Enhancers and super-enhancers in human disease and therapy

By

Heather Ashley Hoke

B.A. Integrated Science, Genetics and Molecular Biology
Northwestern University, 2009

Submitted to the Department of Biology in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy in Biology

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2014

© 2014 Massachusetts Institute of Technology. All rights reserved.

Signature of Author: ____________________________

Signature redacted

Signature of Author: ____________________________

Signature redacted

Certified by: ____________________________

Richard A. Young
Professor of Biology
Thesis Supervisor

Signature redacted

Accepted by: ____________________________

Michael T. Hemann
Associate Professor of Biology
Chairperson, Biology Graduate Committee
Enhancers and super-enhancers in human disease and therapy

By

Heather Ashley Hoke

Submitted to the Department of Biology on May 22, 2014 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Biology

Abstract

The human body is made up of a diverse array of cell types, each with specialized properties and functions that support the organism as a whole. Despite this variability, with few exceptions, these cells contain the same genetic information. The incredible diversity in cellular function arises from differences in gene expression between cell types. Regulation of gene expression involves complex interactions between transcription factors and cofactors and the transcriptional machinery. These interactions occur both at the core promoter and at distal regulatory regions, enhancers, which are looped into close physical proximity with the core promoter. Enhancer regions are typically utilized in a cell-type-specific manner and help to drive the distinct gene expression programs that define diverse cell identities. New technologies have allowed greater ability to map enhancer regions in a variety of cell types in both healthy and diseased tissue. It is becoming increasingly apparent that misregulation of enhancer regions is a major component of many human diseases, including cancer. In addition, the availability of new classes of small molecule inhibitors targeting enhancer-bound transcriptional cofactors suggests that disruption of enhancer function may be an important therapeutic strategy in disease. Identification of super-enhancers, clusters of enhancer regions that are occupied by exceptional levels of many transcriptional coactivators, may further inform our understanding of human development and disease. These super-enhancers are associated with genes that control and define cell state, including many important disease-associated factors. In addition, super-enhancers are particularly vulnerable to perturbation by small molecules targeting enhancer-associated factors. This suggests that super-enhancers may be both biomarkers for disease-critical genes, and Achilles heels, allowing these genes to be targeted by small molecule therapies.

Thesis supervisor: Dr. Richard A. Young

Title: Professor
Acknowledgements

First and foremost, I’d like to thank my advisor Rick Young for always challenging me to pursue the most exciting and interesting projects. His confidence in me has motivated me to push myself and become the scientist that I am today.

It has been my great pleasure to work with the many talented scientists in the Young lab. In particular, I’d like to thank Tony Lee; I couldn’t have asked for a better bay-mate, and will be forever thankful for my bench assignment. His help and advice has been invaluable. I’d also like to thank Jakob Lovén and Charles Lin, who I worked closely with for the main part of my thesis research and have been fantastic colleagues and mentors.

Finally, I’d like to thank my family and friends for their love and support throughout the past five years. My parents have been a huge source of support during my graduate career, and I am so thankful that they were always there for me, both to help me through tough times, and to celebrate happy moments. I am especially proud of my mother, who as a non-scientist, took it upon herself to learn basic biology so that she could better understand my work. Last, I’d like to thank my fiancé Sean for his immeasurable support, love, and patience as I finished my degree.
# Table of Contents

Title page ..................................................................................................................................... 1

Abstract ....................................................................................................................................... 2

Acknowledgements ............................................................................................................. 3

Chapter 1: Introduction ........................................................................................................ 5

Chapter 2: Selective inhibition of tumor oncogenes by disruption of super-enhancers ............. 62

Chapter 3: Conclusions and future directions ........................................................................... 111

Appendix A: Supplementary material for Chapter 2 ................................................................ 124

Appendix B: Enhancer decommissioning by LSD1 during embryonic stem cell differentiation ......................................................................................................................... 157
Chapter 1

Introduction

Preface

Transcriptional enhancer regions play a critical role in driving the cell-type-specific gene expression programs that underlie the incredible diversity in cell function within the human body. These regions are often misregulated in disease, whether through the influence of human sequence variants, somatic mutations, or misregulation of the transcription factors and coactivators that act through enhancer regions. In chapter one of this thesis, I describe the basic stages of the transcriptional cycle, and the transcription factors, cofactors and chromatin regulators that control transcription through their actions at core promoters and enhancers. I then discuss the basic properties of enhancer regions and the identification of super-enhancers, large clusters of transcriptional enhancers that are associated with key cell-type-specific genes. Next, I describe the involvement of enhancer and super-enhancer regions in human disease, and in particular, cancer. Lastly, I discuss the recent development of many inhibitors targeting enhancer-associated transcriptional regulators. Chapter two of this thesis describes the association of super-enhancers with key oncogenes in cancer, and a mechanism by which inhibition of a general transcriptional cofactor, BRD4, can lead to a selective effect on gene expression through the disruption of super-enhancer regions. In chapter three, I offer concluding thoughts on the use of super-enhancers as biomarkers, identifying key cell-state-defining or disease-related genes, and the generality of super-enhancer disruption by transcriptional inhibitors as a therapeutic strategy in human disease.
Transcription regulation

The transcription cycle

The transcription cycle progresses through several stages, each of which is subject to regulation (Reviewed in (Berger, 2007; Kouzarides, 2007; Lee and Young, 2000; Li et al., 2007; Saunders et al., 2006; Shandilya and Roberts, 2012)). Transcription begins by the recruitment of factors to the core promoter, including the RNA Polymerase II holoenzyme (RNA Pol II), general transcription factors (GTFs), and other cofactors such as the Mediator complex. This group of factors is known as the pre-initiation complex (PIC). Following the formation of the PIC at the transcription start site (TSS), transcription generally proceeds for about 25-60 bases before becoming stalled. Release of polymerase from the paused state requires the recruitment of additional factors, leading to the phosphorylation of several pause-factors as well as Pol II itself. After pause release Pol II continues into productive elongation through the gene body. RNA processing occurs concomitantly with transcriptional elongation, and many protein factors are involved in facilitating RNA capping, splicing of introns, and finally cleavage and polyadenylation at the 3' end of the transcript. Transcription usually continues several kilobases after the polyadenylation signal before terminating, and releasing the Pol II from DNA.

Gene expression is influenced by both the core promoter region of a gene, and distal enhancer regions. The core promoter is the region immediately flanking the transcription start site of a gene, while enhancer elements can be located many megabases away from the gene that they control. Both core promoters and enhancers contain binding sites for sequence specific transcription factors that recruit transcriptional cofactors and the transcription machinery. The
many interactions between promoter- and enhancer-bound proteins also mediate looping between the core promoter and enhancer, bringing the regions into close physical proximity, despite their linear distance in the genome. Transcription factors at enhancers and promoters influence transcription through the recruitment of both activating and repressive cofactors that can influence several stages of the transcriptional cycle (Bulger and Groudine, 2011; Ong and Corces, 2011; Shlyueva et al., 2014; Spitz and Furlong, 2012). The following section will provide a brief description of the initiation and elongation stages of the transcriptional cycle, and how these are regulated by the recruitment of key factors to core promoter and enhancer regions.

Initiation

The transcription cycle begins with the assembly of the PIC, which includes RNA Pol II, associated general transcription factors (GTFs), and the Mediator complex (Lee and Young, 2000). Mediator is a large multisubunit complex that acts as a hub, bringing transcription factors, transcriptional coactivators, and signaling pathway components into proximity with the RNA Pol II holoenzyme (Borggre and Yue, 2011; Conaway and Conaway, 2011; Komberg, 2005; Malik and Roeder, 2010; Taatjes, 2010). The many subunits comprising the Mediator complex allow for interactions with a wide variety of sequence-specific transcription factors. Mediator can also interact with a variety of coactivator molecules, including GTFs. In addition, Mediator binds directly to the c-terminal domain (CTD) of Pol II, providing a link between these transcription factors, coactivators, and the RNA Pol II holoenzyme, contributing to the formation of the PIC. Mediator also facilitates looping between enhancer regions and the core promoter through its interaction with Cohesin, a protein complex also involved in the cohesion of sister chromatids (Kagey et al., 2010). In addition, Mediator interacts with additional cofactors that can influence later states of the transcriptional cycle, such as pause release.
Elongation

During transcriptional activation, the progression of RNA Pol II is typically stalled shortly after initiation. This pausing of Pol II just downstream of the TSS appears to be a major regulatory feature in gene expression, with a majority of mammalian genes occupied by paused polymerases (Core and Lis, 2008; Core et al., 2008; Guenther et al., 2007; Rahl et al., 2010). This pausing is mediated by several factors, including NELF (Negative elongation factor) and DSIF (DRB-sensitivity-inducing factor) (Cheng and Price, 2007; Yamaguchi et al., 2013). The transition from paused to elongating Pol II requires the action of the positive transcription elongation factor, P-TEFb. P-TEFb consists of the cyclin-dependent kinase Cdk9, and its partner, cyclin T1, and exists in two forms, an active complex, and an inactive one bound by the inhibitory factors HEXIM1 and the 7SK snRNP. Active P-TEFb promotes elongation through phosphorylation of the Pol II C-terminal domain (CTD) at serine 2, as well as phosphorylation of the negative transcription elongation factors, DSIF and NELF (Peterlin and Price, 2006).

BRD4, a transcriptional coactivator that associates closely with the Mediator complex, is involved in Pol II pause release through its interaction with P-TEFb (Dawson et al., 2011; Dey et al., 2000; Jang et al., 2005; Jiang et al., 1998; Wu and Chiang, 2007; Yang et al., 2005). Through its two bromodomains, BRD4 can bind to acetylated lysine residues on histones, as well as other acetylated proteins, including transcription factors (Huang et al., 2009; LeRoy et al., 2008; Rahman et al., 2011). In addition to these bromodomain-mediated interactions, BRD4 can associate with P-TEFb through a separate C-terminal domain (Jang et al., 2005; Yang et al., 2005). This binding event leads directly to the dissociation of the inhibitory factors HEXIM1 and the 7SK snRNP from P-TEFb, rendering it active and able to promote pause release (Krueger et al., 2010). BRD4 is closely associated with Mediator, and was, in fact, first
identified as an interaction partner of the Mediator complex in murine cells (Jiang et al., 1998). This interaction has been further confirmed in human cells, where Mediator and BRD4 co-occupy both promoter and enhancer regions, together facilitating the recruitment of P-TEFb and release of paused RNA Pol II into productive elongation (Dawson et al., 2011; Loven et al., 2013; Wu and Chiang, 2007).

Additional factors involved in elongation control form a large complex, the super elongation complex (SEC), which has also been implicated in the recruitment of P-TEFb to active genes (Luo et al., 2012; Smith et al., 2011). The SEC is comprised of several elongation factors, including eleven-nineteen lysine-rich leukemia (ELL) proteins, mixed lineage leukemia translocation partners (MLLs), and the P-TEFb complex (Luo et al., 2012). The SEC component ELL3 has been shown to occupy primarily enhancer regions in mouse embryonic stem cells (mESCs), where it may play a role both in facilitating interactions between the enhancer and promoter regions, and in recruiting other SEC components to promote gene activation in later lineages (Lin et al., 2013). Lastly, some transcription factors appear to be involved in elongation control. For example, the MYC transcription factor and proto-oncogene interacts directly with P-TEFb and appears to be involved in mediating transcriptional pause release in a genome-wide manner (Eberhardy and Farnham, 2001, 2002; Gargano et al., 2007; Kanazawa et al., 2003; Rahl et al., 2010). Although MYC is primarily associated with promoter regions, overexpression of MYC can lead to “enhancer invasion”, wherein high levels of MYC lead its association with lower-affinity binding sites at both enhancer and core promoter regions (Lin et al., 2012).

Divergent transcription

One emerging property of transcription is that it occurs in a bidirectional, divergent manner at most mammalian genes (Core et al., 2008; Preker et al., 2008; Seila et al., 2008;
In addition to transcription in the sense, often protein-coding direction, a majority of genes also produce relatively short (50-2,000 nucleotides), unstable antisense transcripts upstream of the promoter (Flynn et al., 2011; Ntini et al., 2013; Preker et al., 2011; Preker et al., 2008). Antisense transcription shares many features with sense transcription, including recruitment of the elongation factor P-TEFb (Flynn et al., 2011). The factors that control promoter directionality are not well understood, but several mechanisms may be at play. First, a minority of human promoter regions contain a TATA element, a motif bound by the TATA-binding protein TBP, which is part of the TFIID GTF involved in the recruitment of RNA Pol II. The sequence-specific binding of TBP can help to direct more precise recruitment of RNA Pol II to the core promoter region, while at TATA-less genes (the majority of human genes), less-specific protein-protein interactions with transcription factors recruit TBP both upstream and downstream of these transcription factors binding sites, contributing to bidirectional transcription (Core and Lis, 2008; Core et al., 2008; Wu and Sharp, 2013). Other factors that may influence transcription directionality include chromatin remodeling, histone deacetylation, and gene-loop formation between the promoter and 3’ region of the gene (Churchman and Weissman, 2011; Tan-Wong et al., 2012; Whitehouse et al., 2007). Control may also occur by recognition of the antisense RNA species by the machinery that typically controls the addition of poly A tracts. The poly A signal motif (PAS) or similar sequences are enriched in the upstream antisense region, and recognition of this sequence can lead to cleavage of the RNA and termination of transcription (Almada et al., 2013; Ntini et al., 2013; Proudfoot, 2011). Divergent transcription may help to increase expression of the sense transcript, for example, by increasing negative supercoiling and DNA unwinding at promoters. Upstream transcription can also lead to an increased propensity for mutagenesis, potentially contributing to
the evolution of new genes (Wu and Sharp, 2013). This divergent transcription also occurs at
eenhancer regions, and may play an important role in enhancer function (discussed later) as well
as the evolution of new genes from enhancer regions (De Santa et al., 2010; Kim et al., 2010;
Sigova et al., 2013; Wu and Sharp, 2013).

Control of transcription by chromatin regulators

In addition to the mechanisms of regulation described above, transcription can also be
influenced by the chromatin state surrounding enhancers and promoters, and by the chromatin
modifying enzymes that act on histone proteins. Genomic DNA does not exist as a naked
molecule, but instead exists as chromatin, consisting of both DNA and closely associated
proteins, histones. Chromatin is comprised of units called nucleosomes: complexes containing
roughly 150 basepairs of DNA wrapped around positively charged histone proteins (Kornberg
and Lorch, 1999).

Histones can influence transcription both by their position, where they may create a
physical barrier to RNA Pol II initiation or elongation, and by posttranslational modifications
that can influence the recruitment of many transcriptional cofactors. To facilitate access of RNA
Pol II to histone-occupied DNA, transcription factors can recruit ATP-dependent chromatin
remodeling complexes, such as the SWI/SNF complex. Nucleosome remodeling complexes can
facilitate, or hinder, transcription, either by sliding histones along the DNA, ejecting them, or
mediating a switch between different histone variants (Hargreaves and Crabtree, 2011; Ho and
Crabtree, 2010).

Histone proteins are the targets of many posttranslational modifications, many of which
occur on the histone tail, 20-40 amino acid segments that extend beyond the main mass of the
nucleosome. Histone modifying enzymes can perform a wide variety of post-translational
modifications, including acetylation, methylation, phosphorylation, and ubiquitination, that are involved in both gene activation and repression. In addition to the "writers" that deposit these modifications, "erasers" such as histone deacetylases or demethylases can remove these marks. "Reader" proteins can bind to specific histone marks, resulting in the recruitment of various regulators of gene expression. Histone modifications are associated with different gene activity states, depending both on the type of modification and the position of the modification on the histone.

Lysine methylation can be associated with both activation and repression of transcription, depending on the nature of the modification. Methylation of histone H3 lysine 4 (H3K4) is generally considered to be a mark of active transcription, with trimethylation (H3K4me3) occurring at promoter regions, while monomethylation (H3K4me1) occurs primarily at enhancer regions. Interestingly, H3K4me1 is present at both active enhancers as well as developmental enhancers that appear to be primed for activation in later developmental stages (Creyghton et al., 2010; Rada-Iglesias et al., 2011; Zentner et al., 2011). Several histone modifications are associated with transcriptional elongation: H3K79me2, which occurs primarily in the 5' region of elongating genes, and H3K36me3, which occurs more 3' in the gene body (Bernstein et al., 2012).

Several histone methylation events are associated with inactive or repressed genes, with different methylation patterns characteristic of different mechanisms of repression (Beisel and Paro, 2011; Cedar and Bergman, 2012; Jones, 2012; Moazed, 2009; Reyes-Turcu and Grewal, 2012). Trimethylation of histone H3 lysine 27 (H3K27me3) is catalyzed by the EZH2 subunit of the Polycomb complex. Genes marked by H3K27me3 are often silent, but poised for activation in later developmental stages. Many of these Polycomb-repressed genes are also marked by
H3K4me3, and are referred to as “bivalent” genes, as they have marks associated with both gene activity and repression (Orkin and Hochedlinger, 2011; Young, 2011). Genomic regions that are more permanently repressed are often associated with trimethylation of H3K9 (H3K9me3), which can be deposited by several histone methyltransferases, including SETDB1, SUV39H1, SUV39H2, EHMT1 (GLP), and EHMT2 (G9A). These regions tend to be gene-poor and contain repetitive elements or retrotransposons (Feng et al., 2010; Lejeune and Allshire, 2011; Mikkelson et al., 2007).

Histone acetylation can occur at many different lysine residues in the histone tail and core and is generally associated with activation of gene expression. These marks are deposited by histone acetyltransferases (HATs) such as p300, GCN5, or TIP60, and removed by histone deacetylases (HDACs). Acetylation of histone H3 lysine 27 (H3K27ac) is associated with active enhancer regions (Bonn et al., 2012; Creyghton et al., 2010; Rada-Iglesias et al., 2011; Zentner et al., 2011). In contrast to H3K4me1, which also occurs at inactive, primed enhancers, H3K27ac appears to be exclusively a mark of active enhancers, making a distinction between these classes of enhancer regions. In keeping with this model, during development, H3K27ac is acquired almost exclusively at regions that are pre-marked by H3K4me1 (Bonn et al., 2012). Unlike lysine methylation, lysine acetylation removes the positive charge associated with the lysine residue, potentially affecting the association of the histone protein with negatively charged DNA in addition to creating a binding site for “reader” proteins.

Interestingly, many active genes are occupied by chromatin modifying enzymes with opposing activities. For example, although histone deacetylation has long been associated with repression of transcription, HDACs tend to be associated with actively transcribed genes (Kurdistani et al., 2002; Wang et al., 2002; Wang et al., 2009). HDACs 1 and 2 exist in several
multisubunit complexes, including the NuRD complex, which also contains the histone
demethylase LSD1, which is responsible for removing the active enhancer mark H3K4me1.

Work from our lab (included as Appendix B) indicates that LSD1 and HDACs 1 and 2 cooperate
at active enhancers to remove H3K4me1 and histone acetyl marks upon differentiation, allowing
for deactivation of pluripotency genes (Whyte et al., 2012). LSD1’s demethylase activity is
hindered by the presence of acetylated histones (Forneris et al., 2005; Lee et al., 2006). We
therefore proposed a mechanism in which dynamic acetylation and deacetylation by HATs and
HDACs results in the maintenance of a level of acetylation that is inhibitory to LSD1
demethylation. Upon differentiation, when HATs are no longer recruited to these enhancers, the
presence of HDACs results in deacetylation, followed by demethylation of the enhancers by
LSD1.

