain at all sites of MLL-FP recruitment. Importantly, data presented here was the first to characterise the degree of MLL-AF4 and MLL-AF6 binding sites that overlap with active uCGIs instead of simply showing a correlation. MLL-FPs did not bind at sites marked by the repressive histone mark H3K27me3 (Figure 3.4A-C). This pattern of binding is consistent with a sampling-based recruitment model whereby recruitment only occurs at uCGIs marked by active chromatin and absent of repressive chromatin. Since polycomb proteins are also recruited to uCGIs it suggests that there may be competition between the fusion protein and polycomb complexes for binding at uCGIs.

This study was the first to look at the genome-wide binding profiles of Menin and LEDGF in MLL-r leukaemia cell lines (Figure 3.5). Both MLL1 and MLL-FPs can interact with Menin in an identical manner (Yokoyama et al., 2005). However, whereas Menin has been reported to have independent gene targets and functions from wild-type MLL1 (Li et al., 2013), data from this D.Phil. reveals that the majority of Menin binding co-localises with binding of the fusion protein. This data is in line with previous studies that showed co-localisation between Menin, LEDGF and MLL-FP binding at select loci and shows that the Menin/LEDGF complex is a legitimate candidate for MLL-FP recruitment (Okuda et al., 2014; Thiel et al., 2010; Yokoyama and Cleary, 2008). ChIP-seq of PAFc in MLL-r leukaemia cell lines showed that binding of PAFc only occurs at approximately 50% of MLL-FP binding sites (Figure 3.6). This argues against a role for PAFc in genome-wide MLL-FP recruitment but PAFc might be required for recruitment of MLL-FPs to specific gene targets. Binding of PAFc was observed at many genes due to its role in transcription. Therefore, the overlap between MLL-FP binding and PAFc binding could be a result of MLL-FPs and PAFc being independently recruited to active genes. The difference between PAFc binding and Menin/LEDGF binding was that the Menin/LEDGF complex had few binding events beyond those that occurred at MLL-FP binding sites.

A Tet-operator (TetO)-based recruitment system was used to determine if the interactions of MLL-AF4 with either PAFc or the Menin/LEDGF complex conferred recruitment *in vivo*. This novel approach allowed for isolated recruitment of MLL-AF4, Menin and PAF1. Whereas MLL-AF4 and Menin could recruit each other to the TetO-array (Figure 4.3), PAF1 was not able to recruit MLL-AF4 to this site, nor was MLL-AF4 able to recruit PAF1 (Figure 4.5). Not only does this rule out the interaction between PAF1 and MLL-AF4 as being involved in recruitment of the fusion protein, it also conflicts with the current understanding of the interaction between Menin and MLL-FPs. Until now, it has been hypothesised that the Menin/LEDGF complex recruits MLL-FPs (Okuda et al., 2014; Yokoyama and Cleary, 2008) but this study shows that the reverse

is also possible, especially since the CXXC domain of the fusion protein alone is sufficient for binding to uCGIs (Figure 4.2). These findings build on a previous study where induction of MLL-ENL resulted in increased binding of Menin to target genes (Caslini et al., 2007). Furthermore, MLL-AF4 recruitment was shown to occur in the absence of LEDGF as well as in the presence of an inhibitor that abolished the Menin-MLL-AF4 interaction (Figure 4.4). The absence of H3K36me3 at sites of MLL-FP recruitment (Figure 3.4) suggests that the Menin/LEDGF complex is not being recruited through binding of this mark as previously proposed (Okuda et al., 2014; Yokoyama and Cleary, 2008). Since MLL-AF4 was shown to recruit Menin to TetO (Figure 4.3), the Menin/LEDGF complex could be recruited to its gene targets by the fusion protein instead of binding to H3K36me3. However, a role for H3K36me2-mediated recruitment or direct binding to DNA cannot be ruled out (Eidahl et al., 2013; Okuda et al., 2014).

It was previously reported that broad domains of MLL-AF4 binding exist at a handful of gene targets (Guenther et al., 2008). ChIP-seq in this D.Phil. showed that this is not unique to MLL-AF4 but is also seen for MLL-ENL, MLL-AF9 and MLL-AF6, a mechanism that has been termed “spreading” here (Figure 5.7). These spreading recruitment domains were identified as being distinct from super enhancers but a subset of previously characterised broad H3K4me3 domains (Figure 5.6) (Benayoun et al., 2014). However, spreading recruitment domains show histone mark and protein occupancy similar to super enhancers, with H3K4me3 instead of H3K4me1. Gene targets of spreading MLL-AF4 correlated with increased expression, increased levels of H3K79me2 and reduced overall survival of MLL-r ALL patients (Figure 5.2), suggesting that these genes play an important role in leukaemia. Further analyses of these spreading peaks in SEM and ML-2 cells showed that they correlate strongly with uCGIs, and revealed an identical spread and increased enrichment of Menin and LEDGF at sites of spreading MLL-FPs (Figure 5.1; 5.3). Furthermore, ENL (a member of the SEC), and BRD4 also showed an overlap with spreading MLL-FP, similar to Menin and LEDGF

(Figure 5.3; 5.6). Truncated mutations of MLL-AF4 revealed that the isolated CXXC domain of MLL-FPs showed no evidence of spreading, whereas addition of the N- or C-terminus of the fusion protein to the CXXC domain increased MLL-FP binding in the gene body indicative of spreading (Figure 5.5). Full-length MLL-AF4 exhibited higher levels of binding in the gene body than either the N- or C-terminal constructs, implicating a synergistic contribution of both halves towards spreading. This suggests that N-terminal interacting proteins (i.e. the Menin/LEDGF complex) or proteins that interact with the fusion partner, such as ENL or BRD4, might facilitate spreading. This possibly explains the increased levels of these proteins at sites of spreading MLL-FPs and potentially implies a new role for them in MLL-FP recruitment. These data have led to a proposed MLL-FP recruitment model whereby spreading across CXXC-mediated binding events is facilitated by the Menin/LEDGF complex and/or proteins within the SEC (Figure 6.1A). This hypothesis was reinforced by observations that CXXC domain-containing proteins that do not interact with these complexes do not show spreading. MLL-FP spreading did not correlate with RNAPII or PAFc suggesting it is not an artefact of transcription.

Previously, MLL-AF4 spreading targets were shown to occur at genes of stem cell identity and self-renewal, implicating them as important for maintaining an undifferentiated leukaemic state (Guenther et al., 2008). In this D.Phil., 14 common gene targets of spreading MLL-AF4, MLL-AF9 and MLL-ENL were identified, in 4 different MLL-r cell lines (Figure 5.7). The majority of these gene targets have been shown to correlate with poor prognosis in MLL-r leukaemia patients, and gene targets of spreading MLL-AF4 correlated with low overall survival (Figure 5.2C). Therefore, gene targets of spreading MLL-FPs may contribute to MLL-r leukaemia progression and outcome, highlighting the mechanism of spreading recruitment as a possible therapeutic target. This would suggest a role in disease similar to that of super enhancers which were shown to mark genes important for disease progression (Lovén et al., 2013). If the

Menin/LEDGF complex or BRD4/SEC are crucial for spreading recruitment it explains the recent findings that have shown good inhibition of leukaemic growth with small molecule inhibitors against these proteins (Borkin et al., 2015; Dawson et al., 2011; Grembecka et al., 2010; Grembecka et al., 2012; He et al., 2015; Zuber et al., 2011). Since spreading appears to be unique to MLL-FPs, it could explain the ability of these drugs to be particularly potent in MLL-r leukaemic cells.

Higher levels of H3K79me2 at spreading MLL-FP gene targets compared to non-spreading MLL-FP gene targets (Figure 5.2B) and the well-known association between MLL-FPs and DOT1L implies that the fusion protein causes the increase in H3K79me2 at spreading target genes. Furthermore, domains of H3K79me2 extend further into the gene body at spreading MLL-FP gene targets compared to non-spreading MLL-FP gene targets, mirroring the spread of fusion protein binding (Figure 3.4G). It will be difficult to determine whether spreading of the fusion protein causes the increased gene expression and broad domains of active histone marks seen at spreading MLL-FP gene targets, or whether they are a consequence of spreading. If spreading MLL-FPs cause these effects it is possible to imagine that MLL-FP binding in the gene body protects the CGIs from methylation as has been proposed previously (Cierpicki et al., 2010; Erfurth et al., 2008). However, the inability of CFP1 and KDM2B to bind at uCGIs where spreading MLL-FPs are recruited (Figure 5.1) argues against CGI protection as a mechanism for establishing the observed chromatin states in the gene body. It is possible that these uCGIs already have an active chromatin state of H3K4me3 and H3K27ac. Data here suggests that actively-marked uCGIs are a pre-requisite for MLL-FP recruitment. The higher binding affinity of the MLL-FP CXXC domain for uCpGs compared to the CXXC domains of CFP1 and KDM2B (Risner et al., 2013) might explain why MLL-FPs bind at these uCGIs whereas CFP1 and KDM2B do not bind at these sites.

Two possible mechanisms of spreading MLL-FP recruitment are discussed below.

1. MLL-FPs may bind at promoter uCGIs with high affinity and spread across neighbouring uCGIs either via the Menin/LEDGF complex binding chromatin marks in the gene body or via travelling with the SEC during transcriptional elongation (Figure 6.1A). This would explain why the isolated MLL-FP CXXC domain could not bind at gene body uCGIs but the addition of SEC- or Menin/LEDGF-interaction domains conferred binding in the gene body. Furthermore, an inhibitor that disrupted the Menin-MLL-AF4 interaction caused a greater reduction in MLL-AF4 binding at gene body uCGIs compared to promoter uCGIs (Figure). However, the absence of the LEDGF substrate H3K36me3 at sites of spreading MLL-FP and no correlation with RNAPII or PAFc would argue against Menin/LEDGF- or elongation-mediated spreading.

2. The CXXC domain of the MLL-FP or the presence of BRD4 at the actively-marked uCGIs under MLL-FP spreading peaks, or a combination of both, may recruit the fusion protein to these uCGIs (Figure 6.1B). If the BRD4-SEC interaction is important in this mechanism it would explain the increased binding of the CXXC-AF4 construct at gene body uCGIs compared to the isolated CXXC domain of the fusion protein. Furthermore, the lack of the isolated CXXC domain at gene body uCGIs suggests that other stabilising factors, possibly the Menin/LEDGF complex, are required. The proximity of the uCGIs in combination with MLL-FP-mediated recruitment of multiple members of the same complexes, e.g. the Menin/LEDGF complex and SEC, might link these recruitment events creating the observed broad domains.

