The role of DOT1L in MLL-AF4 Leukaemia

Laura Godfrey

Linacre College

Molecular Haematology Unit

Weatherall Institute of Molecular Medicine

University of Oxford

Abstract

The role of DOT1L in MLL-AF4 leukaemia

DOT1L is a methyltransferase which has been shown to methylate H3K79 (Ng et al. 2002; Feng et al. 2002; Lacoste et al. 2002). DOT1L and H3K79me have been shown to be involved in active transcription and in particular has been shown to be important for MLL-AF4 leukaemia, where high levels of H3K79me are found at MLL-AF4 gene targets (Krivtsov et al. 2008; Bernt et al 2011). The exact mechanism of how DOT1L is recruited to MLL-AF4 gene targets leading to high levels of H3K79me is currently unknown, especially as it has been demonstrated that AF4 and DOT1L exist in mutually exclusive complexes (Leach et al. 2013; Yokoyama et al. 2004; Biswas et al. 2011). In addition to the recruitment mechanism of DOT1L, the function of H3K79me remains unclear. In this thesis, the recruitment mechanism of DOT1L at MLL-AF4 gene targets was investigated using the *in vivo* TetR-recruitment system (Blackledge et al. 2014). Using this, it was found that DOT1L complex members AF9, ENL and AF10 were sufficient for DOT1L recruitment, in addition to PAF1, a member of the PAF1 elongation complex. To investigate the function of H3K79me at MLL-AF4 gene targets, it was necessary to identify a set of gene targets which were dependent upon H3K79me for transcription. To do this, the DOT1L inhibitor, EPZ-5676, was employed to treat an MLL-AF4 leukaemia cell line (SEM) followed by Nascent RNA seq and ChIP rx seq. From this a set of hypersensitive MLL-AF4 gene targets were identified. Importantly, Capture-C, ATAC seq and ChIP seq revealed a subset of hypersensitive targets with a putative intragenic enhancer. Following EPZ-5676 treatment, Capture-C revealed a disruption in the interaction between the putative intragenic enhancer and promoter of some MLL-AF4 hypersensitive genes targets. This provides evidence for a novel, context dependent role of H3K79me which may be

involved in enhancer-promoter interactions and function at a subset of hypersensitive MLL-AF4 gene targets.

List of publications

Godfrey Laura, Jon Kerry, Ross Thorne, Emmanouela Repapi, James O J Davies, Marta Tapia, Erica Ballabio, *et al*. 'MLL-AF4 binds directly to a BCL-2 specific enhancer and modulates H3K27 acetylation' *Experimental Hematology 2017* March 47: 64–75.

Benito JM, **Godfrey L**, *et al*, 'MLL-rearranged acute lymphoblastic leukaemias activate BCL-2 though H3K79 methylation and are sensitive to the BCL-2 specific antagonist ABT-199' *Cell Reports* 2015 Dec 12:2715-2727

Kerry J, **Godfrey L**, Repapi E, Tapia M, Blackledge NP, Ma H, Ballabio E, O'Byrne S, Ponthan F, Heidenreich O, Roy A, Roberts I, Konopleva M, Klose RJ, Geng H, Milne TA. 'MLL-AF4 spreading identifies binding sites that are distinct from super-enhancers and that govern sensitivity to DOT1L inhibition in Leukaemia.' *Cell Reports* 2017 Jan 18:482-495

Acknowledgements

Firstly, I would like to express my gratitude to my supervisor Tom Milne for his guidance and teaching throughout my DPhil. Thanks go to members of the Milne lab for stimulating scientific conversations and for providing a great and supportive atmosphere to complete my DPhil. Namely, special thanks goes to I-Jun, who has been a special colleague and friend to me in the lab and to Nick for reading my thesis and taking on all of my questions over the last few months! Thanks also goes to Phill for all his help throughout my project. Sincere gratitude goes to the Medical Research Council for the funding of my DPhil project.

I would also like to express thanks to my second supervisor Irene Roberts, for reading my thesis and providing great feedback and encouragement. Great thanks go to Sorcha and Ross, for instrumental help contributing to data in Chapter 4. I am also extremely grateful to the Klose lab for providing me with the TetR system, which forms the basis of my work in Chapter 3. I'd like to acknowledge both Jelena Telenius and Emmanouela Repapi for the generation of bioinformatic pipelines and Nascent RNA seq analysis.

To my friends I have made during my time in Oxford, both at the WIMM and Linacre; Erica (my PhD twin), and the Coffee Club, Iliana, Anna, Caitlin, the rest of the Linacre family, and many more. Thank you for making my time in Oxford full of wonderful memories! To all my amazing pals outside of the Oxford bubble, thanks for always understanding and being there to reassure me along the way.

To my wonderful family; Mum, Dad and Emma. Thank you for your unwavering support, love and providing me with strength and encouragement throughout my DPhil. Last but not least, I want to thank my partner Oscar. Thank you for supporting me through all the highs and lows of my research over the last four years and always being there for a hug at the end of the day.

Table of Contents

List of Figures

List of Tables

Chapter 1 – Introduction

1.1 Overall summary and main objectives

DOT1L is a histone H3K79 methyltransferase which is linked to active gene regulation (Steger et al. 2008). Most active genes are marked with H3K79me at the 5' end of the gene body. It is unclear how DOT1L and H3K79me contribute to transcription. Importantly, we do know that DOT1L and H3K79me are pivotal in MLL rearranged (MLL-r) leukaemia (Bernt & Armstrong 2011; Krivtsov et al. 2008; Aniruddha J Deshpande et al. 2013; M.-J. Chang et al. 2010). MLL-r leukaemia arises from a chromosome translocation between MLL, on chromosome 11 and 1 of over 121 identified MLL fusion partners (Meyer et al. 2013). This generates an MLL fusion gene, which is translated into an MLL-fusion protein (MLL-FP). MLL-FPs bind to gene targets via the MLL portion of the fusion, and it is thought to recruit elongation factors to gene targets where it causes aberrant upregulation of transcription. High levels of H3K79me are observed at MLL-FP gene targets (Krivtsov et al. 2008; Bernt et al 2011; Kerry et al. 2017). Following the perturbation of DOT1L, expression of MLL-FP gene targets is downregulated and the leukaemia is abrogated (Daigle et al. 2013; Daigle et al. 2011). This suggests DOT1L plays an important role in contributing to the transcription of MLL-FP gene targets. Therefore, MLL-r leukaemia's present as an excellent model system to understand the functional role of DOT1L and H3K79me in transcription. Furthermore, MLL-r leukaemia caused by MLL-AF4, have a particularly poor prognosis (Meyer et al. 2013). Even though this is a common MLL-FP, it is under-studied compared to other MLL-FPs. Most other common MLL-FPs contain a fusion partner which can biochemically interact with DOT1L, potentially suggesting a direct

recruitment mechanism to gene targets (Mohan et al. 2010; Biswas et al. 2011; Leach et al. 2013). However, AF4 and DOT1L have not been shown to interact biochemically (Leach et al. 2013; Biswas et al. 2011;Yokoyama et al. 2010; Lin et al. 2010). Therefore, the recruitment and function of DOT1L and H3K79me at MLL-AF4 gene targets is complicated and very poorly understood.

This DPhil project aims to understand how DOT1L is recruited to MLL-AF4 gene targets and furthermore, how H3K79me contributes to transcription of MLL-AF4 gene targets which are dependent upon DOT1L function. Understanding the molecular mechanism of DOT1L at MLL-AF4 gene targets will not only improve to our understanding of how DOT1L might control transcription of MLL-AF4 gene targets, but may also provide insight into the general function of DOT1L and H3K79me in transcription. Finally, given the poor prognosis of MLL-r leukaemias, more targeted and effective treatment is urgently required and should be facilitated by a greater understanding of the molecular mechanism of MLL-AF4 leukaemia.

1.2 Basic organisation of DNA

To understand how transcription is controlled, it is important to discuss how DNA is packaged into the nucleus and how this might influence gene regulation. In order to accommodate the vast amount of DNA humans have in the nucleus of each cell, the DNA has to be packaged in the form of chromatin. The basic subunit of chromatin is a nucleosome, composed of ~145bp of DNA wrapped around an octamer of core histones comprising of doublets H2A, H2B, H3 and H4 and linked together by histone H1 (Richmond et al. 1997; Thoma et al. 1979).

Each core histone has a central globular domain and an unstructured, tail domain which extends beyond the boundary of the nucleosome globular core (Luger & Richmond 1998). Residues in

the histone core and on the tail, can be post-translationally modified. Furthermore, nucleosome remodeling and histone modifications are tightly associated with transcription regulation. Despite the classic solenoid model of chromatin high order structure, recent *in vivo* studies have emerged which demonstrate the 10nm fibre is packaged as a randomized polymer in the nucleus (Ou et al. 2017).

1.3 Spatial organisation of chromatin

Even though DNA is packaged into the nucleus as chromatin, 3D chromatin structure exists which is important for gene regulation. It has been observed that heterochromatin and euchromatin occupy different spatial regions of the nucleus (Lieberman-Aiden et al. 2009; Tamaru 2010). It has recently been suggested that HP1α promotes phase separation of heterochromatic regions, which can exclude factors associated with active genes, this may suggest a potential mechanism of separation between euchromatin and heterochromatin (Larson et al. 2017; Strom et al. 2017).

In addition to the broad separation of active and repressed chromatin, more localised structural domains of metazoan chromosomes have been observed called topological associated domains (TADs) (Lieberman-Aiden et al. 2009; Dixon et al. 2012; Hou et al. 2012; Sexton et al. 2012). TADs have been shown to contain clusters of genes and enhancer elements, delimited by boundary elements such as CTCF and Cohesin (Nichols & Corces 2015; Ghirlando & Felsenfeld 2016; Dixon et al. 2012; Hadjur et al. 2009; Mishiro et al. 2009; Nativio et al. 2009; Tang et al. 2015; Handoko et al. 2011). The interactions between enhancers and promoters within TADs have been shown to differ between cell types and at different stages of differentiation (Beagan et al. 2016; Dixon et al. 2015; Chandra et al. 2015). For example, the α-globin locus lies within a TAD that contains enhancer elements which regulate the α-globin

genes (Hay et al. 2016; Anguita et al. 2002; Davies et al. 2016; Hughes et al. 2013; Hanssen et al. 2017). The interactions between enhancers and promoters of the a-globin genes are observed in an erythroid tissue specific manner (Hughes et al. 2013).

The purpose of TADs has been primarily thought to bring widely separated functional elements, such as enhancers and promoters, into close spatial proximity to further control gene regulation as well as regulating and containing enhancer function (Hnisz et al. 2016; Dowen et al. 2014; Hanssen et al. 2017; Narendra et al. 2016; Oti et al. 2016). This could bring chromatin proteins and RNA polymerase II (RNAPII) into a local compartment where interactions can occur more frequently (Mitchell & Fraser 2008). Disruptions in TAD boundaries have been shown to causes mis-regulation of transcription (Lupiáñez et al. 2015; Amano et al. 2009; Niedermaier et al. 2005; Sagai et al. 2005; Lettice et al. 2011; Giorgio et al. 2015) . This has been identified in some cases as a driver of human disease, demonstrating the importance of spatial chromatin organization in transcription control (Flavahan et al. 2016; Franke et al. 2016; Hnisz et al. 2016).

1.4 Role of Trithorax (Trx) and Polycomb group (PcG) proteins in gene regulation

Transcription is, in part, modulated by a variety of different chromatin proteins which bind to DNA or are stabilised at a gene target by other proteins or histone modifications. The spatial organization of chromatin may create pockets whereby such chromatin proteins and transcription factors are sequestered within a TAD, increasing the tight regulation of transcription (Mitchell & Fraser 2008; Osborne et al. 2004; Osborne et al. 2007) .

A classic example of how chromatin proteins control transcription was originally observed in *Drosophila melanogaster*. Two classes of chromatin proteins important for the control of the expression patterns of homeotic genes (Hox) during development were identified (Ingham 1998; Ingham 1981; Lewis 1978). These are the Trithorax group proteins (TrxG) and Polycomb group proteins (PcG). TrxG proteins were shown to maintain active transcription, whereas, PcG proteins were shown to maintain a repressed transcription state. Many studies have demonstrated an antagonism between TrxG and PcG complexes which may modulate or respond to the active or repressed state of the gene (Tie et al. 2009; Tie et al. 2014; Schmitges et al. 2011; Yuan et al. 2011). Mutation of these proteins were shown to give rise to developmental defects, particularly in the anterior-posterior positioning indicative of mis-expression of Hox genes (Ingham 1981; Ingham 1998; Lewis 1978).

Since their discovery in *D.melanogaster*, TrxG and PcG homologue proteins have been identified in humans. SET domain containing MLL1, along with others, have been shown to be homologous to *D.melanogaster* Trx (Gu et al. 1992; Yu et al. 1998). Polycomb Repressive Complexes 1 and 2 (PRC1/PRC2) are homologous to PcG proteins (Lee et al. 2006; Boyer et al. 2006). Due to the pivotal role of TrxG and PcG proteins in maintaining proper gene regulation, it is unsurprising that mutations and mis-regulation of proteins homologous to Trx and PcG can lead to human disease including cancer (Sauvageau & Sauvageau 2010; Varambally et al. 2002; Morin et al. 2010; Comet et al. 2016; Zinzalla 2016; Kadoch & Crabtree 2015).

1.5 Histone modifications in gene regulation

MLL1 and PRC2 are known to have methyltransferase activity, being able to methylate, H3K4me3 and H3K27me3, respectively (Klymenko et al. 2006; Rea et al. 2000; Shen et al.

2008; Cao et al. 2002). Histone modifications are an example of an epigenetic modification which have been shown, in some cases, to correlate with transcription regulation (Steger et al. 2008; Pokholok et al. 2005; Rao et al. 2005). The proposed function of histone modifications is not completely understood; however, many histone modifications have been shown to act as platforms for 'reader' proteins, which bind to specific modified residues. One example of this is the binding of multiple factors, such as BPTF and TAF3 to H3K4me3 (Lauberth et al. 2013; Wysocka et al. 2006; Ruthenburg et al. 2011; Li et al. 2006). The binding of reader proteins to histone modifications is low affinity relative to some protein:protein interactions as well as DNA:protein interactions observed *in vitro*. One example of this is the binding of chromatin protein AF9 to acetylated lysine residues (K_D = 30-50μM) compared to the interaction between AF9 and AF4 (K_D =1.6nM) (Leach et al. 2013; Li et al. 2014). Nevertheless, low affinity interactions may in part contribute to a high affinity stabilisation of protein complexes which can be created by the additive effect of multivalent interactions (Ruthenburg et al. 2011). Another proposed function of histone modifications is the ability to cause changes in nucleosome structure, for example, lysine acetylation has been demonstrated to disrupt inter-nucleosome interactions via loss of positive charge (Fingerman et al. 2007; Chodaparambil et al. 2007; Shogren-Knaak et al. 2006). Although this does not rule out the possibility of novel functions of histone modifications, the role of histone modifications as platforms and regulators of chromatin accessibility may be pivotal in contributing to gene regulation.

1.6 DNA methylation and PcG/TrxG protein recruitment

DNA methylation occurs at CpG dinucleotides in the human genome and is catalysed by DNA methyltransferases (DNMTs) (Ehrlich et al. 1982; Okano et al. 1999; Holliday & Pugh 1975). Furthermore, promoters are often associated with conserved, high frequency sites of CpG

dinucleotides (CpG islands, or CGIs), which are often unmethylated (uCGIs) at active genes (Bird et al. 1985; Blackledge et al. 2010; Thomson et al. 2010; Long et al. 2013; Illingworth et al. 2010). Although DNA methylation can occur in the gene body of active genes (Hellman & Chess 2007; Ball et al. 2009), DNA methylation observed at CGIs at promoters is associated with gene repression. An examples of this is at silenced genes on the X-chromosome (Venolia & Gartler 1983).

DNA methylation has also been shown to collaborate with repressive histone modifications such as H3K9me at repressed heterochromatin (Mikkelsen et al. 2007; Smallwood et al. 2007; Espada et al. 2004; Fuks et al. 2003; Fujita et al. 2003; Sarraf & Stancheva 2004; Ikegami et al. 2007; Lehnertz et al. 2003; Karimi et al. 2011). H3K9me is generally associated with constitutive heterochromatin, although it has also been observed in repressed regions of euchromatin (Tachibana et al. 2005; Bilodeau et al. 2009). There are multiple KMTs which are responsible for H3K9 mono-, di- or tri-methylation in humans. G9a and GLP have been shown to methylate and bind to H3K9me1/2 (Collins et al. 2008; Tachibana et al. 2005). SETDB1 and SUV39H1/2 have been shown to methylate and bind to H3K9me2/3 (Bilodeau et al. 2009; Karimi et al. 2011). HP1a is a protein shown to bind H3K9me and recruit DNA methyltransferases, thus providing an example of how epigenetic modifications and chromatin proteins can collaborate to maintain the state of chromatin (Lachner et al. 2001; Fuks et al. 2003; Lehnertz et al. 2003).

70% of promoters are associated with unmethylated CpG islands which are linked with increased chromatin accessibility and are associated with gene activation (Bird et al. 1985; Blackledge et al. 2010; Saxonov et al. 2006; Long et al. 2013). Both PRC and MLL complexes contain ZF-CXXC domain proteins, such as KDM2B and MLL, respectively. Both KDM2B

and MLL can bind to CpG islands (Birke et al. 2002; Cierpicki et al. 2010; Farcas et al. 2012; He et al. 2013). As these complexes have been classically linked with maintaining either active or repressive transcription states, it is interesting to consider how this might occur. One proposed mechanism is via competitive binding of CpG islands in a chromatin sampling mechanism (Klose et al. 2013). The antagonism between activating and repressive signals at gene targets could perhaps be extrapolated when understanding how transcription is controlled more generally, whereby activating chromatin proteins and modifications potentially antagonise their repressive counterparts.

1.7 Transcription factor binding

Transcription factors are DNA binding proteins which can bind to short, specific DNA sequences primarily at enhancers and promoters (Wunderlich & Mirny 2009; Small et al. 1992; Yuh et al. 1994; Sandmann et al. 2007; Spitz & Furlong 2012). It has been demonstrated that transcription factors may contribute to transcription by facilitating the recruitment of Mediator and co-activator proteins such as P300, at enhancers and promoters (Johnson et al. 2002; Stumpf et al. 2006; Esnault et al. 2008; Bedford et al. 2010). Furthermore, transcription factors have been shown to be important in facilitating enhancer-promoter interactions (Nolis et al. 2009; Drissen et al. 2004; Bertolino & Singh 2002; Szutorisz et al. 2005; Vakoc et al. 2005; Levasseur et al. 2008). In contrast, some transcription factors have been demonstrated to have repressive roles in transcription (Brenner et al. 2005; Kurland & Tansey 2008; Peukert et al. 1997).

The recruitment of transcription factors to specific binding sites has been shown to occur via facilitated diffusion in *E.coli* (Elf et al. 2007; Bauer & Metzler 2012; Cartailler & Reingruber 2015). This mechanism describes a searching mechanism whereby transcription factors bind

transiently at non-specific sites and more stably at specific DNA sequences (Hammar et al. 2012). Unlike *E.coli*, human DNA is packaged into nucleosomes, meaning the mechanism of transcription factor binding may be different. For most transcription factor binding in humans, it seems that a pre-existing level of chromatin accessibility is required for the binding to specific sites (Biddie et al. 2011; Guertin & Lis 2010). Furthermore, when cells are undergoing differentiation, replication has been shown to be important for initial transcription factor binding for certain key transcription factors, following a delay in H3K27me3 accumulation and more accessible chromatin (Petruk et al. 2017). This window of opportunity may be important for differentiation, where genes normally repressed could be activated due to transcription factor binding.

Combinatorial transcription factor binding at enhancers and promoters has been shown to be important in regulating tissue specific genes (Pilon et al. 2011; Lin et al. 2010). The mechanisms of combinatorial binding are complex but briefly, they may occur via direct transcription factor:transcription factor binding at DNA or by indirect DNA co-occupancy (Jolma et al. 2013; Walhout 2006; Arnosti et al. 1996; Wilson et al. 2010; Voss et al. 2011). Transcription factors that can promote the binding of other transcription factors, by creating a more accessible chromatin state, are termed 'pioneer factors' and are often associated with the control of important tissue specific genes (Adam et al. 2015; Iwafuchi-Doi & Zaret 2016).

Some transcription factors have been demonstrated to compete with nucleosomes for the binding of DNA, one example is FOXA1, which has been shown *in vitro* and *in vivo* to bind DNA via competition with nucleosomes (Holmqvist et al. 2005; Iwafuchi-Doi et al. 2016). This mechanism may promote binding of other transcription factors via increased chromatin accessibility. One way this might be achieved is via the recruitment of nucleosome remodeling

complexes. This has been recently demonstrated, where OCT4 has been shown to recruit BRG1, part of the SWI/SNF chromatin remodeling complex, to gene targets in mouse embryonic stem cells (King & Klose 2017).

1.8 Enhancer function

Enhancers have been characterized as elements which activate transcription of a gene from a distance and independent of orientation (Banerji et al. 1981; Benoist & Chambon 1981). This has been demonstrated more definitively using 3C technology, where physical enhancer-promoter interactions have been observed (Tolhuis et al. 2002; Deng et al. 2012). Several characteristics have been associated with enhancers including DNA hypersensitivity, transcription factor binding and the presence of histone modifications H3K27ac and H3K4me1 (Creyghton et al. 2010; Heintzman et al. 2009; Heintzman et al. 2007; Xi et al. 2007; Thurman et al. 2012; Nord et al. 2013).

H3K27ac is predominantly catalysed by P300 and CREB Binding Protein (CBP) in humans (Tie et al. 2009; Garcia et al. 2007; Suka et al. 2001). Proteins which contain bromodomains and YEATS domains, such as BRD4 and AF9/ENL, have been demonstrated to bind acetylated lysine residues such as H3K27ac (Li et al. 2014; Erb et al. 2017; Wan et al. 2017). H3K4me1 is predominantly catalysed by MLL3 (KMT2C) and MLL4 (KMT2D) in humans (Herz et al. 2012; Hu et al. 2013; Lee et al. 2013; Tie et al. 2014). Although H3K4me1 does mark enhancer elements, it is not always indicative of an active enhancer (Creyghton et al. 2010). The presence of both H3K4me1 and H2K27ac has been demonstrated to be more indicative of an active enhancer (Creyghton et al. 2010). Interestingly, it has recently been demonstrated that catalytically inactive MLL3/4 does not affect transcription, whereas knockout of MLL3/4 did lead to the downregulation of transcription (Dorighi et al. 2017). Furthermore, MLL3/4 have

recently been demonstrated to be important for long-range chromatin interactions within TADs (Yan et al. 2017). This suggests that although H3K4me1 can be used as an indicator of enhancer presence, it's possible that MLL3/4 act as architectural proteins at enhancers which may be functionally more important than the methyltransferase activity.

Interestingly, a class of enhancers has been identified as 'super-enhancers'; these were characterised as enhancers which demonstrated high levels of Mediator, H3K27ac and H3K4me1, by at least one order of magnitude more than a typical enhancer. Furthermore, they were shown to be bound by developmentally important transcription factors and appear important for oncogenic transformation (Hnisz et al. 2013; Whyte et al. 2013).

1.9 Transcription initiation

The process of transcription is very complex and mediated by a multitude of different proteins. Therefore, to generate a clear and simple framework of the events which take place during transcription the process can be split into different stages and explained in a step-wise, deterministic manner (Figure 1.1). However, it is important to keep in mind that the protein interactions and associated functions which control transcription, and the order in which they take place, is extremely dynamic.

To initiate transcription, RNAPII is recruited to the promoter of the gene by the Pre-initiation complex (PIC which contains 6 protein subunits TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH (Kostrewa et al. 2009; He et al. 2013; Tsai & Sigler 2000; Parvin & Sharp 1993; Verrijzer et al. 1995). The promoter contains multiple sequences which are recognised by members of the PIC (Deng & Roberts 2005; Kim et al. 1993). One example of this is the TATA box, which is usually recognized by the transcription factor TATA binding protein (TBP) part

of the TFIID complex (Kim et al. 1993; Littlefield et al. 1999). CpG islands have also been demonstrated as important docking sites at promoters for CXXC-containing proteins which can contribute to transcription initiation (section 1.6). One of these proteins is MLL which deposits H3K4me3, a histone modification associated with active promoters (Chang et al. 2010; Lauberth et al. 2013; Vermeulen et al. 2007). Components of the multi-subunit Mediator complex have also been identified to interact with both the PIC and RNAPII, suggesting a role for Mediator in PIC complex formation and RNAPII recruitment (Baek et al. 2006; Esnault et al. 2008; Jishage et al. 2012; Johnson et al. 2002; Seizl et al. 2011; Plaschka et al. 2015).

An important stage which must occur for RNAPII to leave the promoter and move along the DNA is the separation of the DNA strands to allow "transcription bubble" formation (Holstege et al. 1997; Revyakin et al. 2004). TFIIH, comprised of two DNA helicases and CDK7, has been shown to control this stage (Flores et al. 1992; Glover-Cutter et al. 2009; Shiekhattar et al. 1995). TFIIH helicases have been demonstrated to separate the DNA strands, endorsing RNAPII promoter clearance (Fishburn et al. 2015). The kinase subunit, CDK7, has been demonstrated to phosphorylate Serine 5 on the C-terminal repeat domain (CTD) of RNAPII, further contributing to RNAPII promoter clearance (Feaver et al. 1994; Hengartner et al. 1998).

1.10 Transcription elongation

The initial transition of RNAPII into the gene often leads to the production of short abortive transcripts (Rougvie & Lis 1988). The transition into productive transcription elongation has been shown to be regulated by the DRB sensitivity-inducible factor (DSIF), Negative elongation factor (NELF) and the Positive transcription elongation factor (P-TEFb)(Yamaguchi et al. 1999; Wada et al. 1998; Peterlin & Price 2006; Marshall & Price 1995). P-TEFb is comprised of Cyclin Dependent Kinase 9 (CDK9) and Cyclin T1/T2. P-TEFb

has been shown to be recruited to gene targets via an interaction with BRD4 (Peng et al. 1998; Marshall & Price 1995; Jang et al. 2005; Yang et al. 2005; Peterlin & Price 2006; Yamada et al. 2006), although recent conflicting evidence does suggest P-TEFb may be recruited independently of BRD4 (Winter et al. 2017). CDK9 has been shown to phosphorylate Serine 2 of RNAPII CTD (Marshall et al. 1996), however, it has been suggested that CDK12 may also phosphorylate serine 2 (Yu et al. 2015). In addition to this, CDK9 has been shown to phosphorylate both DSIF and NELF, leading to the dissociation of DSIF and the conversion of NELF into an apparent positive elongation factor (Yamaguchi et al. 1999; Wada et al. 1998; Fujinaga et al. 2004). This cascade of phosphorylation events leads to the pause release of RNAPII and productive transcription elongation. This stage of transcription has been suggested to be a rate limiting stage of many developmentally important gene targets in which RNAPII is poised at proximal promoter regions (Krumm et al. 1995; Liu et al. 2014). This potentially allows for the rapid generation of full length transcripts following signaling cues.

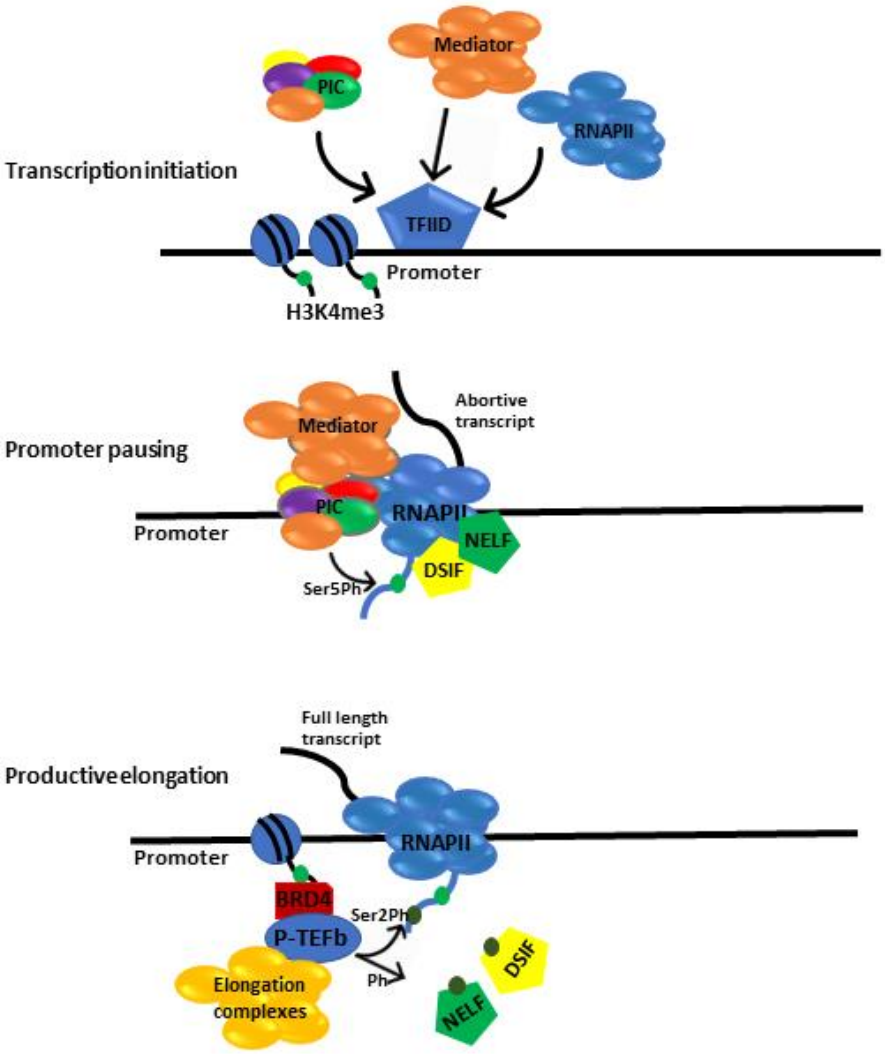

Figure 1-1. Stages of Transcription. **Initiation:** TFIID, containing TBP and TBP-like proteins, bind to specific promoter sequence. Other components of the PIC are recruited to the promoter, as well as components of the mediator complex and RNAPII. H3K4me3 is also located at the promoter region and can help mediate RNAPII stabilization. **Promoter pausing:** TFIIH phosphorylates RNAPII serine 5. RNAPII creates abortive transcripts and is bound by negative regulation factors DSIF and NELF. **Productive elongation:** P-TEFb is recruited via BRD4 (via binding to acetylates lysine residues). P-TEFb phosphorylates RNAPII serine 2, NELF and DSIF. NELF and DSIF release RNAPII. RNAPII begins to create full length transcripts. P-TEFb can recruit elongation complexes to the gene body.

1.10.1 Elongation complexes - EAP/AEP/SEC

P-TEFb has been shown to be part of a large protein complex comprising other proteins which have been demonstrated to enhance transcription elongation and interestingly contain MLL fusion partners and DOT1L (Section 1.16) (Biswas et al. 2011; Yokoyama et al. 2010; Mueller et al. 2007; Lin et al. 2010). One such MLL fusion partner, ENL, was used to purify the ENL Associated Protein (EAP) complex. This macromolecular complex contains ENL, AF4, AFF4, BCoR, CBX8, DOT1L, P-TEFb, LAF4 and RING1 (Mueller et al. 2007). It has been observed that the purified proteins were not present in equi-molar concentrations, suggesting the presence of multiple different sub-complexes with ENL as a common subunit (Mueller et al. 2007).

Further to this, the AF4, ENL P-TEFb (AEP) complex and Super Elongation Complex (SEC) were purified (Yokoyama et al. 2010b; Lin et al. 2010). Neither complex contained DOT1L, demonstrating that even though DOT1L and ENL can interact, DOT1L and AF4 are mutually exclusive (Figure 1.2). Supporting this, it was observed via NMR spectrometry that AF4 and DOT1L compete for the same binding surface on AF9 and ENL (Leach et al. 2013), providing structural evidence for why DOT1L and AF4 do not co-purify, even though they both interact with AF9/ENL. Figure 1.3 describes the known interacting domains of AF4, which include interactions with other members of the purified elongation complexes such as P-TEFb, AF9, ENL and AFF4.

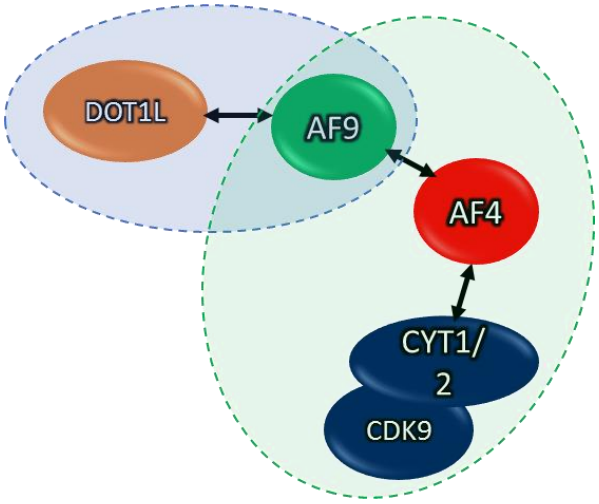

Figure 1-2. AF4 and DOT1L are found in mutually exclusive complexes. (Yokoyama et al., 2010; Lin et al., 2010; Biswas et al., 2011; Leach et al., 2013).

Proteins associated with transcription elongation have been shown biochemically to interact with proteins associated with different stages of transcription. An example of this includes the biochemical purification of AF9 with TBP, demonstrating a potential overlap between transcription initiation and elongation complexes (Biswas et al. 2011). Additionally, the Mediator subunit MED26 has been demonstrated to interact with AF9 and ENL, members of the SEC (Takahashi et al. 2011; Biswas et al. 2011).

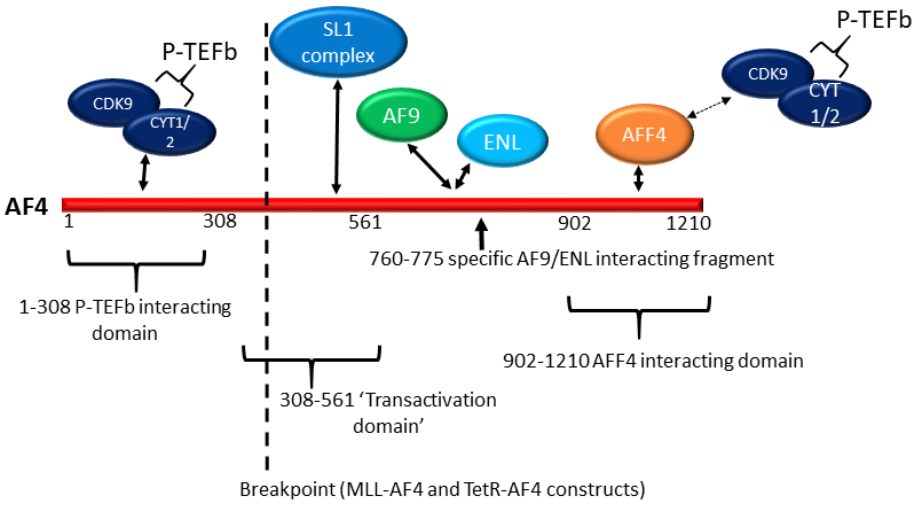

Figure 1-3. Domains of AF4 shown to interact biochemically with interacting proteins. Right of the dashed line represents the portion of AF4 found in the TetR-AF4 fusion protein (Chapter 3) and in MLL-AF4 leukaemia (Biswas et al. 2011; Okuda et al. 2015; Leach et al. 2013; C. Lin et al. 2010;Yokoyama et al. 2010a).

The complex nature of interactions occurring between proteins associated with transcription is highlighted by the multitude of different interactions observed biochemically. These differences in biochemically purified complex compositions could be due to differences in purification method and the type of cells used. For example, the EAP complex was purified from 293T cells (Mueller et al. 2007) whereas the AEP complex was purified from K562 cells using different methods (Yokoyama et al. 2010). Furthermore, even though biochemical purifications do provide an initial understanding of a complex *in vitro*, it cannot be used to fully understand the complex nature of these interactions *in vivo*, at a gene target. Therefore, further clarification *in vivo* is needed if we are to understand how these complexes may assemble at gene targets and contribute to transcription elongation.

1.10.2 PAF1 complex

Another complex which has been associated with transcription elongation, and has also been linked with DOT1L and H3K79me, is the Polymerase Associated Factor 1 (PAF1) complex. The PAF1c is a multi-protein complex originally identified in *S.cerevisiae* (Shi et al. 1997; Wade et al. 1996; Mueller & Jaehning 2002). In humans, the core PAF1 complex (PAF1c) comprises PAF1, LEO1, CTR9, RTF1, CDC73 and SKI8 and has also been shown to interact with RNAPII and promote transcription through chromatin templates *in vitro* (Kim et al. 2010). Additionally, PAF1c, has been demonstrated biochemically to interact with other transcription elongation complexes, including AF9 and ENL, via the YEATS domain (Figure 1.4A) (Squazzo et al. 2002; Xu et al. 2017; He et al. 2011). This suggests there is a link between PAF1 and the SEC, however it is not clear whether this interaction is functional *in vivo* (Squazzo et al. 2002; Xu et al. 2017; He et al. 2011).

In *S.cerevisiae* and humans, it has also been demonstrated that members of the Paf1c interact with the Rad6/Bre1 complex, responsible for H2B ubiquitination (Kim et al. 2009; Van Oss et al. 2016; Wood et al. 2003; Ng et al. 2003). Interestingly, H2BK123ub has been demonstrated as a pre-requisite for high levels of H3K79me and H3K4me in both *S.cerevisiae* and humans (Briggs et al. 2002; Dover et al. 2002; Sun & Allis 2002; Kim et al. 2009; McGinty et al. 2008). Additionally, Paf1c knockout studies have shown a reduction in H2Bub and H3K79me (Ng et al. 2003; Xiao et al. 2005; Wood et al. 2003; Krogan et al. 2003; Van Oss et al. 2016). The mechanism of this and whether this occurs in humans is unclear.

1.11 DOT1L and H3K79 methylation

Dot1 was first identified in *S.cerevisiae* in a genetic screen for proteins whose overexpression disrupted telomeric silencing. This was further shown to occur via antagonism with Silent Information Regulator (Sir) proteins (Singer et al. 1998). Later studies identified DOT1 as a histone methyltransferase (Ng et al. 2002; Feng et al. 2002; Lacoste et al. 2002). Homologous DOT1 methyltransferases have been identified in many other species demonstrating high conservation of DOT1

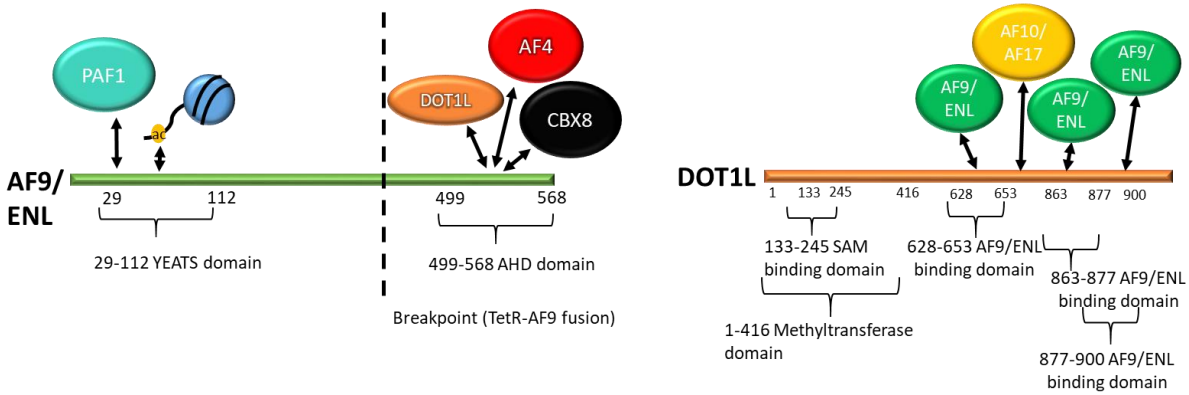

Figure 1-4. Identified biochemical interacting domains of (A) AF9 and ENL (B) DOT1L. DOT1L bind to same binding domain of AF9/ENL rendering these interactions mutually exclusive (He et al. 2011; Li et al. 2014; Wan et al. 2017; Leach et al. 2013; Kuntimaddi et al. 2015; Okada et al. 2005)

DOT1L (human DOT1) is responsible for catalysing mono-, di- and tri methylation of H3K79 in a non-processive manner, dissociating and associating again to deposit di- and tri- H3K79me (Frederiks et al. 2008; Feng et al. 2002). The catalytic domain of DOT1L, corresponding to amino acids 1-416, has been shown to uniquely differ from other lysine methyltransferases (KMTs) as it does not contain a SET domain (Figure 1.4B). (Min et al. 2003; Feng et al. 2002). Furthermore, H3K79 is a residue located in the global core of H3 (Richmond et al. 1997). This may mean this residue is potentially difficult to access which is of particular importance for DOT1L mediated methylation of H3K79 (Lu et al. 2008; Botuyan et al. 2006). However, studies have demonstrated that other neighbouring residues, such as H2BK123ub may enhance H3K79me by facilitating accessibility (Altaf et al. 2007; Jack & Hake 2014).

1.11.2 Readers and Erasers of H3K79 methylation

Most identified modified histone residues, which are associated with transcription, are associated with reader and eraser proteins. Two proteins, both implicated in the DNA damage response, identified as reader proteins of H3K79me are Survivor of motor neurone (SMN) and p53 binding protein (53BP1) (Botuyan et al. 2006; Huyen et al. 2004; Sabra et al. 2013). Although the affinity of 53BP1 for H3K79me is much lower compared to the affinity for H4K20me2 (H3K79me K_D = 2mM and H4K20me2 K_D = 2.2μM), it is possible that H3K79me may play a minor role in the recruitment of 53BP1 to double strand breaks (DSBs) in the DNA repair pathway (Botuyan et al. 2006; Wakeman et al. 2012). In addition to this, there is no currently identified demethylase for H3K79me. This is supported by the fact that loss of H3K79me, following DOT1L inhibition and knockdown, can take up to several days (Daigle et al. 2013). This might suggest that H3K79me is diluted via histone turnover during cell division and histone recycling mechanisms.

1.11.3 DOT1L complex and recruitment mechanism

DOT1L has been purified in several multi-protein complexes associated with transcription elongation (Yokoyama et al. 2010; Biswas et al. 2011; Monroe et al. 2011; Mueller et al. 2007). One of these complexes, termed the DOTCOM complex, was purified from HEK293T cells using a Flag-tagged DOT1L followed by Multidimensional Protein Identification Technology (MUDPIT) analysis. This gave rise to the identification of a 2MDa protein complex comprising of DOT1L, AF10, AF17, ENL, AF9, SKP1, TRRAP and β-catenin (Mohan et al. 2010) (Figure 1.5).

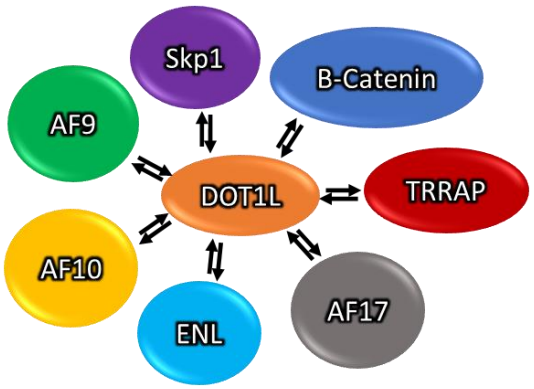

Figure 1.5. Composition of the 'DOTCOM' complex (Mohan et al. 2010)

When considering the recruitment mechanism of DOT1L to genes, it is possible that DOTCOM members can contribute to a stable binding site for DOT1L. Biophysical studies have demonstrated that AF9 and ENL can interact with three independent binding sites of DOT1L, suggesting multiple AF9 or ENL proteins can interact with one DOT1L protein molecule simultaneously (Kuntimaddi et al. 2015) (Figure 1.4B). This perhaps increases the binding affinity of DOT1L at a gene target via multivalent interactions (Ruthenburg et al. 2011). In addition, AF10 and AF17 could also contribute to the stabilisation of DOT1L via the OM-LZ

domain (Chen et al. 2012; Deshpande et al. 2011; Okada et al. 2005). However, exactly how DOT1L is specifically recruited to a gene *in vivo*, likely by some of the purified DOTCOM components, is unknown.

The observed biochemical purification of DOT1L and members of the wnt signalling pathway, Skp1, TRRAP and β-catenin is controversial. Although it was demonstrated that knockdown of DOT1L in *D.melanogaster* led to the decrease in expression of important wingless gene targets (Mohan et al. 2010), more recent studies have demonstrated that DOT1L and H3K79me are not essential for the maintenance of wnt gene target expression in human cells (Ho et al. 2013; Gibbons et al. 2014). This demonstrates how biochemical purifications can identify artefactual interactions which may have no biological relevance *in vivo*.

Even though it was demonstrated biochemically that AF9, ENL, AF10 and AF17 can interact with DOT1L, it is important to understand whether these interactions are important for DOT1L function in the cell. To test this, siRNA knockdowns of DOT1L complex members were carried out to determine the effect upon H3K79me. Upon knockdown of ENL and AF10, a reduction in H3K79 methylation was observed, in addition to a decrease of transcription of DOT1L target genes (Mueller et al. 2007; Deshpande et al. 2014). This suggests the interaction of DOT1L with DOT1L complex members are important for H3K79me, potentially contributing to the recruitment mechanism of DOT1L.

1.12 Functional mechanisms of DOT1L and H3K79me in transcription

DOT1L and H3K79me have primarily been associated with active transcription, with over 90% of the *S.cerevisiae* genome shown to be H3K79 methylated and to be actively transcribed (Turner 2002; Steger et al. 2008; van Welsem et al. 2008). Furthermore, in *D.melanogaster*

and humans H3K79me was restricted to the euchromatic compartment (Schübeler et al. 2004; Steger et al. 2008).

H3K79me2/3 is found at the 5' end of the gene body of actively transcribing genes in mammals, suggesting DOT1L is specifically involved in transcription elongation (Steger et al. 2008). This is further supported by *in vitro* data demonstrating DOT1L and H3K79me associated with transcription elongation complexes (Mohan et al. 2010; Biswas et al. 2011; Kim et al. 2010; Mueller et al. 2007; Leach et al. 2013; Kuntimaddi et al. 2015). Furthermore, transcription was severely impaired following Dot1l knockout, suggesting DOT1L and H3K79me is functionally important in transcription (Steger et al. 2008). Exactly how DOT1L and H3K79 methylation contribute to transcription elongation remains unclear, however, models of DOT1L function in *S.cerevisiae* and human MLL-rearranged leukaemia cells have been proposed.

1.12.1 DOT1L and SIR protein antagonism

The histone deacetylase, Silent information regulator 2 (SIR2) has been shown to antagonise DOT1 to control telomeric silencing in *S.cerevisiae* (Altaf et al. 2007; Fingerman et al. 2007; Kitada et al. 2012). One specific residue which can be acetylated is H4K16. H4K16 has been shown to be located close to H3K79 and it has been suggested that DOT1 activity antagonises the deacetylation of H4K16ac by Sir2 (Fingerman et al. 2007; van Welsem et al. 2008; Altaf et al. 2007). Furthermore, it was demonstrated that deacetylation of H4 is correlated with a decrease in H3K79me (Osborne et al. 2009). This suggests that antagonism between Sir2 and DOT1 may contribute to the maintenance of active or repressed transcription and telomeric silencing in *S.cerevisiae*.

Moreover, recent studies in human MLL-AF9 leukaemia cell lines have demonstrated that DOT1L and H3K79me may antagonise SIRTUIN 1 (SIRT1) at some MLL-fusion protein (MLL-FP) gene targets (Chen et al., 2015). Following inhibition and shRNA mediated knockdown of DOT1L, loss of H3K79me was coupled with the loss of H3K27ac and H3K9ac (Chen et al. 2015). Furthermore, an increase in SIRT1 occupancy was observed. This was coupled with an increase in repressive H3K9me3, potentially catalysed by SUV39H1. Interestingly SIRT1 and SUV39H1 have been biochemically purified together in a separate study, suggesting a potential functional interaction (Vaquero et al. 2007). This suggests that DOT1L may play a role at some gene targets via antagonism of repressive chromatin proteins and modifications (Chen et al., 2015).

1.12.2 DOT1L and BRD4 co-regulate a subset of MLL-FP gene targets

Another recently proposed model of DOT1L function has indicated that DOT1L may create a more permissive chromatin environment at MLL-FP gene targets co-regulated by BRD4. The permissive landscape of these targets was shown to promote CREB1 binding (Gilan et al. 2016). Following DOT1L methyltransferase inhibition, BRD4 and CREB1 binding was lost. CREB1 has been shown to recruit acetyltransferases CBP and P300. Moreover, H4K5ac was shown to be lost following DOT1L inhibition. Therefore, H3K79me potentially creates a permissive chromatin landscape whereby CREB1 can bind to the gene, recruit P300/CBP, thus promoting H4K5ac leading to BRD4 recruitment.

Interestingly, the two models of DOT1L function described are not mutually exclusive and both could occur in concert. This would suggest that DOT1L can antagonise repressive modifications at the same time as promoting an active chromatin state for transcription factor binding. Even though these models depict what the function of DOT1L in transcription might

be, it is unclear whether these mechanisms could be used to describe DOT1L function at all gene targets or whether this is limited to the subset of gene targets studied.

1.13 DOT1L role in embryonic development

So far, it has been demonstrated that DOT1L is important for transcription of some genes but this does not explain whether DOT1L target genes are important for tissue specificity or development, that are both important aspects of understanding the function of DOT1L.

Several models investigating the role of DOT1L in mammalian embryogenesis have been studied, both *in vitro* and *in vivo* using mESCs. This has been modelled using separate approaches including a Cre/LoxP system to inactivate Dot1l methyltransferase activity as well as shRNA knockdown (Jones et al. 2008; Barry 2010; Feng et al. 2010). Both approaches demonstrated severely impaired cellular proliferation and cell cycle arrest at the G2/M stage following differentiation induction. DOT1L is essential for normal embryonic development, since homozygous knock out of Dot1l in transgenic mice is embryonic lethal between E10.5 and E13.5 due to severe anaemia (Feng et al. 2010; Jones et al. 2008). Pre-E10.5 Dot1l null embryos displayed severe anaemia, abnormal blood vessel formation and proliferation defects in erythroid progenitors, supporting an important role for DOT1L for normal blood development.

1.14 DOT1L role in adult haematopoiesis

In addition to embryonic haematopoiesis, DOT1L has been shown to be important for adult haematopoiesis. This was shown using ubiquitous Dot1l knockout in adult mice (Bernt et al. 2011; Nguyen et al. 2011). Mice died 8-12 weeks following Dot1l knockout and displayed

severe anaemia and depletion of haematopoietic stem cells (HSCs) and progenitors. In addition to this, Dot1l knockout was performed specifically in the haematopoietic compartment (Jo et al. 2011). Although the age of death of these mice was not reported, anaemia, hypocellularity of the bone marrow and a reduction in haematopoietic progenitors were observed. These studies together suggest that DOT1L has a role in maintenance of normal adult haematopoiesis. As DOT1L has been shown to play an important role in haematopoiesis in both embryogenesis and adults, it is perhaps not surprising that mis-regulation of DOT1L has been strongly implicated in several types of leukaemia, including acute myeloid leukaemia (AML) and acute lymphoblastic leukaemia (ALL).

1.15 Acute Lymphoblastic Leukaemia (ALL) and MLL-r leukaemia

ALL is the commonest childhood cancer. Although ~90% of childhood ALL is curable by chemotherapy, this treatment also has major side effects (Hunger & Mullighan 2015). The remaining 10% of ALL is particularly aggressive and is mainly caused by rearrangements involving the MLL gene (Pieters et al. 2007; van der Linden et al. 2009). Clearly, new therapies are required to treat this subset. Additionally, more effective and less damaging therapies are also required for the 90% of ALLs which can be treated with standard chemotherapy. To develop new, and better, therapeutic strategies, the molecular mechanisms which drive leukaemias needs to be investigated on a disease-by-disease basis. DOT1L has been shown to be a key driver in MLL-r leukaemia (Bernt & Armstrong 2011; Krivtsov et al. 2008). Therefore, understanding the molecular mechanism of DOT1L in these leukaemias is important for identifying new therapies.

MLL was originally identified from MLL-r leukaemia patient samples and cell lines (Ziemin-van der Poel et al. 1991; Djabali et al. 1992; Tkachuk et al. 1992). MLL-r leukaemia accounts

for approximately 70% of infant leukaemias and, as mentioned, around 10% of childhood leukaemias as well as some adult leukaemia (Meyer et al. 2013). MLL-r leukaemia arises following a chromosome translocation between MLL, on chromosome 11, with 1 of over 121 fusion partners. This creates an in-frame MLL fusion gene composed of the N-terminus of MLL (MLL-N) and the C-terminus of the fusion partner. This is translated into an MLL-FP. 85% of translocations fuse MLL with either AF4, AF9, AF10, AF6, ENL or ELL (Meyer et al. 2013). The breakpoints of MLL span a roughly 8.3kb region, which retains the AT hooks and CXXC domain of MLL and leads to the loss of all other domains, including the methyltransferase domain (Thirman et al. 1993).

Numerous studies have been carried out which show that MLL-FPs can bind to and upregulate common gene targets including the *HOXA* gene cluster and *MEIS1* (Zeisig et al. 2003; Milne et al. 2010; Bach et al. 2010; Milne et al. 2005; Rozovskaia et al. 2001; Orlovsky et al. 2011; Armstrong et al. 2002). Even though common gene targets have been observed, unique gene targets have also been identified in leukaemia driven by different MLL-FPs (Stam et al. 2010; Stumpel et al. 2009). For example, two sub-types of MLL-AF4 leukaemia can be differentiated on the basis of expression vs lack of expression of HOXA genes (Stam et al. 2010). Differential gene targets between MLL-FPs has been further verified with genome wide overlaps between MLL-AF9 and MLL-AF4 gene targets (Guenther et al. 2008a; Bernt et al. 2011).

1.15.2 MLL-AF4 leukaemia

Up to 50% of infant MLL-r ALL and 70% of adult MLL-r ALL are caused by MLL-AF4, the most prevalent MLL-FP (Meyer et al. 2013). The rapid disease onset and aggressive nature of MLL-AF4 ALL contributes to a very poor survival rate compared to other MLL-r leukaemias (Pui et al. 2003). Even though extensive research has been carried out for MLL-FP leukaemia,

most of this has been carried out for MLL-AF9 and MLL-ENL leukaemia. However, treatments and understanding of the molecular mechanisms underlying MLL-AF4 leukaemia are still lacking, indicating why further research to understand the molecular mechanisms driving MLL-AF4 leukaemia is necessary.

1.16 Recruitment of MLL-FPs

MLL-N has been shown to biochemically interact with Menin and Lens epithelium-derived growth factor (LEDGF) which have been shown to be important for MLL-FP recruitment to gene targets (Caslini et al. 2007; Y.-X. Chen et al. 2006; Hughes et al. 2004; Yokoyama et al. 2005; Milne, Martin, et al. 2005). Menin has been demonstrated to connect the interaction between MLL and LEDGF (Yokoyama & Cleary 2008). LEDGF contains a PWWP domain which has been shown to bind to H3K36me2/3 (Hughes et al. 2004; Botbol et al. 2007). Interestingly, the minimal domain of MLL to mediate *HOXA9* activation did not contain the Menin binding domain but did contain the CXXC domain. This suggests that alternative mechanisms of MLL-FP recruitment may be possible via the CXXC domain which has been shown to interact with unmethylated CpG islands and PAF1 complex (Milne et al. 2010). Perturbation of both PAF1 and CpG island interactions led to a complete loss of MLL binding, suggesting both may be in part responsible for MLL recruitment. Recently, it has been shown that PAF1 co-localises with MLL-FPs at only a subset of gene targets and PAF1 is not sufficient to recruit MLL-AF4 using the TetR system which may indicate that PAF1 is not required for MLL-FP recruitment genome wide (Kerry et al. 2017; Milne et al. 2010).

1.17 Models of MLL-AF4 leukaemia

Although MLL-AF4 is the most common MLL-FP and gives rise to an aggressive leukaemia, it has been very difficult to generate mouse models which faithfully recapitulate the human pro-B ALL phenotype typically observed in patients. Several different experimental approaches have been taken to study this. MLL-AF4 knock-in mouse models were initially attempted via mESC engraftment but did not recapitulate the ALL phenotype observed in many MLL-AF4 patients, instead a B-lymphoma was produced (Chen et al., 2006; Metzler et al., 2006). Interestingly, following Mll-AF4 expression in B or T-cell progenitors, a B-lymphoma was always produced, suggesting that Mll-AF4 is sufficient to cause lineage reassignment (Metzler et al. 2006). Moreover, targeting more primitive cell types using Lmo2-Cre gave rise to embryonic lethality suggesting the target cell type is probably not an early stem cell (Metzler et al. 2006). Conversely, MLL-AF4 infant leukaemia, thought to arise *in utero*, is often considered as a separate disease to MLL-AF4 childhood or adult leukaemia (Sanjuan-Pla et al. 2015). Therefore, to try and recapitulate this, Mll-AF4 was expressed in early haematopoietic cells in the mouse embryo (Barrett et al. 2016). This gave rise to a long latency B-lymphoma, again not recapitulating the phenotype observed in the human infant disease. MLL-AF4 leukaemia has been shown to have very few co-operating mutations, however, some cases do have co-existing activating KRAS mutations and/or high expression of *FLT3* (Chillón et al. 2012; Armstrong et al. 2004; Taketani et al. 2004; Liang et al. 2006). However when both *MLL-AF4* and *KRAS* or *FLT3* were overexpressed in mESCs or human CD34+ cord blood cells, enhanced proliferation but not an acute leukaemia was observed (Sanjuan-Pla et al. 2015; Tamai et al. 2011; Montes et al. 2011; Montes et al. 2014; Prieto et al. 2016). Surprisingly, when the reciprocal fusion protein AF4-MLL, sometimes detected in patients, was transduced into Lin-/Sca1+ murine cells, a pro-B-ALL phenotype was observed (Andersson et al. 2015;

Bursen et al. 2010). Even though this faithfully recapitulates the human disease it perhaps does not account for the many patients which do not express the reciprocal fusion protein and the leukaemic capability of MLL-AF4 alone (Kowarz et al. 2007; Thomas et al. 2005; Kumar et al. 2011). Interestingly, it has recently been shown that murine Af4 produces higher viral titres and transduces more efficiently than human AF4 (Lin et al. 2016). Using this approach, human CD34+ cells were retrovirally transduced with MLL-Af4 and successfully generated a pro-B cell ALL in xenografts (Lin et al. 2016). Interestingly, this leukaemia did not express *HOXA9* but did express other MLL-FP gene targets including *RUNX1*, suggesting this model recapitulates a pro-B MLL-AF4 leukaemia which does not express *HOXA9* observed in some patients (Stam et al. 2010). Taken together, these data suggest that after many attempts to model MLL-AF4 leukaemia we are now perhaps getting closer to this, however, there are many aspects of MLL-AF4 leukaemogenesis which are still unclear including understanding the nature of the leukaemia initiating cell and the differences which may occur between infant, childhood and adult MLL-AF4 leukaemia. This highlights the importance of further understanding the molecular mechanisms driving MLL-AF4 leukaemia.

1.18 Molecular mechanisms of MLL-FP gene activation

There are many stages of transcription which could be controlled by MLL-FPs which could lead to the abnormal upregulation of genes. Many of the most common fusion partners have been shown to interact with protein complexes associated with different stages of transcription (Okuda et al. 2016; Biswas et al. 2011). Therefore, it is thought that aberrant recruitment of proteins which can interact with the fusion partner, can contribute to the gene activation mechanism of MLL-FP gene targets. Two most widely reported mechanisms include MLL-FP gene activation via P-TEFb recruitment or DOT1L recruitment.

1.18.1 Role of BRD4 and P-TEFb in MLL-FP leukaemia

P-TEFb has been shown to interact biochemically with common MLL fusion partners including AF4 (Yokoyama et al. 2010; Mueller et al. 2007). Therefore, it is possible that MLL-AF4 may function via recruitment of P-TEFb. MLL-FP leukaemia has been shown to be sensitive to inhibition of CDK9, suggesting that P-TEFb plays a role in MLL leukaemogenesis (Mueller et al. 2009). Furthermore, inhibition of BRD4, shown to recruit P-TEFb, leads to abrogation of MLL-r leukaemias both *in vitro* and *in vivo* (Dawson et al. 2011). In contrast, it has been demonstrated that leukaemic transformation was not achieved via expression of MLL-fused to the P-TEFb interacting domain of AF4 (Yokoyama et al. 2010). Additionally, following MLL-AF4 knockdown, P-TEFb binding was not affected at key MLL-AF4 gene targets such as *BCL2* and *RUNX1* suggesting P-TEFb is not important in regulating these important MLL-AF4 gene targets (Benito et al. 2015; Wilkinson et al. 2013).

1.18.3 Role of DOT1L in MLL-FP leukaemia

It has been observed that over 95% of MLL-FP gene targets generally have aberrantly high levels of H3K79me in the gene body in comparison to other non-MLL-FP targets which exhibit similar expression levels (Bernt et al. 2011; Guenther et al. 2008; Krivtsov et al. 2008). Furthermore, a 7 fold increase in H3K79me was observed at *HOXA9* and *MEIS1* following MLL-ENL binding (Milne et al. 2005). This suggests that high levels of H3K79me observed at MLL-FP genes is a direct consequence of MLL-FP binding.

Moreover, it has also been demonstrated that DOT1L and H3K79me are important for the expression of MLL-FP gene targets *in vivo.* Following DOT1L knockout, *HOXA9* and *MEIS1* were downregulated (Bernt et al., 2011). H3K79me is ubiquitously expressed and marks most

active genes. Interestingly, loss of H3K79me lead to the differential expression of a subset of genes, enriched for MLL-FP gene targets (Bernt et al. 2011; Kerry et al. 2017; Krivtsov et al. 2008; Gilan et al. 2016). The reason why some genes are more reliant upon DOT1L compared to others is poorly understand. Whether the mechanisms used to describe DOT1L function can account for this specificity is unknown.

Several studies have specifically focused upon the regulation of MLL-AF4 gene targets via H3K79me2/3. *BCL2* and *RUNX1* have recently been identified as genes which are directly bound by MLL-AF4 in SEM cells, an MLL-AF4 expressing cell line. Following MLL-AF4 siRNA knockdown and DOT1L inhibition, *BCL2* and *RUNX1* were downregulated (Godfrey et al. 2016; Benito et al. 2015; Wilkinson et al. 2013). Interestingly, H3K27ac was also depleted at the *BCL2* promoter and enhancer following MLL-AF4 siRNA knockdown. This suggests that MLL-AF4 is both modulating H3K79me2/3 and H3K27ac, which may contribute to the downregulation of genes (Godfrey et al. 2016). *PROMININ 1* (*PROM1*), which encodes the cell surface marker CD133, is also regulated via MLL-AF4 and following MLL-AF4 shRNA mediated knockdown, downregulation of *PROM1* was observed. The mechanism of MLL-AF4 gene activation of *PROM1* is unclear, although high levels of H3K79me2/3 are observed in the gene body of *PROM1* (Mak et al. 2012; Guenther et al. 2008; Thomas et al. 2005).

BCL2, RUNX1 and PROM1 are all examples of MLL-AF4 spreading targets (Kerry et al. 2017). Spreading targets, also identified in other MLL-FP leukaemia, were characterised by the spread of large domains of MLL-AF4, Menin, H3K79me3 and unmethylated CpG islands in the gene body coupled with high levels of gene expression. Interestingly, these gene targets were shown to be particularly important for the leukaemia and were also hypersensitive to DOT1L inhibitor treatment (Kerry et al. 2017).

Even though spreading MLL-AF4 gene targets seem to be dependent upon DOT1L, it is unclear how DOT1L is recruited to these targets. AF4 and DOT1L have been observed to be mutually exclusive (Leach et al. 2013; Yokoyama et al. 2010). Therefore, the mechanism of DOT1L recruitment to MLL-AF4 gene targets is not as simple as perhaps the recruitment of DOT1L to MLL-AF9 gene targets where DOT1L interacts directly with the AF9 fusion partner. Therefore, further investigation is required to understand the recruitment mechanism of DOT1L to MLL-AF4 gene targets.