In addition to the writers and erasers that deposit and remove histone modifications,
“reader” proteins can recognize histone marks and lead to the recruitment of transcriptional
regulators to specific chromatin regions. As mentioned previously, BRD4 is an example of the
bromodomain class of acetyl-lysine binding proteins, and is involved in the recruitment of P-
TEFb to active enhancers and promoters marked by acetylated histones. Methylated histone
binding domains include the chromodomain, the WD40 repeat, and the PHD finger, which is
associated with methyl-lysine residues, and the Tudor domain, which can bind both methylated
lysines and arginine (Musselman et al., 2012). One example of a chromodomain protein is
heterochromatin protein 1 (HP1), which recognizes H3K9me3 and facilitates chromatin
compaction at these repressed regions (Bannister et al., 2001; Lachner et al., 2001; Nakayama
and Takami, 2001). At enhancers, the TIP60 acetyltransferase is recruited in part through its
recognition of H3K4me1 through its chromodomain (Jeong et al., 2011).
Conclusion

Cells can achieve specific gene expression programs by controlling transcription through the influence of transcription factors, coactivators, and chromatin regulators. Many of these transcriptional regulators are associated with distal enhancer regions which are looped into close physical proximity with the core promoter region. The following sections describe the functional properties of transcriptional enhancers in greater detail, and discuss how recent improvements in our ability to map these enhancer regions have resulted in a greater understanding of enhancer function in human development and disease.

Transcriptional enhancers

Although transcription requires only a core promoter, enhancers can greatly increase the level of transcription from target genes, and, through binding of cell type-specific transcription factors, can influence the timing and spatial distribution of gene expression within a cell or organism. Early studies established these basic properties of enhancer function, and new genome-wide technologies have extended our knowledge by allowing the mapping of enhancer regions across many cell and tissue types. Estimates indicate that there may be over a million enhancer regions encoded in the human genome (Bernstein et al., 2012). Investigating how these enhancer regions mediate cell-type specific gene expression programs is critical to understanding human development as well as disease. My thesis research led to the identification of a class of enhancers, super-enhancers that consist of large clusters of individual enhancer elements. Super-enhancers are associated with genes that control and define cell state, and thus may act as biomarkers for identifying these key genes.

Historical identification of enhancer elements
The first enhancer element identified was a 72 base pair sequence from the SV40 viral genome. This viral sequence was capable of inducing a two hundred fold increase in the expression of a reporter gene in human HeLa cells (Banerji et al., 1981). The enhancer region retained its function when placed at different distances and orientations to the reporter gene; it was able to enhance expression to a similar degree whether located proximal to the promoter or several kilobases upstream or downstream of the gene (Banerji et al., 1981). Enhancers from additional animal viruses were soon identified and described (de Villiers et al., 1982; Hansen and Sharp, 1983; Schirm et al., 1985; Spandidos and Wilkie, 1983). Some of these enhancer elements showed host-specificity, for example, an enhancer from the mouse polyoma virus was most effective in mouse tissue, rather than primate tissue (de Villiers et al., 1982; Spandidos and Wilkie, 1983).

Subsequently, enhancer elements were identified in metazoan genomes (Banerji et al., 1983). One of the first endogenous mammalian enhancers described was a sequence located in the immunoglobulin heavy chain locus (IGH) that acts as a transcriptional enhancer in reporter assays. The enhancer activity of the IGH enhancer regions was found to be cell type specific; the enhancer functioned in myeloma cells (a B-lymphocyte-derived tumor), but not HeLa cells (derived from a cervical carcinoma) (Banerji et al., 1983; Davidson et al., 1986). In another early example; several enhancer regions were found to control expression of the β-globin gene in a cell-type, and developmental stage specific manner, first in chicken, and later in human cells (Antoniou et al., 1988; Behringer et al., 1987; Choi and Engel, 1986; Hesse et al., 1986; Kollias et al., 1987; Trudel and Costantini, 1987).

Further studies, using various techniques both in vitro and in vivo, indicated that enhancer function depended on the binding of trans-acting protein factors (Augereau and
Chambon, 1986; Davidson et al., 1986; Lee et al., 1987; Mercola et al., 1985; Sassone-Corsi et al., 1984; Scholer and Gruss, 1985; Sen and Baltimore, 1986; Sergeant et al., 1984; Wildeman et al., 1984). These studies also suggested that these trans-factors bound to specific sequences in the enhancer element, and that some of these factors were only present in specific cell types, perhaps accounting for the cell-type-specificity of certain enhancers, such as the IGH or β-globin enhancers (Lee et al., 1987; Mercola et al., 1985; Scholer and Gruss, 1985; Sen and Baltimore, 1986; Wildeman et al., 1984). For example, DNase footprinting experiments, in which transcription factor binding sites are identified based on their protection from DNase I activity, identified four erythroid-specific binding sites in an enhancer in the 3' region of the β-globin gene (Wall et al., 1988). Gel-mobility shift assays revealed that these sites were bound by an erythroid-specific protein, Nuclear Factor Erythroid 1 (NF-E1), later identified and characterized as GATA-1, a transcription factor that is essential for erythroid development (Ferreira et al., 2005; Wall et al., 1988).

**Transcription factor binding**

This binding of sequence-specific transcription factors is a key feature of enhancer regions (reviewed in (Levine, 2010)). As described above, the binding of transcription factors to enhancer regions and subsequent recruitment of the RNA Pol II transcription apparatus and cofactors is a critical step in the activation of gene expression. Enhancers typically contain binding motifs for many distinct classes of sequence-specific transcription factors. Protein-protein interactions between these factors can act in a cooperative manner, increasing their apparent affinity for the enhancer element. Binding of one transcription factor may also indirectly facilitate the binding of another, by altering DNA conformation, or by recruiting cofactors, for example a nucleosome remodeling complex, that mediates the binding of a second
factor. In addition, these transcription factors can act synergistically to recruit transcriptional coactivators or RNA Pol II. These synergistic interactions can result in greater transcriptional activation than could be achieved by independently acting factors (Carey et al., 1990; Giese et al., 1995; Kim and Maniatis, 1997; Thanos and Maniatis, 1995). In addition, the convergence of many transcription factors to an enhancer region contributes to the control of the precise temporal and spatial regulation of gene expression, by requiring a specific combination of transcription factors for gene activation (Ptashne and Gann, 1998). One particularly well-studied enhancer is the β-interferon “enhanceosome”, which is activated by a combination of several transcription factors, c-Jun, ATF2, IRF3, IRF7, and NF-κB (Agalioti et al., 2000; Merika and Thanos, 2001; Thanos and Maniatis, 1995). Crystal structures of this complex reveal the many protein-DNA and protein-protein interactions that contribute to the function of this enhancer (Panne et al., 2007).

**Enhancer promoter looping**

Communication between distal enhancer regions and the core promoter through looping is another critical feature of enhancer function. Looping interactions between different DNA regions were first observed in bacterial systems (Adhya, 1989; Bulger and Groudine, 1999; Matthews, 1992; Ptashne, 1986; Saiz and Vilar, 2006; Schleif, 1992). For example, nitrogen regulatory protein C (NtrC) bound at upstream enhancer regions in the glnA operon can interact with RNA polymerases at the promoter regions, forming a DNA loop (Su et al., 1990). DNA looping interactions can be detected by chromosome conformation capture assays, which rely on proximity-based ligation of DNA fragments, identifying genomic regions that are physically close, regardless of linear distance in the genome (Dekker et al., 2002). These experiments have demonstrated looping interactions between many individual enhancers and promoters in human
cells, including the β-globin enhancers and promoter (Deng et al., 2012; Tolhuis et al., 2002; Vakoc et al., 2005). These looping interactions are likely critical for gene activation; at the β-globin locus, forced looping by tethering of a looping factor was sufficient to activate gene expression (Deng et al., 2012).

Loop formation is stabilized by interactions between the Mediator and cohesin complexes. The cohesin complex, which is also involved in sister chromatid cohesion, is a ring-shaped complex. Although it is unclear precisely how cohesin facilitates looping, the ring structure has the required dimensions to encircle two strands of nucleosome-occupied DNA and it is possible cohesin stabilizes looping interactions in this manner. Cohesin is loaded on to DNA by NIPBL, a protein that is associated with the Mediator complex. Both Mediator and cohesin globally co-occupy cell-type-specific enhancer and promoter regions, likely helping to drive the cell-type-specific expression program by promoting enhancer-promoter loops (Kagey et al., 2010). Global mapping of enhancer-promoter looping interactions is possible through the ChIA-PET technique (chromatin interaction analysis by paired end tag sequencing). ChIA-PET combines chromatin immunoprecipitation and proximity-based ligation to identify loops between regions bound by a particular protein factor. ChIA-PET has been carried out in several systems, identifying enhancer-promoter loops that are highly cell-type-specific, and correlated with transcription levels (Chepelev et al., 2012; Fullwood et al., 2009; Li et al., 2012; Zhang et al., 2013). Given that enhancers can act at large genomic distances from their target genes, identifying the target gene of an enhancer is often difficult, and greater ability to map the physical interactions between promoters and enhancers will improve our ability to understand how enhancers influence gene expression in a given cell type.

*Enhancer RNA*
One emerging property of enhancers is the production of RNA from these regions (reviewed in (Lai and Shiekhattar, 2014; Lam et al., 2014; Orom and Shiekhattar, 2013)). Recent genome-wide sequencing of RNA species has revealed that enhancers are transcribed, producing non-coding RNAs (De Santa et al., 2010; Kim et al., 2010). The level of RNAs produced from enhancer regions is correlated with the expression of nearby genes, and in several systems, induction of protein coding gene expression was preceded by an increase in the production of enhancer RNAs, suggesting that the transcription of RNA from enhancers may play a role in regulating the expression of protein coding genes (Hah et al., 2013; Kim et al., 2010; Koch and Andrau, 2011). There are several hypotheses as to the functional significance of the transcription of enhancer regions. This transcription may be only noise, simply a consequence of the localization of the Pol II transcription apparatus to the looped promoter/enhancer region. Alternatively, the act of transcription itself may facilitate the maintenance of open chromatin at enhancer regions. Recently, evidence has suggested that the RNA produced by enhancer transcription may have a functional role itself. Several studies have indicated that depletion of enhancer RNAs can lead to reduced expression of nearby genes (Lam et al., 2013; Melo et al., 2013; Orom et al., 2010).

The mechanism by which enhancer RNA contributes to enhancer function remains unclear, but many recent studies suggest that RNA may interact with many enhancer-bound proteins, stabilizing their association with the enhancer, and looping to the promoter region. In these studies, depletion of enhancer RNA led to decreased looping interactions between enhancers and promoters, and decreased occupancy of the Mediator coactivator complex (Lai et al., 2013; Li et al., 2012). Evidence suggests that the Mediator complex may interact directly with RNA, and given that many DNA-binding transcription factors also have RNA-binding
capabilities, these physical interactions between enhancer RNA and enhancer-associated protein factors may play a key role in enhancer function (Burdach et al., 2012; Lai et al., 2013; Ng et al., 2013).

**Genome-wide identification of enhancer elements**

With the completion of the human genome project, and ever lowering costs of genomic sequencing, we have been increasingly able to obtain global maps of enhancer regions in many cell types. The work of individual laboratories, as well as large consortiums, such as the ENCODE project, have contributed to the identification of genomic regulatory regions across a wide variety of cell types (Bernstein et al., 2012). Several strategies exist to identify enhancer regions. As described above, enhancers are occupied by a transcription factors, as well as many transcriptional coactivators and the transcriptional apparatus. By determining the genomic localization of these enhancer-bound factors, we can generate a genome-wide map of enhancer regions. One strategy for determining genomic binding sites is chromatin immunoprecipitation followed by massively parallel sequencing (ChIP-seq). In this method, chemical crosslinking is used to link protein factors to their DNA binding sites. Chromatin is then sheared, and the protein factors and their accompanying DNA fragments are immunoprecipitated using an antibody to the desired DNA-associated protein factor. The immunoprecipitated DNA is then purified and sequenced, giving the genome-wide binding sites for the desired factor. Several alternate strategies can be used, including ChIP-exo, in which exonuclease digestion is used to generate DNA fragments centered around the protein bound, protected regions (Rhee and Pugh, 2011). In DNA adenine methyltransferase identification (DamID), the protein of interest is fused to DNA adenine methyltransferase, and these methylated regions are amplified and sequenced (Luo et al., 2011; Wu and Yao, 2013).
When the transcription factors for a particular cell type are known, they can be used to map enhancer regions. For example, in mouse embryonic stem cells (mESC), Oct4, Sox2 and Nanog have been described as master transcription factors (Ng and Surani, 2011; Orkin and Hochedlinger, 2011; Young, 2011). ChIP-seq of Oct4, Sox2 and Nanog can identify putative enhancer regions that are bound by these three factors. Function testing of these regions reveals that binding of the three transcription factors is predictive of enhancer function in a reporter assay (Chen et al., 2008). In addition to transcription factors, enhancers are occupied by many transcriptional cofactors, including the Mediator complex, the bromodomain containing protein BRD4, and the histone acetyltransferase (HAT) p300. ChIP-Seq of each of these factors has been used to identify enhancer regions, and may be particularly useful in cases where the cell-type-specific transcription factors are unknown (Chapuy et al., 2013; Heintzman et al., 2007; Kagey et al., 2010; Rada-Iglesias et al., 2011; Visel et al., 2009).

The binding of transcription factors and associated cofactors creates areas of open chromatin at enhancers, a feature that can be exploited in enhancer identification. One such technique, DNase hypersensitive site sequencing (DNase-Seq), involves digestion of accessible DNA regions by DNase I, followed by sequencing of these hypersensitive regions (Crawford et al., 2006; John et al., 2013; Song and Crawford, 2010). In addition, several newer techniques have been developed based on chromatin accessibility. FAIRE-Seq (Formaldehyde-assisted identification of regulatory elements) uses the fact that formaldehyde crosslinking is more efficient at nucleosome occupied regions to separate these from nucleosome free regions, which are then sequenced (Giresi et al., 2007; Simon et al., 2012). ATAC-seq, (assay for transposase accessible chromatin) utilizes the transposase Tn5 to insert sequencing adapters in to open chromatin regions (Buenrostro et al., 2013).
Although enhancer regions bound by transcription factors are typically depleted of nucleosomes, the nucleosomes immediately surrounding transcription factor binding sites in enhancers often carry stereotypical modification that can be used to identify these regions. As described previously, promoters tend to be trimethylated at histone H3 lysine 4 (H3K3me3); enhancer regions are often monomethylated at this residue (H3K4me1). This H3K4me1 mark occurs both at active enhancers and at enhancer regions that appear to be primed, and are utilized in later lineages. Acetylation of H3K27 appears to mark primarily active enhancers (Creyghton et al., 2010; Rada-Iglesias et al., 2011). Identification of enhancers using histone modifications has been widely used, including by the ENCODE project, to map enhancer regions in a wide variety of organisms and cell types (Bernstein et al., 2012; Bonn et al., 2012; Ernst et al., 2011; Kharchenko et al., 2011; Shen et al., 2012; Wamstad et al., 2012).

In addition to these enhancer-mapping strategies, several new methods for functional identification of enhancers have emerged in recent years. Traditionally, as with the discovery of the SV40 enhancer, enhancer function is confirmed using reporter assays testing the ability of a candidate regulatory sequence to increase the transcription of a reporter gene. Recently, this strategy has been adapted to be used in a high throughput, genome-wide manner, allowing global identification of functional enhancer elements. One such strategy, STARR-seq (self-transcribing active regulatory region sequencing) relies on an enhancer's ability to drive expression in a position- and orientation-independent manner by cloning putative enhancers downstream of a minimal promoter. In this assay, regulatory regions that enhance their own transcription will produce higher levels of corresponding RNA, which is measured by sequencing (Arnold et al., 2013).

**Super-enhancers**
My thesis research, described in chapter two, led to the identification of super-enhancers. Super-enhancers are clusters of transcriptional enhancers, that span large regions of the genome and are occupied by high levels of transcription factors and transcriptional coactivators, as well as the Pol II transcription apparatus itself. Work by others in the Young lab described super-enhancers in mouse embryonic stem cells (mESCs). Here, enhancers were identified as regions occupied by the mESC master transcription factors Oct4, Sox2, and Nanog, determined by ChIP-seq. While most mESC enhancer regions were a few hundred base pairs in size, a small subset of regions were made up of clusters of enhancers that spanned large regions of the genome, up to 50kb. These super-enhancers could be separated from smaller typical enhancers based on their occupancy by exceptional amounts of the Mediator coactivator (Whyte et al., 2013). Out of over 8,000 enhancer regions in mESCs, 231 were identified as super-enhancers, regions whose median length is over an order of magnitude greater than typical enhancers, and which are occupied by an order of magnitude greater level of Mediator. The high level of occupancy of RNA Pol II and coactivators at super-enhancers is concomitant with higher expression levels of genes associated with super-enhancers when compared to genes associated with typical enhancers.

Examination of the genes associated with super-enhancers revealed that many of these genes are known to be important in mESC biology, including Oct4, Sox2 and Nanog themselves, as well as other well-known factors, such as Klf4, Esrrb, and the mir290 cluster. Extending this analysis to other mouse tissues, super-enhancers were again found to be associated with cell-type-specific genes, many which are known to play key roles in defining a particular tissue type. For example, the muscle master transcription factor, MyoD, is associated with a super-enhancer in mouse skeletal muscle myoblasts, while the B-cell-specific transcription factor Pu.1 is
associated with a super-enhancer in mouse B-cells. Therefore, we believe that super-enhancers play a role in sustaining high levels of transcription of key genes that function in defining cell state, across a variety of cell types.

As described above, enhancers function through the cooperative interactions between many transcription factors and cofactors (Carey, 1998; Carey et al., 1990; Giese et al., 1995; Kim and Maniatis, 1997; Thanos and Maniatis, 1995). Enhancers bound by many cooperatively interacting factors can lose activity more rapidly than enhancers with fewer binding sites when the levels of enhancer-bound factors are reduced (Giniger and Ptashne, 1988; Griggs and Johnston, 1991). Super-enhancers, comprising of clusters of enhancer regions highly occupied by transcriptional coactivators, are indeed more sensitive to the loss of the enhancer-associated factors Oct4 and Mediator. Knockdown of these proteins resulted in a greater loss in expression of super-enhancer-associated genes compared to genes associated with typical enhancers. This sensitivity to perturbation may facilitate gene expression changes during development, when a cell must decommission cell-type-specific enhancers to make way for a new gene expression program.

Furthering our study of super-enhancers, we examined these regulatory regions in a panel of 86 human cell types, as well as several human tumor cell lines (Hnisz et al., 2013). In this study, the widely profiled enhancer-associated histone modification H3K27ac was used to identify enhancer regions, and to rank these regions based on ChIP-seq signal to differentiate super- and typical enhancers. Confirming the previous results in mouse tissue, super-enhancers in human tissue were associated with cell type-specific genes linked to biological processes that were important determinants of the cell type and function (Hnisz et al., 2013). Many key cell-type-specific transcription factors are associated with super-enhancers, possibly forming an
interconnected autoregulatory loop. For example, in mESC, the Oct4, Sox2 and Nanog genes are associated with super-enhancers, which are bound by each of these three transcription factors (Whyte et al., 2013). Therefore, identifying super-enhancers may help to identify the master transcription factors, and the regulatory network driving cell state in a variety of cell types.