Figure 6.1

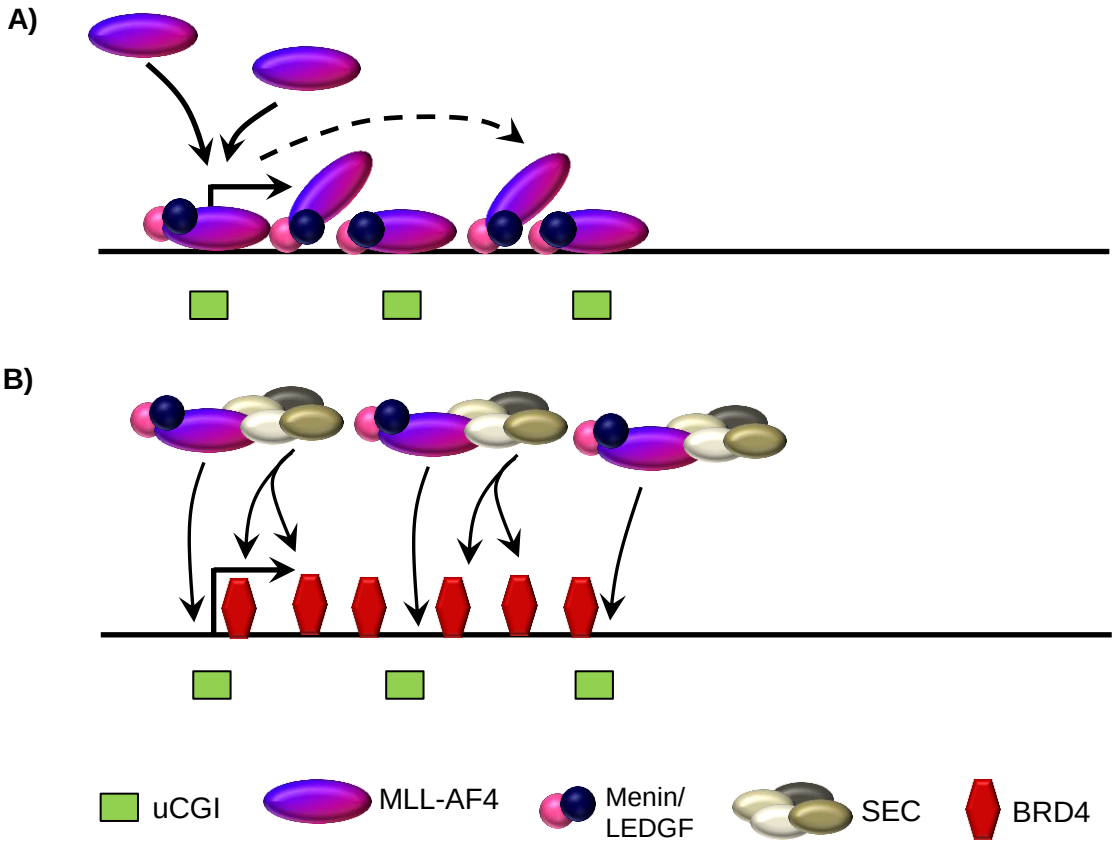


Figure 6.1 Proposed mechanisms of spreading recruitment. A) CXXC domain-mediated recruitment of MLL-AF4 (horizontal MLL-AF4, solid arrows) facilitates multiple binding events of MLL-AF4 to promoter uCGIs. Adjacent to the promoter uCGI, MLL-AF4 binding relies on stabilisation by Menin/LEDGF-mediated chromatin binding (tilted MLL-AF4). If a uCGI is in close proximity, binding can revert to CXXC domain-mediated binding allowing for increased stabilisation of the fusion protein. This generates a spreading effect along the gene body (dashed arrow) until there are no more uCGIs in close proximity. Spreading might also be mediated by interaction with elongation-associated SEC (not shown). **B)** Alternatively, MLL-AF4 is recruited independently to uCGIs and BRD4 binding sites in the gene body via the CXXC domain or SEC, respectively (solid arrows), creating a broad domain of MLL-AF4 binding that might be stabilised by the Menin/LEDGF complex.

The hypothesis that actively marked uCGIs exist prior to MLL-FP recruitment suggests that these gene targets might show increased gene activation before MLL-FP recruitment since the active uCGIs could facilitate binding of RNAPII or transcription initiation complexes. Furthermore, they might that facilitate elongation of RNAPII downstream of the promoter uCGIs. In the same way the addition of the active elongation associated H3K79me2 and SEC members in these regions, likely brought about by the MLL-FP, might further enhance RNAPII elongation, increasing gene transcription even further.

Future work to distinguish between mechanisms of spreading could include mutating CGIs under MLL-FP spreading peaks to see if recruitment is inhibited downstream of the mutated CGI or if independent binding domains arise either side of the mutation. Whilst the term “spreading” has been used here to describe the pattern of recruitment from the promoter into the gene body, the abrupt ending of these domains in the middle of genes argues more for simultaneous recruitment events across broad regions rather than a travelling mechanism. Mutations in MLL-FPs that abolish the interaction with proteins identified here as potentially important for spreading recruitment could determine which factors are involved in spreading recruitment or recruitment restricted to promoters.

Knockdowns of MLL-FPs followed by analysis of nascent transcripts will help to determine if the spreading fusion protein is responsible for the increased gene activation observed at these sites. Knockdown of the gene targets of spreading MLL-FPs, particularly the 14 common targets (Figure 5.7B), in MLL-r cell lines will show if these genes are important in leukaemia development. Correlations and survival analyses presented here point towards an important role for these gene targets in the development of MLL-r leukaemias.

References

Akalin, A., Garrett-Bakelman, F. E., Kormaksson, M., Busuttil, J., Zhang, L., Khrebtukova, I., Milne, T. A., Huang, Y., Biswas, D., Hess, J. L.*, et al.* (2012). Base-Pair Resolution DNA Methylation Sequencing Reveals Profoundly Divergent Epigenetic Landscapes in Acute Myeloid Leukemia. PLoS Genet *8*, e1002781.

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

Altaf, M., Auger, A., Monnet-Saksouk, J., Brodeur, J., Piquet, S., Cramet, M., Bouchard, N., Lacoste, N., Utley, R. T., Gaudreau, L., and Côté, J. (2010). NuA4-dependent Acetylation of Nucleosomal Histones H4 and H2A Directly Stimulates Incorporation of H2A.Z by the SWR1 Complex. Journal of Biological Chemistry *285*, 15966-15977.

Armstrong, S. A., Staunton, J. E., Silverman, L. B., Pieters, R., de Boer, M. L., Minden, M. D., Sallan, S. E., Lander, E. S., Golub, T. R., and Korsmeyer, S. J. (2002). MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. Nature Genetics *30*, 41-47.

Ayton, P. M., Chen, E. H., and Cleary, M. L. (2004). Binding to nonmethylated CpG DNA is essential for target recognition, transactivation, and myeloid transformation by an MLL oncoprotein. Molecular and Cellular Biology *24*, 10470-10478.

Ayton, P. M., and Cleary, M. L. (2003). Transformation of myeloid progenitors by MLL oncoproteins is dependent on Hoxa7 and Hoxa9. Genes & Development *17*, 2298-2307.

Bach, C., Mueller, D., Buhl, S., Garcia-Cuellar, M. P., and Slany, R. K. (2008). Alterations of the CxxC domain preclude oncogenic activation of mixed-lineage leukemia 2. Oncogene *28*, 815-823.

Baek, H. J., Kang, Y. K., and Roeder, R. G. (2006). Human Mediator Enhances Basal Transcription by Facilitating Recruitment of Transcription Factor IIB during Preinitiation Complex Assembly. Journal of Biological Chemistry *281*, 15172-15181.

Bair, E., and Tibshirani, R. (2004). Semi-Supervised Methods to Predict Patient Survival from Gene Expression Data. PLoS Biol *2*, e108.

Balgobind, B. V., Raimondi, S. C., Harbott, J., Zimmermann, M., Alonzo, T. A., Auvrignon, A., Beverloo, H. B., Chang, M., Creutzig, U., Dworzak, M. N.*, et al.* (2009). Novel prognostic subgroups in childhood 11q23/MLL-rearranged acute myeloid leukemia: results of an international retrospective study. Blood *114*, 2489-2496.

Bannister, A. J., Schneider, R., Myers, F. A., Thorne, A. W., Crane-Robinson, C., and Kouzarides, T. (2005). Spatial Distribution of Di- and Tri-methyl Lysine 36 of Histone H3 at Active Genes. Journal of Biological Chemistry *280*, 17732-17736.

Beisel, C., Buness, A., Roustan-Espinosa, I. M., Koch, B., Schmitt, S., Haas, S. A., Hild, M., Katsuyama, T., and Paro, R. (2007). Comparing active and repressed expression states of genes controlled by the Polycomb/Trithorax group proteins. Proceedings of the National Academy of Sciences *104*, 16615-16620.

Benayoun, Bérénice A., Pollina, Elizabeth A., Ucar, D., Mahmoudi, S., Karra, K., Wong, Edith D., Devarajan, K., Daugherty, Aaron C., Kundaje, Anshul B., Mancini, E.*, et al.* (2014). H3K4me3 Breadth Is Linked to Cell Identity and Transcriptional Consistency. Cell *158*, 673-688.

Berg, O. G., Winter, R. B., and Von Hippel, P. H. (1981). Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. Biochemistry *20*, 6929-6948.

Bernstein, B. E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D. K., Huebert, D. J., McMahon, S., Karlsson, E. K., Kulbokas, E. J., Gingeras, T. R.*, et al.* (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. Cell *120*, 169-181.

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K.*, et al.* (2006). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. Cell *125*, 315-326.

Bernt, Kathrin M., Zhu, N., Sinha, Amit U., Vempati, S., Faber, J., Krivtsov, Andrei V., Feng, Z., Punt, N., Daigle, A., Bullinger, L.*, et al.* (2011). MLL-Rearranged Leukemia Is Dependent on Aberrant H3K79 Methylation by DOT1L. Cancer Cell *20*, 66-78.

Bindels, E. M. J., Havermans, M., Lugthart, S., Erpelinck, C., Wocjtowicz, E., Krivtsov, A. V., Rombouts, E., Armstrong, S. A., Taskesen, E., Haanstra, J. R.*, et al.* (2012). EVI1 is critical for the pathogenesis of a subset of MLL-AF9–rearranged AMLs. Blood *119*, 5838-5849.

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

Bitoun, E., Oliver, P. L., and Davies, K. E. (2007). The mixed-lineage leukemia fusion partner AF4 stimulates RNA polymerase II transcriptional elongation and mediates coordinated chromatin remodeling. Human Molecular Genetics *16*, 92-106.

Blackledge, Neil P., Farcas, Anca M., Kondo, T., King, Hamish W., McGouran, Joanna F., Hanssen, Lars L. P., Ito, S., Cooper, S., Kondo, K., Koseki, Y.*, et al.* (2014). Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation. Cell *157*, 1445-1459.

Blackledge, N. P., Long, H. K., Zhou, J. C., Kriaucionis, S., Patient, R., and Klose, R. J. (2012). Bio-CAP: a versatile and highly sensitive technique to purify and characterise regions of non-methylated DNA. Nucleic Acids Research *40*.

Blackledge, N. P., Zhou, J. C., Tolstorukov, M. Y., Farcas, A. M., Park, P. J., and Klose, R. J. (2010). CpG Islands Recruit a Histone H3 Lysine 36 Demethylase. Molecular Cell *38*, 179-190.

Boettiger, A. N., Ralph, P. L., and Evans, S. N. (2011). Transcriptional Regulation: Effects of Promoter Proximal Pausing on Speed, Synchrony and Reliability. PLoS Comput Biol *7*, e1001136.

Borkin, D., He, S., Miao, H., Kempinska, K., Pollock, J., Chase, J., Purohit, T., Malik, B., Zhao, T., Wang, J.*, et al.* (2015). Pharmacologic Inhibition of the Menin-MLL Interaction Blocks Progression of MLL Leukemia In Vivo. Cancer Cell *27*, 589-602.