1.19 DOT1L inhibitors are effective against MLL-r leukaemia and other cancers

MLL-r leukaemia is thought to be a disease which is highly dependent upon transcription of MLL-FP gene targets. Molecular studies of DOT1L in MLL-FP leukaemia have highlighted that DOT1L plays a pivotal role in MLL-FP gene regulation. Therefore, specific inhibitors which block the methyltransferase activity of DOT1L were developed to assess whether this would be an effective therapeutic strategy for treating MLL-r leukaemia.

EPZ-5676, now in clinical trials, and SGC0946 are both small molecule inhibitors which bind to the SAM binding pocket of DOT1L blocking the methyltransferase activity (Yu et al. 2012; Daigle et al. 2013). Both EPZ-5676 and SCG-0946 were shown to abrogate MLL-r leukaemias via inhibition of proliferation and induction of apoptotic pathways *in vivo* and these effects have been shown on leukaemic cells in *vitro* (Daigle et al. 2011; Daigle et al. 2013; Deshpande et al. 2013; Chen et al. 2012).

Furthermore, it has emerged that DOT1L inhibitors could be an effective treatment against other types of leukaemia and cancers. Examples of this include AML containing IDH1/2 and DMNT3a mutations and leukaemias driven by MLL-PTD and CALM-AF10 which are all

sensitive to EPZ-5676 treatment (Rau et al. 2016; Sarkaria et al. 2014; Chen et al. 2012; Meyer et al. 2013; Kühn et al. 2015). Moreover, DOT1L inhibitors have also been reported to have anti-proliferative effects against Oestrogen Receptor (ER)+ breast cancer cell lines (Zhang et al.; Cho et al. 2015).

1.20 DOT1L inhibitors in combinatorial therapy for MLL-r leukaemia

Even though DOT1L inhibitors are in clinical trials, it is already clear that many leukaemias are resistant to these drugs or relapse after an initial therapeutic response (Campbell et al. 2017). This suggests that individually DOT1L inhibitors may not be as effective in the clinic as first thought. As DOT1L is clearly important for MLL-r leukaemia, this may call for the development of better DOT1L inhibitors. Alternatively, current DOT1L inhibitors may be useful in combinatorial treatment. This has been demonstrated with treatment of EPZ-5676 coupled with other chemotherapeutic agents or DNMT inhibitors led to an anti-proliferative effect upon MLL-FP cell lines (Klaus et al. 2014).

In addition, other small molecule inhibitors have been shown to act either synergistically or additively with EPZ-5676. One example of this is ABT-199; a BH3 mimetic which specifically targets BCL2 (Vandenberg & Cory 2013; Souers et al. 2013; Benito et al. 2015). As BCL2 is an important MLL-AF4 gene target, SEM cells were treated with EPZ-5676 and ABT-199. An enhanced anti-proliferative effect was observed using a combination of these inhibitors (Benito et al. 2015). Other epigenetic inhibitors, such as JQ1 and I-BET, which target the BET family of bromodomain proteins, and MI-2-2, a menin inhibitor, have been shown to have enhanced anti-proliferative effect upon MLL-r leukaemias when used in combination with DOT1L inhibitors (Gilan et al. 2016; Dafflon et al. 2017). Taken together, this data suggests that combinatorial treatments with DOT1L inhibitors and other epigenetic drugs may be an

effective treatment approach for MLL-r leukaemias. Therefore, understanding further how DOT1L regulates this leukaemia is key to understanding which combinatorial therapies may be most effective.

The aims of this DPhil project are:

1. How is DOT1L recruited to MLL-AF4 gene targets
2. Which gene targets are dependent upon DOT1L function in MLL-AF4 leukaemia
3. What is the function of DOT1L and H3K79me at MLL-AF4 gene targets

Chapter 2 - Materials and Methods

2.1 Tissue Culture methods

2.1.1 Media and growth conditions of leukaemia cell lines

All media was supplemented with 10% Heat-inactivated Fetal Bovine Serum (FBS #10500-064, Gibco, Life Technologies), 1% GlutaMax (#A12860-01, Gibco, Life Technologies) (Table 2.1).

Table 2-1. Media and components used for cell culture

Media type	**Media component**
IMDM	[+]L-Glutamine, [+]25mM HEPES (#21980-032, Gibco, Life Technologies)
DMEM	[+]4.5g/L D-Glucose, [+]L-Glutamine, [-]Pyruvate DMEM (#41965-039, Gibco, Life Technologies)
RPMI	[+]L-Glutamine (#21875-034, Gibco, Life Technologies)
OPTIMEM	Reduced Serum Medium (#51985-026, Gibco, Life Technologies).

All cell lines were grown at 37°C and 5% CO_2 (Table 2.2). All cells grown in suspension, including SEM, MV4;11 and RS4;11 cells, were maintained between $0.5x10^6$-$2x10^6$ cells/ml. Cell densities were determined using a haemocytometer in addition to using Trypan blue staining to determine live cells. Based upon cell densities cells were passaged usually every 2-3 days.

2.1.2 Growth conditions of mouse embryonic stem cells (mESCs)

mESCs were grown in round 10cm dishes and passaged at a 1:10 ratio. The dishes were coated for 10 minutes in 0.1% PBS-Gelatin (made in-house and autoclaved (Gelatin Solution, Type B, 2% in H_2O, #G1393-100ML, Sigma, Life Science; PBS pH7.4, #10010-015, Gibco, Life Technologies). Gelatin was then aspirated and 10ml of DMEM plus appropriate supplements were added to the dish pre-emptively. mESCs were passaged at 90% confluency and diluted 1/10 using diluted 0.1% trypsin diluted in PBS (0.5% Trypsin-EDTA (10X), #15400-054, Gibco, Life Technologies). This was then incubated for 3 minutes and quenched with FBS supplemented DMEM.

2.1.3 Human Fetal Bone Marrow Samples (FBM)

Human Fetal Bone Marrow cells (FBM) were processed by Sorcha O'Byrne in the laboratory of Irene Roberts and Andi Roy. Second trimester FBM cells were used for H3K79me3 ChIP seq experiments and were attained through the Human Developmental Biology Resource (www.hdbr.org). FBM cells were extracted by repetitive flushing of fetal long bones with DMEM. FBM cells were then red cell depleted by density gradient separation with Ficoll-Paque Premium (GE Healthcare) to isolate Mono-Nuclear cells (MNCs). MNCs were then fixed as per the ChIP seq protocol and stored at -80ºC.

2.1.4 Patient-derived MLL-AF4 primograft cells

Patient-derived MLL-AF4 primograft cells L826 patient cells were obtained from Newcastle Haematological BioBank and are covered by a generic approval given by the Newcastle & North Tyneside Ethics Committee (REC reference number: 07/H0906/109+5). The mouse transplantation experiments are covered by a Home Office Project licence PPL 60/4552. L826 cells were initially derived from a diagnostic ALL patient sample, which were engrafted into NSG mice recipients prior to isolation from the spleen. Cells were thawed and maintained in RPMI with 20% serum before being fixed for ChIP seq experiments.

2.1.5 Mycoplasma testing

All cell lines generated or obtained were tested for mycoplasma using the Lonza MycoAlert mycoplasma detection kit. All cell lines used in this thesis were mycoplasma free.

Table 2-2. Cell line information

Cell line	**Species**	**Sub-type/phenotype**	**Media**	**Supplements**
Embryonic Stem Cells	Mouse	E14, TOT2N variant (Blackledge *et al*., 2014)	DMEM	10% FBS, 1% GlutaMax, 1% Penicillin (100Units/ml), Streptomycin (100Units/ml), 1% LIF (made in house), β-mercaptoethanol
SEM	Human	MLL-AF4 Pre-B cell ALL	IMDM	10% FBS, 1% GlutaMax
MV4;11	Human	MLL-AF4 Monocytic AML	IMDM	10% FBS, 1% GlutaMax
RS4;11	Human	MLL-AF4 Pre-B cell ALL	RPMI	10% FBS, 1% GlutaMax
MLL-AF4 primograft cells	Human	MLL-AF4 ALL	IMDM	10% FBS, 1% GlutaMax

2.1.6 Freezing and recovering cells

All cell types were frozen down at a density of $1x10^7$ in 10% Dimethyl-Sulphoxide (DMSO Hybri-MAX®, #D2650, Sigma, Life Science). Cells were spun at 1000rpm for 5 minutes and resuspended in warm media plus 10% DMSO and aliquoted into freezing vials. Cells were transferred to a freezing unit and stored at -80°C until being stored long term in liquid nitrogen cryo-storage. Upon recovery, cells were transferred from liquid nitrogen on dry ice and thawed

at room temperature. Cells were then spun at 1000rpm for 5 minutes to remove DMSO and resuspended into the necessary volume of media and incubated.

2.2 RNAi and cell transfection methods

2.2.1 siRNA knockdown assays using electroporation

siRNA knockdown assays were performed using a rectangle pulse EPI 2500 electroporator (Fischer Heidelberg, Germany (Martinez et al. 2004)), $5x10^6$ cells at 350V for 0.1 seconds in a 2mm cuvette. All siRNA knockdown assays carried out in this thesis were upon SEM cells using 2µg of siRNA (Table 2.3). Following electroporation, incubation at room temperature for 15 minutes was carried out. The cells were then transferred back into 5ml of warm IMDM (supplemented with 10% FBS and 1% GlutaMax). Cells were then cultured for 48 hours. After 48 hours, cells were counted and the electroporation was repeated. Following an additional 48 hours, cells were harvested and processed for downstream applications.

Table 2-3. Specific siRNAs used

siRNA target	**Company/Catalogue number**
DOT1L	Ambion/s39010/s39012
AF9	Ambion/s8826/s8828
ENL	Ambion/s8820/s8822
PROM1	Ambion/s16879/s16880
Negative targeting control	Ambion/4390847

2.2.2 Lipofectamine transfection and generation of stable mESC lines

mESCs were grown to 60% confluency in a 10cm dish. Media was removed and the cells were washed using PBS. Cells were replenished with 9ml of fresh media. 28µg of DNA (as per the lipofectamine 2000 protocol) and 80µl of lipofectamine 2000 (#11668-019, Invitrogen, Life Technologies) were mixed together with 1ml of optimem. This mixture was incubated for 5 minutes at room temperature. Following incubation, the mixture was gently pipetted onto mESCs. Cells were incubated for 48 hours and were either harvested for downstream applications or selected to make a stable cell line. For stable selection, 1µg/ml of puromycin (Sigma) was added to the cells. After 24 hours, media and puromycin were replenished and cells were washed. Puromycin selection was continued for 7-10 days until colonies formed. Colonies were picked and resuspended into 150µl media (+1µg/ml puromycin) in a gelatinised 96-well dish. Cells were passaged and expanded to a 6-well dish. Western blot analysis was performed to verify the expression of the integrant.

2.2.3 Generation of inducible shRNA SEM cell lines

SEM cells were lentivirally transduced using spinoculation with inducible shRNA plasmid. This work was carried out by Ryan Beveridge (funded by the BRC). Following 24 hours, cells were selected for using 1µg/ml puromycin. Cells were plated onto methylcult media (H4100; STEMCELL technologies) for 14 days and colonies were picked and expanded. Following expansion, clones were induced with 2µg/ml doxycycline followed by FACs analysis to determine GFP expression, and processed for downstream applications.

2.3 EPZ-5676 assays

2.3.1 EPZ-5676 treatment

Cells were seeded at $0.2x10^6$ cells/ml in 50ml of media. Cells were treated with EPZ-5676 (Epizyme) 0.5µM, 1µM and 2µM or with 0µM (DMSO only) control. Cells were grown for 7 days, with a change of fresh media and EPZ-5676 at day 3 and 6, where the cells were counted and split to $0.5x10^6$cell/ml and $0.7x10^6$cells/ml, respectively. At day 7, EPZ-5676 treated cells were harvested and processed for downstream applications.

2.3.2 Cell viability assay following EPZ-5676 treatment

Cells were assayed for viability using the Cell Titre Glo 2.0 kit (Promega G9241) as per the manufacturer's protocol. Cells were grown as indicated in section 2.3.1. $1x10^6$ cells were aliquoted into a 96 well opaque plate. An equal volume of Cell Titre glo reagents was added to the wells. A medium only well was used as a background control. Reagent plus cells were mixed and incubated at room temperature for 10 minutes. Following this, the luminescence was read and analysed.

2.3.3 Colony assay coupled with EPZ-5676 treatment

SEM cells were treated with EPZ-5676 for 7 days (section 2.3.2). At day 7, SEM cells were plated at a density of $1x10^3$ cells, in triplicate in Methylcult media (supplemented with 20% FBS). All colonies were grown for 14 days (37°C, 5% CO_2) and counted.

2.4 Cloning methods

2.4.1 Polymerase Chain Reaction (PCR)

All PCR used in the cloning part of this project was done using KAPA HiFi HotStart mix. 10ng plasmid DNA was mixed with 1.5µL 10mM forward/reverse primers and 12.5µL 2X KAPA HiFi HotStart Ready Mix and rest with water to a total of 25µl. This was then ran using the thermos cycler (T100 Thermal Cycler, Bio-RAD) (Table 2.4).

Table 2-4. PCR amplification programme

Step	**Temperature (°C)**	**Time (min)**	**Cycles**
1	95	3.00	1
2	98	0.20	-
3	65	0.15	-
4	72	1.00	Repeat step 2-4 x30
5	72	2.00	1
6	4	Hold	-

This programme is a rough guide to what is generally used but can vary depending upon the template being amplified and the primers used. Depending upon downstream application, PCR products were purified either using the Qiagen PCR purification kit (Cat No: 28104) or Qiagen gel extraction kit (Cat No: 28104) as per the manufacturer's protocol.

2.4.2 Restriction enzyme digestion

Depending upon the restriction enzyme used and company (either NEB or Thermo scientific) conditions for each reaction varied to follow the manufacturer's protocol. Usually, 1µg of DNA was used in the reaction. The correct size fragment was isolated using a TAE agarose gel of an appropriate percentage and purified using Qiagen gel extraction kit. To ligate a vector and insert following restriction digest, T4 DNA ligase (400,000units/mL, #M0202S, NEB) was used overnight at 16°C. Different ratios of Vector:Insert were used depending upon the size of each.

2.4.3 Ligation Independent cloning (LIC)

The vector was initially digested with restriction enzyme BaeI (Table 2.5).

Table 2-5. LIC digestion reaction mix

Component	**Amount**
LIC vector	5µg
10x cutsmart buffer	1x (5µl)
BaeI (NEB #R0613L)	1µl
SAM (NEB #B9003S)	1µl
MQ	Up to 50µl

The reaction was incubated at 25°C overnight and supplemented with an extra 1µl of both BaeI and SAM during this incubation. The correct size band was then excised using Qiagen gel extraction kit. The insert was PCR amplified and purified using primers containing 15bp LIC sites, complementary to sites flanking the BaeI-digestion sites on the LIC vector. The LIC reaction was then carried out on both vector and insert (Table 2.6).

Table 2-6. LIC reaction mix

Vector reaction	**Volume (µl)**	**Insert reaction**	**Volume (µl)**
T4 DNA polymerase (NEB#M0203L)	3	T4 DNA polymerase	3
dGTPs (Thermo#R0161)	2	dCTPs (Thermo #R0151)	2
Buffer 2.1 (NEB#B7202S)	5	Buffer 2.1	5
Vector	10	Insert	10
Water	25	Water	25
	50µl total		50µl total

This reaction was carried out for 30 minutes at 22°C followed by 20 minutes at 75°C for heat inactivation. This reaction uses the 3' to 5' exonuclease activity of T4 DNA polymerase to give rise to complementary overhangs between vector and insert. The addition of dGTPs or dCTPs prevents the exonuclease activity from chewing back the DNA more than the 15bp overhang. The fragments are then purified using Qiagen PCR purification kit. The vector and insert were then incubated together using different molar ratios at room temperature for 30 minutes. The ligation mix was then transformed into chemically competent bacteria.

2.4.4 Site directed mutagenesis

Site directed mutagenesis was carried out for on TetR-AF9 plasmid following NEB Q5 site directed mutagenesis protocol (Cat. no. E0554).

2.4.5 Bacterial transformations

High transformation efficiency *E.coli* HST08 strains (Stellar Competent Cells, #636763, Clontech Laboratories Inc.) were thawed on ice. 5µL DNA from ligation reactions was added to 50µL cells and incubated on ice for 30 minutes. Maxiprep or miniprep DNA was used at much lower concentrations. The bacteria were heat-shocked at 42˚C for 45 seconds and immediately incubated on ice for 2 minutes. 450µl of SOC media was added to the bacteria and this was incubated at 37°C for 1 hour at 220rpm. 50µl of bacteria was spread onto agar plates containing 50µg/ml of the appropriate antibiotic. Plates were then incubated at 37°C overnight. Colonies were then pick and grown and DNA was purified using either the Qiagen miniprep or maxiprep kit as per the manufacturers protocol.

2.5. Protein methods

2.5.1 Whole cell extracts

Whole cell extracts were performed upon $1x10^6$ cells. Cells were washed using PBS and resuspended in 2x NuPAGE® LDS Sample Buffer (NP0008). Samples were heated at 95°C for 5 minutes, cooled and sonicated using a Bioruptor (Diagenode) at high frequency, 30s on/30s off for 5 minutes. Samples were stored at -20°C for downstream applications, such as western blotting.

2.5.2 Salt soluble cell extraction

$1x10^6$ cells were washed with PBS and pelleted. Cells were resuspended in 50µl of BC300 (20mM Tris HCl pH 7.5; 20% glycerol; 300mM KCl; 5mM EDTA) + 0.5% NP40 + Protease inhibitor cocktail (Roche) and incubated on ice for 30 minutes. Cells were then spun down at 13000rpm for 10 minutes at 4°C. 50µl of 2 x NuPAGE LBS sample buffer plus 0.1% β-mercaptoethanol was then added to the supernatant. The pellet was retained for histone acid extraction protocol if required.

2.5.3 Histone acid extraction

The pellet was retained (Section 2.5.2) from $1x10^6$ cells. 50µl of 0.4M HCl was added to the pellet and incubated on ice for 30 minutes. This was spun at 13000rpm for 5 mins at 4°C. 1ml ice cold acetone was added to the supernatant and stored at -20°c overnight. The sample was then spun at 13000rpm, 5 minutes, 4°C and the acetone removed. The pellet was then washed x2 with ice cold acetone and spun down 13000rpm, 5minutes, 4°C. The pellet was then airdried at room temperature and resuspended in NuPAGE LBS sample buffer plus 4mol Urea to aid resuspension. This was then stored at -20°C or ran immediately on an SDS-PAGE gel.

2.5.4 SDS PAGE

Pre-cast 4-12% Bis-Tris gels (Novex, Life Technologies) were used for proteins up to 200kDa with exception of histone protein detection where 12% Bis-Tris gels were used. 1X MOPS SDS Running Buffer (NuPAGE® MOPS SDS Running Buffer (20X), #NP0001, Novex, Life Technologies) was used for detection of non-histone proteins. 1x MES SDS Running buffer was used for the detection of histone proteins (NuPAGE MES SDS Running buffer (20X) #NP0002, Novex, Life technologies). Gels were run at 180V for approximately 1 hour. The

protein molecular weight ladder used was the Precision Plus Protein Kaleidoscope (Biorad #110375).

2.5.5 Western blotting

Proteins were transferred to a PVDF membrane (Immobilon®-P Transfer Membrane, #IPVH00010, EMD Millipore) using a wet transfer setup. Gel and membrane were sandwiched between filter paper and 2x sponges and submerged in a tank filled with transfer buffer (192mM glycine; 25mM Trizma® base; 10% methanol). 100V was supplied to the tank for 1 hour at 4°C. For DOT1L western blotting, 0.05% SDS was added to the transfer buffer to encourage optimal transfer. The transfer was also performed overnight at 25V, 4°C.

Following transfer, membranes were incubated in Ponceau stain (SIGMA-ALDRICH P-7170-1L) for 5 minutes and washed x3 in 100% methanol. The membrane was then dried, re-hydrated and blocked in 5% milk for 1 hour. Membranes were then incubated with antibody overnight at 4°C. The following day the membrane was washed for 5 minutes x3 in 0.05% TBS-Tween and subsequently incubated in 5% milk with the appropriate secondary antibody. The membrane was then washed for a further 5 minutes x3 with 0.05% TBS-tween and then probed using ECL. (ECL Prime Western Blotting Detection Reagent, #RPN2232, Amersham, GE Healthcare) Solutions A & B were mixed in a 1:1 ratio and added over the membrane and incubated for 5 minutes. Excess solution was removed and the membrane was developed using chemiluminescence film (Amersham Hyperfilm ECL, #28906837, GE Healthcare) and Xograph Imaging.

Table 2-7. Antibody information

Antibody name	**Ordering information**	**Species**	**Working concentrations (W.B = western blot)**
DOT1L	Bethyl A300-593A/594A	Rabbit	1:1000 W.B and 1:5000 ChIP
AF9	Bethyl A300-597A	Rabbit	1:5000 W.B and ChIP
ENL	Bethyl A302-267A	Rabbit	1:5000 W.B and ChIP
AF17	Bethyl A302-198A	Rabbit	1:5000 ChIP
PAF1	Bethyl A300-172A	Rabbit	1:5000 W.B and ChIP
TetR (FS2)	Klose lab	Rabbit	1:1000000 W.B and 1:5000 ChIP
AF4	Abcam ab31812	Rabbit	1:5000 ChIP
AFF4	Bethyl A302-538A	Rabbit	1:5000 ChIP
H3K79me3	Diagenode C15410068	Rabbit	1:10000 W.B and 1:5000 ChIP
H3K27ac	Diagenode C15410196	Rabbit	1:10000 W.B and 1:5000 ChIP
H3K9ac	Abcam ab4441	Rabbit	1:10000 W.B and 1:5000 ChIP
H3K9me2	Abcam ab1220	Mouse	1:10000 W.B and 1:5000 ChIP
H3K9me3	Abcam ab8988	Rabbit	1:10000 W.B and 1:5000 ChIP
CBX8	Bethyl A300-882	Rabbit	1:5000 ChIP
TBP	Bethyl A301-229A	Rabbit	1:5000 ChIP
RNAPII serine 2 phos	Abcam ab5095	Rabbit	1:2000 ChIP
H3	Abcam ab1791	Rabbit	1:1000000 W.B

H4	Abcam ab10158	Rabbit	1:1000000 W.B
GAPDH	Bethyl A300-641	Rabbit	1:10000 W.B and 1:5000 ChIP
Anti-Rabbit HRP (2°)	Sigma A6667	Goat	1:10000 W.B
Anti-Mouse HRP (2°)	Sigma A4416	Goat	1:10000 W.B

2.6 Chromatin Immunoprecipitation (ChIP)

2.6.1 Fixation

Cells were counted, spun down (1000rpm, 5 minutes) and washed with PBS. All ChIP qPCR experiments were carried out on $1x10^7$ cells unless indicated otherwise. For ChIP seq experiments $4x10^7$ SEM cells were used (fixed as separate $1x10^7$ cell pellets). Cells were then either double or single (histone modifications only) fixed.

For double fixation, cells were fixed in 1mL 2mM disuccinimidyl glutarate (DSG, #80424-50MG-F, Sigma) and rotated at room temperature for 30 minutes. This was followed by centrifugation (13000rpm, 1 minutes) and removal of DSG. The pellet was then resuspended in 1mL 1% formaldehyde (FA) and rotated at room temperature for 30 minutes. For single fixation, cells were fixed in 1mL 1% FA and rotated at room temperature for 10 minutes. Fixed material was spun (13,000rpm, 1 minute) and the pellet was rinsed in PBS before being frozen on dry ice.

2.6.2 Day 1

Fixed pellets were thawed on ice and resuspended in 100µL SDS lysis buffer (50mM Tris-HCl pH8.1; 10mM EDTA; 1% SDS) plus 1:1000 protease inhibitor cocktail and syringe-passage through a 27-gage (27G) needle. Lysed sample was transferred to a small Covaris tube (microTUBE AFA Fiber Pre-Slit Snap-CAP 6x16mm, #520045, Covaris Inc) for use in a

Covaris sonicator (S220, Covaris Inc.). Tubes were loaded into the degassed water at 4˚C and run using the Covaris SonoLite software. The parameters used were: Duty cycle = 10% Intensity = 5 Cycles / burst = 200 Duration = 300s. For single-fixed material this setup was run once whereas for double-fixed material this was done twice. After sonication, samples were transferred to Eppendorf tubes and spun at 13,000rpm for 10 minutes at 4˚C.

Supernatant was retained and diluted 1:10 in ChIP dilution buffer (16.7mM Tris-HCl pH8.1; 167mM NaCl; 1.2mM EDTA; 0.01% SDS; 1.1% Triton X-100 (#X100-100ML, Sigma)) (CDB). Protein A/G Dynabeads (#10001/4D, Novex, Life Technologies) were resuspended in CDB and added to the chromatin in CDB for pre-clearing. This was incubated at 4˚C for 30 minutes, with rotation. After pre-clearing, samples were spun at 3,200rpm for 2 minutes and chromatin was split into 1mL aliquots with 2, aliquots of 50µL being retained for input material and for sonication efficiency testing. Antibodies were added to the 1mL aliquots and incubated with rotation at 4˚C overnight (See table 2.7 for antibodies and dilutions).

2.6.3 Day 2

10µL of Protein A/G Dynabeads (resuspended in CDB) were added to each 1mL ChIP aliquot and incubated at 4˚C for 3-5 hours, with rotation. Beads were washed twice in RIPA wash buffer (50mM HEPES-KOH pH7.6; 500mM LiCl; 1mM EDTA; 1% NP-40; 0.7% Na-Deoxycholate) using a magnetic stand, the material was transferred into a new tube following the first wash. Beads were washed in TE + 50mM NaCl and spun at 3200rpm for 2 minutes. Following this, supernatant was removed and beads were eluted in 105µL elution buffer and incubated at 65˚C for 30 minutes. Beads were spun at 13000rpm for 1 minute and 100µL eluted material was transferred to new Eppendorf tubes. Samples were then incubated at 65°C overnight for de-crosslinking. Input samples were diluted to 100µL and put at 65˚C overnight.

2.6.4 Day 3

1µL RNAse A (10mg/mL, #EN0531, Thermo Scientific) was added to each sample and incubated at 37˚C for 30 minutes. 1µL Proteinase K (20mg/mL, #EO0491, Thermo Scientific) was added to each sample and incubated at 45˚C for 30 minutes. DNA was extracted using the Qiagen PCR Purification kit and eluted in 200µl water for ChIP qPCR. For ChIP seq experiments, DNA was eluted into 51µl and samples were quantified using the qubit DNA quantification protocol as per manufactures protocol. Samples were stored at -20˚C.

2.6.5 RT-qPCR

ChIP material 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 of DNA was added to 20µL reaction mix (12.5µL Fast SYBR® Green Master Mix (#4385612, Applied Biosystems); 0.25µL forward primer (10µM); 0.25µL reverse primer (10µM); 7µL water) in a single well of a 96-well PCR plate (MicroAmp Fast Optical 96-well Reaction Plate, #4346906, Applied Biosystems). For TaqMan primers, 5µL DNA was added to 15µL reaction mix (10µL TaqMan® Fast Advanced Master Mix (#4444557, Applied Biosystems); 1µl primer/probe mix (Invitrogen) 4µL water) in a single well. Plates were sealed with film (MicroAmp Adhesive Film, #4311971, Applied Biosystems) and spun at 1000rpm for 1 minute 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 (ΔΔCT; Fast; DNA) (Table 2.8).

Table 2-8. qPCR protocol

Temperature (^{0} C)	**Time (s)**	**Cycles**
95	20	1 (initial denaturation)
95	3	40 (Denaturation, Annealing and extension)
60	30	
95	15	1 (Melt curve stage, optional)
60	60	
95	15	
60	15	

Primers for RT-PCR were designed and ordered using IDT primer quest tool (SYBR green only). Primer lengths of between 20-25bp and a melting temperature of between 55-60°C were chosen in addition to a product size of 100bp. ChIP qPCR input material was used to validate the efficiency of primers before use. Melt cure analysis of all new primers was carried out to validate no off-target binding and primer dimers. Primer sequences used in this thesis are available upon request.

2.6.6 ChIP rx

ChIP rx was performed exactly as described for ChIP seq experiments apart from the addition of Drosophila melanogaster S2 cells (kindly provided by the Fulga lab) during the lysis step of the protocol. S2 cells were fixed as described above for ChIP qPCR and ChIP seq applications and added to the human samples prepared for ChIP seq in a 1:4 ratio during the lysis step, before sonication. The normal ChIP seq protocol was followed from this point onwards. ChIP

rx was performed on SEM cells treated with EPZ-5676. ChIP rx was validated using dm3 specific primers (available upon request).

2.6.7 Preparation of libraries for ChIP seq and ChIP rx

Following QC of the ChIP experiment (including qubit quantification and qPCR) next generation sequencing libraries were made. The NEBNext® Ultra™ DNA Library Prep Kit for Illumina® (Cat no. E7370L) and NEBNext Multiplex oligos for illumina (Cat. no. E7335L/E7500L) were used for preparing all ChIP seq and ChIP rx libraries, as per manufactures instruction.

2.7 RNA methods

2.7.1 Total RNA Extraction

RNA was extracted using $1x10^6$ cells. To extract whole cell RNA the Qiagen RNeasy Mini Kit (#74106) was used as per the manufacturer's instructions. RNA was quantified using the NanoDrop (Thermo scientific) and stored at -80˚C.

2.7.2 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 5 minutes before being snap-cooled on ice for 2 minutes. When working with multiple samples, the same amount of RNA was inputted into the cDNA reaction. The mixture was added to 10µL reaction master mix (Table 2.9). The reaction was run using a thermocycler (BioRad) (Table 2.10).

Table 2-9. RT-PCR reaction mix

Component	Amount
5x First strand buffer (Invitrogen #02321)	4µl (1x)
0.1M DTT (Invitrogen #00147)	2µl
RNaseOUT (40 units/µl Invitrogen #51535)	1µl
Superscript III reverse transcriptase (200 units/µl Invitrogen #56575)	1µl
MQ water	2µl (10µl total)

T**able 2-10.** RT-PCR protocol

Step	**Temperature (°C)**	**Time (min)**	**Cycles**
1	95	3.00	1
2	98	0.20	-
3	65	0.15	-
4	72	1.00	Repeat step 2-4 x30
5	72	2.00	1
6	4	Hold	-

The PCR reaction mixture was diluted with 200µL water and cDNA was stored at -20˚C for downstream applications.

2.8 Nascent RNA extraction

$1x10^8$ cells were treated with DMSO, 0.5µM, 1µM and 2µM of EPZ-5676 and harvested on day 7 (section 2.3), for nascent RNA extraction.

2.8.1 Trizol RNA preparation

50µM of 4-thiouridine (Sigma, T4509 4-thiouridine, 100mg) was added to the cells in media. Cells were incubated for 1 hour to allow labelling of nascent transcripts to occur. Cells were spun down, 1000rpm for 5 minutes. Media was removed and 10ml TriZol (Thermo Scientific #15596026) was added to the cell pellet, shaken vigorously and incubated at room temperature for 5 minutes. 2ml Chloroform was added to the TriZol mixture and vigorously shook for 15 sec, followed by a 3 minute incubation at room temperature. Samples were centrifuged at 13,000xg for 15 minutes at 4°C. The aqueous upper phase was transferred into a fresh 15ml falcon and 1x volume of isopropanol was added. Samples were mixed and incubated at room temperature for 10 minutes, and spun at 13,000xg, 10 minutes at 4°C. The supernatant was removed, and the pellet was washed with 10ml of 75% ethanol. Samples were centrifuged 13,000 g for 10 minutes at 4°C and the ethanol was removed. Samples were resuspended in 87ul of RNase free water and incubated at 65°C for 10 minutes. Samples were treated with the TURBO DNA-free™ Kit (Thermo-fisher #AM1907) following the manufactures instructions.

2.8.2 Biotinylation assay

A Biotinylation reaction mix was generated (Table 2.11).

Table 2-11. Biotinylation reaction mix

Component	**Amount**
RNA	100µg
1mg/ml Biotin-HPDP (pierce, 50mg EZ-link Biotin-HPDP #21341)	200µl
10x Biotinylation buffer (100mM Tris pH 7.4;10mM EDTA)	100µl
RNAse free water	700µl

The reaction was incubated at room temperature for 1.5 hours with rotation. Following this, an equal volume of Chloroform/Isoamylalcohol (24:1) was added, mixed vigorously and incubated for 2–3 minutes at room temperature in phase lock gel heavy tubes (Eppendorf 0032 005.152). Samples were centrifuged at 130000xg for 5 minutes. This step was then repeated and then transferred into a 1.5ml tube. Precipitation of RNA was carried out by adding 1:10 reaction volume of 5M NaCl, 1x volume isopropanol and centrifuged at 13 0000g for 20 minutes at 4°C. Supernatant was removed and pellets were washed with 75% ethanol and resuspended in 100µl 1x TE.

2.8.3 Isolation of labelled RNA

RNA samples were heated to 65°C for 10 minutes. Samples were mixed with 100µL streptavidin beads (µMacs Streptavidin Kit, Miltenyi, Order No. 130-074- 101) and incubated with rotation for 15 minutes. uMacs columns were then equilibrated on a magnetic stand using wash buffer (100mM Tris pH 7.5; 10mM EDTA; 1M NaCl; 0.1% Tween20). RNA samples were then added directly to the column and flow through discarded. Columns were washed 3x 900µl, 65°C wash buffer and then 3x with room temperature wash buffer. RNA was eluted into 700µl buffer RLT by adding 10µl 100mM DTT. This step was repeated. The RNeasy MiniElute cleanup kit (#74204) was used as a final step as per the manufacturer's instructions and eluted with 40µl RNAse free water.

2.8.4 Nascent RNA sequencing library prep

Nascent RNA seq libraries were prepared using NEBNext Ultra directional RNA library prep kit for illumina and NEBNext Multiplex olgios for illumina (Cat. no. E7335L/E7500L) (NEB #E7530), following manufacturer's instructions.

2.9 ATAC seq

The ATAC seq protocol was adapted from (Buenrostro et al. 2013). 50 000 SEM cells were harvested and washed in PBS and resuspended gently in 50µl cold lysis buffer (10mM Tris-HCl, pH 7.4; 10mM NaCl; 3mM MgCl2; 0.1% IGEPAL CA-630). Cells were spun down immediately at 500xg for 10min at 4°C. The supernatant was discarded and the pellet was set on ice and resuspended in a transposase reaction mix (Table 2.12).

Table 2-12. Transposase reaction mix

Component	**Amount**
2x TD Buffer (Illumina Cat #FC-121-1030)	25µL
Tn5 Transposes (Illumina Cat #FC-121-1030)	2.5µL
Nuclease Free H2O	22.5µL

The reaction was incubated at 37°C for 30 minutes. DNA was purified using a Qiagen MinElute Kit as per manufactures instruction. DNA was eluted in 10µL 10mM Tris pH 8.

2.9.1 ATAC seq library preparation

The DNA fragments were amplified in a PCR reaction (Table 2.13). The amplified library was purified using Qiagen PCR clean up kit and eluted in 20µL water.

Table 2-13. ATAC library prep reaction mix

Components	**Amount**
Transposed DNA	10µl
Customised universal primer	2.5µl
Customised index primer	2.5µl
2x NEBNext high-fidelity 2x PCR master mix	25µl
Nuclease free water	10µl

Table 2-14. ATAC library prep PCR amplification protocol

Step	Time (min)	Temperature (ºC)
1	5.00	72
2	0.30	98
3	0.10	98
4	0.30	63
5	1.00	72 (Repeat steps 3-5 x12)
6	Hold	4

2.10 Capture-C

For in-depth description of this protocol please refer to (Davies et al. 2016).

2.10.1 3C library preparation

$2x10^7$ SEM cells were fixed with 2% formaldehyde for 10 minutes with rotation. The fixed cells were then resuspended in lysis buffer and frozen at -80°C. The cell pellet was then thawed and lysis buffer removed. The pellet was resuspended in DpnII buffer and dounce homogenisation was performed with a total of 50 strokes. The homogenate was then spun down and supernatant discarded. The pellet was resuspended in DpnII buffer and digestion reactions were set up with the DpnII enzyme. Necessary controls were also set up with undigested chromatin. Digests were incubated overnight at 37°C, shaking. Additional DpnII was added twice throughout the digest. Following this, a ligation reaction was then carried out, adding DNA ligase and ligation buffer to the reaction and overnight incubation. The material was then de-crosslinked (65ºC, overnight) and then DNA was extraction using a phenol-chloroform extraction method.

2.10.2 Addition of Illumina sequencing adapters

10-12µg of 3C library was sonicated using the following settings: Duty cycle: 10%, Intensity: 5, Cycles per burst: 200, Time: 360 seconds. This was followed by an Ampure XP SPRI bead clean at a 1:1.9 ratio of material to beads as per manufacturer's protocol. Libraries were then made using the NEBNext DNA library preparation kit, as per the manufacturer's protocol.

2.10.3 Capture step

Capture probe oligonucleotides (see table 2.15 for list of capture oligos used in this experiment) are reconstituted to 2.9nM total concentration. 1-2µg of differentially indexed samples were mixed at a ratio of 1:1 to generate a pooled library. The hybridisation reaction was then performed. 1-2µg of pooled indexed library was mixed with 5µg of COT DNA (Nimblegen), 1nmol of TS-HE Universal Oligo and 1nmol TS-HE, all supplied from the Nimblegen kit. This mixture was vacuum centrifuged at 50ºC until all liquid was evaporated. To this, 7.5µl 2x hybridisation buffer and 3µl hybridisation component was added (per library in the pooled sample). This was then denatured at 95ºC for 10 minutes and added to Biotinylated capture oligos at 47ºC for 72 hours. Following 72 hours material was washed using Nimblegen wash kit. Streptavidin beads were prepared and washed and added to captured material (100µl per library in the pooled sample). This was mixed at 47ºC, 600rpm for 45 minutes. This material was then washed using Nimblegen wash buffers. DNA was then amplified (on the beads) using KAPA HiFi HotStart ready mix, followed by an Ampure XP SPRI bead clean up.

Table 2-15. List of oligo probes and positions used in Capture-C experiments

Probe Name	Fragment Position (hg19)		
NA	Chromosome	Start	Stop
JUN	1	59249099	59250691
GADD45A	1	68150066	68151771
SPI1	11	47399928	47400398
FOS	14	75744773	75746081
BCL2	18	60986641	60988918
ELL2	5	95297309	95297978
MYB	6	135501733	135502770
MYC	8	128748253	128748439
MSL3	X	11774388	11776889
PROM1_SI	4	16077645	16078021
PROM1_LI	4	16084464	16085645
LMO4_1	1	87793884	87795360
SMYD2_1	1	214454064	214456235
JMJD1C_2	10	65226055	65226262
JARID1A_1	12	498164	499589
FLT3_1	13	28674038	28675050
SUZ12_1	17	29058515	29059304
MED13_1	17	60141972	60142862
MED13_2	17	60142859	60143711
BCL11A	2	60778931	60782448
BAZ2B_2	2	160472165	160472912
EP300_1	22	41487595	41490012
MBNL1_2	3	151986212	151986427
MBNL1_3	3	151986424	151989271
TAPT1_1	4	16226868	16228644
TET2_2	4	106066532	106067212
TET2_1	4	106067209	106069662
MEF2C_2	5	88180024	88180319
SUPT3H_1	6	45345424	45345962
MYB_1	6	135501733	135502770
ARID1B_2	6	157099820	157100601
CDK6_1	7	92462911	92463843
EZH2_2	7	148580112	148580317
FUT10_1	8	33330317	33330678
ASH2L_1	8	37961940	37963209
OGT_1	X	70752518	70753336

2.10.4 Double Capture

Up to 2µg of DNA from the first capture was used and the protocol from the first capture was repeated. This time only a single volume of hybridisation reaction is used. The reaction was incubated at 47ºC for 24 hours. This was followed by a second round of PCR amplification and bead clean up. The final Capture-C library was then quantified using KAPA library quantification kit and stored at -20ºC for sequencing.

2.11 Sequencing

Prior to sequencing all libraries were quality controlled by analysis on a 2200 TapeStation instrument (Agilent technologies) as per the manufactures instructions to check for amplification and lack of adapter dimer. Libraries were quantified using KAPA Universal Complete Kit (KK4824, Kappa Biosystems) as per manufactures instruction. All libraries were sequenced using the NextSeq 500 platform (illumina), using paired end reads. For ChIP seq and ATAC seq, a 75 cycles high kit was used. For Nascent RNA seq a 150 cycles high kit was used. For Capture-C, a 300 cycles high kit was used.

2.12 Bioinformatics analysis of ChIP seq, ChIP rx and ATAC seq data

2.12.1 Mapping and quality control

Pair-ended ChIP seq/ATAC seq data was mapped using a pipeline developed by Jelena Telenius. Version 10.1 of the "DnaseANDChIP" pipeline was used. In summary, the pipeline performs quality control checks of ChIP seq fastq files, trims adaptors and merges reads if required. Reads were then mapped to hg19 (or hg19 and dm3 for ChIP rx seq) and duplicates were removed. It employs the following modules: samtools/0.1.19, bedtools/2.17.0, fastqc/0.10.1 bowtie/1.0.0, cutadapt/1.2.1 and trim_galore/0.3.1. ATAC seq data was mapped

against hg19 with the same parameters except the distance between paired end reads was set to 1000bp using different illumina adapters.

2.12.2 Visualisation of data

The HOMER tool (vs 4.7) (Heinz et al. 2010) was used for generation of bigwig files for visualisation in the UCSC genome browser. For HOMER to process the data, the bam file must be converted into a TagDirectory. This was performed using the makeTagDirectory tool. To visualise the data on UCSC, bed graph or bigwig files were generated from the Tag directories. This was done using the makeUCSCfile tool or makebigWig tool. Importantly, all files were normalised to $1x10^7$ million reads, except from ChIP rx files which were normalised using calculated normalisation factor.

2.12.3 Determination of the Normalization Factor in ChIP rx.

For a detailed description of the basis and derivation of the ChIP rx normalisation factors, please refer to Orlando *et al*., 2014. In brief, the *D.melanogaster* normalisation factor is calculated based upon the number of reads mapping to dm3 in each sample. Because the same amount of S2 cells were added to each sample, the resulting signal should be equilibrated across experiments. The reference normalisation factor is calculated as 1/number of reads mapping to dm3 per million. This contrasts with traditional normalisation which used 1/number of reads mapping to hg19 per million.

2.12.4 Peak calling

H3K79me2/3 peak calling was performed by Ross Thorne for SEM, FBM and MLL-AF4 patient samples. This was done using the Homer package and the findPeaks command.

Furthermore, the histone modification -style histone flag was used to identify broad domains of H3K79me.

2.12.5 Nascent RNA-seq analysis

Quality control using the fastQC package was performed (http://www.bioinformatics.babraham.ac.uk/projects/fastqc), reads were aligned using STAR (Dobin et al., 2013) against hg19. Gene expression levels were quantified as number of reads using the featureCounts function from the Subread package with default parameters (Liao et al. 2014). The read counts were used for the identification of global differential gene expression between specified populations using the edgeR package (Robinson et al. 2010). RPKM values were also generated using the edgeR package. Genes were considered differentially expressed between samples if they had an adjusted p-value (FDR) of less than 0.05.

2.12.6 Capture-C analysis

Capture-C data was analysed using an in-house pipeline generated by Jelena Telenius, James Davies and Marieke Oudelaar. Capture-C probes were designed using the CapSequm tool (http://apps.molbiol.ox.ac.uk/Cap)

Chapter 3 - DOT1L complex stabilisation at MLL-AF4 gene targets

3.1 Introduction

3.1.1 Understanding protein interactions and recruitment

The concept of recruitment is often used in the field to refer to the localisation of specific proteins at specific genes. However, the use of the word "recruitment" would seem to imply that this must be a directed or deterministic process at the molecular level. It has been shown that transcription factors find binding sites via stochastic facilitated diffusion, constantly sampling the DNA (Hammar et al. 2012; Hammar et al. 2014; Elf et al. 2007; Marklund et al. 2013). This mechanism is driven by non-specific interactions which have short residency times and specific interactions which have comparatively longer residency times (Elf et al. 2007; Hammar et al. 2012). This suggests that the recruitment of a protein at a gene may be a consequence of stochastic facilitated diffusion until pausing at a specific binding site leads to a stable binding event. Therefore, to understand recruitment, it is important to consider that stable binding and increased residency time of a protein at a specific binding site is generated via the dissociation constant or K_D which can potentially be established through multiple interactions (Ruthenburg et al. 2011).

The binding strengths of protein interactions are measured *in vitro* by a dissociation constant (K_D) concentration at which a protein:protein interaction is half maximal. Higher affinity interactions are associated with a lower K_D value and lower affinity interactions are associated with a higher K_D value. Unless the affinities of different protein interactions are compared in a context dependent manner it is difficult to gain insight into the biological relevance of the in

vitro calculated K_D. One of the higher affinity interactions observed in the context of transcription is the interaction between a transcription factor and DNA, although binding affinities vary depending upon the specific transcription factor, reports have indicated that the interaction between a transcription factor and DNA can occur at pM-nM K_D (Wang et al. 2009). In comparison, protein:protein interactions are of lower affinity, an example of this is the measured affinity of the interaction between AF9 and DOT1L, which has an observed K_D of 1.6nM (Leach et al., 2015). Furthermore, the interactions between the YEATS domain of AF9 and ENL and H3 acetylated lysine residues is much weaker, with a K_D of 30-50μM (Wan et al., 2017, Andrews et al., 2016). This demonstrates the variety of different binding affinities which might occur at a gene during transcription. Therefore, it might be assumed that a TF is more stably bound, with higher affinity, at a gene compared to a protein:protein interaction at a gene. However, it has been demonstrated that multiple lower affinity interactions, may together create a higher affinity, stable binding site for a protein, a concept referred to as multivalency (Ruthenburg et al. 2011). The stable binding of a protein at chromatin then enables functionality at a gene, for example, the recruitment of DOT1L enables it to methylate H3K79 of spatially close histones.

3.1.2 Studying protein interactions

Studying protein interactions which occur at genes is challenging because transcription is a very complex process controlled by many different proteins. Therefore, to observe individual protein interactions in detail it is important to study these interactions in relative isolation of transcription, whilst still maintaining the physiological relevance which is key to understanding whether these interactions are occurring at DNA within the nucleus of cells. There are many different *in vitro* techniques which are used to analyse protein:protein interactions, and although they can be used to initially identify and test interactions *in vitro*, the interactions

which are identified lack cellular context and furthermore, it is difficult to identify whether any of the interactions are artefacts of the techniques used.

3.1.3 The Tet Repressor (TetR) system

One system which does address whether interactions demonstrate cellular context in mammalian cells *in vivo* is the TetR system. Forms of the TetR system have been used coupled with immunofluorescence, however, a novel system has now been generated in mES cells using ChIP qPCR as a detection tool (Pollex et al. 2013; Blackledge et al. 2014). This novel TetR system depends upon the interaction between the TetR and the Tet Operator (TetO). The TetO, comprising 14 TetO repeats, was integrated into the mESC genome on chromosome 8, a gene desert region. The TetO array flanked by human chromatin from chromosome 7 contains no genes, CpG islands or histone modifications (histone modifications tested include H3K4me3, H3K27me3, H3K9ac, H2AK119ub, H3K36me3, H4K20me2 and H3K9me3) and is not bound by RNAPII (Blackledge et al. 2014). A protein of interest (X) can be tethered to the TetR and stably expressed, meaning that the TetR-X fusion is stably bound at the TetO with high affinity (Figure 3.1). Therefore, it is possible, to ask which recruited protein (Y) is detectable at the TetO due to anchoring protein X. This system allows testing whether anchoring a specific protein is sufficient for the stable assembly of a complex free from the potential added complexities of transcription at a gene locus.

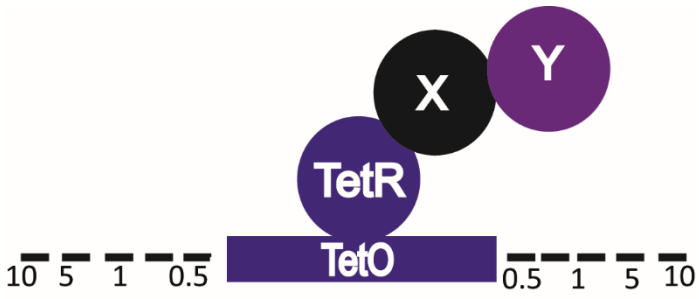

Figure 3-1. Schematic diagram of TetR system. TetR-X is bound at the TetO array and recruits protein Y.

Another important advantage of this system is that, upon doxycycline treatment, TetR-X can no longer bind to the TetO. This serves as an important control for demonstrating the recruitment of protein Y is due to the presence of TetR-X. Furthermore, it can also be used to understand the binding dynamics of the recruited proteins following the loss of the original tethered protein. To interpret TetR ChIP qPCR results, an amplified ChIP qPCR signal determines that a protein is detectable at the TetO. An absence of amplified signal determines that a protein may not be detectable at the TetO, although it is difficult to definitively rule out the absence of a protein using this method, this may be due to problems with crosslinking or antibody quality. The differences in ChIP signal observed in the different systems used may be due to variation of the ChIP protocol and therefore, the TetR system will be used in a binary fashion to determine the detectability of a protein or not.

3.1.4 Understanding DOT1L recruitment

DOT1L is a chromatin modifying protein which is involved in active transcription, however, it is unclear how DOT1L contributes to transcription (Steger et al. 2008). A key aspect in understanding the function of DOT1L in transcription is understanding which proteins it can interact with and how it is recruited to genes requiring H3K79me. DOT1L has no identified DNA binding domain. So, it is likely that other proteins present at a gene are required to

generate a high affinity binding site for DOT1L. Very little is known about how a high affinity binding site is created for DOT1L, as none of the biochemically purified complex members of DOT1L (AF9, ENL, AF10 or AF17) have been demonstrated to bind to DNA (Mohan et al. 2010). β-catenin is a DNA binding protein which has been co-purified with DOT1L (Mohan et al. 2010). However, this interaction has more recently been shown to be non-specific (Ho et al. 2013). Another context in which DOT1L has been shown to interact with a DNA binding protein is in MLL-r leukaemia, where a member of the DOT1L complex is often fused to MLL (i.e.MLL-AF9). MLL-FPs can bind to DNA and so this is one way in which DOT1L could be directly recruited to a gene, although this has yet to be confirmed *in vivo*. MLL-AF4 leukaemia, is another MLL-r leukaemia where high levels of DOT1L and H3K79me3 have been found at MLL-AF4 gene targets. Furthermore, gene regulation of MLL-AF4 gene targets is highly dependent upon DOT1L and H3K79me (Krivtsov et al. 2008). Interestingly, AF4, has not been co-purified with DOT1L (Yokoyama et al. 2010). Moreover, further *in vitro* investigation has demonstrated that AF4 and DOT1L bind to AF9 in a mutually exclusive interaction (Figure 3.3) (Leach et al. 2013). Therefore, even though MLL-AF4 leukaemias are highly dependent upon DOT1L, DOT1L is not known to interact with AF4. The high levels of H3K79me3 observed at MLL-AF4 gene targets could be explained in several ways: firstly, it is possible that MLL-AF4 is in some way increasing the activity of DOT1L or secondly, it might create a high affinity, more stabilised binding site for DOT1L. Furthermore, a combination of both mechanisms may also exist. In this chapter, it is investigated whether MLL-AF4 is responsible for the recruitment of DOT1L and what the mechanism of stabilisation may be.

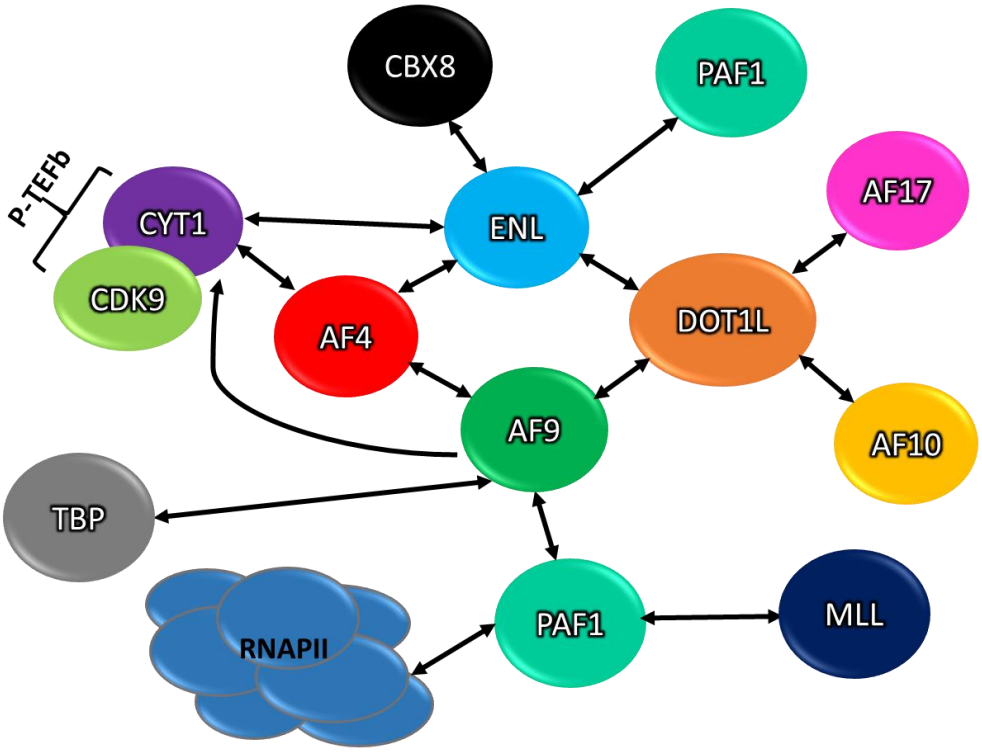

Figure 3-2. Network of in vitro biochemically purified interactions. Protein:protein interactions associated with DOT1L and DOT1L core complex members (AF9, ENL, AF10 and AF17) observed via biochemical purifications. AF4:ENL/AF9 and ENL/AF9:DOT1l have been observed as mutually exclusive interactions (Biswas et al. 2011; Mohan et al. 2010; Leach et al. 2013; Milne et al. 2010; Sobhian et al. 2010; He et al. 2011; Yokoyama et al. 2010).

As many interactions have been observed biochemically with DOT1L and MLL-AF4 *in vitro*, a network of hypothesised interactions which might occur *in vivo* can be established (Figure 3.2-3). Using this it is possible to create a framework which can be tested to understand the mechanism of DOT1L recruitment at MLL-AF4 gene targets.

3.2 Aims

1. To determine whether components of the DOT1L complex are sufficient to create a stable binding site for DOT1L *in vivo*.
2. To determine if DOT1L recruitment is sufficient for H3K79me in the absence of transcription.
3. To determine if MLL-AF4 is sufficient for DOT1L recruitment *in vivo*.

3.3 Results

3.2.1 DOT1L complex members are sufficient for DOT1L recruitment to chromatin

To test whether DOT1L complex members are sufficient for DOT1L recruitment, TetR fusion proteins were generated with AF9, ENL and AF10. Most of the work presented in this thesis concerns human MLL-AF4 and DOT1L, but the TetR system is based in mESCs. It was therefore necessary to consider whether it was appropriate to use human fusion proteins in these cells. Mouse protein sequences of DOT1L complex members were compared to the human protein sequence. All mouse complex members had above 70% identity to their human counterpart (Table 3.1).

Table 3-1. List of % amino acid identity of human vs mouse DOT1L complex members in addition to AF4 and PAF1.

Protein	% amino acid identity (Mouse Vs human)
DOT1L	84%
ENL	87%
AF9	96%
AF10	93%
AF17	89%
AF4	70%
PAF1	98%

In addition to this human AF9, ENL and AF10 are all sufficient to create leukaemia in a mouse when fused to mouse or human MLL indicating human versions of these proteins are functional in the mouse (Chen et al., 2010, Demartino et al., 2002, Krivstov et al., 2006, Somervaille et al., 2006, Takacova et al., 2012). Due to this similarity, it was reasoned that human protein:protein interactions would likely be conserved between each human protein and the endogenous mouse homologues of its binding partners. Therefore, human proteins were used

in all TetR fusions unless otherwise indicated. All TetR fusions were generated using cDNA already present in the lab or cDNA made by RT-PCR amplification from SEM RNA.

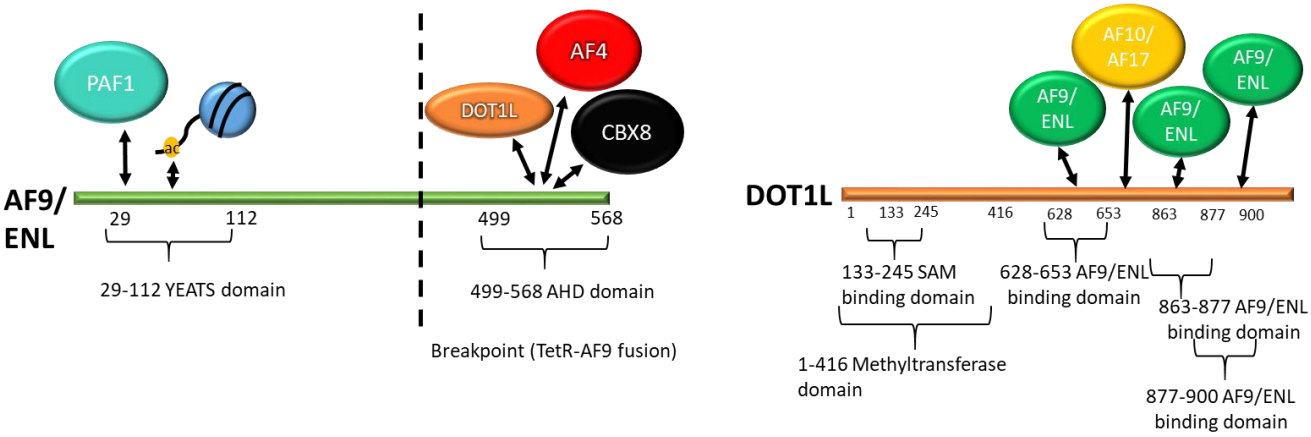

Figure 3-3. Identified biochemical interacting domains of (A) AF9 and ENL (B) DOT1L. DOT1L bind to same binding domain of AF9/ENL reandering these interactions mutually exclusive (He et al. 2011; Li et al. 2014; Wan et al. 2017; Leach et al. 2013; Kuntimaddi et al. 2015; Okada et al. 2005).

3.2.1 AF9 is sufficient for Dot1l recruitment

AF9, and its orthologue ENL, are YEATS domain proteins which share many similarities in protein domain structure (Figure 3.3). AF9 and ENL have been shown in vitro to interact with DOT1L via the ANC homology domain (AHD) (Figure 3.3). At the beginning of this project, the most widely available biochemical and structural data of a DOT1L interacting protein was AF9. Therefore, AF9 was chosen to study DOT1L recruitment initially. A TetR-AF9 fusion was synthesised using the AHD domain of AF9 and stably expressed in mESCs containing a TetO array (Figure 3.4A). The expression of TetR-AF9 was confirmed by western blot analysis (Figure 3.4B).

To test whether TetR-AF9 was bound at the TetO, ChIP qPCR for TetR (using an antibody recognising an N-terminal FS2 tag) and AF9 was carried out in the TetR-AF9 mESC cell line. TetR and AF9 were both detectable at the TetO (Figure 3.4C-D). As an important control, in

all experiments, TetR fusions were compared to a cell line stably expressing TetR-only. In this example, TetR is detectable in TetR-only cells, but AF9 is not (Figure 3.4C-D).

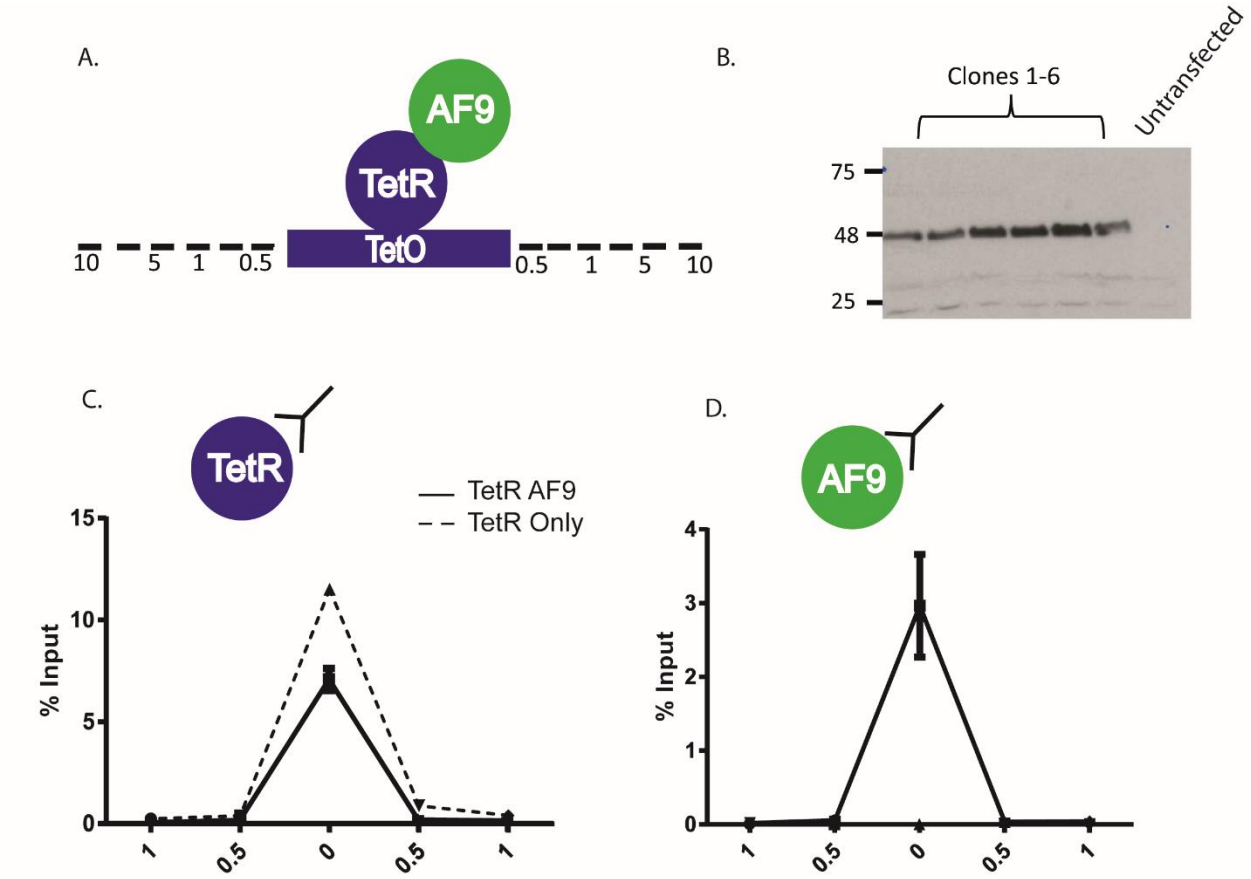

Figure 3-4. Validation of TetR AF9 system (A) Schematic of TetR-AF9 bound at the TetO. Numbers represent in Kb the distance from the TetO. (B) Western blot analysis demonstrating the expression of 6 clonal cell lines plus untransfected control (C-D) ChIP qPCR using antibody against TetR (N-terminal FS2 tag) and AF9 demonstrating TetR-AF9 bound in the TetR-AF9 cell line (full line) compared to TetR only cell line (dashed line). Error bars in C-D represent standard deviation from three biological replicates.

To test whether AF9 was sufficient for stable of binding of Dot1l at DNA, ChIP qPCR was carried out for Dot1l. Dot1l was detectable at the TetO demonstrating that the affinity of the AF9 Dot1l interaction is sufficient for Dot1l stabilisation at DNA (Figure 3.5A). In addition to Dot1l binding specifically at the TetO, H3K79me3 was detected at the regions spanning the TetO up to +/-5kb from the TetO (Figure 3.5B). The dip in H3K79me3 at the TetO is likely due to this region being nucleosome depleted, which has been observed with other histone modifications, H3K27me3 and H2AK119ub (Blackledge et al. 2014). This demonstrates that Dot1l recruitment to DNA, in the absence of gene regulation, is sufficient for the functional methyltransferase activity of Dot1l in the absence of transcription at chromatin. In addition to this, H3K79me3 is detected within the regions spanning the TetO (where Dot1l is not detectable). This suggests a dynamic interaction of the TetO with the surrounding chromatin. Internal controls for H3K79me3 ChIP qPCR were carried out at H3K79me3 marked genes, SOX2 and OCT4. Of note, the recovery of H3K79me3 at SOX2 and OCT4 was higher than that observed at the TetO, suggesting H3K79me3 levels are higher at natural gene targets (Figure 3.5C).