Some super-enhancers overlap with other historically characterized examples of large enhancer regions, for example, the locus control regions (LCRs), corresponding to the IGH enhancer and β-globin enhancer (Forrester et al., 1990; Grosveld et al., 1987; Li et al., 2002; Madisen and Groudine, 1994). More recent global analyses have identified similar large enhancer regions. Examples include transcription initiation platforms (TIPs), characterized by binding of RNA Pol II and GTFs (Koch and Andrau, 2011); stretch enhancers, enhancer regions defined by bioinformatics analysis of chromatin state (Parker et al., 2013); and DNA methylation valleys, large regions of hypomethylated DNA (Xie et al., 2013). The identification, by different strategies, of these large cell-type-specific enhancer regions supports the idea that large, super-enhancer regions are critical for regulating genes that define and control cell state.

Conclusion

Enhancers are now well recognized as key drivers of cell-type-specific expression, helping to specify expression programs through the integration of many cooperatively acting sequence-specific transcription factors and coactivators. Other key properties of enhancers include their ability to act at great genomic distances, and without regard to the orientation of their target genes. This function is facilitated by looping interactions that bring enhancers and promoters into close physical proximity. In addition, transcription of enhancers themselves, and the RNA produced, may prove to be an important aspect of enhancer function. Recent advances allowing the genome-wide mapping of enhancer regions underscores the importance of these
elements in controlling cell-type-specific gene expression programs. Although each cell type typically contains thousands of active enhancer regions, the identification of super-enhancers may help to further define the critical genes and regulatory regions that control and define cell state.

Dysregulation of enhancer function in human disease

*Early examples of enhancer involvement in disease*

Early work in deciphering the causes of genetic disease uncovered several examples in which mutations in regulatory regions, rather than coding sequence, was associated with disease phenotypes (reviewed in (Kleinjan and Lettice, 2008; Kleinjan and van Heyningen, 2005)). For example, thalassemia, a blood disorder resulting from imbalanced levels of the oxygen-carrying factors α- and β-globin, can be caused by mutations in the protein coding regions of these genes, or by alterations that affect regulatory regions. For example, β-thalassemia can be caused by translocations that remove the β-globin locus control region, a well-characterized distal enhancer responsible for driving high level expression of the β-globin gene (Driscoll et al., 1989; Kioussis et al., 1983). In another example, point mutations or translocations affecting a regulatory region upstream of the sonic hedgehog (*SHH*) gene, an important regulator of limb and brain development, were found to be associated with inherited preaxial polydactyly (Lettice et al., 2003). Interestingly, mutations or deletions of the coding region of the *SHH* gene cause a different disease, holoprosencephaly, a developmental disorder of the brain (Belloni et al., 1996; Roessler et al., 1996; Roessler et al., 1997). The mutations causing preaxial polydactyly occur in a limb-specific enhancer, resulting in limb, rather than brain defects (Kleinjan and Lettice, 2008; Lettice et al., 2003).

*Genetic sequence variation in enhancers contributes to complex human traits and disease*
Many of these early studies investigated relatively simple traits and disorders, involving single gene, Mendelian inheritance patterns. More recently, genome-wide, high throughput studies have allowed greater ability to identify the involvement of regulatory regions in disease. Genome-wide association studies (GWAS) identify common genetic sequence variants (typically single nucleotide polymorphisms, SNPs) that are associated with complex human traits or diseases (Stranger et al., 2011). Many hundreds of GWASs have been performed, assessing a wide spectrum of traits and disorders. In these studies, the vast majority of trait- or disease-associated SNPs occur in non-coding regions of the genome (Hindorff et al., 2009; Manolio, 2010). Although many early studies focused on SNPs within coding regions, it has become apparent that sequence variation in non-coding regions plays an important role in regulating gene expression. Many disease- or trait- associated SNPs have been mapped to regions termed expression quantitative trait loci (eQTLs), genomic regions that influence the expression of a particular gene or genes (reviewed in (Cookson et al., 2009)).

With the ability to obtain global maps of regulatory regions in many cell types, it has become apparent that the majority of these non-coding SNPs occur in regulatory regions of the genome, including enhancers (Maurano et al., 2012). Further, disease-associated SNPs tend to occur in regulatory elements found in cell types that are related to disease pathology. For example, SNPs associated with diseases such as Alzheimer’s disease or bipolar disorder tend to occur in regulatory regions in brain tissue, while SNPs associated with heart-related traits and disorders, such as coronary heart disease, tend to occur in heart-specific regulatory regions (Maurano et al., 2012). The cell-type specific nature of enhancer function explains how genomic variants, which are present in all cells in the body, can cause cell- or tissue-specific phenotypes. GWA studies provide only a statistical association between a particular genome region and a trait
or disease, and it is often difficult to extend this knowledge to gain an understanding of the underlying biology resulting in phenotypic differences (Frazer et al., 2009; Fugger et al., 2012). Overlaying maps of enhancer regions in relevant cell types with disease-associated loci identified by GWAS may help identify the functional variants (Schaub et al., 2012).

Although unraveling the functional significance of disease- or trait-associated SNPs remains difficult, several cases have demonstrated that sequence variants can disrupt transcription factor binding sites within enhancer regions, resulting in altered expression of a nearby gene (Bauer et al., 2013; Schaub et al., 2012). For example, BCL11A is a well-studied regulator of fetal hemoglobin expression (Esteghamat et al., 2013; Sankaran et al., 2008; Sankaran et al., 2009; Xu et al., 2013; Xu et al., 2010). Several studies examining fetal hemoglobin expression levels identified trait-associated SNPs within the second intron of BCL11A (Bhatnagar et al., 2011; Lettre et al., 2008; Menzel et al., 2007; Nuinoon et al., 2010; Solovieff et al., 2010). These SNPs coincided with DNaseI HS regions that were also marked by typical enhancer-associated histone modifications, and bound by the transcription factors GATA1 and TAL1 in erythroid cells. Closer examination of these enhancer SNP sites revealed that this variant disrupted a binding motif recognized by GATA1 and TAL1, resulting in reduced binding of these transcription factors, concomitant with modestly altered expression of BCL11A and fetal hemoglobin (Bauer et al., 2013).

*Genetic sequence variants in enhancer elements in cancer*

Many of these GWA studies have highlighted genomic variants related to cancer. Again, many of these cancer susceptibility-related SNPs occur in regulatory regions of the genome, including promoters, regulatory regions within introns, and distal enhancers that may be many hundreds of kilobases away from the affected gene. Disease associated sequence variants in
regulatory regions, including enhancers, have been identified in a variety of tumor types, including colorectal (Broderick et al., 2007; Lubbe et al., 2012; Tomlinson et al., 2008), prostate (Demichelis et al., 2012; Haiman et al., 2007), breast (Easton et al., 2007; Jiang et al., 2011), nasopharangeal (Yew et al., 2012), renal (Schodel et al., 2012), lung (Liu et al., 2011), and melanoma (Horn et al., 2013; Huang et al., 2013).

Of particular interest in cancer, many cancer-associated SNPs are located in the 8q24 gene desert, which is located near the MYC oncogene (reviewed in (Sur et al., 2013)). Variations associated with cancer risk have been mapped to this location in a variety of tumor types, including prostate, breast, colorectal, bladder, lung, ovary, pancreas, kidney and brain (Ahmadiyeh et al., 2010; Al Olama et al., 2009; Amundadottir et al., 2006; Couch et al., 2009; Domagk et al., 2007; Fletcher et al., 2008; Ghoussaini et al., 2008; Gruber et al., 2007; Gudmundsson et al., 2013; Gudmundsson et al., 2007; Haiman et al., 2007; Kiemeney et al., 2008; Li et al., 2008; Park et al., 2008; Pomerantz et al., 2009; Poynter et al., 2007; Shete et al., 2009; Tenesa et al., 2008; Tomlinson et al., 2008; Turnbull et al., 2010; Tuupanen et al., 2009; Yeager et al., 2009; Zanke et al., 2007; Zhang et al., 2014).

In colorectal cancer, several studies have further investigated the mechanisms by which sequence variation in the 8q24 regions leads to cancer risk (Jia et al., 2009; Pomerantz et al., 2009; Tuupanen et al., 2009). In one example, ChIP assays found that the colorectal cancer predisposition SNP rs6983267 was located in an enhancer region, marked by H3K4me1 and p300 binding in a colorectal cancer cell line (Pomerantz et al., 2009). Reporter assays indicated that this region had active enhancer function, and that the risk allele led to significantly greater gene activation. Further examining the sequence context of this SNP found it was located in a binding site for the transcription factor TCF7L2 (also known as TCF4), part of the Wnt signaling
pathway, known to be important in colon cancer pathogenesis (Polakis, 2000; Pomerantz et al., 2009). The risk allele corresponds to the consensus motif for this TCF7L2, and ChIP of TCF7L2 indicated greater binding to the risk allele. Although this enhancer region is located nearly 400kb away from the MYC gene, physical interaction between the two was confirmed by chromosome conformation capture (3C) experiments. Interestingly, mice lacking this enhancer region are resistant to developing intestinal tumors, further underscoring the importance of this enhancer in the development of colon cancer (Sur et al., 2012). This data provides another example of the molecular mechanism by which sequence variation may lead to phenotypic differences, in this case, altering binding of TCF7L2, leading to differential activity of an enhancer.

**Somatic changes in enhancers in tumor cells**

In addition to these links between human sequence variation and cancer, there are several examples of somatic mutation involving enhancer regions in cancer. Many B-cell tumors have translocations linking immunoglobulin enhancer regions to oncogenes, such as MYC, and BCL2. These translocation events can be the result of aberrant activation of the immunoglobulin gene remodeling process, involving V(D)J recombination, somatic hypermutation, and class switch recombination (Aplan, 2006; Kuppers and Dalla-Favera, 2001). Burkitt lymphoma, for example, is characterized by translocations between the MYC oncogene and either the immunoglobulin heavy locus, or less frequently, the immunoglobulin light chain κ or λ loci (Boerma et al., 2009; Hecht and Aster, 2000).

Several recent studies have indicated that epigenetic changes in the enhancer landscape may play a significant role in tumor biology. One study examined the variation in enhancer regions between colorectal carcinoma and normal colon epithelial cells by ChIP-seq of
H3K4me1 (Akhtar-Zaidi et al., 2012). This identified thousands of enhancer loci that were either gained or lost in colon cancer cells, corresponding to the expression levels of nearby genes. These variant enhancer loci (VELs) were highly enriched in SNPs linked to colon cancer risk, suggesting that the changes in enhancer regions may be a key factor in colon cancer pathogenesis. Another study examined patterns of DNA methylation in cancer cells and found that hypomethylation of enhancer regions was highly correlated with upregulation of cancer-related genes, while hypermethylation of enhancer sites was correlated with down-regulation of expression of cell-type-specific genes (Aran and Hellman, 2013; Aran et al., 2013). Surprisingly, the methylation state of enhancer regions was better correlated with gene expression changes than the methylation state of promoters, further underscoring the importance of enhancer regions in driving gene expression changes in cancer.

**Dysregulation of enhancer-acting factors in disease**

In addition to sequence variation and mutation within enhancer sequences themselves, many diseases have been linked to mutations or dysregulation of the transcription factors and transcriptional coactivators that function at enhancer regions (reviewed in (Herz et al., 2014; Lee and Young, 2013)). Many oncogenes overexpressed in cancer are transcription factors. For example, the *TAL1* transcription factor is overexpressed in half of T cell acute lymphoblastic leukemias (T-ALLs), and drives the expression of a key network of genes involved in T-ALL pathogenesis (Sanda et al., 2012). The *MYC* oncogene, which is the most frequently amplified oncogene in cancer, is a basic helix loop helix (bHLH) transcription factor, which binds to a six nucleotide E-box sequence (Beroukhim et al., 2010; Blackwood et al., 1991; Nesbit et al., 1999). *MYC* tends to bind to the core promoter region of genes, but when it is overexpressed, it begins to bind to lower affinity E-box-like regions in both enhancer and promoter (Lin et al., 2012).
This increased promoter binding and enhancer invasion of MYC is accompanied by global amplification of transcription in the tumor cell, possibly providing more raw material needed to drive tumor growth and proliferation (Lin et al., 2012; Nie et al., 2012).

Cofactors and chromatin regulators that act at enhancers can also be deregulated in cancer. Mutations in the Mediator coactivator complex can also occur in cancer; the MED12 subunit is frequently mutated in uterine leiomyomas and leiomyosarcomas, and prostate cancer (Barbieri et al., 2012; Makinen et al., 2011a; Makinen et al., 2011b). CDK8, part of the same Mediator module as MED12, is amplified, and acts as an oncogene in colon cancer and melanoma (Firestein et al., 2008; Kapoor et al., 2010; Morris et al., 2008). MLL3 and MLL4, histone methyltransferases that catalyze H3K4me1 at enhancers (Herz et al., 2014; Hu et al., 2013), are frequently mutated in a wide range of cancers, including colorectal cancer, breast cancer, liver cancer, bladder carcinoma, medulloblastoma, non-Hodgkin lymphoma, and gastric cancer (Ashktorab et al., 2010; Ellis and Perou, 2013; Fujimoto et al., 2012; Gui et al., 2011; Jones et al., 2012; Kandoth et al., 2013; Morin et al., 2011; Parsons et al., 2011; Watanabe et al., 2011; Zhang et al., 2012).

**Super-enhancers in human disease**

Disease-associated sequence variation and somatic changes in super-enhancer regions may play an important role in human disease. Disease- and trait-associated SNPs are particularly enriched in the super-enhancers of relevant cell types, when compared to typical enhancers (Hnisz et al., 2013). In one example, two SNPs associated with Alzheimer’s disease risk are found in a super-enhancer associated with the BIN1 gene in brain tissue. BIN1 expression has recently been linked to Alzheimer’s disease risk, and in the same study, an insertion in the BIN1 super-enhancer was also linked to Alzheimer’s risk (Chapuis et al., 2013). This enrichment of
disease-associated variation in super-enhancer regions was supported by work from Parker and colleagues, who found an enrichment of type 2 diabetes-associated variation in stretch enhancers, large enhancer regions, of islet cells (Parker et al., 2013).

As discussed further in chapter two, super-enhancers in tumor cells are associated with many key oncogenes, including MYC, the most commonly overexpressed oncogene in cancer (Beroukhim et al., 2010; Loven et al., 2013). Comparisons of super-enhancers between tumor cells and corresponding healthy tissue revealed that a majority of the genes acquired by tumor cells had functions that were related to cancer hallmark traits, such as evasion of growth suppressors, or activation of invasion and metastasis (Hnisz et al., 2013). Tumor cells can acquire super-enhancers through several mechanisms, including translocation, focal amplification, and overexpression of transcription factors. Examples of these cases include multiple myeloma, where the highly active IGH enhancer can be translocated near the MYC oncogene (Dib et al., 2008; Hnisz et al., 2013; Shou et al., 2000); T cell acute lymphoblastic leukemia (T-ALL), where the master transcription factor TALI is overexpressed, and is bound to a super-enhancer region near MYC (Bash et al., 1995; Brown et al., 1990; Hnisz et al., 2013); and small cell lung cancer (SCLC), where focal amplifications of the MYC gene and surrounding enhancer regions commonly occur (Iwakawa et al., 2013).

Many recent studies have identified super-enhancers in tumor cells, supporting the observation that key oncogenes are often associated with super-enhancer regions. Chapuy et al., found that super-enhancers in diffuse large b-cell lymphoma (DLBCL) were associated with several known oncogenes, and also identified OCA-B as a previously unknown oncogene driver of DLBCL (Chapuy et al., 2013). Shi et al. identified a larger super-enhancer in the 8q24 locus downstream of the MYC gene in acute myeloid leukemia (AML); this region was frequently
subject to focal amplifications in patient samples (Shi et al., 2013). Groeschel et al. found that a translocation in AML places the EVII gene under the control of a super-enhancer from the GATA2 gene, leading to aberrant expression of EVII (Groeschel et al., 2014). Finally, Walker et al. found that translocations at the 8q24 region in multiple myeloma often place super-enhancers near the MYC gene, increasing expression of the MYC oncogene. In addition to the commonly translocated immunoglobulin enhancers, including IGH, they found that other translocations involve important myeloma genes, such as XBP1, along with their super-enhancers (Walker et al., 2014).

Targeting enhancer function through small molecule inhibitors of transcriptional coactivators

Given the recent appreciation that disruptions in transcriptional regulators and regulatory regions can play a key role in disease, there has been increasing interest in developing therapeutics that target transcriptional regulators (reviewed in (Dawson and Kouzarides, 2012; Stellrecht and Chen, 2011; Villicana et al., 2014)). Many small molecules have been developed targeting chromatin modifying enzymes. Several of these types of inhibitors are FDA approved for cancer treatment, including the DNA methyltransferase inhibitors azacytidine and decitabine, and the histone deacetylase inhibitors vorinostat and romidepsin. Many additional small molecules are under investigation, including additional DMNT and HDAC inhibitors, as well as inhibitors of histone acetyltransferases (HATs) such as p300, lysine methyltransferases such as the H3K79 methyltransferase DOT1L, or the Polycomb complex member EZH2, and lysine demethylases such as LSD1 (Reviewed in (Helin and Dhanak, 2013; Villicana et al., 2014). Small molecule inhibitors targeting the histone reader, BRD4, have also shown promise as
therapeutic agents in many cancer types (Dawson et al., 2011; Delmore et al., 2011; Filippakopoulos et al., 2010; Mertz et al., 2011; Zuber et al., 2011).

In addition to targeting chromatin regulators, other transcriptional inhibitors target the general transcription factors and other complexes closely associated with RNA Pol II. Triptolide, and its analog minnelide, inhibits the activity of the TFIIH general transcription factor complex by binding covalently to the XPB subunit (Iben et al., 2002; Manzo et al., 2012; Titov et al., 2011; Vispe et al., 2009). Many inhibitors target CDK kinases involved in the transcriptional cycle, including CDK7, a component of TFIIH, CDK8, which forms a complex with Mediator, and CDK9, the kinase component of P-TEFb (Villicana et al., 2014). Flavopiridol, for example, is a pan-CDK inhibitor that has been investigated in the treatment of chronic lymphocytic leukemia (CLL) (Christian et al., 2009). Many other CDK inhibitors are under clinical investigation in a variety of cancers (Villicana et al., 2014). Most of these small molecules are relatively non-specific inhibitors of CDK activity, and many can target non-CDK kinases. Further research focused on the development of inhibitors selective for individual CDKs may potentially limit side effects due to off target inhibition of non-transcriptional kinases (Villicana et al., 2014; Wesierska-Gadek and Krystof, 2009).

A critical feature of cancer therapy is the ability to selectively target tumor cells, and the oncogenic properties that define them. This raises the conundrum of how global inhibitors of transcription can achieve a gene-selective and tumor-selective effect. Despite this, many of these inhibitors, including small molecules targeting BRD4, appear able to selectively inhibit the expression of key oncogenes in tumor cells, such as MYC (Bandopadhayay et al., 2014; Dawson and Kouzarides, 2012; Dawson et al., 2011; Delmore et al., 2011; Filippakopoulos et al., 2010; Mertz et al., 2011; Zuber et al., 2011). This selective effect on the MYC oncogene is especially.
important, given the difficulty in targeting transcription factors, such as MYC, with conventional small molecule therapies (Darnell, 2002).