Bradley, William D., Arora, S., Busby, J., Balasubramanian, S., Gehling, Victor S., Nasveschuk, Christopher G., Vaswani, Rishi G., Yuan, C.-C., Hatton, C., Zhao, F.*, et al.* (2014). EZH2 Inhibitor Efficacy in Non-Hodgkin's Lymphoma Does Not Require Suppression of H3K27 Monomethylation. Chemistry & Biology *21*, 1463-1475.

Cao, F., Townsend, Elizabeth C., Karatas, H., Xu, J., Li, L., Lee, S., Liu, L., Chen, Y., Ouillette, P., Zhu, J.*, et al.* (2014). Targeting MLL1 H3K4 Methyltransferase Activity in Mixed-Lineage Leukemia. Molecular Cell *53*, 247-261.

Cao, R., Tsukada, Y.-i., and Zhang, Y. (2005). Role of Bmi-1 and Ring1A in H2A Ubiquitylation and Hox Gene Silencing. Molecular Cell *20*, 845-854.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R. S., and Zhang, Y. (2002). Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. Science *298*, 1039-1043.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C. A. M., Taylor, M. S., Engstrom, P. G., Frith, M. C.*, et al.* (2006). Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet *38*, 626-635.

Caslini, C., Yang, Z., El-Osta, M., Milne, T. A., Slany, R. K., and Hess, J. L. (2007). Interaction of MLL Amino Terminal Sequences with Menin Is Required for Transformation. Cancer Research *67*, 7275-7283.

Chang, M. J., Wu, H. Y., Achille, N. J., Reisenauer, M. R., Chou, C. W., Zeleznik-Le, N. J., Hemenway, C. S., and Zhang, W. Z. (2010). Histone H3 Lysine 79 Methyltransferase Dot1 Is Required for Immortalization by MLL Oncogenes. Cancer Research *70*, 10234-10242.

Chen, J., Santillan, D. A., Koonce, M., Wei, W., Luo, R., Thirman, M. J., Zeleznik-Le, N. J., and Diaz, M. O. (2008a). Loss of MLL PHD finger 3 is necessary for MLL-ENL-induced hematopoietic stem cell immortalization. Cancer Research *68*, 6199-6207.

Chen, W., Kumar, A. R., Hudson, W. A., Li, Q., Wu, B., Staggs, R. A., Lund, E. A., Sam, T. N., and Kersey, J. H. (2008b). Malignant Transformation Initiated by Mll-AF9: Gene Dosage and Critical Target Cells. Cancer Cell *13*, 432-440.

Chen, Y.-X., Yan, J., Keeshan, K., Tubbs, A. T., Wang, H., Silva, A., Brown, E. J., Hess, J. L., Pear, W. S., and Hua, X. (2006). The tumor suppressor menin regulates hematopoiesis and myeloid transformation by influencing Hox gene expression. Proceedings of the National Academy of Sciences of the United States of America *103*, 1018-1023.

Chessells, J. M., Harrison, C. J., Kempski, H., Webb, D. K. H., Wheatley, K., Hann, I. M., Stevens, R. F., Harrison, G., and Gibson, B. E. (2002). Clinical features, cytogenetics and outcome in acute lymphoblastic and myeloid leukaemia of infancy: Report from the MRC childhood leukemia working party. Leukemia *16*, 776-784.

Cierpicki, T., Risner, L. E., Grembecka, J., Lukasik, S. M., Popovic, R., Omonkowska, M., Shultis, D. D., Zeleznik-Le, N. J., and Bushweller, J. H. (2010). Structure of the MLL CXXC domain-DNA complex and its functional role in MLL-AF9 leukemia. Nat Struct Mol Biol *17*, 62-68.

Clouaire, T., Webb, S., and Bird, A. (2014). Cfp1 is required for gene expression-dependent H3K4 trimethylation and H3K9 acetylation in embryonic stem cells. Genome Biology *15*, 451.

Clouaire, T., Webb, S., Skene, P., Illingworth, R., Kerr, A., Andrews, R., Lee, J.-H., Skalnik, D., and Bird, A. (2012). Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. Genes & Development *26*, 1714-1728.

Conerly, M. L., Teves, S. S., Diolaiti, D., Ulrich, M., Eisenman, R. N., and Henikoff, S. (2010). Changes in H2A.Z occupancy and DNA methylation during B-cell lymphomagenesis. Genome Research *20*, 1383-1390.

Conter, V., Bartram, C. R., Valsecchi, M. G., Schrauder, A., Panzer-Grümayer, R., Möricke, A., Aricò, M., Zimmermann, M., Mann, G., De Rossi, G.*, et al.* (2010). Molecular response to treatment redefines all prognostic factors in children and adolescents with B-cell precursor acute lymphoblastic leukemia: results in 3184 patients of the AIEOP-BFM ALL 2000 study. Blood *115*, 3206-3214.

Core, L., Waterfall, J., and Lis, J. (2008a). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845 - 1848.

Core, L. J., Waterfall, J. J., and Lis, J. T. (2008b). Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. Science *322*, 1845-1848.

Daigle, S. R., Olhava, E. J., Therkelsen, C. A., Basavapathruni, A., Jin, L., Boriack-Sjodin, P. A., Allain, C. J., Klaus, C. R., Raimondi, A., Scott, M. P.*, et al.* (2013). Potent inhibition of DOT1L as treatment of MLL-fusion leukemia. Blood *122*, 1017-1025.

Daigle, Scott R., Olhava, Edward J., Therkelsen, Carly A., Majer, Christina R., Sneeringer, Christopher J., Song, J., Johnston, L. D., Scott, Margaret P., Smith, Jesse J., Xiao, Y.*, et al.* (2011). Selective Killing of Mixed Lineage Leukemia Cells by a Potent Small-Molecule DOT1L Inhibitor. Cancer Cell *20*, 53-65.

Daugaard, M., Baude, A., Fugger, K., Povlsen, L. K., Beck, H., Sørensen, C. S., Petersen, N. H. T., Sorensen, P. H. B., Lukas, C., Bartek, J.*, et al.* (2012). LEDGF (p75) promotes DNA-end resection and homologous recombination. Nat Struct Mol Biol *19*, 803-810.

Dawson, M. A., Gudgin, E. J., Horton, S. J., Giotopoulos, G., Meduri, E., Robson, S., Cannizzaro, E., Osaki, H., Wiese, M., Putwain, S.*, et al.* (2014). Recurrent mutations, including NPM1c, activate a BRD4-dependent core transcriptional program in acute myeloid leukemia. Leukemia *28*, 311-320.

Dawson, M. A., Prinjha, R. K., Dittmann, A., Giotopoulos, G., Bantscheff, M., Chan, W.-I., Robson, S. C., Chung, C.-w., Hopf, C., Savitski, M. M.*, et al.* (2011). Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. Nature *478*, 529-533.

de Napoles, M., Mermoud, J. E., Wakao, R., Tang, Y. A., Endoh, M., Appanah, R., Nesterova, T. B., Silva, J., Otte, A. P., Vidal, M.*, et al.* (2004). Polycomb Group Proteins Ring1A/B Link Ubiquitylation of Histone H2A to Heritable Gene Silencing and X Inactivation. Developmental Cell *7*, 663-676.

Deaton, A., and Bird, A. (2010). CpG islands and the regulation of transcription. Genes Dev *25*, 1010 - 1022.

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

Deshpande, A. J., Chen, L., Fazio, M., Sinha, A. U., Bernt, K. M., Banka, D., Dias, S., Chang, J., Olhava, E. J., Daigle, S. R*., et al.* (2013). Leukemic transformation by the MLL-AF6 fusion oncogene requires the H3K79 methyltransferase Dot1l. Blood *121*, 2533-2541.

Deshpande, Aniruddha J., Deshpande, A., Sinha, Amit U., Chen, L., Chang, J., Cihan, A., Fazio, M., Chen, C.-w., Zhu, N., Koche, R*., et al.* (2014). AF10 Regulates Progressive H3K79 Methylation and HOX Gene Expression in Diverse AML Subtypes. Cancer Cell *26*, 896-908.

Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R. Z., Ragozin, S., and Jeltsch, A. (2010). The Dnmt3a PWWP Domain Reads Histone 3 Lysine 36 Trimethylation and Guides DNA Methylation. Journal of Biological Chemistry *285*, 26114-26120.

Dobson, C. L., Warren, A. J., Pannell, R., Forster, A., and Rabbitts, T. H. (2000). Tumorigenesis in mice with a fusion of the leukaemia oncogene Mll and the bacterial lacZ gene. The EMBO Journal *19*, 843-851.

Dou, Y., Milne, T. A., Ruthenburg, A. J., Lee, S., Lee, J. W., Verdine, G. L., Allis, C. D., and Roeder, R. G. (2006). Regulation of MLL1 H3K4 methyltransferase activity by its core components. Nat Struct Mol Biol *13*, 713-719.

Dou, Y., Milne, T. A., Tackett, A. J., Smith, E. R., Fukuda, A., Wysocka, J., Allis, C. D., Chait, B. T., Hess, J. L., and Roeder, R. G. (2005). Physical Association and Coordinate Function of the H3 K4 Methyltransferase MLL1 and the H4 K16 Acetyltransferase MOF. Cell *121*, 873-885.

Dreyer, Z. E., Hilden, J. M., Jones, T. L., Devidas, M., Winick, N. J., Willman, C. L., Harvey, R. C., Chen, I. M., Behm, F. G., Pullen, J*., et al.* (2015). Intensified chemotherapy without SCT in infant ALL: Results from COG P9407 (Cohort 3). Pediatric Blood & Cancer *62*, 419-426.

Driessen, E. M. C., van Roon, E. H. J., Spijkers-Hagelstein, J. A. P., Schneider, P., de Lorenzo, P., Valsecchi, M. G., Pieters, R., and Stam, R. W. (2013). Frequencies and prognostic impact of RAS mutations in MLL-rearranged acute lymphoblastic leukemia in infants. Haematologica *98*, 937-944.

Eidahl, J. O., Crowe, B. L., North, J. A., McKee, C. J., Shkriabai, N., Feng, L., Plumb, M., Graham, R. L., Gorelick, R. J., Hess, S*., et al.* (2013). Structural basis for high-affinity binding of LEDGF PWWP to mononucleosomes. Nucleic Acids Research *41*, 3924-3936.

Elf, J., Li, G.-W., and Xie, X. S. (2007). Probing Transcription Factor Dynamics at the Single-Molecule Level in a Living Cell. Science *316*, 1191-1194.

Enderle, D., Beisel, C., Stadler, M. B., Gerstung, M., Athri, P., and Paro, R. (2011). Polycomb preferentially targets stalled promoters of coding and noncoding transcripts. Genome Research *21*, 216-226.

Erfurth, F. E., Popovic, R., Grembecka, J., Cierpicki, T., Theisler, C., Xia, Z.-B., Stuart, T., Diaz, M. O., Bushweller, J. H., and Zeleznik-Le, N. J. (2008). MLL protects CpG clusters from methylation within the Hoxa9 gene, maintaining transcript expression. Proceedings of the National Academy of Sciences *105*, 7517-7522.

Esnault, C., Ghavi-Helm, Y., Brun, S., Soutourina, J., Van Berkum, N., Boschiero, C., Holstege, F., and Werner, M. (2008). Mediator-Dependent Recruitment of TFIIH Modules in Preinitiation Complex. Molecular Cell *31*, 337-346.