Following six hours of 2µg/ml doxycycline treatment, which disrupts the interaction between the TetR and TetO, TetR-AF9 is no longer detectable by ChIP qPCR (Figure 3.5D-E). Additionally, Dot1l is no longer detectable at the TetO (Figure 3.5F), suggesting a requirement for the continued presence of AF9. Interestingly, H3K79me3 is still detectable after six hours' doxycycline treatment (figure 3.5G), suggesting that H3K79me3 is maintained following loss of Dot1l.

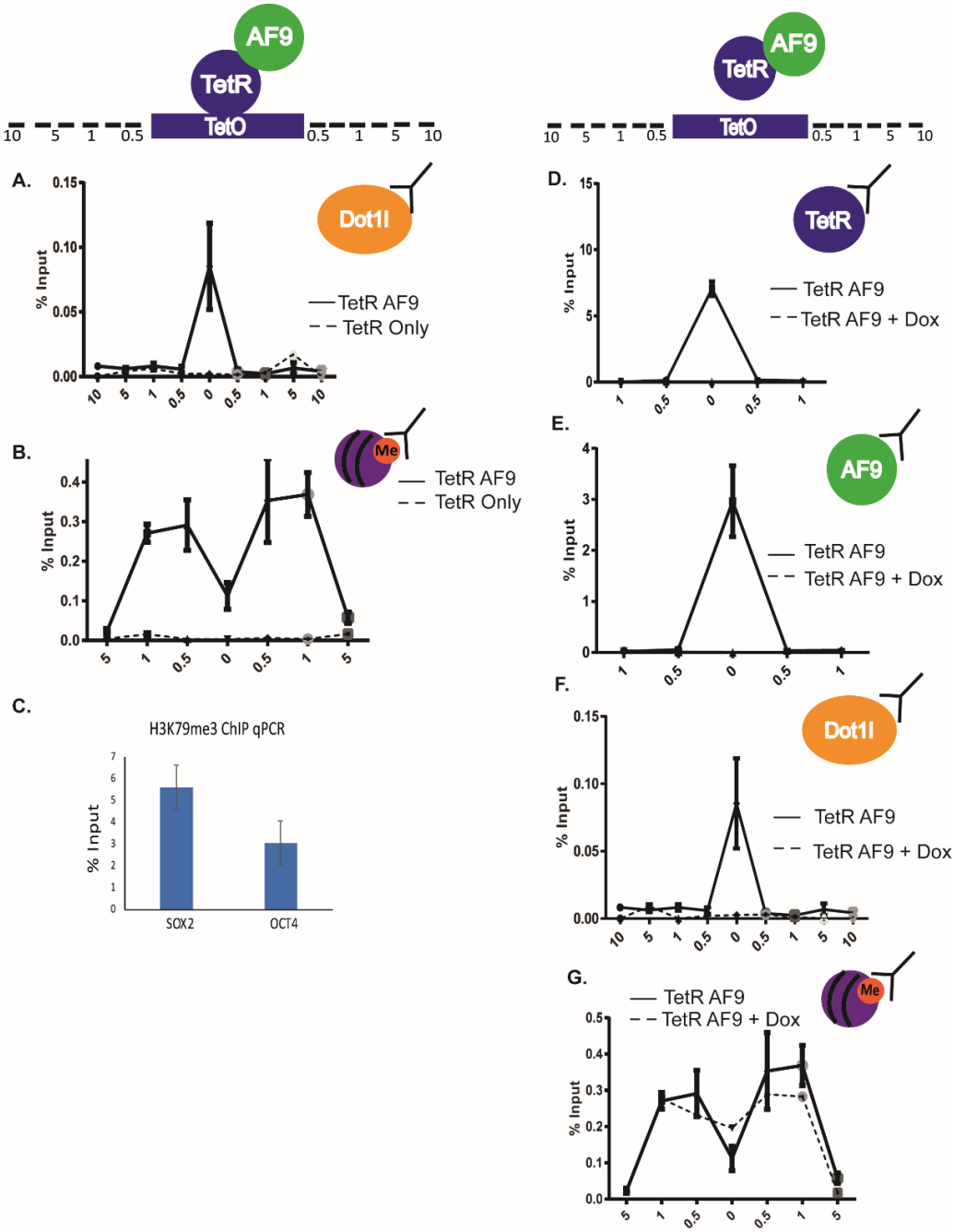

Figure 3-5. AF9 is sufficient for Dot1l recruitment (A) ChIP qPCR for Dot1l at the TetO in the TetR-AF9 cell line (full line) compared to TetR only cell line (dashed line) (B) H3K79me3 is detectable at the TetO by ChIP qPCR in the TetR-AF9 cell line (full line) compared to TetR only cell line (dashed line) (C) H3K79me3 ChIP qPCR at Sox2 and Oct4 (D-F) TetR , AF9 and Dot1l ChIP qPCR in the TetR-AF9 cell line following six hours doxycycline treatment (full line represents TetR-AF9 cell line, dashed line represents TetR-AF9 cell line plus doxycycline). (G) H3K79me3 ChIP qPCR in the TetR-AF9 cell line following six hours doxycycline treatment. In all figures, error bars represent standard deviation of three biological replicates where available.

To determine the length of time H3K79me3 is stable after removal of Dot1l, H3K79me3 ChIP qPCR was performed at a series of time points following doxycycline treatment. H3K79me3 remained detectable up to 48 hours post doxycycline treatment and was undetectable at 72 hours, compared to minus doxycycline control (Figure 3.6A-D). A demethylase for H3K79me has not been identified and therefore loss of H3K79me may be due to nucleosome turnover and recycling rather than an active process of removal. In fact, H3K79me3 was depleted from MLL-ENL gene targets 48 hours following the loss of MLL-ENL binding (Milne et al. 2005).

To observe H3K79me3 dynamics at the TetO using an alternative approach, the TetR-AF9 cell line was treated with 2µM of DOT1L methyltransferase inhibitor, EPZ-5676. H3K79me3 was analysed via ChIP qPCR over a time course assay compared to a 0µM control (Figure 3.6E-G). H3K79me3 was still detectable following 96 hours of EPZ-5676 treatment (Figure 3.6E-G). This suggests there is difference between loss of H3K79me3 following doxycycline treatment and EPZ-5676 treatment. Further repeats of this experiment would be needed to confirm this. Nevertheless, this difference, if significant, could be due to the relatively low efficiency of EPZ-5676 inhibition, supported by data in this thesis (Chapter 4) and published reports (Daigle et al. 2013).

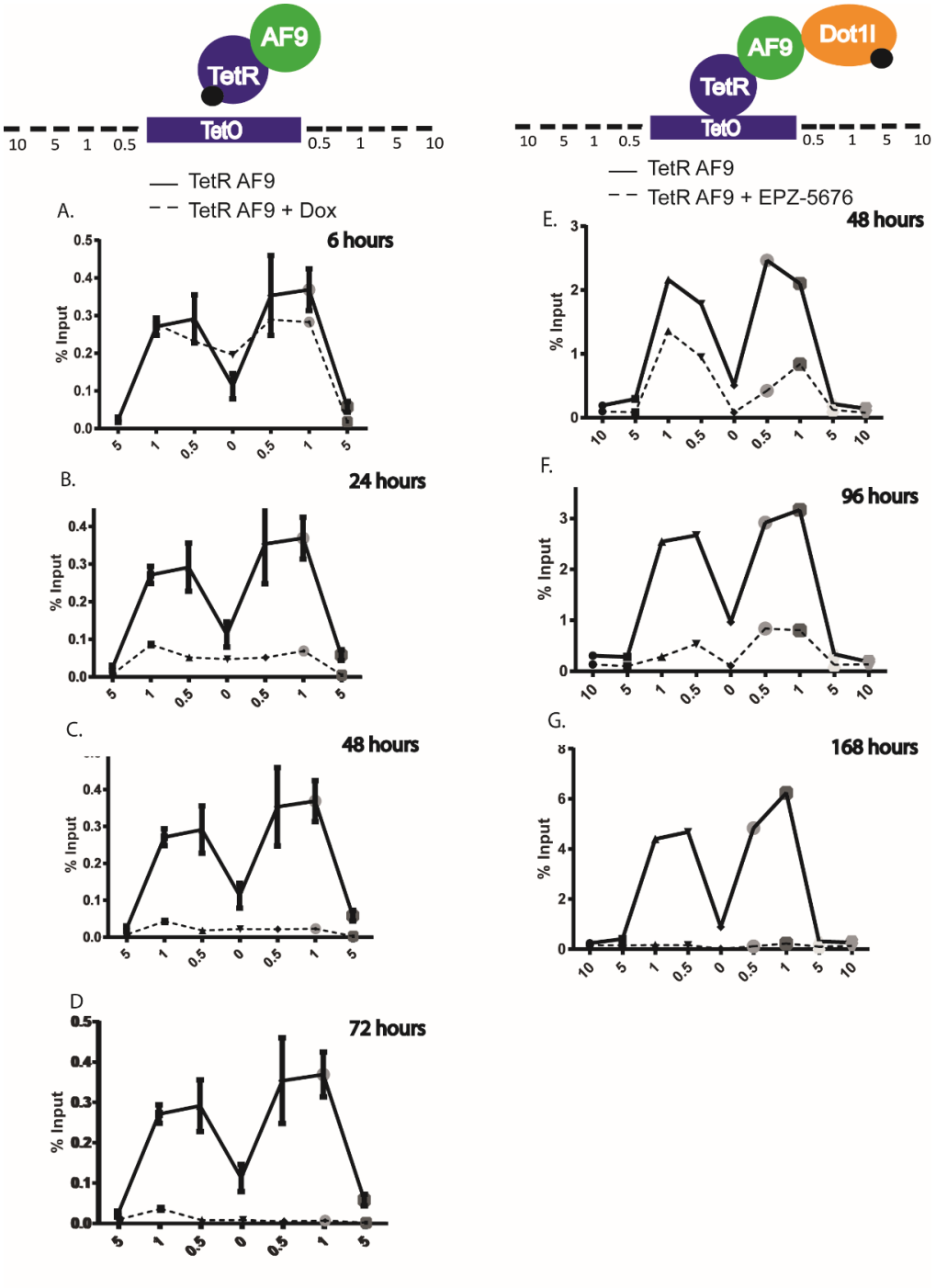

Figure 3-6. H3K79me3 dynamics at the TetO. (A-D) H3K79me3 ChIP qPCR in the TetR-AF9 cell line following 6, 24, 48 and 72 hours 2µg/ml doxycycline treatment (dashed line) compared to TetR-AF9 no doxycycline control (filled line) (E-G) H3K79me3 ChIP qPCR in the TetR AF9 cell line following 48, 96 and 168 hours of 2µM EPZ-5676 treatment (dashed line) compared to TetR-AF9 cell line untreated control (filled line). Error bars represent standard deviation of three biological replicates. All other figures are n=1

3.2.2 AF9 D546R mutation abolishes Dot1l binding

To further investigate whether Dot1l recruitment was specifically due to the AF9 interaction, a TetR-AF9 mutant (D546R) which changes an aspartic acid residue to an arginine residue, was generated. This mutation changes the conformation of the DOT1L binding pocket of AF9, abolishing the interaction between AF9 and DOT1L (Kuntimaddi et al. 2015). TetR-AF9 was still detectable, as in the wild-type TetR-AF9 cell line, at the TetO by ChIP qPCR (Figure 3.7A-B, 5E-F). However, Dot1l and H3K79me3 were not detected by ChIP qPCR in the TetR-AF9 D546R cell line at the TetO or surrounding regions, compared to the wildtype TetR-AF9 cell line (Figure 3.7C-D, 5G-H). This is supported by published NMR structural data of AF9 and DOT1L (Leach et al., 2015) and it is likely that AF9 and DOT1L interact directly *in vivo*.

3.2.3 ENL and AF10 are sufficient for DOT1L recruitment

DOT1L has been shown to biochemically interact with other proteins in the DOT1L complex such as ENL and AF10 (Mohan et al. 2010; Deshpande et al. 2014; Leach et al. 2013). In addition to this, both ENL and AF10 are common fusion partners of MLL and therefore may be directly responsible for DOT1L recruitment at MLL-ENL and MLL-AF10 gene targets *in vivo* (Deshpande et al. 2014; Zeisig et al. 2005). To test whether ENL and AF10 were sufficient for DOT1L recruitment *in vivo*, TetR-ENL and TetR-AF10 fusion proteins were generated, using the AHD domain of ENL and full length AF10, and transiently transfected in mESCs containing the TetO array.

Both TetR-ENL and TetR-AF10 were detectable at the TetO by ChIP qPCR (Figure 3.8A, 3.8F-G). There is currently no good available antibody for AF10 for ChIP applications, therefore, the presence of TetR-AF10 was determined using TetR and Af17 (another member

of the DOT1L complex which AF10 has been shown to interact with) ChIP qPCR (Figure 3.8A-B). In both the TetR-ENL and TetR-AF10 fusions, Dot1l and H3K79me3 were detectable (Figure 3.8C-D, 3.8H-I). These data suggest both ENL and AF10 are sufficient for the functional recruitment of Dot1l and other DOT1L complex members to chromatin *in vivo*.

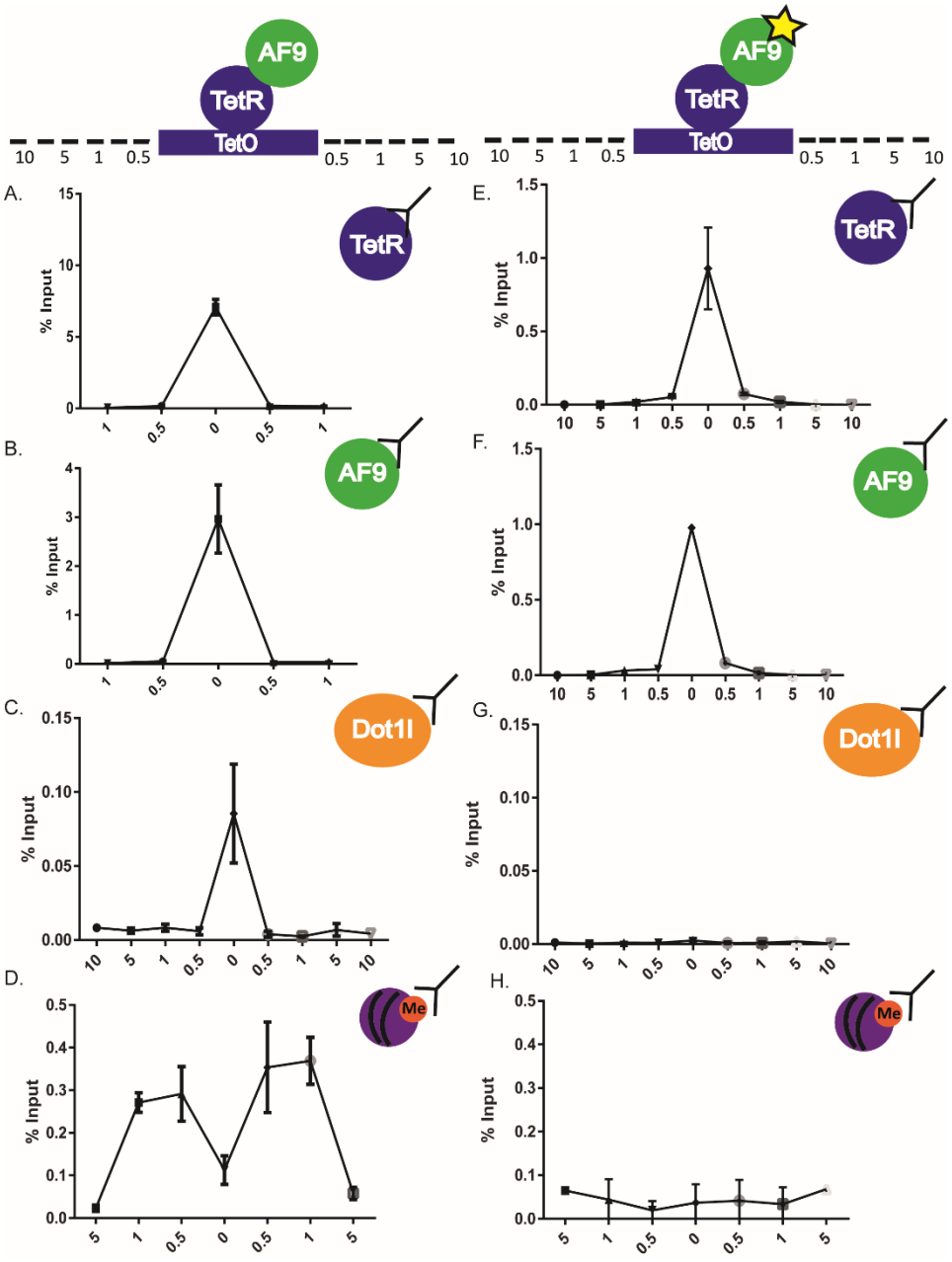

Figure 3-7. AF9 D546R disrupts Dot1l binding. (A-B) TetR-AF9 is detectable at TetO in TetR-AF9 cell line by ChIP qPCR for TetR and AF9 (C-D) Dot1l and H3K79me3 in TetR-AF9 cell line by ChIP qPCR (E-F) ChIP qPCR for TetR and AF9 in TetR-AF9 D546R transfected mES base cells (containing TetO) (G-H) Dot1l and H3K79me3 are not detectable in TetR-AF9 D564R transfected mES base cells. Error bars represent standard deviation of two biological replicates.

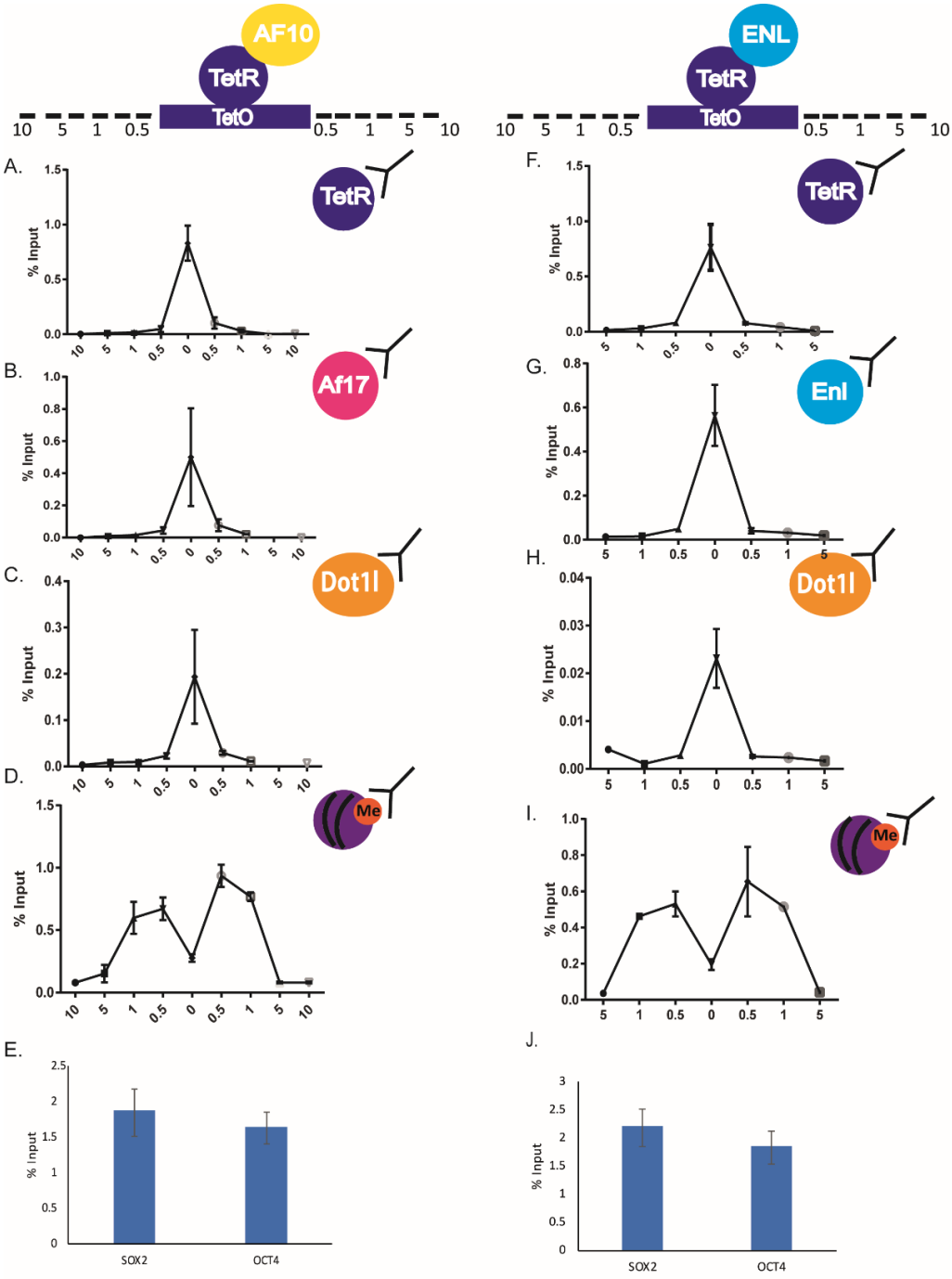

Figure 3-8. ENL and AF10 are sufficient for Dot1l recruitment. (A) TetR ChIP qPCR in TetR-AF10 transiently transfected mES base cells (B) AF17 ChIP qPCR in TetR-AF10 transiently transfected mES base cells (C-D) Dot1l and H3K79me3 ChIP qPCR in TetR-AF10 transfected mES base cells. (E) H3K79me3 ChIP qPCR at control gene targets Sox2 and Oct4 in TetR-AF10 transfected mES base cells (F-G) TetR and ENL ChIP qPCR in TetR-ENL transiently transfected mES base cells (H-I) Dot1l and H3K79me3 ChIP qPCR in TetR-ENL transiently transfected mES base cells (J) H3K79me3 ChIP qPCR at control gene targets SOX2 and OCT4 in TetR-ENL transient cells. All error bars represent standard deviation of 3 biological replicates.

3.2.4 AF4 is not sufficient for DOT1L recruitment to chromatin

To determine whether AF4 is sufficient for recruitment of Dot1l in the TetR system, the domain of AF4 found in the MLL-AF4 fusion protein, which retains the AF9/ENL and AFF4 binding domains, was fused to TetR (Figure 3.9).

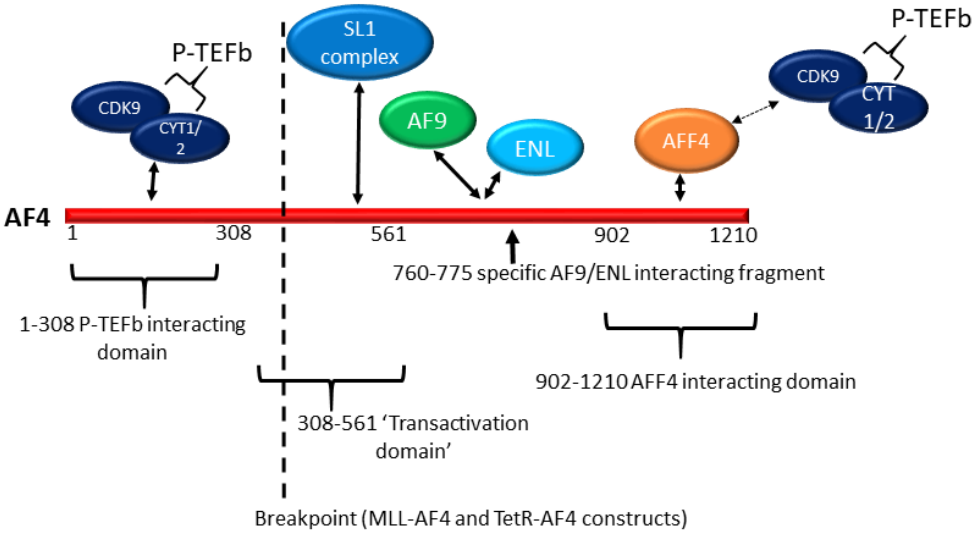

Figure 3-9. Domains of AF4 shown to interact biochemically with interacting proteins. Right of the dashed line represents the portion of AF4 found in the TetR-AF4 fusion protein (Short TetR-AF4 fusion contains amino acids 561-902) and in MLL-AF4 leukaemia (Biswas et al. 2011; Okuda et al. 2015; Leach et al. 2013; C. Lin et al. 2010;Yokoyama et al. 2010a).

Efforts at establishing a stable cell line were made, however, this was not possible due to cell death, potentially due to toxicity of human AF4 (Lin et al. 2016). As an alternative, transient transfections followed by ChIP qPCR were carried out. Initially, to test whether AF4 maintained its binding ability when fused to TetR, ChIP qPCR was carried out for Af9 and Enl. TetR-AF4, Af9 and Enl were all detectable at the TetO (Figure 3.10A-D). To test whether AF4 is sufficient for DOT1L recruitment, Dot1l and H3K79me3 ChIP qPCR was carried out. Neither Dot1l nor H3K79me3 were detectable at the TetO or regions spanning the TetO,

suggesting AF4 is not sufficient for DOT1L and H3K79me3 recruitment (Figure 3.10E-F). As a positive control, H3K79me3 was detectable at SOX2 and OCT4 in this experiment (Figure 3.10G). However, as the transfection efficiency of mESCs using lipofectamine is around 10% (data not shown), it cannot be excluded that Dot1l and H3K79me3 were present at levels too low for detection by ChIP.

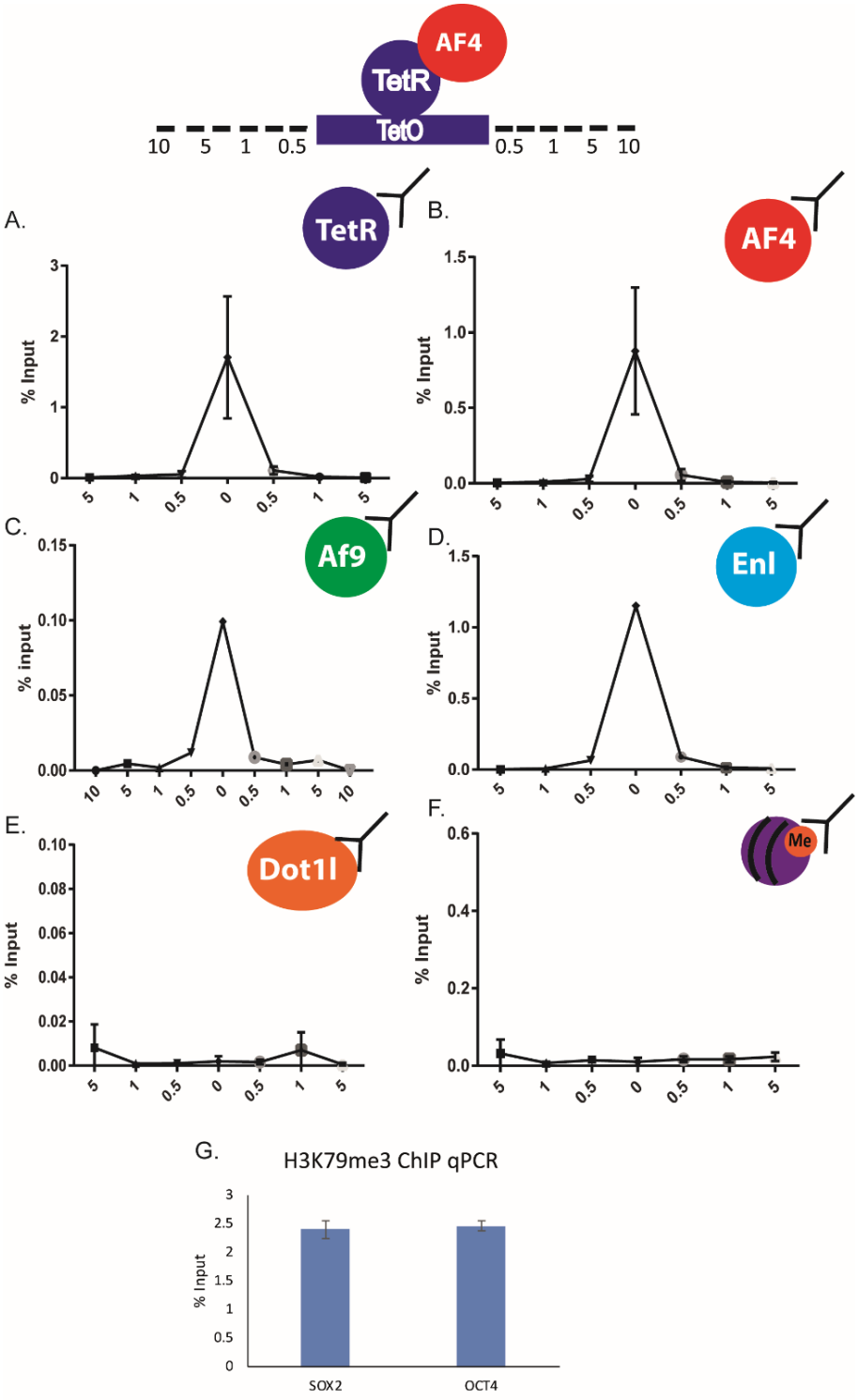

Figure 3-10. AF4 is not sufficient for DOT1L recruitment (A-B) ChIP qPCR for TetR and AF4 at the TetO in TetR-AF4 transiently transfected mES base cells (C-D) ChIP qPCR for Af9 and Enl at the TetO in TetR-Af4 transfected cells (E-F) ChIP qPCR for Dot1L and H3K79me3 at the TetO in TetR-AF4 transfected mES base cells. Error bars on all figures represent standard deviation from three biological replicates

The TetR-AF4 fusion contains an AF9/ENL binding domain in addition to an AFF4 binding domain, which in turn can interact with P-TEFb (Figure 3.9). One potential reason why AF4 is toxic to the cells is that P-TEFb may be sequestered by the TetR-AF4 fusion away from genes where it is required for transcription. To address this problem, a truncated version of AF4, which contained only the AF9 and ENL binding domain, was fused to the TetR (Figure 3.9, amino acids 561-902). Supporting this explanation, a stable cell line was successfully generated and TetR-AF4 (561-902) was detectable at the TetO (Figure 3.11A). In addition to this, Enl was also detected at the TetO, suggesting truncated AF4 can still interact with Enl (Figure 3.11B). Dot1l and H3K79me3 were again not detectable in this system, suggesting that this domain of AF4 does not recruit Dot1l (Figure 3.11C-D). Moreover, H3K79me3 ChIP qPCR at SOX2 and OCT4 demonstrated this ChIP was successful (Figure 3.11E). Note that it cannot be ruled out that other domains of AF4 may be sufficient for DOT1L recruitment. Nevertheless, these data suggest that MLL-AF4 may not directly stabilise DOT1L at gene targets, at least to the same degree as AF9 and ENL. Given the important role of DOT1L at MLL-AF4 target genes, this suggests there is no direct link between MLL-AF4 and DOT1L, therefore, other protein complexes recruited by MLL-AF4 may be responsible for DOT1L stabilisation.

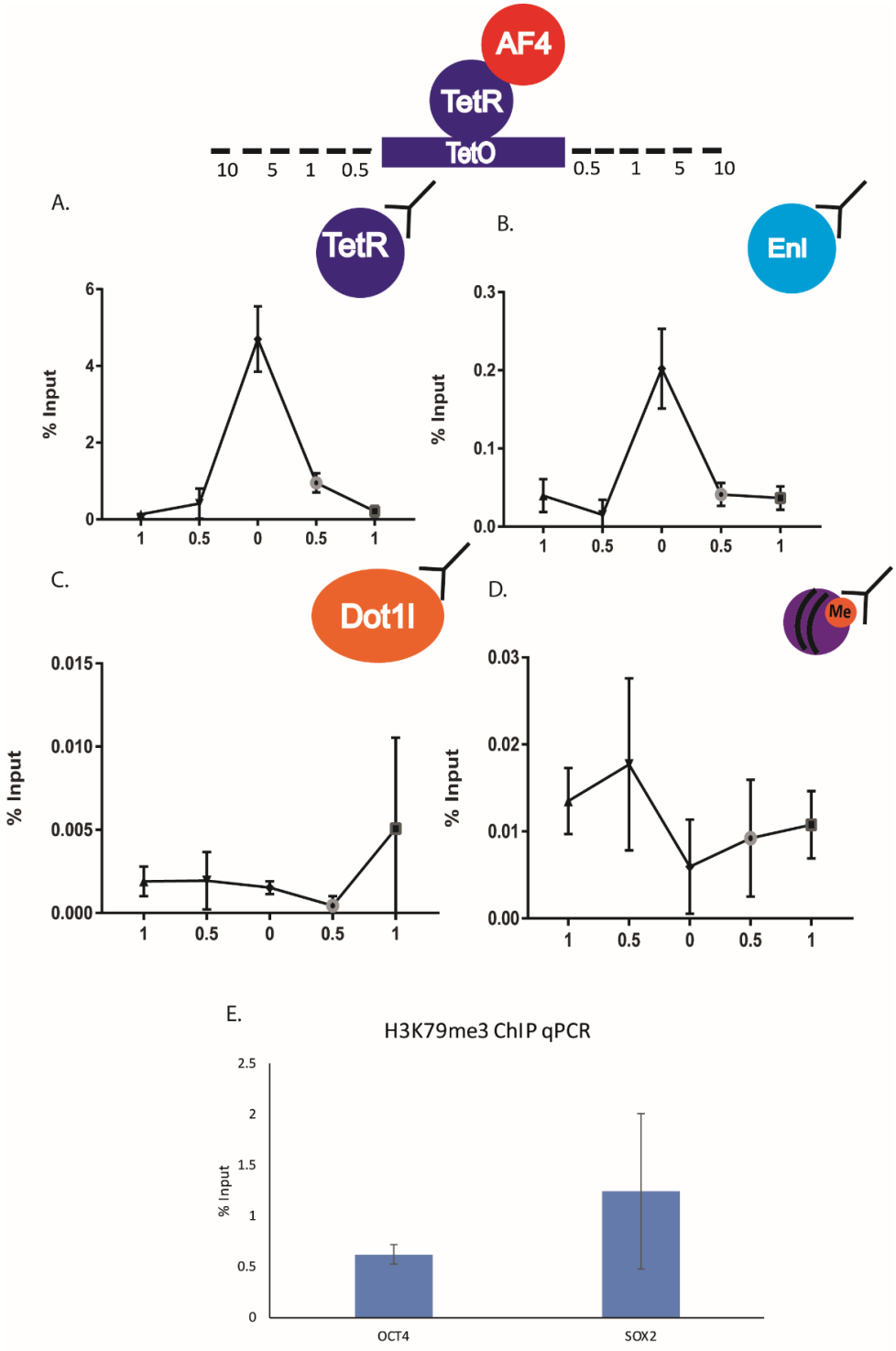

Figure 3-11. TetR-AF4(561-902) is not sufficient for DOT1L recruitment (A) ChIP qPCR for TetR at the TetO in stably expressing TetR-AF4 (561-902) cell line (B) ChIP qPCR for Enl at the TetO in TetR-AF4 (561-902) cell line (C-D) ChIP qPCR for Dot1l and H3K79me3 at the TetO in the TetR-AF4 (561-902) cell line (E) H3K79me3 ChIP qPCR at SOX2 and OCT4 in the TetR-AF4 (561-902) cell line. Error bars in all figures represent standard deviation from three biological replicates.

3.2.5 PAF1 is sufficient for DOT1L recruitment

As AF4 is not sufficient for DOT1L recruitment but is sufficient for the recruitment of DOT1L complex members, it is possible that MLL-AF4 could recruit DOT1L to gene targets via an indirect mechanism. To explore whether an indirect mechanism of DOT1L recruitment may be orchestrated by MLL-AF4, biochemical interactions of AF4 and DOT1L were revisited and investigated *in vivo* (Figure 3.2). It has been demonstrated that AF9 and ENL can interact biochemically with TBP, part of the SL1 complex (Biswas et al. 2011). It has been proposed that at MLL-AF4 gene targets, the SL1 complex is important for gene activation (Okuda et al. 2015). In addition to this, PAF1, part of the PAF1 elongation complex, has been shown *in vitro* to interact with AF9 and ENL, and MLL (Kim et al. 2010; Van Oss et al. 2016; He et al. 2011; Milne et al. 2010). Furthermore, PAF1 has been linked to H3K79me in yeast, whereby a reduction of H3K79me is observed following PAF1 knockdown (Krogan et al., 2003 and Wood et al., 2003). As these protein complexes can potentially interact with MLL-AF4 and the DOT1L complex, it is possible that they could contribute to a stable binding site for DOT1L at MLL-AF4 gene targets.

If Paf1 and Tbp could contribute to DOT1L stabilisation at MLL-AF4 gene targets, it might be expected that they can interact with AF4 and AF9 *in vivo*. To address this, Tbp and Paf1 ChIP qPCR was performed in the TetR-AF4 and TetR-AF9 cell lines. AF9 was sufficient for both Paf1 and Tbp recruitment to chromatin in the TetR system, with Tbp binding abolished following doxycycline treatment (Figure 3.12A-B). Similarly, AF4 was sufficient for Paf1 recruitment to chromatin in the TetR system (Figure 3.12C).

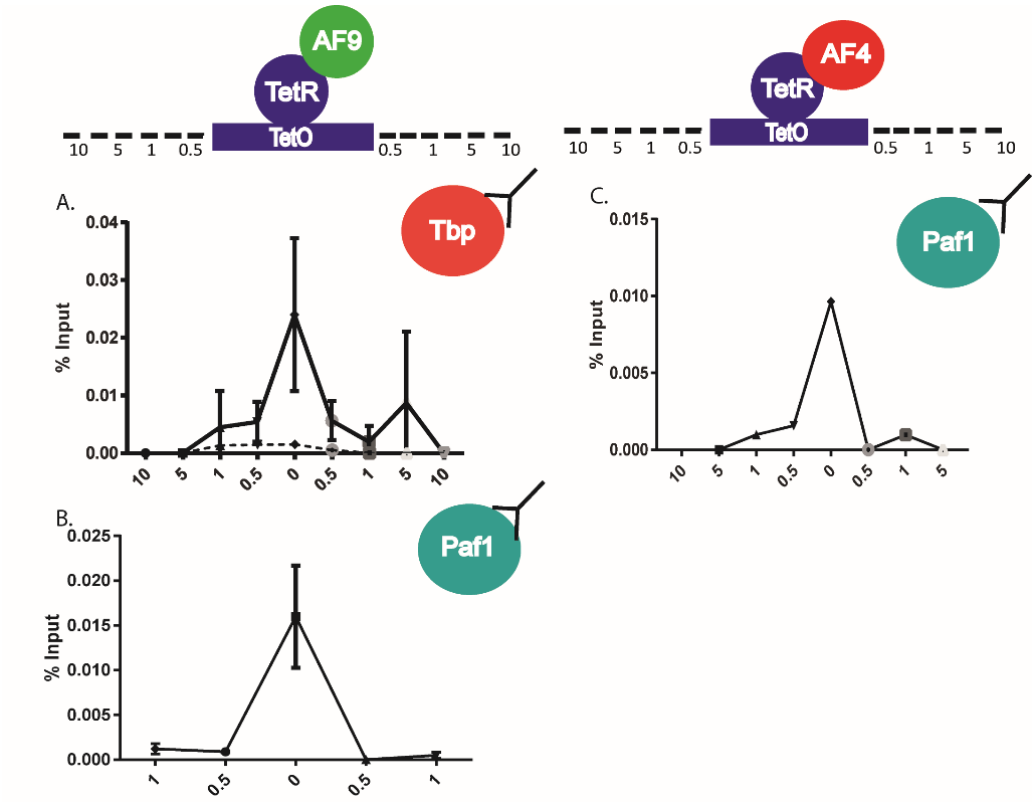

Figure 3-12. AF9 and AF4 are sufficient for the recruitment of other transcription proteins (A) ChIP qPCR for Tbp at the TetO in the TetR AF9 cell line (full line) compared to TetR only cell line (dashed line) (B) ChIP qPCR for Paf1 at the TetO in the TetR AF9 cell line (C) ChIP qPCR for Paf1 at the TetO in the TetR-AF4 transfected mES base cell line. Error bars where available represent standard deviation of three biological replicates.

PAF1 has been shown to interact with the YEATS domain of AF9. DOT1L and AF4 have been demonstrated to interact with the AHD domain of AF9. This suggests that these interactions can occur simultaneously. Therefore, it is possible that both AF9 and PAF1 can be recruited directly via the MLL-AF4 fusion protein. To test whether PAF1 was sufficient for DOT1L and AF9 recruitment, a stable cell line was generated using a TetR-PAF1 fusion, already present in the lab, which contained full length PAF1. Initially, TetR-PAF1 and Af9 binding at the TetO was confirmed via TetR, PAF1 and Af9 ChIP qPCR (Figure 3.13A-B and E). To investigate

whether PAF1 was sufficient for Dot1l and H3K79me3 recruitment to chromatin, Dot1l and H3K79me3 ChIP qPCR were carried out in the TetR-PAF1 cell line. Both Dot1l and H3K79me3 were detectable at the TetO and absent following 48 hour doxycycline treatment (Figure 3.13C-D). Taken together, this suggests that PAF1 is sufficient for Dot1l and H3K79me3 recruitment to chromatin *in vivo*. Therefore, at a gene PAF1 and AF9 may stabilise DOT1L binding.

Interestingly, PAF1 has been shown to interact with RNAPII, another DNA binding protein complex. Therefore, it is possible that DOT1L is additionally stabilised at MLL-AF4 gene targets via the PAF1-RNAPII interaction. Therefore, to test whether PAF1 was sufficient for RNAPII recruitment in vivo, Ser2Ph RNAPII ChIP qPCR was carried out in the TetR-PAF1 cell line. RNAPII was detectable in the TetR-PAF1 cell line and binding was abolished following 48 hours doxycycline treatment (Figure 3.13F), suggesting that PAF1 is sufficient for RNAPII recruitment to chromatin in the absence of a gene. Moreover, this may aid in illustrating how a stable DOT1L binding site might be created, via the accumulation of multiple protein complex interactions.

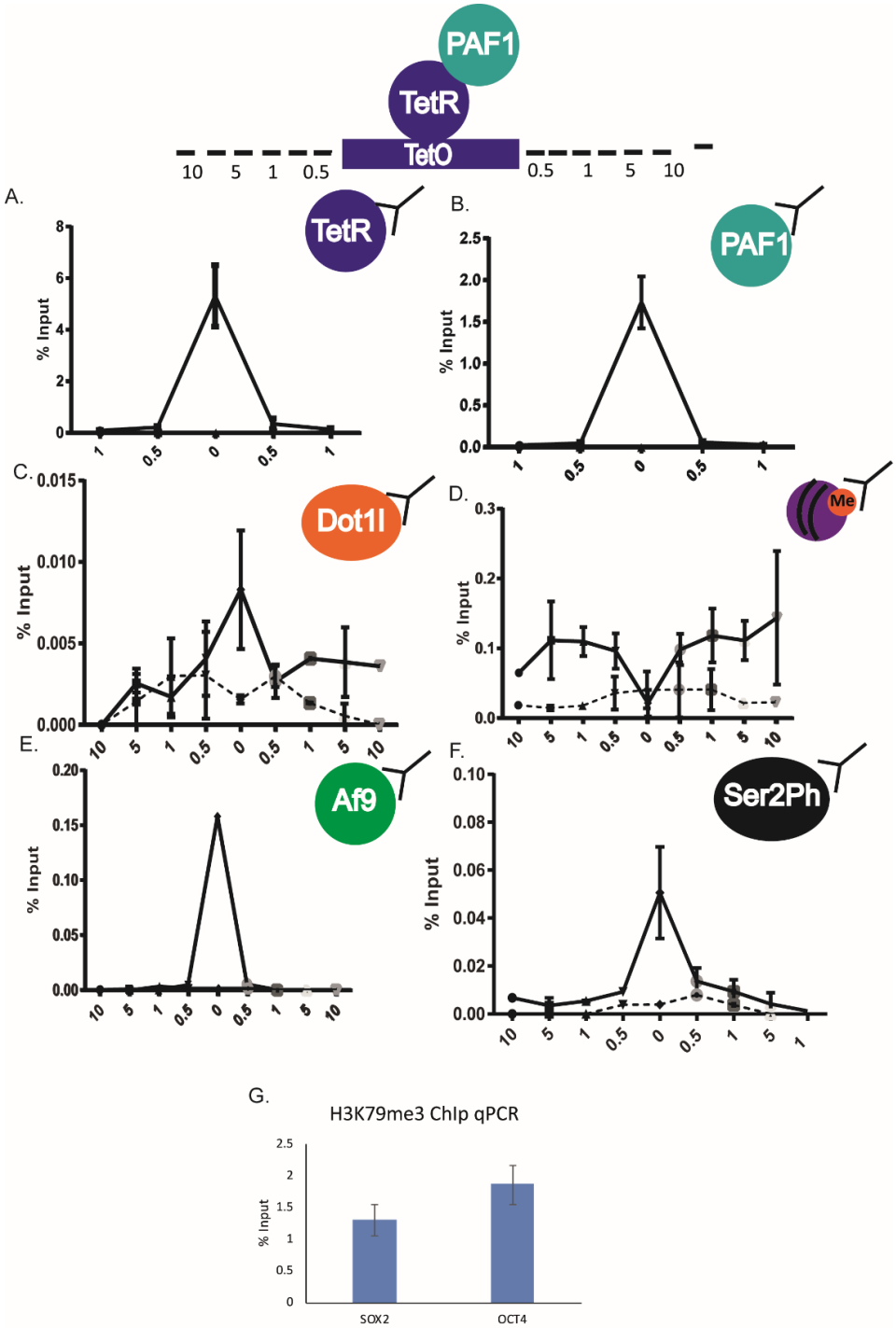

Figure 3-13. PAF1 is sufficient for DOT1L recruitment. (A-B) TetR and PAF1 ChIP qPCR at the TetO in TetR-PAF1 cell line (C-D) Dot1l and H3K79me3 ChIP qPCR at the TetO in TetR-PAF1 cell line (full line) compared to 2µg/ml 6 hours doxycycline treatment in TetR PAF1 cell line (dashed line) (E-F) Enl and RNAPII Serine2Ph ChIP qPCR at TetO in TetR-PAF1 cell line (full line) compared to 2µg/ml 6 hours doxycycline treatment in TetR-PAF1 cell line (dashed line). Error bars where available represent standard deviation from three biological replicates.

One important consideration with the different TetR cell lines is whether they express the fusion protein to the same degree. To address this, whole cell extracts were prepared from $1x10^6$ cells from each cell line followed by western blotting. This analysis revealed that all cell lines express the respective TetR fusion protein to a similar degree demonstrating that binary recruitment of a protein in one system can be compared to another (Figure 3.14A). Furthermore, H3K79me3 levels are comparable across experiments at control genes (Figure 3.14B).

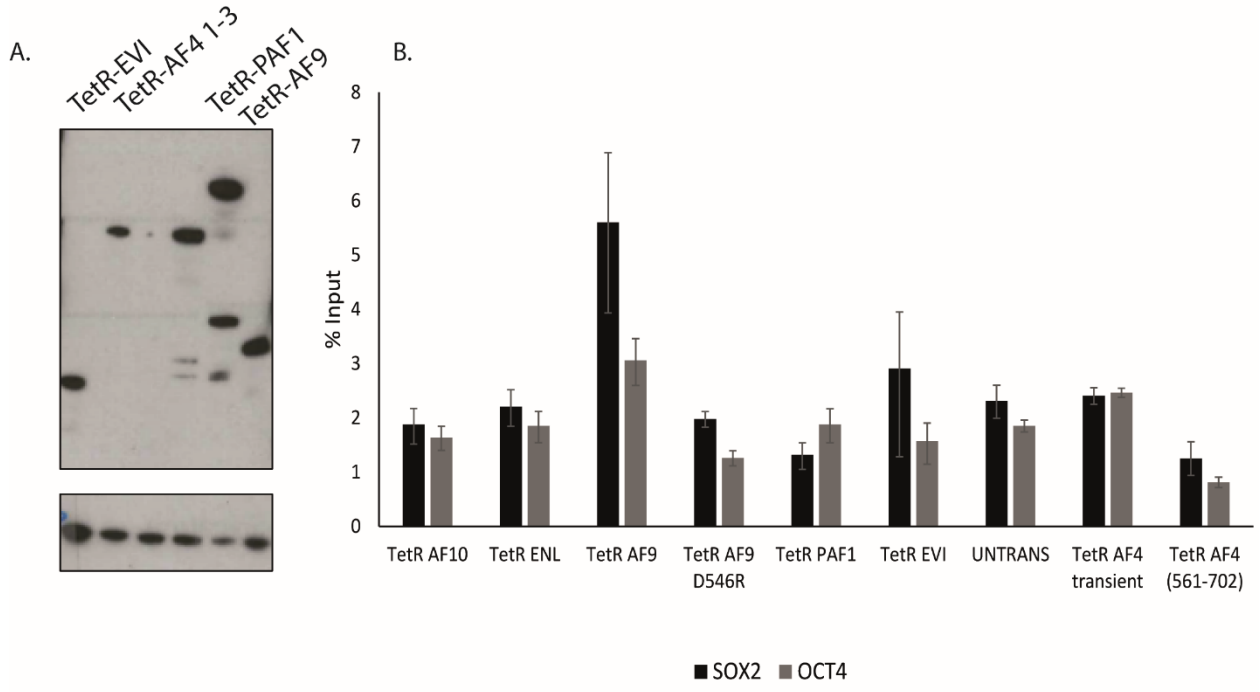

Figure 3-14. TetR fusion protein expression and H3K79me3 positive ChIP qPCR controls (A) Western blot analysis demonstrating the expression levels of all TetR fusion proteins from each stable cell line. Western blot membrane probed using anti-FS2 antibody and developed using ECL. (B) Positive controls for all H3K79me3 ChIP qPCR performed at SOX2 and OCT4.

3.2.6 ENL may be necessary for DOT1L recruitment at natural gene targets

Several proteins were successfully identified that are sufficient for DOT1L recruitment to the TetO. We wanted to expand this analysis to better understand how stable DOT1L binding is controlled at important gene targets. As ENL has been demonstrated to be sufficient for Dot1l recruitment *in vivo* (Figure 3.8) it was important to test whether ENL was necessary for DOT1L recruitment at genes in human MLL-AF4 leukaemia cells. To test this, siRNA mediated knockdown of ENL was carried out in SEM cells. Two different ENL siRNAs were tested, and both demonstrated a clear knockdown of ENL after 4 days of transfection on both the protein and RNA level (Figure 3.15A-B). However, ENL knockdown also had a pronounced effect upon DOT1L mRNA levels (Figure 3.15B), which suggests that ENL may regulate DOT1L at the level of transcription or RNA stability.

Interestingly, ENL knockdown leads to a reduction in mRNA levels of the MLL-AF4 DOT1L-dependent genes, *PROM1* and *BCL11A*, (characterised by downregulation following DOT1L inhibition, see Chapter 4) (Figure 3.14B). To understand the effect of knockdown on DOT1L localisation to these genes, H3K79me3 ChIP qPCR was carried out 5 days post initial knockdown. A reduction in H3K79me3 was observed at *PROM1* and *BCL11A* (Figure 3.15C), suggesting reduced DOT1L stabilisation. This is in line with data from the literature demonstrating ENL siRNA knockdown leads to a reduction in global and local H3K79me levels (Mueller et al., 2007). However, owing to the effect of ENL knockdown on DOT1L mRNA, it is possible that the reduction in H3K79me3 is due to lower DOT1L protein levels, rather than the loss of its interaction with ENL.

3.2.7 ENL is important for MLL-AF4 leukaemia

As ENL knockdown seems to lead to downregulation of some DOT1L dependent genes, it was possible that ENL was also important for leukaemic growth of SEM cells. To determine this, siRNA knockdown of ENL coupled with colony assays were conducted. Interestingly, ENL knockdown demonstrated a 40% reduction in colony number in comparison to scrambled siRNA control (Figure 3.15D). This suggests that ENL is important for the maintenance of MLL-AF4 leukaemia, consistent with published data suggesting ENL is important for acute myeloid leukaemia, and may be involved in the regulation of important leukaemia gene targets, potentially via DOT1L and H3K79me (Wan et al., 2017 and Erb et al., 2017).

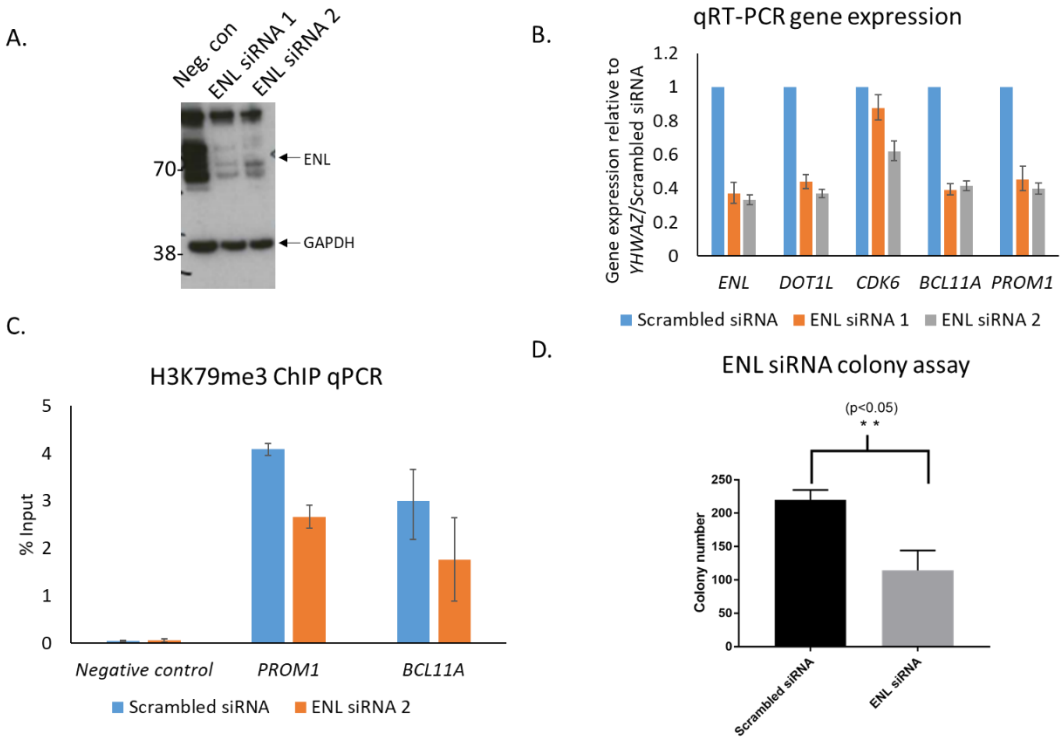

Figure 3-15. ENL may be necessary for DOT1L recruitment (A) Western blot analysis on WCEs made from SEM cells treated with two independent ENL siRNAs, using anti-ENL antibody and anti-GAPDH as loading control (B) qRT-PCR from RNA purified from SEM cells treated with two independent ENL siRNAs (Orange and Grey) and scrambled siRNA control (Blue). Gene expression normalised to YHWAZ. Error bars represent standard deviation of two biological replicates (C) H3K79me3 ChIP qPCR at DOT1L gene targets on SEM cells treated with scrambled and ENL siRNAs. Error bars represent standard deviation of two biological replicates. (D) Colony assay performed on SEM cells treated with scrambled and ENL siRNA knockdown. Error bars represent standard error between three technical replicates. $p<0.05$ following unpaired parametric t-test.

3.2.8 PAF1 does not affect H3K79me at MLL-AF4 gene targets following 5 days siRNA/shRNA knockdown

As PAF1 has been demonstrated to be sufficient for Dot1l recruitment to TetO *in vivo* (Figure 3.13) it was important to test whether PAF1 was necessary for DOT1L recruitment at genes in human MLL-AF4 leukaemia cells. To establish this, siRNA-mediated knockdown of PAF1 was performed, with a reduction on both the protein and RNA levels observed after four days (Figure 3.16A-B) and with no effect on DOT1L expression (Figure 3.16B). However, there were no observed effects on the DOT1L dependent gene targets *CDK6*, *BCL11A* or *PROM1* (Figure 3.16B). To determine if PAF1 knockdown affected DOT1L recruitment, H3K79me3 ChIP qPCR was performed four days after initial knockdown. Consistent with the expression analysis, no apparent effect was observed upon H3K79me3 levels at *PROM1* or *BCL11A* (Figure 3.16C). This indicates that PAF1 may not be necessary for the recruitment of DOT1L at these specific genes or alternatively, the PAF1 knockdowns were insufficient. Given the stability of H3K79me, it is also possible that a longer knockdown is necessary to see the effect of PAF1 depletion on transcription of these genes.

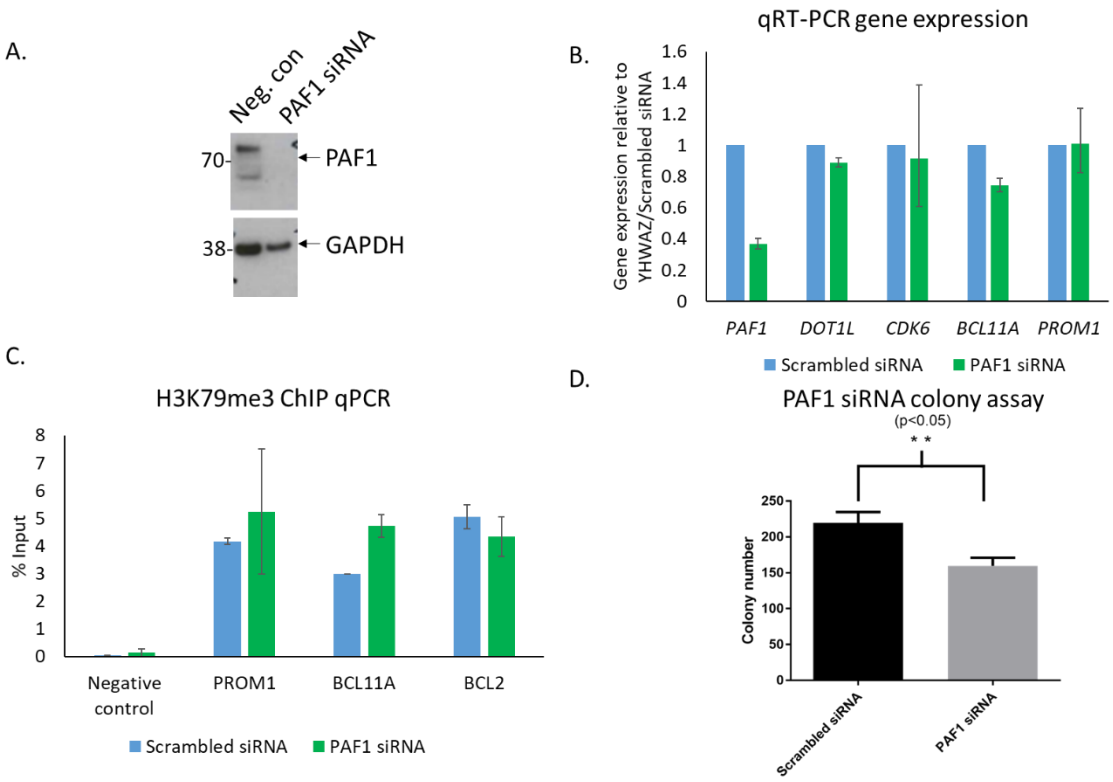

Figure 3-16. PAF1 knockdown may not be necessary for DOT1L recruitment. (A) Western blot analysis WCEs of PAF1 siRNA treatment compared to scrambled siRNA control in SEM cells. Western blot membrane probed with anti-PAF1 antibody and anti-GAPDH as loading control (B) qRT-PCR analysis from RNA purified from SEM cells treated with scrambled siRNA control (blue) and PAF1 siRNA (green) normalised to internal control YHWAZ. Error bars represent standard deviation of two biological replicates (C) H3K79me3 ChIP qPCR at DOT1L gene targets in SEM cells treated with scrambled and ENL siRNAs. Error bars represent standard deviation between two biological replicates (D) Colony forming assay performed on SEM cells treated with scrambled control siRNA and PAF1 siRNA. Error bars represent standard error between three technical replicates. $p<0.05$ following unpaired parametric t-test.

Due to the potential need to observe the changes in H3K79me over a longer PAF1 knockdown time course, an inducible PAF1 shRNA approach was used (Zuber et al. 2011). The inducible PAF1 shRNA provides a method to observe the effect on H3K79me over a longer period without performing multiple electroporation experiments on the same cells. In addition to this, it provides a consistent level of knockdown which is comparable amongst biological replicates.

SEM cells were lentivirally transduced with an inducible PAF1 shRNA construct and clonal lines were established (Chapter 2). Cell lines were induced with doxycycline (2µg/ml), and a minus doxycycline control was grown in parallel. PAF1 RNA levels were reduced to 20% after 24 hours and remained low at 48 hours (Figure 3.17A). Similarly, a reduction on the protein level was observed after 24 hours, with the reduction maintained at 48 hours (Figure 3.17B). To determine whether reduction of PAF1 led to a decrease in binding at gene targets, ChIP qPCR for LEO1, a PAF1 complex member, was performed six days post shRNA induction, due to poor ChIP signal using a PAF1 antibody. LEO1 binding was reduced by approximately 50% following PAF1 knockdown (Figure 3.17C). To determine the consequence of this on DOT1L recruitment, H3K79me3 ChIP qPCR was also carried out six days post shRNA induction. No reduction in H3K79me3 was observed at MLL-AF4 DOT1L-dependent gene targets (Figure 3.17D). These data perhaps indicate that further investigation would be required to verify the necessity of PAF1 in DOT1L recruitment. As this is not a complete knockdown and only a 50% reduction is observed in PAF1 complex occupancy at genes, it is possible that the effects on DOT1L recruitment cannot be observed. Additionally, it may be that a longer period is required to observe an effect upon H3K79me, due to the slow turnover of H3K79me, even if DOT1L recruitment is affected.

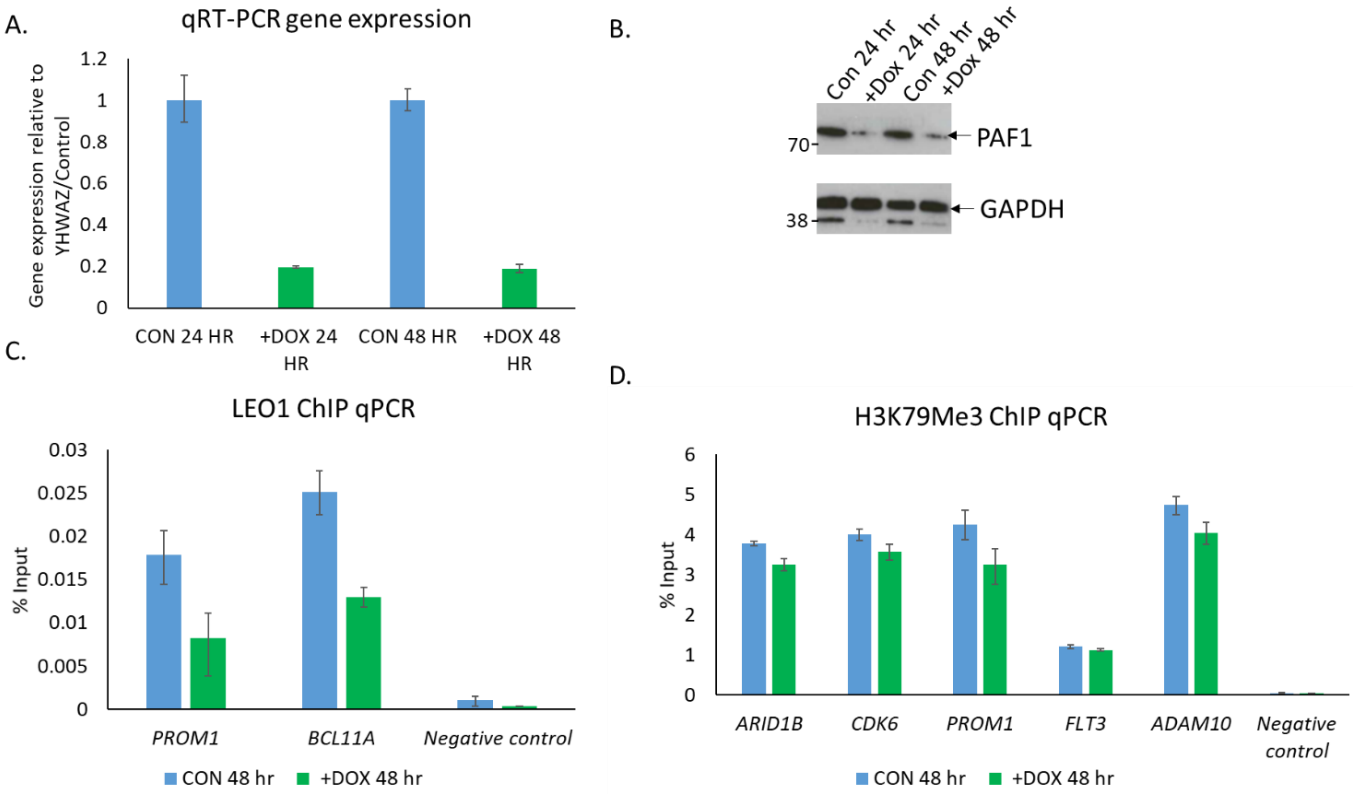

Figure 3-17. PAF1 shRNA knockdown does not lead to loss of H3K79me (A) qRT-qPCR of PAF1 following PAF1 shRNA knockdown in SEM cells following 2µg/ml doxycycline compared to no doxycycline control, 24 and 48 hours post induction. Error bars represent 2 biological replicates. (B) Western blot analysis on WCEs from SEMs following 48hr PAF1 shRNA knockdown (C)LEO1 ChIP qPCR following 48hr PAF1 shRNA knockdown in SEM cells at DOT1L gene targets (D) H3K79me3 ChIP qPCR in SEM cells 48 hours post PAF1 shRNA knockdown at DOT1L genes.