There are several properties of oncogenes, and of tumor cells, that may allow transcriptional inhibitors to act in a gene-specific and tumor-specific manner. First, many oncogenes and pro-apoptotic factors have short half-lives, resulting in rapid down-regulation of these transcripts upon global inhibition of transcription. Tumor cells often become addicted to certain oncogenes, rendering tumor cells particularly sensitive to reduced levels of oncogene expression. In addition, tumor cells may require higher sustained levels of transcription than normal cells, as suggested by the transcriptional amplification effected by the MYC oncogene.

Lastly, in the second chapter of this thesis, I describe a mechanism by which disruption of super-enhancers by the BRD4 inhibitor JQ1 leads to the selective downregulation of tumor oncogenes. In multiple myeloma cells, key oncogenes, including MYC, are associated with large, BRD4-occupied super-enhancers. These super-enhancers are particularly sensitive to JQ1 treatment, resulting in a disproportionate loss of BRD4 occupancy at super-enhancer regions. Loss of BRD4 binding at super-enhancers is concomitant with a loss of transcriptional elongation at associated genes, and a decrease in mRNA levels of these key oncogenes. Again, recent work has corroborated these findings, lending support to the idea that some transcriptional inhibitors may be able to selectively target super-enhancers and associated genes. As described above, Chapuy et al. investigated super-enhancer regions in DLBCL. In this study, they observed that super-enhancer-associated genes, including OCA-B, were particularly sensitive to BRD4 inhibition by JQ1 (Chapuy et al., 2013). Groeschel et al. (2014) found that the expression of EVII was especially sensitive to JQ1 treatment in AML cell lines in which EVII is driven by a translocated, highly BRD4-occupied super-enhancer. They also observed greater loss of BRD4
occupancy in response to JQ1 at super-enhancers when compared to typical enhancers in these cell lines (Groschel et al., 2014). Wang et al. examined the response of T-lymphoblastic leukemia (T-LL) to a gamma-secretase inhibitor (GSI), which prevents nuclear localization of the transcription factor Notch. They found that regions showing a dynamic change in Notch occupancy upon GSI treatment overlapped significantly with super-enhancers. These dynamic super-enhancer regions in turn correlated with changes in gene expression of associated genes upon drug treatment (Wang et al., 2014).

Conclusion

Together, this evidence provides a strong impetus for further study of enhancer regions, and super-enhancers in particular, as a therapeutic target in cancer and other human disease, which will be discussed further in chapter three. Enhancer regions play an important role in controlling gene expression programs through their association with sequence-specific transcription factors and looping interactions with promoter regions. New sequencing technologies have allowed for the genome-wide mapping of enhancer regions in a wide range of cell types, giving new appreciation to the role of these regions in human development and disease. The identification of super-enhancers may serve to further focus future research by marking the critical genes involved in cell or disease-state. Because super-enhancers are associated with many critical tumor oncogenes in the cancers thus far investigated, identifying the super-enhancers in a given tumor may provide valuable insight into its biology. In addition, given the greater ability to target transcriptional regulators and other enhancer-associated factors, and the particular sensitivity super-enhancers display in response to many perturbations, this strategy may be an effective way to selectively target important super-enhancer-associated oncogenes.
References


Chapter 2

Selective Inhibition of Tumor Oncogenes by Disruption of Super-enhancers

Jakob Loven*, Heather A. Hoke1,2*, Charles Y. Lin1,3,5*, Ashley Lau1,2, David A. Orlando1, Christopher R. Vakoc4, James E. Bradner5,6, Tong Ihn Lee1, and Richard A. Young1,2

1Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, 2Department of Biology, MIT, Cambridge, MA 02139, 3Computational and Systems Biology Program, Massachusetts Institute of Technology, Cambridge, MA 02139, 4Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA, 5Department of Medical Oncology, Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, USA, 6Department of Medicine, Harvard Medical School, 25 Shattuck Street, Boston, MA 02115, USA

* These authors contributed equally

This chapter originally appeared in Cell 2013 vol. 153 (2) pp. 320-334. Corresponding Supplementary Material is appended. Supplementary tables and data are available online at http://dx.doi.org/10.1016/j.cell.2013.03.036.
Personal Contribution to the Project

This work was a close collaboration between Jakob Lovén, Charles Y. Lin and myself. I performed the majority of experiments, with assistance from Jakob Lovén and Ashley Lau. Charles Y. Lin performed the analysis. I wrote the manuscript in collaboration with Jakob Lovén, Charles Y. Lin, and Richard A. Young.
Summary

Chromatin regulators have become attractive targets for cancer therapy, but it is unclear why inhibition of these ubiquitous regulators should have gene-specific effects in tumor cells. Here we investigate how inhibition of the widely-expressed transcriptional coactivator BRD4 leads to selective inhibition of the MYC oncogene in multiple myeloma (MM). BRD4 and Mediator were found to co-occupy thousands of enhancers associated with active genes. They also co-occupied a small set of exceptionally large super-enhancers associated with genes that feature prominently in MM biology, including the MYC oncogene. Treatment of MM tumor cells with the BET-bromodomain inhibitor JQ1 led to preferential loss of BRD4 at super-enhancers and consequent transcription elongation defects that preferentially impacted genes with super-enhancers, including MYC. Super-enhancers were found at key oncogenic drivers in many other tumor cells. These observations have implications for the discovery of novel cancer therapeutics directed at components of super-enhancers in diverse tumor types.
Control
Tumor cells

BRD4 inhibited
Growth arrest

Typical enhancer driven gene

Coactivators:
BRD4, Mediator

Enhancer elements

Super-enhancer driven gene

Oncogene

Graphical Abstract
Introduction

Chromatin regulators are attractive as therapeutic targets for cancer because they are deregulated in numerous cancers (Baylin and Jones, 2011; Elsasser et al., 2011; Esteller, 2008; Feinberg and Tycko, 2004; You and Jones, 2012) and amenable to small molecule inhibition (Cole, 2008; Dawson and Kouzarides, 2012; Geutjes et al., 2012). Inhibition of some chromatin regulators has already proven to be efficacious for treatment of certain cancers (Issa and Kantarjian, 2009; Marks and Xu, 2009). Most chromatin regulators, however, are expressed in a broad range of healthy cells and contribute generally to gene expression, so inhibition of these important genome-associated proteins might be expected to adversely affect global gene expression in healthy cells and thus produce highly toxic effects. Nonetheless, inhibitors of some chromatin regulators, such as BRD4, have been shown to selectively inhibit transcription of key oncogenic drivers such as c-MYC (hereafter referred to as MYC) in multiple tumor types (Dawson et al., 2011; Delmore et al., 2011; Mertz et al., 2011; Zuber et al., 2011). It is important to understand how inhibition of a widely expressed, general regulator such as BRD4 can exert a selective effect on the expression of a small number of genes in specific cells.

BRD4 is a member of the bromodomain and extraterminal (BET) subfamily of human bromodomain proteins, which include BRDT, BRD2, BRD3, and BRD4. These proteins associate with acetylated chromatin and facilitate transcriptional activation (LeRoy et al., 2008; Rahman et al., 2011). BRD4 was first identified as an interaction partner of the murine Mediator coactivator complex (Jiang et al., 1998), and was subsequently shown to associate with Mediator in a variety of human cells (Dawson et al., 2011; Wu and Chiang, 2007). BRD4 is involved in the control of transcriptional elongation by RNA polymerase II (RNA Pol II) through its recruitment of the positive transcription elongation factor P-TEFb (Jang et al., 2005; Yang et al.,
Almost all human cells express the **BRD4** gene, based on analysis of human tissue expression data across 90 distinct tissue types (Human body index - transcriptional profiling, see Supplementary Methods), and BRD4 is found associated with a large population of active genes in CD4+ T cells (Zhang et al., 2012). It is not yet clear whether the BRD4 protein is generally involved in the transcription of active genes in tumor cells or if it is selectively associated with a subset of these genes.

Two recently developed bromodomain inhibitors, JQ1 and iBET, selectively bind to the amino-terminal twin bromodomains of BRD4 (Filippakopoulos et al., 2010; Nicodeme et al., 2010). These BET inhibitors cause selective repression of the potent **MYC** oncogene in a range of tumors, including multiple myeloma (MM), Burkitt’s lymphoma (BL), acute myeloid leukemia (AML), and acute lymphoblastic leukemia (ALL) (Dawson et al., 2011; Delmore et al., 2011; Mertz et al., 2011; Ott et al., 2012; Zuber et al., 2011). The inhibition of **MYC** apparently occurs as a consequence of BRD4 depletion at the enhancers that drive **MYC** expression (Delmore et al., 2011). Although BRD4 is widely expressed in mouse tissues, mice are reasonably tolerant of the levels of BET bromodomain inhibition that inhibit certain tumors in mouse models (Dawson et al., 2011; Delmore et al., 2011; Filippakopoulos et al., 2010; Mertz et al., 2011; Zuber et al., 2011).

The multiple myeloma cell line (MM1.S) used to study the effects of JQ1 has an **IgH–MYC** rearrangement and **MYC** gene expression is driven by factors associated with the **IgH** enhancer (Dib et al., 2008; Shou et al., 2000). Enhancers function through co-operative and synergistic interactions between multiple transcription factors and coactivators (Carey et al., 1990; Giese et al., 1995; Kim and Maniatis, 1997; Thanos and Maniatis, 1995). Cooperative binding and synergistic activation confer increased sensitivity, so that small changes in activator
concentration can lead to dramatic changes in activator binding and transcription of associated genes (Carey, 1998). Furthermore, enhancers with large numbers of transcription factor binding sites can be more sensitive to small changes in factor concentration than those with smaller numbers of binding sites (Giniger and Ptashne, 1988; Griggs and Johnston, 1991). This concept led us to postulate that some feature of the \emph{IgH} enhancer might account for the selective effect of BRD4 inhibition.

We show here that BRD4 and Mediator are associated with most active enhancers and promoters in MM1.S tumor cells, but exceptionally high levels of these cofactors occur at a small set of large enhancer regions, which we call super-enhancers. Super-enhancers are associated with \emph{MYC} and other key genes that feature prominently in the biology of MM, including many lineage-specific survival genes. Treatment of MM tumor cells with the BRD4 inhibitor JQ1 caused a preferential loss of BRD4, Mediator and P-TEFb at super-enhancers and preferential loss of transcription at super-enhancer-associated genes, including the \emph{MYC} oncogene. Tumor cell addiction to high-level expression of these oncogenes may then contribute to their vulnerability to super-enhancer disruption (Chin et al., 1999; Felsher and Bishop, 1999; Jain et al., 2002; Weinstein, 2002). We find super-enhancers in additional tumor types, where they are similarly associated with key oncogenes. Thus, key oncogene drivers of tumor cells are regulated by super-enhancers, which can confer disproportionate sensitivity to loss of the BRD4 coactivator and thus cause selective inhibition of transcription.
Results

**BRD4 and Mediator co-occupy promoters of active genes in multiple myeloma**

Transcription factors bind to enhancers and recruit the Mediator coactivator, which in turn becomes associated with RNA polymerase II at the transcription start site, thus forming DNA loops between enhancers and core promoters (Kagey et al., 2010). BRD4 is known to associate with Mediator in some mammalian cells (Dawson et al., 2011; Jiang et al., 1998; Wu et al., 2003). To identify active promoter and enhancer elements and determine how BRD4 and Mediator occupy the genome in MM1.S multiple myeloma cells, we used chromatin immunoprecipitation coupled to high-throughput sequencing (ChIP-Seq) with antibodies against the Mediator subunit MED1, BRD4, the enhancer-associated histone modification H3K27Ac, and the transcription start site (TSS)-associated histone modification H3K4Me3 (Figure 1).

ChIP-Seq signals for both Mediator and the histone modification H3K27Ac have previously been shown to occur at both enhancers and TSSs (Creyghton et al., 2010; Heintzman et al., 2009; Rada-Iglesias et al., 2011), and enhancers can be distinguished from TSSs by the absence of TSS annotation and relatively low levels of H3K4Me3. We found that BRD4 co-occupied enhancers and TSSs with MED1 throughout the genome (Figure 1A and 1B) and that the levels of BRD4 and MED1 were strongly correlated (Figure S1).

To confirm that BRD4 and Mediator are generally associated with active genes in MM1.S cells, we compared the ChIP-Seq data for these regulators with that for RNA Pol II and the histone modification H3K4Me3. The levels of BRD4 and Mediator correlated with the levels of RNA Pol II genome-wide (Figure 1C). Signals for BRD4 and Mediator were found together with those for the histone modification H3K4Me3 and RNA Pol II at ~10,000 annotated TSSs and these were considered active TSSs (Table S1). Signals for BRD4 and the enhancer-
associated histone modification H3K27Ac were found in ~8,000 Mediator-occupied regions either lacking TSSs or extending beyond the immediate vicinity of the TSS, and these were considered enhancer regions (Table S2, Data S1, Supplementary Methods).
Figure 1
Figure 1: Mediator and BRD4 co-occupy promoters of active genes in multiple myeloma

A) Gene tracks of MED1, BRD4, H3K27Ac, and H3K4Me3 ChIP-Seq occupancy at the enhancer (left) and promoter (right) of SMARCA4 in MM1.S multiple myeloma cells. The x-axis shows genomic position and enhancer containing regions are depicted with a white box. The y-axis shows signal of ChIP-Seq occupancy in units of reads per million mapped reads per base pair (rpm/bp).

B) Meta-gene representation of global MED1, BRD4, H3K27Ac, and H3K4Me3 occupancy at enhancers and promoters. The x-axis shows the +/- 2.5kb region flanking either the center of enhancer regions (left) or the TSS of active genes (right). The y-axis shows the average background subtracted ChIP-Seq signal in units of rpm/bp.

C) Median MED1 and BRD4 levels in the +/- 1kb region around the TSSs of actively transcribed genes ranked by increasing RNA Polymerase II (RNA Pol II) occupancy in MM1.S cells. Levels are in units of rpm/bp with the left y-axis showing levels of MED1 and the right y-axis showing levels of BRD4. Promoters were binned (50/bin) and a smoothing function was applied to median levels. See also Figure S1.
Super-enhancers are associated with key multiple myeloma genes

Further analysis of the ~8,000 enhancer regions revealed that the MED1 signal at 308 enhancers was significantly greater than at all other enhancers and promoters (Figure 2A; Figure S2A; Table S2). These 308 “super-enhancers” differed from typical enhancers in both size and Mediator levels (Figure 2B). Remarkably, ~40% of all enhancer-bound Mediator and BRD4 occupied these 308 super-enhancers. While the typical enhancer had a median size of 1.3kb, the super-enhancers had a median size of 19.4kb. These super-enhancers were thus 15-fold larger than typical enhancers and were occupied, based on ChIP-Seq signal, by 18-fold more Mediator and 16-fold more BRD4. Similarly high levels of H3K27Ac were observed in these large regions (Figure 2B). Examples of gene tracks showing super-enhancers at either end of the spectrum of Mediator occupancy (Figure 2A) are shown in Figure 2C. The largest super-enhancer was found associated with the IGLL5 gene, which encodes an immunoglobulin lambda peptide expressed at high levels in these cells.

We next sought to identify the complete set of MM1.S genes that are most likely associated with super-enhancers. Enhancers tend to loop to and associate with adjacent genes in order to activate their transcription (Gondor and Ohlsson, 2009; Lelli et al., 2012; Ong and Corces, 2011; Spitz and Furlong, 2012). Most of these interactions occur within a distance of ~50kb of the enhancer (Chepelev et al., 2012). Using a simple proximity rule, we assigned all transcriptionally active genes (TSSs) to super-enhancers within a 50kb window, a method shown to identify a large proportion of true enhancer/promoter interactions in embryonic stem cells (Dixon et al., 2012). This identified 681 genes associated with super-enhancers (Supplemental Table S3), and 307 of these had a super-enhancer overlapping a portion of the gene, as shown for CCND2 in Figure 2C.
Super-enhancer-associated genes were generally expressed at higher levels than genes with typical enhancers and tended to be specifically expressed in MM1.S cells (Figure 2D). To test whether components of super-enhancers confer stronger activity compared to typical enhancers, we cloned representative super-enhancer or typical enhancer fragments of similar size into luciferase reporter constructs and transfected these into MM1.S cells. Cloned sequence fragments from super-enhancers generated 2-3 fold higher luciferase activity compared to typical enhancers of similar size (Figure 2E; Supplemental Methods). These results are consistent with the notion that super-enhancers help activate high levels of transcription of key genes that regulate and enforce the MM1.S cancer cell state.

The super-enhancer-associated genes included most genes that have previously been shown to have important roles in multiple myeloma biology, including MYC, IRF4, PRDM1/BLIMP-1 and XBP1 (Figure 3A). MYC is a key oncogenic driver in MM (Chng et al., 2011; Dib et al., 2008; Holien et al., 2012; Shou et al., 2000) and the MM1.S MYC locus contains a chromosomal rearrangement that places MYC under the control of the IgH enhancer, which qualifies as a super-enhancer in MM1.S cells. The IRF4 gene encodes a key plasma cell transcription factor that is frequently deregulated in MM (Shaffer et al., 2008). PRDM1/BLIMP-1 encodes a transcription factor considered a master regulator of plasma cell development, and is required for the formation of plasma cell tumors in a mouse model (Shapiro-Shelef et al., 2003; Turner et al., 1994). XBP1 encodes a basic-region leucine zipper (bZIP) transcription factor of the CREB-ATF family that governs plasma cell differentiation (Reimold et al., 2001). XBP1 is frequently overexpressed in human MM and can drive the development of MM in a mouse model (Carrasco et al., 2007; Claudio et al., 2002).
Super-enhancers were associated with many additional genes that have important roles in cancer pathogenesis more generally (Figure 3B). Cyclin D2 (CCND2) is deregulated in many human cancers, including MM (Bergsagel et al., 2005; Musgrove et al., 2011). The PIM1 kinase has been implicated in the biology of many different cancers (Shah et al., 2008). MCL1 and BCL-xL, members of the BCL-2 family of apoptosis regulators, are frequently deregulated in cancer, promoting cell survival and chemoresistance (Beroukhim et al., 2010). We conclude that super-enhancers are frequently associated with genes that feature prominently in the biology of multiple myeloma and other human cancers.
Enhancers ranked by increasing MED1 signal

Genome-wide average

Typical enhancers
Super-enhancer

Number: 7,985
Median size: 1,330 bp

Total Density at signal constituents
MED1: 1 unit 1 unit
BRD4: 1 unit 1 unit
H3K27Ac: 1 unit 1 unit

Top1 Enhancer
IGLL5 Super-enhancer

TOP1 Enhancer
IGLL5 Super-enhancer

SMARCA4 Enhancer
CCND2 Super-enhancer

Figure 2
**Figure 2: Super-enhancers identified in multiple myeloma**

A) Total MED1 ChIP-Seq signal in units of reads per million in enhancer regions for all enhancers in MM1.S. Enhancers are ranked by increasing MED1 ChIP-Seq signal.

B) Meta-gene representation of global MED1 (red line) and BRD4 (blue line) occupancy at typical enhancers and super-enhancers. The x-axis shows the start and end of the enhancer (left) or super-enhancer (right) regions flanked by +/- 5kb of adjacent sequence. Enhancer and super-enhancer regions on the x-axis are relatively scaled. The y-axis shows the average signal in units of rpm/bp.

C) Gene tracks of MED1 (top) and BRD4 (bottom) ChIP-Seq occupancy at the typical enhancer upstream of *TOP1*, the super-enhancer downstream of *IGLL5*, the typical enhancer upstream of *SMARCA4*, and the super-enhancer overlapping the *CCND2* gene TSS. The x-axis shows genomic position and super-enhancer containing regions are depicted with a grey box. The y-axis shows signal of ChIP-Seq occupancy in units of rpm/bp.