Faber, J., Krivtsov, A. V., Stubbs, M. C., Wright, R., Davis, T. N., van den Heuvel-Eibrink, M., Zwaan, C. M., Kung, A. L., and Armstrong, S. A. (2009). HOXA9 is required for survival in human MLL-rearranged acute leukemias. Blood *113*, 2375-2385.

Farcas, A. M., Blackledge, N. P., Sudbery, I., Long, H. K., McGouran, J. F., Rose, N. R., Lee, S., Sims, D., Cerase, A., Sheahan, T. W.*, et al.* (2012). KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. eLife *1*.

Ferrando, A. A., Armstrong, S. A., Neuberg, D. S., Sallan, S. E., Silverman, L. B., Korsmeyer, S. J., and Look, A. T. (2003). Gene expression signatures in MLL-rearranged T-lineage and B-precursor acute leukemias: dominance of HOX dysregulation. Blood *102*, 262-268.

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

Fitzpatrick, G. V., Soloway, P. D., and Higgins, M. J. (2002). Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. Nat Genet *32*, 426-431.

Fujinaga, K., Irwin, D., Huang, Y., Taube, R., Kurosu, T., and Peterlin, B. M. (2004). Dynamics of Human Immunodeficiency Virus Transcription: P-TEFb Phosphorylates RD and Dissociates Negative Effectors from the Transactivation Response Element. Molecular and Cellular Biology *24*, 787-795.

Gallo, A., Agnese, S., Esposito, I., Galgani, M., and Avvedimento, V. E. (2010). Menin stimulates homology-directed DNA repair. FEBS Letters *584*, 4531-4536.

Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y., and Reinberg, D. (2012). PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes. Molecular Cell *45*, 344-356.

Gardiner-Garden, M., and Frommer, M. (1987). CpG Islands in vertebrate genomes. Journal of Molecular Biology *196*, 261-282.

Garriga, J., and Graña, X. (2004). Cellular control of gene expression by T-type cyclin/CDK9 complexes. Gene *337*, 15-23.

Geng, H., Brennan, S., Milne, T. A., Chen, W.-Y., Li, Y., Hurtz, C., Kweon, S.-M., Zickl, L., Shojaee, S., Neuberg, D.*, et al.* (2012). Integrative Epigenomic Analysis Identifies

Biomarkers and Therapeutic Targets in Adult B-Acute Lymphoblastic Leukemia. Cancer Discovery *2*, 1004-1023.

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

Goel, M. K., Khanna, P., and Kishore, J. (2010). Understanding survival analysis: Kaplan-Meier estimate. International Journal of Ayurveda Research *1*, 274-278.

Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A.*, et al.* (1999). Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring. Science *286*, 531-537.

Grembecka, J., Belcher, A. M., Hartley, T., and Cierpicki, T. (2010). Molecular Basis of the Mixed Lineage Leukemia-Menin Interaction: IMPLICATIONS FOR TARGETING MIXED LINEAGE LEUKEMIAS. Journal of Biological Chemistry *285*, 40690-40698.

Grembecka, J., He, S. H., Shi, A. B., Purohit, T., Muntean, A. G., Sorenson, R. J., Showalter, H. D., Murai, M. J., Belcher, A. M., Hartley, T.*, et al.* (2012). Menin-MLL inhibitors reverse oncogenic activity of MLL fusion proteins in leukemia. Nature Chemical Biology *8*, 277-284.

Guenther, M. G., Jenner, R. G., Chevalier, B., Nakamura, T., Croce, C. M., Canaani, E., and Young, R. A. (2005). Global and Hox-specific roles for the MLL1 methyltransferase. Proceedings of the National Academy of Sciences of the United States of America *102*, 8603-8608.

Guenther, M. G., Lawton, L. N., Rozovskaia, T., Frampton, G. M., Levine, S. S., Volkert, T. L., Croce, C. M., Nakamura, T., Canaani, E., and Young, R. A. (2008). Aberrant chromatin at genes encoding stem cell regulators in human mixed-lineage leukemia. Genes & Development *22*, 3403-3408.

Guo, Junjie U., Su, Y., Zhong, C., Ming, G.-l., and Song, H. (2011). Hydroxylation of 5-Methylcytosine by TET1 Promotes Active DNA Demethylation in the Adult Brain. Cell *145*, 423-434.

He, J., Shen, L., Wan, M., Taranova, O., Wu, H., and Zhang, Y. (2013). Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. Nat Cell Biol *15*, 373-384.

He, N., Liu, M., Hsu, J., Xue, Y., Chou, S., Burlingame, A., Krogan, N. J., Alber, T., and Zhou, Q. (2010). HIV-1 Tat and Host AFF4 Recruit Two Transcription Elongation Factors into a Bifunctional Complex for Coordinated Activation of HIV-1 Transcription. Molecular Cell *38*, 428-438.

He, S., Malik, B., Borkin, D., Miao, H., Shukla, S., Kempinska, K., Purohit, T., Wang, J., Chen, L., Parkin, B.*, et al.* (2015). Menin-MLL inhibitors block oncogenic transformation by MLL-fusion proteins in a fusion partner-independent manner. Leukemia.

He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L.*, et al.* (2011). Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. Science *333*, 1303-1307.

Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., Ye, Z., Lee, L. K., Stuart, R. K., Ching, C. W.*, et al.* (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108-112.

Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A.*, et al.* (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet *39*, 311-318.

Hellman, A., and Chess, A. (2007). Gene Body-Specific Methylation on the Active X Chromosome. Science *315*, 1141-1143.

Herz, H., Mohan, M., Garruss, A., Liang, K., Takahashi, Y., Mickey, K., Voets, O., Verrijzer, C., and Shilatifard, A. (2012). Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. Genes Dev *26*, 2604 - 2620.

Hilden, J. M., Dinndorf, P. A., Meerbaum, S. O., Sather, H., Villaluna, D., Heerema, N. A., McGlennen, R., Smith, F. O., Woods, W. G., Salzer, W. L.*, et al.* (2006). Analysis of prognostic factors of acute lymphoblastic leukemia in infants: report on CCG 1953 from the Children's Oncology Group. Blood *108*, 441-451.

Holstege, F. C. P., Jennings, E. G., Wyrick, J. J., Lee, T. I., Hengartner, C. J., Green, M. R., Golub, T. R., Lander, E. S., and Young, R. A. (1998). Dissecting the Regulatory Circuitry of a Eukaryotic Genome. Cell *95*, 717-728.

Hu, D., Gao, X., Morgan, M., Herz, H., Smith, E., and Shilatifard, A. (2013a). The MLL3/MLL4 branch of the COMPASS family is a major H3K4 monomethylase at enhancers. Mol Cell Biol *33*, 4745 - 4754.

Hu, D., Garruss, A. S., Gao, X., Morgan, M. A., Cook, M., Smith, E. R., and Shilatifard, A. (2013b). The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. Nat Struct Mol Biol *20*, 1093-1097.

Hughes, C. M., Rozenblatt-Rosen, O., Milne, T. A., Copeland, T. D., Levine, S. S., Lee, J. C., Hayes, D. N., Shanmugam, K. S., Bhattacharjee, A., Biondi, C. A.*, et al.* (2004). Menin Associates with a Trithorax Family Histone Methyltransferase Complex and with the Hoxc8 Locus. Molecular Cell *13*, 587-597.

Hunger, S. P., Lu, X., Devidas, M., Camitta, B. M., Gaynon, P. S., Winick, N. J., Reaman, G. H., and Carroll, W. L. (2012). Improved Survival for Children and Adolescents With Acute Lymphoblastic Leukemia Between 1990 and 2005: A Report From the Children's Oncology Group. Journal of Clinical Oncology *30*, 1663-1669.

Illingworth, R., Kerr, A., DeSousa, D., Jørgensen, H., Ellis, P., Stalker, J., Jackson, D., Clee, C., Plumb, R., Rogers, J.*, et al.* (2008). A Novel CpG Island Set Identifies Tissue-Specific Methylation at Developmental Gene Loci. PLoS Biol *6*, e22.

Illingworth, R. S., Gruenewald-Schneider, U., Webb, S., Kerr, A. R. W., James, K. D., Turner, D. J., Smith, C., Harrison, D. J., Andrews, R., and Bird, A. P. (2010). Orphan CpG Islands Identify Numerous Conserved Promoters in the Mammalian Genome. PLoS Genet *6*, e1001134.

Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C., and Zhang, Y. (2011). Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. Science *333*, 1300-1303.

Jang, M. K., Mochizuki, K., Zhou, M., Jeong, H.-S., Brady, J. N., and Ozato, K. (2005). The Bromodomain Protein Brd4 Is a Positive Regulatory Component of P-TEFb and Stimulates RNA Polymerase II-Dependent Transcription. Molecular Cell *19*, 523-534.

Jeong, K. W., Kim, K., Situ, A. J., Ulmer, T. S., An, W., and Stallcup, M. R. (2011). Recognition of enhancer element–specific histone methylation by TIP60 in transcriptional activation. Nat Struct Mol Biol *18*, 1358-1365.

Jiang, D., Kong, N. C., Gu, X., Li, Z., and He, Y. (2011a). <italic>Arabidopsis</italic> COMPASS-Like Complexes Mediate Histone H3 Lysine-4 Trimethylation to Control Floral Transition and Plant Development. PLoS Genet *7*, e1001330.

Jiang, H., Shukla, A., Wang, X., Chen, W.-y., Bernstein, B. E., and Roeder, R. G. (2011b). Role for Dpy-30 in ES Cell-Fate Specification by Regulation of H3K4 Methylation within Bivalent Domains. Cell *144*, 513-525.

Jishage, M., Malik, S., Wagner, U., Uberheide, B., Ishihama, Y., Hu, X., Chait, Brian T., Gnatt, A., Ren, B., and Roeder, Robert G. (2012). Transcriptional Regulation by Pol II(G) Involving Mediator and Competitive Interactions of Gdown1 and TFIIF with Pol II. Molecular Cell *45*, 51-63.

Johnson, K. M., and Carey, M. (2003). Assembly of a Mediator/TFIID/TFIIA Complex Bypasses the Need for an Activator. Current Biology *13*, 772-777.

Jones, B., Su, H., Bhat, A., Lei, H., Bajko, J., Hevi, S., Baltus, G. A., Kadam, S., Zhai, H., Valdez, R.*, et al.* (2008). The Histone H3K79 Methyltransferase Dot1L Is Essential for Mammalian Development and Heterochromatin Structure. PLoS Genet *4*, e1000190.

Jones, P. A., and Liang, G. (2009). Rethinking how DNA methylation patterns are maintained. Nat Rev Genet *10*, 805-811.

Jonkers, I., Kwak, H., and Lis, J. T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. eLife *3*, e02407.

Jude, C. D., Climer, L., Xu, D., Artinger, E., Fisher, J. K., and Ernst, P. (2007). Unique and Independent Roles for MLL in Adult Hematopoietic Stem Cells and Progenitors. Cell Stem Cell *1*, 324-337.

Karlić, R., Chung, H.-R., Lasserre, J., Vlahoviček, K., and Vingron, M. (2010). Histone modification levels are predictive for gene expression. Proceedings of the National Academy of Sciences *107*, 2926-2931.