3.4 Discussion

In this chapter, it has been shown that DOT1L complex members, AF9, ENL and AF10 are all sufficient for Dot1l and H3K79me recruitment to chromatin *in vivo*. As AF9, ENL and AF10 have all been demonstrated to fuse to MLL in MLL-r leukaemia, it is probable that a direct recruitment mechanism of DOT1L occurs at these MLL-FP gene targets. AF17, another co-purified protein with DOT1L, was not investigated in this project, but based upon its homology to AF10, it is likely that AF17 is also sufficient for DOT1L recruitment. As AF9, ENL and AF10 have all been demonstrated to fuse to MLL in MLL-r leukaemia, it is probable that a direct recruitment mechanism of DOT1L occurs at these MLL-FP gene targets. Furthermore, we have demonstrated that following DOT1L recruitment, a broad domain of H3K79me3 domain can be established at a region of chromatin completely devoid of transcription. This highlights that DOT1L function is not dependent upon transcription.

However, this does not fully explain how DOT1L can be recruited to MLL-AF4 gene targets. Importantly, we have shown, for the first time, that AF4 is not sufficient for Dot1l and H3K79me recruitment to chromatin *in vivo*, consistent with *in vitro* data demonstrating that AF4 and DOT1L are found in mutually exclusive protein complexes (Biswas et al. 2011; Yokoyama et al. 2010; Milne et al. 2005). However, it has recently been observed that MLL-Af4 and DOT1L can interact via *in vitro* Co-IP experiments (Lin et al. 2016). This observation is surprising as this interaction has not been shown previously. One key difference between this observation and others, is that, mouse Af4 is fused to MLL. The sequence conservation of mouse and human Af4 is 70%. Therefore, it is possible that the 30% difference may account for why MLL-Af4 may recruit Dot1l. However, preliminary data from our lab has shown that

TetR-MLL-Af4 does not lead to the recruitment of DOT1L or H3K79me at the TetO *in vivo* (data not shown).

As MLL-AF4 does not likely recruit DOT1L directly to gene targets, a potential indirect recruitment mechanism of DOT1L could exist. AF9 interacts with both AF4 and DOT1L, however, structural studies have demonstrated that both interactions involve the same binding pocket of AF9 (Leach et al., 2013). Thus, these binding events are mutually exclusive. The K_D values of AF9 with either AF4 or DOT1L are 0.17nM and 1.6nM, respectively. Therefore, even though AF4 and DOT1L might compete for binding with AF9, it is possible that AF9 interacts with AF4 with higher affinity. Interestingly, since AF9, AF4 and DOT1L have intrinsically disordered domains (IDDs), it has been suggested that that the interaction of AF9 with either AF4 or DOT1L could occur dynamically and may overlap (Leach et al. 2013). This suggests that AF9 could in part contribute to the stability of DOT1L at MLL-AF4 gene targets whilst still interacting with MLL-AF4. However, our results with the TetR system demonstrate that AF4 cannot directly recruit Dot1l so this hypothesis does not seem likely.

To investigate this further, we proposed that an indirect mechanism of DOT1L recruitment might occur at MLL-AF4 gene targets. Both TetR-AF4 and TetR-AF9 were sufficient for Paf1 recruitment *in vivo* and furthermore, TetR-PAF1 was sufficient for DOT1L recruitment *in vivo*. Importantly PAF1 interacts with a different binding domain of AF9 compared to AF4 and DOT1L and therefore, could interact with both simultaneously. Interestingly, this raises the question why is not Dot1l not detectable in the TetR-AF4 system even though Af9 and Paf1 are detectable? This could be due to Dot1l being below the detection threshold in the TetR-AF4 system, due to the number of interactions which take place between AF4 and DOT1L (Figure 3.19).

In addition to this, PAF1 was also sufficient for Af9 and RNAPII recruitment and as RNAPII interacts directly with the DNA, this provides an additional explanation of how DOT1L might be stabilised at active gene targets. Interestingly, the PAF1 complex has been functionally associated with DOT1L via promotion of H2Bub, and although the link between H2Bub and H3K79me has not been addressed in this thesis, the recruitment of DOT1L via PAF1 may explain this observed connection (Xiao et al. 2005; Wood et al. 2003; He et al. 2011). Taken together, this suggests that DOT1L may be recruited to MLL-AF4 gene targets via an indirect mechanism involving PAF1 and members of the DOT1L complex, AF9 and ENL. Taking these data from this chapter together, an updated version of the in vitro biochemical interactions can be generated which highlights those interactions which occur in vivo at chromatin (Figure 3.18).

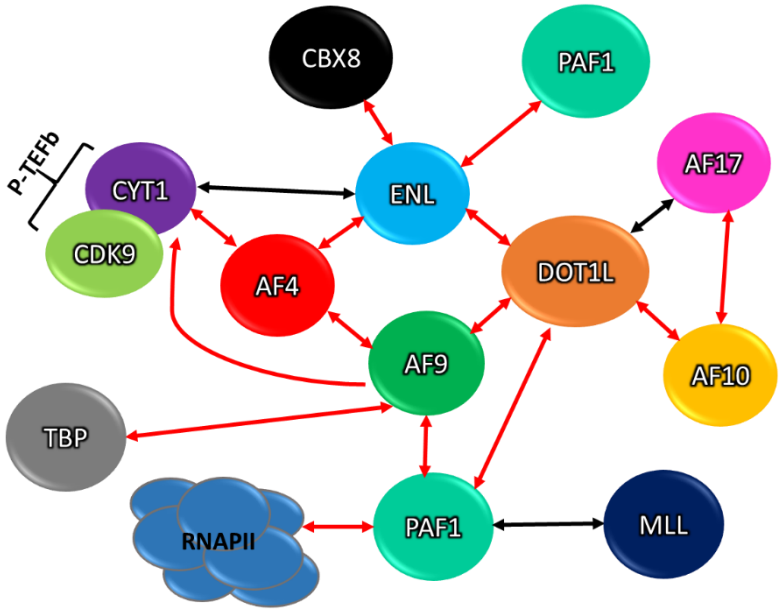

Figure 3-18. Network of proteins verified in vivo using the TetR system. Red arrows demonstrate interactions observed using the TetR system. Black arrows represent interactions not tested in this study.

3.4.4 A working model of DOT1L recruitment at MLL-AF4 gene targets

Taking data from this chapter together, a working hypothesis can be generated to explain an indirect model for DOT1L recruitment at MLL-AF4 gene targets. As MLL-AF4 can interact directly with AF9 and ENL, this can potentially create a local concentration of AF9 and ENL at MLL-AF4 gene targets. Furthermore, MLL-AF4, along with RNAPII and other members of the PAF1 complex, may create a stabile binding site for PAF1, via the interaction with AF9 and ENL. This increased stabilisation of proteins which are sufficient for DOT1L recruitment may lead to the recruitment of DOT1L at MLL-AF4 gene targets, leading to high levels of H3K79me observed within the gene body (Figure 3.19).

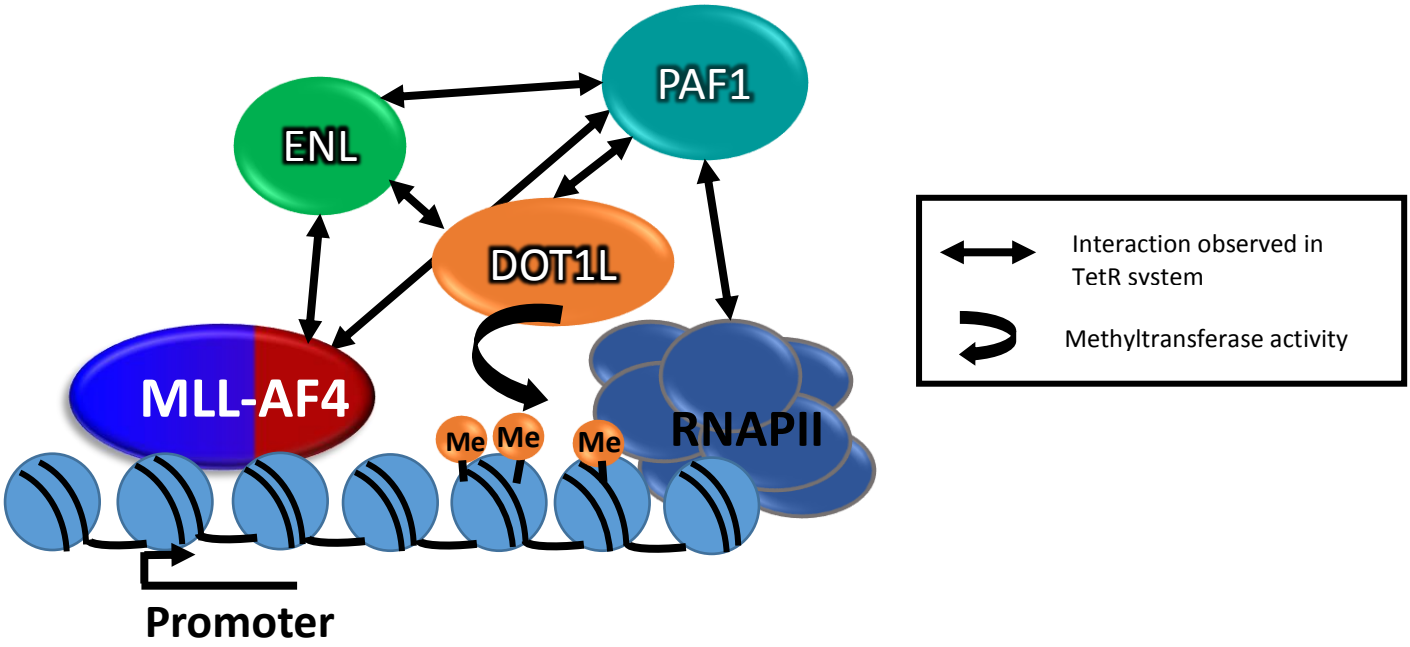

Figure 3-19. Model of DOT1L recruitment at MLL-AF4 gene targets. MLL-AF4 may recruit DOT1L via ENL/AF9 and PAF1. This creates a stable binding site for DOT1L. Double sided arrow represent interactions which have been observed in the TetR system. PAF1 may further stabilise DOT1L via its interaction with RNAPII. This stabilisation may promote H3K79me in the gene body (curved arrow).

Chapter 4 – Defining a set of hypersensitive DOT1L gene targets in MLL-AF4 leukaemia

4.1 Introduction

MLL-AF4 leukaemias arise from a chromosome translocation which creates an MLL-AF4 fusion protein (Moore et al. 2005). MLL-AF4 has been shown to bind and cause aberrant upregulation of gene targets leading to a block in differentiation and maintenance of a stem cell-like state, thus causing leukaemia. The mechanism by which MLL-AF4 leads to the upregulation of gene targets is poorly understood.

One of the most prevalent models is that MLL-AF4 activates gene targets via H3K79me (Krivtsov et al. 2008; Okada et al. 2005). Interestingly, high levels of H3K79me3 have been observed at MLL-AF4 gene targets in comparison to equally expressed targets and relative to the same gene targets expressed in normal CD34/19+ cells (Bernt et al. 2011; Krivtsov et al. 2008). In addition to the high levels observed, H3K79me is functionally important for the expression of certain MLL-AF4 gene targets, including *BCL2* and *RUNX1*, which demonstrate reduced expression following DOT1L inhibition (Wilkinson et al. 2013; Benito et al. 2015)(See introduction). Furthermore, following the inhibition or knockdown of DOT1L, leukaemia is abrogated in MLL-r leukaemia cell lines, mouse and rat models (Krivtsov et al. 2008; Daigle et al. 2011; Daigle et al. 2013; Bernt et al. 2011). Due to the dependence of MLL-r leukaemias in general on DOT1L, EPZ-5676, a SAM analogue which blocks DOT1L methyltransferase activity, was developed and is currently in stage I clinical trials (Waters et al. 2014; Stein et al. 2014). Taken together, H3K79me plays an important role in MLL-AF4 leukaemia. Even though it is known that H3K79me may be important in regulating a subset of MLL-AF4 gene

targets, how universal the requirement is for its activity is unclear (Krivtsov et al. 2008). By identifying DOT1L dependent gene targets in MLL-AF4 leukaemia, further investigation into the functional mechanism of DOT1L at MLL-AF4 gene targets can be carried out.

4.2 Aims

1. Optimisation of EPZ-5676 in MLL-AF4 ALL leukaemia cell line SEM to study DOT1L function at MLL-AF4 gene targets
2. Identify a set of DOT1L dependent MLL-AF4 gene targets in SEM cells
3. Compare DOT1L dependent gene targets to H3K79me gene targets in other cell types

4.3 Results

4.3.1 Inhibition of DOT1L leads to reduction in cell proliferation

EPZ-5676 has been shown to have 37,000-fold selectivity for DOT1L compared to 15 other protein methyltransferases, and did not affect 8 other histone modifications tested by western blot from MLL-r leukaemia cell lines. This suggests off target effects on other protein methyltransferases are minimal (Daigle et al. 2013). Thus EPZ-5676 is an excellent molecular tool to study the specific function of DOT1L and H3K79me in the context of MLL-AF4 leukaemia. As EPZ-5676 has been used in MLL-AF4 leukaemia cell lines extensively by several different labs, a literature review was carried out to investigate the conditions used in published data.

Daigle et al., 2013 extensively characterized EPZ-5676 in the MLL-AF4 leukaemia cell line, MV4;11. Following 6-7 days of treatment with 1μM EPZ-5676, a quantitative ELISA demonstrated that around 80% of global H3K79me2 was lost (Daigle et al. 2013). In addition

to this, expression of known MLL-AF4 gene targets controlled by H3K79me, such as *HOXA9* and *MEIS1*, demonstrated an 80% reduction in expression after 6-7 days of 1μM EPZ-5676 treatment. Finally, it was also demonstrated in this study that anti-proliferative effects were not observed until after day 7 of EPZ-5676 treatment, suggesting that any effect seen on the level of RNA at day 7 was not likely to be due to anti-proliferative or cell death secondary effects.

In this chapter, SEM cells, an MLL-AF4 ALL cell line, is used as a model system. SEM cells were chosen due to the abundance of genomic data already held for SEM cells in our lab and in published databases. Viable, healthy cells produce more ATP than dying or non-proliferating cells. Therefore, to determine the effect of EPZ-5676 treatment on SEM cell viability, an ATP luciferase assay was employed. It is important to note that this assay does not identify the specific cause of a decrease in cell viability. For example, it does not reveal whether the cells are undergoing apoptosis or reduced proliferation. However, due to published data demonstrating both processes occur following EPZ-5676 treatment, an ATP assay gives a broad indication of when an affect upon cell viability in general may occur. SEM cells were treated with 2μM EPZ-5676 compared to 0uM (DMSO only) control for 13 days (Figure 4.1A). A reduction in proliferation was not observed until day 8 of EPZ-5676 treatment, with more obvious effects observed at day 9 (Figure 4.1A).

To determine the dynamic range of the inhibitor, SEM cells were treated with a range of doses of EPZ-5676 of 0μM, 0.5μM (low), 1μM and 2μM (high). As no effect was observed on SEM cells until day 8 with 2μM EPZ-5676, it might be expected that at lower concentrations, 0.5μM and 1μM, an effect upon viability may be delayed. To test this, an ATP assay was carried out at day 9 of EPZ-5676 treatment (Figure 4.1B). As expected, 0.5μM and 1μM show very little effect upon proliferation at this timepoint, so a longer treatment at these concentrations may be

required for the effect on proliferation to become visible. As there is no effect until day 8 on cell viability following any concentration of EPZ-5676 tested (0.5μM, 1μM or 2μM), it was reasoned that a suitable timepoint to observe the effect of DOT1L inhibition on gene expression may be just before these effects become observable, at day 7. This minimises secondary effects related to a decrease in viability, whilst still maximising the effect EPZ-5676 has upon transcription. By using a range of dosages, initial changes in transcription can be observed in the lower concentrations in comparison to more pronounced changes in transcription which may occur in the higher concentration.

4.3.2 DOT1L activity is important for leukaemogenesis of MLL-AF4 leukaemia

To test whether the leukaemic phenotype of SEM cells was perturbed following treatment with EPZ-5676 a colony assay was employed following 7 days treatment with 2μM EPZ-5676 on SEM cells (see methods). Even though there is no obvious effect on proliferation of the bulk culture after 7 days of treatment (Figure 1A), plating these cells and growing them on methylcellulose for 14 days produced a large reduction in colony formation, with just 20% of the colonies formed compared to a 0μM control (Figure 4.1C). In addition, there was a marked difference in colony morphology. 0μM colonies displayed clear, round colonies, representing dividing, stem cell-like cells (Figure 1D). EPZ-5676 treated cells produced very sparse and few colonies, representing cellular differentiation (Gordon 1993) (Figure 4.1D). This suggests that treatment with EPZ-5676 led to an abrogation of the leukaemic phenotype in SEM cells.

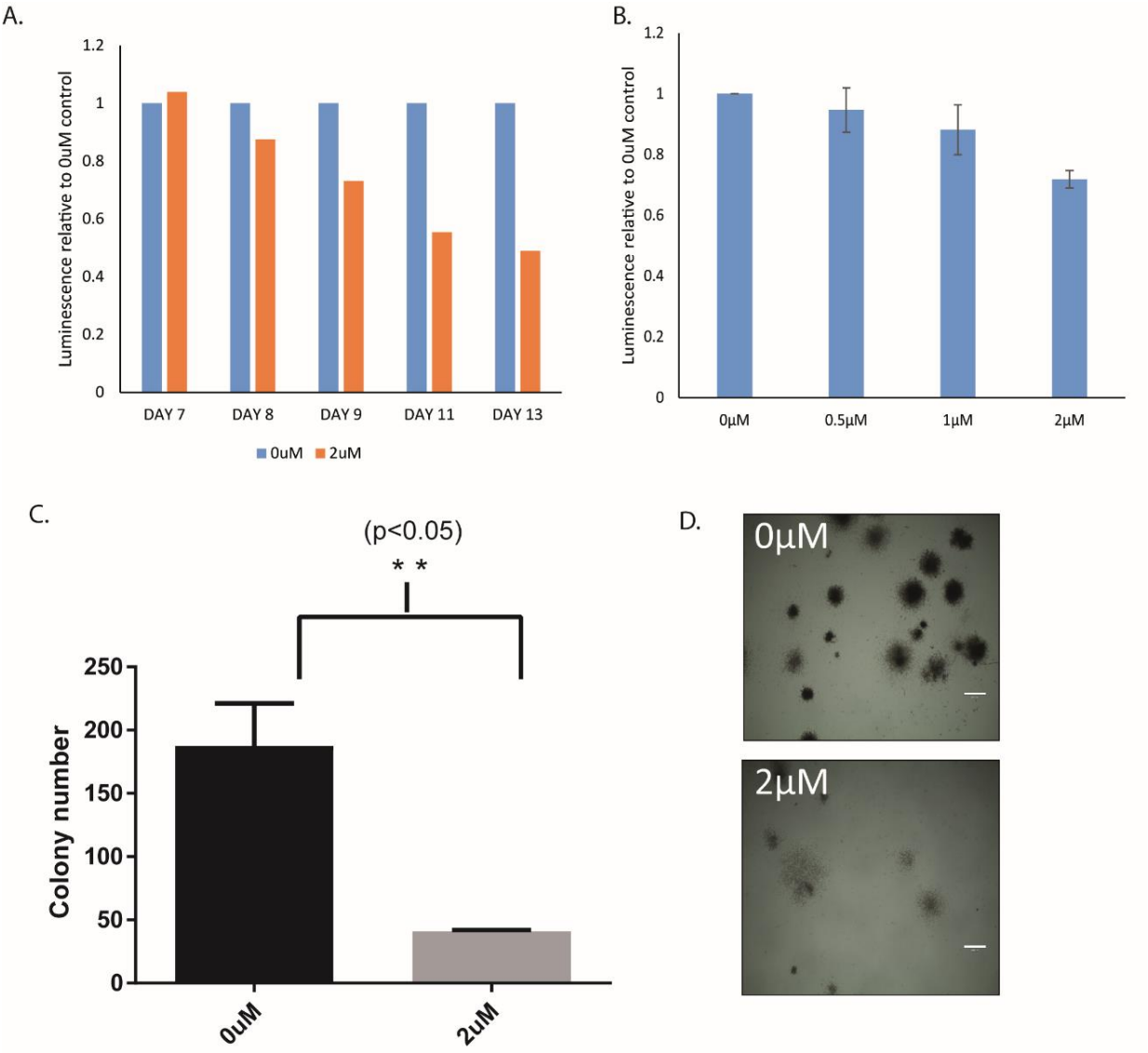

Figure 4-1. SEM cells are sensitive to EPZ-5676 treatment. (A) CellTitre glo ATP assay displaying luminescence of SEM cells following 2µM EPZ-5676 treatment between day 7 and 12 of treatment (B) CellTitre glo ATP assay on SEM cells following 9 days of 0.5µM, 1uM and 2µM EPZ-5676 treatment. Error bars represent standard deviation of three biological replicates (C) Colony assay displaying colony numbers of SEM cells treated with 0µM control and 2µM EPZ-5676. P<0.05 following unpaired parametric t-test (D) Differences in colony morphology between 0µM and 2µM EPZ-5676 treated SEM cells.

4.3.3 Inhibition of DOT1L leads to loss of H3K79 methylation

As EPZ-5676 treatment for 7 days did not affect proliferation, it was necessary to determine how much of a reduction in H3K79me occurs. As described, published studies have shown a gradual and slow reduction in H3K79me following EPZ-5676 treatment of MV4;11 cells (Daigle et al. 2011). To address this, western blot analysis was carried out to determine the levels of bulk H3K79me3 following 7 days of EPZ-5676 treatment at 0μM, 0.5μM, 1μM and 2μM (Figure 4.2A). H3K79me3 demonstrates a dose dependent decrease, with 0.5μM showing a subtle effect and 2μM showing a more obvious effect on bulk H3K79me3 levels. Therefore, treatment with a range of EPZ-5676 dosages coupled with collecting cells at the same timepoint may provide a way to distinguish the effects of transcription following the reduction of H3K79me3 to varying levels.

To determine changes in the level of H3K79me3 at the gene level, H3K79me3 ChIP qPCR was carried out at 0μM, 0.5μM, 1μM and 2μM EPZ-5676 treated SEM cells (Figure 4.2B). H3K79me3 levels demonstrated a decrease in all treatment conditions in comparison to the 0μM control, at *BCL2*, a known H3K79me3-dependent MLL-AF4 gene target. Notably, a large decrease in H3K79me3 is observed at 0.5μM (Figure 4.2B), despite the small change observed for 0.5μM at the bulk level as shown by western blot analysis (Figure 4.2A). This could be due to a locus specific effect at *BCL2* which may not be representative of the effects occurring genome wide or differences in sensitivity between techniques. Furthermore, it is also possible that there is a low level of H3K79me3 genome-wide which contributes to the level observed at bulk histone. Despite this, a dose dependent decrease is observable between 0.5μM, 1μM and 2μM, with an above-background level of H3K79me3 still being present at 2μM.

4.2.4 Loss of DOT1L activity causes reduced gene expression of downstream targets

To determine whether the loss of H3K79me leads to a reduction in gene expression of specific MLL-AF4 gene targets, RT-qPCR was carried out following 7 days of 0.5μM, 1μM and 2μM EPZ-5676 treatment. A dose dependent decrease of the level of mRNA is observed for two known DOT1L sensitive targets; *HOXA9* and *BCL2* (Figure 4.2C). Levels of these two transcripts were reduced to 50-60% following 2μM treatment compared to the 0μM control.

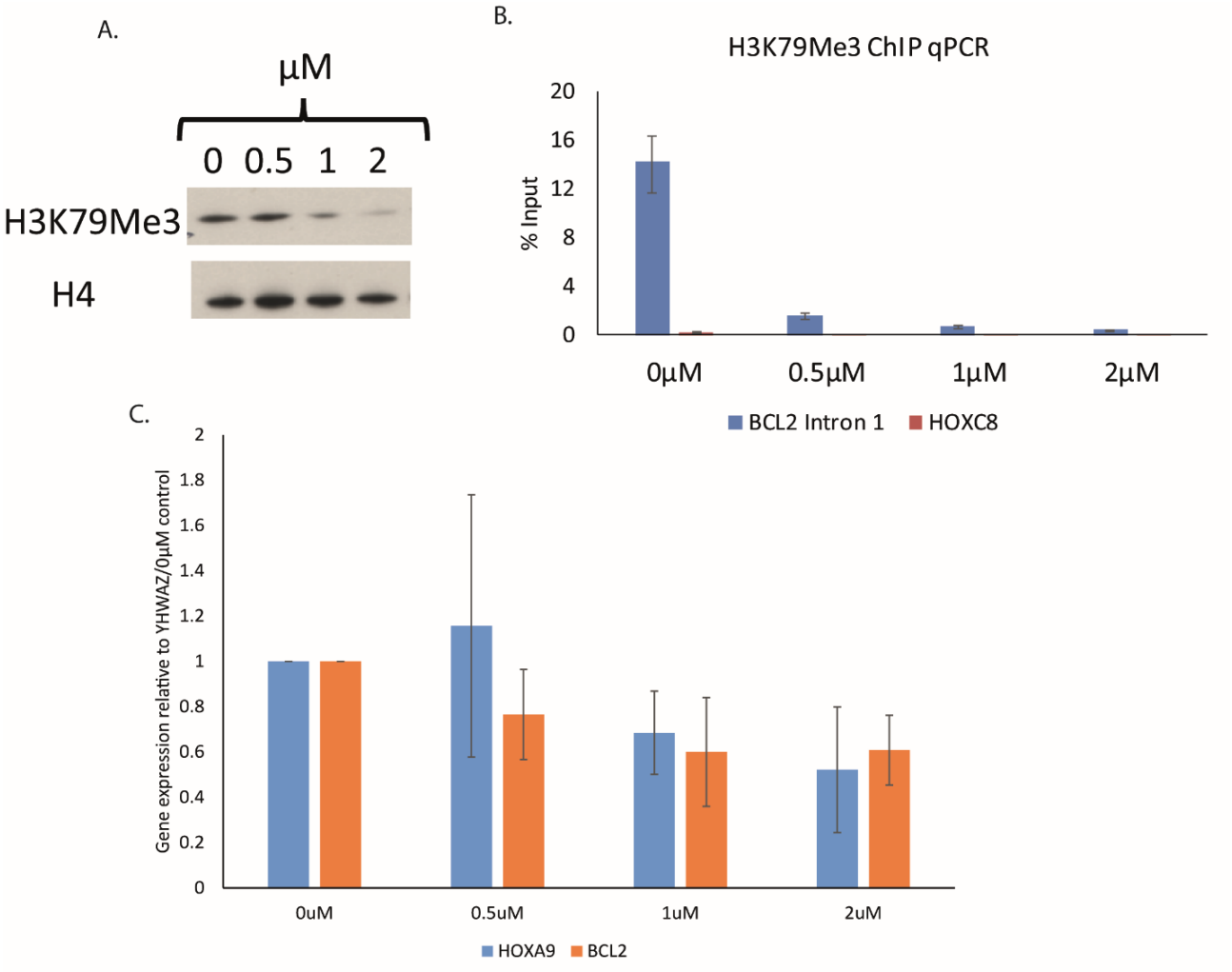

Figure 4-2. H3K79me3 is lost following 7 days EPZ-5676 treatment (A) Western blot analysis of whole cell extracts taken from SEM cells treated with 0.5μM, 1μM and 2μM EPZ-5676 compared to 0μM control. Western blot membrane probed with anti-H3K79me3 antibody and anti-H4 is used as loading control. Representative of 3 biological replicates. (B) ChIP qPCR using antibodies against H3K79me3 in 0μM, 0.5μM, 1μM and 2μM treatments at *BCL2*. *HOXC8* is also shown as a negative control for H3K79me3 signal (C) qRT-PCR generated using cDNA from SEM cells treated with 0.5μM, 1μM and 2μM EPZ-5676 for 7 days at *BCL2* and *HOXA9*. Error bars are representative of three biological replicates.

4.3.5 EPZ-5676 treatment reveals up- and downregulated gene targets by nascent RNA seq

Nascent RNA seq was chosen to identify genes which were dependent upon H3K79me. This technique was chosen because dynamic differential changes occurring at the level of transcription can be masked by steady state levels of RNA. In brief, nascent RNA seq uses 4-thiouridine (4-SU) labelling of nascent transcripts for one hour followed by total RNA extraction (see methods for full details). Following this, the 4-SU-labelled transcripts are purified from the total RNA and sequenced. Nascent RNA sequencing was carried out following 7 days of EPZ-5676 treatment at 0µM, 0.5µM, 1µM and 2µM, in triplicate allowing for differential gene expression analysis using featureCounts and the edgeR statistical package. Differential gene expression analysis was performed by Emmanouela Repapi (Computational Biology Research Group). QC analysis for each experiment was undertaken including western blot, ChIP qPCR, and ATP assay to verify the inhibition and effect upon proliferation (data not shown).

Following EPZ-5676 treatment, up-regulated, down-regulated and unchanged genes were observed in all concentrations in comparison to the 0µM control (Figure 4.3). Differentially expressed genes were defined as those whose expression showed a log2 fold change of more than 0.5 and an adjusted *p*-Value (FDR) of less than 0.05. With increasing concentrations of EPZ-5676, additional genes became differentially expressed. In the 0.5µM treatment, there were 278 gene targets downregulated and 708 targets upregulated. In the 1µM treatment, there were 583 downregulated gene targets and 1387 upregulated targets. Finally, in the 2µM, there were 2462 gene targets downregulated and 2600 upregulated gene targets. In all concentrations, there was a large proportion of insensitive gene targets (Figure 4.4B).

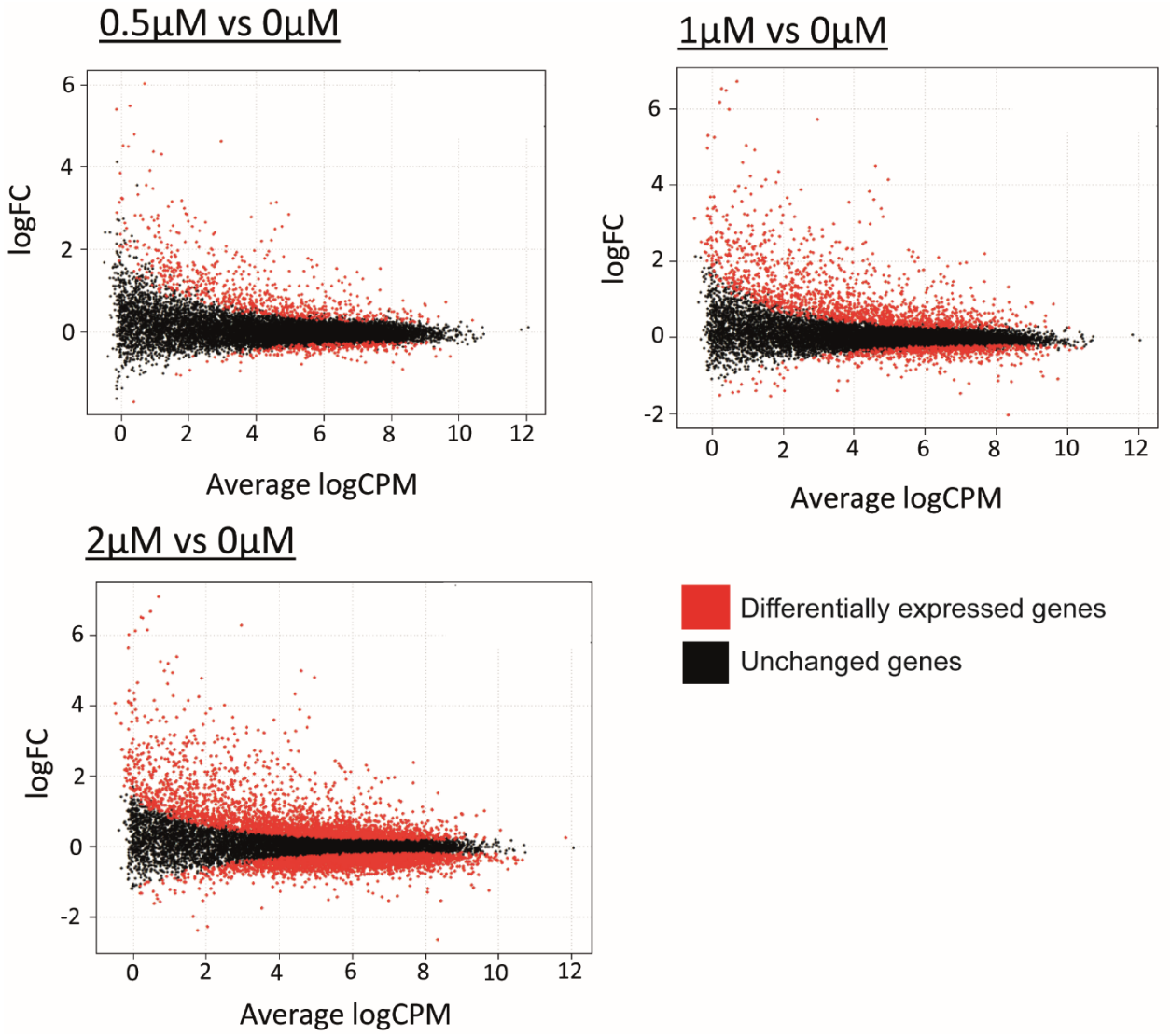

Figure 4-3. Nascent RNA seq reveals differentially expressed genes. SEM cells were treated for 7 days with 0.5µM, 1µM and 2µM EPZ-5676. Nascent RNA was extracted and analysed to determine differentially expressed genes. Correlation plots displaying LogFC on Y-axis and the average logCPM on x-axis for each treatment compared to 0µM control. In red are differentially expressed genes ($p<0.05$) and in black are unchanged genes, with each point representative of one gene (N=3).

4.3.6 Downregulated gene targets are enriched for H3K79me targets

It is possible that many of the differentially expressed genes observed were the result of secondary effects. Furthermore, as H3K79me3 is linked with gene activation, a prediction which could be made is that EPZ-5676 treatment will lead to the downregulation of genes, genome-wide. One caveat of RNA sequencing analysis is that it assumes that the vast majority of gene expression does not change and normalises the data from each sample to achieve this. Therefore, global changes in gene expression between samples, which might be biologically relevant, could be masked by normalisation and redistribution of the data. By adjusting the distribution profile, many downregulated genes may be mis-assigned as unchanged, and many unchanged genes may be mis-assigned as upregulated. Therefore, although genes which are downregulated are almost certainly downregulated it is difficult to pinpoint whether genes are upregulated and unchanged. To demonstrate this, exogenous normalisation, similar to that conducted using ChIP-rx, would be required.

To control for these caveats, targets that were associated with detectable H3K79me levels and were therefore more likely to be directly regulated by DOT1L were focused upon to establish a core set of H3K79me gene targets. H3K79me2 and H3K79me3 ChIP seq datasets generated previously in the lab were peak called using homer software (see methods). H3K79me2 and H3K79me3 data sets identified 11547 and 9269 gene targets, respectively (Figure 4.4A). Differences may be due to technical differences in the ChIP protocol, such as antibody efficiency, or may suggest more gene targets are marked with H3K79me2 rather than the H3K79me3, consistent with current models of DOT1L as a non-processive methyltransferase (Frederiks et al. 2008a). Overlap between the two sets identified a core set of 9193 H3K79me targets common to both H3K79me2 and H3K79me3 datasets (Figure 4.4A).

Up-, down-regulated and insensitive gene targets from all EPZ-5676 treatment conditions were compared to the core set of H3K79me targets. In all conditions, 80-85% of downregulated gene targets were H3K79me targets (Figure 4.4B). In contrast, only a minority of the upregulated gene targets (18-34%) were H3K79me targets. Although this small subset of upregulated targets may be interesting, we chose to initially focus upon the downregulated gene targets, which is consistent with the association of H3K79me with gene activation (Steger et al. 2008). In addition, many upregulated gene targets seem to be low expressing targets which become upregulated by a small amount. Interestingly, a large proportion (56-62%) of the insensitive gene targets in all conditions were H3K79me targets. This could suggest that H3K79me is not sufficiently depleted at these targets to influence transcription or that only a subset of H3K79me genes are sensitive to the loss of H3K79me.

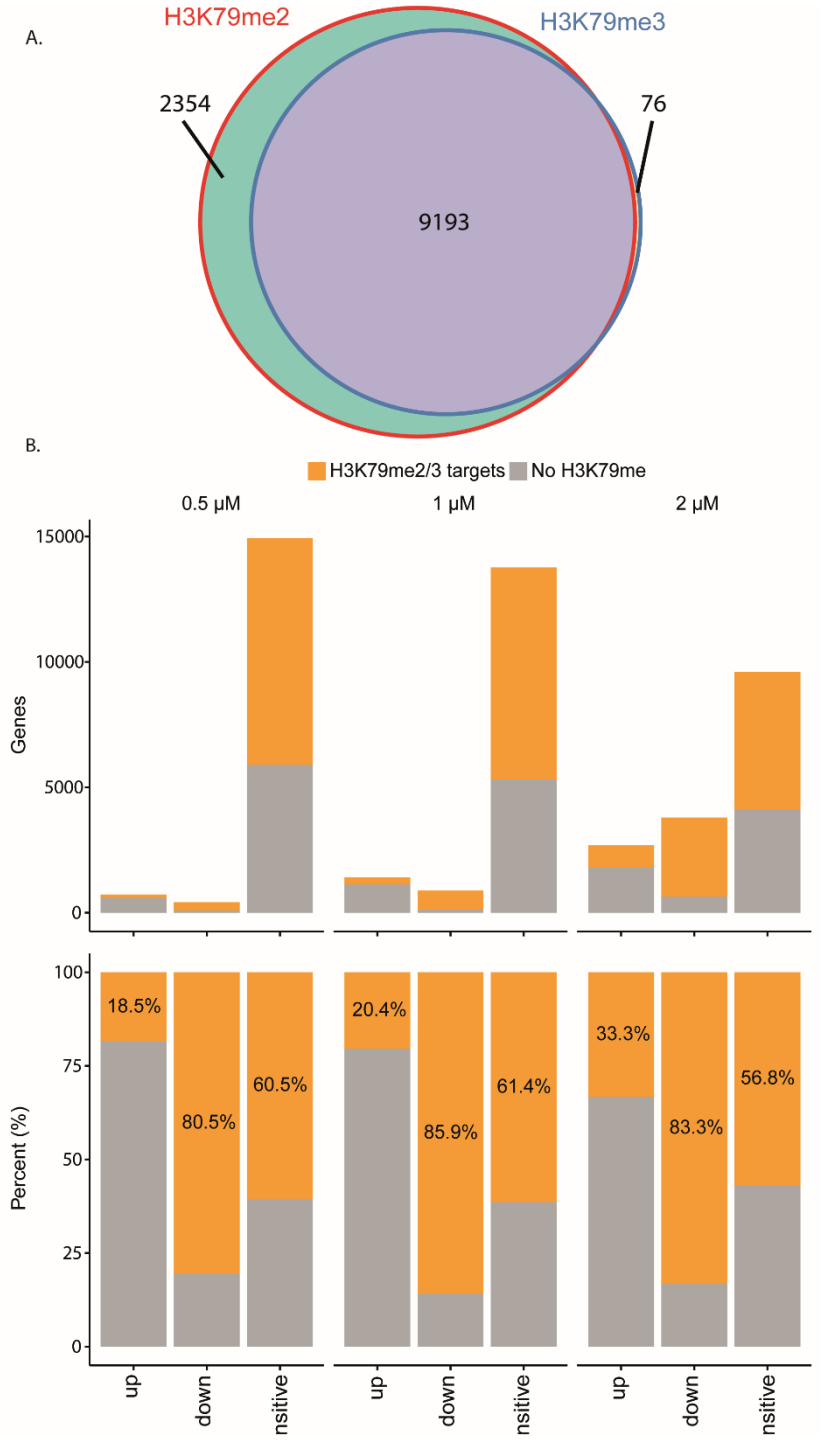

Figure 4-4. H3K79me2/3 marks a subset of downregulated genes. (A) Overlap between H3K79me2 and H3K79me3 ChIP seq data sets currently in the lab performed by Tom Milne and Erica Ballabio (B) Top: Number of genes in each category; upregulated, downregulated and insensitive H3K79me gene targets in 0.5µM, 1µM and 2µM concentrations. Bottom: percentage of genes in each category; upregulated, downregulated and insensitive H3K79me gene targets in 0.5µM, 1µM and 2µM concentrations which are H3K79me direct targets.

4.3.7 Defining a set of hypersensitive DOT1L gene targets

To study DOT1L function at MLL-AF4 gene targets, it is important to define a set of genes which are most responsive to EPZ-5676. To do this a set of 'hypersensitive' genes were defined as those genes which are downregulated in all dosages of inhibitor. 182 hypersensitive gene targets were downregulated in all treatments (Figure 4.5A). Within the group of 182 hypersensitive targets, 71% (130) were direct H3K79me targets (Figure 4.5B). Overall, although it still cannot be totally ruled out that some of these hypersensitive targets may be downregulated due to secondary effects, there is a much higher confidence that these 130 targets are downregulated directly due to the loss of H3K79me. As these gene targets have been identified using nascent RNA sequencing, it is important to verify that the downregulation or insensitivity is not an artefact of the sequencing technique. To test this, qRT-PCR was performed using Nascent RNA at *PROM1*, a hypersensitive gene, and *FLT3*, an insensitive gene (Figure 4.5C).

4.3.8 Hypersensitive targets have initial high H3K79me and high expression levels

As a proportion of both hypersensitive and insensitive gene targets are marked by H3K79me, it is possible that there is something different about these two groups which determine that one is sensitive and one is not. One obvious difference could be that levels of H3K79me and levels of expression may be different between these two groups. To test this, the level of H3K79me and expression at 0μM were compared between the 130 hypersensitive H3K79me marked genes and H3K79me genes that were insensitive to 2μM EPZ-5676 (Figure 4.6A). From this analysis, hypersensitive H3K79me gene targets seem to have slightly higher levels of H3K79me than the insensitive targets. In addition to this, hypersensitive H3K79me gene

targets are slightly more highly expressed than the insensitive targets (Figure 4.6B). This may be significant in understanding why these targets are more sensitive to EPZ-5676 treatment.

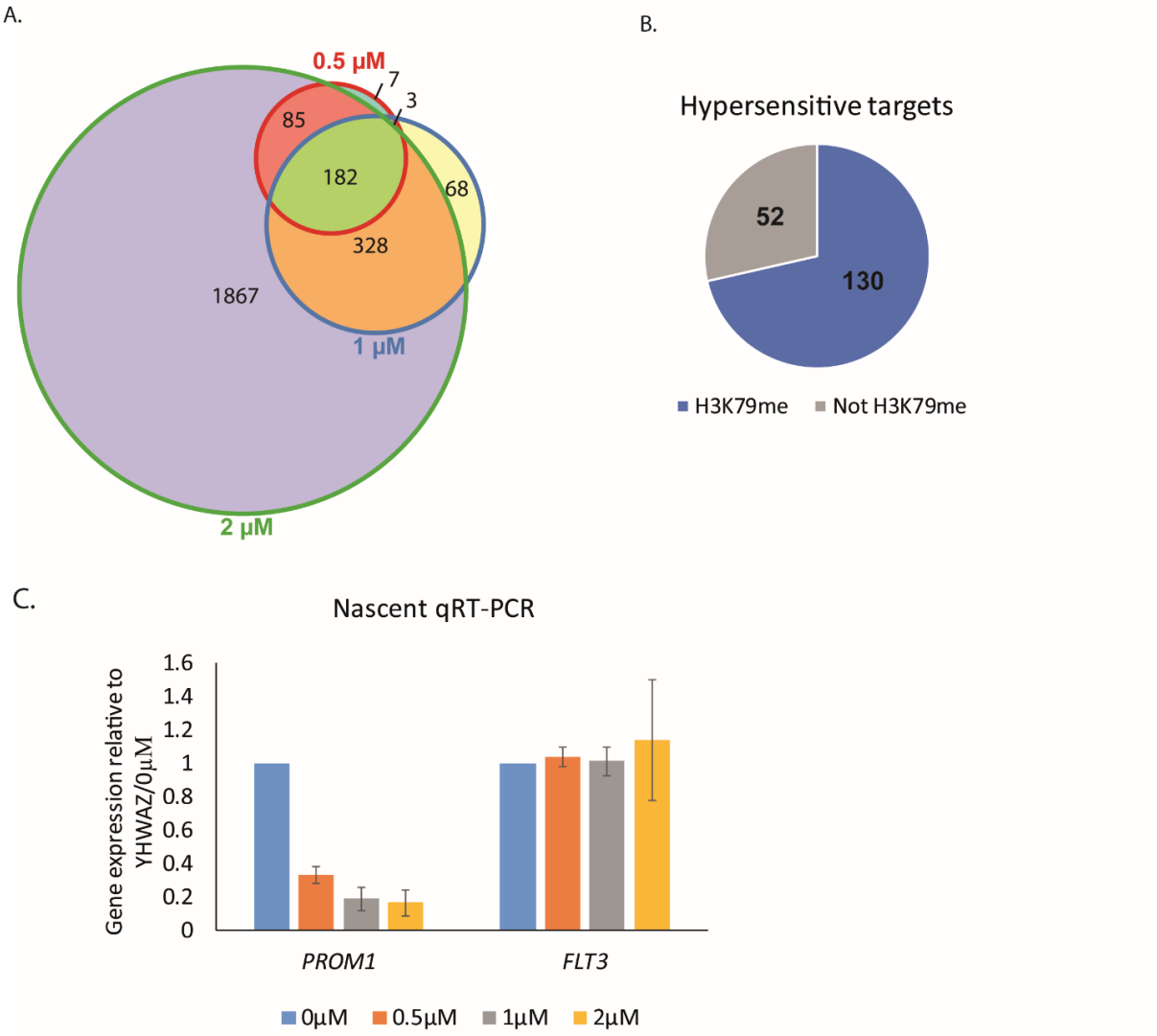

Figure 4-5. Identification of a subset of hypersensitive downregulated gene targets (A) Overlap between all downregulated gene targets from 0.5µM, 1µM and 2µM EPZ-5676 treatments reveals 182 hypersensitive gene targets. (B) Proportion of hypersensitive genes which are H3K79me2/3 direct targets. (C) qRT-PCR performed on cDNA generated from nascent RNA following 7 days EPZ-5676 treatment at hypersensitive gene *PROM1* and insensitive gene *FLT3*. Error bars represent three biological replicates.

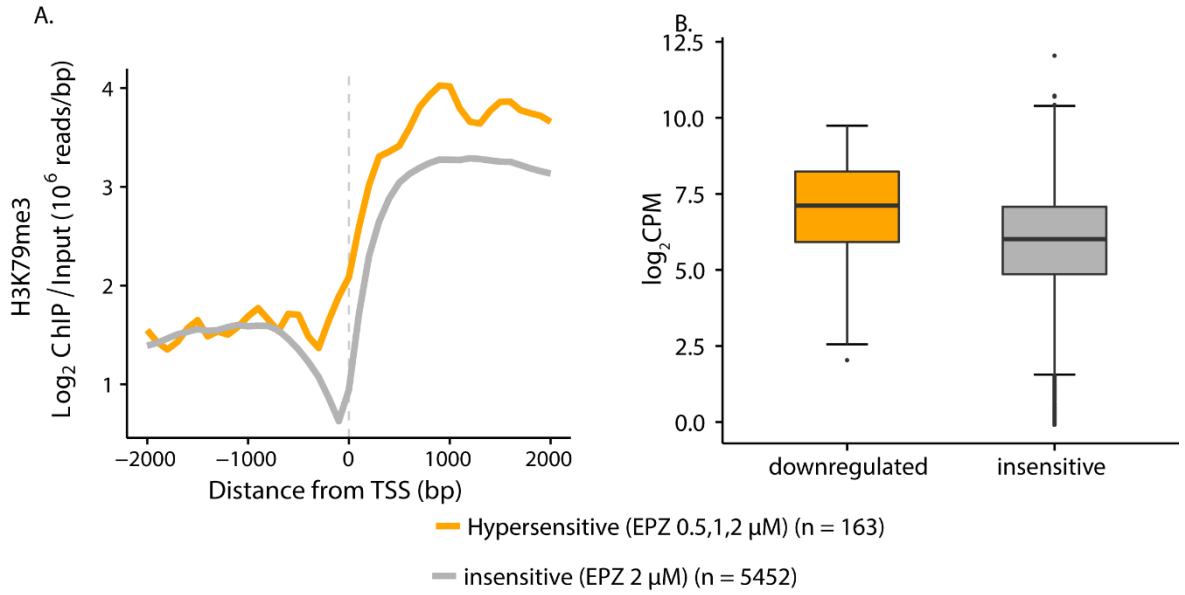

Figure 4-6. Hypersensitive gene targets demonstrate high levels of expression and H3K79me3 levels (A) Histogram demonstrating the level of H3K79me3 hypersensitive targets (orange) compared to insensitive gene targets from 2µM EPZ-5676 treatment (grey). Y-axis displays H3K79me3 signal and x-axis displays +/- 2000bp from the TSS (B) Box plot displaying logCPM of hypersensitive gene targets compared to insensitive targets from the 2µM EPZ-5676 treatment

4.3.9 Hypersensitive gene targets are important for gene regulation

EPZ-5676 treatment of SEM cells results in increased PARP cleavage, indicating initiation of programmed cell death (Godfrey et al. 2016). This suggests that differentially expressed genes identified following EPZ-5676 treatment may be important for maintenance of the leukaemia. To determine whether H3K79me marked hypersensitive genes (130) are enriched for certain molecular pathways or functions, a Metacore analysis was conducted. Hypersensitive genes were enriched for transcription and DNA binding functions, consistent with the role of DOT1L in active transcription (Table 4.1). This highlights that following EPZ-5676 treatment

downstream targets associated with transcription are affected which may contribute to the abrogation of the leukaemia.

Table 4-1. Metacore analysis of H3K79me marked hypersensitive genes (Top) Process analysis (Bottom) Molecular function analysis.

#	Processes	pValue	FDR
1	positive regulation of nucleobase-containing compound metabolic process	9.924E-15	4.443E-11
2	positive regulation of nitrogen compound metabolic process	2.383E-14	5.335E-11
3	positive regulation of RNA metabolic process	3.764E-14	5.617E-11
4	negative regulation of biosynthetic process	9.256E-14	9.553E-11
5	positive regulation of nucleic acid-templated transcription	1.280E-13	9.553E-11
6	positive regulation of transcription, DNA-templated	1.280E-13	9.553E-11
7	negative regulation of macromolecule metabolic process	1.860E-13	1.190E-10
8	positive regulation of gene expression	2.187E-13	1.222E-10
9	positive regulation of RNA biosynthetic process	2.456E-13	1.222E-10
10	negative regulation of macromolecule biosynthetic process	4.024E-13	1.801E-10

#	Molecular functions	pValue	FDR
1	protein binding	1.427E-09	1.012E-06
2	transcription regulatory region DNA binding	1.283E-08	2.419E-06
3	regulatory region DNA binding	1.416E-08	2.419E-06
4	regulatory region nucleic acid binding	1.525E-08	2.419E-06
5	nucleic acid binding transcription factor activity	2.047E-08	2.419E-06
6	sequence-specific DNA binding transcription factor activity	2.047E-08	2.419E-06
7	chromatin DNA binding	5.830E-08	5.905E-06
8	sequence-specific DNA binding	5.051E-07	4.477E-05
9	transcription factor binding transcription factor activity	2.914E-06	2.265E-04
10	protein binding transcription factor activity	3.194E-06	2.265E-04

4.3.10 A subset of hypersensitive genes are MLL-AF4 gene targets

To further understand the role of DOT1L at MLL-AF4 gene targets it was important to identify the proportion of differentially expressed MLL-AF4 gene targets. To do this, a set of MLL-AF4 bound genes (previously established in the lab using an overlap between ChIP seq datasets of MLL (N) bound targets and AF4 bound targets (Wilkinson et al. 2013; Kerry et al. 2017) were compared to the up-regulated, downregulated and insensitive gene targets in each EPZ-5676 treatment (Figure 4.7A). A similar proportion (20-30%) of MLL-AF4 gene targets was observed in most categories and conditions. Slightly more were observed in the downregulated genes in 0.5μM and 1μM conditions of between 45-50%, with only 28% observed in the downregulated 2μM condition. This may be due to an increase in differentially expressed gene targets in the 2μM condition, with more non-MLL-AF4 gene targets becoming downregulated. To focus specifically on the directly DOT1L regulated MLL-AF4 gene targets, the overlap between MLL-AF4 gene targets and H3K79me marked hypersensitive genes were assessed. Strikingly, 66% of the hypersensitive H3K79me genes were bound by MLL-AF4 (Figure 4.7B). Therefore, whilst in general MLL-AF4 gene targets are not enriched in the total downregulated genes compared to total up-regulated or total insensitive targets, hypersensitive targets are strongly enriched with gene targets bound by MLL-AF4. This highlights the potential importance of the hypersensitive targets for leukaemogenesis. Furthermore, as MLL-AF4 gene targets have been observed to have extremely high levels of H3K79me (Krivtsov et al. 2008), this may explain the difference in levels of H3K79me observed in the hypersensitive targets (Figure 4.6A) which may in part come from the MLL-AF4 targets within this group. Amongst the hypersensitive MLL-AF4 gene targets were *HOXA9, CDK6* and other targets, which have previously been shown to have high levels of H3K79me (Krivtsov et al. 2008; Bernt & Armstrong 2011; Okada et al. 2005). In addition, two of the most sensitive gene targets

identified were the MLL-AF4 targets *PROM1*, which encodes the cell surface marker CD133, and *BCL11A*, which encodes a transcription factor important for normal B cell development. *PROM1* has been shown to be downregulated following AF4 knockdown as well as being sensitive to MLL-AF4 knockdown (Mak et al. 2012; Thomas et al. 2005; Kerry et al. 2017). Interestingly, other known MLL-AF4 gene targets, such as *FLT3*, were insensitive to EPZ-5676 treatment and showed no differential change in nascent RNA level, suggesting MLL-AF4 may contribute to the regulation of these genes via an alternative mechanism.

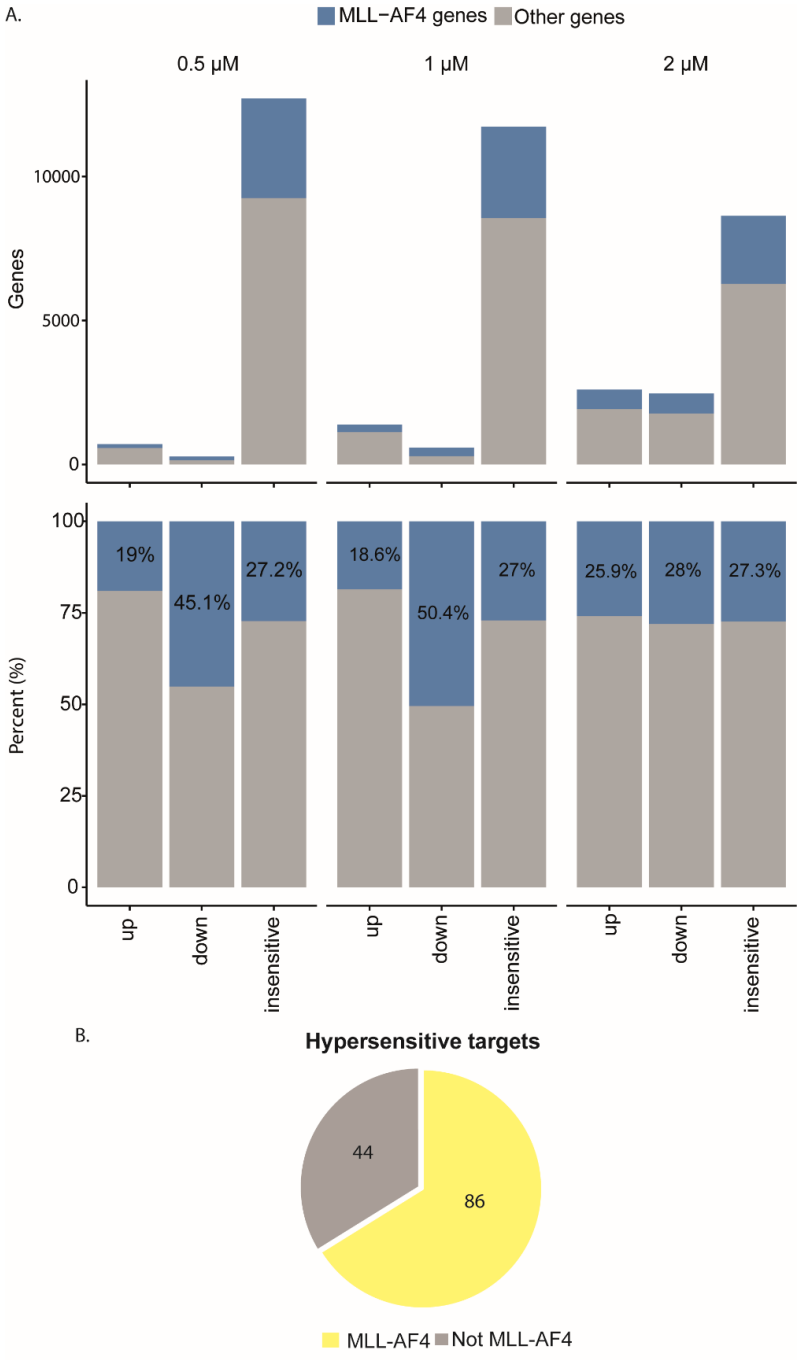

Figure 4-7. Proportion of hypersensitive targets are MLL-AF4 targets (A) Number of genes in each category; upregulated, downregulated and insensitive which are bound by MLL-AF4. MLL and AF4 ChIP seq performed in the lab by Tom Milne. Bottom: percentage of genes in each category: upregulated, downregulated and insensitive in 0.5µM, 1µM and 2µM concentrations which are MLL-AF4 direct targets (B) proportion of 182 hypersensitive targets (Figure 4.5A) which are bound by MLL-AF4.

4.3.11 H3K79me3 is lost genome-wide following EPZ-5676 treatment

As there is a large proportion of H3K79me gene targets which are insensitive, one possibility is that EPZ-5676 treatment did not lead to full H3K79me depletion genome-wide. To address this, a quantitative H3K79me3 ChIP seq method was employed.

4.3.11.1 Quantitative ChIP seq

ChIP seq is traditionally normalised by Reads Per Million (RPM), which normalises for the total read count but does not always account for differences observed between samples, even when using the same antibody. For example, sample A may have 100% of histone modification present compared to sample B which only has 50% of the amount of histone modification present in sample A. These differences may be masked when using traditional normalisation because, even though sample B has overall less signal, it can be sequenced to a greater depth due to the lower complexity of the library being sequenced. Sample A, which has 100% modification will not be sequenced as deeply due to the larger amount of global modification and higher library complexity. Following normalisation to total number of reads, this may mean that sample A and B look very similar following sequencing (Figure 4.8A). To account for these differences, an external reference genome normalisation method can be used. By normalizing to a reference genome, which is present at the same amount in all samples, these differences in amount of modification can be revealed (Figure 4.8A). In addition to this, technical differences in the ChIP protocol are also accounted for as the reference DNA undergoes the same process within each sample.

4.3.11.2 Observing the loss of H3K79me3 using ChIP rx

To observe global changes in H3K79me the ChIP rx method was used (described in Orlando *et al*., 2015). A known number of fixed *D.melanogaster* S2 cells are 'spiked in' at a 4:1 ratio (human vs *D D.melanogaster* at the lysis step of the ChIP protocol. The experiment is then processed as per the normal ChIP protocol (see methods). Following sequencing, a normalisation factor can be calculated based upon the number of *D.melanogaster* reads present in each sample. The input of each ChIP sample should give rise to a similar number of *D.melanogaster* reads, due to the constant number of *D.melanogaster* cells used in each ChIP. Differences observed in the number of *D.melanogaster* reads in the IP samples are due to differences in amount of material pulled down which is an indication of the differences in amounts of modification present.

As the *D.melanogaster* cells are lysed at the same time as the human cells and taken through the IP, the antibody used in the IP must recognise the *D.melanogaster* histone modification. To test this, ChIP qPCR was carried out on S2 cells using an anti-H3K79me3 antibody. Primers were designed for known active and repressed genes in *D.melanogaster*. H3K79me3 was detectable by qPCR at the active gene (*rp49*) and not at the repressed gene (*xm5a*) (Figure 4.8B).

Following antibody verification, SEM cells were treated with 0.5µM, 1µM and 2µM EPZ-5676 for 7 days, compared to 0µM control and single-fixed for ChIP. Following sequencing, reads were mapped and filtered to hg19 and dm3 using an in-house pipeline developed by Jelena Telenius. The normalisation factors were calculated and reads were normalised using the homer

makeUCSCfile command using the –norm flag. Traditional normalisation was also carried out to highlight the differences observed between these two methods.

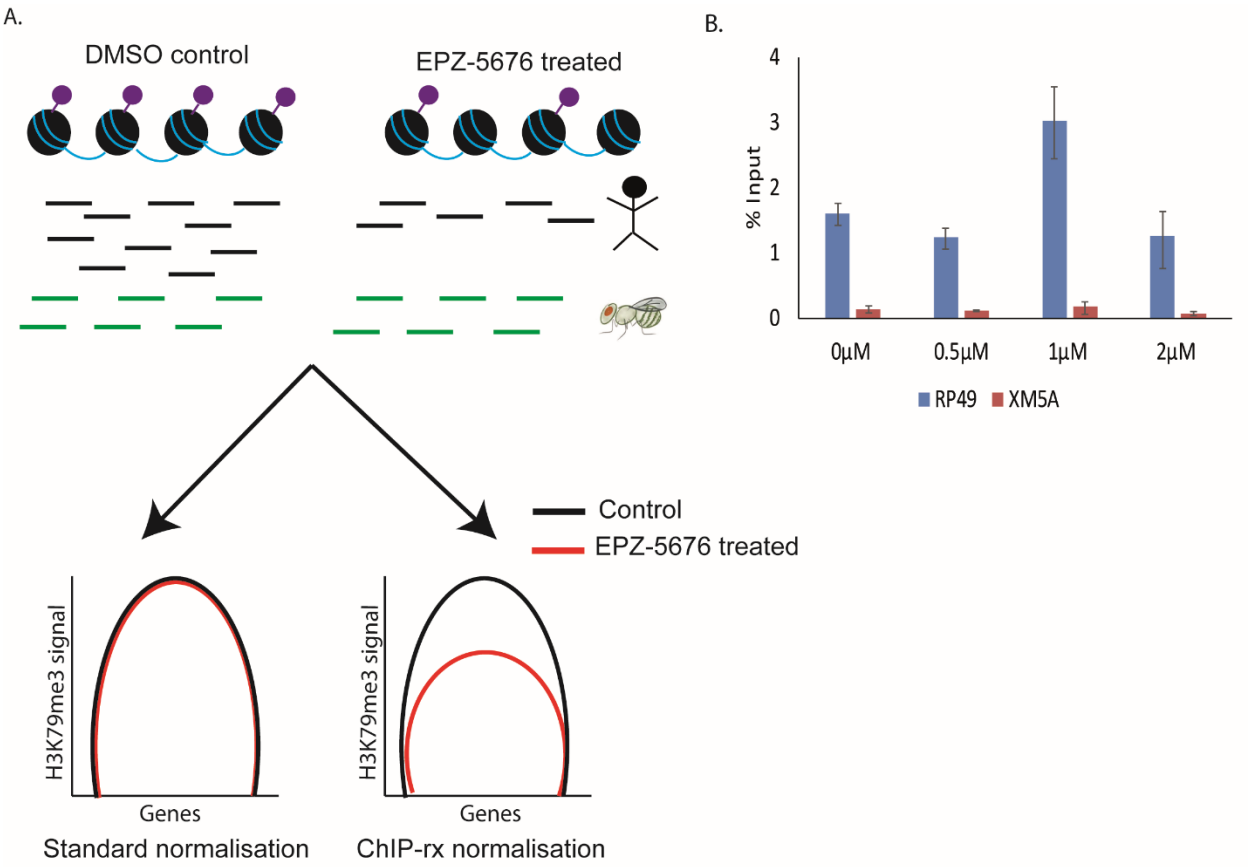

Figure 4-8. Using ChIP rx to study loss of H3K79me following EPZ-5676 treatment. (A) Schematic describing the ChIP rx principle. Adapted from Orlando et al., 2014. (B) H3K79me3 ChIP qPCR in Drosophila S2 cells at an expressed gene (rp49) and a repressed gene (xm5a) from spike-in experiment with SEM cells treated with 0.5μM, 1μM and 2μM EPZ-5676 for 7 days.