D) Left: Boxplots of expression values for genes with proximal typical enhancers (white), or with proximal super-enhancers (pink). The y-axis shows expression value in Log2 arbitrary units. Right: Boxplots of cell type specificity values for genes with proximal typical enhancers (white), or with proximal super-enhancers (purple). The y-axis shows the Z-score of the JS divergence statistic for genes, with higher values corresponding to a more cell type specific pattern of expression. Changes between expression levels are significant (two tailed Welch's t-test, p-value < 2e-16) as are changes between cell type specificity levels (two tailed Welch's t-test, p-value = 1e-14)
E) Bar graph depicting luciferase activity of reporter constructs containing cloned fragments of typical enhancers and super-enhancers in MM1.S cells. 2kb fragments of three super-enhancers, IGLL5, DUSP5, and SUB1, and three typical enhancers, PDHX, SERPINB8, and TOP1, ranked 1, 129, 227, 2352, 4203, and 4794, respectively, in terms of MED1 occupancy, were cloned into reporter plasmids downstream of the luciferase gene, driven by a minimal MYC promoter. Luciferase activity is represented as fold over empty vector. Error bars represent standard deviation of triplicate experiments. See also Figure S2 and Data S1.
A

Super-enhancers:

- 15.0 IgH insertion der3(3;8)

- 15.0 BRD4

- 15.0 MED1

- 15.0 IRF4

- 15.0 MED1

- 15.0 PIM1

- 15.0 XBP1

B

Super-enhancers:

- 15.0 MED1

- 15.0 CCND2

- 15.0 PIM1

- 15.0 BCL-xL

- 15.0 MCL1

Figure 3
Figure 3: Super-enhancers are associated with key multiple myeloma genes

A,B) Gene tracks of MED1 and BRD4 ChIP-Seq occupancy at super-enhancers near genes with important roles in multiple myeloma biology A) or genes with important roles in cancer B). Super-enhancers are depicted in gray boxes over the gene tracks. The x-axis shows genomic position and super-enhancer containing regions are depicted with a grey box. The y-axis shows signal of ChIP-Seq occupancy in units of rpm/bp.
Inhibition of BRD4 leads to displacement of BRD4 genome-wide

BRD4 interacts with chromatin-associated proteins such as transcription factors, the Mediator complex and acetylated histones (Dawson et al., 2011; Dey et al., 2003; Jang et al., 2005; Jiang et al., 1998; Wu and Chiang, 2007; Wu et al., 2013). Previous studies have shown that treatment of MM1.S cells with JQ1 leads to reduced levels of BRD4 at the IgH enhancer that drives MYC expression (Delmore et al., 2011), but it is not clear whether such treatment causes a general reduction in the levels of BRD4 associated with the genome. We found that treatment of MM1.S cells with 500nM JQ1 for 6 hours reduced the levels of BRD4 genome-wide by approximately 70% (Figure 4A and 4B). This reduction in BRD4 occupancy was evident both by inspection of individual gene tracks (Figure 4C) and through global analysis of the average effects at enhancers and TSSs (Figure 4D). JQ1 treatment led to ~60% reduction in BRD4 signal at enhancers and ~90% reduction at promoters (Figure 4D). The reduction in BRD4 was more profound at super-enhancers such as those associated with IgH–MYC and CCND2 (Figure 4E), where the loss of BRD4 was nearly complete. We conclude that BET bromodomain inhibition of BRD4 leads to reduced levels of BRD4 at enhancers and promoters throughout the genome in MM1.S cells.
Figure 4
Figure 4: Inhibition of BRD4 leads to loss of BRD4 genome-wide

A) Tracks showing BRD4 ChIP-Seq occupancy on the 35mb right arm of chromosome 21 after DMSO (top) or 500nM JQ1 (bottom) treatment. The chromosome 21 ideogram is displayed above the gene tracks with the relevant region highlighted in blue. The x-axis of the gene tracks shows genomic position and the y-axis shows BRD4 ChIP-Seq signal in units of rpm/bp.

B) Boxplot showing the distributions of BRD4 ChIP-Seq signal at BRD4 enriched regions after DMSO (left) or 500nM JQ1 (right) treatment. BRD4 enriched regions were defined in MM1.S cells treated with DMSO. The y-axis shows BRD4 ChIP-Seq signal in units of rpm/bp. The loss of BRD4 occupancy at BRD4 enriched regions after JQ1 is highly significant (p-value < 1e-16).

C) Gene tracks of BRD4 ChIP-Seq occupancy at the enhancer (left) and promoter (right) of SMARCA4 in MM1.S cells after DMSO (top) or 500nM JQ1 (bottom) treatment for 6 hours. The x-axis shows genomic position and enhancer containing regions are depicted with a white box. The y-axis shows signal of ChIP-Seq occupancy in units of rpm/bp.

D) Meta-gene representation of global BRD4 occupancy at enhancers and promoters after DMSO (solid line) or 500nM JQ1 (dotted line) treatment. The x-axis shows the +/- 2.5kb region flanking either the center of enhancer regions (left) or the TSS of active genes. The y-axis shows the average background subtracted ChIP-Seq signal in units of rpm/bp.

E) Gene tracks of BRD4 binding at super-enhancers after DMSO (top) or 500nM JQ1 (bottom) treatment. The x-axis shows genomic position and super-enhancer containing regions are depicted with a grey box. The y-axis shows signal of ChIP-Seq occupancy in units of rpm/bp.
Transcription of super-enhancer-associated genes is highly sensitive to BRD4 inhibition

Enhancers are formed through co-operative and synergistic binding of multiple transcription factors and coactivators (Carey, 1998; Carey et al., 1990; Giese et al., 1995; Kim and Maniatis, 1997; Thanos and Maniatis, 1995). As a consequence of this binding behavior, enhancers bound by many cooperatively-interacting factors lose activity more rapidly than enhancers bound by fewer factors when the levels of enhancer-bound factors are reduced (Giniger and Ptashne, 1988; Griggs and Johnston, 1991). The presence of super-enhancers at MYC and other key genes associated with multiple myeloma led us to consider the hypothesis that super-enhancers are more sensitive to reduced levels of BRD4 than typical enhancers and that genes associated with super-enhancers might then experience a greater reduction of transcription than genes with average enhancers when BRD4 is inhibited (Figure 5A).

To test this hypothesis, we first examined the effects of various concentrations of JQ1 on BRD4 occupancy genome-wide (Figure 5B). JQ1 had little effect on MM1.S cell viability when treated for 6 hours at these various concentrations, while at later time points JQ1 had a significant anti-proliferative effect (Figure 5C). As expected, MYC protein levels were significantly depleted by exposure of MM1.S cells to 50nM or greater doses of JQ1 for 6 hrs (Figure 5D) (Delmore et al., 2011). In contrast, JQ1 did not affect total BRD4 protein levels within the cells, and did not significantly reduce ChIP efficiency (Figure 5E). When BRD4 occupancy was examined genome-wide in cells exposed to the increasing concentrations of JQ1, it was evident that super-enhancers showed a greater loss of BRD4 occupancy than typical enhancer regions (Figure 5F). For example, the IgH super-enhancer showed significantly greater reduction in BRD4 occupancy in cells treated with 5nM or 50nM JQ1 than typical enhancer regions such as that upstream of SMARCA4 (Figure 5G). Ultimately, virtually all BRD4
occupancy was lost at the *IgH* super-enhancer (97% reduction versus DMSO control) after treatment with 500nM JQ1, while loss of BRD4 occupancy at the typical enhancer for *SMARCA4* was less pronounced (71% reduction versus DMSO control) (Figure 5G).

We next investigated whether genes associated with super-enhancers might experience a greater reduction of transcription than genes with average enhancers when BRD4 is inhibited. As expected, treatment of MM1.S cells with 500nM JQ1 led to progressive reduction in global mRNA levels over time (Figure 6A; Figure S3A). Similarly, treatment with increasing concentrations of JQ1 caused progressive reductions in global mRNA levels (Figure 6A; Figure S3B). There was a selective depletion of mRNAs from super-enhancer-associated genes that occurred in both a temporal (Figure 6B) and concentration-dependent manner (Figure 6C). Notably, *MYC* and *IRF4* mRNA levels were more rapidly depleted than other mRNAs that are expressed at similar levels (Figure 6D). The levels of transcripts from super-enhancer-associated genes were somewhat more affected than those from genes that have multiple typical enhancers bound by BRD4 (Figure S3C and S3D). Thus, BET bromodomain inhibition preferentially impacts transcription of super-enhancer-driven genes.

To further test the model that super-enhancers are responsible for the special sensitivity to BRD4 inhibition, we transfected MM1.S cells with luciferase reporter constructs containing super-enhancer and typical enhancer fragments and examined the effects of various JQ1 concentrations on luciferase activity. Upon treatment with JQ1, MM1.S cells transfected with a super-enhancer reporter experienced a greater reduction in luciferase activity than those transfected with a typical enhancer reporter (Figure 6E). Interestingly, the dose-response curve observed for luciferase activity of the super-enhancer construct is consistent with that expected for enhancers that are bound cooperatively by multiple factors (Figure 5A) (Giniger and Ptashne,
1988; Griggs and Johnston, 1991). These results are also consistent with the model that super-enhancers are responsible for the special sensitivity of gene transcription to BRD4 inhibition.
Figure 5
Figure 5: BRD4 occupancy at super-enhancers is highly sensitive to bromodomain inhibition

A) Schematic example of how cooperative interactions of enhancer-associated factors at super-enhancers leads to both higher transcriptional output and increased sensitivity to factor concentration.

B) Measuring the effects of various concentrations of JQ1 on genome-wide on BRD4 occupancy. Schematic depicting the experimental procedure.

C) Short-term JQ1 treatment (6 hours) has little effect on MM1.S cell viability. JQ1 sensitivity of MM1.S cells by measurement of ATP levels (CellTiterGlo) after 6, 24, 48 and 72 hours of treatment with JQ1 (5, 50, 500 or 5000nM) or vehicle (DMSO, 0.05%).

D) Western blot of relative MYC levels after 6 hours of JQ1 or DMSO treatment.

E) Western blot of relative BRD4 levels after 6 hours of JQ1 or DMSO treatment. ChIP-Western blot of the relative levels of immunoprecipitated BRD4 after 6 hours of JQ1 or DMSO treatment.

F) Line graph showing the percentage of BRD4 occupancy remaining after 6 hour treatment at various JQ1 concentrations for typical enhancers (grey line) or super-enhancers (red line). The y-axis shows the fraction of BRD4 occupancy remaining versus DMSO. The x-axis shows different JQ1 concentrations (DMSO (none), 5nM, 50nM, and 500nM). Error bars represent 95% confidence intervals of the mean (95% CI).

G) Gene tracks of BRD4 ChIP-Seq occupancy after various concentrations of JQ1 treatment at the IgH-MYC associated super-enhancer (left) and the SMARCA4-associated typical enhancer (right). The x-axis shows genomic position and grey boxes depict super-enhancer regions. The
y-axis shows signal of ChIP-Seq occupancy in units of rpm/bp. The percent BRD4 remaining after each concentration of JQ1 treatment is annotated to the right of the gene tracks.
**BRD4 inhibition and transcription elongation**

At active genes, enhancers and core promoters are brought into close proximity, so factors associated with enhancers can act on the transcription apparatus in the vicinity of transcription start sites and thereby influence initiation or elongation. BRD4 is known to interact with Mediator and P-TEFb and to be involved in the control of transcriptional elongation by RNA polymerase II (Conaway and Conaway, 2011; Dawson et al., 2011; Jang et al., 2005; Krueger et al., 2010; Rahman et al., 2011; Yang et al., 2005). This suggests that the preferential loss of BRD4 from super-enhancers might affect the levels of Mediator and P-TEFb at these sites and furthermore, that the reduced levels of mRNAs from super-enhancer-associated genes might be due to an effect on transcription elongation.

To test these predictions, we carried out ChIP-Seq for the Mediator component MED1, and the catalytic subunit of the P-TEFb complex CDK9, in MM1.S cells treated with DMSO or 500nM JQ1 for 6 hours. In control cells, MED1 and CDK9 were found at enhancers and promoters of active genes throughout the MM genome, as expected (Figure 1A and 1B; Figure S3E). In cells treated with JQ1, reduced levels of MED1 and CDK9 were observed primarily at enhancers, with the greatest loss at super-enhancers (Figure 6F). As many super-enhancers span contiguous regions that encompass or overlap the TSS, we analyzed MED1 and CDK9 loss in either TSS proximal or TSS distal regions of super-enhancers and again observed loss of MED1 and CDK9 predominantly at TSS distal regions (Figure S3F). We conclude that inhibition of BRD4 genomic binding leads to a marked reduction in the levels of Mediator and P-TEFb at genomic regions distal to TSSs, with the greatest reduction occurring at super-enhancers.

To determine whether reduced levels of BRD4 lead to changes in transcription elongation, we quantified changes in transcription elongation by performing ChIP-Seq of RNA
Pol II before and after treatment of MM1.S cells with 500nM JQ1. We then calculated the fold loss of RNA Pol II occupancy in the gene body regions for all transcriptionally active genes and found that more than half of these genes show a decrease in elongating RNA Pol II density after JQ1 treatment (Figure 6G). Importantly, genes associated with super-enhancers showed a greater decrease of RNA Pol II in their elongating gene body regions compared to genes associated with typical enhancers (Figure 6H; Figure S3G). Inspection of individual gene tracks revealed pronounced elongation defects at super-enhancer-associated genes such as MYC and IRF4, with the greatest effects observed with MYC (Figure 6I and 6J). Thus, the selective effects of JQ1 on the transcription of MYC and other super-enhancer-associated genes can be explained, at least in part, by the sensitivity of super-enhancers to reduced levels of BRD4, which leads to a pronounced effect on pause release and transcription elongation.
Figure 6

Figure 6: JQ1 causes disproportionate loss of transcription at super-enhancer genes

A) Boxplots showing the Log₂ change in gene expression for all actively transcribed genes in JQ1 treated vs. control cells for a time course of cells treated with 500nM JQ1 (left) or for a concentration course of cells treated for 6 hours with varying amounts of JQ1 (right). The Y-axis shows the Log₂ change in gene expression vs. untreated control cells (left graph) or control cells treated with DMSO for 6 hours (right graph).

B,C) Line graph showing the Log₂ change in gene expression vs. control cells after JQ1 treatment in a time B) or dose C) dependent manner for genes associated with typical enhancers (grey line) or genes associated with super-enhancers (red line). The Y-axis shows the Log₂ change in gene expression of JQ1 treated vs. untreated control cells. X-axis shows time of 500nM JQ1 treatment B) or JQ1 treatment concentration at 6 hours C).

D) Graph showing the Log₂ change in gene expression after JQ1 treatment over time for genes ranked in the top 10% of expression in MM1.S cells. Each line represents a single gene with the MYC and IRF4 genes drawn in red. The Y-axis shows the Log₂ change in gene expression of JQ1 treated vs. untreated control cells. The X-axis shows time of 500nM JQ1 treatment.

E) Line graph showing luciferase activity after JQ1 treatment at various concentrations for luciferase reporter constructs containing either a fragment from the IGLL5 super-enhancer (red line) or the PDHX typical enhancer (grey line). Y-axis represents relative luciferase activity in arbitrary units. X-axis shows JQ1 concentrations. Error bars are standard error of the mean.

F) Bar graphs showing the percentage loss of either MED1 (top, red) or CDK9 (bottom, green) at promoters, typical enhancers, and super-enhancers. Error bars represent 95% CI.
G) Graph of loss of RNA Pol II density in the elongating gene body region for all transcriptionally active genes in MM1.S cells after 6 hour 500nM JQ1 treatment. Genes are ordered by decrease in elongating RNA Pol II in units of \( \log_2 \) fold loss. Genes with a greater than 0.5 \( \log_2 \) fold change in elongating RNA Pol II are shaded in green (loss) or red (gain). The amount of RNA Pol II loss is indicated for select genes.

H) Bar graph showing the \( \log_2 \) fold change in RNA Pol II density in elongating gene body regions after 6 hour 500nM JQ1 treatment for genes with typical enhancers (left, grey) or genes with super-enhancers (red, right). Error bars represent 95% CI.

I,J) Gene tracks of RNA Pol II ChIP-Seq occupancy after DMSO (black) or 500nM JQ1 treatment (red) at the super-enhancer proximal \( MYC \) gene I) and \( IRF4 \) gene J). The y-axis shows signal of ChIP-Seq occupancy in units of rpm/bp. See also Figure S3.
Super-enhancers are associated with disease-critical genes in other cancers

To map enhancers and determine whether super-enhancers occur in additional tumor types, we investigated the genome-wide occupancy of Mediator (MED1), BRD4, and the enhancer-associated histone modification H3K27Ac using ChIP-Seq in glioblastoma multiforme (GBM) and small cell lung cancer (SCLC) (Figure 7). Mediator (MED1) occupancy was used to identify enhancer elements because enhancer-bound transcription factors bind directly to Mediator (Borggrefe and Yue, 2011; Conaway and Conaway, 2011; Kornberg, 2005; Malik and Roeder, 2010; Taatjes, 2010) and because it has proven to produce high-quality evidence for enhancers in mammalian cells (Kagey et al., 2010). Global occupancy of BRD4 and H3K27Ac were used as corroborative evidence to identify enhancer elements (Figure S4; Table S4). Analysis of the regions occupied by Mediator revealed that, as in MM1.S cells, large genomic domains were occupied by this coactivator in both GBM and SCLC (Figure 7A, 7B, 7D, and 7E). The median super-enhancer was 30kb in GBM cells and 11kb in SCLC cells (Figure 7B and 7E). As in MM1.S cells, these GBM and SCLC super-enhancers were an order of magnitude larger and showed a commensurate increase in MED1, BRD4 and H3K27Ac levels when compared to normal enhancers (Figure 7B and 7E).

The super-enhancers in GBM and SCLC were found to be associated with many well-known tumor-associated genes (Figure 7C and 7F; Table S5). In GBM, super-enhancers were associated with genes encoding three transcription factors (RUNX1, FOSL2, BHLHE40) critical for mesenchymal transformation of brain tumors (Carro et al., 2010); the super-enhancers associated with BHLHE40 are shown in Figure 7C. BCL3, which associates with NFkB and is deregulated in many blood and solid tumor types, is associated with an upstream, and an intergenic super-enhancer in GBM (Figure 7C) (Maldonado and Melendez-Zajgla, 2011). In
SCLC, a super-enhancer is associated with the *INSM1* gene, which encodes a transcription factor involved in neuronal development that is highly expressed in neuroendocrine tissue and tumors such as SCLC (Figure 7F) (Pedersen et al., 2003). A super-enhancer is also associated with the *ID2* gene, which is highly expressed in SCLCs and encodes a protein that interacts with the well-known retinoblastoma tumor suppressor (Figure 7F) (Pedersen et al., 2003; Perk et al., 2005). These results indicate that super-enhancers are likely to associate with critical tumor oncogenes in diverse tumor types.
Figure 7
Figure 7: Super-enhancers are associated with key genes in other cancers

A,D) Total MED1 ChIP-Seq signal in units of reads per million in enhancer regions for all enhancers in A) the glioblastoma multiforme (GBM) cell line U-87 MG or D) the small cell lung cancer (SCLC) cell line H2171. Enhancers are ranked by increasing MED1 ChIP-Seq signal.