Kim, H., Lee, J. E., Cho, E. J., Liu, J. O., and Youn, H. D. (2003). Menin, a tumor suppressor, represses JunD-mediated transcriptional activity by association with an mSin3A-histone deacetylase complex. Cancer Research *63*, 6135-6139.

Kim, J., Guermah, M., and Roeder, R. G. (2010). The Human PAF1 Complex Acts in Chromatin Transcription Elongation Both Independently and Cooperatively with SII/TFIIS. Cell *140*, 491-503.

Kim, J., Hake, S. B., and Roeder, R. G. (2005). The Human Homolog of Yeast BRE1 Functions as a Transcriptional Coactivator through Direct Activator Interactions. Molecular Cell *20*, 759-770.

Kim, J. B., and Sharp, P. A. (2001). Positive Transcription Elongation Factor b Phosphorylates hSPT5 and RNA Polymerase II Carboxyl-terminal Domain Independently of Cyclin-dependent Kinase-activating Kinase. Journal of Biological Chemistry *276*, 12317-12323.

Kim, S.-K., Jung, I., Lee, H., Kang, K., Kim, M., Jeong, K., Kwon, C. S., Han, Y.-M., Kim, Y. S., Kim, D., and Lee, D. (2012). Human Histone H3K79 Methyltransferase DOT1L Methyltransferase Binds Actively Transcribing RNA Polymerase II to Regulate Gene Expression. Journal of Biological Chemistry *287*, 39698-39709.

Kim, Y. S., Burns, A. L., Goldsmith, P. K., Heppner, C., Park, S. Y., Chandrasekharappa, S. C., Collins, F. S., Spiegel, A. M., and Marx, S. J. (1999). Stable overexpression of MEN1 suppresses tumorigenicity of RAS. Oncogene *18*, 5936-5942.

Klose, R. J., Cooper, S., Farcas, A. M., Blackledge, N. P., and Brockdorff, N. (2013). Chromatin Sampling—An Emerging Perspective on Targeting Polycomb Repressor Proteins. PLoS Genet *9*, e1003717.

Knutson, S. K., Kawano, S., Minoshima, Y., Warholic, N. M., Huang, K.-C., Xiao, Y., Kadowaki, T., Uesugi, M., Kuznetsov, G., Kumar, N.*, et al.* (2014). Selective Inhibition of EZH2 by EPZ-6438 Leads to Potent Antitumor Activity in EZH2-Mutant Non-Hodgkin Lymphoma. Molecular Cancer Therapeutics *13*, 842-854.

Knutson, S. K., Wigle, T. J., Warholic, N. M., Sneeringer, C. J., Allain, C. J., Klaus, C. R., Sacks, J. D., Raimondi, A., Majer, C. R., Song, J.*, et al.* (2012). A selective inhibitor of EZH2 blocks H3K27 methylation and kills mutant lymphoma cells. Nat Chem Biol *8*, 890-896.

Koh, K., Tomizawa, D., Moriya Saito, A., Watanabe, T., Miyamura, T., Hirayama, M., Takahashi, Y., Ogawa, A., Kato, K., Sugita, K.*, et al.* (2015). Early use of allogeneic hematopoietic stem cell transplantation for infants with MLL gene-rearrangement-positive acute lymphoblastic leukemia. Leukemia *29*, 290-296.

Krivtsov, A. V., Feng, Z., Lemieux, M. E., Faber, J., Vempati, S., Sinha, A. U., Xia, X., Jesneck, J., Bracken, A. P., Silverman, L. B.*, et al.* (2008). H3K79 Methylation Profiles Define Murine and Human MLL-AF4 Leukemias. Cancer Cell *14*, 355-368.

Krivtsov, A. V., Figueroa, M. E., Sinha, A. U., Stubbs, M. C., Feng, Z., Valk, P. J. M., Delwel, R., Dohner, K., Bullinger, L., Kung, A. L.*, et al.* (2013). Cell of origin determines clinically relevant subtypes of MLL-rearranged AML. Leukemia *27*, 852-860.

Krivtsov, A. V., Twomey, D., Feng, Z., Stubbs, M. C., Wang, Y., Faber, J., Levine, J. E., Wang, J., Hahn, W. C., Gilliland, D. G.*, et al.* (2006). Transformation from committed progenitor to leukaemia stem cell initiated by MLL–AF9. Nature *442*, 818-822.

Ku, M., Koche, R. P., Rheinbay, E., Mendenhall, E. M., Endoh, M., Mikkelsen, T. S., Presser, A., Nusbaum, C., Xie, X., Chi, A. S.*, et al.* (2008). Genomewide Analysis of PRC1 and PRC2 Occupancy Identifies Two Classes of Bivalent Domains. PLoS Genet *4*, e1000242.

Kuntimaddi, A., Achille, Nicholas J., Thorpe, J., Lokken, Alyson A., Singh, R., Hemenway, Charles S., Adli, M., Zeleznik-Le, Nancy J., and Bushweller, John H. (2015). Degree of Recruitment of DOT1L to MLL-AF9 Defines Level of H3K79 Di- and Tri-methylation on Target Genes and Transformation Potential. Cell Reports *11*, 808-820.

Lagha, M., Bothma, Jacques P., Esposito, E., Ng, S., Stefanik, L., Tsui, C., Johnston, J., Chen, K., Gilmour, David S., Zeitlinger, J., and Levine, Michael S. (2013). Paused Pol II Coordinates Tissue Morphogenesis in the Drosophila Embryo. Cell *153*, 976-987.

Lauberth, Shannon M., Nakayama, T., Wu, X., Ferris, Andrea L., Tang, Z., Hughes, Stephen H., and Roeder, Robert G. (2013). H3K4me3 Interactions with TAF3 Regulate Preinitiation Complex Assembly and Selective Gene Activation. Cell *152*, 1021-1036.

Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C. T., Low, H. M., Kin Sung, K. W., Rigoutsos, I., Loring, J., and Wei, C.-L. (2010). Dynamic changes in the human methylome during differentiation. Genome Research *20*, 320-331.

Lee, J.-E., Wang, C., Xu, S., Cho, Y.-W., Wang, L., Feng, X., Baldridge, A., Sartorelli, V., Zhuang, L., Peng, W., and Ge, K. (2013). H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. eLife *2*.

Lee, J.-H., Tate, C. M., You, J.-S., and Skalnik, D. G. (2007). Identification and Characterization of the Human Set1B Histone H3-Lys4 Methyltransferase Complex. Journal of Biological Chemistry *282*, 13419-13428.

Lee, J.-S., and Shilatifard, A. (2007). A site to remember: H3K36 methylation a mark for histone deacetylation. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis *618*, 130-134.

Lee, J., and Skalnik, D. (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, S., Lee, D.-K., Dou, Y., Lee, J., Lee, B., Kwak, E., Kong, Y.-Y., Lee, S.-K., Roeder, R. G., and Lee, J. W. (2006). Coactivator as a target gene specificity determinant for histone H3 lysine 4 methyltransferases. Proceedings of the National Academy of Sciences *103*, 15392-15397.

Li, B., Howe, L., Anderson, S., Yates, J. R., and Workman, J. L. (2003). The Set2 Histone Methyltransferase Functions through the Phosphorylated Carboxyl-terminal Domain of RNA Polymerase II. Journal of Biological Chemistry *278*, 8897-8903.

Li, B. E., Gan, T., Meyerson, M., Rabbitts, T. H., and Ernst, P. (2013). Distinct pathways regulated by menin and by MLL1 in hematopoietic stem cells and developing B cells. Blood *122*, 2039-2046.

Li, J., Moazed, D., and Gygi, S. P. (2002). Association of the Histone Methyltransferase Set2 with RNA Polymerase II Plays a Role in Transcription Elongation. Journal of Biological Chemistry *277*, 49383-49388.

Lin, C., Garrett, A. S., De Kumar, B., Smith, E. R., Gogol, M., Seidel, C., Krumlauf, R., and Shilatifard, A. (2011). Dynamic transcriptional events in embryonic stem cells mediated by the super elongation complex (SEC). Genes & Development *25*, 1486-1498.

Lin, C., Smith, E. R., Takahashi, H., Lai, K. C., Martin-Brown, S., Florens, L., Washburn, M. P., Conaway, J. W., Conaway, R. C., and Shilatifard, A. (2010). AFF4, a Component of the ELL/P-TEFb Elongation Complex and a Shared Subunit of MLL Chimeras, Can Link Transcription Elongation to Leukemia. Molecular Cell *37*, 429-437.

Liu, L., Jin, G., and Zhou, X. (2015). Modeling the relationship of epigenetic modifications to transcription factor binding. Nucleic Acids Research *43*, 3873-3885.

Long, Hannah K., Blackledge, Neil P., and Klose, Robert J. (2013a). ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. Biochemical Society Transactions *41*, 727-740.

Long, H. K., Sims, D., Heger, A., Blackledge, N. P., Kutter, C., Wright, M. L., Grützner, F., Odom, D. T., Patient, R., Ponting, C. P., and Klose, R. J. (2013b). Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates. eLife *2*.

Lovén, J., Hoke, Heather A., Lin, Charles Y., Lau, A., Orlando, David A., Vakoc, Christopher R., Bradner, James E., Lee, Tong I., and Young, Richard A. (2013). Selective Inhibition of Tumor Oncogenes by Disruption of Super-Enhancers. Cell *153*, 320-334.

Lund, K., Adams, P. D., and Copland, M. (2014). EZH2 in normal and malignant hematopoiesis. Leukemia *28*, 44-49.

Lynch, M. D., Smith, A. J. H., De Gobbi, M., Flenley, M., Hughes, J. R., Vernimmen, D., Ayyub, H., Sharpe, J. A., Sloane-Stanley, J. A., Sutherland, L.*, et al.* (2012). An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. EMBO J *31*, 317-329.

Mak, W., Baxter, J., Silva, J., Newall, A. E., Otte, A. P., and Brockdorff, N. (2002). Mitotically Stable Association of Polycomb Group Proteins Eed and Enx1 with the Inactive X Chromosome in Trophoblast Stem Cells. Current Biology *12*, 1016-1020.

Margueron, R., Li, G., Sarma, K., Blais, A., Zavadil, J., Woodcock, C. L., Dynlacht, B. D., and Reinberg, D. (2008). Ezh1 and Ezh2 Maintain Repressive Chromatin through Different Mechanisms. Molecular Cell *32*, 503-518.

Martin, M. E., Milne, T. A., Bloyer, S., Galoian, K., Shen, W., Gibbs, D., Brock, H. W., Slany, R., and Hess, J. L. (2003). Dimerization of MLL fusion proteins immortalizes hematopoietic cells. Cancer Cell *4*, 197-207.

Matsuo, H., Kajihara, M., Tomizawa, D., Watanabe, T., Saito, A. M., Fujimoto, J., Horibe, K., Kodama, K., Tokumasu, M., Itoh, H.*, et al.* (2014). EVI1 overexpression is a poor prognostic factor in pediatric patients with mixed lineage leukemia-AF9 rearranged acute myeloid leukemia. Haematologica *99*, e225-e227.

McCabe, M. T., Ott, H. M., Ganji, G., Korenchuk, S., Thompson, C., Van Aller, G. S., Liu, Y., Graves, A. P., Iii, A. D. P., Diaz, E.*, et al.* (2012). EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. Nature *492*, 108-112.