A genome wide analysis was carried out observing all genes in all treatments using both traditional and ChIP rx normalisation factors (Figure 4.9A-B). Using ChIP rx normalisation it can be observed that H3K79me3 is lost genome wide in 0.5μM, 1μM and 2μM conditions compared to the 0μM control. This loss is striking compared to the traditional normalisation which demonstrates very little change in H3K79me3, and even an increase in the 1μM

treatment genome wide. This highlights the importance of using reference normalisation to demonstrate the loss of H3K79me genome wide which would be otherwise masked using traditional normalisation. Furthermore, this demonstrates that both sensitive and insensitive gene targets lose H3K79me3. This can be further illustrated by visualising the loss of H3K79me3 both at *PROM1*, a hypersensitive gene and *FLT3*, an insensitive gene (Figure 4.9C-D). Again, the difference between traditional and ChIP rx normalised data can be clearly observed.

Even though slight differences are observed between concentrations at the ChIP rx level, these changes contrast with those observed at the bulk histone level by western blot (Figure 4.2A). This could be due to sensitivity of ChIP seq compared to western blot analysis, or perhaps western blot detects a pool of H3K79me marked histones that are not incorporated into chromatin. Nascent RNA sequencing tracks demonstrate that although H3K79me3 is lost at both *PROM1* and *FLT3*, only *PROM1* expression is sensitive to the loss of H3K79me3 and *FLT3* is insensitive, showing no reduction in transcription (Figure 4.10).

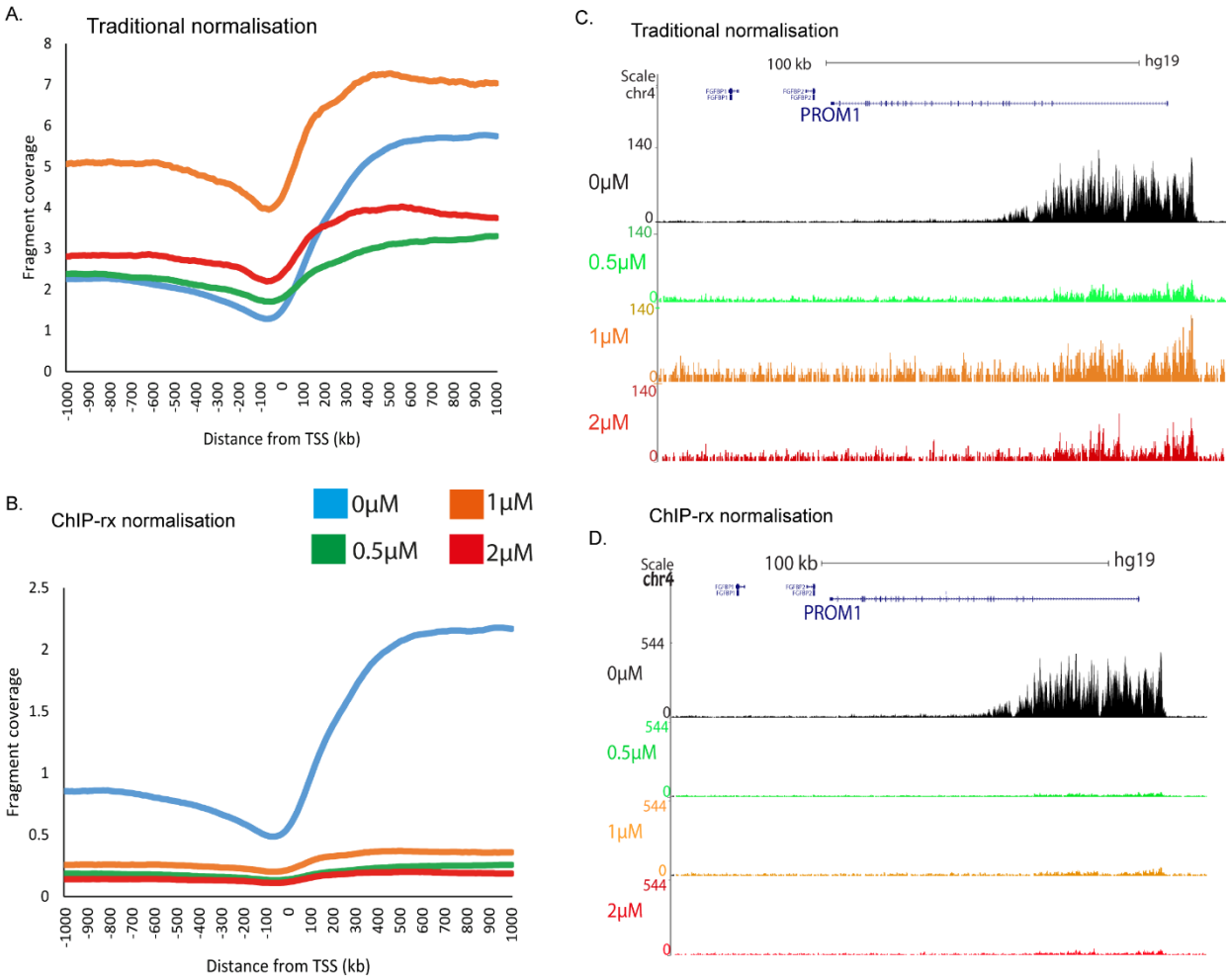

Figure 4-9. H3K79me3 is lost genome wide following EPZ-5676 treatment (A-B) UCSC genome browser tracks display H3K79me3 levels at hypersensitive gene targets using a traditional normalisation method and ChIP rx normalisation method from SEM cells treated with 0.5μM, 1μM and 2μM EPZ-5676 (C-D) Histogram demonstrating genome wide difference in H3K79me3 following 7 days 0.5μM, 1μM and 2μM EPZ-5676 treatment using traditional and ChIP rx normalisation methods.

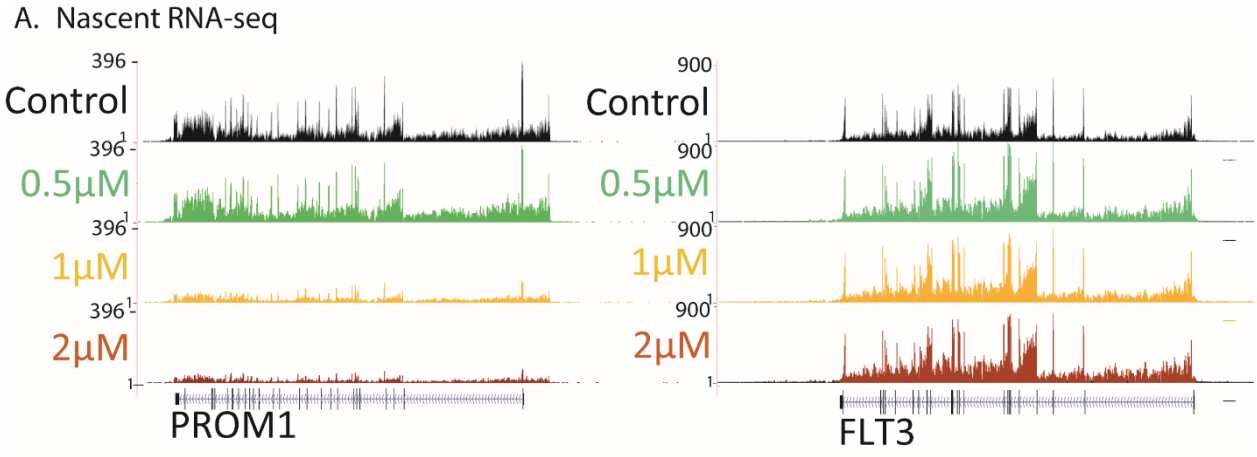

Figure 4-10. Examples of sensitive and insensitive targets. UCSC tracks displaying Nascent RNA seq data, normalised to $1x10^7$ reads, in 0µM control, 0.5µM, 1µM and 2µM EPZ-5676 treatments of SEM cells at *PROM1* and *FLT3*.

4.3.12 DOT1L is important for the regulation of MLL-AF4 gene target, *PROM1*

As the most downregulated gene target, *PROM1*, codes for the cell surface marker CD133, it was interesting to determine whether downregulation of this gene target leads to a reduction of CD133 on the cell surface, particularly because changes in nascent RNA are reduced but not completely lost. FACS analysis was carried out on SEM cells, showing a reduction in surface CD133 following EPZ-5676 treatment (Figure 4.11A, performed by Sorcha O'Byrne). Complementary to this, western blot analysis demonstrated a reduction of intracellular PROM1 (Figure 4.11B).

To determine whether CD133 was important for leukaemogenesis of MLL-AF4 ALL leukaemia (*PROM1* is not expressed in MLL-AF4 AML), *PROM1* siRNA knockdown coupled with colony assays was performed. FACS analysis demonstrated that CD133 was reduced following siRNA knockdown (data not shown), however no difference in colony formation or morphology was observed between scrambled siRNA control and *PROM1* siRNA (Figure

4.11C). This suggests *PROM1* may not necessary for the maintenance of MLL-AF4 ALL, however, it was possible that *PROM1* levels had recovered to wildtype levels. To address this issue, *PROM1* shRNA cell lines were generated. Unfortunately, the cells died within 24 hours of shRNA induction, therefore, whether expression had been reduced was unclear.

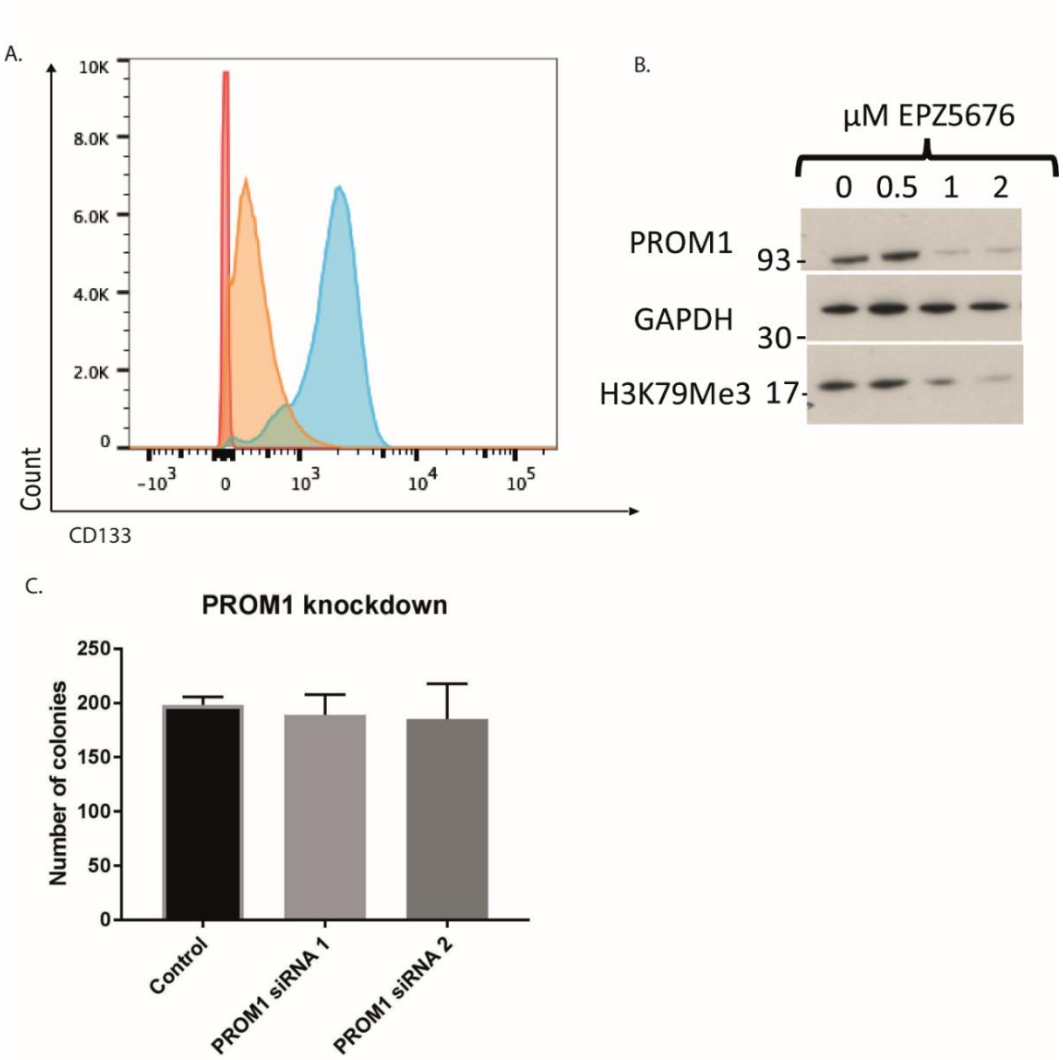

Figure 4-11. CD133 expression is lost following EPZ-5676 treatment (A) CD133 FACS analysis of SEM cells treated with 2µM EPZ-5676 for 7 days (orange) compared to 0µM control (blue) and negative control (red) (B) Western blot analysis of WCEs generated from SEM cells treated with 2µM EPZ-5676 for 7 days compared to 0µM control. Western blot membrane probed with anti-PROM1 antibody, Anti-H3K79me3 and anti-GAPDH as loading control (C) Two independent PROM1 siRNA mediated knockdowns of SEM cells compared to scrambled siRNA control coupled with colony assays. Y-axis indicates number of colonies counted. Error bars represent standard error (n=3).

4.3.13 DOT1L knockdown validates hypersensitive gene targets

Following identification of a set of hypersensitive gene targets using EPZ-5676, a complementary approach was needed to validate these targets. To do this a DOT1L specific siRNA was employed. SEM cells were treated with DOT1L siRNA for 6 days. After 2 days of knockdown, DOT1L levels were reduced on the protein level (Figure 4.12B) and RNA levels remain reduced at day 6 (Figure 4.12A). In addition to this H3K79me3 shows a clear reduction on the bulk histone level by western blot at day 4 and 6 (Figure 4.12C). This is further verified by H3K79me3 ChIP qPCR at day 6, demonstrating a loss of H3K79me3 at sensitive targets (*PROM1*, *CDK6* and *BCL11A*) and an insensitive target (*FLT3*) compared to the scrambled siRNA control (Figure 4.12D). Interestingly, consistent with EPZ-5676 treatment, hypersensitive targets *PROM1*, *BCL11A* and *CDK6* RNA levels were reduced following DOT1L knockdown by around 20% while the insensitive gene *FLT3* did not show any reductions in total RNA level (Figure 4.12A). The disparity in the level of reduction between these sensitive gene targets in DOT1L knockdown and EPZ-5676 treatment may be due to the difference in nascent RNA vs total RNA changes. Moreover, a substantial level of H3K79me was still detectable by ChIP qPCR at *BCL2* following DOT1L knockdown, compared to the near-complete loss following inhibition by EPZ-5676. Therefore, the residual amount of H3K79me may be enough for transcription to occur to a level close to normal. Nevertheless, a consistent change in expression of the sensitive gene targets is observed between EPZ-5676 treatment and DOT1L knockdown, indicating these targets are most likely dependent upon DOT1L, rather than being affected by off-target effects of the EPZ-5676.

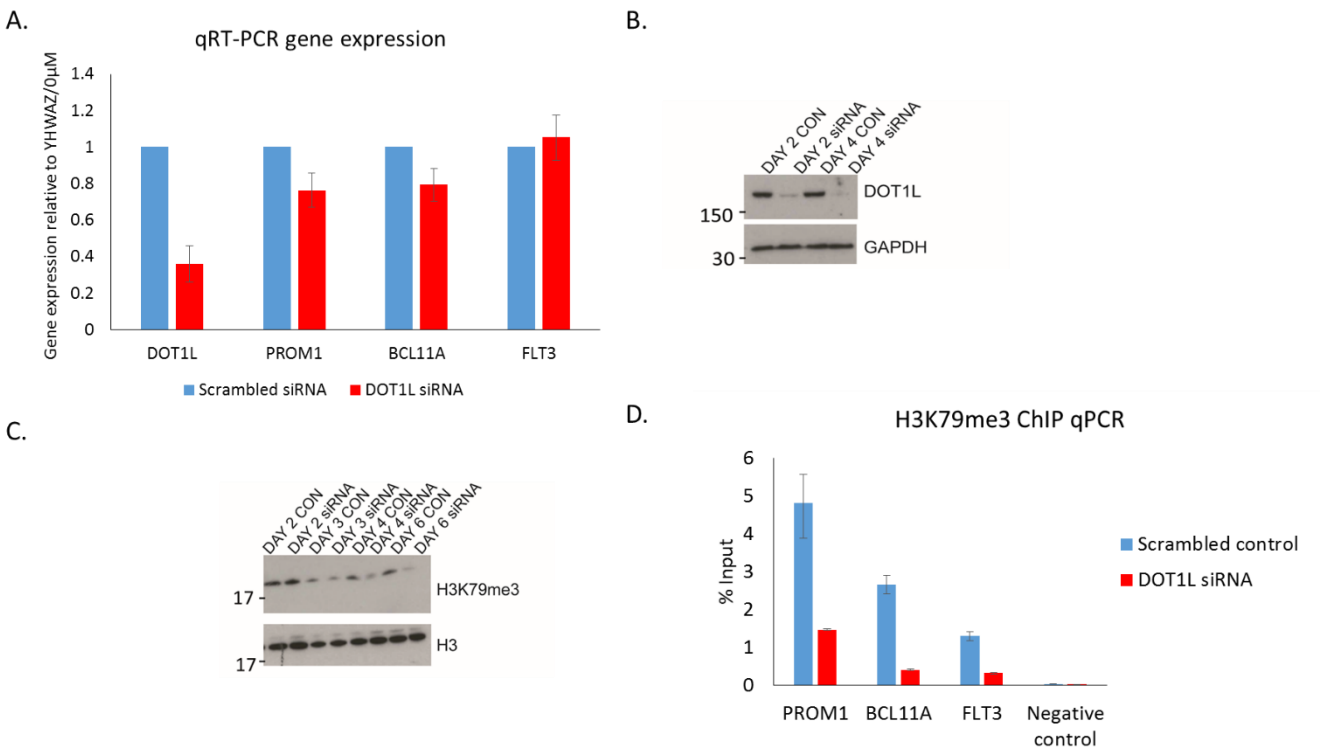

Figure 4-12. DOT1L knockdown validates hypersensitive genes (A) Following DOT1L mediated siRNA knockdown in SEM cells, qRT-PCR shows a reduction in DOT1L expression and hypersensitive genes relative to internal housekeeping gene *YHWAZ* relative to scrambled siRNA control. No reduction was observed in the insensitive gene *FLT3*. Error bars represent standard deviation of two biological replicates (B and C) Western blot analysis of WCEs and histone extraction of SEM cells treated with various timepoints of DOT1L siRNA knockdown compared to scrambled siRNA control. Western blot membrane probed using anti-DOT1L and anti-H3K79me3 antibody with anti-GAPDH and anti-H3 as loading controls (D) H3K79me3 ChIP qPCR at *PROM1*, *BCL11A*, *FLT3* and negative control region in SEM cells treated with DOT1L siRNA compared to scrambled siRNA control.

4.3.14 Verification of MLL-AF4 and H3K79 methylation targets in other MLL-AF4 cell lines

One important consideration when studying gene regulation in cell lines is that over time mutations in other genes can accumulate. This may distort gene regulation, meaning they may not be representative of what is occurring inside the leukaemic cells of a patient with MLL-AF4. In addition to this, even if the cell line does not accumulate additional mutations, the cell line was established from one patient and may not be typical of cells from other MLL-AF4 patients. To investigate whether the identified hypersensitive genes were specific for SEM cells, two other MLL-AF4 cell lines; MV4;11 (MLL-AF4 AML) and RS4;11 (MLL-AF4 ALL) were treated for 7 days with EPZ-5676. RS4;11 cells demonstrated a loss of H3K79me3 and a reduction in gene expression of *BCL2, PROM1* and *BCL11A* (Figure 4.13C-D). However, *CDK6* and *FLT3* were insensitive (Figure 4.13C-D). Furthermore, *PROM1* is not expressed in myeloid cells and therefore was not detectable in MV4;11 cells, however, *CDK6* and *BCL2* were downregulated and demonstrated loss of H3K79me3 by ChIP qPCR (Figure 4.13A-B). Interestingly, *FLT3* was downregulated in MV4;11 cells, in contrast to RS4;11 and SEM cells. Taken together these data demonstrate, reassuringly, that there are similarities in DOT1L dependent gene targets between MLL-AF4 cell lines, further validating some of the hypersensitive targets observed in this chapter. However, cell line specific and ALL vs AML-specific gene regulation is also occurring.

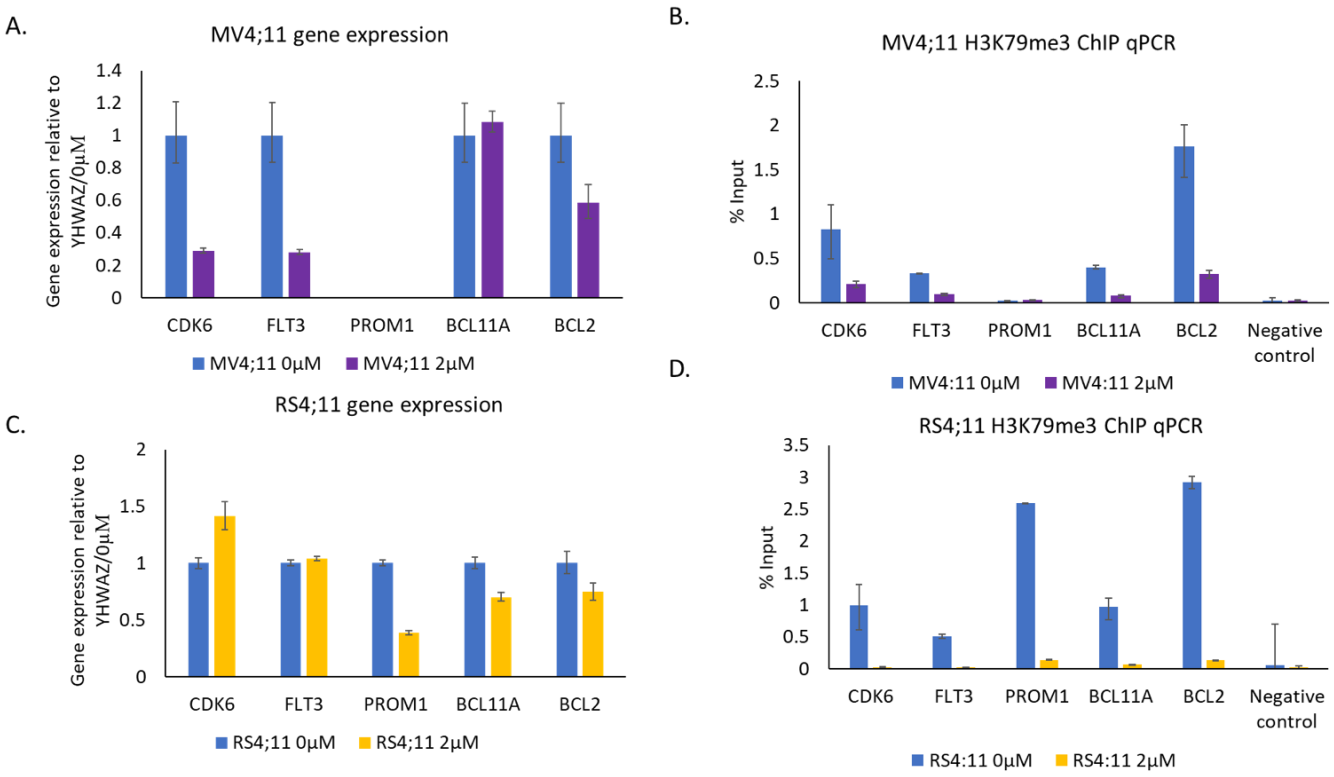

Figure 4-13. H3K79me3 hypersensitive targets are common to MLL-AF4 cell lines. (A) qRT-PCR in MV4;11 cells treated with 2µM EPZ-5676 and 0µM control for 7 days at sensitive and insensitive genes. (B) H3K79me3 ChIP qPCR performed in MV4;11 cells (ALL cell line) following 2µM EPZ-5676 (orange bars) and 0µM (Blue bars) for 7 days. (C) qRT-PCR performed in RS4;11 cells following 2µM EPZ-5676 and 0µM control. (D) H3K79me3 ChIP qPCR performed in RS4;11 cells (AML cell line) following 2µM EPZ-5676 and 0µM control for 7 days. Error bars in all figures represent standard deviation from two biological replicates.

4.3.15 H3K79me gene targets are found in primary MLL-AF4 ALL cells and normal fetal bone marrow (FBM) cells

Even though H3K79me genes targets have been identified in multiple MLL-AF4 cell lines, this does not address how similar these genes are compared to cells from patients with MLL-AF4 leukaemia. This is particularly important when considering the treatment of MLL-AF4 leukaemia and designing therapies that will not only work in cell lines, but importantly within patients. Interestingly, it has been suggested that MLL-AF4 infant ALL may arise *in utero,* potentially from the transformation of early foetal haematopoietic progenitor cells, found in the foetal bone marrow (FBM) (Barrett et al. 2016). MLL-AF4 leukaemia is dependent upon H3K79me (Krivtsov et al. 2008). Therefore, comparing H3K79me3 gene targets within FBM cells to MLL-AF4 leukaemia may provide some insight into the dynamics of H3K79me3 at genes pre- and post leukaemogenesis.

To highlight important similarities or differences between H3K79me3 targets in leukaemia vs normal cells, H3K79me3 ChIP sequencing was performed upon (1) MLL-AF4 primograft cells, originally derived from a diagnostic ALL patient sample (2) human FBM cells (see methods). Following sequencing H3K79me3 gene targets were identified and compared between SEM cells, the MLL-AF4 primograft sample and FBM (Figure 4.14A). Specific examples demonstrate that shared and different gene targets between FBM and MLL-AF4 leukaemia (patient sample and SEM) are present (Figure 4.14B). A total of 8870 marked H3K79me3 genes targets were observed in the MLL-AF4 patient sample, compared to 9193 found in SEM cells. Reassuringly, 81% of SEM H3K79me genes were observed in the MLL-AF4 patient cells, suggesting a strong similarity between the two. Even though this is compared to just one patient, the results indicate that H3K79me3 targets in SEM cells are generally

representative of MLL-AF4 ALL patient gene targets. 12,789 H3K79me3 gene targets were identified in FBM, around 40% more than observed in SEM cells or the patient sample. As FBM represents a population containing different cell types, this may explain the larger number of H3K79me3 genes observed. However, what is striking is that 93% of H3K79me3 found in SEM cells and the MLL-AF4 patient sample, are also found in FBM. As H3K79me marks normal gene targets which are not aberrantly controlled by MLL-AF4, it is likely that most of these targets are non MLL-AF4 genes. Nevertheless, it does indicate that gene targets which are regulated by MLL-AF4 in the leukaemia, may already be H3K79 methylated in the original cell.

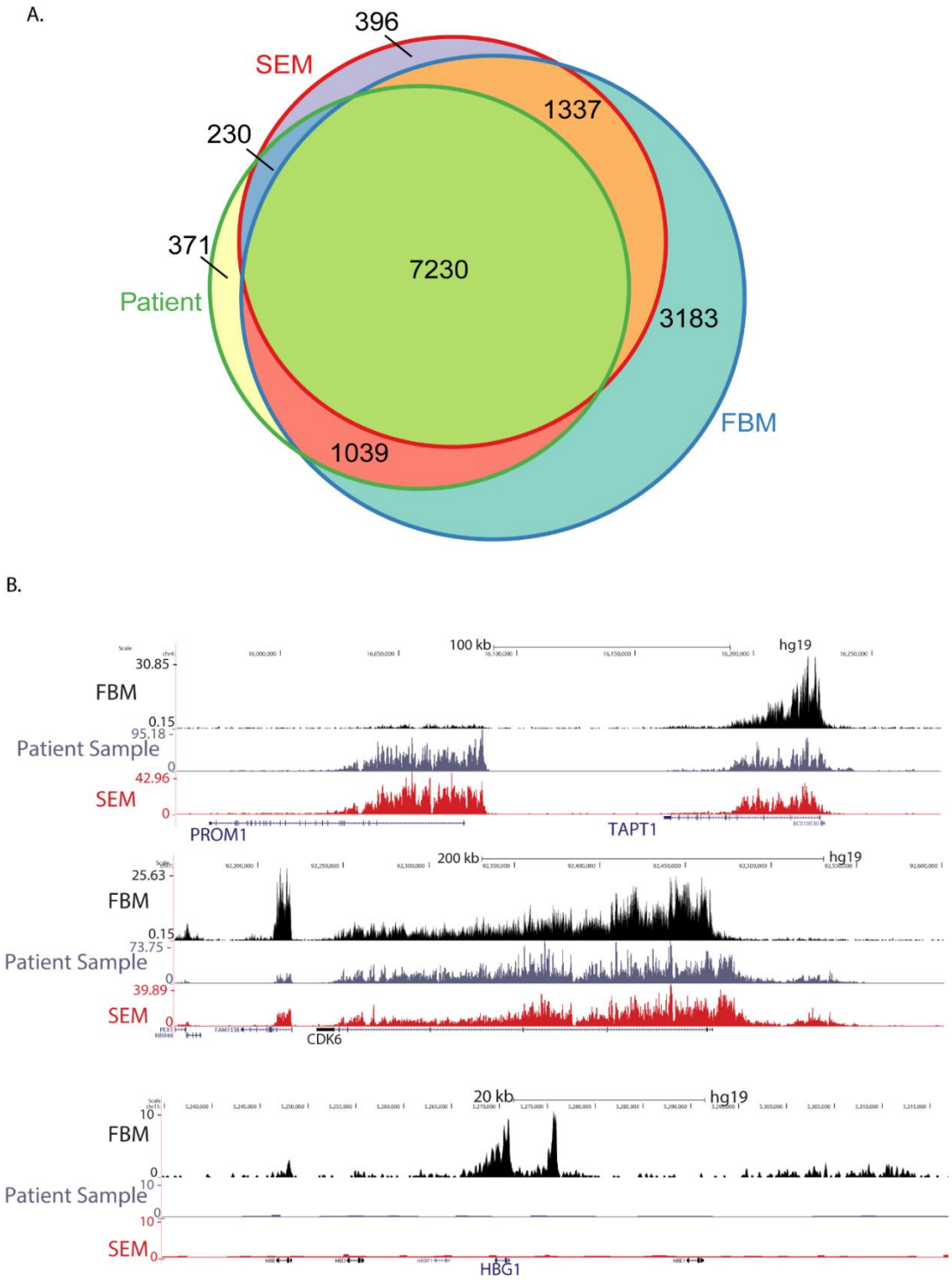

Figure 4-14. H3K79me3 found in an MLL-AF4 patient sample and FBM cells. (A) Venn diagram representing overlap of H3K79me3 gene targets in SEM cells, MLL-AF4 patient cells and FBM cells (B) UCSC ChIP seq tracks demonstrating shared and independent H3K79me3 gene targets in FBM, SEM and patient sample.

4.4 Discussion

4.4.1 DOT1L regulates a subset of gene targets

In this chapter it has been observed that H3K79me3 controls a subset of hypersensitive MLL-AF4 gene targets in MLL-AF4 leukaemia cell lines. There are two currently proposed mechanisms of DOT1L function at MLL-FP gene targets. One of these has described DOT1L as a SIRT1 antagonist and the other has proposed that H3K79me can modulate transcription factor binding and histone modification deposition (Chen et al. 2015; Gilan et al. 2016). As only a subset of H3K79me targets are downregulated following EPZ-5676 treatment, it is unclear whether these mechanisms can be used to describe the specificity of H3K79me function at hypersensitive MLL-AF4 gene targets without further analysis. This will be addressed in Chapter 5.

Conversely, many gene targets remained insensitive to EPZ-5676 treatment despite the loss of H3K79me3. This suggests that MLL-AF4 may regulate these insensitive gene targets via a DOT1L and H3K79me independent mechanism. It also suggests that H3K79me is not generically required for transcription per se. As described in chapter 1, MLL-FPs have been demonstrated to regulate gene targets via different mechanisms. Examples includes via P-TEFb, which is involved in early transcription elongation, or via the SL1 complex, involved in transcription initiation (Garcia-Cuellar et al. 2016; Okuda et al. 2015).

The validation of downregulation observed when using the inhibitor is very important. This is because due to the nature of H3K79me3 and the relatively long residency time, off target secondary effects are likely. To validate the hypersensitive gene targets observed following EPZ-5676 treatment, a DOT1L siRNA was employed. DOT1L siRNA-mediated knockdown confirmed that a reduction of H3K79me3 led to the downregulation of specific hypersensitive

gene targets tested. Although the effect upon expression of hypersensitive genes such as *PROM1* was subtle, this may be a consequence of remaining DOT1L and H3K79me still present at genes. Importantly, it is also plausible that other proteins, besides H3K79, are methylated by DOT1L. EPZ-5676 has been shown to specifically inhibit DOT1L methyltransferase activity by in vitro screening (Daigle et al. 2013) but it is possible it has other in vivo effects. Therefore, when using EPZ-5676 it may be more accurate to consider H3K79me as an output of DOT1L methyltransferase activity.

4.4.2 Comparison of H3K79me gene targets in different cell types

Following FBM H3K79me ChIP seq, a large proportion of H3K79me3 targets identified were also found in SEM and in MLL-AF4 ALL patient cells. It has been proposed that MLL-AF4 infant ALL may arise via a translocation occurring within a foetal hematopoietic progenitor cell, found within the FBM population (Montes et al. 2011; Barrett et al. 2016). Following transformation, it is possible that H3K79me3 gene targets are then aberrantly maintained via MLL-AF4. Furthermore, MLL-AF4 may increase H3K79me at these gene targets, consistent with the high levels of H3K79me3 observed at MLL-AF4 gene targets (Krivtsov et al. 2008). This could occur via the increased stability of DOT1L binding at these targets by mechanisms described in chapter 3.

Reassuringly, H3K79me3 gene targets observed in SEM display a large overlap with H3K79me3 gene targets in the MLL-AF4 ALL patient sample. Furthermore MLL-AF4 ALL cell lines, SEM and RS4;11, did show some similarities in H3K79me3 hypersensitive targets. Taken together, these data suggest that H3K79me3 hypersensitive genes observed in SEM cells are likely hypersensitive in MLL-AF4 ALL patients. Therefore, studying the function of

DOT1L and H3K79me in MLL-AF4 cell lines can be extrapolated to understanding what may be occurring at gene targets in patients with MLL-AF4 ALL.

Chapter 5 – Understanding the function of H3K79me at MLL-AF4 hypersensitive gene targets

5.1 Introduction

DOT1L and H3K79me play an important role in the transcription of many genes, including some which are directly bound by MLL-AF4 and may be important for the leukaemia (Identified in Chapter 4). Following a global reduction in H3K79me, only a subset of H3K79me-marked gene targets demonstrates a reduction in transcription. This suggests that only a subset of H3K79me marked gene targets are dependent upon H3K79me for transcription, thus H3K79me may play a specific regulatory role at these gene targets.

Very little is known about exactly how H3K79me can contribute to transcription and much is speculated, in part due to the difficulty in studying H3K79me. Due to the location of H3K79me within the nucleosome core, *in vitro* reconstituted H3K79me modified nucleosomes are difficult to generate, precluding extensive biophysical and biochemical analyses. However, experiments have provided some evidence of how H3K79me can mechanistically contribute to transcription.

As noted in the introduction, histone modifications have been demonstrated to act as platforms which chromatin proteins can bind and have been shown to biophysically alter nucleosome structure. Two reader proteins, 53BP1 and SMN, have been identified for H3K79me *in vitro.* Both proteins have been shown to interact with H3K79me via a tudor domain (Wysocki et al. 2005; Sabra et al. 2013). Although neither 53BP1 or SMN are associated with transcription,

the existence of H3K79me-binding proteins suggests that this mark may be able to interact with a similar transcription-associated protein (Huyen et al. 2004). In addition to this, H3K79me has been shown to cause a conformational change on the surface of the nucleosome (Lu et al. 2008). Although it is inconclusive whether this might directly affect nucleosome structure, this could potentially affect transcription by modulating the accessibility of chromatin and directly controlling transcription factor binding.

5.1.2 Current models of DOT1L function

Two models have recently been proposed to explain the function of DOT1L in transcription. Whilst there are differences between these models, they mainly focus upon the interplay between H3K79 methylation and other histone modifying enzymes and the modifications associated with them. As described in chapter 1 in detail, one of these models suggests DOT1L may act as an antagonist to the H3K9 deacetylase, SIRT1, in MLL-AF9 leukaemia cell lines and is consistent with studies carried out in yeast demonstrating that DOT1L can antagonize Sir proteins (Takahashi et al. 2011; Chen et al. 2015). Following the loss of H3K79me, a reduction in H3K9ac and subsequent increase in H3K9me was observed, dependent upon the H3K9 methyltransferase, SUV39H1. This loss of activating histone modifications and gain of repressive histone modifications associated with facultative heterochromatin was proposed as a functional mechanism which may lead to the downregulation of H3K79me marked gene targets.

In addition to this, it has been shown that DOT1L can act in co-operation with BRD4 in MLL-FP cell lines (Gilan et al. 2016). This model suggests that H3K79me facilitates the binding of transcription factor CREB1. CREB1 has been shown to recruit the acetyltransferase p300, one of the substrates of which is H4. An important reader of acetylated lysine residues, including

H4K5ac, is BRD4. Following loss of H3K79me, H4K5ac and BRD4 binding are reduced at H3K79me gene targets, in addition to a reduction in CREB1 and P300 binding. Therefore, this model suggests that H3K79me facilitates transcription factor binding leading to the recruitment of P300 and thus BRD4.

Taken together, these models suggest that following a loss of H3K79me, a reduction in histone acetylation and an increase repressive H3K9me is observed, potentially due to changes in transcription factor binding such as CREB1, and antagonism of repressive modifications directly by DOT1L and H3K79me. Loss of acetylation leads to the loss of BRD4 at these loci and may lead to a loss of other proteins such as AF9 and ENL, which also bind to acetylated lysine residues (Li et al. 2014; Wan et al. 2017; Erb et al. 2017). Disruption of interactions with multiple elongation components may therefore lead to a decrease in transcription.

Data from this thesis and others have demonstrated that the requirement for H3K79me is not universal and only a subset of H3K79me gene targets are downregulated following DOT1L inhibition. This poses the question of what mediates this specificity. This question will be addressed in this chapter by comparing hypersensitive H3K79me marked MLL-AF4 gene targets vs insensitive H3K79me marked MLL-AF4 gene targets.

5.2 Aims

1. Determine the effects upon other histone modifications following EPZ-5676 treatment.
2. Investigate changes in chromatin accessibility following EPZ-5676 treatment.
3. Understand whether chromatin structure is affected by loss of H3K79me.

5.2 Results

5.2.1 Loss of H3K79me leads to changes in histone modifications

According to the models described one of the ways transcription is affected is via changes in histone modifications. We therefore wanted to test as a first pass whether changes in histone modifications could dictate the specificity of the hypersensitive MLL-AF4 gene targets. If this was the case, we would expect that changes in histone modifications would only occur at the hypersensitive gene targets and not at the insensitive gene targets.

Following treatment of SEM cells with EPZ-5676 for 7 days, global loss of H3K79me was detected by western blot analysis as expected (Figure 5.1A). No changes in the activating histone modifications H3K27ac and H3K9ac, or repressive modifications H3K9me2 and H3K9me3, were detected at the bulk histone level (Figure 5.1A). To test whether changes were observed on a gene-by-gene basis, ChIP qPCR was employed. Both H3K27ac and H3K9ac showed a reduction at hypersensitive gene targets, including *PROM1* and *BCL11A* (Figure 5.1B-C). In addition to this, an increase in H3K9me2 and H3K9me3 was observed at these loci (Figure 5.1D-E). However, these changes were not limited to hypersensitive gene targets, as similar effects were observed at insensitive gene targets, including *FLT3* and *ADAM10* (Figure 5.1B-E). Furthermore, an increase in H3K9me2 and H3K9me3 was also observed at the silent gene target *HOXC8* (Figure 5.1D-E).

As an important control, it is worth establishing whether EPZ-5676 treatment affects the expression of any histone modifying enzymes, which could potentially lead to changes in histone modifications at gene targets. A table including the Log Fold Change (LogFC) of nascent RNA of histone modifying enzymes associated with the modifications examined in this thesis was compiled (Table 5.1). Even though there are observable changes in 4 enzymes

at the level of nascent RNA, most histone modifying enzymes do not show a change in gene expression upon EZP5676 treatment. Together these data suggest that both hypersensitive and insensitive gene targets display seemingly non-specific changes in histone modification upon loss of H3K79me. Although this analysis is not genome wide, decreases in H3K27ac, H3K9ac and increases in H3K9me2/3 seem to occur generally, therefore cannot alone explain the specificity of the observed downregulation of transcription.

Table 5-1. Changes in nascent RNA levels of histone modifying enzymes.

Protein	Substrate residues	LogFC	FDR
Lysine acetyltransferases			
GCN5	K9	-0.4	0.01
P300/KAT3	K27/K18/H4K5	0.07	0.44
CBP	K27/K18/H4K5	-0.14	0.15
MOZ/KAT6	K9/K14	-0.02	0.8
MORF	K9/K14	0.0039	0.9
PCAF/KAT2B	K9	0.73	0.003
Lysine deacetylases			
HDAC1		-0.32	0.001
HDAC2		-0.03	0.79
HDAC3		-0.07	0.6
HDAC4		-0.17	0.15
SIRT1	K9	0.14	0.19
SIRT2		0.41	0.19
SIRT3		0.39	0.01
Methyltransferases			
SUV39H1/KMT	H3K9	-0.43	0.7
G9A	H3K9	-0.3	0.02
SETDB1	H3K9	0.19	0.06
Demethylases			
KDM3A	H3K9	0.09	0.3
KDM3B	H3K9	0.11	0.3
KDM4A	H3K9	0.08	0.4
KDM4B		0.12	0.4
KDM4C		0.11	0.3
KDM4D		-0.23	0.3

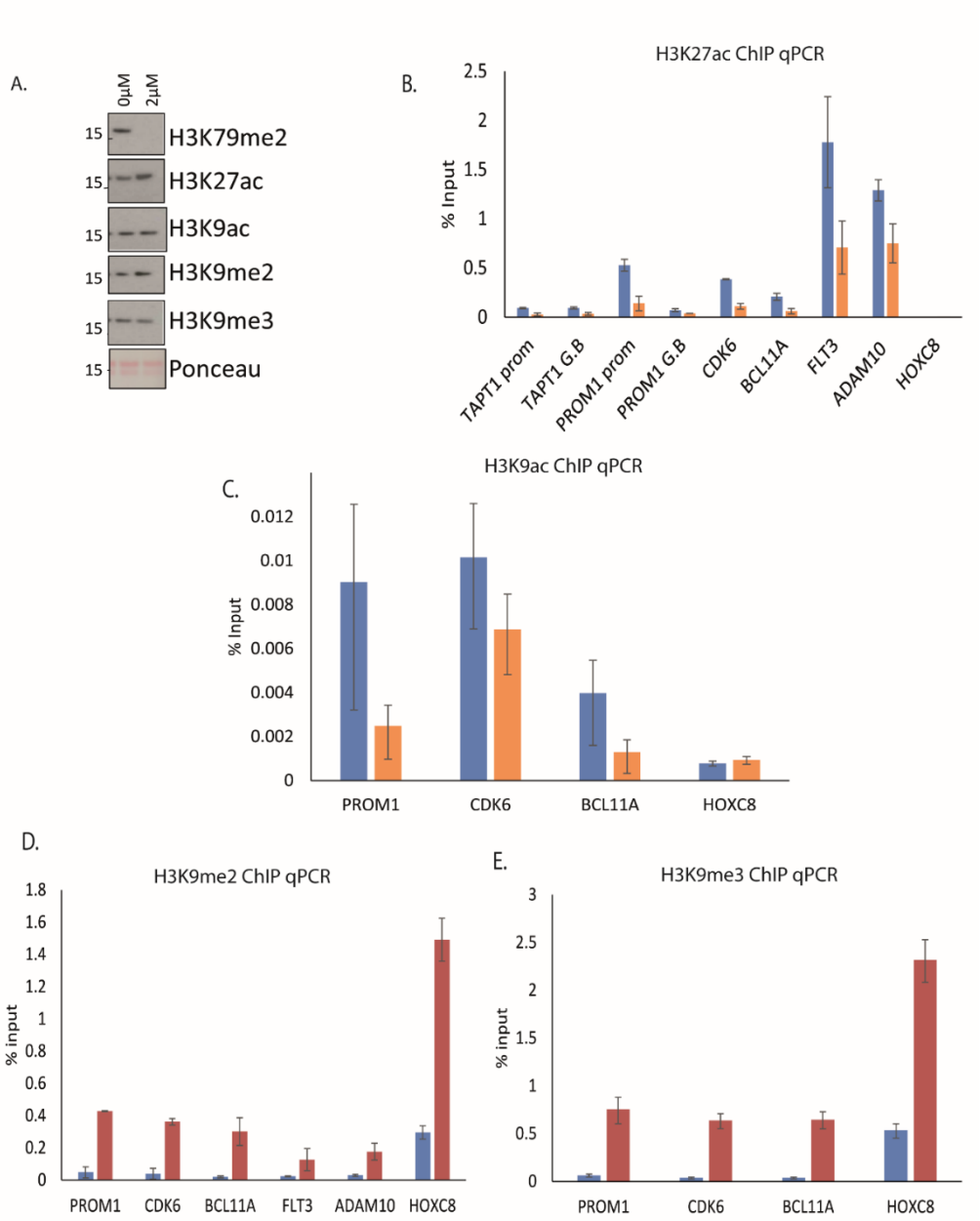

Figure 5-1. Loss of H3K79me leads to a change in histone modifications (A) Western blot analysis on purified histones from SEM cells treated with 2µM EPZ-5676 for 7 days compared to 0µM control. (B-C) H3K27ac and H3K9ac ChIP qPCR following 7 days 0µM and 2µM EPZ-5676 treatment of SEM cells at sensitive (*PROM1* (prom = promoter G.B= gene body), *TAPT1*, *BCL11A*, *CDK6*) and insensitive (*FLT3*, *ADAM10*) genes plus negative control (*HOXC8*). Error bars represent standard deviation from two biological replicates (D-E) H3K9me2 and H3K9me3 ChIP qPCR following 7 days 0µM and 2µM EPZ-5676 treatment of SEM cells at sensitive (*PROM1*, *TAPT1*, *BCL11A*, *CDK6*) and insensitive (*FLT3*, *ADAM10*) genes plus negative control (*HOXC8*). Error bars represent standard deviation from two biological replicates.

5.2.2 Loss of H3K79me leads to subtle changes in chromatin accessibility

So far, these results have indicated that changes in histone modifications due to loss of H3K79me appear to be general and non-specific. Therefore, this cannot explain specific transcription changes. To explain specific transcription changes, more general effects upon histone modifications must in some way translate into a specific effect at hypersensitive gene targets. What could provide specificity at gene targets? One possibility is specific transcription factor binding. Therefore, there are two possibilities. Firstly, H3K79me may promote chromatin accessibility and promote binding of transcription factors, and following loss of H3K79me transcription factor binding affinity is reduced, this is supported by published data (Gilan et al. 2016). Secondly, the general increase in repressive histone modifications across the genome, upon H3K79me reduction, may have a repressive effect upon transcription factor binding to particular regions, for example; at potential enhancers of hypersensitive genes. Both mechanisms could work synergistically and in both cases a loss of transcription factor binding would be expected following loss of H3K79me. To test this, ATAC seq was employed following EPZ-5676 treatment of SEM cells.

5.2.3 Measuring chromatin accessibility using ATAC seq

Assay for Transposase Accessible Chromatin (ATAC) seq is a method used to map chromatin accessibility genome wide (Buenrostro *et al*., 2015). Briefly, this technique relies upon the Tn5 transposase to insert sequencing adapters into accessible regions of chromatin. Fragments generated can be sequenced, mapped and visualized on a genome browser, such as UCSC, to observe fragments which were accessible to the transposase at a population level. Therefore, visualised peaks represent DNA which is not bound by DNA binding proteins such as transcription factors or histones. Peaks are often observed directly around a transcription factor

binding site, due to a lack of nucleosomes, and therefore even though the peak is truly displaying regions of chromatin accessibility it can be used to demonstrate where transcription factor binding is occurring.

If H3K79me is modulating transcription factor binding and/or antagonising repressive chromatin proteins, which has been proposed in published data (Chen et al. 2015; Gilan et al. 2016), then it would be expected that upon loss of H3K79me, chromatin accessibility would change.

5.2.4 Hypersensitive and Insensitive gene targets display changes in chromatin accessibility

ATAC seq revealed subtle but clear reductions in chromatin accessibility at some hypersensitive gene targets including *PROM1*, *TAPT1*, *BCL11A*, and *ARID1B* following 7 days of 2μM EPZ-5676 treatment (Figure 5.2 and 5.3). ATAC peaks were either lost or demonstrated a reduced intensity within the 5' end of the gene body at these hypersensitive targets. The reductions in ATAC peaks overlapped with H3K27ac and H3K4me1 in the gene body. These modifications are associated with active enhancers and therefore may indicate these genes contain putative intragenic enhancers (Heintzman et al. 2009; Creyghton et al. 2010). In addition to this, ATAC, H3K27ac and H3K4me1 also co-localise with MLL-AF4 binding and H3K79me (Figures 5.8C and 5.9C). Some hypersensitive gene targets, such as *MYC* and *BCL2*, did not demonstrate prominent reductions in chromatin accessibility and ATAC peaks did not overlap with both H3K27ac and H3K4me1 in the 5' end of the gene body (Figure 5.4). Insensitive gene targets such as *FLT3* and *LMO4* did not demonstrate any obvious changes in chromatin accessibility (Figure 5.5). In addition to this, no clear changes in chromatin accessibility were observed at the promoters of any gene targets observed suggesting

this change in chromatin accessibility within the gene body may be specifically occurring at intragenic enhancer regions.

Although changes seemed, by eye, more prominent at a subset of hypersensitive targets which contained putative intragenic enhancers (as defined by histone modification and ATAC), there are small increases and decreases in ATAC signal at all gene targets, both hypersensitive and insensitive. This could be due to sensitivity and technical variation of the ATAC technique. The significance and quantification of the changes in chromatin accessibility at hypersensitive and insensitive gene targets has not been addressed in this thesis, which would require more biological replicates. Taking this into account, it can be suggested from this data that chromatin accessibility is reduced at some hypersensitive targets which may be indicative of reductions in transcription factor binding.

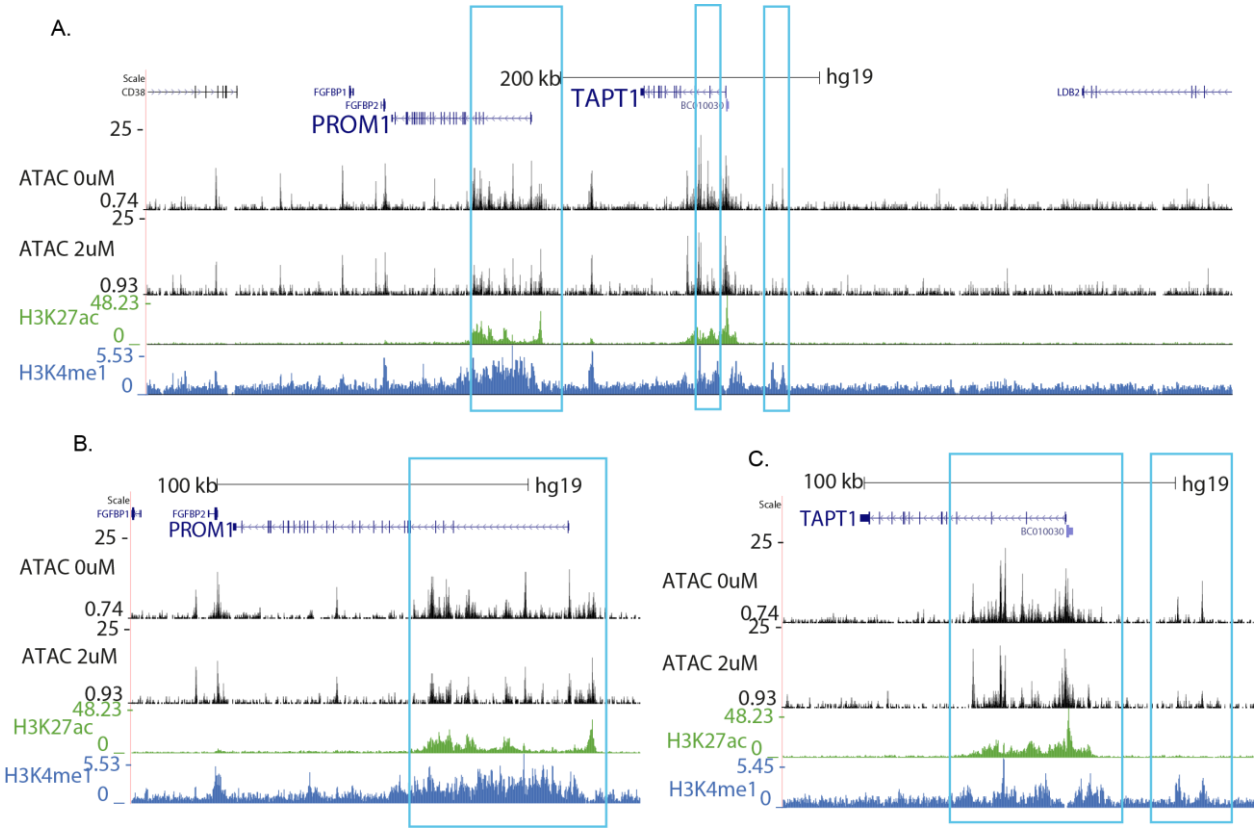

Figure 5-2. Reductions in chromatin accessibility observed at *PROM1* and *TAPT1* following EPZ-5676 treatment. (A) ATAC seq following 7 days 0μM and 2μM EPZ-5676 treatment of SEM cells. Blue windows represent reductions observed in chromatin accessibility corresponding with H3K27ac and H3K4me1 ChIP seq. Representative of 2 biological replicates (B) Magnified version of (A) representing ATAC and ChIP seq at PROM1 (C) Magnified version of (A) representing ATAC and ChIP at *TAPT1*.

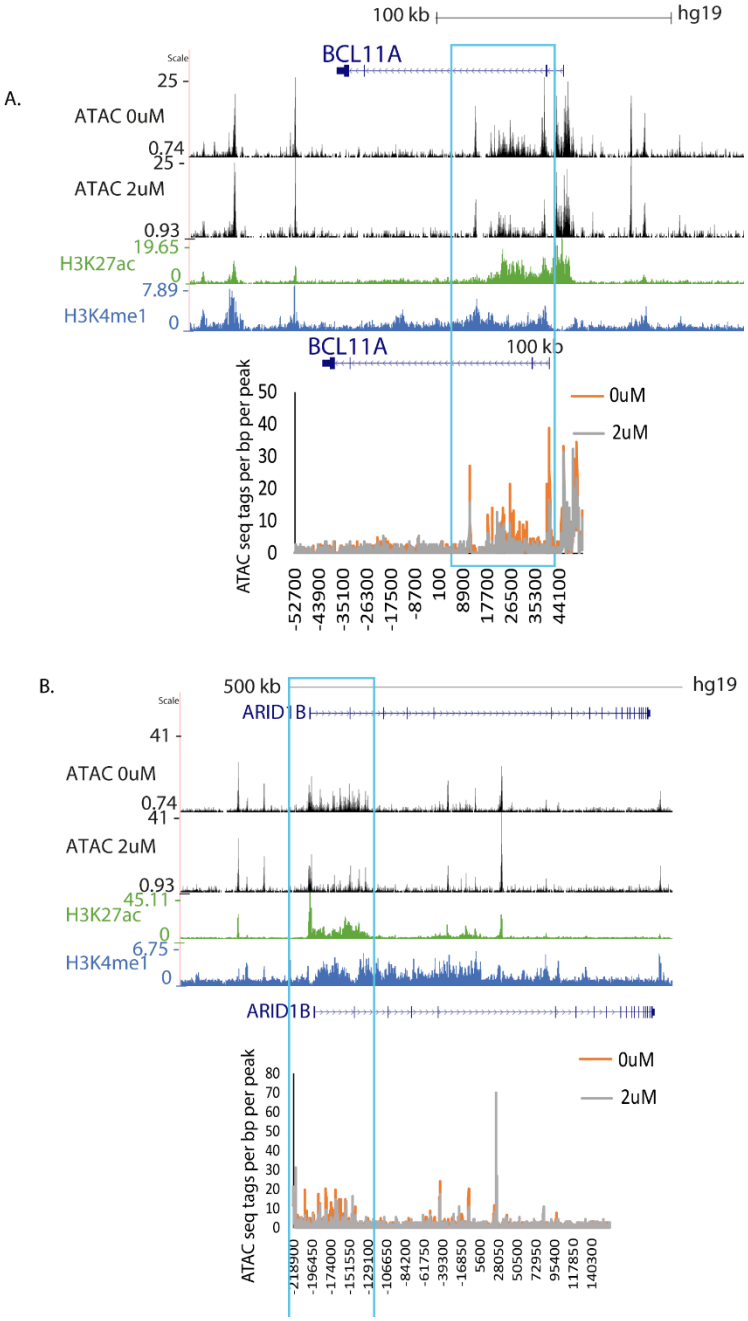

Figure 5-3. Reductions in chromatin accessibility at hypersensitive gene targets following EPZ-5676 treatment. (A) ATAC seq following 7 days of 0µM and 2µM EPZ-5676 treatment of SEM cells at *BCL11A*. Blue windows represent reductions observed in chromatin accessibility corresponding with H3K27ac and H3K4me1 ChIP seq. Histogram represents zoomed in view of ATAC seq signal at *BCL11A* in 0µM (Orange) and 2 µM (Blue) Representative of two biological replicates (B) ATAC seq following 7 days of 0µM and 2µM EPZ-5676 treatment of SEM cells at *ARID1B*. Blue windows represent reductions observed in chromatin accessibility corresponding with H3K27ac and H3K4me1 ChIP seq. Histogram represents zoomed in view of ATAC seq signal at *ARID1B* in 0µM (Orange) and 2 µM (Blue). Representative of two biological replicates.

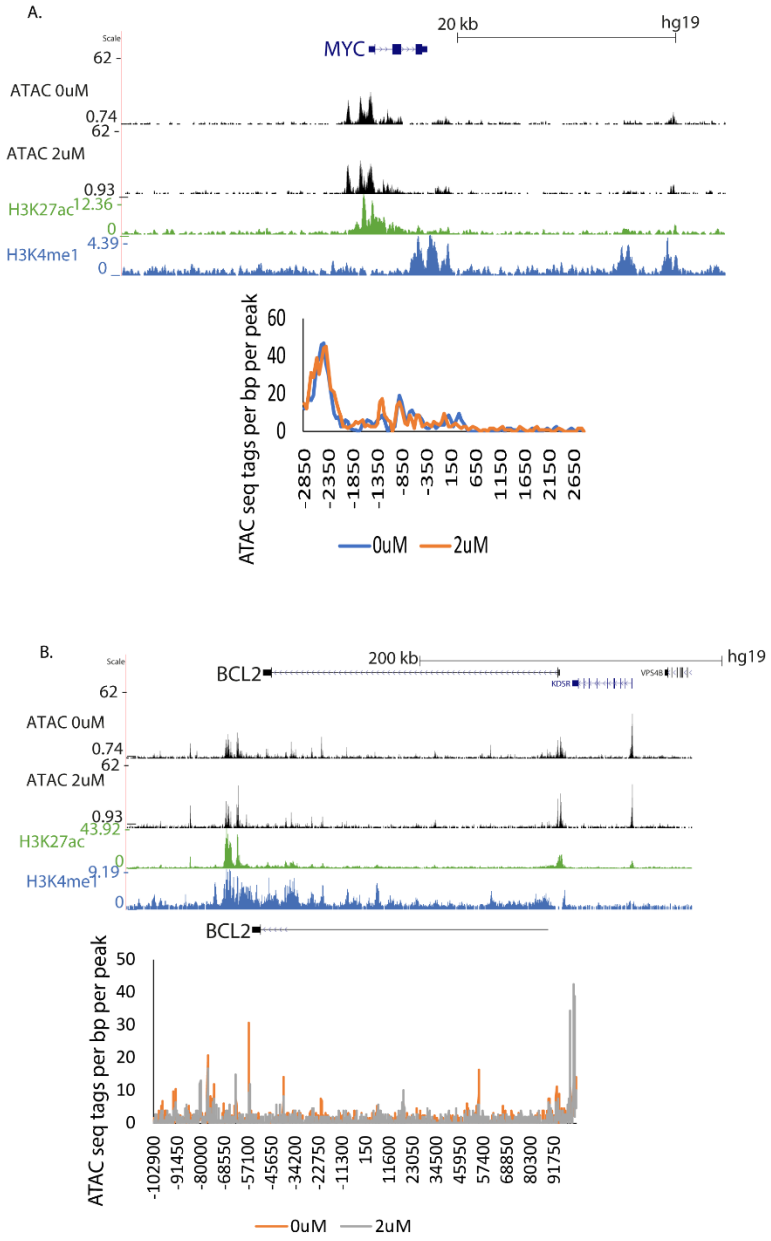

Figure 5-4. No prominent changes in chromatin accessibility observed at some hypersensitive gene targets following EPZ-5676 treatment. (A) ATAC seq following 7 days of 0µM and 2µM EPZ-5676 treatment of SEM cells at *MYC*. ATAC seq peaks correspond to H3K27ac and H3K4me1 ChIP seq. Histogram showing zoomed in ATAC seq peaks for *MYC* in the 0µM (Orange) and 2µM (Blue) EPZ5676 treatment (B) ATAC seq following 7 days of 0µM and 2µM EPZ-5676 treatment at *BCL2*. ATAC seq peaks correspond to H3K27ac and H3K4me1 ChIP seq. Histogram showing zoomed in ATAC seq peaks for *BCL2* in the 0µM (Orange) and 2µM (Blue) EPZ5676 treatment. ATAC seq representative of two biological replicates.