B,E) Meta-gene representation of global MED1 and BRD4 occupancy at B) typical GMB enhancers and super-enhancers or E) typical SCLC enhancers and super-enhancers. The x-axis shows the start and end of the enhancer (left) or super-enhancer (right) regions flanked by +/- 5kb of adjacent sequence. Enhancer and super-enhancer regions on the x-axis are relatively scaled. The y-axis shows the average signal in units of rpm/bp.

C,F) Gene tracks of MED1 and BRD4 ChIP-Seq occupancy at C) super-enhancers near BHLHE40 and BCL3, genes with important roles in GBM or at F) super-enhancers near INSM1 and ID2, genes with important roles in SCLC. Super-enhancers are depicted in gray boxes over the gene tracks. See also Figure S4.
Discussion

Chromatin regulators have become attractive targets for cancer therapy, but many of these regulators are expressed in a broad range of healthy cells and contribute generally to gene expression. Thus, it is unclear how inhibition of a global chromatin regulator such as BRD4 might produce selective effects, such as at the MYC oncogene (Delmore et al., 2011). We have found that key regulators of tumor cell state in MM1.S cells are associated with large enhancer domains, characterized by disproportionately high levels of BRD4 and Mediator. These super-enhancers are more sensitive to perturbation than typical enhancers and the expression of the genes associated with super-enhancers is preferentially affected. Thus, the preferential loss of BRD4 at super-enhancers associated with the MYC oncogene and other key tumor-associated genes can explain the gene-selective effects of JQ1 treatment in these cells.

BRD4 is an excellent example of a chromatin regulator that is expressed in a broad range of healthy cells and contributes generally to gene expression. Most cell types for which RNA-Seq data is available express the BRD4 gene. ChIP-seq data revealed that BRD4 generally occupies the enhancer and promoter elements of active genes with the Mediator coactivator in MM1.S cells (Figure 1). These results eliminate the model that BRD4 is exclusively associated with a small set of genes that are thereby rendered inactive by the BRD4 inhibitor JQ1 and instead suggest that the gene-specific effects of the small molecule have other causes.

We have found that ~3% of the enhancers in MM1.S cells are exceptionally large and are occupied by remarkably high amounts of BRD4 and Mediator. These super-enhancers are generally an order of magnitude larger and contain an order of magnitude more BRD4, Mediator and histone marks associated with enhancers (H3K27Ac) than typical enhancers. Our results suggest that super-enhancers are collections of closely spaced enhancers that can collectively
facilitate high levels of transcription from adjacent genes. Importantly, the super-enhancers are associated with the MYC oncogene and additional genes such as IGLL5, IRF4, PRDM1/BLIMP-1, and XBP1 that feature prominently in MM biology.

Co-operative and synergistic binding of multiple transcription factors and coactivators occurs at enhancers. Enhancers bound by many cooperatively-interacting factors can lose activity more rapidly than enhancers bound by fewer factors when the levels of enhancer-bound factors are reduced (Giniger and Ptashne, 1988; Griggs and Johnston, 1991). The presence of super-enhancers at MYC and other key genes associated with multiple myeloma led us to test the hypothesis that super-enhancers are more sensitive to reduced levels of BRD4 than average enhancers. We found that treatment of these tumor cells with the BET-bromodomain inhibitor JQ1 leads to preferential loss of BRD4 at super-enhancers. In addition, this decrease in BRD4 occupancy is accompanied by a corresponding loss of MED1 and CDK9 at super-enhancers. Consequent transcription elongation defects and mRNA decreases preferentially impact super-enhancer-associated genes, with an especially profound effect at the MYC oncogene.

Super-enhancers are not restricted to MM cells. We have identified super-enhancers in two additional tumor types, small cell lung cancer and glioblastoma multiforme. Super-enhancers identified in these cell types have characteristics similar to those found in MM1.S; they span large genomic regions and contain exceptional amounts of Mediator and BRD4. These super-enhancers are also associated with important tumor genes in both cell types. In GBM cells, BHLHE40 and BCL3 are known to be important in tumor biology, and are each associated with super-enhancers in this cell type. In H2171 SCLC cells, super-enhancers are associated with INSM1 and ID2, which are frequently overexpressed in SCLC.
Our results demonstrate that super-enhancers occupied by BRD4 regulate critical oncogenic drivers in multiple myeloma and show that BRD4 inhibition leads to preferential disruption of these super-enhancers. This insight into the mechanism by which BRD4 inhibition causes selective loss of oncogene expression in this highly malignant blood cancer may have implications for future drug development in oncology. Tumor cells frequently become addicted to oncogenes, thus becoming unusually reliant on high level expression of these genes (Cheung et al., 2011; Chin et al., 1999; Felsher and Bishop, 1999; Garraway and Sellers, 2006; Garraway et al., 2005; Jain et al., 2002; Weinstein, 2002). Thus, preferential disruption of super-enhancer function may be a general approach to selectively inhibiting the oncogenic drivers of many tumor cells.

Experimental Procedures

Cell Culture

MM1.S multiple myeloma cells (CRL-2974 ATCC) and U-87 MG glioblastoma cells (HTB-14 ATCC) were purchased from ATCC. H2171 small cell lung carcinoma cells (CRL-5929 ATCC) were kindly provided by John Minna, UT Southwestern. MM1.S and H2171 cells were propagated in RPMI-1640 supplemented with 10% fetal bovine serum and 1% GlutaMAX (Invitrogen, 35050-061). U-87 MG cells were cultured in Eagle's Minimum Essential Medium (EMEM) modified to contain Earles Balanced Salt Solution, nonessential amino acids, 2 mM L-glutamine, 1 mM sodium pyruvate, and 1500 mg/l sodium bicarbonate. Cells were grown at 37°C and 5% CO₂.
For JQ1 treatment experiments, cells were resuspended in fresh media containing JQ1 (5nM, 50nM, 500nM, 5000nM) or vehicle (DMSO, 0.05%) and treated for a duration of 6 hours, unless otherwise indicated.

**ChIP-Seq**

ChIP was carried out as described in Lin et al. (2012). Additional details are provided in supplemental methods. Antibodies used are as follows: total RNA Pol II (Rpb1 N-terminus): Santa Cruz sc-899 lot# K0111; MED1: Bethyl Labs A300-793A lot#A300-793A-2; BRD4: Bethyl Labs A301-985A lot# A301-985A-1; CDK9: Santa Cruz Biotechnology sc-484, lot D1612. ChIP-Seq datasets of H3K4Me3 and H3K27Ac in MML.S and MEDI and H3K27Ac in U-87 MG and H2717 were previously published (Lin et al., 2012).

**Luciferase Reporter Assays**

A minimal *Myc* promoter was amplified from human genomic DNA and cloned into the SacI and HindIII sites of the pGL3 basic vector (Promega). Enhancer fragments were likewise amplified from human genomic DNA and cloned into the BamHI and SalI sites of the pGL3-pMyc vector. All cloning primers are listed in Table S6. Constructs were transfected into MML.S cells using Lipofectamine 2000 (Invitrogen). The pRL-SV40 plasmid (Promega) was cotransfected as a normalization control. Cells were incubated for 24 hours, and luciferase activity was measured using the Dual-Luciferase Reporter Assay System (Promega). For the JQ1 concentration course, cells were resuspended in fresh media containing various concentrations of JQ1 24 hours after transfection, and incubated for an additional 6 hours before harvesting. Luminescence measurements were made using the Dual-Luciferase Reporter Assay System (Promega) on a Wallac EnVision (Perkin Elmer) plate reader.
Cell Viability Assays

Cell viability was measured using the CellTiterGlo Assay kit (Promega, G7571). MM1.S cells were resuspended in fresh media containing JQ1 (5nM, 50nM, 500nM, 1000nM) or vehicle (DMSO, 0.05%), then plated in 96-well plates at 10,000 cells/well in a volume of 100uL. Viability was measured after 6, 24, 48 and 72 hour incubations by addition of CellTiter Glo reagent and luminescence measurement on a Tecan Safire² plate reader.

Western Blotting

Western blots were carried out using standard protocols. Antibodies used are as follows: c-Myc (Epitomics, cat. #: 1472-1), BRD4 (Epitomics, cat.#: 5716-1) or β-Actin (Sigma, clone AC-15, A5441).

Data Analysis

All ChIP-Seq datasets were aligned using Bowtie (version 0.12.2) (Langmead et al., 2009) to build version NCBI36/HG18 of the human genome. Aligned and raw data can be found online associated with the GEO Accession ID GSE42355 (www.ncbi.nlm.nih.gov/geo/). Individual dataset GEO Accession IDs and background datasets used can be found in Table S7.

ChIP-Seq read densities in genomic regions was calculated as in (Lin et al. 2012). We used the MACS version 1.4.1 (Model based analysis of ChIP-Seq) (Zhang et al., 2008) peak finding algorithm to identify regions of ChIP-Seq enrichment over background. A p-value threshold of enrichment of 1e-9 was used for all datasets.
Active enhancers were defined as regions of ChIP-Seq enrichment for the Mediator complex component MED1 outside of promoters (e.g. a region not contained within +/- 2.5kb region flanking the promoter). In order to accurately capture dense clusters of enhancers, we allowed MED1 regions within 12.5kb of one another to be stitched together. To identify super-enhancers, we first ranked all enhancers by increasing total background subtracted ChIP-Seq occupancy of MED1 (x-axis), and plotted the total background subtracted ChIP-Seq occupancy of MED1 in units of total rpm (y-axis). This representation revealed a clear inflection point in the distribution of MED1 at enhancers. We geometrically defined the inflection point and used it to establish the cut off for super-enhancers (See supplemental methods).

Acknowledgments

We thank Zi Peng Fan for bioinformatics support; Michael R. McKeown for help with cell viability assays; Jun Qi for providing JQ1; Tom Volkert, Jennifer Love, Sumeet Gupta, and Jeong-Ah Kwon at the Whitehead Genome Technologies Core for Solexa sequencing; and members of the Young, Bradner and Vakoc labs for helpful discussion. This work was supported by a Swedish Research Council Postdoctoral Fellowship VR-B0086301 (J.L.), the Damon-Runyon Cancer Research Foundation (J.E.B), a Burroughs-Wellcome CAMS award and NCI Cancer Center Support Grant DevelopmentFund CA45508 (C.R.V.), and National Institutes of Health grants HG002668 (R.A.Y) and CA146445 (R.A.Y, T.L).
References


Chapter 3

Conclusions and future directions

Transcriptional enhancer regions play a critical role in human evolution, development, and disease. In the previous chapters, I discussed the functional properties of enhancer regions, how they influence transcription, and how these regions can be misregulated in disease. I described the identification of super-enhancers, and how these hubs of transcriptional regulation may be targets for therapeutic intervention in cancer. In this chapter, I describe further how super-enhancer regions may be utilized to identify key genes and to map the core regulatory circuitry in both healthy and diseased tissue, and how these enhancer regions may play an important role in the mechanism of action of drugs targeting transcription. Given the increasing promise of many inhibitors against chromatin regulators and transcriptional cofactors, understanding how super-enhancer regions influence drug response may inform and improve our ability to use these drugs in the treatment of human disease.

Super-enhancers mark key genes

Markers of key genes in healthy tissue

In both healthy and diseased tissues, super-enhancers are associated with many key genes that are known to control and define cell state. These include both transcription factors involved in directing cell-state-specific gene expression programs as well as genes involved directly in cell function. For example, in heart tissue, super-enhancer-associated genes include three transcription factors known to be key regulators of cardiac development (GATA4, Nkx2.5, and TBX5), as well as several genes involved in cardiac muscle function, such as MYH6, a cardiac-
specific component of the motor protein, myosin; myoglobin (MB), the oxygen carrying protein in muscle cells; and KCNJ1, a voltage-gated potassium channel involved in repolarization of cardiac muscle (Hnisz et al., 2013; McCulley and Black, 2012; Nerbonne and Kass, 2005). In another example, the nuclear receptor transcription factor PPARG and C/EBP transcription factor family members, including CEBPB and CEBPD, are important regulators of adipocyte differentiation, and are associated with super-enhancers in adipose tissue, as is lipoprotein lipase (LPL), an enzyme involved in the import of fatty acids for storage in adipose tissue (Hnisz et al., 2013; Rosen and MacDougald, 2006; Wang and Eckel, 2009). Thus, identification of super-enhancers in cell or tissue types where the functional and regulatory characteristics are less well understood may help us to uncover the basic biology of these tissues.

**Super-enhancers mark key genes in cancer**

Super-enhancers are also associated with key genes in cancer cells. Many known oncogenes, anti-apoptotic genes, cell cycle regulators, as well as genes involved in functions associated with the cell or origin are driven by super-enhancer regions. For example, in multiple myeloma, a cancer originating from antibody-producing plasma cells, the immunoglobulin light chain gene IGLL5 is associated with a super-enhancer, as is the transcription factor IRF4, which is involved in plasma cell development and is also important in multiple myeloma biology (Loven et al., 2013). In addition, many cancer-related genes appear to acquire super-enhancers. For example, anti-apoptotic factors, such as the Bcl2 family member BCL2L1 (Bcl-xL), are associated with a super-enhancer in T cell acute lymphoblastic leukemia (T-ALL), but not normal T cells (Hnisz et al., 2013). Genes for cell cycle regulators, such as cyclins, may also acquire super-enhancers: in colon carcinoma, CCND1 (cyclin D1) gains a super-enhancer, while in T-ALL, CCND3 (cyclin D3) becomes super-enhancer-associated (Hnisz et al., 2013). These
classes of genes are also associated with super-enhancers in multiple myeloma, with super-enhancers at both BCL2 and BCL2L1 as well as CCND2 (cyclin D2) (Loven et al., 2013). Acquired super-enhancers in tumor cells often fall into categories of genes involved in cancer hallmark traits, such as immune avoidance, induction of angiogenesis, evasion of growth suppressors, or activation of invasion or metastasis (Hnisz et al., 2013). For example, the highest-ranked super-enhancer acquired in the colon carcinoma cell line HCT-116 is PCDH7, protocadherin-7, an integral membrane protein involved in cell adhesion that has been strongly linked to metastasis in breast cancer (Bos et al., 2009; Hnisz et al., 2013; Li et al., 2013; Yoshida et al., 1998). Therefore, profiling of super-enhancers in cancer cells may help to identify key genes involved in tumor function and pathogenesis.

The identification of super-enhancers in tumor cells may improve our understanding of tumor cell function, and also suggest existing or novel therapeutic targets. In glioblastoma multiforme (GBM), for example, the super-enhancer-associated vascular endothelial growth factor (VEGFA) is targeted by the clinically approved monoclonal antibody Bevacizumab (brand name Avastin) (Friedman et al., 2009; Hnisz et al., 2013). In acute myeloid leukemia (AML), the highest-ranked super-enhancer is associated with CD47, which encodes a transmembrane protein that acts as a “don’t eat me” signal to macrophages, allowing tumor cells to avoid immune destruction (analysis unpublished, data available from (Shi et al., 2013); (Grimsley and Ravichandran, 2003)). Monoclonal antibodies targeting CD47 have shown promise in models of human AML and may soon be investigated clinically (Majeti et al., 2009). Thus, identifying the roughly 100-1000 super-enhancer-associated genes in a given tumor type may help to reduce the search space when investigating potential therapeutic targets.

*Identifying super-enhancers in patient samples*
Typical ChIP-seq experiments, such as those thus far used to identify super-enhancers, require the use of as many as 10 million cells, far more than possible to obtain from typical patient tumor samples. Continuing improvements in ChIP-seq technology, allowing for the use of small sample sizes and high throughput analysis, may improve our ability to identify super-enhancers in primary tumor cells (Adli and Bernstein, 2011; Blecher-Gonen et al., 2013; Gilfillan et al., 2012). In addition, new technologies that can be used for identifying enhancer regions, such as ATAC-Seq, may allow for the profiling of super-enhancers in as few as 500 cells. Modification of these techniques to allow for the identification of enhancer and super-enhancer regions in formalin-fixed, paraffin-embedded (FFPE) samples would also greatly increase the ability to assess patient samples (Fanelli et al., 2011; Fanelli et al., 2010). Lastly, it may be possible to identify super-enhancers through alternate means, such as the measurement of RNA produced by these regions. Given that single-cell methods of RNA sequencing are currently being investigated, this type of technology may dramatically improve our ability to use super-enhancers to understand basic tumor biology and cellular heterogeneity in the context of clinically relevant samples (Picelli et al., 2013; Ramskold et al., 2012).

**Mapping regulatory circuitry with super-enhancers**

As described previously, and in chapter 1, super-enhancers are often associated with key transcription factors that are known to regulate cell-state-specific gene expression programs, such as the well-characterized embryonic stem cell master transcription factors Oct4, Sox2, and Nanog. Being able to decipher the transcriptional regulatory circuitry in less well-understood tissue types would not only contribute to our understanding of the basic biology of these cells, but would also be useful in applications such as the reprogramming of cells for regenerative medicine. The human genome encodes over one thousand transcription factors, hundreds of
which may be expressed in a given cell type, making it difficult to predict which of these may form a small set of master regulators (Vaquerizas et al., 2009). Using super-enhancer regions may help to focus on the key transcription factors for a given cell type, facilitating the prediction of core circuitry in any cell type.

In mouse embryonic stem cells, the Oct4, Sox2 and Nanog master transcription factors form an interconnected autoregulatory loop that contributes to the regulation of the pluripotent state (Boyer et al., 2005). In this loop, each transcription factor is directly involved in activation of its own expression and that of the others, forming a core regulatory circuit. This type of regulatory motif can also occur in cancer. For example, in T-ALL, the oncogenic transcription factor TAL1 forms a core circuit with the transcription factors GATA3 and RUNX1, driving the malignant state in these cells (Sanda et al., 2012). Many known master transcription factors are driven by super-enhancers (Hnisz et al., 2013). As observed in mouse embryonic stem cells, Oct4, Sox2 and Nanog are associated with super-enhancer regions, which are in turn occupied by each of these transcription factors (Whyte et al., 2013). Thus, identification of super-enhancer regions may aid in building maps of the core regulatory circuitry in other cell types by identifying the set of transcription factors whose genes are associated with super-enhancers, and who occupy their own super-enhancers and those of the other transcription factors in the circuit. In lieu of obtaining ChIP-seq data for each transcription factor, it may also be possible to use the presence of a particular transcription factor’s binding motif in a super-enhancer region to indicate binding. This type of prediction would allow us to use super-enhancers to map the core regulatory circuitry in both healthy and diseased tissue, helping to identify the master regulators of gene expression in any cell type.