McMahon, K. A., Hiew, S. Y. L., Hadjur, S., Veiga-Fernandes, H., Menzel, U., Price, A. J., Kioussis, D., Williams, O., and Brady, H. J. M. (2007). Mll Has a Critical Role in Fetal and Adult Hematopoietic Stem Cell Self-Renewal. Cell Stem Cell *1*, 338-345.

Mendenhall, E. M., Koche, R. P., Truong, T., Zhou, V. W., Issac, B., Chi, A. S., Ku, M., and Bernstein, B. E. (2010). GC-Rich Sequence Elements Recruit PRC2 in Mammalian ES Cells. PLoS Genet *6*, e1001244.

Metzler, M., Forster, A., Pannell, R., Arends, M. J., Daser, A., Lobato, M. N., and Rabbitts, T. H. (2006). A conditional model of MLL-AF4 B-cell tumourigenesis using invertor technology. Oncogene *25*, 3093-3103.

Meyer, C., Hofmann, J., Burmeister, T., Groger, D., Park, T. S., Emerenciano, M., Pombo de Oliveira, M., Renneville, A., Villarese, P., Macintyre, E.*, et al.* (2013). The MLL recombinome of acute leukemias in 2013. Leukemia *27*, 2165-2176.

Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R. P.*, et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

Milne, T. A., Briggs, S. D., Brock, H. W., Martin, M. E., Gibbs, D., Allis, C. D., and Hess, J. L. (2002). MLL Targets SET Domain Methyltransferase Activity to Hox Gene Promoters. Molecular Cell *10*, 1107-1117.

Milne, T. A., Hughes, C. M., Lloyd, R., Yang, Z., Rozenblatt-Rosen, O., Dou, Y., Schnepp, R. W., Krankel, C., LiVolsi, V. A., Gibbs, D.*, et al.* (2005a). Menin and MLL cooperatively regulate expression of cyclin-dependent kinase inhibitors. Proceedings of the National Academy of Sciences of the United States of America *102*, 749-754.

Milne, T. A., Kim, J., Wang, G. G., Stadler, S. C., Basrur, V., Whitcomb, S. J., Wang, Z., Ruthenburg, A. J., Elenitoba-Johnson, K. S. J., Roeder, R. G., and Allis, C. D. (2010). Multiple Interactions Recruit MLL1 and MLL1 Fusion Proteins to the HOXA9 Locus in Leukemogenesis. Molecular Cell *38*, 853-863.

Milne, T. A., Martin, M. E., Brock, H. W., Slany, R. K., and Hess, J. L. (2005b). Leukemogenic MLL fusion proteins bind across a broad region of the Hox a9 locus, promoting transcription and multiple histone modifications. Cancer Research *65*, 11367-11374.

Mishra, Bibhu P., Zaffuto, Kristin M., Artinger, Erika L., Org, T., Mikkola, Hanna K. A., Cheng, C., Djabali, M., and Ernst, P. (2014). The Histone Methyltransferase Activity of MLL1 Is Dispensable for Hematopoiesis and Leukemogenesis. Cell Reports *7*, 1239-1247.

Mohan, M., Herz, H.-M., Takahashi, Y.-H., Lin, C., Lai, K. C., Zhang, Y., Washburn, M. P., Florens, L., and Shilatifard, A. (2010). Linking H3K79 trimethylation to Wnt signaling through a novel Dot1-containing complex (DotCom). Genes & Development *24*, 574-589.

Mohan, M., Herz, H., Smith, E., Zhang, Y., Jackson, J., Washburn, M., Florens, L., Eissenberg, J., and Shilatifard, A. (2011). The COMPASS family of H3K4 methylases in Drosophila. Mol Cell Biol *31*, 4310 - 4318.

Mohn, F., Weber, M., Rebhan, M., Roloff, T. C., Richter, J., Stadler, M. B., Bibel, M., and Schübeler, D. (2008). Lineage-Specific Polycomb Targets and De Novo DNA Methylation Define Restriction and Potential of Neuronal Progenitors. Molecular Cell *30*, 755-766.

Monroe, S. C., Jo, S. Y., Sanders, D. S., Basrur, V., Elenitoba-Johnson, K. S., Slany, R. K., and Hess, J. L. (2011). MLL-AF9 and MLL-ENL alter the dynamic association of

transcriptional regulators with genes critical for leukemia. Experimental Hematology *39*, 77-86.e75.

Mueller, C. L., Porter, S. E., Hoffman, M. G., and Jaehning, J. A. (2004). The Paf1 Complex Has Functions Independent of Actively Transcribing RNA Polymerase II. Molecular Cell *14*, 447-456.

Mueller, D., Bach, C., Zeisig, D., Garcia-Cuellar, M.-P., Monroe, S., Sreekumar, A., Zhou, R., Nesvizhskii, A., Chinnaiyan, A., Hess, J. L., and Slany, R. K. (2007). A role for the MLL fusion partner ENL in transcriptional elongation and chromatin modification. Blood *110*, 4445-4454.

Muntean, A. G., Giannola, D., Udager, A. M., and Hess, J. L. (2008). The PHD fingers of MLL block MLL fusion protein-mediated transformation. Blood *112*, 4690-4693.

Muntean, A. G., and Hess, J. L. (2012). The Pathogenesis of Mixed-Lineage Leukemia. Annual Review of Pathology: Mechanisms of Disease *7*, 283-301.

Muntean, A. G., Tan, J., Sitwala, K., Huang, Y., Bronstein, J., Connelly, J. A., Basrur, V., Elenitoba-Johnson, K. S. J., and Hess, J. L. (2010). The PAF Complex Synergizes with MLL Fusion Proteins at HOX Loci to Promote Leukemogenesis. Cancer Cell *17*, 609-621.

Musselman, C. A., and Kutateladze, T. G. (2011). Handpicking epigenetic marks with PHD fingers. Nucleic Acids Research *39*, 9061-9071.

Myers, L. C., Gustafsson, C. M., Hayashibara, K. C., Brown, P. O., and Kornberg, R. D. (1999). Mediator protein mutations that selectively abolish activated transcription. Proceedings of the National Academy of Sciences *96*, 67-72.

Nakamura, T., Mori, T., Tada, S., Krajewski, W., Rozovskaia, T., Wassell, R., Dubois, G., Mazo, A., Croce, C. M., and Canaani, E. (2002). ALL-1 Is a Histone Methyltransferase that Assembles a Supercomplex of Proteins Involved in Transcriptional Regulation. Molecular Cell *10*, 1119-1128.

Neff, T., Sinha, A. U., Kluk, M. J., Zhu, N., Khattab, M. H., Stein, L., Xie, H., Orkin, S. H., and Armstrong, S. A. (2012). Polycomb repressive complex 2 is required for MLL-AF9 leukemia. Proceedings of the National Academy of Sciences *109*, 5028-5033.

Nguyen, A. T., Taranova, O., He, J., and Zhang, Y. (2011). DOT1L, the H3K79 methyltransferase, is required for MLL-AF9–mediated leukemogenesis. Blood *117*, 6912-6922.

Nock, A., Ascano, Janice M., Barrero, Maria J., and Malik, S. (2012). Mediator-Regulated Transcription through the +1 Nucleosome. Molecular Cell *48*, 837-848.

Okada, Y., Feng, Q., Lin, Y., Jiang, Q., Li, Y., Coffield, V. M., Su, L., Xu, G., and Zhang, Y. (2005). hDOT1L Links Histone Methylation to Leukemogenesis. Cell *121*, 167-178.

Okada, Y., Jiang, Q., Lemieux, M., Jeannotte, L., Su, L., and Zhang, Y. (2006). Leukaemic transformation by CALM-AF10 involves upregulation of Hoxa5 by hDOT1L. Nat Cell Biol *8*, 1017-1024.

Okuda, H., Kawaguchi, M., Kanai, A., Matsui, H., Kawamura, T., Inaba, T., Kitabayashi, I., and Yokoyama, A. (2014). MLL fusion proteins link transcriptional coactivators to previously active CpG-rich promoters. Nucleic Acids Research *42*, 4241-4256.

Ooi, S. K. T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C. D*., et al.* (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature *448*, 714-717.

Pandey, R. R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-DiNardo, D., and Kanduri, C. (2008). Kcnq1ot1 Antisense Noncoding RNA Mediates Lineage-Specific Transcriptional Silencing through Chromatin-Level Regulation. Molecular Cell *32*, 232-246.

Payer, B., and Lee, J. T. (2008). X Chromosome Dosage Compensation: How Mammals Keep the Balance. Annual Review of Genetics *42*, 733-772.

Pieters, R., den Boer, M. L., Durian, M., Janka, G., Schmiegelow, K., Kaspers, G. J. L., van Wering, E. R., and Veerman, A. J. P. (1998). Relation between age, immunophenotype and in vitro drug resistance in 395 children with acute lymphoblastic leukemia - implications for treatment of infants. Leukemia *12*, 1344-1348.

Pieters, R., Schrappe, M., De Lorenzo, P., Hann, I., De Rossi, G., Felice, M., Hovi, L., LeBlanc, T., Szczepanski, T., Ferster, A*., et al.* (2007). A treatment protocol for infants younger than 1 year with acute lymphoblastic leukaemia (Interfant-99): an observational study and a multicentre randomised trial. The Lancet *370*, 240-250.

Pigazzi, M., Masetti, R., Bresolin, S., Beghin, A., Di Meglio, A., Gelain, S., Trentin, L., Baron, E., Giordan, M., Zangrando, A*., et al.* (2011). MLL partner genes drive distinct gene expression profiles and genomic alterations in pediatric acute myeloid leukemia: an AIEOP study. Leukemia *25*, 560-563.

Plath, K., Fang, J., Mlynarczyk-Evans, S. K., Cao, R., Worringer, K. A., Wang, H., de la Cruz, C. C., Otte, A. P., Panning, B., and Zhang, Y. (2003). Role of Histone H3 Lysine 27 Methylation in X Inactivation. Science *300*, 131-135.

Pradeepa, M. M., Sutherland, H. G., Ule, J., Grimes, G. R., and Bickmore, W. A. (2012). Psip1/Ledgf p52 Binds Methylated Histone H3K36 and Splicing Factors and Contributes to the Regulation of Alternative Splicing. PLoS Genet *8*, e1002717.

Price, D. H. (2000). P-TEFb, a Cyclin-Dependent Kinase Controlling Elongation by RNA Polymerase II. Molecular and Cellular Biology *20*, 2629-2634.

Pui, C.-H., Campana, D., Pei, D., Bowman, W. P., Sandlund, J. T., Kaste, S. C., Ribeiro, R. C., Rubnitz, J. E., Raimondi, S. C., Onciu, M*., et al.* (2009). Treating Childhood Acute Lymphoblastic Leukemia without Cranial Irradiation. New England Journal of Medicine *360*, 2730-2741.

Pui, C. H., Chessells, J. M., Camitta, B., Baruchel, A., Biondi, A., Boyett, J. M., Carroll, A., Eden, O. B., Evans, W. E., Gadner, H*., et al.* (2003). Clinical heterogeneity in childhood acute lymphoblastic leukemia with 11q23 rearrangements. Leukemia *17*, 700-706.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279-283.