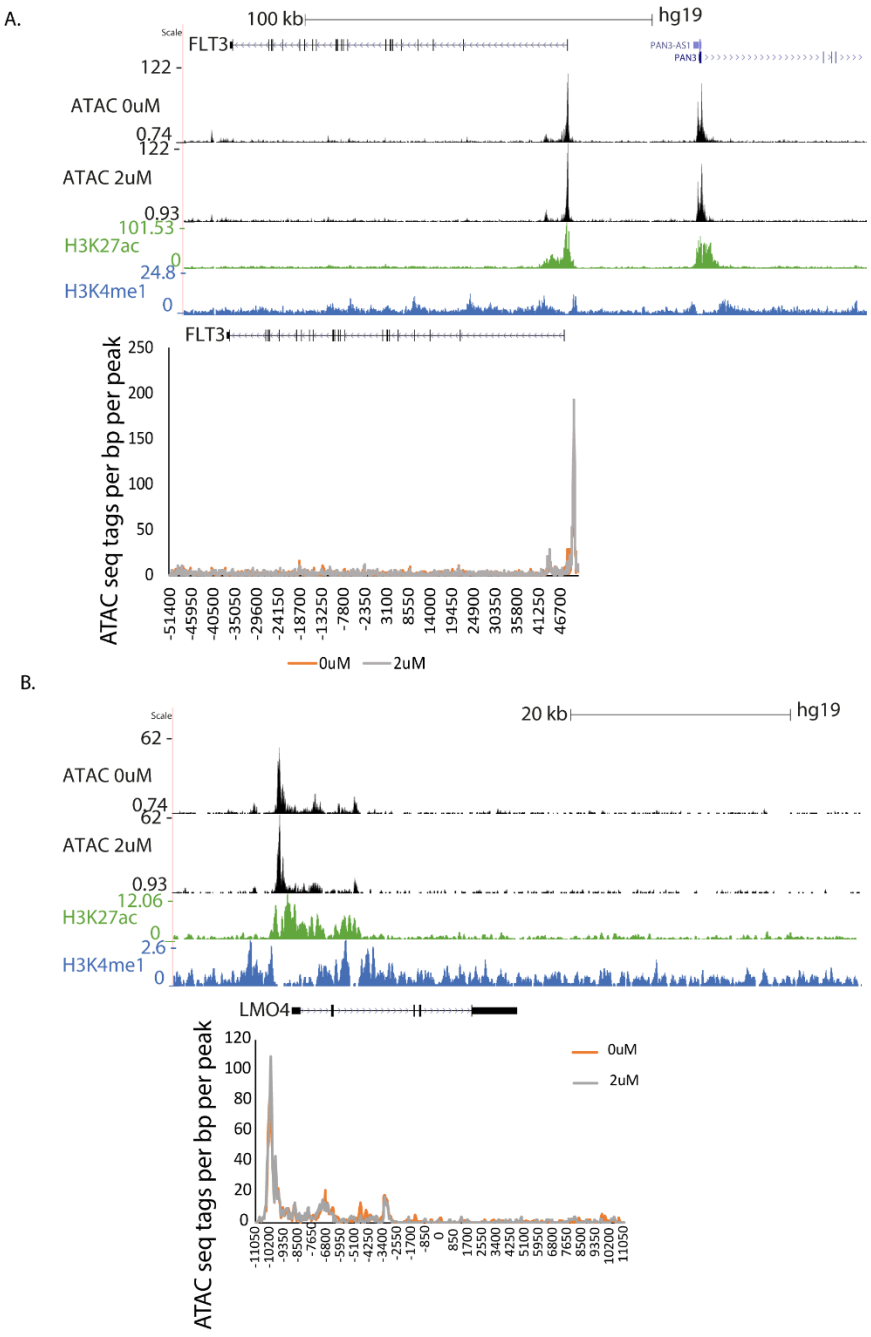

Figure 5-5. No prominent changes in chromatin accessibility observed at insensitive gene targets following EPZ-5676 treatment. (A) ATAC seq following 7 days of 0μM and 2μM EPZ-5676 treatment of SEM cells at *FLT3*. ATAC seq peaks correspond to H3K27ac and H3K4me1 ChIP seq. Histogram showing zoomed in ATAC seq peaks for *FLT3* in the 0μM (Orange) and 2μM (Blue) EPZ5676 treatment (B) ATAC seq following 7 days of 0μM and 2μM EPZ-5676 treatment of SEM cells at *LMO4*. ATAC seq peaks correspond to H3K27ac and H3K4me1 ChIP seq. Histogram showing zoomed in ATAC seq peaks for *LMO4* in the 0μM (Orange) and 2μM (Blue) EPZ5676 treatment. ATAC seq representative of two biological replicates.

5.2.5 Differential transcription factor binding may determine reductions in enhancer-promoter interactions

ATAC seq analysis indicates that loss of H3K79me at sensitive gene targets could affect transcription in part by altering transcription factor binding at these targets, but it does not indicate which specific transcription factors are involved. One way of identifying which transcription factors could potentially regulate a specific gene target is by analyzing the transcription factor motifs in the underlying DNA sequence of a gene target. To assess whether hypersensitive gene targets are enriched for specific transcription factor motifs a motif analysis was carried out on gene targets downregulated following EPZ-5676 treatment. This analysis includes all genes downregulated following 2μM EPZ-5676 treatment (including those which are not MLL-AF4 targets), as by using a larger set of gene targets sensitive to EPZ-5676 we may observe differences in transcription factor binding motifs more clearly. This analysis was also used to test whether, CREB1, a transcription factor whose binding has been shown to be modulated by H3K79me in MLL-r leukaemia cell lines (Gilan et al. 2016), was enriched at sensitive targets. It would perhaps be expected that if CREB1 was modulated by H3K79me, it would be enriched at genes targets dependent upon H3K79me.

Several transcription factor binding motifs were found to be enriched at sensitive genes relative to all genes, and many of these transcription factors are known to be expressed in SEM cells. Interestingly, CREB1, was not enriched in this set of gene targets. Whilst this does not rule out a role for CREB1 at a subset of these gene targets, it does suggest that enriched transcription factors identified may be important for the transcription of a larger set of sensitive gene targets. Many of the motifs enriched included ETS family transcription factors, such as ETS1, FLI1 and ERG, in addition to MYB, a transcription factor shown to be necessary for leukaemogenesis (Zuber et al. 2011) (Figure 5.6A). FPKM values were taken from Nascent

RNA seq to establish whether the enriched transcription factors were expressed in SEM cells (Figure 5.7B). From this, MYC, MYB and ETS transcription factors are both enriched and more highly expressed in SEM cells compared to other transcription factors (Figure 5.6B). Notably, transcription of the vast majority of these transcription factors is not affected by EPZ-5676 treatment, arguing that these gene targets are not rendered hypersensitive merely by changes in expression of their regulators. This has not been investigated further in this thesis but could be addressed in future work.

A.

Transcription factor motif	p-value
ETS	1.00E-19
Elk1	1.00E-17
YY1	1.00E-17
Elk4	1.00E-15
ELF1	1.00E-14
E2F4	1.00E-13
GABPA	1.00E-10
c-Myc	1.00E-10
NRF	1.00E-07
CLOCK	1.00E-07
Sp1	1.00E-07
Max	1.00E-07
USF1	1.00E-06
ETV1	1.00E-06
n-Myc	1.00E-05
ETS1	1.00E-05
ZNF711	1.00E-05
MITF	1.00E-05
Fli1	1.00E-04

Transcription factor motif	p-Value
NFY	1.00E-04
ZFX	1.00E-04
GFX	1.00E-04
CRX	1.00E-03
Nr5a2	1.00E-03
Usf2	1.00E-03
E-box	1.00E-03
E2F7	1.00E-03
GSC	1.00E-03
NRF1	1.00E-03
BMYB	1.00E-03
bHLHE40	1.00E-03
HIF-1b	1.00E-03
ELF5	1.00E-02
Esrrb	1.00E-02
HOXA2	1.00E-02
ZNF143	1.00E-02
KLF5	1.00E-02
OTX2	1.00E-02
E2F1	1.00E-02
ERG	1.00E-02
MYB	1.00E-02

B.

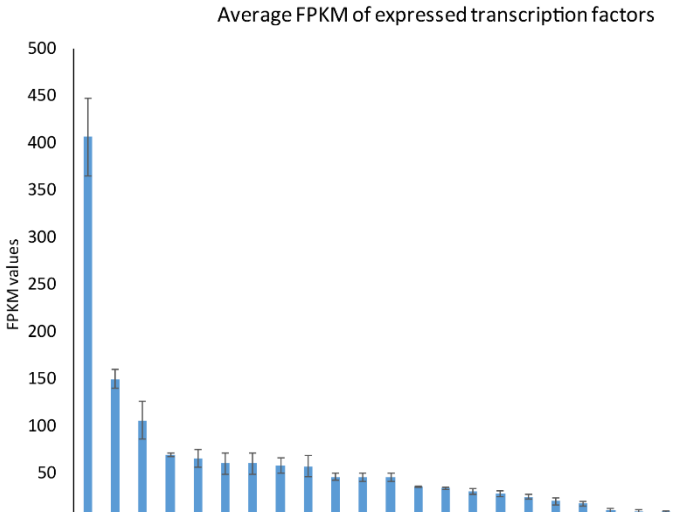

Figure 5-6. Transcription factor motifs enriched in downregulated genes following EPZ-5676 treatment of SEM cells (A) P-values in order of significance of enriched transcription factor motifs in downregulated gene targets following 7 days 2μM EPZ-5676 treatment compared to all active genes. (B) Average Fragment Per Kilobase per Million (FPKM) values of transcription factors whose motifs were enriched in (A). Ordered to display the transcription factors most high expressed in SEM cells. Error bars represent standard deviation of three biological replicates.

5.2.6 Loss of H3K79me leads to a reduction of putative enhancer-promoter interactions

As discussed earlier, there are two main possibilities of how H3K79me might function at the putative intragenic enhancers to modulate transcription factor binding. Firstly, it is possible that H3K79me promotes chromatin accessibility for transcription factor binding to occur. Secondly, H3K79me may antagonize repressive modifications. Repressive modifications have been demonstrated to create a local domain of inaccessible chromatin, in particular H3K9me has been associated with accumulation of HP1 and the phase separation of heterochromatin (Strom et al. 2017; Larson et al. 2017). This in turn may modulate chromatin accessibility and may impact transcription factor binding (Fitzgerald & Bender 2001).

Both proposed mechanisms are not mutually exclusive and both may lead to perturbation of transcription factor binding following loss of H3K79me. It is possible that disruption of transcription factor binding at the identified potential intragenic enhancers could account for the downregulation in transcription observed at hypersensitive targets. Enhancers have been shown to regulate transcription by physical interaction with the receptive promoter, and transcription factor binding at enhancers and promoters has been demonstrated to be important for this interaction, with the assistance of other co-factors (Nolis et al. 2009; Vakoc et al. 2005; Jing et al. 2008). Enhancer-promoter interactions contribute to the 3D structure and organisation of chromatin in the nucleus, known as TADs and smaller sub-TADs (Dixon et al. 2012). Therefore, if transcription factor binding at intragenic enhancer regions is reduced following loss of H3K79me, the structure of the TAD or sub-TAD may be perturbed.

If the putative embedded enhancer elements identified in hypersensitive gene targets are dependent upon H3K79me for transcription factor binding and/or antagonism of repressive modifications, then it may be expected that following loss of H3K79me, the enhancer-promoter

interaction occurring within a domain may become less stable and therefore a decrease in frequency of this interaction may be observed. Furthermore, if this mechanism is specific only to hypersensitive gene targets which contain embedded enhancer elements, then no changes would be expected at insensitive gene targets.

The structure of TADs has previously been studied at the globin locus using 3C techniques, including the high resolution Capture-C technique, coupled with ChIP seq analysis (Vernimmen et al. 2009; Kowalczyk et al. 2012; Hanssen et al. 2017). Therefore, Capture-C was employed here to determine the effect of loss of H3K79me, following EPZ-5676 treatment, on the overall TAD structure and the interactions occurring within the domain.

5.2.7 Using Capture-C to study domain structure

In brief, Capture-C relies upon the principles of digestion and re-ligation of DNA fragments which are physically close to each other, and uses an oligonucleotide probe to enrich for interactions with a specific capture point, typically a promoter or enhancer, followed by high-throughput sequencing to identify interaction sites. Data gathered from the globin locus has shown that two main pieces of information can be uncovered using Capture-C (Hughes et al. 2013). Firstly, it describes the extent of a domain, indicating where the boundaries are. Secondly, it highlights the frequency of interactions which occur with the capture point, within the defined domain. As Capture-C is performed on a population of cells, it describes the relative frequency of the interactions occurring with the capture point within a population of cells. Higher frequency interactions, indicated by relatively higher peak signal, are potentially indicative of more stable interactions, for example between enhancer and promoter elements. Lower frequency interactions, indicated by a lower peak signal, are potentially indicative of

more transient interactions within the domain and have been suggested to be a result of random 'sampling' of the chromatin within a domain (Davies et al. 2016).

5.2.8 Capture-C reveals reduced enhancer-promoter interactions at some hypersensitive gene targets

Ten hypersensitive gene targets and twelve insensitive gene targets were investigated, for each of which a high-quality Capture-C probe was designed for the promoter. These gene targets were chosen based upon their importance for leukaemia and the level of downregulation at the nascent RNA level following EPZ-5676 treatment (Figure 5.7). For all 22 gene targets the overall domain structure was not perturbed following loss of H3K79me3 (Figure 5.8A-5.11A, S1-S14). This suggests that H3K79me does not affect the boundaries of the domains of any gene target, irrespective of transcription.

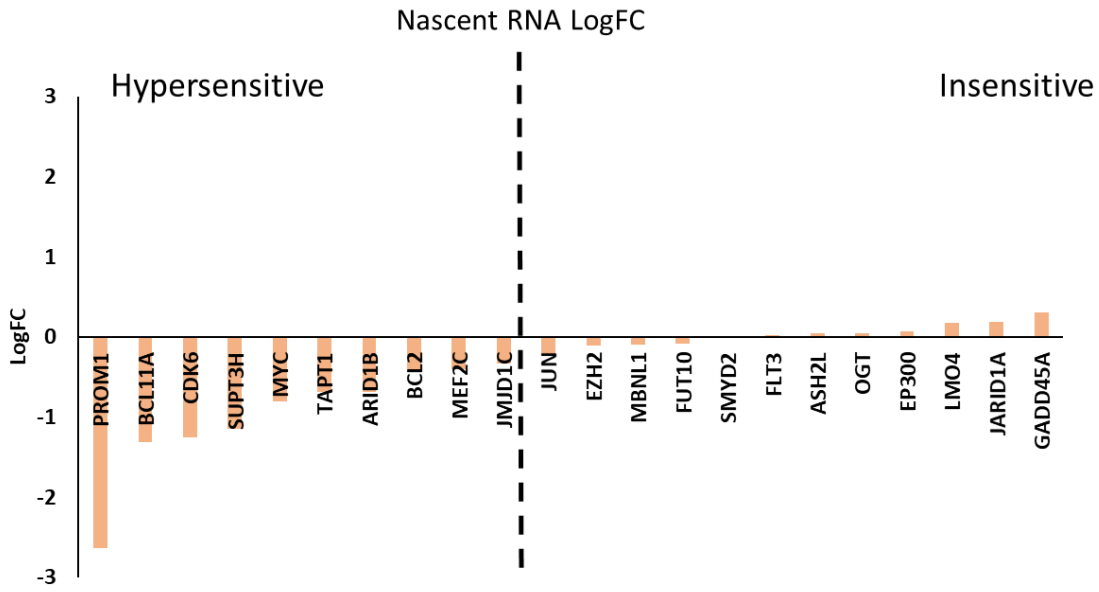

Figure 5-7. LogFC of hypersensitive and insensitive genes focused upon in this chapter. LogFC values calculated relative to 0µM control and averaged over three biological replicates. LogFC over +/- 0.5 represents significant differential expression.

Six of the ten hypersensitive gene targets demonstrated the presence of a putative intragenic enhancer (defined by ATAC, H3K27ac and H3K4me1 peaks in the gene body, Table 5.2), of which five (including *PROM1*, *TAPT1*, *ARID1B* and *BCL11A)* demonstrated a reduction of the high frequency interactions between the promoter and the intragenic enhancer by Capture-C, following 2µM EPZ-5676 treatment compared to the DMSO control (Figure 8A-9A, Table 5.2).

Table 5-2. Changes in Nascent RNA (LogFC), ATAC and Capture-C and presence of intragenic H3K27ac and H3K4me1 observed at hypersensitive and insensitive genes.

Hypersensitive genes

Gene target	LogFC	Change in ATAC	Intragenic enhancer (H3K27ac and H3K4me1 in gene body)	Obvious change in Capture-C
PROM1	-2.63	✓	✓	✓
BCL11A	-1.31	✓	✓	✓
CDK6	-1.25	✓	✓	✓
TAPT1	-0.69	✓	✓	✓
ARID1B	-0.68	✓	✓	✓
SUPT3H	-1.15	✓	x	x
MYC	-0.8	✓	x	x
BCL2	-0.4	✓	x	x
MEF2C	-0.4	✓	✓	x
JMJD1C	-0.33	✓	x	x

Insensitive genes

Gene target	Change in ATAC	Intragenic enhancer (H3K27ac and H3K4me1 in gene body)	Obvious change in Capture-C
ASH2L	✓	x	x
EP300	✓	x	x
EZH2	✓	x	x
FLT3	✓	x	x
FUT10	✓	x	x
GADD45A	✓	x	x
JARID1A	✓	x	x
JUN	✓	x	x
LMO4	✓	x	x
MBNL1	✓	x	x
OGT	✓	x	x
SMYD2	✓	x	x

The presence of a putative intragenic enhancer was not observed in the other four hypersensitive gene targets, *BCL2*, *MYC*, *JMJD1C* and *SUPT3H* (Figure 5.10C, S2, S4, Table 5.2). None of these four hypersensitive gene targets demonstrated an obvious reduction in high frequency interaction within the domain following EPZ-5676 treatment (Figure 5.10A, S2-S4, Table 5.2). This suggests that transcription at these hypersensitive gene targets may be controlled by H3K79me via an alternative mechanism. Finally, to determine whether reductions in chromatin interactions are specific to a subset of hypersensitive gene targets and not a general consequence of loss of H3K79me, Capture-C was performed at twelve insensitive targets following EPZ-5676 treatment. None of the twelve insensitive gene targets demonstrated the presence of an intragenic enhancer and no obvious changes in the high frequency interactions within the domains were observed (Figure 5.11, S5-14, Table 5.2). To verify this completely a genome wide analysis would be required to establish whether putative intragenic enhancers are specific to hypersensitive targets per se or whether any insensitive do display this feature, however, these results indicate that this effect may be unique to hypersensitive and/or intragenic enhancer-containing genes.

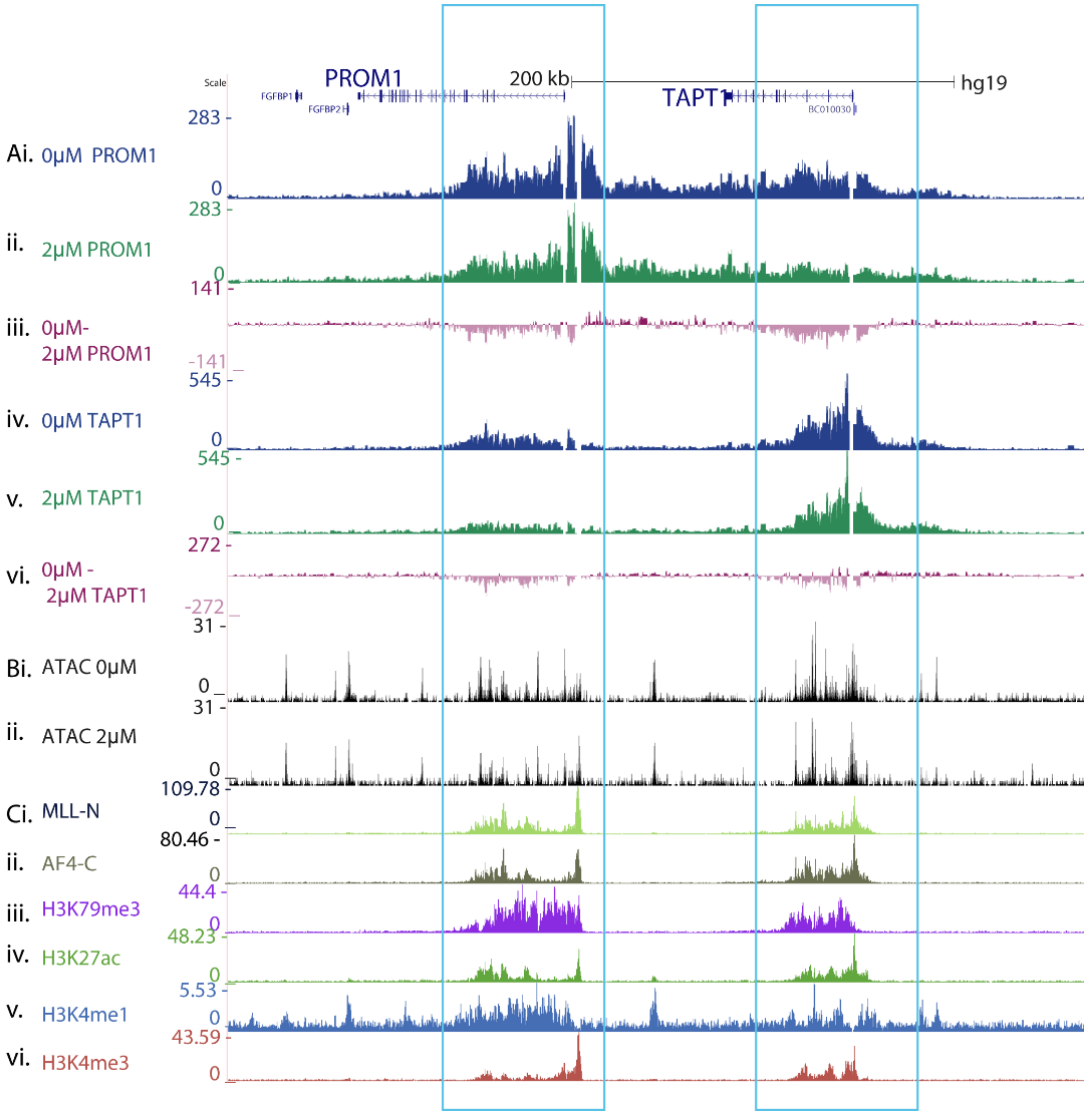

Figure 5-8. High-throughput analysis depicting chromatin features at *PROM1* and *TAPT1*. (A) Capture-C from both *PROM1* and *TAPT1* promoters following 7 days 0μM control and 2μM EPZ-5676 treatment of SEM cells. Pink tracks represent subtraction of 2μM from 0μM treatment. All tracks represent average of 3 biological replicates. (B) ATAC seq at *PROM1* and *TAPT1* following 7 days 0μM and 2μM EPZ-5676 treatment of SEM cells, representation of two biological replicates. Blue windows represent changes in chromatin interactions and accessibility observed using Capture-C and ATAC seq. (C) ChIP seq tracks at *PROM1* and *TAPT1* for relevant chromatin features including MLL-N, AF4-C, H3K79me3, H3K27ac, H3K4me1 and H3K4me3. ChIP seq generated in the lab by E.Ballabio, J.Kerry and T.Milne.

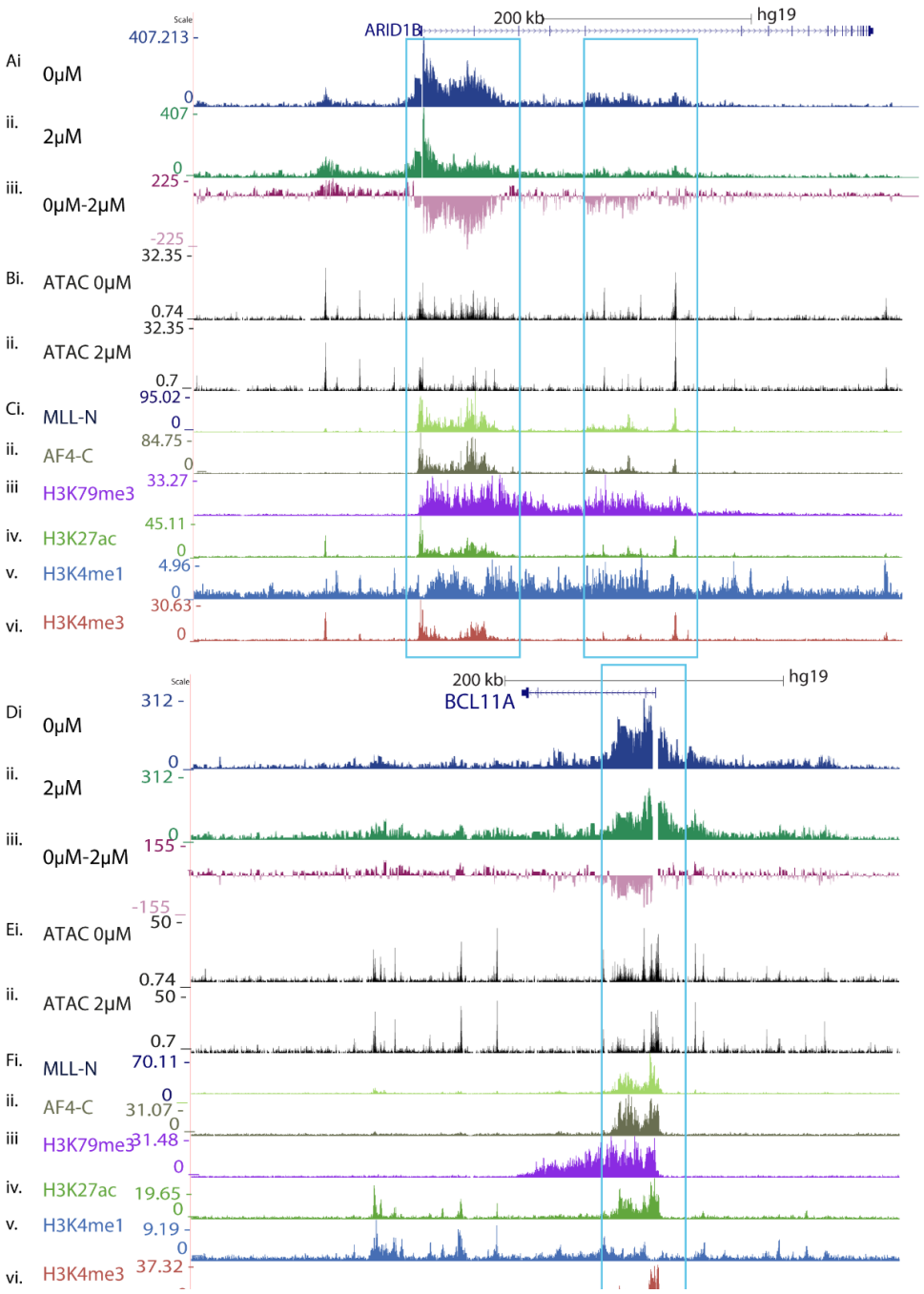

Figure 5-9. High through-put analysis depicting chromatin features at *ARID1B* and *BCL11A*. (A and D) Capture-C from both *ARID1B* and *BCL11A* promoters following 7 days 0μM control and 2μM EPZ-5676 treatment of SEM cells. Pink tracks represent subtraction of 2μM from 0μM treatment. All tracks represent average of three biological replicates. (B and E) ATAC seq at *ARID1B* and *BCL11A* following 7 days 0μM and 2μM EPZ-5676 treatment of SEM cells, representation of two biological replicates. Blue windows represent changes in chromatin interactions and accessibility observed using Capture-C and ATAC seq. (C and F) ChIP seq tracks at *ARID1B* and *BCL11A* for relevant chromatin features including MLL-N, AF4-C, H3K79me3, H3K27ac, H3K4me1 and H3K4me3. ChIP seq generated in the lab by E.Ballabio, J.Kerry and T.Milne.

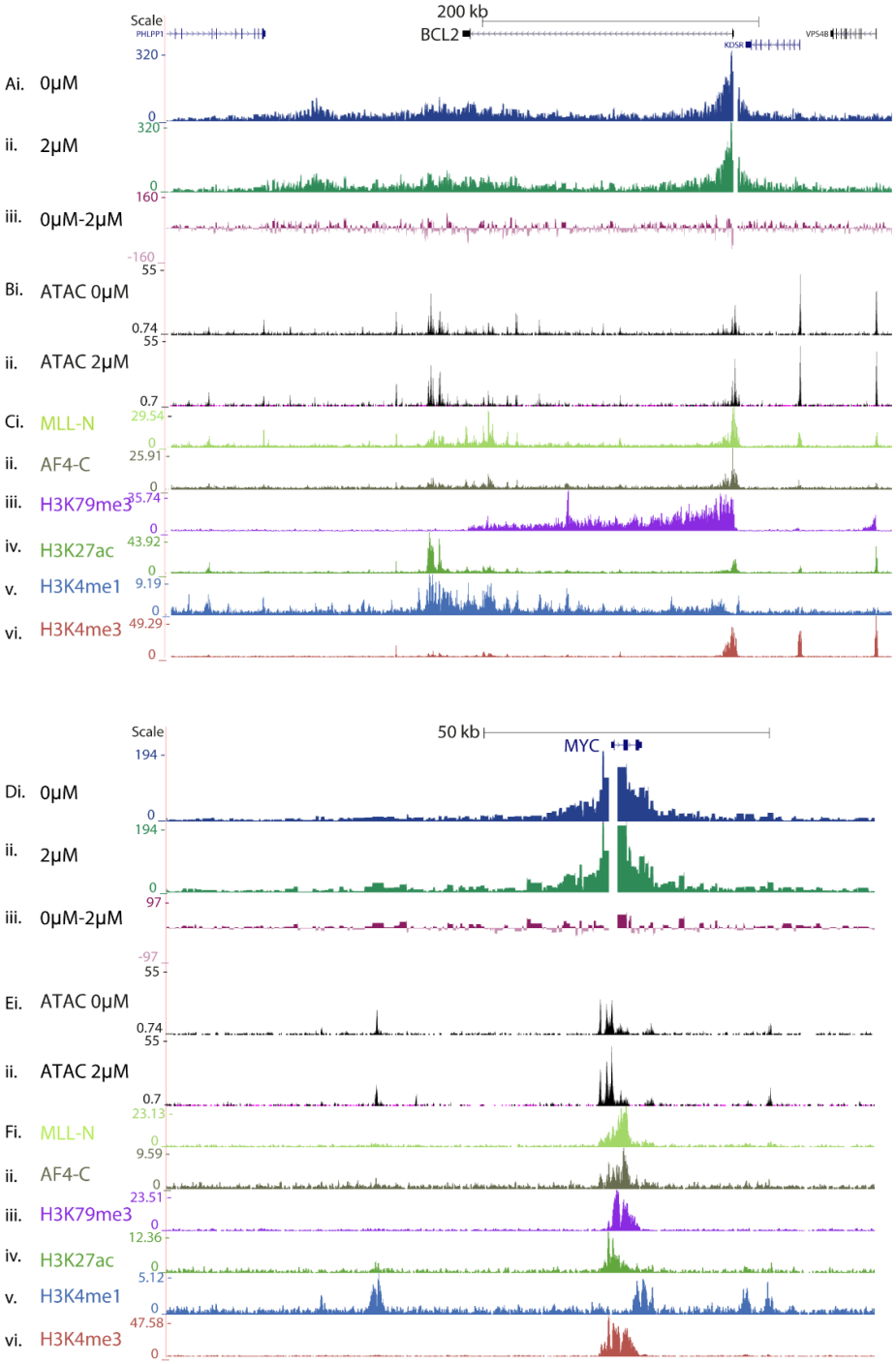

Figure 5-10. High-throughput analysis of chromatin features at BCL2 and MYC (A and D) Capture-C from BCL2 and MYC promoters following 7 days 0μM and 2μM EPZ-5676 treatment of SEM cells. Pink tracks represent subtraction of 2μM from 0μM treatment. All tracks represent average of three biological replicates. (B and E) ATAC seq at BCL2 and MYC following 7 days 0μM and 2μM EPZ-5676 treatment of SEM cells, representation of two biological replicates (C and F) ChIP seq tracks at BCL2 and MYC for relevant chromatin features including MLL-N, AF4-C, H3K79me3, H3K27ac, H3K4me1, H3K4me3. ChIP seq generated in the lab by E.Ballabio, J.Kerry and T.Milne.

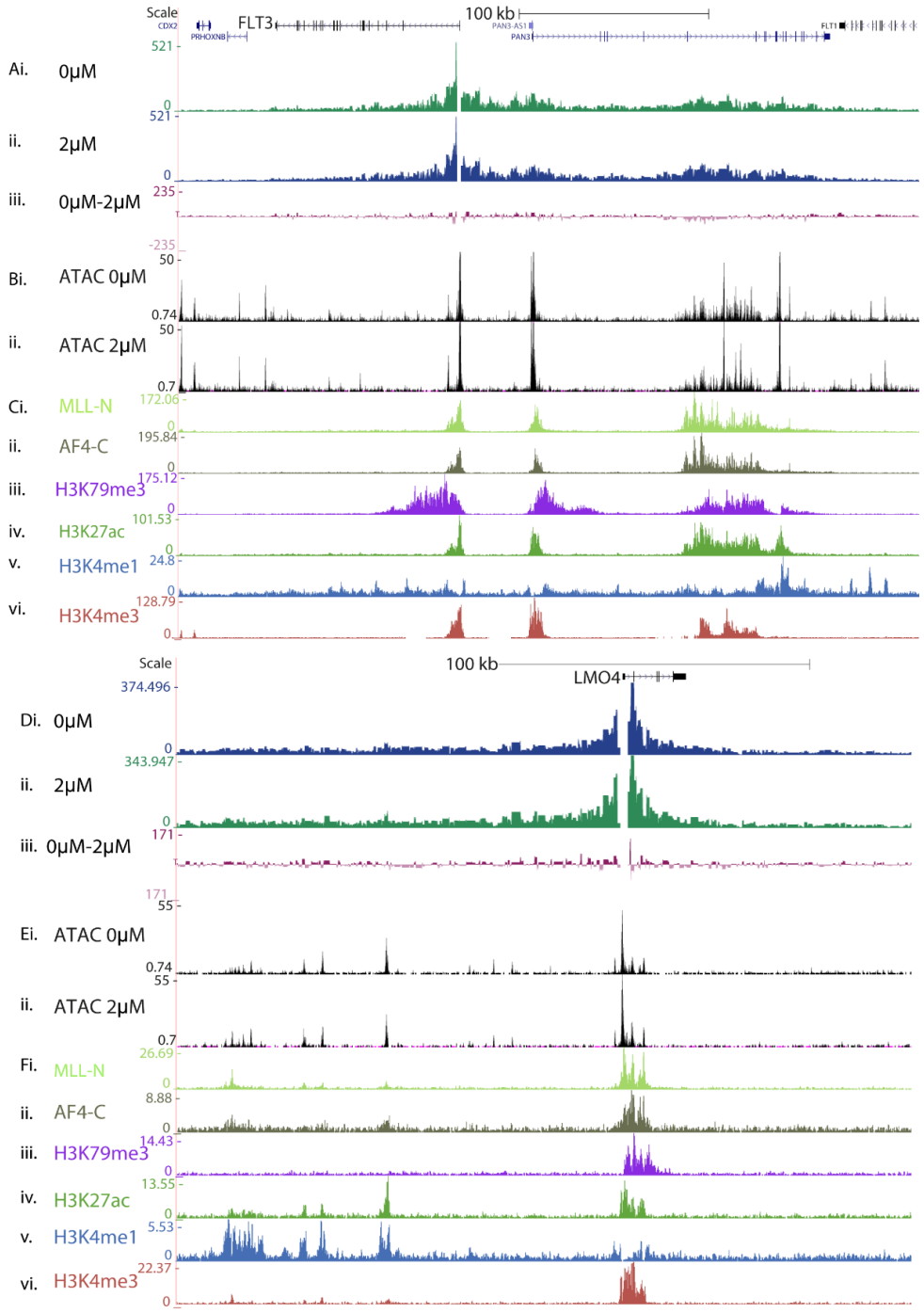

Figure 5-11 High throughput analysis of chromatin features at *FLT3* and *LMO4* (A and D) Capture-C from *FLT3* and *LMO4* promoters following 7 days 0μM and 2μM EPZ-5676 treatment of SEM cells. Pink tracks represent subtraction of 2μM from 0μM treatment. All tracks represent average of three biological replicates. (B and E) ATAC seq at *FLT3* and *LMO4* following 7 days 0μM and 2μM EPZ-5676 treatment of SEM cells, representation of two biological replicates (C and F) ChIP seq tracks at *FLT3* and *LMO4* for relevant chromatin features including MLL-N, AF4-C, H3K79me3, H3K27ac, H3K4me1, H3K4me3. ChIP seq generated in the lab by E.Ballabio, J.Kerry and T.Milne.

5.3 Summary

In summary, a potential novel mechanism of H3K79me has been identified which may occur in a context dependent manner. The loss of H3K79me leads to the downregulation of a subset of H3K79me gene targets. Following this loss, changes in histone modifications at both sensitive and insensitive gene targets occur, indicating that these changes alone cannot explain the effects on transcription. Upon further investigation, reductions in chromatin accessibility were observed in the gene body of some hypersensitive gene targets and not at insensitive gene targets. Furthermore, putative intragenic enhancers, marked with H3K27ac, H3K4me1 and H3K79me, were identified in the hypersensitive gene targets which demonstrated a reduction in chromatin accessibility in the gene body. Following loss of H3K79me, a reduction in the interaction between the putative intragenic enhancer and promoter was observed at hypersensitive targets which demonstrated this feature. This suggests a possible new role for H3K79me as regulator of intragenic enhancer-promoter interactions, probably in a specific context.

5.4 Discussion

5.4.1 A role of H3K79me in intragenic enhancer function

H3K79me is classically associated with transcription elongation, with H3K79me being found in the gene body of most active genes (Steger et al. 2008). However, an additional role has been suggested in *D. melanogaster*, where H3K79me has been shown to be present at a subset of enhancers (Bonn et al. 2012). Here, it is shown that H3K79me may mark a set of intragenic enhancers, which are found within the gene body of a subset of gene targets. Importantly it is worth emphasising that physical perturbation is observed between the enhancer and promoter following loss of H3K79me, indicating that this is not just a correlation between H3K79me and putative enhancer localisation, it is a potential new functional mechanism of DOT1L and H3K79me. As H3K79me is found in the gene body of most active genes, it may therefore serve as a dual regulator of both elongation and enhancer function at these targets, due to the location of the enhancer.

This raises an important question about the nature of the difference between intergenic and intragenic enhancers. Some hypersensitive gene targets which do not appear to have intragenic enhancers include *BCL2* and *MYC*. Both *BCL2* and *MYC* have intergenic distal enhancers (Godfrey et al. 2016; Uslu et al. 2014). No changes in interaction are observed between the intergenic enhancer and promoter of *BCL2* or *MYC* by Capture-C. Likewise, there are no differences in chromatin accessibility at *BCL2* and *MYC* enhancers following EPZ-5676 treatment. This could be due to the presence of H3K79me and other elongation complexes in addition to enhancer complexes at intragenic enhancers.

Even though *BCL2* and *MYC* do not demonstrate the presence of an intragenic enhancer, these genes are still downregulated following EPZ-5676 treatment. This suggests that these targets are controlled by H3K79me by an alternative mechanism, such as ongoing antagonism of repressive chromatin proteins in the gene body, allowing transcription elongation to occur. Alternatively, there is always the possibility that they are regulated by an indirect effect due to the loss of H3K79me.

One example of a protein that could affect elongation is AF9, a protein shown to interact directly with DOT1L and which is purified in the super elongation complex (Lin et al. 2011), that can also bind to H3K9ac (Wan et al. 2017; Li et al. 2014), as well as MED26, a subunit of the mediator complex (Takahashi et al. 2011; Biswas et al. 2011). As H3K9ac is reduced following loss of H3K79me, destabilisation of AF9 may influence mediator stabilisation. This highlights one way in which disrupting H3K79me may perturb the protein:protein interactions occurring between enhancers and promoters, especially when transcription elongation and enhancer-promoter interactions are brought into close proximity. It also emphasizes a way in which histone modifications may contribute to gene regulation in a context dependent manner.

5.4.2 The context dependent role of repressive histone modifications

Another way histone modifications could lead to the downregulation of gene targets in a context dependent manner is in the case of the accumulation of repressive H3K9me. H3K9me has recently been shown to cause phase separation of heterochromatin via recruitment of HP1, which may drive the formation of a heterochromatic state (Strom et al. 2017; Larson et al. 2017). In addition to this, the repressive modification H3K27me3 has been shown to block transcription factor binding (Petruk et al. 2017). Therefore, the accumulation of repressive

modifications may lead to a chromatin structure which perturbs activating transcription factor binding at intragenic enhancers.

In section 5.2.1 it was observed that most of the histone modifying enzymes that were checked were not perturbed at the level of nascent RNA following EPZ-5676 treatment. However, 4 enzymes, including GCN5, HDAC1, G9A and SIRT3 were significantly de-regulated (Table 5.1). Changes in the level of nascent RNA may be translated into de-regulation at the protein level. The changes in protein levels could provide a simple explanation of why changes in histone modifications are observed nonspecifically. In particular, SIRT3 showed an increase in expression and this may be partly responsible for the apparent global decrease in acetylation and increase in H3K9me seen at all targets tested including silent ones such as HOXC8. The fact that HOXC8 is not expressed or marked with H3K79me in SEM cells but an increase in H3K9me is observed following EPZ-5676 treatment is interesting. This further supports the non-specific changes in histone modifications are not responsible for specific changes in transcription observed.

5.4.3 H3K79me and the modulation of transcription factors

If H3K79me is ultimately modulating transcription factor binding at the subset of hypersensitive targets, it might be expected that specific transcription factors are found at these targets compared to insensitive genes. Whilst time constraints have not allowed this to be fully addressed in this thesis, preliminary analysis showed that one transcription factor motif which is enriched at the downregulated targets is the motif for ERG. ERG is an ETS transcription factor which has previously been implicated in acute leukaemia (Coskun et al. 2011) and has been shown to interact with the H3K9 methyltransferase SETDB1 (Yang et al. 2002). Therefore, H3K79me may antagonize SETDB1 which may prevent binding of repressive

transcription factors. In addition to this, it is also possible that H3K79me is stabilizing activating transcription factors, such as MYB, which is supported by the observed reductions in chromatin accessibility at the putative enhancer regions following EPZ-5676 treatment.

Chapter 6 – General Discussion

MLL-AF4 leukaemia is the most prevalent MLL-r leukaemia. It is also understudied and has a very poor prognosis (Meyer et al. 2013). Understanding the mechanism by which MLL-AF4 controls the transcription of genes is paramount to understanding the molecular mechanism of this disease, in the anticipation of identifying better therapeutic strategies. DOT1L and H3K79me have been shown to be important for both MLL-AF4 gene targets expression and leukaemogenesis (Okada et al. 2005; Bernt et al. 2011). Although other pathways of MLL-AF4 gene activation have been studied, aberrant levels of H3K79me are found at over 95% of MLL-AF4 target genes and there are currently DOT1L inhibitors in clinical trials (Bernt et al. 2011; Krivtsov et al. 2008). Despite this, very little was previously known about (1) how DOT1L is initially recruited and stabilised to MLL-AF4 gene targets and (2) once there, how DOT1L and H3K79me contribute to the high levels of transcription observed.

6.1 Model of DOT1L recruitment and function at hypersensitive MLL-AF4 gene targets

Taking data from this thesis we've developed a working model of DOT1L recruitment and function at hypersensitive MLL-AF4 gene targets. As AF4 is not sufficient for DOT1L recruitment, it is probable MLL-AF4 may control a subset of genes via indirect recruitment and stabilisation of DOT1L. Of note, the AF9-AF4 and AF9-DOT1L interactions are mutually exclusive and therefore, cannot occur simultaneously at an MLL-AF4 gene targets, this has been demonstrated in this thesis and in published literature (Leach et al. 2013; Biswas et al. 2011; Yokoyama et al. 2010). New data from this thesis has demonstrated that both AF9 and

PAF1 are sufficient for DOT1L recruitment *in vivo* (Chapter 3), in line with published data demonstrating the interaction between AF9-DOT1L and an association between PAF1 and DOT1L *in vitro* (Kim et al. 2010; Leach et al. 2013; Yokoyama et al. 2010; Biswas et al. 2011). Together, these interactions could lead to high local concentrations of PAF1 and AF9 or ENL at MLL-AF4 gene targets which may generate a high affinity, stable binding site for DOT1L (Figure 6.1A). Following DOT1L stabilisation at a gene, H3K79me is deposited in the gene body and simultaneously at the putative intragenic enhancer of hypersensitive MLL-AF4 genes. H3K79me may antagonise repressive proteins, such as SIRT1, thereby creating a potentially permissive chromatin state. This may promote high affinity transcription factor binding which may in part facilitate the interaction between the enhancer and promoter and enhance transcription (Figure 6.1A).

Following loss of H3K79me, the binding affinity of SIRT1 and other repressive chromatin modifying enzymes could increase thus causing activating modifications such as H3 acetylation to decrease (Figure 6.1B). This may cause a decrease in chromatin accessibility leading to a reduction in the binding affinity of activating transcription factors at the putative intragenic enhancer regions (Figure 6.1B). The loss of activating transcription factor binding may lead to the reduction in enhancer-promoter interaction and reduction in transcription observed (Figure 6.1B).

6.2 How does this model account for specificity of H3K79me function?

It has been observed that following loss of H3K79me, downregulation of a subset of H3K79me marked gene targets occurs, demonstrating that H3K79me has a specific function at these genes. Therefore, it was important to address what accounted for this specificity and whether proposed models may explain this. Loss of H3K79me3 did demonstrate reduction in histone

acetylation and an increase in repressive modifications, in line with published results, however, this was observed at both hypersensitive and insensitive genes. Therefore, this does not alone account for the specific transcriptional changes we observe (Chapter 5). Changes in histone modifications have been shown to effect transcription factor binding, which do confer specificity by binding to specific DNA sequences. Examples of this include H3K27me3 and H3K9me3, which have been demonstrated to perturb transcription factor binding (Petruk et al. 2017; Soufi et al. 2012; Lupien et al. 2008). Furthermore, it has been suggested that repressive complexes and histone modifications cause a decreases in chromatin accessibility which may disrupt the ability of DNA binding proteins to access their binding sites (Fitzgerald & Bender 2001; Francis et al. 2001). Several hypersensitive genes have been demonstrated to contain putative embedded enhancers which are directly marked with H3K79me3. Therefore, one way which H3K79me3 could be required specifically for hypersensitive genes is if the binding of particular transcription factors that regulate specific genes was disrupted by the creation of a repressive chromatin domain. CREB1 binding has been suggested to be dependent upon H3K79me (Gilan et al. 2016), however, analysis from this thesis demonstrated that CREB1 binding sites were not enriched at downregulated gene targets. Instead, other transcription factors whose motifs are enriched at downregulated genes and that are expressed in SEM cells, include ETS family transcription factors (such as ELF1 and ERG) (Figure 5.6). Therefore, the binding of these specific transcription factors to the embedded putative enhancer, marked with H3K79me, could confer specificity of H3K79me3 function at hypersensitive gene targets. Thus the specificity of H3K79me loss could in part be controlled by the particular dependency these gene targets have on the binding of ETS factors to intragenic enhancers.

6.4 Future work – Does H3K79me modulate specific transcription factor binding?

Taken together, the next most important question to address is whether H3K79me modulates the binding of ETS transcription factors such as ELF1 and ERG. To initially test this, ChIP qPCR for ELF1 and ERG can be conducted to determine whether they bind to hypersensitive genes. If transcription factor binding is dependent upon H3K79me at hypersensitive genes, then following loss of H3K79me, it would be expected that transcription factor binding is reduced. To test this ChIP qPCR following EPZ-5676 treatment, or a complementary approach such as DOT1L knockdown, at hypersensitive genes can be conducted.

If transcription factor binding is reduced following loss of H3K79me, then it is possible that the reduction in transcription and reduction of interaction between the enhancer and promoter is due to the specific reduction in transcription factor binding and subsequent reduction in activity of the enhancer at hypersensitive genes. To test this, siRNA mediated knockdown of specific transcription factors could be used coupled with Capture-C and ATAC seq. Even though this does not rule out the potential of co-operative transcription factor binding, and further work potentially involving combinatorial transcription factor knockdown would be needed to address this, it provides an initial way to test this hypothesis.

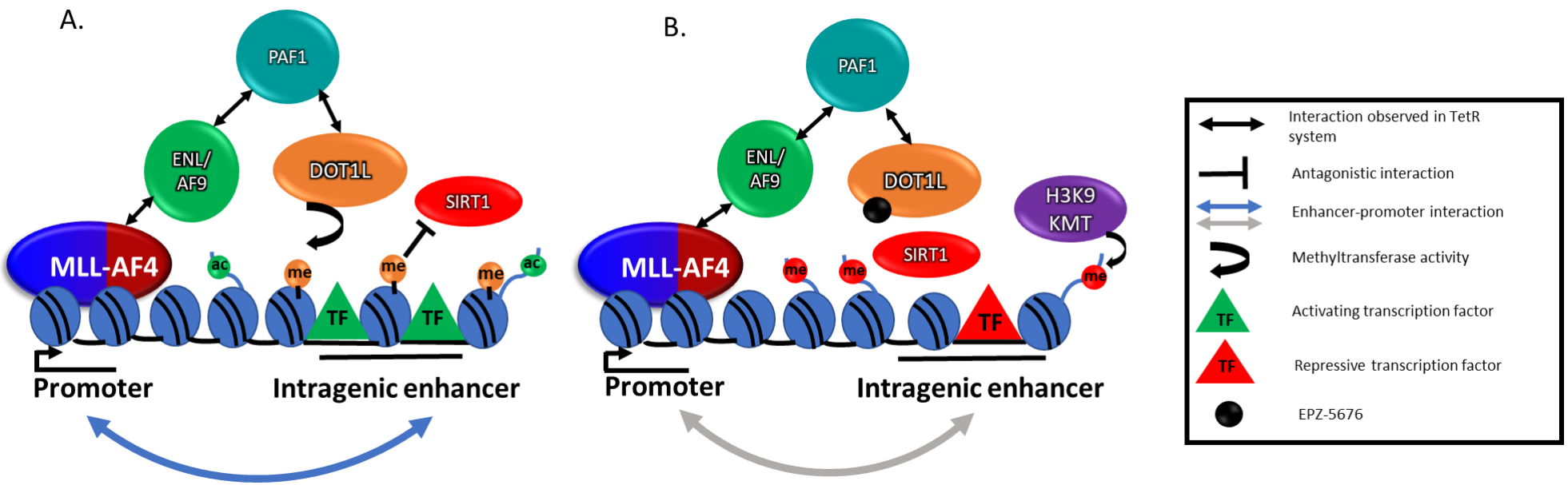

Figure 6-1 Model of DOT1L recruitment and function at MLL-AF4 hypersensitive genes (A) DOT1L may be recruited to MLL-AF4 gene targets via stabilisation of AF9/ENL and PAF1. Stable recruitment of DOT1L leads to H3K79me found in the gene body and putative intragenic enhancer. H3K79me may contribute to active transcription and enhancer function via antagonism of repressive histone modifying enzymes. This may lead to an increase in chromatin accessibility and transcription factor binding which may facilitate promoter-enhancer interactions (blue arrow) (B) Following DOT1L inhibition, H3K79me is lost from the gene. Therefore, antagonism of repressive chromatin modifying enzymes may be perturbed. This may lead to an increase of repressive chromatin modifications. Furthermore, this may result in reduced or altered chromatin accessibility which may prevent and promote activating and repressive transcription factor binding, respectively. Overall, this may lead to perturbations in the interaction between the promoter and enhancer (grey arrow) and a subsequent reduction in transcription at hypersensitive genes

Supplementary data

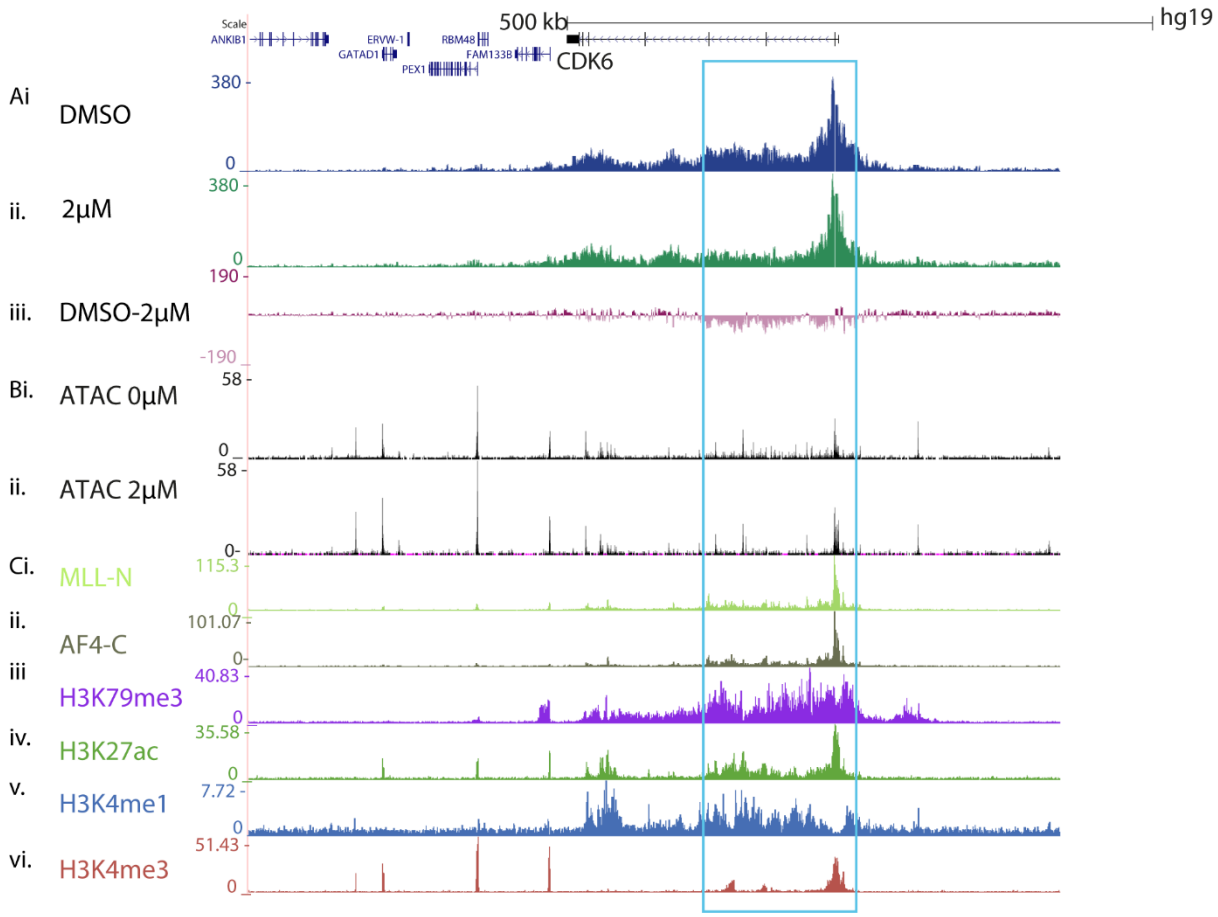

Figure S1. High throughput analysis of chromatin features at *CDK6* (A) Capture-C from *CDK6* promoter following 7 days 0μM and 2μM EPZ-5676 treatment of SEM cells. Pink tracks represent subtraction of 2μM from 0μM treatment. All tracks represent average of three biological replicates. (B) ATAC-seq at *CDK6* following 7 days 0μM and 2μM EPZ-5676 treatment of SEM cells, representation of two biological replicates (C) ChIP seq tracks at *CDK6* of relevant chromatin features including MLL-N, AF4-C, H3K79me3, H3K27ac, H3K4me1, H3K4me3. ChIP seq generated in the lab by E.Ballabio, J.Kerry and T.Milne

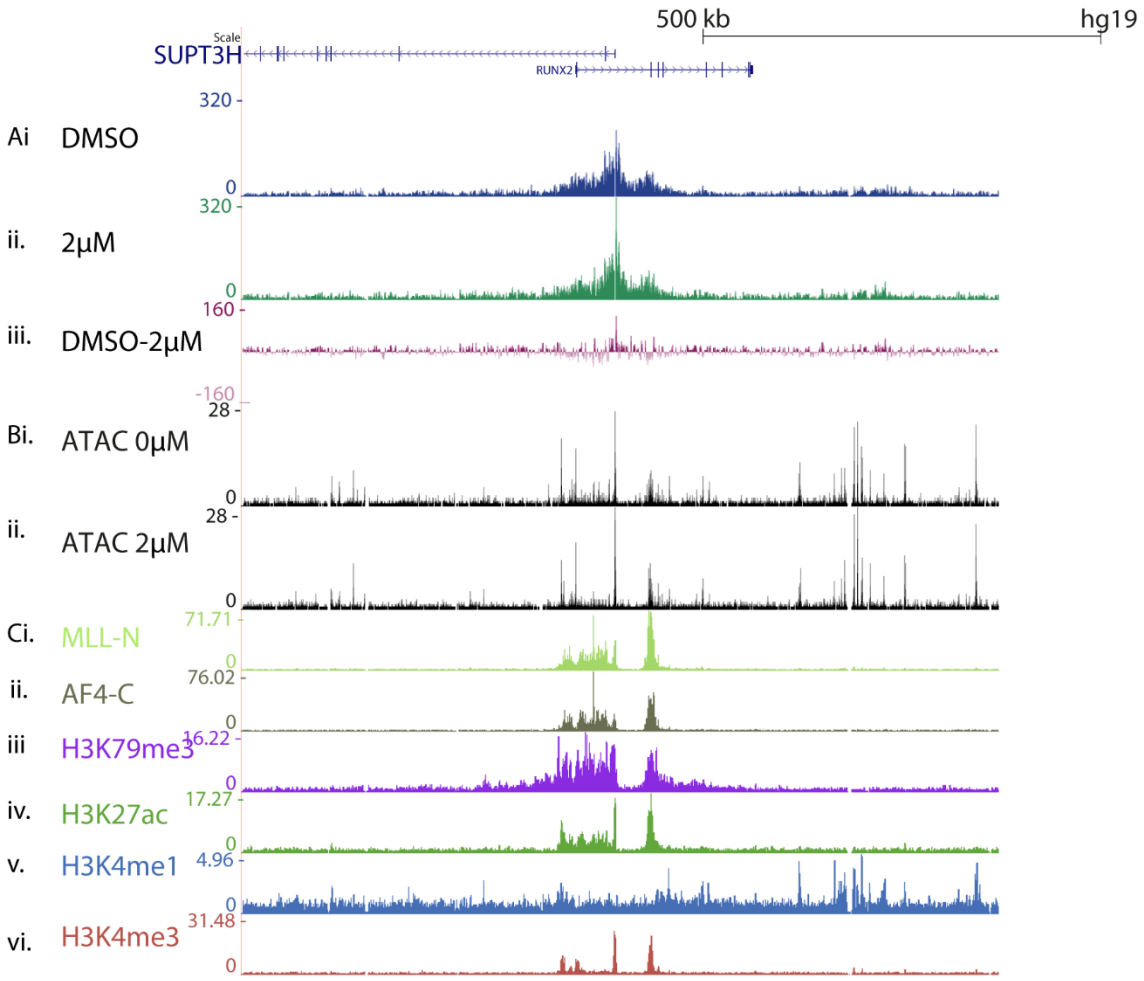

Figure S2. High throughput analysis of chromatin features at *SUPT3H* (A) Capture-C from *SUPT3H* promoter following 7 days 0µM and 2µM EPZ-5676 treatment of SEM cells. Pink tracks represent subtraction of 2µM from 0µM treatment. All tracks represent average of three biological replicates. (B) ATAC seq at *SUPT3H* following 7 days 0µM and 2µM EPZ-5676 treatment of SEM cells, representation of two biological replicates (C) ChIP seq tracks at *SUPT3H* of relevant chromatin features including MLL-N, AF4-C, H3K79me3, H3K27ac, H3K4me1, H3K4me3. ChIP seq generated in the lab by E.Ballabio, J.Kerry and T.Milne.

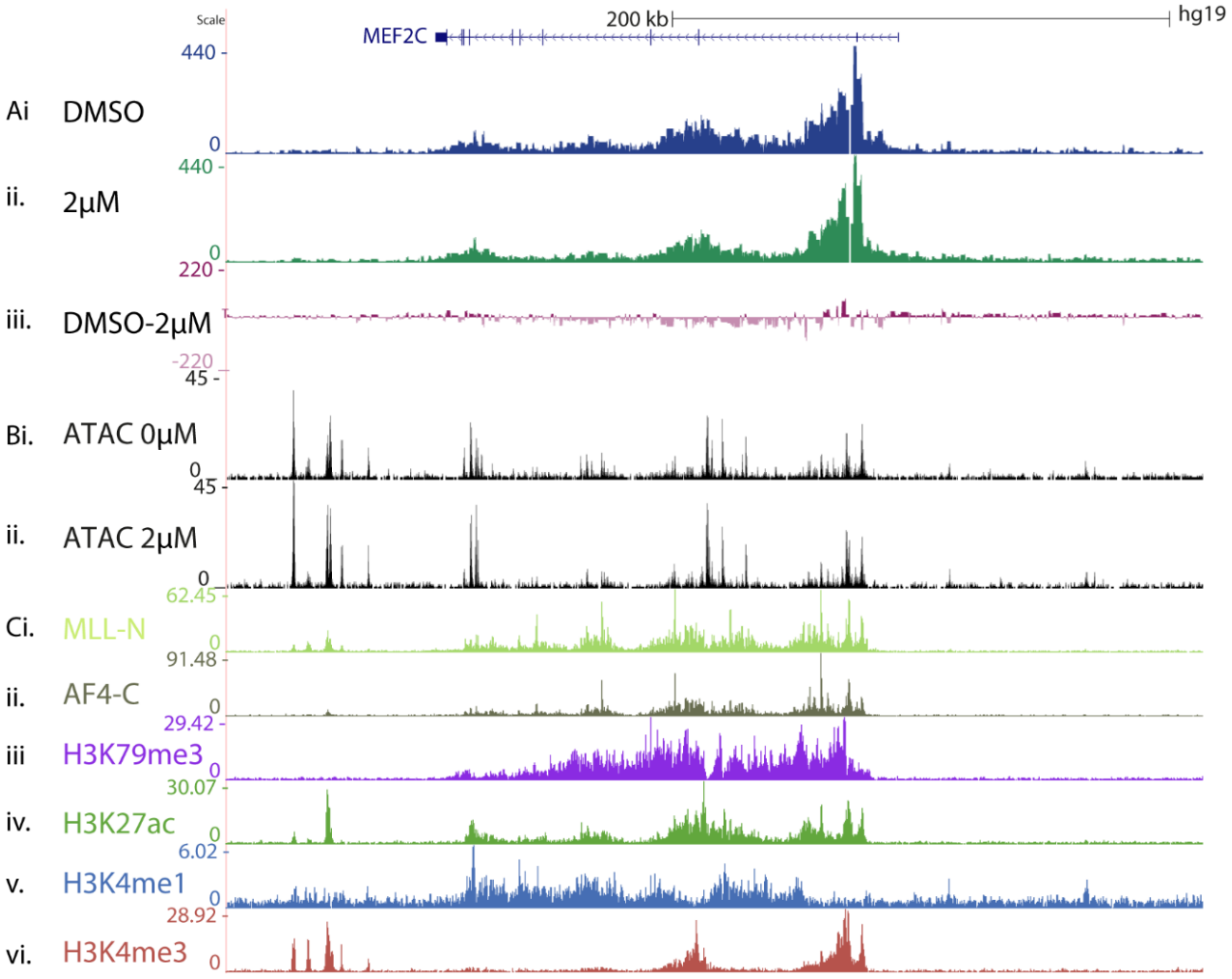

Figure S9. High throughput analysis of chromatin features at *MEF2C* (A) Capture-C from *MEF2C* promoter following 7 days 0µM and 2µM EPZ-5676 treatment of SEM cells. Pink tracks represent subtraction of 2µM from 0µM treatment. All tracks represent average of three biological replicates. (B) ATAC seq at *MEF2C* following 7 days 0µM and 2µM EPZ-5676 treatment of SEM cells, representation of two biological replicates (C) ChIP seq tracks at *MEF2C* of relevant chromatin features including MLL-N, AF4-C, H3K79me3, H3K27ac, H3K4me1, H3K4me3. ChIP seq generated in the lab by E.Ballabio, J.Kerry and T.Milne.