Super-enhancers and transcriptional therapeutics
Disruption of super-enhancers by bromodomain inhibitors in multiple myeloma

In addition to using super-enhancers to identify potential therapeutic targets, these enhancers may themselves be targets for therapeutic intervention through drugs acting on the transcriptional regulators that function at super-enhancers. As I described in chapter 2, in multiple myeloma, the BET bromodomain inhibitor, JQ1, appears to function through the selective disruption of highly BRD4-occupied super-enhancer regions, possibly due to the cooperative and synergistic interactions between the cofactors and transcriptional apparatus likely occurring at these large, clustered enhancer regions (Loven et al., 2013). Preferential loss of BRD4 at these regions is accompanied by a loss of transcriptional elongation at super-enhancer-associated genes and a concomitant loss in mRNA expression levels. The MYC oncogene is associated with the large IGH super-enhancers in multiple myeloma, and experiences a profound loss in transcription elongation, as evidenced by loss of RNA Pol II binding in the gene body, as well as a dramatic reduction in mRNA levels. Overexpression of MYC is capable of partially rescuing tumor cells from the effects of JQ1 treatment, suggesting that the loss of this potent oncogene plays an important role in JQ1 response, although it is likely that other factors are involved (Delmore et al., 2011). Because super-enhancers are associated with many genes involved in tumor pathogenesis, it is likely that loss of several of these important genes may contribute to the effect of JQ1. These results are significant, in part because of the difficulty of directly targeting an oncogenic transcription factor such as c-MYC with conventional small molecule inhibitors. Despite the fact that JQ1 targets transcription elongation, a general transcriptional process, this inhibitor appears to have a relatively selective effect on gene expression, with a profound effect on expression of the MYC oncogene in multiple myeloma (Delmore et al., 2011; Loven et al., 2013).
Bromodomain inhibitors act on super-enhancers in many tumor types

The mechanism described in chapter 2 in multiple myeloma may be a general principle for understanding the mechanism of action of bromodomain inhibitors in other cell types. For example, similar results were found in diffuse large B cell lymphoma (DLBCL), where MYC and several other tumor-associated genes, including BCL6 and POU2AF1 (OCA-B), are associated with super-enhancers, and are also down-regulated by JQ1 treatment (Chapuy et al., 2013). Similarly, in AML, the EVII gene is driven by the GATA2 super-enhancer through a translocation event, and both the enhancer region and the EVII gene are profoundly affected by JQ1, experiencing a loss in BRD4 occupancy as well as transcription (Groeschel et al., 2014). In other AML cell lines, where EVII was expressed, but not occupied by a super-enhancer containing such high levels of BRD4, the gene was not sensitive to JQ1 treatment (Groeschel et al., 2014). Therefore, bromodomain inhibitors may act through the disruption of super-enhancers in many other cancer types in addition to multiple myeloma, as described in chapter 2.

The role of super-enhancers in other transcriptional therapeutics

As described in previous chapters, there is currently great interest in developing small molecule inhibitors targeted against factors involved in the transcription cycle. In addition to bromodomain inhibitors, such as JQ1, many inhibitors target other chromatin regulators, including histone deacetylases (HDACs), histone methyl transferases such as EZH2 or DOT1L, and histone demethylases such as LSD1. In addition, many drugs are being developed to target kinases involved in the transcriptional cycle, including CDK7, CDK8, and CDK9 (Stellrecht and Chen, 2011; Villicana et al., 2014). Understanding how these drugs affect transcription will be
important in making further improvements in designing these inhibitors and employing them in cancer therapy.

A major challenge in the use of these therapeutics is understanding how transcriptional inhibition can achieve a favorable therapeutic index, allowing for the specific targeting of tumor cells. Several mechanisms have been suggested by which these transcriptional inhibitors may act on tumor cells. First, tumor cells could be generally dependent on high levels of transcription, as evidenced by the transcriptional amplification effect induced by over expression of the c-MYC oncogene. Second, many oncogenes have short-lived mRNA and protein products, so these may be among the first, and most profoundly affected genes in response to inhibition of transcription. For example, the c-MYC oncogene produces exceptionally short-lived products, with both mRNA and protein half-lives of around 30 minutes (Dani et al., 1984; Ramsay et al., 1984). However, our evidence suggests that certain transcriptional inhibitors, such as JQ1, may not act by profoundly downregulating global transcription, but rather function by targeting super-enhancer regions in the genome, leading to a more selective down-regulation of specific genes. As described previously, JQ1 treatment induces defects in transcription elongation that are more pronounced at super-enhancer-associated genes, suggesting that the disruption of super-enhancer regions has a role in JQ1’s action, and mRNA and protein half-life alone may not account for the selective effect of this inhibitor (Loven et al., 2013). At a cellular level, tumor cells may become addicted or dependent on certain oncogenes or genes involved in other tumor-promoting pathways, as described above (Luo et al., 2009). Because super-enhancers are associated with many of these key categories of tumor genes, their disruption by JQ1 may lead to adverse effects on tumor cells.
This mechanism may extend to other transcripational inhibitors. For example, inhibition of the Notch transcription factor by a gamma secretase inhibitor (GSI) produced similar results in T-lymphoblastic leukemia (T-LL), where super-enhancer regions and their associated genes were the most affected by GSI treatment (Wang et al., 2014). Thus, super-enhancers may represent regions of the genome with heightened requirements for transcriptional activators, and heightened sensitivity to inhibition of these factors. ChIP-Seq profiling of transcriptional drug targets, and identification of super-enhancers based on these factors, may help to key in on genomic regions and genes that may be particularly sensitive to inhibitor treatment. In addition, the recent development of Chem-Seq technology allows the profiling of drug-DNA interactions, which may aid in finding those regions of the genome that are highly targeted by transcriptional inhibitors (Anders et al., 2014). Examining the effect of drug treatment on transcription itself, either by profiling RNA Pol II occupancy, or nascent transcription by GRO-Seq, may also improve our understanding of how these drugs act (Core et al., 2008). Measuring transcription directly rather than measuring the resultant steady-state mRNA levels may reveal any selective effects of these inhibitors that are separate from the half-life of the gene products. This further investigation of the many promising transcriptional inhibitors should indicate the extent to which super-enhancers play a common role in the mechanism of action of each of the varying classes of these drugs.

Conclusion

In summary, super-enhancer regions are associated with key genes that have important roles in cell function and in the control of cell-type-specific gene expression programs. In addition, super-enhancer regions can be particularly sensitive to perturbation by inhibitors targeting transcription. Super-enhancers may therefore act as biomarkers, allowing us to identify
genes that play critical roles in cellular function, both in healthy tissues and in tumor cells. In addition, by identifying transcription factors that are both driven by super-enhancers, and act at super-enhancers, we may be able to map the core regulatory circuitry in any human cell type. Thus, profiling super-enhancers could contribute to our knowledge of the basic function and regulation of a particular cell or tissue type, and in tumor cells, may suggest important targets for therapeutic intervention. Super-enhancers themselves may play an important role in the mechanism of action of transcriptional inhibitors, a major new category of potential cancer therapeutics. Better understanding of the basic properties of super-enhancers, their location in healthy and diseased tissues, and how these regions are affected by transcriptional inhibitors may lead to an improved ability to characterize and treat human disease.
References


Appendix A

Supplementary material for chapter 2

SUPPLEMENTAL INFORMATION

Supplemental Data

Table S1: Actively transcribed genes in studied cell lines

Table S2: Active enhancers in multiple myeloma

Table S3: Super-enhancer-associated genes in multiple myeloma

Table S4: Active enhancers in glioblastoma multiforme and small cell lung cancer

Table S5: Super-enhancer-associated genes in glioblastoma multiforme and small cell lung cancer

Table S6: List of primers used to clone enhancer regions

Table S7: ChIP-Seq datasets and backgrounds

Data S1: Enhancer and super-enhancer .bed file

Figure S1: Mediator and BRD4 co-occupy promoters of active genes in multiple myeloma

Figure S2: Super-enhancers identified in multiple myeloma

Figure S3: Transcription of super-enhancer-associated genes is highly sensitive to BRD4 inhibition

Figure S4: Super-enhancers are associated with key genes in other cancers

Supplemental Experimental Procedures

124
Cell Culture

Chromatin Immunoprecipitation (ChIP)

Illumina Sequencing and Library Generation

Luciferase Reporter Assays

Cell Viability Assays

Western Blotting

RNA Extraction and Synthetic RNA Spike-In

Microarray Sample Preparation and Analysis

Data Analysis

Supplemental References
Tables and Figures

Table S1, related to Figures 1, 2, 4, 6 and 7: Actively transcribed genes in multiple myeloma

Table of actively transcribed genes in the multiple myeloma cell line MM1.S.

Table S2, related to Figures 1, 2, 3, 4, 5, and 6: Active enhancers in multiple myeloma

Table of active enhancers in the multiple myeloma cell line MM1.S. Genomic coordinates are provided for all identified enhancers. Enhancers that have been classified as super-enhancers are indicated. Total background subtracted mediator (MED1) signal at each enhancer is also provided.

Table S3, related to Figures 2, 3, and 6: Super-enhancer-associated genes in multiple myeloma

Table of super-enhancer-associated actively transcribed genes in the multiple myeloma cell line MM1.S (see supplemental methods). For each gene, the coordinates of the proximal super-enhancers are provided. Genes that are annotated as altered in cancer in either (Patel et al., 2013) or (Futreal et al., 2004) are indicated.

Table S4, related to Figure 7: Active enhancers in glioblastoma multiforme and small cell lung cancer

Table of active enhancers in the glioblastoma multiforme (GBM) cell line U-87 and the small cell lung cancer (SCLC) cell line H2171. Genomic coordinates are provided for all identified enhancers. Enhancers that have been classified as super-enhancers are indicated. Total background subtracted mediator (MED1) signal at each enhancer is also provided.
Table S5, related to Figure 7: Super-enhancer-associated genes in glioblastoma multiforme and small cell lung cancer

Table of super-enhancer-associated actively transcribed genes in the glioblastoma multiforme (GBM) cell line U-87 and the small cell lung cancer (SCLC) cell line H2171 (see supplemental methods). For each gene, the coordinates of the proximal super-enhancers are provided. Genes that are annotated as altered in cancer in either (Patel et al., 2013) or (Futreal et al., 2004) are indicated.

Table S6, related to Figure 2 and Figure 6: Primers used to clone enhancer regions

This table lists primer sequences for all cloned regions used in reporter assays.

Table S7, related to Figures 1 through Figure 7: ChIP-Seq datasets and backgrounds

This table lists all ChIP-Seq datasets used in this study, their GEO accession number, and the appropriate background used to define enriched regions.

Data S1: Enhancer and super-enhancer .bed file

This file is a UCSC genome browser .bed format file (genome.ucsc.edu) containing the genomic coordinates for all enhancers and super-enhancers in profiled cell lines. Coordinates are in genome build hg18. For each cell line, individual tracks are provided for all enhancers (colored in black) or super-enhancers only (colored in red).
Figure S1, related to Figure 1: Mediator and BRD4 co-occupy promoters of active genes in multiple myeloma

A) Scatter plots showing the relationship between MED1 (x-axis) and BRD4 (y-axis) ChIP-Seq occupancy at individual enhancers in MM1.S cells. Units are in total ChIP-Seq reads per million in enhancer regions. The Pearson correlation score between MED1 and BRD4 is displayed ($\rho = 0.91$).

B) Scatter plots showing the relationship between MED1 (x-axis) and BRD4 (y-axis) ChIP-Seq occupancy at individual promoters of actively transcribed genes in MM1.S cells. Units are in total ChIP-Seq reads per million in enhancer regions. The Pearson correlation score between MED1 and BRD4 is displayed ($\rho = 0.81$).

C) Pie chart showing the overlap of BRD4 regions with either MED1 or H3K27Ac regions. BRD4 regions that overlap either a MED1 or H3K27Ac enriched region are shaded in red. BRD4 regions without any overlap are shaded in grey.

Figure S2, related to Figure 2: Super-enhancers identified in multiple myeloma

A) Left: Total MED1 ChIP-Seq signal in units of reads per million in enhancer regions for all enhancers in MM1.S. Enhancers are ranked by increasing MED1 ChIP-Seq signal. Right: Total MED1 ChIP-Seq signal in units of reads per million in promoter regions for all promoters of actively transcribed genes in MM1.S that do not overlap with a super-enhancer. Promoters are ranked by increasing MED1 ChIP-Seq signal. In both plots, Y-axis shows MED1 signal from 0 to 500,000.
Figure S3, related to Figure 6: Transcription of super-enhancer-associated genes is highly sensitive to BRD4 inhibition

A,B) Heatmap of changes in gene expression for all actively transcribed genes after treatment with 500nM JQ1 at various time points A) or after treatment with various doses of JQ1 for 6 hours B). Each line in the heatmap shows a single gene and is shaded relative to the change in gene expression compared to control untreated cells. Genes are ranked by average fold loss across the time course. The key on the right shows the relationship between color and change in gene expression with green indicative of loss of gene expression.

C,D) Line graph showing the Log₂ change in gene expression vs. control cells after JQ1 treatment in a time C) or dose D) dependent manner for genes associated with multiple typical enhancers (grey line) or genes associated with super-enhancers (red line). The Y-axis shows the Log₂ change in gene expression of JQ1 treated vs. untreated control cells. X-axis shows time of 500nM JQ1 treatment C) or JQ1 treatment concentration at 6 hours D).

E) Meta-gene representation of global CDK9 occupancy at enhancers and promoters. The x-axis shows the +/- 2.5kb region flanking either the center of enhancer regions (left) or the TSS of active genes. The y-axis shows the average background subtracted ChIP-Seq signal in units of rpm/bp.

F) Bar graphs showing the percentage loss of either MED1 (left, red) or CDK9 (right, green) at super-enhancer regions that are TSS proximal (left) or TSS distal (right). Error bars represent 95% confidence intervals of the mean.

G) Histograms showing the distribution of changes in RNA Pol II density in the elongating gene body region of genes for typical enhancer-associated genes (top, grey) or super-enhancer-
associated genes (bottom, red). The median of the distribution is shown with a line. X-axis shows the Log2 change in RNA Pol II density in the elongating gene body region. Y-axis shows probability density of the distribution. The difference is distributions is significant by a Kolmogorov-Smirnov (KS) test (p-value < 2e-16).

Figure S4, related to Figure 7: Super-enhancers are associated with key genes in other cancers

A) Meta-gene representation of global MED1, BRD4, H3K27Ac, and H3K4Me3 occupancy at enhancers and promoters in the glioblastoma multiforme cell line U-87. The x-axis shows the +/- 2.5kb region flanking either the center of enhancer regions (left) or the transcription start site (TSS) of active genes. The y-axis shows the average background subtracted ChIP-Seq signal in units of rpm/bp.

B) Meta-gene representation of global MED1, BRD4, H3K27Ac, and H3K4Me3 occupancy at enhancers and promoters in the small cell lung cancer cell line H2171. The x-axis shows the +/- 2.5kb region flanking either the center of enhancer regions (left) or the transcription start site (TSS) of active genes. The y-axis shows the average background subtracted ChIP-Seq signal in units of rpm/bp.
**Figure S1**

(A) BRD4 signal at enhancers (rpm) vs. MED1 signal at enhancers (rpm) with a correlation coefficient $p = 0.91$.

(B) BRD4 signal at promoters (rpm) vs. MED1 signal at promoters (rpm) with a correlation coefficient $p = 0.81$.

(C) Pie chart showing 18,726 BRD4 regions. 12% are BRD4 only, and 88% overlap with MED1 or H3K27Ac.
Enhancers ranked by increasing MED1 signal

Promoters ranked by increasing MED1 signal

Figure S2
Figure S3
Figure S4
Materials and Methods

Experimental Procedures

Cell Culture

MM1. S multiple myeloma cells (CRL-2974 ATCC) and U-87 MG glioblastoma cells (HTB-14 ATCC) were purchased from ATCC. H2171 small cell lung carcinoma cells (CRL-5929 ATCC) were kindly provided by John Minna, UT Southwestern. MM1. S and H2171 cells were propagated in RPMI-1640 supplemented with 10% fetal bovine serum and 1% GlutaMAX (Invitrogen, 35050-061). U-87 MG cells were cultured in Eagle’s Minimum Essential Medium (EMEM) modified to contain Earles Balanced Salt Solution, nonessential amino acids, 2 mM L-glutamine, 1 mM sodium pyruvate, and 1500 mg/l sodium bicarbonate. Cells were grown at 37°C and 5% CO₂.

JQ1 was kindly provided by James Bradner, Dana Farber Cancer Institute. For JQ1 dose dependence experiments, cells were resuspended in fresh media containing JQ1 (5nM, 50nM, 500nM, 5000nM) or vehicle (DMSO, 0.05%) for a duration of 6 hours, unless otherwise indicated. For JQ1 time course experiments, cells were resuspended in fresh media containing 500nM JQ1 and sampled at various time points.

Chromatin Immunoprecipitation (ChIP)
Cells were crosslinked for 10 minutes at room temperature by the addition of one-tenth of the volume of 11% formaldehyde solution (11% formaldehyde, 50mM Hepes pH 7.3, 100mM NaCl, 1mM EDTA pH 8.0, 0.5mM EGTA pH 8.0) to the growth media followed by 5 minutes quenching with 100mM glycine. Cells were washed twice with PBS, then the supernatant was aspirated and the cell pellet was flash frozen in liquid nitrogen. Frozen crosslinked cells were stored at -80°C. 100ul of Dynal magnetic beads (Sigma) were blocked with 0.5% BSA (w/v) in PBS. Magnetic beads were bound with 10ug of the indicated antibody.

For BRD4 occupied genomic regions, we performed ChIP-Seq experiments using a Bethyl Laboratories (A301-985A, lot A301-985A-1) antibody. The affinity purified antibody was raised in rabbit against an epitope corresponding to amino acids 1312-1362 of human BRD4. Antibody specificity was previously determined (Dawson et al., 2011). For MED1 (CRSP1/TRAP220) occupied genomic regions, we performed ChIP-Seq experiments using a Bethyl Laboratories (A300-793A, lot A300-783A) antibody. The affinity purified antibody was raised in rabbit against an epitope corresponding to amino acids 1523-1581 mapping at the C-terminus of human MED1. For CDK9 occupied genomic regions, we performed ChIP-Seq experiments using a Santa Cruz Biotechnology (sc-484, lot D1612) antibody. For RNA polymerase II occupied genomic regions, we performed ChIP-Seq experiments using a Santa Cruz Biotechnology (sc-899, lot K0111) antibody. The affinity purified antibody was raised in rabbit against an epitope mapping to the N-terminus of murine RBP1, the largest subunit of RNA Pol II.
For MM1.S, crosslinked cells were lysed with lysis buffer 1 (50 mM Hepes pH 7.3, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, and 0.25% Triton X-100) and resuspended and sonicated in sonication buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA pH 8.0, 0.1% SDS, and 1% TritonX-100). Cells were sonicated for 10 cycles at 30 seconds each on ice (18-21 watts) with 60 seconds on ice between cycles. Sonicated lysates were cleared and incubated overnight at 4°C with magnetic beads bound with antibody to enrich for DNA fragments bound by the indicated factor. Beads were washed two times with sonication buffer, one time with sonication buffer with 500 mM NaCl, one time with LiCl wash buffer (10 mM Tris pH 8.0, 1 mM EDTA, 250 mM LiCl, 1% NP-40,) and one time with TE with 50 mM NaCl. DNA was eluted in elution buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS). Cross-links were reversed overnight. RNA and protein were digested using RNase A and Proteinase K, respectively, and DNA was purified with phenol chloroform extraction and ethanol precipitation.