Rahl, P. B., Lin, C. Y., Seila, A. C., Flynn, R. A., McCuine, S., Burge, C. B., Sharp, P. A., and Young, R. A. (2010). c-Myc Regulates Transcriptional Pause Release. Cell *141*, 432-445.

Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., Goodnough, L. H., Helms, J. A., Farnham, P. J., Segal, E., and Chang, H. Y. (2007). Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. Cell *129*, 1311-1323.

Risner, L. E., Kuntimaddi, A., Lokken, A. A., Achille, N. J., Birch, N. W., Schoenfelt, K., Bushweller, J. H., and Zeleznik-Le, N. J. (2013). Functional Specificity of CpG DNA-binding CXXC Domains in Mixed Lineage Leukemia. The Journal of Biological Chemistry *288*, 29901-29910.

Ross, M. E., Mahfouz, R., Onciu, M., Liu, H.-C., Zhou, X., Song, G., Shurtleff, S. A., Pounds, S., Cheng, C., Ma, J.*, et al.* (2004). Gene expression profiling of pediatric acute myelogenous leukemia. Blood *104*, 3679-3687.

Rubnitz, J. E., Raimondi, S. C., Tong, X., Srivastava, D. K., Razzouk, B. I., Shurtleff, S. A., Downing, J. R., Pui, C.-H., Ribeiro, R. C., and Behm, F. G. (2002). Favorable Impact of the t(9;11) in Childhood Acute Myeloid Leukemia. Journal of Clinical Oncology *20*, 2302-2309.

Schilling, E., and Rehli, M. (2007). Global, comparative analysis of tissue-specific promoter CpG methylation. Genomics *90*, 314-323.

Schmitges, Frank W., Prusty, Archana B., Faty, M., Stützer, A., Lingaraju, Gondichatnahalli M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S.*, et al.* (2011). Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. Molecular Cell *42*, 330-341.

Seila, A. C., Calabrese, J. M., Levine, S. S., Yeo, G. W., Rahl, P. B., Flynn, R. A., Young, R. A., and Sharp, P. A. (2008). Divergent Transcription from Active Promoters. Science *322*, 1849-1851.

Seizl, M., Larivière, L., Pfaffeneder, T., Wenzeck, L., and Cramer, P. (2011). Mediator head subcomplex Med11/22 contains a common helix bundle building block with a specific function in transcription initiation complex stabilization. Nucleic Acids Research *39*, 6291-6304.

Sengupta, D., Kannan, A., Kern, M., Moreno, M. A., Vural, E., Stack, B., Suen, J. Y., Tackett, A. J., and Gao, L. (2015). Disruption of BRD4 at H3K27Ac-enriched enhancer region correlates with decreased c-Myc expression in Merkel cell carcinoma. Epigenetics *10*, 460-466.

Shen, L., Kondo, Y., Guo, Y., Zhang, J., Zhang, L., Ahmed, S., Shu, J., Chen, X., Waterland, R. A., and Issa, J.-P. J. (2007). Genome-Wide Profiling of DNA Methylation Reveals a Class of Normally Methylated CpG Island Promoters. PLoS Genet *3*, e181.

Shi, J., Wang, E., Zuber, J., Rappaport, A., Taylor, M., Johns, C., Lowe, S. W., and Vakoc, C. R. (2013). The Polycomb complex PRC2 supports aberrant self-renewal in a mouse model of MLL-AF9;NrasG12D acute myeloid leukemia. Oncogene *32*, 930-938.

Shi, X., Kachirskaia, I., Walter, K. L., Kuo, J.-H. A., Lake, A., Davrazou, F., Chan, S. M., Martin, D. G. E., Fingerman, I. M., Briggs, S. D.*, et al.* (2007). Proteome-wide Analysis in

Saccharomyces cerevisiae Identifies Several PHD Fingers as Novel Direct and Selective Binding Modules of Histone H3 Methylated at Either Lysine 4 or Lysine 36. Journal of Biological Chemistry *282*, 2450-2455.

Shilatifard, A. (1998). Identification and Purification of the Holo-ELL Complex: EVIDENCE FOR THE PRESENCE OF ELL-ASSOCIATED PROTEINS THAT SUPPRESS THE TRANSCRIPTIONAL INHIBITORY ACTIVITY OF ELL. Journal of Biological Chemistry *273*, 11212-11217.

Shilatifard, A. (2012). The COMPASS Family of Histone H3K4 Methylases: Mechanisms of Regulation in Development and Disease Pathogenesis. Annual Review of Biochemistry *81*, 65-95.

Shilatifard, A., Lane, W. S., Jackson, K. W., Conaway, R. C., and Conaway, J. W. (1996). An RNA Polymerase II Elongation Factor Encoded by the Human ELL Gene. Science *271*, 1873-1876.

Shin Voo, K., Carlone, D. L., Jacobsen, B. M., Flodin, A., and Skalnik, D. G. (2000). Cloning of a Mammalian Transcriptional Activator That Binds Unmethylated CpG Motifs and Shares a CXXC Domain with DNA Methyltransferase, Human Trithorax, and Methyl-CpG Binding Domain Protein 1. Molecular and Cellular Biology *20*, 2108-2121.

Shun, M.-C., Botbol, Y., Li, X., Di Nunzio, F., Daigle, J. E., Yan, N., Lieberman, J., Lavigne, M., and Engelman, A. (2008). Identification and Characterization of PWWP Domain Residues Critical for LEDGF/p75 Chromatin Binding and Human Immunodeficiency Virus Type 1 Infectivity. Journal of Virology *82*, 11555-11567.

Silva, J., Mak, W., Zvetkova, I., Appanah, R., Nesterova, T. B., Webster, Z., Peters, A. H. F. M., Jenuwein, T., Otte, A. P., and Brockdorff, N. (2003). Establishment of Histone H3 Methylation on the Inactive X Chromosome Requires Transient Recruitment of Eed-Enx1 Polycomb Group Complexes. Developmental Cell *4*, 481-495.

Simon, Jeffrey A., and Kingston, Robert E. (2013). Occupying Chromatin: Polycomb Mechanisms for Getting to Genomic Targets, Stopping Transcriptional Traffic, and Staying Put. Molecular Cell *49*, 808-824.

Slany, R. K., Lavau, C., and Cleary, M. L. (1998). The Oncogenic Capacity of HRX-ENL Requires the Transcriptional Transactivation Activity of ENL and the DNA Binding Motifs of HRX. Molecular and Cellular Biology *18*, 122-129.

So, C. W., Karsunky, H., Wong, P., Weissman, I. L., and Cleary, M. L. (2004). Leukemic transformation of hematopoietic progenitors by MLL-GAS7 in the absence of Hoxa7 or Hoxa9. Blood *103*, 3192-3199.

Stam, R. W., den Boer, M. L., Meijerink, J. P. P., Ebus, M. E. G., Peters, G. J., Noordhuis, P., Janka-Schaub, G. E., Armstrong, S. A., Korsmeyer, S. J., and Pieters, R. (2003). Differential mRNA expression of Ara-C–metabolizing enzymes explains Ara-C sensitivity in MLL gene–rearranged infant acute lymphoblastic leukemia. Blood *101*, 1270-1276.

Stam, R. W., Schneider, P., Hagelstein, J. A. P., van der Linden, M. H., Stumpel, D., de Menezes, R. X., de Lorenzo, P., Valsecchi, M. G., and Pieters, R. (2010). Gene expression profiling-based dissection of MLL translocated and MLL germline acute lymphoblastic leukemia in infants. Blood *115*, 2835-2844.

Steger, D. J., Lefterova, M. I., Ying, L., Stonestrom, A. J., Schupp, M., Zhuo, D., Vakoc, A. L., Kim, J.-E., Chen, J., Lazar, M. A.*, et al.* (2008). DOT1L/KMT4 Recruitment and H3K79 Methylation Are Ubiquitously Coupled with Gene Transcription in Mammalian Cells. Molecular and Cellular Biology *28*, 2825-2839.

Stein, R., Razin, A., and Cedar, H. (1982). In vitro methylation of the hamster adenine phosphoribosyltransferase gene inhibits its expression in mouse L cells. Proceedings of the National Academy of Sciences of the United States of America *79*, 3418-3422.

Stumpel, D. J. P. M., Schneider, P., van Roon, E. H. J., Boer, J. M., de Lorenzo, P., Valsecchi, M. G., de Menezes, R. X., Pieters, R., and Stam, R. W. (2009). Specific promoter methylation identifies different subgroups of MLL-rearranged infant acute lymphoblastic leukemia, influences clinical outcome, and provides therapeutic options. Blood *114*, 5490-5498.

Stumpel, D. J. P. M., Schotte, D., Lange-Turenhout, E. A. M., Schneider, P., Seslija, L., de Menezes, R. X., Marquez, V. E., Pieters, R., den Boer, M. L., and Stam, R. W. (2011). Hypermethylation of specific microRNA genes in MLL-rearranged infant acute lymphoblastic leukemia: major matters at a micro scale. Leukemia *25*, 429-439.

Svejstrup, J. Q., Li, Y., Fellows, J., Gnatt, A., Bjorklund, S., and Kornberg, R. D. (1997). Evidence for a mediator cycle at the initiation of transcription. Proceedings of the National Academy of Sciences *94*, 6075-6078.

Takahashi, H., Parmely, Tari J., Sato, S., Tomomori-Sato, C., Banks, Charles A. S., Kong, Stephanie E., Szutorisz, H., Swanson, Selene K., Martin-Brown, S., Washburn, Michael P.*, et al.* (2011). Human Mediator Subunit MED26 Functions as a Docking Site for Transcription Elongation Factors. Cell *146*, 92-104.

Tanaka, S., Miyagi, S., Sashida, G., Chiba, T., Yuan, J., Mochizuki-Kashio, M., Suzuki, Y., Sugano, S., Nakaseko, C., Yokote, K.*, et al.* (2012). Ezh2 augments leukemogenicity by reinforcing differentiation blockage in acute myeloid leukemia. Blood *120*, 1107-1117.

Tavares, L., Dimitrova, E., Oxley, D., Webster, J., Poot, R., Demmers, J., Bezstarosti, K., Taylor, S., Ura, H., Koide, H.*, et al.* (2012). RYBP-PRC1 Complexes Mediate H2A Ubiquitylation at Polycomb Target Sites Independently of PRC2 and H3K27me3. Cell *148*, 664-678.

Tazi, J., and Bird, A. (1990). Alternative chromatin structure at CpG islands. Cell *60*, 909-920.

Terranova, R., Agherbi, H., Boned, A., Meresse, S., and Djabali, M. (2006). Histone and DNA methylation defects at Hox genes in mice expressing a SET domain-truncated form of Mll. Proceedings of the National Academy of Sciences of the United States of America *103*, 6629-6634.

Thiel, A. T., Blessington, P., Zou, T., Feather, D., Wu, X., Yan, J., Zhang, H., Liu, Z., Ernst, P., Koretzky, G. A., and Hua, X. (2010). MLL-AF9-Induced Leukemogenesis Requires Coexpression of the Wild-Type Mll Allele. Cancer Cell *17*, 148-159.

Thiel, A. T., Huang, J., Lei, M., and Hua, X. (2012). Menin as a hub controlling mixed lineage leukemia. BioEssays *34*, 771-780.

Thomson, J. P., Skene, P. J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A. R. W., Deaton, A., Andrews, R., James, K. D.*, et al.* (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature *464*, 1082-U1162.