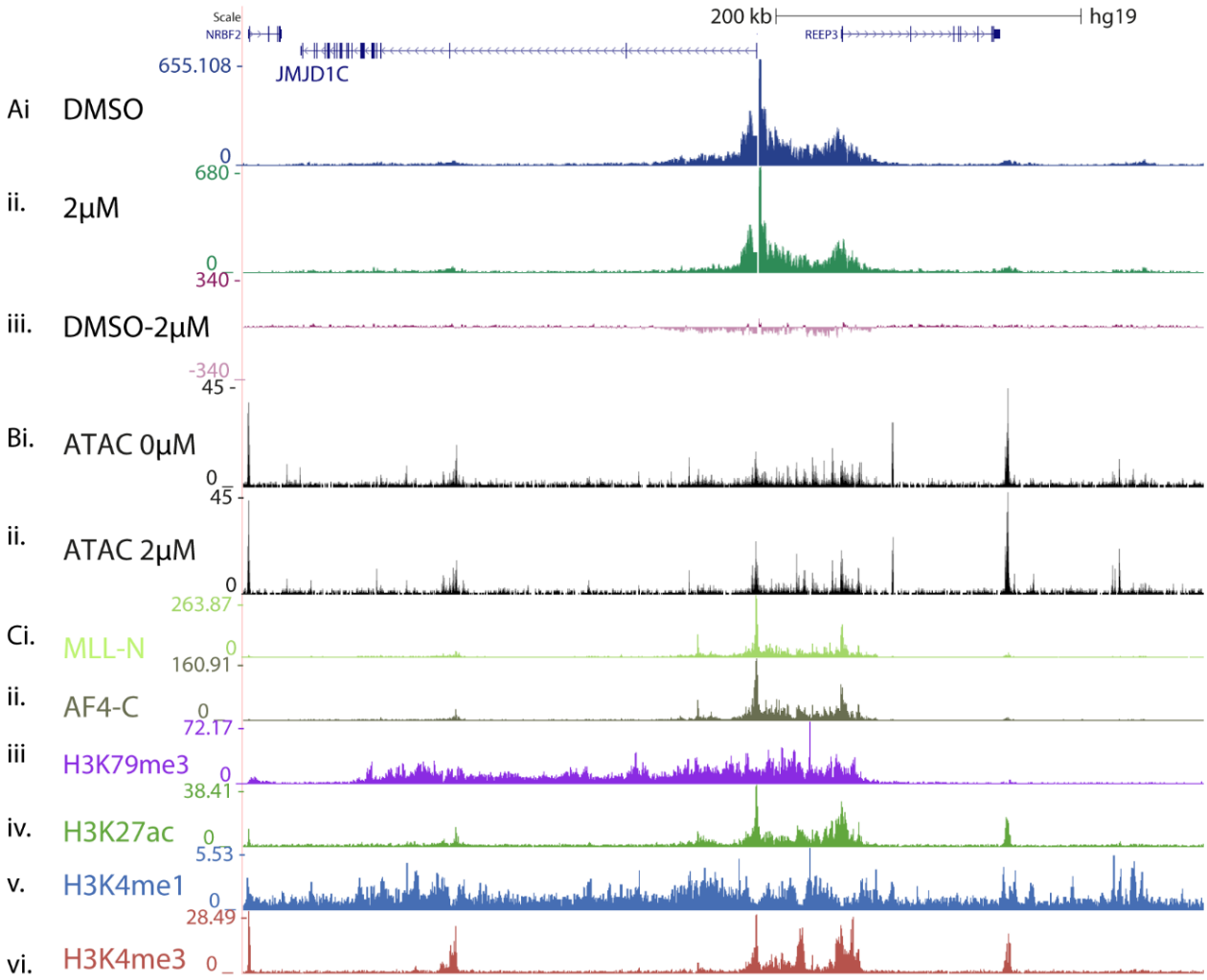

Figure S10. High throughput analysis of chromatin features at *JMJD1C* (A) Capture-C from *JMJD1C* promoter following 7 days 0µM and 2µM EPZ-5676 treatment of SEM cells. Pink tracks represent subtraction of 2µM from 0µM treatment. All tracks represent average of three biological replicates. (B) ATAC seq at *JMJD1C* following 7 days 0µM and 2µM EPZ-5676 treatment of SEM cells, representation of two biological replicates (C) ChIP seq tracks at *JMJD1C* of relevant chromatin features including MLL-N, AF4-C, H3K79me3, H3K27ac, H3K4me1, H3K4me3. ChIP seq generated in the lab by E.Ballabio, J.Kerry and T.Milne.

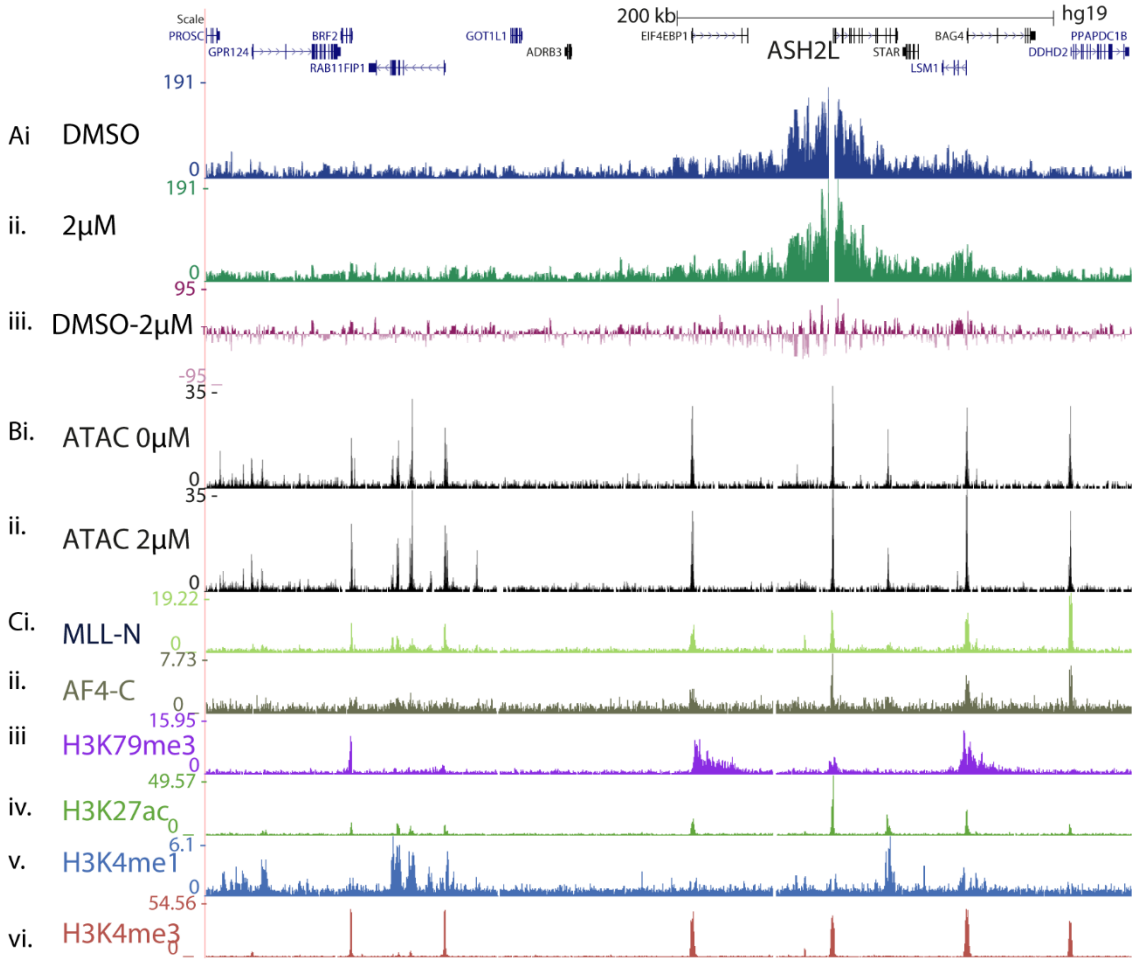

Figure S11. High throughput analysis of chromatin features at *ASH2L* (A) Capture-C from *ASH2L* promoter following 7 days 0µM and 2µM EPZ-5676 treatment of SEM cells. Pink tracks represent subtraction of 2µM from 0µM treatment. All tracks represent average of three biological replicates. (B) ATAC seq at *ASH2L* following 7 days 0µM and 2µM EPZ-5676 treatment of SEM cells, representation of two biological replicates (C) ChIP seq tracks at *ASH2L* of relevant chromatin features including MLL-N, AF4-C, H3K79me3, H3K27ac, H3K4me1, H3K4me3. ChIP seq generated in the lab by E.Ballabio, J.Kerry and T.Milne.

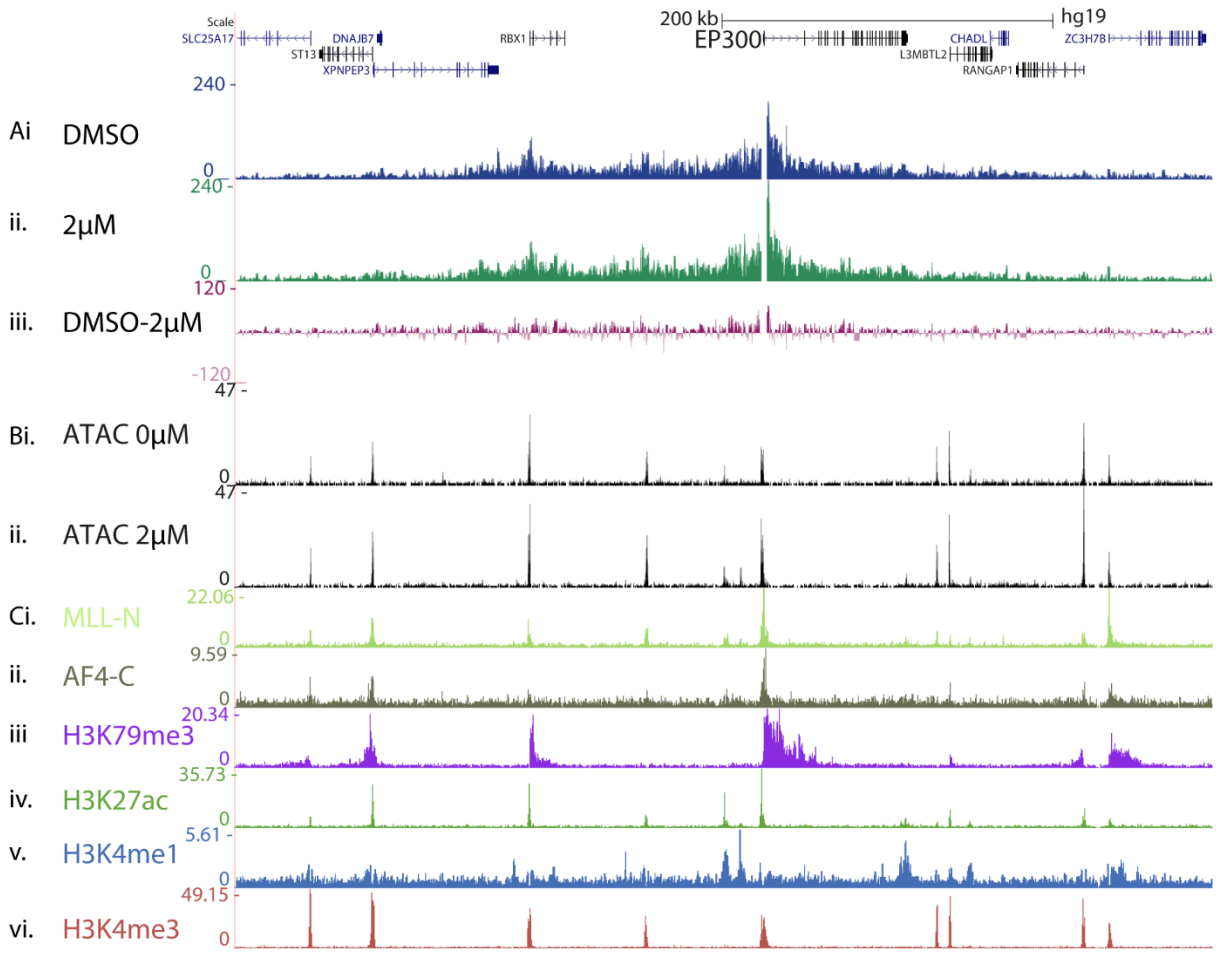

Figure S12. High throughput analysis of chromatin features at *EP300* (A) Capture-C from *EP300* promoter following 7 days 0μM and 2μM EPZ-5676 treatment of SEM cells. Pink tracks represent subtraction of 2μM from 0μM treatment. All tracks represent average of three biological replicates. (B) ATAC seq at *EP300* following 7 days 0μM and 2μM EPZ-5676 treatment of SEM cells, representation of two biological replicates (C) ChIP seq tracks at *EP300* of relevant chromatin features including MLL-N, AF4-C, H3K79me3, H3K27ac, H3K4me1, H3K4me3. ChIP seq generated in the lab by E.Ballabio, J.Kerry and T.Milne.

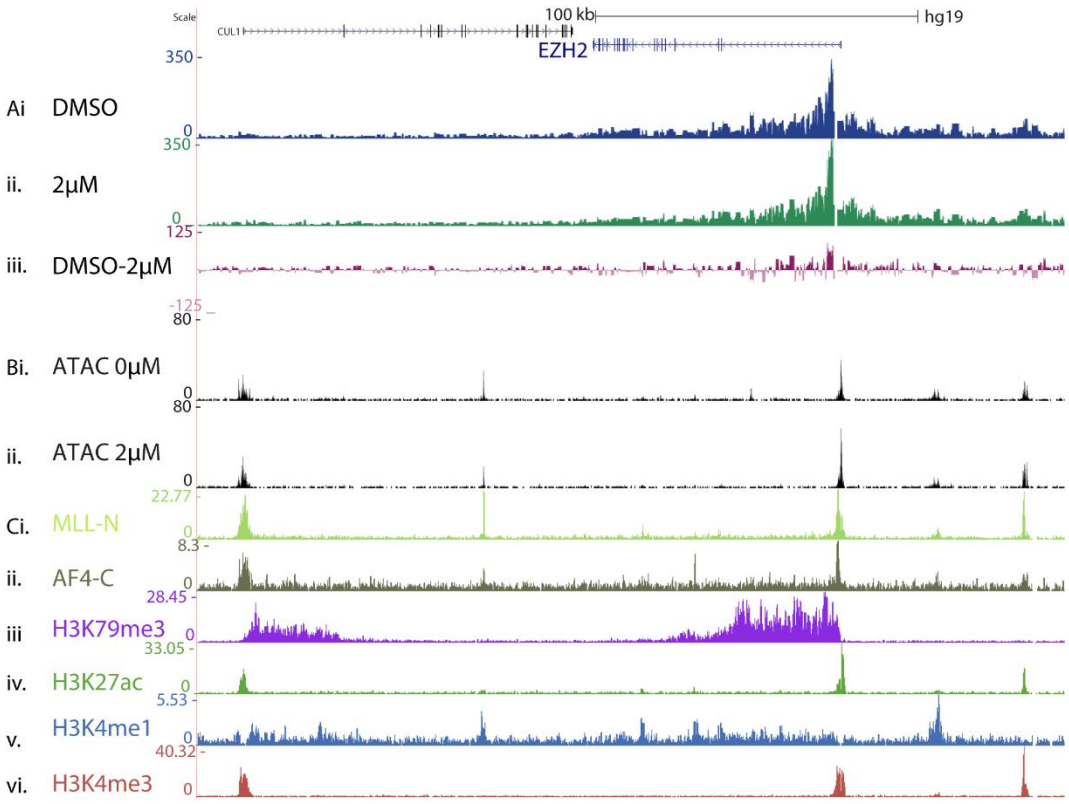

Figure S13. High throughput analysis of chromatin features at *EZH2* (A) Capture-C from *EZH2* promoter following 7 days 0μM and 2μM EPZ-5676 treatment of SEM cells. Pink tracks represent subtraction of 2μM from 0μM treatment. All tracks represent average of three biological replicates. (B) ATAC seq at *EZH2* following 7 days 0μM and 2μM EPZ-5676 treatment of SEM cells, representation of two biological replicates (C) ChIP seq tracks at *EZH2* of relevant chromatin features including MLL-N, AF4-C, H3K79me3, H3K27ac, H3K4me1, H3K4me3. ChIP seq generated in the lab by E.Ballabio, J.Kerry and T.Milne.

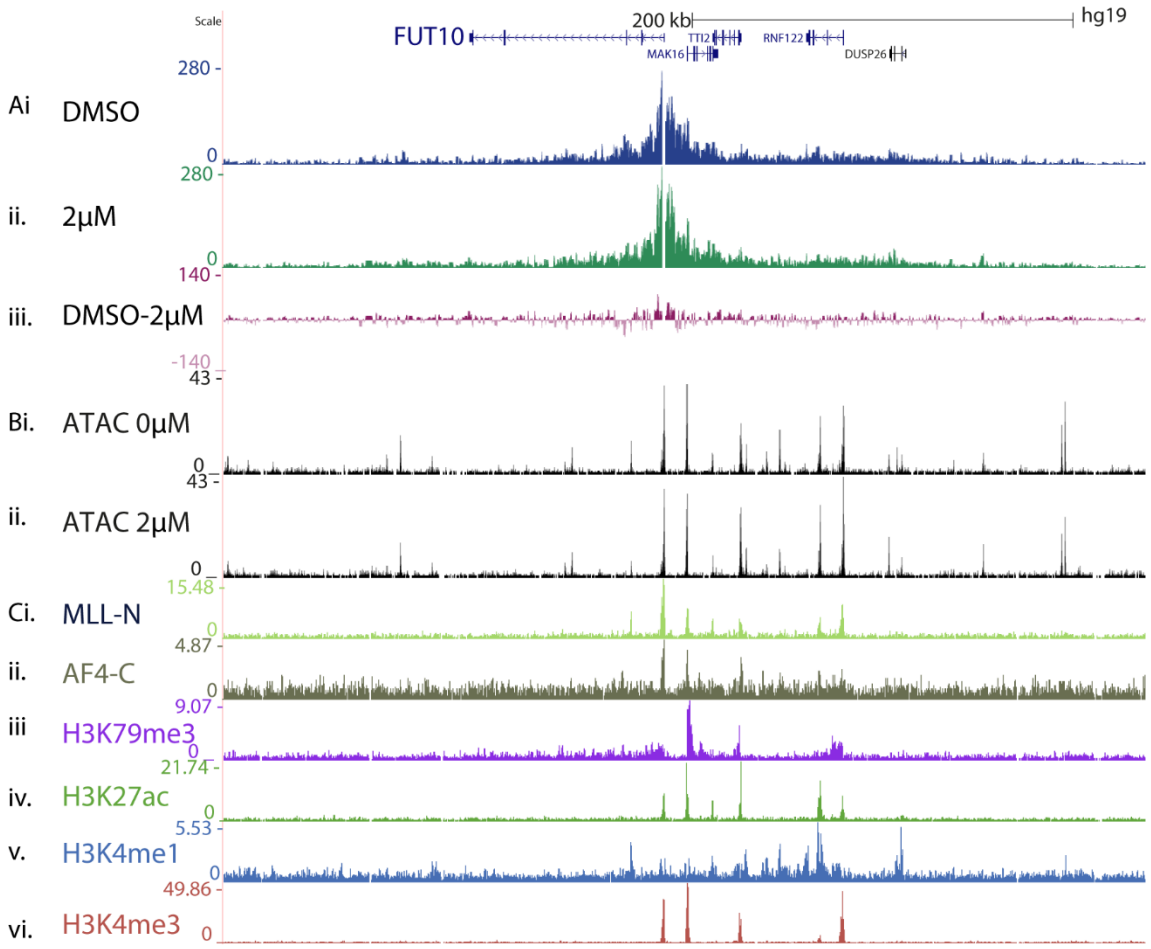

Figure S14. High throughput analysis of chromatin features at *FUT10* (A) Capture-C from *FUT10* promoter following 7 days 0µM and 2µM EPZ-5676 treatment of SEM cells. Pink tracks represent subtraction of 2µM from 0µM treatment. All tracks represent average of three biological replicates. (B) ATAC seq at *FUT10* following 7 days 0µM and 2µM EPZ-5676 treatment of SEM cells, representation of two biological replicates (C) ChIP seq tracks at *FUT10* of relevant chromatin features including MLL-N, AF4-C, H3K79me3, H3K27ac, H3K4me1, H3K4me3. ChIP seq generated in the lab by E.Ballabio, J.Kerry and T.Milne.

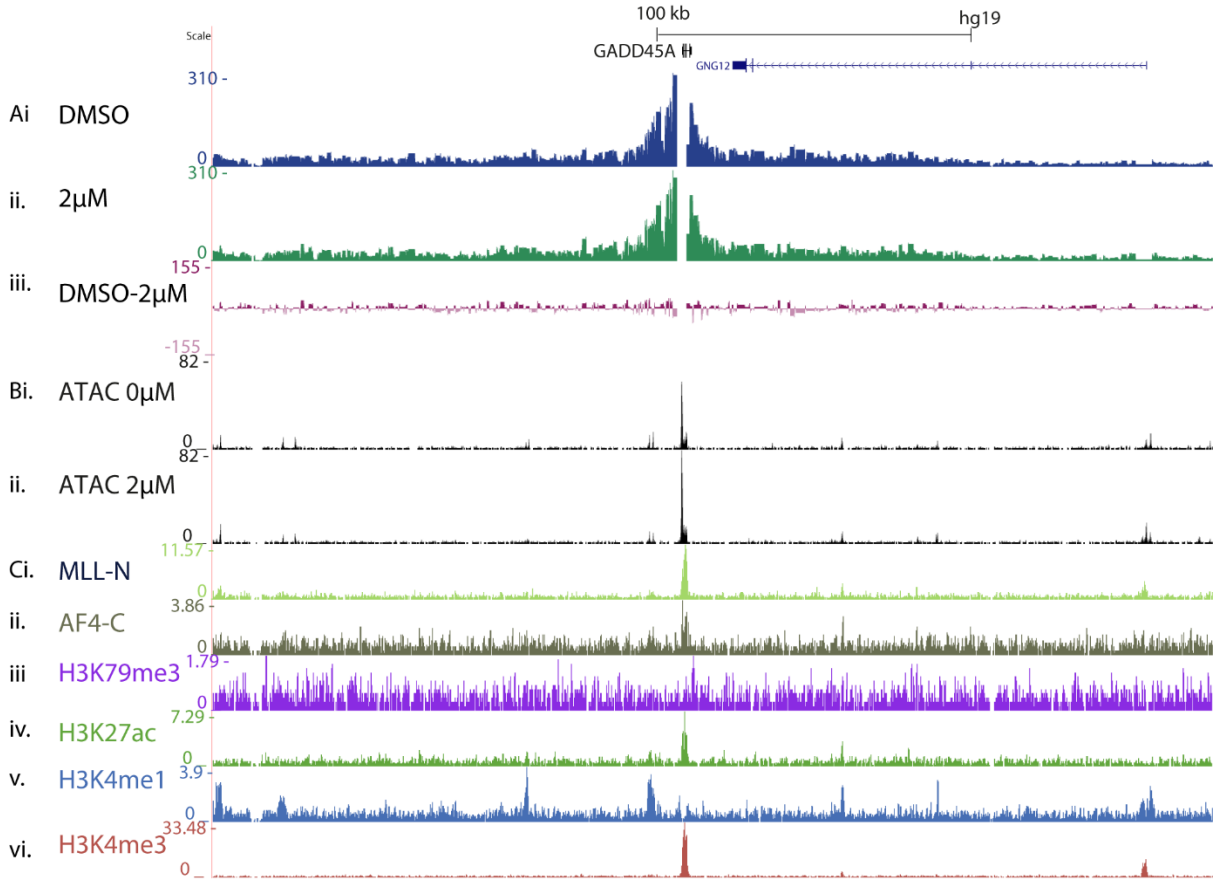

Figure S15. High throughput analysis of chromatin features at *GADD45A* (A) Capture-C from *GADD45A* promoter following 7 days 0μM and 2μM EPZ-5676 treatment of SEM cells. Pink tracks represent subtraction of 2μM from 0μM treatment. All tracks represent average of three biological replicates. (B) ATAC seq at *GADD45A* following 7 days 0μM and 2μM EPZ-5676 treatment of SEM cells, representation of two biological replicates (C) ChIP seq tracks at *GADD45A* of relevant chromatin features including MLL-N, AF4-C, H3K79me3, H3K27ac, H3K4me1, H3K4me3. ChIP seq generated in the lab by E.Ballabio, J.Kerry and T.Milne.

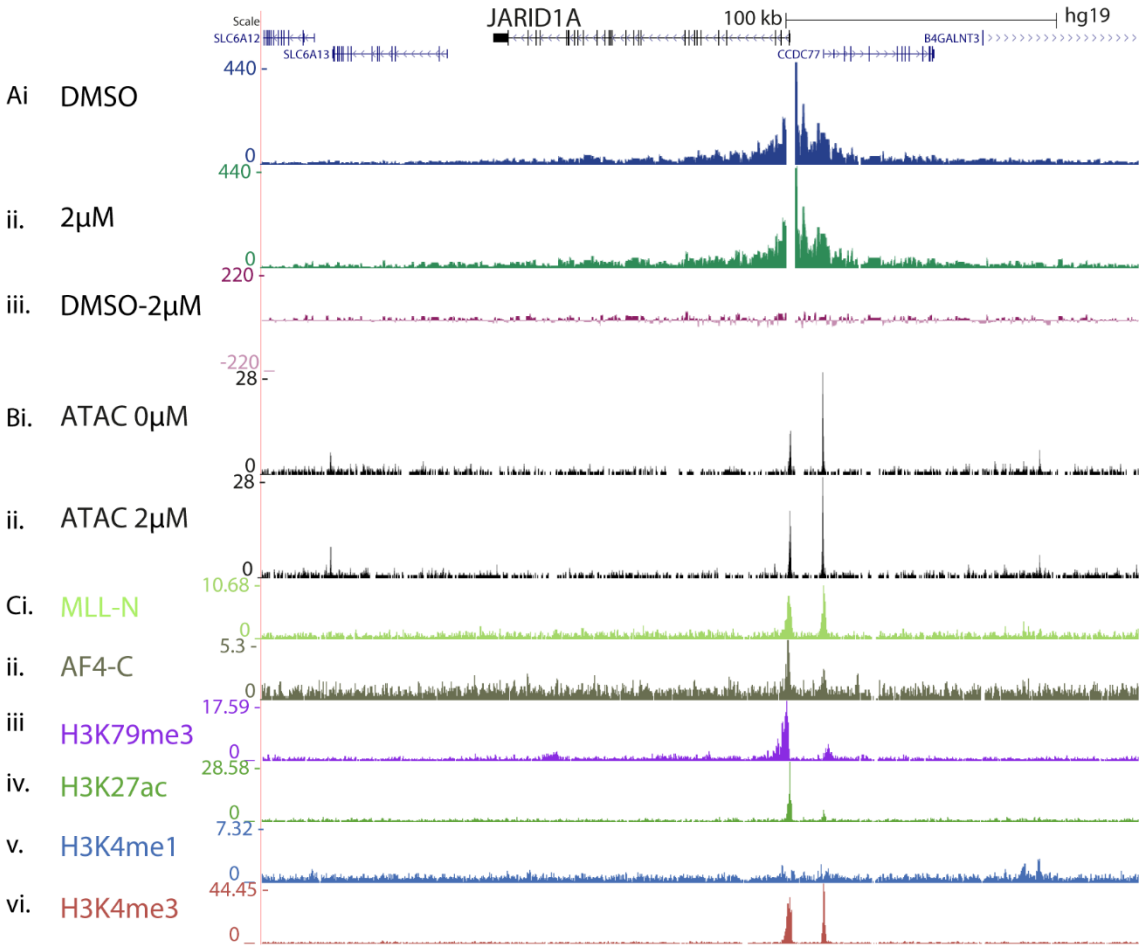

Figure S16. High throughput analysis of chromatin features at *JARID1A* (A) Capture-C from *JARID1A* promoter following 7 days 0µM and 2µM EPZ-5676 treatment of SEM cells. Pink tracks represent subtraction of 2µM from 0µM treatment. All tracks represent average of three biological replicates. (B) ATAC seq at *JARID1A* following 7 days 0µM and 2µM EPZ-5676 treatment of SEM cells, representation of two biological replicates (C) ChIP seq tracks at *JARID1A* of relevant chromatin features including MLL-N, AF4-C, H3K79me3, H3K27ac, H3K4me1, H3K4me3. ChIP seq generated in the lab by E.Ballabio, J.Kerry and T.Milne.

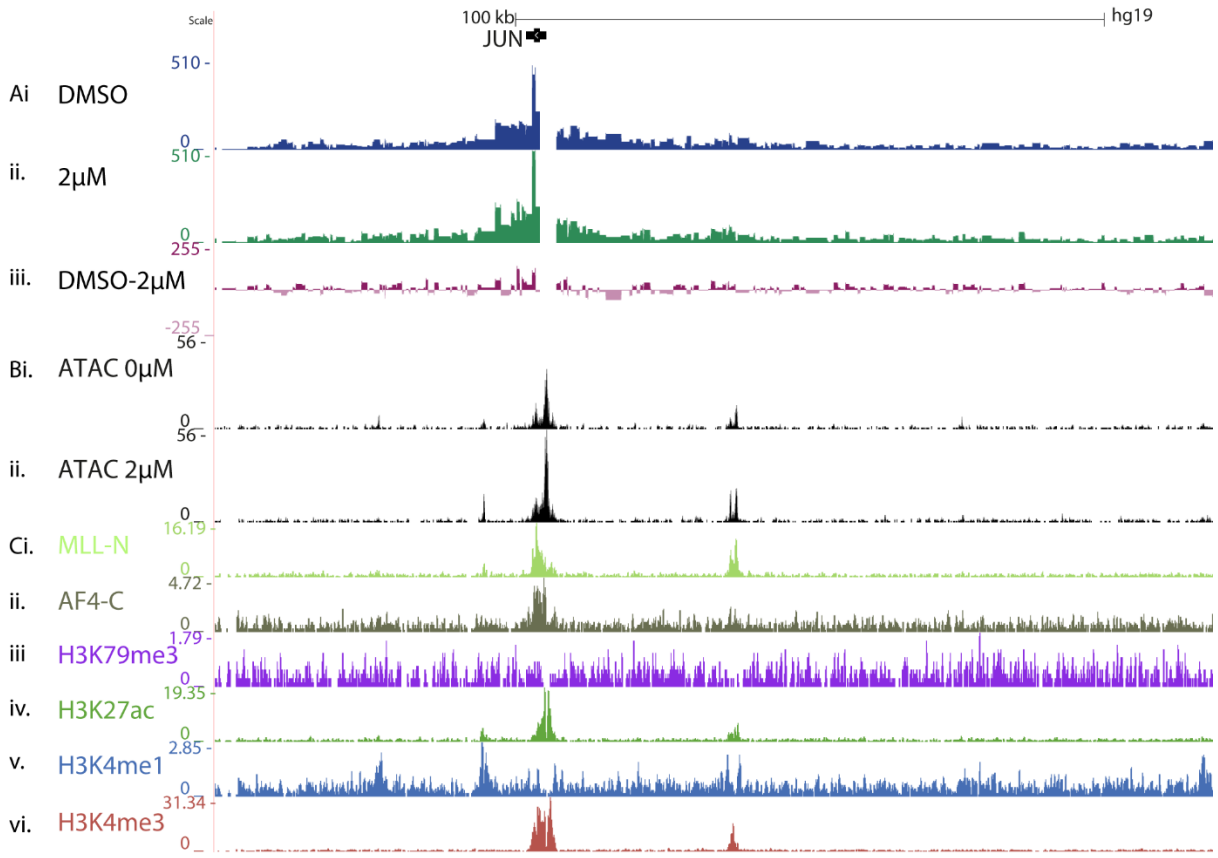

Figure S17. High throughput analysis of chromatin features at *JUN* (A) Capture-C from *JUN* promoter following 7 days 0μM and 2μM EPZ-5676 treatment of SEM cells. Pink tracks represent subtraction of 2μM from 0μM treatment. All tracks represent average of three biological replicates. (B) ATAC seq at *JUN* following 7 days 0μM and 2μM EPZ-5676 treatment of SEM cells, representation of two biological replicates (C) ChIP seq tracks at *JUN* of relevant chromatin features including MLL-N, AF4-C, H3K79me3, H3K27ac, H3K4me1, H3K4me3. ChIP seq generated in the lab by E.Ballabio, J.Kerry and T.Milne.

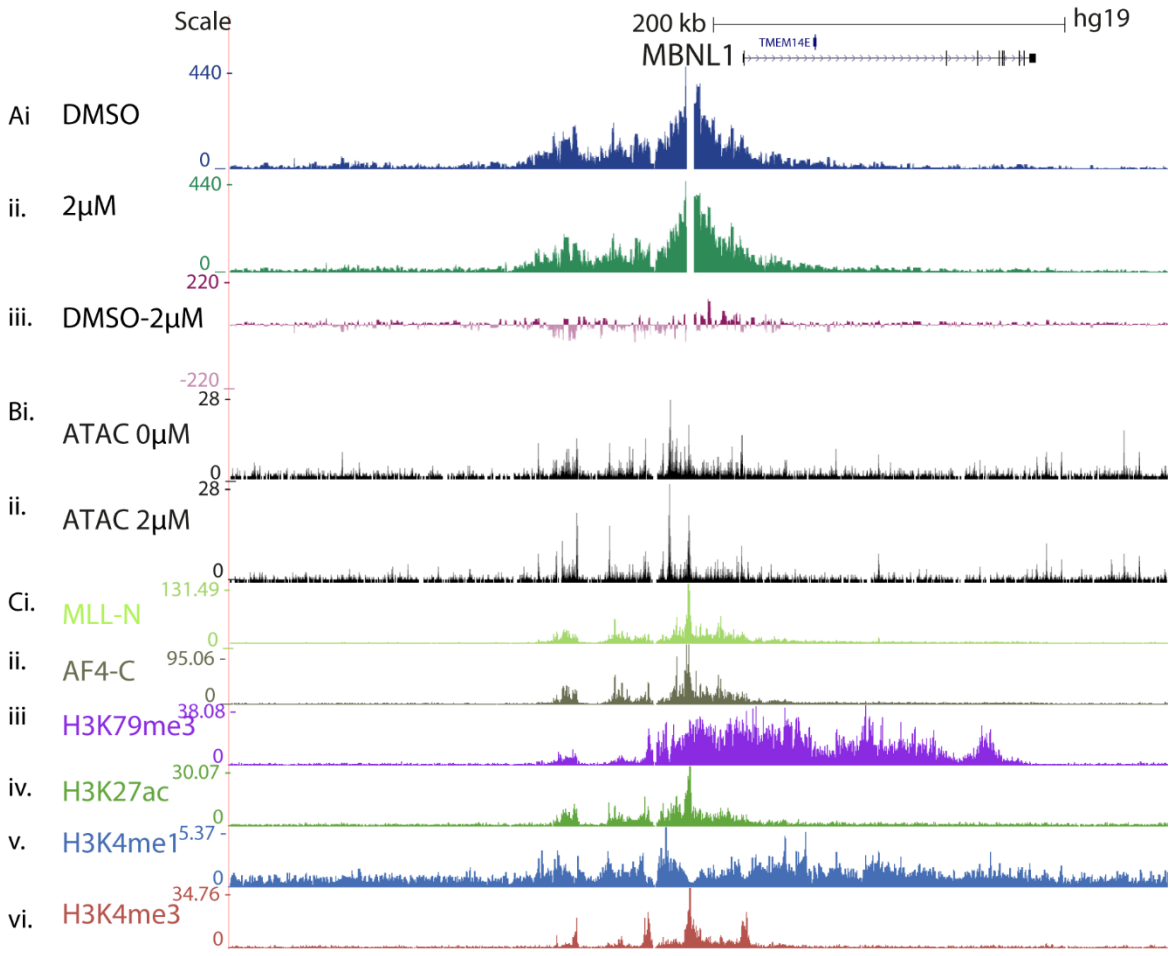

Figure S18. High throughput analysis of chromatin features at *MBNL1* (A) Capture-C from *MBNL1* promoter following 7 days 0µM and 2µM EPZ-5676 treatment of SEM cells. Pink tracks represent subtraction of 2µM from 0µM treatment. All tracks represent average of three biological replicates. (B) ATAC seq at *MBNL1* following 7 days 0µM and 2µM EPZ-5676 treatment of SEM cells, representation of two biological replicates (C) ChIP seq tracks at *MBNL1* of relevant chromatin features including MLL-N, AF4-C, H3K79me3, H3K27ac, H3K4me1, H3K4me3. ChIP seq generated in the lab by E.Ballabio, J.Kerry and T.Milne.

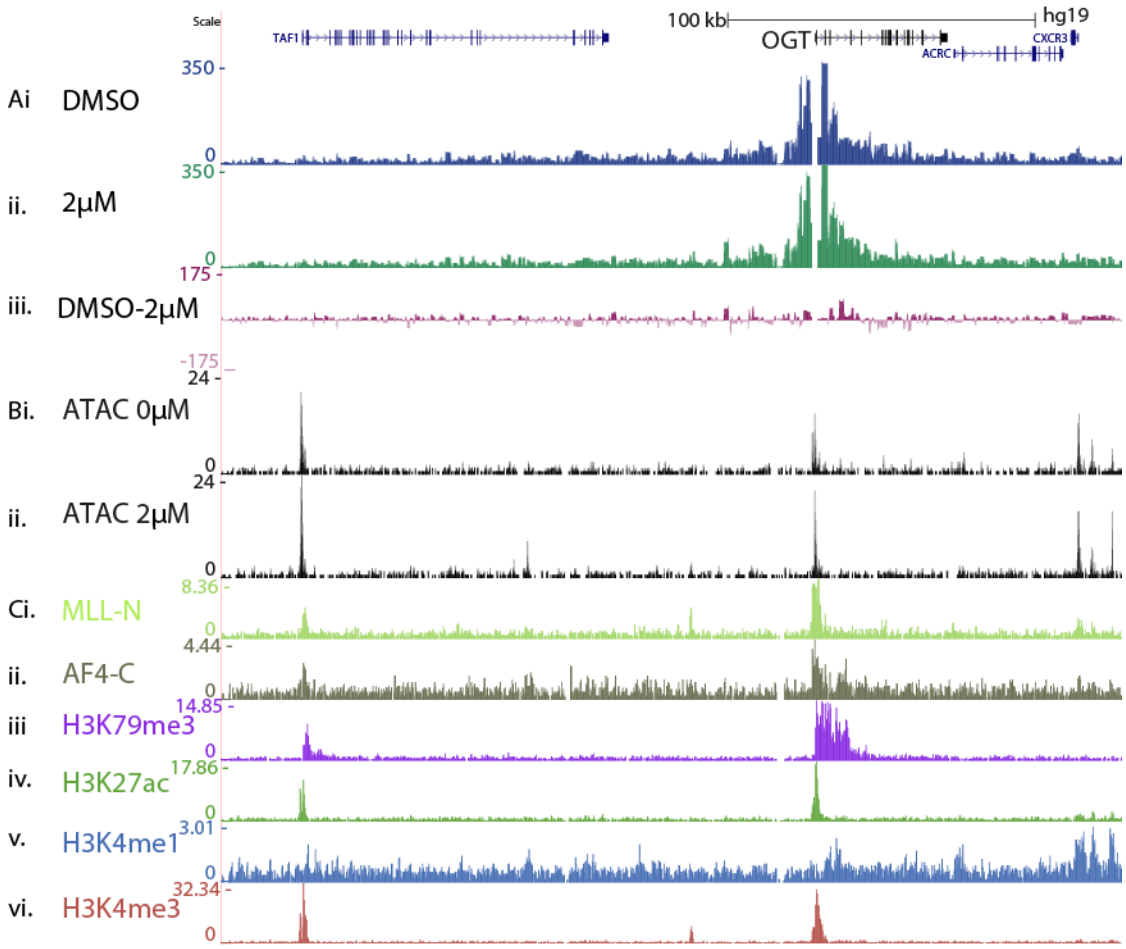

Figure S19. High throughput analysis of chromatin features at *OGT* (A) Capture-C from *OGT* promoter following 7 days 0μM and 2μM EPZ-5676 treatment of SEM cells. Pink tracks represent subtraction of 2μM from 0μM treatment. All tracks represent average of three biological replicates. (B) ATAC seq at *OGT* following 7 days 0μM and 2μM EPZ-5676 treatment of SEM cells, representation of two biological replicates (C) ChIP seq tracks at *OGT* of relevant chromatin features including MLL-N, AF4-C, H3K79me3, H3K27ac, H3K4me1, H3K4me3. ChIP-seq generated in the lab by E.Ballabio, J.Kerry and T.Milne.

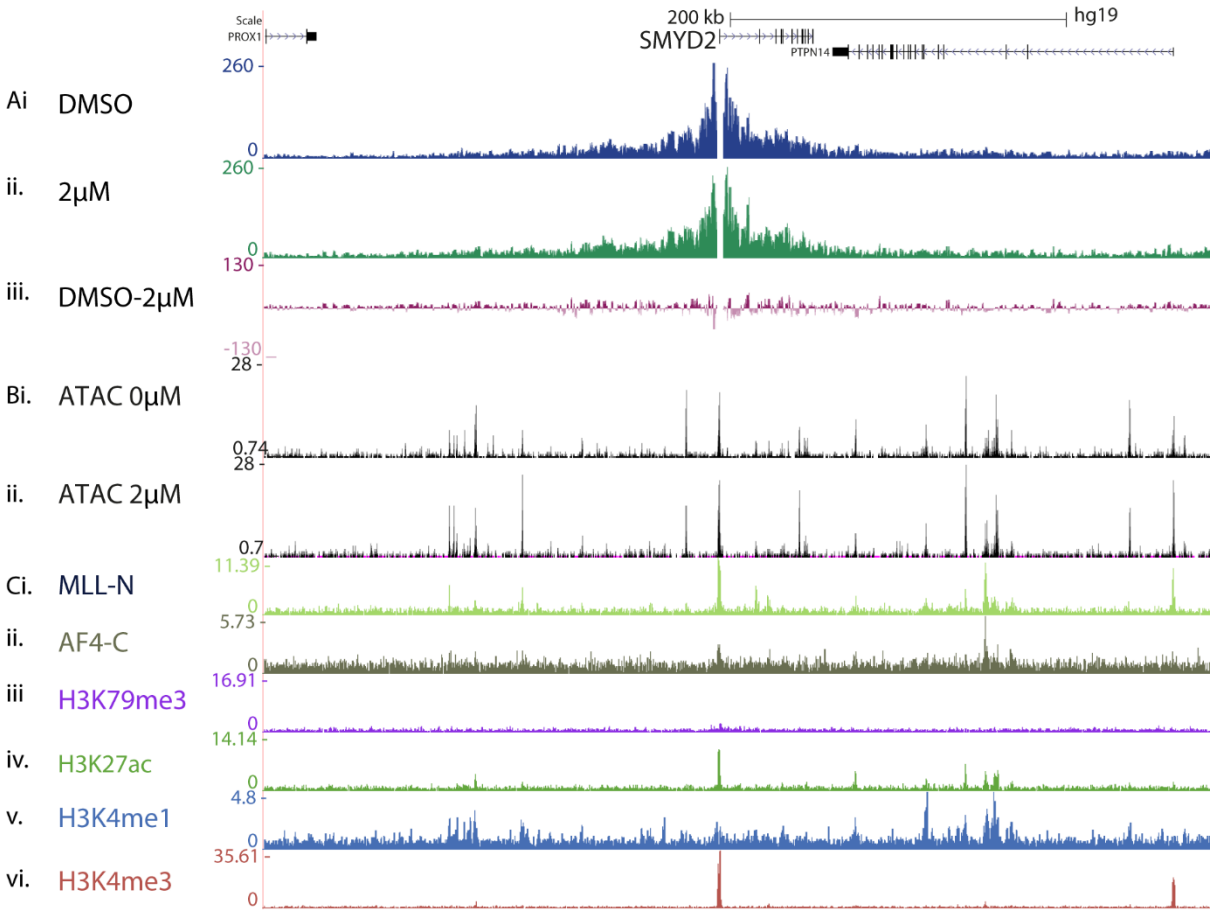

Figure S20. High throughput analysis of chromatin features at *SMYD2* (A) Capture-C from S*MDY2* promoter following 7 days 0μM and 2μM EPZ-5676 treatment of SEM cells. Pink tracks represent subtraction of 2μM from 0μM treatment. All tracks represent average of three biological replicates. (B) ATAC seq at *SMDY2* following 7 days 0μM and 2μM EPZ-5676 treatment of SEM cells, representation of two biological replicates (C) ChIP seq tracks at *SMYD2* of relevant chromatin features including MLL-N, AF4-C, H3K79me3, H3K27ac, H3K4me1, H3K4me3. ChIP seq generated in the lab by E.Ballabio, J.Kerry and T.Milne.

References

Adam, R.C. et al., 2015. Pioneer factors govern super-enhancer dynamics in stem cell plasticity and lineage choice. *Nature*, 521(7552), pp.366–370. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25799994 [Accessed September 18, 2017].

Altaf, M. et al., 2007. Interplay of chromatin modifiers on a short basic patch of histone H4 tail defines the boundary of telomeric heterochromatin. *Molecular cell*, 28(6), pp.1002–14. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2610362&tool=pmcentrez&rendertype=abstract [Accessed December 25, 2014].

Amano, T. et al., 2009. Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. *Developmental cell*, 16(1), pp.47–57. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19097946 [Accessed August 6, 2017].

Andersson, A.K. et al., 2015. The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias. *Nature genetics*, 47(4), pp.330–7. Available at: http://www.nature.com/doifinder/10.1038/ng.3230 [Accessed August 22, 2017].

Anguita, E. et al., 2002. Deletion of the mouse α-globin regulatory element (HS −26) has an unexpectedly mild phenotype. *Blood*, 100(10). Available at: http://www.bloodjournal.org/content/100/10/3450.long?sso-checked=true [Accessed August 21, 2017].

Armstrong, S.A. et al., 2004. FLT3 mutations in childhood acute lymphoblastic leukemia. *Blood*, 103(9), pp.3544–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14670924 [Accessed September 25, 2017].

Armstrong, S.A. et al., 2002. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nature genetics*, 30(1), pp.41–7. Available at: http://www.nature.com/doifinder/10.1038/ng765 [Accessed September 18, 2017].

Arnosti, D.N. et al., 1996. The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. *Development*, 122(1).

Bach, C. et al., 2010. Leukemogenic transformation by HOXA cluster genes. *Blood*, 115(14), pp.2910–8. Available at: http://www.bloodjournal.org/content/115/14/2910.abstract [Accessed September 18, 2014].

Baek, H.J., Kang, Y.K. & Roeder, R.G., 2006. Human Mediator enhances basal transcription by facilitating recruitment of transcription factor IIB during preinitiation complex assembly. *The Journal of biological chemistry*, 281(22), pp.15172–81. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16595664 [Accessed August 30, 2017].

Ball, M.P. et al., 2009. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nature biotechnology*, 27(4), pp.361–8. Available at: http://www.nature.com/doifinder/10.1038/nbt.1533 [Accessed September 15, 2017].

Banerji, J., Rusconi, S. & Schaffner, W., 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. *Cell*, 27(2), pp.299–308. Available at: http://linkinghub.elsevier.com/retrieve/pii/009286748190413X [Accessed August 18, 2017].

Barrett, N.A. et al., 2016. Mll-AF4 Confers Enhanced Self-Renewal and Lymphoid Potential during a Restricted Window in Development. *Cell Reports*, 16(4), pp.1039–1054. Available at: http://linkinghub.elsevier.com/retrieve/pii/S2211124716308002 [Accessed August 15, 2017].

Barry, E., 2010. Targeting DOT1L action and interactions in leukemia: the role of DOT1L in transformation and development. *Expert opinion on …*, 14(4), pp.405–418. Available at: http://informahealthcare.com/doi/abs/10.1517/14728221003623241 [Accessed February 13, 2014].

Bauer, M. & Metzler, R., 2012. Generalized facilitated diffusion model for DNA-binding proteins with search and recognition states. *Biophysical journal*, 102(10), pp.2321–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22677385 [Accessed September 18, 2017].

Beagan, J.A. et al., 2016. Local Genome Topology Can Exhibit an Incompletely Rewired 3D-Folding State during Somatic Cell Reprogramming. *Cell Stem Cell*, 18(5), pp.611–624. Available at: http://www.sciencedirect.com/science/article/pii/S1934590916300479 [Accessed September 15, 2017].

Bedford, D.C. et al., 2010. Target gene context influences the transcriptional requirement for the KAT3 family of CBP and p300 histone acetyltransferases. *Epigenetics*, 5(1), pp.9–15. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20110770 [Accessed September 18, 2017].

Benito, J.M. et al., 2015. MLL-Rearranged Acute Lymphoblastic Leukemias Activate BCL-2 through H3K79 Methylation and Are Sensitive to the BCL-2-Specific Antagonist ABT-199. *Cell reports*, 13(12), pp.2715–2727. Available at: http://www.cell.com/article/S2211124715014151/fulltext [Accessed December 30, 2015].

Benoist, C. & Chambon, P., 1981. In vivo sequence requirements of the SV40 early promotor region. *Nature*, 290(5804), pp.304–10. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6259538 [Accessed September 18, 2017].

Bernt, K.M., Zhu, N., Sinha, A.U., Vempati, S., Faber, J., Krivtsov, A. V, Feng, Z., Punt, N., Daigle, A., Bullinger, L., Pollock, R.M., Richon, V.M., Kung, A.L. & Armstrong, S.A., 2011. MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L. *Cancer cell*, 20(1), pp.66–78. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21741597 [Accessed August 2, 2017].

Bernt, K.M. & Armstrong, S. a, 2011. A role for DOT1L in MLL-rearranged leukemias. *Epigenomics*, 3(6), pp.667–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22126283 [Accessed February 13, 2014].

Bertolino, E. & Singh, H., 2002. POU/TBP cooperativity: a mechanism for enhancer action from a distance. *Molecular cell*, 10(2), pp.397–407. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12191484 [Accessed September 18, 2017].

Biddie, S.C. et al., 2011. Transcription Factor AP1 Potentiates Chromatin Accessibility and Glucocorticoid Receptor Binding. *Molecular Cell*, 43(1), pp.145–155. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1097276511004606 [Accessed August 24, 2017].

Bilodeau, S. et al., 2009. SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes & development*, 23(21), pp.2484–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19884255 [Accessed August 13, 2017].

Bird, A. et al., 1985. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell*, 40(1), pp.91–99. Available at: http://linkinghub.elsevier.com/retrieve/pii/0092867485903125 [Accessed August 23, 2017].

Birke, M. et al., 2002. The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. *Nucleic acids research*, 30(4), pp.958–65. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11842107 [Accessed August 21, 2017].

Biswas, D. et al., 2011. Function of leukemogenic mixed lineage leukemia 1 (MLL) fusion proteins through distinct partner protein complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 108(38), pp.15751–6. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3179097&tool=pmcentrez&rendertype=abstract [Accessed January 28, 2014].

Blackledge, N.P. et al., 2010. CpG Islands Recruit a Histone H3 Lysine 36 Demethylase. *Molecular Cell*, 38(2), pp.179–190. Available at: http://www.sciencedirect.com/science/article/pii/S1097276510002868 [Accessed August 23, 2017].

Blackledge, N.P. et al., 2014. Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation. *Cell*, 157(6), pp.1445–59. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4048464&tool=pmcentrez&rendertype=abstract [Accessed July 15, 2014].

Bonn, S. et al., 2012. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nature genetics*, 44(2), pp.148–56. Available at: http://dx.doi.org/10.1038/ng.1064 [Accessed January 31, 2014].

Botbol, Y. et al., 2007. Chromatinized templates reveal the requirement for the LEDGF/p75 PWWP domain during HIV-1 integration in vitro. *Nucleic Acids Research*, 36(4), pp.1237–1246. Available at: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkm1127 [Accessed August 18, 2017].

Botuyan, M.V. et al., 2006. Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. *Cell*, 127(7), pp.1361–73. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1804291&tool=pmcentrez&rendertype=abstract [Accessed December 21, 2014].

Boyer, L.A. et al., 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature, Published online: 19 April 2006; | doi:10.1038/nature04733*, 441(7091), p.349. Available at: https://www.nature.com/nature/journal/v441/n7091/full/nature04733.html [Accessed September 18,

2017].

Brenner, C. et al., 2005. Myc represses transcription through recruitment of DNA methyltransferase corepressor. *The EMBO journal*, 24(2), pp.336–46. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15616584 [Accessed September 22, 2017].

Briggs, S.D. et al., 2002. Gene silencing: Trans-histone regulatory pathway in chromatin. *Nature*, 418(6897), pp.498–498. Available at: http://www.nature.com/doifinder/10.1038/nature00970 [Accessed August 24, 2017].

Buenrostro, J.D. et al., 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods*, 10(12), pp.1213–1218. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24097267 [Accessed October 4, 2017].

Bursen, A. et al., 2010. The AF4·MLL fusion protein is capable of inducing ALL in mice without requirement of MLL·AF4. *Blood*, 115(17). Available at: http://www.bloodjournal.org/content/115/17/3570.long?sso-checked=true [Accessed August 15, 2017].

Campbell, C.T. et al., 2017. Mechanisms of Pinometostat (EPZ-5676) Treatment–Emergent Resistance in MLL-Rearranged Leukemia. *Molecular Cancer Therapeutics*, 16(8). Available at: http://mct.aacrjournals.org/content/16/8/1669.long [Accessed August 16, 2017].

Cao, R. et al., 2002. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science (New York, N.Y.)*, 298(5595), pp.1039–43. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12351676 [Accessed September 18, 2017].

Cartailler, J. & Reingruber, J., 2015. Facilitated diffusion framework for transcription factor search with conformational changes. *Physical Biology*, 12(4), p.46012. Available at: http://stacks.iop.org/1478-3975/12/i=4/a=046012?key=crossref.d7313c31860e10e17cbb7f0f8c56c394 [Accessed September 18, 2017].

Caslini, C. et al., 2007. Interaction of MLL Amino Terminal Sequences with Menin Is Required for Transformation. *Cancer Research*, 67(15). Available at: http://cancerres.aacrjournals.org/content/67/15/7275.long [Accessed August 18, 2017].

Chandra, T. et al., 2015. Global reorganization of the nuclear landscape in senescent cells. *Cell reports*, 10(4), pp.471–83. Available at: http://linkinghub.elsevier.com/retrieve/pii/S221112471401122X [Accessed September 15, 2017].

Chang, M.-J. et al., 2010. Histone H3 lysine 79 methyltransferase Dot1 is required for immortalization by MLL oncogenes. *Cancer research*, 70(24), pp.10234–10242. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3040779&tool=pmcentrez&rendertype=abstract [Accessed February 2, 2014].

Chang, P.-Y. et al., 2010. Binding of the MLL PHD3 finger to histone H3K4me3 is required for MLL-dependent gene transcription. *Journal of molecular biology*, 400(2), pp.137–44. Available at:

http://www.sciencedirect.com/science/article/pii/S0022283610004717 [Accessed January 24, 2014].

Chen, C.-W. et al., 2015. DOT1L inhibits SIRT1-mediated epigenetic silencing to maintain leukemic gene expression in MLL-rearranged leukemia. *Nature medicine*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25822366 [Accessed April 1, 2015].

Chen, L. et al., 2012. Abrogation of MLL–AF10 and CALM–AF10-mediated transformation through genetic inactivation or pharmacological inhibition of the H3K79 methyltransferase Dot1l. *Leukemia*, 27(4), pp.813–822. Available at: http://dx.doi.org/10.1038/leu.2012.327 [Accessed October 7, 2015].

Chen, W. et al., 2006. A murine Mll-AF4 knock-in model results in lymphoid and myeloid deregulation and hematologic malignancy. *Blood*, 108(2). Available at: http://www.bloodjournal.org/content/108/2/669.long?sso-checked=true [Accessed August 15, 2017].

Chen, Y.-X. et al., 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(4), pp.1018–23. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16415155 [Accessed August 18, 2017].

Chillón, M.C. et al., 2012. Prognostic significance of FLT3 mutational status and expression levels in MLL-AF4+ and MLL-germline acute lymphoblastic leukemia. *Leukemia*, 26(11), pp.2360–6. Available at: http://www.nature.com/doifinder/10.1038/leu.2012.161 [Accessed September 25, 2017].

Cho, M.-H. et al., 2015. DOT1L cooperates with the c-Myc-p300 complex to epigenetically derepress CDH1 transcription factors in breast cancer progression. *Nature communications*, 6, p.7821. Available at: http://www.nature.com/doifinder/10.1038/ncomms8821 [Accessed July 31, 2017].

Chodaparambil, J. V et al., 2007. A charged and contoured surface on the nucleosome regulates chromatin compaction. *Nature Structural & Molecular Biology*, 14(11), pp.1105–1107. Available at: http://www.nature.com/doifinder/10.1038/nsmb1334 [Accessed August 2, 2017].

Cierpicki, T. et al., 2010. Structure of the MLL CXXC domain-DNA complex and its functional role in MLL-AF9 leukemia. *Nature structural & molecular biology*, 17(1), pp.62–8. Available at: http://www.nature.com/doifinder/10.1038/nsmb.1714 [Accessed August 21, 2017].

Collins, R.E. et al., 2008. The ankyrin repeats of G9a and GLP histone methyltransferases are mono- and dimethyllysine binding modules. *Nature structural & molecular biology*, 15(3), pp.245–50. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18264113 [Accessed August 21, 2017].

Comet, I. et al., 2016. Maintaining cell identity: PRC2-mediated regulation of transcription and cancer. *Nature reviews. Cancer*, 16(12), pp.803–810. Available at: http://www.nature.com/doifinder/10.1038/nrc.2016.83 [Accessed September 28, 2017].

Coskun, E. et al., 2011. The role of microRNA-196a and microRNA-196b as ERG regulators in acute myeloid leukemia and acute T-lymphoblastic leukemia. *Leukemia Research*, 35(2), pp.208–213. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0145212610002511 [Accessed August 4,

2017].

Creyghton, M.P. et al., 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences*, 107(50), pp.21931–21936. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21106759 [Accessed August 7, 2017].

Dafflon, C. et al., 2017. Complementary activities of DOT1L and Menin inhibitors in MLL-rearranged leukemia. *Leukemia*, 31(6), pp.1269–1277. Available at: http://www.nature.com/doifinder/10.1038/leu.2016.327 [Accessed August 18, 2017].

Daigle, S.R. et al., 2013. Potent inhibition of DOT1L as treatment of MLL-fusion leukemia. *Blood*, 122(6), pp.1017–25. Available at: http://www.bloodjournal.org/content/122/6/1017.abstract [Accessed January 28, 2014].

Daigle, S.R. et al., 2011. Selective killing of mixed lineage leukemia cells by a potent small-molecule DOT1L inhibitor. *Cancer cell*, 20(1), pp.53–65. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21741596 [Accessed January 28, 2014].

Davies, J.O.J. et al., 2016. Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *doi.org*, 13(1), pp.74–80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26595209 [Accessed August 7, 2017].

Dawson, M.A. et al., 2011. Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. *Nature*, 478(7370), pp.529–33. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21964340 [Accessed August 2, 2017].

Deng, W. et al., 2012. Controlling Long-Range Genomic Interactions at a Native Locus by Targeted Tethering of a Looping Factor. *Cell*, 149(6), pp.1233–1244. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0092867412005880 [Accessed August 18, 2017].

Deng, W. & Roberts, S.G.E., 2005. A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes & development*, 19(20), pp.2418–23. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16230532 [Accessed September 18, 2017].

Deshpande, A.J. et al., 2014. AF10 Regulates Progressive H3K79 Methylation and HOX Gene Expression in Diverse AML Subtypes. *Cancer Cell*, 26(6), pp.896–908. Available at: http://www.sciencedirect.com/science/article/pii/S1535610814004139 [Accessed November 21, 2014].

Deshpande, A.J. et al., 2013. Leukemic transformation by the MLL-AF6 fusion oncogene requires the H3K79 methyltransferase Dot1l. *Blood*, 121(13). Available at: http://www.bloodjournal.org/content/121/13/2533.long?sso-checked=true [Accessed July 31, 2017].

Deshpande, A.J. et al., 2013. Leukemic transformation by the MLL-AF6 fusion oncogene requires the H3K79 methyltransferase Dot1l. *Blood*, 121(13), pp.2533–41. Available at: http://bloodjournal.org/content/121/13/2533.abstract [Accessed May 25, 2014].

Deshpande, A.J. et al., 2011. The clathrin-binding domain of CALM and the OM-LZ domain of AF10 are sufficient to induce acute myeloid leukemia in mice. *Leukemia*, 25(11), pp.1718–27. Available at: http://dx.doi.org/10.1038/leu.2011.153 [Accessed October 7, 2015].

Dixon, J.R. et al., 2015. Chromatin architecture reorganization during stem cell differentiation. *Nature*, 518(7539), pp.331–336. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25693564 [Accessed September 15, 2017].

Dixon, J.R. et al., 2012a. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398), pp.376–80. Available at: http://www.nature.com/doifinder/10.1038/nature11082 [Accessed August 18, 2017].

Djabali, M. et al., 1992. A trithorax–like gene is interrupted by chromosome 11q23 translocations in acute leukaemias. *Nature Genetics*, 2(2), pp.113–118. Available at: http://www.nature.com/doifinder/10.1038/ng1092-113 [Accessed August 15, 2017].

Dorighi, K.M. et al., 2017. *Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation*, Available at: http://www.sciencedirect.com/science/article/pii/S1097276517302745 [Accessed August 22, 2017].

Dover, J. et al., 2002. Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6. *The Journal of biological chemistry*, 277(32), pp.28368–71. Available at: http://www.jbc.org/lookup/doi/10.1074/jbc.C200348200 [Accessed August 24, 2017].

Dowen, J.M. et al., 2014. Control of Cell Identity Genes Occurs in Insulated Neighborhoods in Mammalian Chromosomes. *Cell*, 159(2), pp.374–387. Available at: http://www.sciencedirect.com/science/article/pii/S0092867414011799 [Accessed September 18, 2017].

Drissen, R. et al., 2004. The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes & development*, 18(20), pp.2485–90. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15489291 [Accessed July 17, 2017].

Ehrlich, M. et al., 1982. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic acids research*, 10(8), pp.2709–21. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7079182 [Accessed August 30, 2017].

Elf, J., Li, G.-W. & Xie, X.S., 2007. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science (New York, N.Y.)*, 316(5828), pp.1191–4. Available at: http://www.sciencemag.org/content/316/5828/1191.long [Accessed August 27, 2015].

Erb, M.A. et al., 2017. Transcription control by the ENL YEATS domain in acute leukaemia. *Nature*, 543(7644), pp.270–274. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28241139 [Accessed August 7, 2017].

Esnault, C. et al., 2008. Mediator-Dependent Recruitment of TFIIH Modules in Preinitiation Complex. *Molecular Cell*, 31(3), pp.337–346. Available at:

http://www.sciencedirect.com/science/article/pii/S109727650800470X [Accessed August 30, 2017].

Espada, J. et al., 2004. Human DNA methyltransferase 1 is required for maintenance of the histone H3 modification pattern. *The Journal of biological chemistry*, 279(35), pp.37175–84. Available at: http://www.jbc.org/lookup/doi/10.1074/jbc.M404842200 [Accessed September 28, 2017].

Farcas, A.M. et al., 2012. KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *eLife*, 1, pp.479–489–489. Available at: http://elifesciences.org/lookup/doi/10.7554/eLife.00205 [Accessed August 25, 2017].

Feaver, W.J. et al., 1994. Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIH/TFIIK. *Cell*, 79(6), pp.1103–1109. Available at: http://linkinghub.elsevier.com/retrieve/pii/009286749490040X [Accessed July 31, 2017].

Feng, Q. et al., 2002. Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Current biology : CB*, 12(12), pp.1052–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12123582 [Accessed December 17, 2014].

Feng, Y. et al., 2010. Early mammalian erythropoiesis requires the Dot1L methyltransferase. *Blood*, 116(22), pp.4483–91. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3321834&tool=pmcentrez&rendertype=abstract [Accessed January 5, 2015].

Fingerman, I.M., Li, H.-C. & Briggs, S.D., 2007. A charge-based interaction between histone H4 and Dot1 is required for H3K79 methylation and telomere silencing: identification of a new trans-histone pathway. *Genes & development*, 21(16), pp.2018–29. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1948857&tool=pmcentrez&rendertype=abstract [Accessed January 8, 2015].

Fishburn, J. et al., 2015. Double-stranded DNA translocase activity of transcription factor TFIIH and the mechanism of RNA polymerase II open complex formation. *Proceedings of the National Academy of Sciences of the United States of America*, 112(13), pp.3961–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25775526 [Accessed August 30, 2017].

Fitzgerald, D.P. & Bender, W., 2001. Polycomb group repression reduces DNA accessibility. *Molecular and cellular biology*, 21(19), pp.6585–97. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11533246 [Accessed August 20, 2017].

Flavahan, W.A. et al., 2016. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature*, 529(7584), pp.110–4. Available at: http://www.nature.com/doifinder/10.1038/nature16490 [Accessed August 21, 2017].

Flores, O., Lu, H. & Reinberg, D., 1992. Factors involved in specific transcription by mammalian RNA polymerase II. Identification and characterization of factor IIH. *The Journal of biological chemistry*, 267(4), pp.2786–93. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1733973 [Accessed August 24, 2017].

Francis, N.J. et al., 2001. Reconstitution of a functional core polycomb repressive complex. *Molecular cell*, 8(3), pp.545–56. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11583617 [Accessed September 27, 2017].

Franke, M. et al., 2016. Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature*, 538(7624), pp.265–269. Available at: http://www.nature.com/doifinder/10.1038/nature19800 [Accessed August 21, 2017].