Tie, F., Banerjee, R., Saiakhova, A. R., Howard, B., Monteith, K. E., Scacheri, P. C., Cosgrove, M. S., and Harte, P. J. (2014). Trithorax monomethylates histone H3K4 and interacts directly with CBP to promote H3K27 acetylation and antagonize Polycomb silencing. Development *141*, 1129-1139.

Tomizawa, D., Koh, K., Sato, T., Kinukawa, N., Morimoto, A., Isoyama, K., Kosaka, Y., Oda, T., Oda, M., Hayashi, Y.*, et al.* (2007). Outcome of risk-based therapy for infant acute lymphoblastic leukemia with or without an MLL gene rearrangement, with emphasis on late effects: a final report of two consecutive studies, MLL96 and MLL98, of the Japan Infant Leukemia Study Group. Leukemia *21*, 2258-2263.

Vakoc, C. R., Sachdeva, M. M., Wang, H., and Blobel, G. A. (2006). Profile of Histone Lysine Methylation across Transcribed Mammalian Chromatin. Molecular and Cellular Biology *26*, 9185-9195.

van Nuland, R., Smits, A. H., Pallaki, P., Jansen, P. W. T. C., Vermeulen, M., and Timmers, H. T. M. (2013). Quantitative Dissection and Stoichiometry Determination of the Human SET1/MLL Histone Methyltransferase Complexes. Molecular and Cellular Biology *33*, 2067-2077.

Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F.*, et al.* (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature *457*, 854-858.

Voigt, P., Tee, W.-W., and Reinberg, D. (2013). A double take on bivalent promoters. Genes & Development *27*, 1318-1338.

Vora, A., Goulden, N., Wade, R., Mitchell, C., Hancock, J., Hough, R., Rowntree, C., and Richards, S. (2013). Treatment reduction for children and young adults with low-risk acute lymphoblastic leukaemia defined by minimal residual disease (UKALL 2003): a randomised controlled trial. The Lancet Oncology *14*, 199-209.

Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G. A., Winston, F.*, et al.* (1998). DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. Genes & Development *12*, 343-356.

Wang, E., Kawaoka, S., Yu, M., Shi, J., Ni, T., Yang, W., Zhu, J., Roeder, R. G., and Vakoc, C. R. (2013). Histone H2B ubiquitin ligase RNF20 is required for MLL-rearranged leukemia. Proceedings of the National Academy of Sciences of the United States of America *110*, 3901-3906.

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R. S., and Zhang, Y. (2004). Role of histone H2A ubiquitination in Polycomb silencing. Nature *431*, 873-878.

Wang, P., Lin, C., Smith, E. R., Guo, H., Sanderson, B. W., Wu, M., Gogol, M., Alexander, T., Seidel, C., Wiedemann, L. M.*, et al.* (2009). Global Analysis of H3K4 Methylation Defines MLL Family Member Targets and Points to a Role for MLL1-Mediated H3K4 Methylation in the Regulation of Transcriptional Initiation by RNA Polymerase II. Molecular and Cellular Biology *29*, 6074-6085.

Wang, Q.-f., Wu, G., Mi, S., He, F., Wu, J., Dong, J., Luo, R. T., Mattison, R., Kaberlein, J. J., Prabhakar, S.*, et al.* (2011). MLL fusion proteins preferentially regulate a subset of wild-type MLL target genes in the leukemic genome. Blood *117*, 6895-6905.

Wang, Z., Song, J., Milne, T. A., Wang, G. G., Li, H., Allis, C. D., and Patel, D. J. (2010). Pro Isomerization in MLL1 PHD3-Bromo Cassette Connects H3K4me Readout to CyP33 and HDAC-Mediated Repression. Cell *141*, 1183-1194.

Weber, M., Hellmann, I., Stadler, M. B., Ramos, L., Paabo, S., Rebhan, M., and Schubeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet *39*, 457-466.

Whyte, Warren A., Orlando, David A., Hnisz, D., Abraham, Brian J., Lin, Charles Y., Kagey, Michael H., Rahl, Peter B., Lee, Tong I., and Young, Richard A. (2013). Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. Cell *153*, 307-319.

Wilkinson, Adam C., Ballabio, E., Geng, H., North, P., Tapia, M., Kerry, J., Biswas, D., Roeder, Robert G., Allis, C. D., Melnick, A.*, et al.* (2013). RUNX1 Is a Key Target in t(4;11) Leukemias that Contributes to Gene Activation through an AF4-MLL Complex Interaction. Cell Reports *3*, 116-127.

Wolf, S. F., Jolly, D. J., Lunnen, K. D., Friedmann, T., and Migeon, B. R. (1984). Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: implications for X-chromosome inactivation. Proceedings of the National Academy of Sciences of the United States of America *81*, 2806-2810.

Wu, M., Wang, P., Lee, J., Martin-Brown, S., Florens, L., Washburn, M., and Shilatifard, A. (2008). Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS. Mol Cell Biol *28*, 7337 - 7344.

Wu, X., Johansen, Jens V., and Helin, K. (2013). Fbxl10/Kdm2b Recruits Polycomb Repressive Complex 1 to CpG Islands and Regulates H2A Ubiquitylation. Molecular Cell *49*, 1134-1146.

Wysocka, J., Myers, M. P., Laherty, C. D., Eisenman, R. N., and Herr, W. (2003). Human Sin3 deacetylase and trithorax-related Set1/Ash2 histone H3-K4 methyltransferase are tethered together selectively by the cell-proliferation factor HCF-1. Genes & Development *17*, 896-911.

Wysocka, J., Swigut, T., Milne, T. A., Dou, Y., Zhang, X., Burlingame, A. L., Roeder, R. G., Brivanlou, A. H., and Allis, C. D. (2005). WDR5 Associates with Histone H3 Methylated at K4 and Is Essential for H3 K4 Methylation and Vertebrate Development. Cell *121*, 859-872.

Xia, Z.-B., Popovic, R., Chen, J., Theisler, C., Stuart, T., Santillan, D. A., Erfurth, F., Diaz, M. O., and Zeleznik-Le, N. J. (2005). The MLL fusion gene, MLL-AF4, regulates cyclin-dependent kinase inhibitor CDKN1B (p27kip1) expression. Proceedings of the National Academy of Sciences of the United States of America *102*, 14028-14033.

Xu, B., On, D. M., Ma, A., Parton, T., Konze, K. D., Pattenden, S. G., Allison, D. F., Cai, L., Rockowitz, S., Liu, S.*, et al.* (2015). Selective inhibition of EZH2 and EZH1 enzymatic activity by a small molecule suppresses MLL-rearranged leukemia. Blood *125*, 346-357.

Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T., and Handa, H. (2006). P-TEFb-Mediated Phosphorylation of hSpt5 C-Terminal Repeats Is Critical for Processive Transcription Elongation. Molecular Cell *21*, 227-237.

Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., and Handa, H. (1999). NELF, a Multisubunit Complex Containing RD, Cooperates with DSIF to Repress RNA Polymerase II Elongation. Cell *97*, 41-51.

Yang, Z., Yik, J. H. N., Chen, R., He, N., Jang, M. K., Ozato, K., and Zhou, Q. (2005). Recruitment of P-TEFb for Stimulation of Transcriptional Elongation by the Bromodomain Protein Brd4. Molecular Cell *19*, 535-545.

Yokoyama, A., and Cleary, M. L. (2008). Menin Critically Links MLL Proteins with LEDGF on Cancer-Associated Target Genes. Cancer Cell *14*, 36-46.

Yokoyama, A., Lin, M., Naresh, A., Kitabayashi, I., and Cleary, M. L. (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., Somervaille, T. C. P., Smith, K. S., Rozenblatt-Rosen, O., Meyerson, M., and Cleary, M. L. (2005). The Menin Tumor Suppressor Protein Is an Essential Oncogenic Cofactor for MLL-Associated Leukemogenesis. Cell *123*, 207-218.

Yokoyama, A., Wang, Z., Wysocka, J., Sanyal, M., Aufiero, D. J., Kitabayashi, I., Herr, W., and Cleary, M. L. (2004). Leukemia Proto-Oncoprotein MLL Forms a SET1-Like Histone Methyltransferase Complex with Menin To Regulate Hox Gene Expression. Molecular and Cellular Biology *24*, 5639-5649.

Yu, B. D., Hanson, R. D., Hess, J. L., Horning, S. E., and Korsmeyer, S. J. (1998). MLL, a mammalian trithorax-group gene, functions as a transcriptional maintenance factor in morphogenesis. Proceedings of the National Academy of Sciences of the United States of America *95*, 10632-10636.

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

Yuan, W., Xie, J., Long, C., Erdjument-Bromage, H., Ding, X., Zheng, Y., Tempst, P., Chen, S., Zhu, B., and Reinberg, D. (2009). Heterogeneous Nuclear Ribonucleoprotein L Is a Subunit of Human KMT3a/Set2 Complex Required for H3 Lys-36 Trimethylation Activity in Vivo. Journal of Biological Chemistry *284*, 15701-15707.

Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S., and Zhu, B. (2011). H3K36 Methylation Antagonizes PRC2-mediated H3K27 Methylation. Journal of Biological Chemistry *286*, 7983-7989.

Zeisig, B. B., Milne, T., Garcia-Cuellar, M. P., Schreiner, S., Martin, M. E., Fuchs, U., Borkhardt, A., Chanda, S. K., Walker, J., Soden, R.*, et al.* (2004). Hoxa9 and Meis1 are key targets for MLL-ENL-mediated cellular immortalization. Molecular and Cellular Biology *24*, 617-628.

Zhang, W., Prakash, C., Sum, C., Gong, Y., Li, Y., Kwok, J. J. T., Thiessen, N., Pettersson, S., Jones, S. J. M., Knapp, S.*, et al.* (2012a). Bromodomain-containing Protein 4 (BRD4) Regulates RNA Polymerase II Serine 2 Phosphorylation in Human CD4+ T Cells. Journal of Biological Chemistry *287*, 43137-43155.

Zhang, X., Zhao, X., Fiskus, W., Lin, J., Lwin, T., Rao, R., Zhang, Y., Chan, John C., Fu, K., Marquez, Victor E.*, et al.* (2012b). Coordinated Silencing of MYC-Mediated miR-29 by HDAC3 and EZH2 as a Therapeutic Target of Histone Modification in Aggressive B-Cell Lymphomas. Cancer Cell *22*, 506-523.

Zhao, J., Sun, B. K., Erwin, J. A., Song, J.-J., and Lee, J. T. (2008). Polycomb Proteins Targeted by a Short Repeat RNA to the Mouse X Chromosome. Science *322*, 750-756.

Zilberman, D., Coleman-Derr, D., Ballinger, T., and Henikoff, S. (2008). Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. Nature *456*, 125-129.

Zippo, A., Serafini, R., Rocchigiani, M., Pennacchini, S., Krepelova, A., and Oliviero, S. (2009). Histone Crosstalk between H3S10ph and H4K16ac Generates a Histone Code that Mediates Transcription Elongation. Cell *138*, 1122-1136.

Zuber, J., Shi, J., Wang, E., Rappaport, A. R., Herrmann, H., Sison, E. A., Magoon, D., Qi, J., Blatt, K., Wunderlich, M.*, et al.* (2011). RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. Nature *478*, 524-528.