Frederiks, F. et al., 2010. Heterologous expression reveals distinct enzymatic activities of two DOT1 histone methyltransferases of Trypanosoma brucei. *Journal of cell science*, 123(Pt 23), pp.4019–23. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21084562 [Accessed January 6, 2015].

Frederiks, F. et al., 2008a. Nonprocessive methylation by Dot1 leads to functional redundancy of histone H3K79 methylation states. *Nature Structural & Molecular Biology*, 15(6), pp.550–557. Available at: http://www.nature.com/doifinder/10.1038/nsmb.1432 [Accessed July 20, 2017].

Fujinaga, K. et al., 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(2), pp.787–95. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14701750 [Accessed August 24, 2017].

Fujita, N. et al., 2003. Methyl-CpG binding domain 1 (MBD1) interacts with the Suv39h1-HP1 heterochromatic complex for DNA methylation-based transcriptional repression. *The Journal of biological chemistry*, 278(26), pp.24132–8. Available at: http://www.jbc.org/lookup/doi/10.1074/jbc.M302283200 [Accessed September 28, 2017].

Fuks, F. et al., 2003. The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic acids research*, 31(9), pp.2305–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12711675 [Accessed August 21, 2017].

Garcia-Cuellar, M.-P. et al., 2016. Leukemogenic MLL-ENL Fusions Induce Alternative Chromatin States to Drive a Functionally Dichotomous Group of Target Genes. *Cell Reports*, 15(2), pp.310–322.

Garcia, B.A. et al., 2007. Organismal differences in post-translational modifications in histones H3 and H4. *The Journal of biological chemistry*, 282(10), pp.7641–55. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17194708 [Accessed August 21, 2017].

Ghirlando, R. & Felsenfeld, G., 2016. CTCF: making the right connections. *Genes & development*, 30(8), pp.881–91. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27083996 [Accessed August 21, 2017].

Gibbons, G.S. et al., 2014. Regulation of Wnt Signaling Target Gene Expression by the Histone Methyltransferase DOT1L. *ACS chemical biology*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25361163 [Accessed November 6, 2014].

Gilan, O. et al., 2016. Functional interdependence of BRD4 and DOT1L in MLL leukemia. *Nature Structural & Molecular Biology*, 23(7), pp.673–681. Available at:

http://www.nature.com/doifinder/10.1038/nsmb.3249 [Accessed July 21, 2016].

Giorgio, E. et al., 2015. A large genomic deletion leads to enhancer adoption by the lamin B1 gene: a second path to autosomal dominant adult-onset demyelinating leukodystrophy (ADLD). *Human Molecular Genetics*, 24(11), pp.3143–3154. Available at: https://academic.oup.com/hmg/article-lookup/doi/10.1093/hmg/ddv065 [Accessed September 18, 2017].

Glover-Cutter, K. et al., 2009. TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II. *Molecular and cellular biology*, 29(20), pp.5455–64. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19667075 [Accessed July 31, 2017].

Godfrey, L. et al., 2016. MLL-AF4 binds directly to a BCL-2 specific enhancer and modulates H3K27 acetylation. *Experimental hematology*, 0(0). Available at: http://www.ncbi.nlm.nih.gov/pubmed/27856324 [Accessed February 15, 2017].

Gordon, M., 1993. Human haemopoietic stem cell assays. *Blood Reviews*, 7(3), pp.190–197. Available at: http://linkinghub.elsevier.com/retrieve/pii/0268960X9390005O [Accessed September 13, 2017].

Gu, Y. et al., 1992. The t(4;11) chromosome translocation of human acute leukemias fuses the ALL-1 gene, related to Drosophila trithorax, to the AF-4 gene. *Cell*, 71(4), pp.701–708. Available at: http://linkinghub.elsevier.com/retrieve/pii/009286749290603A [Accessed August 15, 2017].

Guenther, M.G. et al., 2008a. Aberrant chromatin at genes encoding stem cell regulators in human mixed-lineage leukemia. *Genes & development*, 22(24), pp.3403–8. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2607073&tool=pmcentrez&rendertype=abstract [Accessed March 10, 2015].

Guertin, M.J. & Lis, J.T., 2010. Chromatin landscape dictates HSF binding to target DNA elements. M. Lichten, ed. *PLoS genetics*, 6(9), p.e1001114. Available at: http://dx.plos.org/10.1371/journal.pgen.1001114 [Accessed August 24, 2017].

Hadjur, S. et al., 2009. Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature*, 460(7253), pp.410–3. Available at: http://www.nature.com/doifinder/10.1038/nature08079 [Accessed September 15, 2017].

Hammar, P. et al., 2014. Direct measurement of transcription factor dissociation excludes a simple operator occupancy model for gene regulation. *Nature Genetics*, 46(4), pp.405–408. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24562187 [Accessed September 7, 2017].

Hammar, P. et al., 2012. The lac repressor displays facilitated diffusion in living cells. *Science (New York, N.Y.)*, 336(6088), pp.1595–8. Available at: http://www.sciencemag.org/cgi/doi/10.1126/science.1221648 [Accessed August 24, 2017].

Handoko, L. et al., 2011. CTCF-mediated functional chromatin interactome in pluripotent cells. *Nature genetics*, 43(7), pp.630–8. Available at: http://www.nature.com/doifinder/10.1038/ng.857 [Accessed

September 15, 2017].

Hanssen, L.L.P. et al., 2017. Tissue-specific CTCF–cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. *Nature Cell Biology*, 19(8), pp.952–961. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28737770 [Accessed August 13, 2017].

Hay, D. et al., 2016. Genetic dissection of the α-globin super-enhancer in vivo. *Nature genetics*, 48(8), pp.895–903. Available at: http://www.nature.com/doifinder/10.1038/ng.3605 [Accessed August 21, 2017].

He, J. et al., 2013. Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. *Nature cell biology*, 15(4), pp.373–84. Available at: http://www.nature.com/doifinder/10.1038/ncb2702 [Accessed August 25, 2017].

He, N. et al., 2011. Human Polymerase-Associated Factor complex (PAFc) connects the Super Elongation Complex (SEC) to RNA polymerase II on chromatin. *Proceedings of the National Academy of Sciences of the United States of America*, 108(36), pp.E636-45. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21873227 [Accessed August 2, 2017].

He, Y. et al., 2013. Structural visualization of key steps in human transcription initiation. *Nature*, 495(7442), pp.481–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23446344 [Accessed September 18, 2017].

Heintzman, N.D. et al., 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics*, 39(3), pp.311–318. Available at: http://www.nature.com/doifinder/10.1038/ng1966 [Accessed August 18, 2017].

Heintzman, N.D. et al., 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, 459(7243), pp.108–112. Available at: http://www.nature.com/doifinder/10.1038/nature07829 [Accessed August 7, 2017].

Heinz, S. et al., 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell*, 38(4), pp.576–89. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1097276510003667 [Accessed October 4, 2017].

Hellman, A. & Chess, A., 2007. Gene body-specific methylation on the active X chromosome. *Science (New York, N.Y.)*, 315(5815), pp.1141–3. Available at: http://www.sciencemag.org/cgi/doi/10.1126/science.1136352 [Accessed September 15, 2017].

Hengartner, C.J. et al., 1998. Temporal Regulation of RNA Polymerase II by Srb10 and Kin28 Cyclin-Dependent Kinases. *Molecular Cell*, 2(1), pp.43–53. Available at: http://www.sciencedirect.com/science/article/pii/S1097276500801124 [Accessed July 31, 2017].

Herz, H.-M. et al., 2012. Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. *Genes & development*, 26(23), pp.2604–20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23166019 [Accessed September 18, 2017].

Hnisz, D. et al., 2016. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science*, 351(6280). Available at: http://science.sciencemag.org/content/351/6280/1454/tab-pdf [Accessed August 21, 2017].

Hnisz, D. et al., 2013. Super-Enhancers in the Control of Cell Identity and Disease. *Cell*, 155(4), pp.934–947. Available at: http://www.sciencedirect.com/science/article/pii/S0092867413012270 [Accessed August 18, 2017].

Ho, L.-L. et al., 2013. DOT1L-mediated H3K79 methylation in chromatin is dispensable for Wnt pathway-specific and other intestinal epithelial functions. *Molecular and cellular biology*, 33(9), pp.1735–45. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23428873 [Accessed February 6, 2014].

Holliday, R. & Pugh, J.E., 1975. DNA modification mechanisms and gene activity during development. *Science (New York, N.Y.)*, 187(4173), pp.226–32. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1111098 [Accessed September 15, 2017].

Holmqvist, P.-H. et al., 2005. FoxA1 binding to the MMTV LTR modulates chromatin structure and transcription. *Experimental cell research*, 304(2), pp.593–603. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0014482704007207 [Accessed August 24, 2017].

Holstege, F.C., Fiedler, U. & Timmers, H.T., 1997. Three transitions in the RNA polymerase II transcription complex during initiation. *The EMBO journal*, 16(24), pp.7468–80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9405375 [Accessed September 18, 2017].

Hou, C. et al., 2012. Gene Density, Transcription, and Insulators Contribute to the Partition of the Drosophila Genome into Physical Domains. *Molecular Cell*, 48(3), pp.471–484. Available at: http://www.sciencedirect.com/science/article/pii/S1097276512007757 [Accessed September 15, 2017].

Hu, D. et al., 2013. The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. *Molecular and cellular biology*, 33(23), pp.4745–54. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24081332 [Accessed September 18, 2017].

Hughes, C.M. et al., 2004. Menin associates with a trithorax family histone methyltransferase complex and with the hoxc8 locus. *Molecular cell*, 13(4), pp.587–97. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14992727 [Accessed August 18, 2017].

Hughes, J.R. et al., 2013. High-resolution analysis of cis-acting regulatory networks at the α-globin locus. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1620). Available at: http://rstb.royalsocietypublishing.org/content/368/1620/20120361.long [Accessed August 6, 2017].

Hunger, S.P. & Mullighan, C.G., 2015. Acute Lymphoblastic Leukemia in Children D. L. Longo, ed. *New England Journal of Medicine*, 373(16), pp.1541–1552. Available at: http://www.nejm.org/doi/10.1056/NEJMra1400972 [Accessed September 18, 2017].

Huyen, Y. et al., 2004. Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. *Nature*, 432(7015), pp.406–11. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15525939

[Accessed December 22, 2014].

Ikegami, K. et al., 2007. Genome-wide and locus-specific DNA hypomethylation in G9a deficient mouse embryonic stem cells. *Genes to cells : devoted to molecular & cellular mechanisms*, 12(1), pp.1–11. Available at: http://doi.wiley.com/10.1111/j.1365-2443.2006.01029.x [Accessed September 28, 2017].

Illingworth, R.S. et al., 2010. Orphan CpG islands identify numerous conserved promoters in the mammalian genome. W. Reik, ed. *PLoS genetics*, 6(9), p.e1001134. Available at: http://dx.plos.org/10.1371/journal.pgen.1001134 [Accessed September 15, 2017].

Ingham, P.W., 1981. Trithorax: A new homoeotic mutation ofDrosophila melanogaster : II. The role oftrx (+) after embryogenesis. *Wilhelm Roux's archives of developmental biology*, 190(6), pp.365–369. Available at: http://link.springer.com/10.1007/BF00863275 [Accessed August 24, 2017].

Ingham, P.W., 1998. trithorax and the regulation of homeotic gene expression in Drosophila: a historical perspective. *The International journal of developmental biology*, 42(3), pp.423–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9654027 [Accessed August 24, 2017].

Iwafuchi-Doi, M. et al., 2016. The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. *Molecular cell*, 62(1), pp.79–91. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1097276516001799 [Accessed August 24, 2017].

Iwafuchi-Doi, M. & Zaret, K.S., 2016. Cell fate control by pioneer transcription factors. *Development (Cambridge, England)*, 143(11), pp.1833–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27246709 [Accessed September 18, 2017].

Jack, A.P.M. & Hake, S.B., 2014. Getting down to the core of histone modifications. *Chromosoma*, 123(4), pp.355–71. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24789118 [Accessed September 13, 2014].

Jang, M.K. et al., 2005. The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Molecular cell*, 19(4), pp.523–34. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16109376 [Accessed August 2, 2017].

Jing, H. et al., 2008. Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus. *Molecular cell*, 29(2), pp.232–42. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18243117 [Accessed August 7, 2017].

Jishage, M. et al., 2012. Transcriptional Regulation by Pol II(G) Involving Mediator and Competitive Interactions of Gdown1 and TFIIF with Pol II. *Molecular Cell*, 45(1), pp.51–63. Available at: http://www.sciencedirect.com/science/article/pii/S1097276511009580 [Accessed August 30, 2017].

Jo, S.Y. et al., 2011. Requirement for Dot1l in murine postnatal hematopoiesis and leukemogenesis by MLL translocation. *Blood*, 117(18), pp.4759–68. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-79955948524&partnerID=tZOtx3y1 [Accessed

February 13, 2014].

Johnson, K.M. et al., 2002. TFIID and human mediator coactivator complexes assemble cooperatively on promoter DNA. *Genes & development*, 16(14), pp.1852–63. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12130544 [Accessed August 24, 2017].

Jolma, A. et al., 2013. DNA-Binding Specificities of Human Transcription Factors. *Cell*, 152(1–2), pp.327–339. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23332764 [Accessed August 24, 2017].

Jones, B. et al., 2008. The histone H3K79 methyltransferase Dot1L is essential for mammalian development and heterochromatin structure. *PLoS genetics*, 4(9), p.e1000190. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2527135&tool=pmcentrez&rendertype=abstract [Accessed January 28, 2014].

Kadoch, C. & Crabtree, G.R., 2015. Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Science advances*, 1(5), p.e1500447. Available at: http://advances.sciencemag.org/cgi/doi/10.1126/sciadv.1500447 [Accessed September 28, 2017].

Karimi, M.M. et al., 2011. DNA Methylation and SETDB1/H3K9me3 Regulate Predominantly Distinct Sets of Genes, Retroelements, and Chimeric Transcripts in mESCs. *Cell Stem Cell*, 8(6), pp.676–687. Available at: http://www.sciencedirect.com/science/article/pii/S193459091100169X [Accessed August 21, 2017].

Kerry, J. et al., 2017a. MLL-AF4 Spreading Identifies Binding Sites that Are Distinct from Super-Enhancers and that Govern Sensitivity to DOT1L Inhibition in Leukemia. *Cell Reports*, 18(2), pp.482–495.

Kim, J. et al., 2009. RAD6-Mediated Transcription-Coupled H2B Ubiquitylation Directly Stimulates H3K4 Methylation in Human Cells. *Cell*, 137(3), pp.459–471. Available at: http://www.sciencedirect.com/science/article/pii/S0092867409002037 [Accessed August 24, 2017].

Kim, J., Guermah, M. & Roeder, R.G., 2010. The Human PAF1 Complex Acts in Chromatin Transcription Elongation Both Independently and Cooperatively with SII/TFIIS. *Cell*, 140(4), pp.491–503.

Kim, J.L., Nikolov, D.B. & Burley, S.K., 1993. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature*, 365(6446), pp.520–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8413605 [Accessed September 18, 2017].

King, H.W. & Klose, R.J., 2017. The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. *eLife*, 6. Available at: http://elifesciences.org/lookup/doi/10.7554/eLife.22631 [Accessed August 24, 2017].

Kitada, T. et al., 2012. Mechanism for epigenetic variegation of gene expression at yeast telomeric heterochromatin. *Genes & development*, 26(21), pp.2443–55. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23124068 [Accessed October 4, 2017].

Klaus, C.R. et al., 2014. DOT1L inhibitor EPZ-5676 displays synergistic antiproliferative activity in

combination with standard of care drugs and hypomethylating agents in MLL-rearranged leukemia cells. *The Journal of pharmacology and experimental therapeutics*, 350(3), pp.646–56. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24993360 [Accessed December 17, 2014].

Klose, R.J. et al., 2013. Chromatin sampling--an emerging perspective on targeting polycomb repressor proteins. *PLoS genetics*, 9(8), p.e1003717. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3749931&tool=pmcentrez&rendertype=abstract [Accessed January 21, 2014].

Klymenko, T. et al., 2006. A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. *Genes & development*, 20(9), pp.1110–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16618800 [Accessed September 18, 2017].

Kostrewa, D. et al., 2009. RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature*, 462(7271), pp.323–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19820686 [Accessed September 18, 2017].

Kowalczyk, M.S. et al., 2012. Intragenic enhancers act as alternative promoters. *Molecular cell*, 45(4), pp.447–58. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22264824 [Accessed May 18, 2015].

Kowarz, E. et al., 2007. Complex MLL rearrangements in t(4;11) leukemia patients with absent AF4·MLL fusion allele. *Leukemia*, 21(6), pp.1232–1238. Available at: http://www.nature.com/doifinder/10.1038/sj.leu.2404686 [Accessed August 22, 2017].

Krivtsov, A. V et al., 2008. H3K79 methylation profiles define murine and human MLL-AF4 leukemias. *Cancer cell*, 14(5), pp.355–68. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2591932&tool=pmcentrez&rendertype=abstract [Accessed January 19, 2015].

Krogan, N.J. et al., 2003. The Paf1 Complex Is Required for Histone H3 Methylation by COMPASS and Dot1p: Linking Transcriptional Elongation to Histone Methylation. *Molecular Cell*, 11(3), pp.721–729. Available at: http://www.cell.com/article/S1097276503000911/fulltext [Accessed September 17, 2015].

Krumm, A., Hickey, L.B. & Groudine, M., 1995. Promoter-proximal pausing of RNA polymerase II defines a general rate-limiting step after transcription initiation. *Genes & development*, 9(5), pp.559–72. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7698646 [Accessed September 18, 2017].

Kühn, M.W.M. et al., 2015. MLL partial tandem duplication leukemia cells are sensitive to small molecule DOT1L inhibition. *Haematologica*, 100(5). Available at: http://www.haematologica.org/content/100/5/e190.long [Accessed July 31, 2017].

Kumar, A.R. et al., 2011. t(4;11) leukemias display addiction to MLL-AF4 but not to AF4-MLL. *Leukemia Research*, 35(3), pp.305–309. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0145212610004169 [Accessed August 22, 2017].

Kuntimaddi, A. et al., 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*. Available at: http://www.sciencedirect.com/science/article/pii/S2211124715003794 [Accessed April 26, 2015].

Kurland, J.F. & Tansey, W.P., 2008. Myc-mediated transcriptional repression by recruitment of histone deacetylase. *Cancer research*, 68(10), pp.3624–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18483244 [Accessed September 22, 2017].

Lachner, M. et al., 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature*, 410(6824), pp.116–120. Available at: http://www.nature.com/doifinder/10.1038/35065132 [Accessed July 25, 2017].

Lacoste, N. et al., 2002. Disruptor of telomeric silencing-1 is a chromatin-specific histone H3 methyltransferase. *The Journal of biological chemistry*, 277(34), pp.30421–4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12097318 [Accessed January 6, 2015].

Larson, A.G. et al., 2017. Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature*, 547(7662), pp.236–240. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28636604 [Accessed August 20, 2017].

Lauberth, S.M. et al., 2013. H3K4me3 Interactions with TAF3 Regulate Preinitiation Complex Assembly and Selective Gene Activation. *Cell*, 152(5), pp.1021–1036. Available at: http://www.sciencedirect.com/science/article/pii/S009286741300144X [Accessed August 24, 2017].

Leach, B.I. et al., 2013. Leukemia fusion target AF9 is an intrinsically disordered transcriptional regulator that recruits multiple partners via coupled folding and binding. *Structure (London, England : 1993)*, 21(1), pp.176–83. Available at: http://www.sciencedirect.com/science/article/pii/S0969212612004261 [Accessed February 13, 2014].

Lee, J.-E. et al., 2013. H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. *eLife*, 2, p.e01503. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24368734 [Accessed September 18, 2017].

Lee, T.I. et al., 2006. Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells. *Cell*, 125(2), pp.301–313. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0092867406003849 [Accessed August 25, 2017].

Lehnertz, B. et al., 2003. Suv39h-Mediated Histone H3 Lysine 9 Methylation Directs DNA Methylation to Major Satellite Repeats at Pericentric Heterochromatin. *Current Biology*, 13(14), pp.1192–1200. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0960982203004329 [Accessed August 21, 2017].

Lettice, L.A. et al., 2011. Enhancer-adoption as a mechanism of human developmental disease. *Human Mutation*, 32(12), pp.1492–1499. Available at: http://onlinelibrary.wiley.com/doi/10.1002/humu.21615/abstract;jsessionid=AB65EEF9BB4F35B7B5C5650BA248F908.f03t01 [Accessed September 18, 2017].

Levasseur, D.N. et al., 2008. Oct4 dependence of chromatin structure within the extended Nanog locus in ES cells. *Genes & development*, 22(5), pp.575–80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18283123 [Accessed September 18, 2017].

Lewis, E.B., 1978. A gene complex controlling segmentation in Drosophila. *Nature*, 276(5688), pp.565–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/103000 [Accessed August 24, 2017].

Li, H. et al., 2006. Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature*, 442(7098), pp.91–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16728978 [Accessed September 28, 2017].

Li, Y., Wen, H., Xi, Y., Tanaka, K., Wang, H., Peng, D., Ren, Y., Jin, Q., Dent, S.Y.R., et al., 2014. AF9 YEATS Domain Links Histone Acetylation to DOT1L-Mediated H3K79 Methylation. *Cell*, 159(3), pp.558–571. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0092867414012379 [Accessed August 2, 2017].

Liang, D.-C. et al., 2006. K-ras mutations and N-ras mutations in childhood acute leukemias with or without mixed-lineage leukemia gene rearrangements. *Cancer*, 106(4), pp.950–956. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16404744 [Accessed September 25, 2017].

Liao, Y., Smyth, G.K. & Shi, W., 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7), pp.923–930. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24227677 [Accessed September 20, 2017].

Lieberman-Aiden, E. et al., 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (New York, N.Y.)*, 326(5950), pp.289–93. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19815776 [Accessed July 5, 2017].

Lin, C. et al., 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(3), pp.429–37. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20159561 [Accessed August 2, 2017].

Lin, C. et al., 2011. Dynamic transcriptional events in embryonic stem cells mediated by the super elongation complex (SEC). *Genes & development*, 25(14), pp.1486–98. Available at: http://genesdev.cshlp.org/content/25/14/1486.long [Accessed September 25, 2014].

Lin, S., Luo, R.T., et al., 2016. Instructive Role of MLL-Fusion Proteins Revealed by a Model of t(4;11) Pro-B Acute Lymphoblastic Leukemia. *Cancer Cell*, 30(5), pp.737–749.

Lin, Y.C. et al., 2010. A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. *Nature Immunology*, 11(7), pp.635–643. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20543837 [Accessed August 24, 2017].

van der Linden, M.H. et al., 2009. Outcome of congenital acute lymphoblastic leukemia treated on the Interfant-99 protocol. *Blood*, 114(18), pp.3764–8. Available at: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2009-02-204214 [Accessed September 18, 2017].

Littlefield, O., Korkhin, Y. & Sigler, P.B., 1999. The structural basis for the oriented assembly of a TBP/TFB/promoter complex. *Proceedings of the National Academy of Sciences of the United States of America*, 96(24), pp.13668–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10570130 [Accessed September 18, 2017].

Liu, L. et al., 2014. Transcriptional Pause Release Is a Rate-Limiting Step for Somatic Cell Reprogramming. *Cell Stem Cell*, 15(5), pp.574–588. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25312495 [Accessed September 18, 2017].

Long, H.K., Sims, D., et al., 2013. Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates. *eLife*, 2. Available at: http://elifesciences.org/lookup/doi/10.7554/eLife.00348 [Accessed August 23, 2017].

Long, H.K., Blackledge, N.P. & Klose, R.J., 2013. ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. *Biochemical Society transactions*, 41(3), pp.727–40. Available at: http://biochemsoctrans.org/lookup/doi/10.1042/BST20130028 [Accessed September 15, 2017].

Lu, X. et al., 2008. The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure. *Nature Structural & Molecular Biology*, 15(10), pp.1122–1124. Available at: http://www.nature.com/doifinder/10.1038/nsmb.1489 [Accessed July 21, 2017].

Luger, K. & Richmond, T.J., 1998. The histone tails of the nucleosome. *Current opinion in genetics & development*, 8(2), pp.140–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9610403 [Accessed September 24, 2014].

Lupiáñez, D.G. et al., 2015. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, 161(5), pp.1012–1025. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0092867415003773 [Accessed September 15, 2017].

Lupien, M. et al., 2008. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell*, 132(6), pp.958–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18358809 [Accessed September 27, 2017].

Mak, A.B., Nixon, A.M.L. & Moffat, J., 2012. The Mixed Lineage Leukemia (MLL) Fusion–Associated Gene AF4 Promotes CD133 Transcription. *Cancer Research*, 72(8). Available at: http://cancerres.aacrjournals.org/content/72/8/1929.long [Accessed July 14, 2017].

Marklund, E.G. et al., 2013. Transcription-factor binding and sliding on DNA studied using micro- and macroscopic models. *Proceedings of the National Academy of Sciences of the United States of America*, 110(49), pp.19796–801. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24222688 [Accessed September 7, 2017].

Marshall, N.F. et al., 1996. Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *The Journal of biological chemistry*, 271(43), pp.27176–83. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8900211 [Accessed August 24, 2017].

Marshall, N.F. & Price, D.H., 1995. Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *The Journal of biological chemistry*, 270(21), pp.12335–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7759473 [Accessed August 24, 2017].

Martinez, N. et al., 2004. The oncogenic fusion protein RUNX1-CBFA2T1 supports proliferation and inhibits senescence in t(8;21)-positive leukaemic cells. *BMC Cancer*, 4(1), p.44. Available at: http://bmccancer.biomedcentral.com/articles/10.1186/1471-2407-4-44 [Accessed September 20, 2017].

McGinty, R.K. et al., 2008. Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. *Nature*, 453(7196), pp.812–816. Available at: http://www.nature.com/doifinder/10.1038/nature06906 [Accessed September 18, 2017].

Metzler, M. et al., 2006. A conditional model of MLL-AF4 B-cell tumourigenesis using invertor technology. *Oncogene*, 25(22), pp.3093–3103. Available at: http://www.nature.com/doifinder/10.1038/sj.onc.1209636 [Accessed August 15, 2017].

Meyer, C. et al., 2013. The MLL recombinome of acute leukemias in 2013. *Leukemia*, 27(11), pp.2165–76. Available at: http://dx.doi.org/10.1038/leu.2013.135 [Accessed July 17, 2014].

Mikkelsen, T.S. et al., 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, 448(7153), pp.553–560. Available at: http://www.nature.com/doifinder/10.1038/nature06008 [Accessed August 22, 2017].

Milne, T.A., Martin, M.E., et al., 2005. Leukemogenic MLL fusion proteins bind across a broad region of the Hox a9 locus, promoting transcription and multiple histone modifications. *Cancer research*, 65(24), pp.11367–74. Available at: http://cancerres.aacrjournals.org/content/65/24/11367.short [Accessed February 13, 2014].

Milne, T.A., Hughes, C.M., et al., 2005. 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(3), pp.749–54. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15640349 [Accessed August 21, 2017].

Milne, T.A. et al., 2010. Multiple Interactions Recruit MLL1 and MLL1 Fusion Proteins to the HOXA9 Locus in Leukemogenesis. *Molecular Cell*, 38(6), pp.853–863. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1097276510003734 [Accessed August 21, 2017].

Min, J. et al., 2003. Structure of the Catalytic Domain of Human DOT1L , a Non-SET Domain Nucleosomal Histone Methyltransferase University of North Carolina at Chapel Hill. , 112, pp.711–723.

Mishiro, T. et al., 2009. Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. *The EMBO journal*, 28(9), pp.1234–45. Available at: http://emboj.embopress.org/cgi/doi/10.1038/emboj.2009.81 [Accessed September 15, 2017].

Mitchell, J.A. & Fraser, P., 2008. Transcription factories are nuclear subcompartments that remain in the absence of transcription. *Genes & development*, 22(1), pp.20–5. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/18172162 [Accessed July 17, 2017].

Mohan, M., Lin, C., et al., 2010. Licensed to elongate: a molecular mechanism for MLL-based leukaemogenesis. *Nature reviews. Cancer*, 10(10), pp.721–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20844554 [Accessed February 6, 2014].

Mohan, M., Herz, H.-M., et al., 2010. Linking H3K79 trimethylation to Wnt signaling through a novel Dot1-containing complex (DotCom). *Genes & development*, 24(6), pp.574–589. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2841335&tool=pmcentrez&rendertype=abstract [Accessed February 2, 2014].

Monroe, S.C. et al., 2011. MLL-AF9 and MLL-ENL alter the dynamic association of transcriptional regulators with genes critical for leukemia. *Experimental hematology*, 39(1), pp.77-86–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20854876 [Accessed July 13, 2017].

Montes, R. et al., 2011. Enforced expression of MLL-AF4 fusion in cord blood CD34+ cells enhances the hematopoietic repopulating cell function and clonogenic potential but is not sufficient to initiate leukemia. *Blood*, 117(18), pp.4746–58. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21389315 [Accessed February 6, 2014].

Montes, R. et al., 2014. Ligand-independent FLT3 activation does not cooperate with MLL-AF4 to immortalize/transform cord blood CD34+ cells. *Leukemia*, 28(3), pp.666–674. Available at: http://www.nature.com/doifinder/10.1038/leu.2013.346 [Accessed August 22, 2017].

Moore, S.D.P., Strehl, S. & Dal Cin, P., 2005. Acute myelocytic leukemia with t(11;17)(q23;q12-q21) involves a fusion of MLL and AF17. *Cancer genetics and cytogenetics*, 157(1), pp.87–9. Available at: http://www.sciencedirect.com/science/article/pii/S0165460804002754 [Accessed September 26, 2014].

Morin, R.D. et al., 2010. Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nature genetics*, 42(2), pp.181–5. Available at: http://www.nature.com/doifinder/10.1038/ng.518 [Accessed September 28, 2017].

Mueller, C.L. & Jaehning, J.A., 2002. Ctr9, Rtf1, and Leo1 are components of the Paf1/RNA polymerase II complex. *Molecular and cellular biology*, 22(7), pp.1971–80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11884586 [Accessed August 9, 2017].

Mueller, D., Bach, C., Zeisig, D., Garcia-Cuellar, M.-P., Monroe, S., Sreekumar, A., Zhou, R., Nesvizhskii, A., Chinnaiyan, A., Hess, J.L., et al., 2007. A role for the MLL fusion partner ENL in transcriptional elongation and chromatin modification. *Blood*, 110(13). Available at: http://www.bloodjournal.org/content/110/13/4445.long?sso-checked=true [Accessed September 1, 2017].

Mueller, D. et al., 2009. Misguided Transcriptional Elongation Causes Mixed Lineage Leukemia N. Zeleznik-Le, ed. *PLoS Biology*, 7(11), p.e1000249. Available at: http://dx.plos.org/10.1371/journal.pbio.1000249 [Accessed August 18, 2017].

Narendra, V. et al., 2016. CTCF-mediated topological boundaries during development foster appropriate gene regulation. *Genes & development*, 30(24), pp.2657–2662. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28087711 [Accessed September 18, 2017].

Nativio, R. et al., 2009. Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus. W. A. Bickmore, ed. *PLoS genetics*, 5(11), p.e1000739. Available at: http://dx.plos.org/10.1371/journal.pgen.1000739 [Accessed September 15, 2017].

Ng, H.H. et al., 2002. Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association. *Genes & development*, 16(12), pp.1518–1527. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=186335&tool=pmcentrez&rendertype=abstract [Accessed January 21, 2014].

Ng, H.H., Dole, S. & Struhl, K., 2003a. The Rtf1 component of the Paf1 transcriptional elongation complex is required for ubiquitination of histone H2B. *The Journal of biological chemistry*, 278(36), pp.33625–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12876293 [Accessed August 24, 2017].

Nguyen, A.T. et al., 2011. Essential role of DOT1L in maintaining normal adult hematopoiesis. *Cell research*, 21(9), pp.1370–3. Available at: http://dx.doi.org/10.1038/cr.2011.115 [Accessed June 1, 2016].

Nichols, M.H. & Corces, V.G., 2015. A CTCF Code for 3D Genome Architecture. *Cell*, 162(4), pp.703–705. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0092867415009691 [Accessed August 21, 2017].

Niedermaier, M. et al., 2005. An inversion involving the mouse Shh locus results in brachydactyly through dysregulation of Shh expression. *The Journal of clinical investigation*, 115(4), pp.900–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15841179 [Accessed September 18, 2017].

Nolis, I.K. et al., 2009a. Transcription factors mediate long-range enhancer-promoter interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 106(48), pp.20222–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19923429 [Accessed August 2, 2017].

Nord, A.S. et al., 2013. Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell*, 155(7), pp.1521–31. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24360275 [Accessed September 18, 2017].

Okada, Y. et al., 2005. hDOT1L links histone methylation to leukemogenesis. *Cell*, 121(2), pp.167–78. Available at: http://www.sciencedirect.com/science/article/pii/S0092867405001868 [Accessed February 2, 2014].

Okano, M. et al., 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, 99(3), pp.247–57. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10555141 [Accessed September 15, 2017].

Okuda, H. et al., 2015. AF4 uses the SL1 components of RNAP1 machinery to initiate MLL fusion- and AEP-

dependent transcription. *Nature Communications*, 6, p.8869. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26593443 [Accessed August 17, 2017].

Okuda, H. et al., 2016. TBP loading by AF4 through SL1 is the major rate-limiting step in MLL fusion-dependent transcription. *Cell Cycle*, 15(20), pp.2712–2722. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27564129 [Accessed August 17, 2017].

Orlovsky, K. et al., 2011. Down-regulation of homeobox genes MEIS1 and HOXA in MLL-rearranged acute leukemia impairs engraftment and reduces proliferation. *Proceedings of the National Academy of Sciences*, 108(19), pp.7956–7961. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21518888 [Accessed September 18, 2017].

Osborne, C.S. et al., 2004. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nature Genetics*, 36(10), pp.1065–1071. Available at: http://www.nature.com/doifinder/10.1038/ng1423 [Accessed August 18, 2017].

Osborne, C.S. et al., 2007. Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. Susan M Gasser, ed. *PLoS biology*, 5(8), p.e192. Available at: http://dx.plos.org/10.1371/journal.pbio.0050192 [Accessed September 28, 2017].

Osborne, E.A., Dudoit, S. & Rine, J., 2009. The establishment of gene silencing at single-cell resolution. *Nature genetics*, 41(7), pp.800–6. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2739733&tool=pmcentrez&rendertype=abstract [Accessed January 9, 2015].

Van Oss, S.B. et al., 2016. The Histone Modification Domain of Paf1 Complex Subunit Rtf1 Directly Stimulates H2B Ubiquitylation through an Interaction with Rad6. *Molecular Cell*, 64(4), pp.815–825. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27840029 [Accessed June 26, 2017].

Oti, M. et al., 2016. CTCF-mediated chromatin loops enclose inducible gene regulatory domains. *BMC Genomics 2016 17:1*, 17(1), p.252. Available at: https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-016-2516-6 [Accessed September 18, 2017].

Ou, H.D. et al., 2017. ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science*, 357(6349). Available at: http://science.sciencemag.org/content/357/6349/eaag0025.long [Accessed September 15, 2017].

Parvin, J.D. & Sharp, P.A., 1993. DNA topology and a minimal set of basal factors for transcription by RNA polymerase II. *Cell*, 73(3), pp.533–40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8490964 [Accessed September 18, 2017].

Peng, J. et al., 1998. Identification of multiple cyclin subunits of human P-TEFb. *Genes & development*, 12(5), pp.755–62. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9499409 [Accessed August 24, 2017].

Peterlin, B.M. & Price, D.H., 2006. Controlling the elongation phase of transcription with P-TEFb. *Molecular cell*, 23(3), pp.297–305. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16885020 [Accessed August 2, 2017].

Petruk, S., Cai, J., et al., 2017. Delayed Accumulation of H3K27me3 on Nascent DNA Is Essential for Recruitment of Transcription Factors at Early Stages of Stem Cell Differentiation. *Molecular Cell*, 66(2), p.247–257.e5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28410996 [Accessed August 7, 2017].

Petruk, S., Mariani, S.A., et al., 2017. Structure of Nascent Chromatin Is Essential for Hematopoietic Lineage Specification. *Cell reports*, 19(2), pp.295–306. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28402853 [Accessed September 27, 2017].

Peukert, K. et al., 1997. An alternative pathway for gene regulation by Myc. *The EMBO Journal*, 16(18), pp.5672–5686. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9312026 [Accessed September 22, 2017].

Pieters, R. 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. *Lancet (London, England)*, 370(9583), pp.240–50. Available at: http://linkinghub.elsevier.com/retrieve/pii/S014067360761126X [Accessed September 18, 2017].

Pilon, A.M. et al., 2011. Genome-wide ChIP-Seq reveals a dramatic shift in the binding of the transcription factor erythroid Kruppel-like factor during erythrocyte differentiation. *Blood*, 118(17), pp.e139-48. Available at: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2011-05-355107 [Accessed August 24, 2017].

Plaschka, C. et al., 2015. Architecture of the RNA polymerase II–Mediator core initiation complex. *Nature*, 518(7539), pp.376–380. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25652824 [Accessed August 30, 2017].

Pokholok, D.K. et al., 2005. Genome-wide Map of Nucleosome Acetylation and Methylation in Yeast. *Cell*, 122(4), pp.517–527. Available at: http://www.sciencedirect.com/science/article/pii/S0092867405006458 [Accessed August 22, 2017].

Pollex, T., Piolot, T. & Heard, E., 2013. Live-Cell Imaging Combined with Immunofluorescence, RNA, or DNA FISH to Study the Nuclear Dynamics and Expression of the X-Inactivation Center. In Humana Press, Totowa, NJ, pp. 13–31. Available at: http://link.springer.com/10.1007/978-1-62703-526-2_2 [Accessed August 31, 2017].

Prieto, C. et al., 2016. Activated KRAS Cooperates with MLL-AF4 to Promote Extramedullary Engraftment and Migration of Cord Blood CD34+ HSPC But Is Insufficient to Initiate Leukemia. *Cancer Research*, 76(8). Available at: http://cancerres.aacrjournals.org/content/76/8/2478.long [Accessed August 15, 2017].

Pui, C.-H. et al., 2003. Clinical heterogeneity in childhood acute lymphoblastic leukemia with 11q23 rearrangements. *Leukemia*, 17(4), pp.700–706. Available at:

http://www.nature.com/doifinder/10.1038/sj.leu.2402883 [Accessed August 22, 2017].

Rao, B. et al., 2005. Dimethylation of histone H3 at lysine 36 demarcates regulatory and nonregulatory chromatin genome-wide. *Molecular and cellular biology*, 25(21), pp.9447–59. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16227595 [Accessed August 22, 2017].

Rau, R.E. et al., 2016. DOT1L as a therapeutic target for the treatment of DNMT3A-mutant acute myeloid leukemia. *Blood*, 128(7). Available at: http://www.bloodjournal.org/content/128/7/971.long?sso-checked=true [Accessed July 31, 2017].

Rea, S. et al., 2000. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature, Published online: 10 August 2000; | doi:10.1038/35020506*, 406(6796), p.593. Available at: https://www.nature.com/nature/journal/v406/n6796/full/406593a0.html [Accessed September 18, 2017].

Revyakin, A., Ebright, R.H. & Strick, T.R., 2004. Promoter unwinding and promoter clearance by RNA polymerase: detection by single-molecule DNA nanomanipulation. *Proceedings of the National Academy of Sciences of the United States of America*, 101(14), pp.4776–80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15037753 [Accessed September 18, 2017].

Richmond, T.J. et al., 1997. Crystal structure of the nucleosome core particle at 2.8|[thinsp]||[Aring]| resolution. *Nature*, 389(6648), pp.251–260. Available at: http://www.nature.com/doifinder/10.1038/38444 [Accessed July 31, 2017].

Robinson, M.D., McCarthy, D.J. & Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), pp.139–140. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19910308 [Accessed September 20, 2017].

Rougvie, A.E. & Lis, J.T., 1988. The RNA polymerase II molecule at the 5′ end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. *Cell*, 54(6), pp.795–804. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0092867488910872 [Accessed July 31, 2017].

Rozovskaia, T. et al., 2001. Upregulation of Meis1 and HoxA9 in acute lymphocytic leukemias with the t(4 : 11) abnormality. *Oncogene*, 20(7), pp.874–8. Available at: http://www.nature.com/doifinder/10.1038/sj.onc.1204174 [Accessed September 18, 2017].

Ruthenburg, A.J. et al., 2011. Recognition of a Mononucleosomal Histone Modification Pattern by BPTF via Multivalent Interactions. *Cell*, 145(5), pp.692–706. Available at: http://www.sciencedirect.com/science/article/pii/S0092867411004351 [Accessed August 22, 2017].

Sabra, M. et al., 2013a. The Tudor protein survival motor neuron (SMN) is a chromatin-binding protein that interacts with methylated lysine 79 of histone H3. *Journal of Cell Science*, 126(16). Available at: http://jcs.biologists.org/content/126/16/3664.long [Accessed August 9, 2017].

Sagai, T. et al., 2005. Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development (Cambridge, England)*, 132(4), pp.797–

803. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8026324 [Accessed September 18, 2017].

Sandmann, T. et al., 2007. A core transcriptional network for early mesoderm development in Drosophila melanogaster. *Genes & development*, 21(4), pp.436–49. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17322403 [Accessed September 28, 2017].

Sanjuan-Pla, A. et al., 2015. Revisiting the biology of infant t(4;11)/MLL-AF4+ B-cell acute lymphoblastic leukemia. *Blood*, 126(25). Available at: http://www.bloodjournal.org/content/126/25/2676.long?sso-checked=true [Accessed July 14, 2017].

Sarkaria, S.M. et al., 2014. Primary acute myeloid leukemia cells with IDH1 or IDH2 mutations respond to a DOT1L inhibitor in vitro. *Leukemia*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25092143 [Accessed October 21, 2014].

Sarraf, S.A. & Stancheva, I., 2004. Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Molecular cell*, 15(4), pp.595–605. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1097276504004046 [Accessed September 28, 2017].

Sauvageau, M. & Sauvageau, G., 2010. Polycomb group proteins: multi-faceted regulators of somatic stem cells and cancer. *Cell stem cell*, 7(3), pp.299–313. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1934590910003942 [Accessed August 24, 2017].

Saxonov, S., Berg, P. & Brutlag, D.L., 2006. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences of the United States of America*, 103(5), pp.1412–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16432200 [Accessed August 30, 2017].

Schmitges, F.W. et al., 2011. Histone methylation by PRC2 is inhibited by active chromatin marks. *Molecular cell*, 42(3), pp.330–41. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21549310 [Accessed September 28, 2017].

Schübeler, D. et al., 2004. The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes & development*, 18(11), pp.1263–71. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=420352&tool=pmcentrez&rendertype=abstract [Accessed December 28, 2014].

Seizl, M. et al., 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(14), pp.6291–6304. Available at: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkr229 [Accessed August 30, 2017].

Sexton, T. et al., 2012. Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. *Cell*, 148(3), pp.458–472. Available at: http://www.sciencedirect.com/science/article/pii/S0092867412000165 [Accessed August 21, 2017].

Shanower, G.A. et al., 2005. Characterization of the grappa gene, the Drosophila histone H3 lysine 79 methyltransferase. *Genetics*, 169(1), pp.173–84. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1448877&tool=pmcentrez&rendertype=abstract [Accessed January 5, 2015].

Shen, X. et al., 2008. EZH1 Mediates Methylation on Histone H3 Lysine 27 and Complements EZH2 in Maintaining Stem Cell Identity and Executing Pluripotency. *Molecular Cell*, 32(4), pp.491–502. Available at: http://www.sciencedirect.com/science/article/pii/S1097276508007326 [Accessed September 18, 2017].

Shi, X. et al., 1997. Cdc73p and Paf1p are found in a novel RNA polymerase II-containing complex distinct from the Srbp-containing holoenzyme. *Molecular and cellular biology*, 17(3), pp.1160–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9032243 [Accessed August 9, 2017].

Shiekhattar, R. et al., 1995. Cdk-activating kinase complex is a component of human transcription factor TFIIH. *Nature*, 374(6519), pp.283–287. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7533895 [Accessed August 24, 2017].

Shogren-Knaak, M. et al., 2006. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science (New York, N.Y.)*, 311(5762), pp.844–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16469925 [Accessed September 18, 2017].

Singer, M.S. et al., 1998. Identification of high-copy disruptors of telomeric silencing in Saccharomyces cerevisiae. *Genetics*, 150(2), pp.613–32. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1460361&tool=pmcentrez&rendertype=abstract [Accessed September 24, 2014].

Small, S., Blair, A. & Levine, M., 1992. Regulation of even-skipped stripe 2 in the Drosophila embryo. *The EMBO journal*, 11(11), pp.4047–57. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1327756 [Accessed September 28, 2017].

Smallwood, A. et al., 2007. Functional cooperation between HP1 and DNMT1 mediates gene silencing. *Genes & development*, 21(10), pp.1169–78. Available at: http://www.genesdev.org/cgi/doi/10.1101/gad.1536807 [Accessed September 28, 2017].

Sobhian, B. et al., 2010. HIV-1 Tat assembles a multifunctional transcription elongation complex and stably associates with the 7SK snRNP. *Molecular cell*, 38(3), pp.439–51. Available at: http://www.cell.com/article/S1097276510003114/fulltext [Accessed November 17, 2015].

Souers, A.J. et al., 2013. ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. *Nature medicine*, 19(2), pp.202–8. Available at: http://www.nature.com/doifinder/10.1038/nm.3048 [Accessed August 25, 2017].

Soufi, A., Donahue, G. & Zaret, K.S., 2012. Facilitators and Impediments of the Pluripotency Reprogramming Factors' Initial Engagement with the Genome. *Cell*, 151(5), pp.994–1004. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0092867412012986 [Accessed September 27, 2017].

Spitz, F. & Furlong, E.E.M., 2012. Transcription factors: from enhancer binding to developmental control. *Nature Reviews Genetics*, 13(9), pp.613–626. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22868264 [Accessed September 28, 2017].

Squazzo, S.L. et al., 2002. The Paf1 complex physically and functionally associates with transcription elongation factors in vivo. *The EMBO journal*, 21(7), pp.1764–74. Available at: http://emboj.embopress.org/cgi/doi/10.1093/emboj/21.7.1764 [Accessed August 9, 2017].

Stam, R.W. et al., 2010. Gene expression profiling–based dissection of MLL translocated and MLL germline acute lymphoblastic leukemia in infants. *Blood*, 115(14). Available at: http://www.bloodjournal.org/content/115/14/2835.long?sso-checked=true [Accessed August 22, 2017].

Steger, D.J. et al., 2008. DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. *Molecular and cellular biology*, 28(8), pp.2825–39. Available at: http://mcb.asm.org/content/28/8/2825.short [Accessed February 1, 2014].

Stein, E.M. et al., 2014. The DOT1L Inhibitor EPZ-5676: Safety and Activity in Relapsed/Refractory Patients with MLL-Rearranged Leukemia. *Blood*, 124(21).

Strom, A.R. et al., 2017. Phase separation drives heterochromatin domain formation. *Nature*, 547(7662), pp.241–245. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28636597 [Accessed August 7, 2017].

Stumpel, D.J.P.M. et al., 2009. Specific promoter methylation identifies different subgroups of MLL-rearranged infant acute lymphoblastic leukemia, influences clinical outcome, and provides therapeutic options. *Blood*, 114(27), pp.5490–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19855078 [Accessed September 18, 2017].

Stumpf, M. et al., 2006. The mediator complex functions as a coactivator for GATA-1 in erythropoiesis via subunit Med1/TRAP220. *Proceedings of the National Academy of Sciences of the United States of America*, 103(49), pp.18504–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17132730 [Accessed September 18, 2017].

Suka, N. et al., 2001. Highly Specific Antibodies Determine Histone Acetylation Site Usage in Yeast Heterochromatin and Euchromatin. *Molecular Cell*, 8(2), pp.473–479. Available at: http://www.sciencedirect.com/science/article/pii/S109727650100301X [Accessed August 21, 2017].

Sun, Z.-W. & Allis, C.D., 2002. Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. *Nature*, 418(6893), pp.104–108. Available at: http://www.nature.com/doifinder/10.1038/nature00883 [Accessed August 24, 2017].

Szutorisz, H., Dillon, N. & Tora, L., 2005. The role of enhancers as centres for general transcription factor recruitment. *Trends in biochemical sciences*, 30(11), pp.593–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16126390 [Accessed September 18, 2017].

Tachibana, M. et al., 2005. Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. *Genes & development*, 19(7), pp.815–26.

Available at: http://www.ncbi.nlm.nih.gov/pubmed/15774718 [Accessed August 21, 2017].

Takahashi, H., Parmely, T.J., et al., 2011. Human Mediator Subunit MED26 Functions as a Docking Site for Transcription Elongation Factors. *Cell*, 146(1), pp.92–104. Available at: http://www.sciencedirect.com/science/article/pii/S0092867411006477 [Accessed July 31, 2017].

Takahashi, Y.-H. et al., 2011. Dot1 and histone H3K79 methylation in natural telomeric and HM silencing. *Molecular cell*, 42(1), pp.118–26. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3085244&tool=pmcentrez&rendertype=abstract [Accessed January 8, 2015].

Taketani, T. et al., 2004. FLT3 mutations in the activation loop of tyrosine kinase domain are frequently found in infant ALL with MLL rearrangements and pediatric ALL with hyperdiploidy. *Blood*, 103(3), pp.1085–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14504097 [Accessed September 25, 2017].

Tamai, H. et al., 2011. Activated K-Ras protein accelerates human MLL/AF4-induced leukemo-lymphomogenicity in a transgenic mouse model. *Leukemia*, 25(5), pp.888–891. Available at: http://www.nature.com/doifinder/10.1038/leu.2011.15 [Accessed August 15, 2017].

Tamaru, H., 2010. Confining euchromatin/heterochromatin territory: jumonji crosses the line. *Genes & development*, 24(14), pp.1465–78. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20634313 [Accessed September 18, 2017].

Tang, Z. et al., 2015. CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell*, 163(7), pp.1611–27. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0092867415015044 [Accessed September 15, 2017].

Thirman, M.J. et al., 1993. Rearrangement of the MLL Gene in Acute Lymphoblastic and Acute Myeloid Leukemias with 11q23 Chromosomal Translocations. *New England Journal of Medicine*, 329(13), pp.909–914. Available at: http://www.nejm.org/doi/abs/10.1056/NEJM199309233291302 [Accessed August 21, 2017].

Thoma, F., Koller, T. & Klug, A., 1979. Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *The Journal of Cell Biology*, 83(2). Available at: http://jcb.rupress.org/content/83/2/403.long [Accessed August 22, 2017].

Thomas, M. et al., 2005. Targeting MLL-AF4 with short interfering RNAs inhibits clonogenicity and engraftment of t(4;11)-positive human leukemic cells. *Blood*, 106(10). Available at: http://www.bloodjournal.org/content/106/10/3559.long?sso-checked=true [Accessed July 14, 2017].

Thomson, J.P. et al., 2010. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature*, 464(7291), pp.1082–1086. Available at: http://www.nature.com/doifinder/10.1038/nature08924 [Accessed August 23, 2017].

Thurman, R.E. et al., 2012. The accessible chromatin landscape of the human genome. *Nature*, 489(7414),

pp.75–82. Available at: http://www.nature.com/doifinder/10.1038/nature11232 [Accessed September 18, 2017].

Tie, F. et al., 2009. CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. *Development (Cambridge, England)*, 136(18), pp.3131–41. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19700617 [Accessed August 21, 2017].

Tie, F. et al., 2014. Trithorax monomethylates histone H3K4 and interacts directly with CBP to promote H3K27 acetylation and antagonize Polycomb silencing. *Development (Cambridge, England)*, 141(5), pp.1129–39. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24550119 [Accessed September 18, 2017].

Tkachuk, D.C., Kohler, S. & Cleary, M.L., 1992. Involvement of a homolog of Drosophila trithorax by 11q23 chromosomal translocations in acute leukemias. *Cell*, 71(4), pp.691–700. Available at: http://linkinghub.elsevier.com/retrieve/pii/0092867492906029 [Accessed August 15, 2017].

Tolhuis, B. et al., 2002. Looping and Interaction between Hypersensitive Sites in the Active β-globin Locus. *Molecular Cell*, 10(6), pp.1453–1465. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1097276502007815 [Accessed August 18, 2017].

Tsai, F.T. & Sigler, P.B., 2000. Structural basis of preinitiation complex assembly on human pol II promoters. *The EMBO journal*, 19(1), pp.25–36. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10619841 [Accessed September 18, 2017].

Turner, B.M., 2002. Cellular memory and the histone code. *Cell*, 111(3), pp.285–91. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12419240 [Accessed January 15, 2015].

Uslu, V.V. et al., 2014. Long-range enhancers regulating Myc expression are required for normal facial morphogenesis. *Nature Genetics*, 46(7), pp.753–758. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24859337 [Accessed August 16, 2017].

Vakoc, C.R. et al., 2005. Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Molecular cell*, 17(3), pp.453–62. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15694345 [Accessed August 7, 2017].

Vandenberg, C.J. & Cory, S., 2013. ABT-199, a new Bcl-2-specific BH3 mimetic, has in vivo efficacy against aggressive Myc-driven mouse lymphomas without provoking thrombocytopenia. *Blood*, 121(12), pp.2285–8. Available at: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2013-01-475855 [Accessed August 25, 2017].

Vaquero, A. et al., 2007. SIRT1 regulates the histone methyl-transferase SUV39H1 during heterochromatin formation. *Nature*, 450(7168), pp.440–444. Available at: http://www.nature.com/doifinder/10.1038/nature06268 [Accessed September 27, 2017].

Varambally, S. et al., 2002. The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature*, 419(6907), pp.624–629. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12374981

[Accessed September 28, 2017].

Venolia, L. & Gartler, S.M., 1983. Comparison of transformation efficiency of human active and inactive X-chromosomal DNA. *Nature*, 302(5903), pp.82–83. Available at: http://www.nature.com/doifinder/10.1038/302082a0 [Accessed September 15, 2017].

Vermeulen, M. et al., 2007. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell*, 131(1), pp.58–69. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17884155 [Accessed August 1, 2017].

Vernimmen, D. et al., 2009. Chromosome looping at the human α-globin locus is mediated via the major upstream regulatory element (HS −40). *Blood*, 114(19). Available at: http://www.bloodjournal.org/content/114/19/4253.long?sso-checked=true [Accessed August 6, 2017].

Verrijzer, C.P. et al., 1995. Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell*, 81(7), pp.1115–25. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7600579 [Accessed September 18, 2017].

Voss, T.C. et al., 2011. Dynamic exchange at regulatory elements during chromatin remodeling underlies assisted loading mechanism. *Cell*, 146(4), pp.544–54. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0092867411007616 [Accessed September 28, 2017].

Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., 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(3), pp.343–56. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9450929 [Accessed August 24, 2017].

Wada, T., Takagi, T., Yamaguchi, Y., Watanabe, D., et al., 1998. Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription in vitro. *The EMBO journal*, 17(24), pp.7395–403. Available at: http://emboj.embopress.org/cgi/doi/10.1093/emboj/17.24.7395 [Accessed August 24, 2017].

Wade, P.A. et al., 1996. A Novel Collection of Accessory Factors Associated with Yeast RNA Polymerase II. *Protein Expression and Purification*, 8(1), pp.85–90. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1046592896900777 [Accessed August 9, 2017].

Wakeman, T.P. et al., 2012. Bat3 facilitates H3K79 dimethylation by DOT1L and promotes DNA damage-induced 53BP1 foci at G1/G2 cell-cycle phases. *The EMBO journal*, 31(9), pp.2169–81. Available at: http://emboj.embopress.org/content/31/9/2169.abstract [Accessed January 5, 2015].

Walhout, A.J.M., 2006. Unraveling transcription regulatory networks by protein-DNA and protein-protein interaction mapping. *Genome research*, 16(12), pp.1445–54. Available at: http://www.genome.org/cgi/doi/10.1101/gr.5321506 [Accessed September 25, 2017].

Wan, L. et al., 2017. ENL links histone acetylation to oncogenic gene expression in acute myeloid leukaemia. *Nature*, 543(7644), pp.265–269. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28241141

[Accessed August 7, 2017].

Wang, Y. et al., 2009. Quantitative transcription factor binding kinetics at the single-molecule level. *Biophysical journal*, 96(2), pp.609–20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19167308 [Accessed September 8, 2017].

Waters, N.J. et al., 2014. Pediatric Dose Determinations for the Phase I Study of the DOT1L Inhibitor, EPZ-5676, in MLL-r Acute Leukemia: Leveraging Early Clinical Data in Adults through Physiologically-Based Pharmacokinetic Modeling. *Blood*, 124(21). Available at: http://www.bloodjournal.org/content/124/21/3619 [Accessed September 22, 2017].

van Welsem, T. et al., 2008. Synthetic lethal screens identify gene silencing processes in yeast and implicate the acetylated amino terminus of Sir3 in recognition of the nucleosome core. *Molecular and cellular biology*, 28(11), pp.3861–72. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2423298&tool=pmcentrez&rendertype=abstract [Accessed December 24, 2014].

Whyte, W.A. et al., 2013. Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. *Cell*, 153(2), pp.307–319. Available at: http://www.sciencedirect.com/science/article/pii/S0092867413003929 [Accessed August 18, 2017].

Wilkinson, A.C. 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(1), pp.116–27. Available at: http://www.sciencedirect.com/science/article/pii/S221112471300003X [Accessed January 19, 2015].

Wilson, N.K. et al., 2010. Combinatorial Transcriptional Control In Blood Stem/Progenitor Cells: Genome-wide Analysis of Ten Major Transcriptional Regulators. *Cell Stem Cell*, 7(4), pp.532–544. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1934590910004406 [Accessed September 25, 2017].

Winter, G.E. et al., 2017. BET Bromodomain Proteins Function as Master Transcription Elongation Factors Independent of CDK9 Recruitment. *Molecular Cell*, 67(1), p.5–18.e19. Available at: http://www.sciencedirect.com/science/article/pii/S1097276517304069 [Accessed September 15, 2017].

Wood, A. et al., 2003. The Paf1 complex is essential for histone monoubiquitination by the Rad6-Bre1 complex, which signals for histone methylation by COMPASS and Dot1p. *The Journal of biological chemistry*, 278(37), pp.34739–42. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12876294 [Accessed August 9, 2017].

Wunderlich, Z. & Mirny, L.A., 2009. Different gene regulation strategies revealed by analysis of binding motifs. *Trends in Genetics*, 25(10), pp.434–440. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0168952509001656 [Accessed August 24, 2017].

Wysocka, J. et al., 2006. A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature*, 442(7098), p.86. Available at: http://www.nature.com/doifinder/10.1038/nature04815 [Accessed August 22, 2017].

Wysocki, R. et al., 2005. Role of Dot1-dependent histone H3 methylation in G1 and S phase DNA damage checkpoint functions of Rad9. *Molecular and cellular biology*, 25(19), pp.8430–43. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1265753&tool=pmcentrez&rendertype=abstract [Accessed January 6, 2015].

Xi, H. et al., 2007. Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. *PLoS genetics*, 3(8), p.e136. Available at: http://dx.plos.org/10.1371/journal.pgen.0030136 [Accessed September 18, 2017].

Xiao, T. et al., 2005a. Histone H2B ubiquitylation is associated with elongating RNA polymerase II. *Molecular and cellular biology*, 25(2), pp.637–51. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15632065 [Accessed August 24, 2017].

Xu, Y. et al., 2017. Architecture of the RNA polymerase II-Paf1C-TFIIS transcription elongation complex. *Nature Communications*, 8, p.15741. Available at: http://www.nature.com/doifinder/10.1038/ncomms15741 [Accessed August 9, 2017].

Yamada, T. et al., 2006. P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Molecular cell*, 21(2), pp.227–37. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1097276505018125 [Accessed September 18, 2017].

Yamaguchi, Y. et al., 1999. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell*, 97(1), pp.41–51. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10199401 [Accessed August 24, 2017].

Yan, J. et al., 2017. Histone H3 Lysine 4 methyltransferases MLL3 and MLL4 Modulate Long-range Chromatin Interactions at Enhancers. *bioRxiv*. Available at: http://www.biorxiv.org/content/early/2017/02/21/110239 [Accessed August 22, 2017].

Yang, L. et al., 2002. Molecular cloning of ESET, a novel histone H3-specific methyltransferase that interacts with ERG transcription factor. *Oncogene*, 21(1), pp.148–152. Available at: http://www.nature.com/doifinder/10.1038/sj.onc.1204998 [Accessed August 7, 2017].

Yang, Z. et al., 2005. Recruitment of P-TEFb for Stimulation of Transcriptional Elongation by the Bromodomain Protein Brd4. *Molecular Cell*, 19(4), pp.535–545. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1097276505014310 [Accessed August 21, 2017].

Yokoyama, A. et al., 2010. A higher-order complex containing AF4 and ENL family proteins with P-TEFb facilitates oncogenic and physiologic MLL-dependent transcription. *Cancer Cell*, 17(5), pp.198–212. Available at: http://www.cell.com/cancer-cell/pdf/S1535-6108(16)30495-0.pdf [Accessed August 8, 2017].

Yokoyama, A. et al., 2004. Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression. *Molecular and cellular biology*, 24(13), pp.5639–49. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15199122 [Accessed August 21, 2017].

Yokoyama, A. et al., 2005. The menin tumor suppressor protein is an essential oncogenic cofactor for MLL-associated leukemogenesis. *Cell*, 123(2), pp.207–18. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16239140 [Accessed August 18, 2017].

Yokoyama, A. & Cleary, M.L., 2008. Menin critically links MLL proteins with LEDGF on cancer-associated target genes. *Cancer cell*, 14(1), pp.36–46. Available at: http://www.sciencedirect.com/science/article/pii/S153561080800158X [Accessed January 23, 2014].

Yu, B.D. et al., 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(18), pp.10632–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9724755 [Accessed August 15, 2017].

Yu, M. et al., 2015. RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. *Science (New York, N.Y.)*, 350(6266), pp.1383–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26659056 [Accessed September 21, 2017].

Yu, W. et al., 2012. Catalytic site remodelling of the DOT1L methyltransferase by selective inhibitors. *Nature communications*, 3, p.1288. Available at: http://dx.doi.org/10.1038/ncomms2304 [Accessed January 29, 2014].

Yuan, W. et al., 2011. H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *The Journal of biological chemistry*, 286(10), pp.7983–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21239496 [Accessed September 28, 2017].

Yuh, C.H. et al., 1994. Complexity and organization of DNA-protein interactions in the 5'-regulatory region of an endoderm-specific marker gene in the sea urchin embryo. *Mechanisms of development*, 47(2), pp.165–86. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7811639 [Accessed September 28, 2017].

Zeisig, B.B. et al., 2003. Hoxa9 and Meis1 Are Key Targets for MLL-ENL-Mediated Cellular Immortalization. *Molecular and Cellular Biology*, 24(2), pp.617–628. Available at: http://mcb.asm.org/content/24/2/617.short [Accessed February 13, 2014].

Zeisig, D.T. et al., 2005. The eleven-nineteen-leukemia protein ENL connects nuclear MLL fusion partners with chromatin. *Oncogene*, 24(35), pp.5525–5532. Available at: http://www.nature.com/doifinder/10.1038/sj.onc.1208699 [Accessed August 8, 2017].

Zhang, L. et al., Inhibition of histone H3K79 methylation selectively inhibits proliferation, self-renewal and metastatic potential of breast cancer. *Oncotarget*, 5(21), pp.10665–10677. Available at: http://www.impactjournals.com/oncotarget/index.php?journal=oncotarget&page=article&op=view&path%5B%5D=2496&path%5B%5D=5189 [Accessed February 26, 2015].

Ziemin-van der Poel, S. et al., 1991. Identification of a gene, MLL, that spans the breakpoint in 11q23 translocations associated with human leukemias. *Proceedings of the National Academy of Sciences of the United States of America*, 88(23), pp.10735–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=53005&tool=pmcentrez&rendertype=abstra

ct [Accessed September 26, 2014].

Zinzalla, G., 2016. A New Way Forward in Cancer Drug Discovery: Inhibiting the SWI/SNF Chromatin Remodelling Complex. *Chembiochem : a European journal of chemical biology*, 17(8), pp.677–82. Available at: http://doi.wiley.com/10.1002/cbic.201500565 [Accessed September 28, 2017].

Zuber, J. et al., 2011. An integrated approach to dissecting oncogene addiction implicates a Myb-coordinated self-renewal program as essential for leukemia maintenance. *Genes & development*, 25(15), pp.1628–40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21828272 [Accessed August 4, 2017].