For BRD4 ChIPs in U-87 and H2171, crosslinked cells were lysed with lysis buffer 1 (50 mM HEPES [pH 7.3], 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, and 0.25% Triton X-100) and washed with lysis buffer 2 (10 mM Tris-HCl [pH 8.0], 200 mM NaCl, 1 mM EDTA [pH 8.0] and 0.5 mM EGTA [pH 8.0]). Cells were resuspended and sonicated in sonication buffer (50 mM Tris-HCl [pH 7.5], 140 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% Triton X-100, 0.1% SDS) for nine cycles at 30 s each on ice (18–21 W) with 60 s on ice between cycles. Sonicated lysates were cleared and incubated overnight at 4°C with magnetic beads bound with antibody to enrich for DNA fragments bound by the indicated factor. Beads were washed three times with sonication buffer, one time with sonication buffer with 500 mM NaCl, one time with LiCl wash buffer (20 mM Tris [pH 8.0], 1 mM EDTA, 250 mM LiCl, 0.5% NP-40, 0.5% Na-
deoxycholate) and one time with TE. DNA was eluted in elution buffer. Crosslinks were reversed overnight. RNA and protein were digested using RNase A and Proteinase K, respectively, and DNA was purified with phenol chloroform extraction and ethanol precipitation.

ChIP-Seq datasets of H3K4Me3 and H3K27Ac in MM1.S, and MED1 and H3K27Ac in U-87 MG and H2171 were previously published (Lin et al., 2012).

**Illumina Sequencing and Library Generation**

Purified ChIP DNA was used to prepare Illumina multiplexed sequencing libraries. Libraries for Illumina sequencing were prepared following the Illumina TruSeq™ DNA Sample Preparation v2 kit protocol with the following exceptions. After end-repair and A-tailing, Immunoprecipitated DNA (~10-50ng) or Whole Cell Extract DNA (50ng) was ligated to a 1:50 dilution of Illumina Adapter Oligo Mix assigning one of 24 unique indexes in the kit to each sample. Following ligation, libraries were amplified by 18 cycles of PCR using the HiFi NGS Library Amplification kit from KAPA Biosystems. Amplified libraries were then size-selected using a 2% gel cassette in the Pippin Prep™ system from Sage Science set to capture fragments between 200 and 400 bp. Libraries were quantified by qPCR using the KAPA Biosystems Illumina Library Quantification kit according to kit protocols. Libraries with distinct TruSeq indexes were multiplexed by mixing at equimolar ratios and running together in a lane on the Illumina HiSeq 2000 for 40 bases in single read mode.

**Luciferase Reporter Assays**
A minimal Myc promoter was amplified from human genomic DNA and cloned into the SacI and HindIII sites of the pGL3 basic vector (Promega). Enhancer fragments were likewise amplified from human genomic DNA and cloned into the BamHI and SalI sites of the pGL3-pMyc vector. All cloning primers are listed in Table S6. Constructs were transfected into MM1.S cells using Lipofectamine 2000 (Invitrogen). The pRL-SV40 plasmid (Promega) was cotransfected as a normalization control. Cells were incubated for 24 hours, and luciferase activity was measured using the Dual-Luciferase Reporter Assay System (Promega). For the JQ1 concentration course, cells were resuspended in fresh media containing various concentrations of JQ1 24 hours after transfection, and incubated for an additional 6 hours before harvesting. Luminescence measurements were made using the Dual-Luciferase Reporter Assay System (Promega) on a Wallac EnVision (Perkin Elmer) plate reader.

Cell Viability Assays
Cell viability was measured using the CellTiterGlo Assay kit (Promega, G7571). MM1.S cells were resuspended in fresh media containing JQ1 (5nM, 50nM, 500nM, 1000nM) or vehicle (DMSO, 0.05%), then plated in 96-well plates at 10,000 cells/well in a volume of 100uL. Viability was measured after 6, 24, 48 and 72 hour incubations by addition of CellTiter Glo reagent and luminescence measurement on a Tecan Safire2 plate reader.

Western Blotting
Total cell lysates for immunoblotting were prepared by pelleting 1x10^6 cells from each cell line at 4°C (1,200 rpm) for 5 min using a Sorvall Legend centrifuge (Thermo Fisher Scientific). After collecting the cells the pellets were washed 1X with ice-cold 1X PBS and the pellet after the
final wash was resuspended in RIPA lysis buffer (Sigma, R0278) containing 150 mM NaCl, 1.0% IGEPAL® CA-630, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris [pH 8.0], 2X Halt Protease inhibitors (Pierce, Thermo Fisher Scientific), and 2X Phosphatase inhibitor cocktail 2 (Sigma, P5726) and 3 (Sigma, P0044). Total protein lysates were collected and snap-frozen in liquid nitrogen before being stored at -80°C. Protein concentrations were determined by using the Bradford protein assay (Bio-Rad, #500-0006). Total cell lysates were loaded per well onto NuPAGE NOVEX 4%-12% Bis-Tris Mini Gel (Invitrogen, Carlsbad, CA) and separated by electrophoreses at 200 V for 1 hr. The gels were then transferred onto a PVDF membrane (Immobilon-P; Millipore, Billerica, MA) by semi-dry transfer system (TransBlot SD; Bio-Rad) and blocked by incubation with 5% dry milk in TBST (TBS with 0.2% Tween-20). Membranes were probed using antibodies raised against c-Myc (Epitomics, cat. #: 1472-1), BRD4 (Epitomics, cat.#: 5716-1) or β-Actin (Sigma, clone AC-15, A5441). Chemiluminescent detection was performed with appropriate secondary antibodies and developed using high-resolution BioMax MR film (Kodak, cat. #: 870 13012).

**RNA Extraction and Synthetic RNA Spike-In**

Total RNA and sample preparation was performed as previously described (Loven et al., 2012). Briefly, MM1.S cells were were incubated in JQ1-containing media (5nM, 50nM, 500nM, 5000nM) or vehicle (DMSO, 0.05%) for a duration of 6 hours. Cell numbers were determined by manually counting cells using C-Chip disposable hemocytometers (Digital Bio, DHC-N01) prior to lysis and RNA extraction. Biological duplicates (equivalent to ten million cells per replicate) were subsequently collected and homogenized in 1 mL of TRIzol Reagent (Life Technologies, 15596-026), purified using the mirVANA miRNA isolation kit (Ambion, AM1560) following
the manufacturer's instructions and re-suspended in 100μL nuclease-free water (Ambion, AM9938). Total RNA was spiked-in with ERCC RNA Spike-In Mix (Ambion, 4456740), treated with DNA-free™ DNase I (Ambion, AM1906) and analyzed on Agilent 2100 Bioanalyzer for integrity. RNA with the RNA Integrity Number (RIN) above 9.8 was hybridized to GeneChip® PrimeView Human Gene Expression Arrays (Affymetrix).

Microarray Sample Preparation and Analysis

For microarray analysis, 100ng of total RNA containing ERCC RNA Spike-In Mix (see above) was used to prepare biotinylated aRNA (cRNA) according to the manufacturer’s protocol (3’ IVT Express Kit, Affymetrix 901228). GeneChip arrays (Primeview, Affymetrix 901837) were hybridized and scanned according to standard Affymetrix protocols. All samples were processed in technical duplicate. Images were extracted with Affymetrix GeneChip Command Console (AGCC), and analyzed using GeneChip Expression Console. A Primeview CDF that included probe information for the ERCC controls, provided by Affymetrix, was used to generate .CEL files. We processed the CEL files using standard tools available within the affy package in R. The CEL files were processed with the expresso command to convert the raw probe intensities to probeset expression values. The parameters of the expresso command were set to generate Affymetrix MAS5-normalized probeset values. We used a loess regression to re-normalize these MAS5 normalized probeset values, using only the spike-in probesets to fit the loess. The affy package provides a function, loess.normalize, which will perform loess regression on a matrix of values (defined using the parameter mat) and allows for the user to specify which subset of data to use when fitting the loess (defined using the parameter subset, see the affy package documentation for further details). For this application the parameters mat and subset were set as
the MAS5-normalized values and the row-indices of the ERCC control probesets, respectively. The default settings for all other parameters were used. The result of this was a matrix of expression values normalized to the control ERCC probes. The probeset values from the duplicates were averaged together and the $\log_2$ fold change comparing the control to the JQ1 treated samples are shown. Spike-in normalized gene expression tables can be found online associated with the GEO Accession ID GSE44929 (www.ncbi.nlm.nih.gov/geo/).
Data Analysis

Accessing data generated in this manuscript

All ChIP-Seq and expression data generated in this manuscript can be found online associated with GEO Accession ID GSE44931 (www.ncbi.nlm.nih.gov/geo/).

Gene sets and annotations

All analysis was performed using RefSeq (NCBI36/HG18) (Pruitt et al., 2007) human gene annotations.

BRD4 expression in human tissues

Processed gene expression values for 90 distinct normal tissues were retrieved from the Human Body Index (GSE7307). BRD4 expression values were calculated by averaging expression values for three distinct probes that aligned to the sense strand of the BRD4 transcript. Since in any given tissue, approximately 50% of all genes are expressed (Ramskold et al., 2009), we classified BRD4 as expressed in a given tissue if the average expression values of its probes was greater than the median expressed probe in the respective tissue. Using this metric, BRD4 was determined to be expressed in 95% of all human tissues.

ChIP-Seq data processing
All ChIP-Seq datasets were aligned using Bowtie (version 0.12.2) (Langmead et al., 2009) to build version NCBI36/HG18 of the human genome. Alignments were performed using the following criteria: -n2, -e70, -m1, -k1, --best. These criteria preserved only reads that mapped uniquely to the genome with 1 or fewer mismatches. Aligned and raw data can be found online associated with the GEO Accession ID GSE42355 (www.ncbi.nlm.nih.gov/geo/).

Calculating read density

We developed a simple method to calculate the normalized read density of a ChIP-Seq dataset in any region. ChIP-Seq reads aligning to the region were extended by 200bp and the density of reads per basepair (bp) was calculated. In order to eliminate PCR bias, multiple reads of the exact same sequence aligning to a single position were collapsed into a single read. Only positions with at least 2 overlapping extended reads contributed to the overall region density. The density of reads in each region was normalized to the total number of million mapped reads producing read density in units of reads per million mapped reads per bp (rpm/bp).

Identifying ChIP-Seq enriched regions

We used the MACS version 1.4.1 (Model based analysis of ChIP-Seq) (Zhang et al., 2008) peak finding algorithm to identify regions of ChIP-Seq enrichment over background. A p-value threshold of enrichment of 1e-9 was used for all datasets. The GEO accession number and
background used for each dataset can be found in the accompanying supplementary file (Table S6).

Defining actively transcribed genes

In MM1.S and U-87 cells, a gene was defined as actively transcribed if enriched regions for H3K4me3, RNA Polymerase II (RNA Pol II), BRD4, and Mediator were located within +/- 5kb of the TSS. In H2171 cells, a gene was defined as actively transcribed if enriched regions for H3K4me3, RNA Polymerase II (RNA Pol II), and BRD4, were located within +/- 5kb of the TSS. Mediator was omitted in H2171 transcribed gene definitions due to the lower numbers of enriched regions for the factor. H3K4me3 is a histone modification associated with transcription initiation (Guenther et al., 2007) and BRD4 and Mediator have been shown to also occupy promoters of active genes (Kagey et al., 2010; Zhang et al., 2012). Using these criteria, we identified transcriptionally active genes in MM1.S, U-87, and H2171 cells (Table S1).

Defining active enhancers

Active enhancers form loops with promoters that are facilitated by the Mediator complex (Kagey et al., 2010), thus we used regions of ChIP-Seq enrichment for the Mediator complex component MED1 outside of promoters (e.g. a region not contained within +/- 2.5kb region flanking the promoter) to define active enhancers. In order to accurately capture dense clusters of enhancers, we allowed MED1 regions within 12.5kb of one another to be stitched together (distance criteria established in the accompanying mouse embryonic stem cell manuscript). We noticed that in
some cases stitched regions in gene dense areas spanned multiple transcriptionally active genes and were likely not representative of enhancer clusters. Thus, stitched regions spanning more than two promoters of active genes were excluded from the analysis. This analysis defined enhancer regions in the MM1.S genome (Table S2). The same method was used to identify enhancers in the glioblastoma multiforme cell line U-87 and in the small cell lung cancer cell line H2171 (Table S4). Data S1 contains hg18 genome coordinate tracks for all enhancers in all profiled cell types in the form of a UCSC genome browser .bed file that can be uploaded and viewed at genome.ucsc.edu (Data S1).

Creating meta representations of ChIP-Seq occupancy at active enhancers and promoters

Genome-wide average “meta” representations of ChIP-Seq occupancy at active enhancers and promoters were created by mapping ChIP-Seq read density to the 5kb regions flanking the center of active enhancer regions or transcription start sites (TSS) of active genes. Each active enhancer or TSS region was split into one hundred 50bp bins. All active enhancer or TSS regions were then aligned and the average background subtracted ChIP-Seq factor density in each bin was calculated to create a meta genome-wide average in units of rpm/bp. See Figure 1B; Figure 4D; Figure S3E; Figure S4A and S4B.

Correlating BRD4 and MED1 occupancy with RNA Pol II at promoters

All genes were ranked by increasing density of RNA Pol II ChIP-Seq reads in the promoter region (+/- 1kb of transcription start site (TSS) and binned in increments of 100 genes. The
median ChIP-Seq density for promoter regions within each bin was calculated in rpm/bp for both RNA Pol II and BRD4 and plotted (BRD4) or shaded by binding density (RNA Pol II). The same method was applied for MED1 in place of BRD4. See Figure 1C.

*Defining promoter boundaries in MM1.S cells*

Promoter regions are typically defined as the +/- several kb region centered on the TSS. In order to produce a more accurate and gene specific definition of promoters, we used the promoter centric H3K4Me3 histone modification mark to delineate the span of promoters for actively transcribed genes in MM1.S. At these genes, the boundaries of the directly overlapping H3K4Me3 ChIP-Seq enriched region were used to define the upstream/downstream ends of the promoter.

*Calculating total ChIP-Seq occupancy signal at enhancers and promoters*

The total ChIP-Seq occupancy signal at enhancers was calculated by first determining the average ChIP-Seq read density in the entire enhancer region (rpm/bp). This value was multiplied by the length of the region to produce total ChIP-Seq occupancy in units of total rpm. See Figure 2A; Figure S2A; Figure 7A and 7B. For all cell lines used in this study the total MED1 ChIP-Seq occupancy signal at enhancers can be found in Table S2 (MM1.S) or Table S4 (GBM and SCLC).
The total ChIP-Seq occupancy signal at promoters was calculated similarly. In order to accurately compare signal at promoters to that at enhancers, promoters that overlapped super-enhancers were excluded from analysis (Figure S2A).

Correlations of total BRD4 and MED1 occupancy signal at enhancers and promoters were calculated using a Pearson correlation statistic (Figure S1A and S1B).

**Identifying super-enhancers**

We observed disproportionately high occupancy of factors such as MED1 and BRD4 at a subset of enhancers that we termed “super-enhancers”. To formally identify super-enhancers, we first ranked all enhancers by increasing total background subtracted ChIP-Seq occupancy of MED1 (x-axis), and plotted the total background subtracted ChIP-Seq occupancy of MED1 in units of total rpm (y-axis). This representation revealed a clear inflection point in the distribution of MED1 at enhancers (Figure 2A; Figure 7A and 7D). We geometrically defined the inflection point and used it to establish the cut off for super-enhancers. Enhancers that were classified as super enhancers are noted in Table S2 (MM1.S) and Table S4 (GBM and SCLC).

**Creating meta representations of ChIP-Seq occupancy at typical enhancers and super-enhancers**

Genome-wide average “meta” representations of MED1 and BRD4 ChIP-Seq occupancy at typical enhancers and super-enhancers were created by mapping MED1 and BRD4 ChIP-Seq read density to the enhancer regions and their corresponding +/-5kb flanking regions. Each
enhancer region was split into 100 equally sized bins. The flanking +/- 5kb regions were split into 50bp bins. This split all enhancer regions, regardless of their size, into 300 bins. All typical enhancer or super-enhancer regions were then aligned and the average MED1 and BRD4 ChIP-Seq density in each bin was calculated to create a meta genome-wide average in units of rpm/bp. In order to visualize the length disparity between typical and super-enhancer regions, the enhancer region (between its actual start and end) was scaled relative to its average length. See Figure 2B (MM1.S), Figure 7B (GBM) and 7E (SCLC).

**Characterizing super-enhancers**

We calculated the total ChIP-Seq occupancy for MED1, BRD4, and H3K27Ac at super-enhancers and typical enhancers and determined the fold difference between the average signal at super-enhancers versus typical enhancers. ChIP-Seq density for MED1, BRD4, and H3K27Ac were determined by calculating the average density of each factor in constituent regions of every enhancer. See Figure 2B (MM1.S), Figure 7B (GBM) and 7E (SCLC).

**Assigning genes to super-enhancers**

Transcriptionally active genes were assigned to enhancers using the following method: Enhancers tend to loop to and associate with adjacent genes in order to activate their transcription (Ong and Corces, 2011). Most of these interactions occur within a distance of ~50kb of the enhancer (Chepelev et al., 2012). Using a simple proximity rule, we assigned all transcriptionally active genes (TSSs) to super-enhancers within a 50kb window, a method shown.
to identify a large proportion of true enhancer/promoter interactions in embryonic stem cells (Dixon et al., 2012). This identified 681 genes associated with super-enhancers in MM1.S cells (Table S3). The same method was applied to identify super-enhancer-associated genes in GBM and SCLC (Table S5).

Calculating the cell type specificity of enhancer-associated genes

Spike-in normalized gene expression levels for all genes in MM1.S cells were determined as previously described (RNA Extraction and Synthetic RNA Spike-In and Microarray Sample Preparation and Analysis sections). The expression levels of genes associated with typical enhancers or super-enhancers were compared using a Welch’s t-test and found be statistically significant (p-value < 2e-16) (Figure 2D).

To evaluate the cell-type specificity of SE and typical enhancer-associated genes, we compared MM1.S gene expression from the dataset (GSE17385) (Carrasco et al., 2007) to the largest matched microarray expression dataset of normal human tissues (GSE7307) on Affymetrix U133 plus 2.0 arrays from Gene Expression Omnibus (GEO). We used the three positive controls from GSE17385 and the 504 samples of normal human tissues from GSE7307 for the analysis. All .CEL files were MAS5 normalized together and probe level intensities were reported using standard Affymetrix CDF. We used the entropy-based measure Jensen-Shannon (JS) divergence to quantify cell-type specificity (Cabili et al., 2011). For each triplicate of MM1.S expression dataset from GSE7307, the JS divergence quantified the similarity between a probe’s expression pattern across the MM1.S dataset and the 504 samples of normal human tissues and a pre-defined pattern that represents an extreme case in which the probe is expressed only in MM1.S.
The JS divergence statistic was converted to a Z-score to give a relative measure of the cell type specificity of genes in MM1.S compared to the average gene. A similar comparison of cell type specificity Z-scores also showed a statistically significant difference between genes associated with typical or super-enhancers (p-value = 1e-14) (Figure 2D).

Quantifying effects of JQ1 treatment on BRD4 occupancy genome wide

In order to quantify the effects of JQ1 treatment on the genomic occupancy of BRD4, we first calculated average BRD4 occupancy (in units of rpm/bp) at BRD4 enriched regions in DMSO treated cells. We next calculated the average BRD4 occupancy in those same regions in cells treated with 500nM JQ1 for 6 hours. The distributions of BRD4 at enriched regions are plotted (Figure 4B). The difference in the distributions of BRD4 was compared using a Welch’s t-test and found to be statistically significant (p-value < 1e-16) (Figure 4B)

Quantifying effects of JQ1 treatment on BRD4 occupancy at active enhancers

In order to quantify the effects of JQ1 treatment on the genomic occupancy of BRD4 at active enhancers, we first calculated the BRD4 occupancy at typical enhancers or super-enhancers in DMSO and various concentrations of JQ1. The loss of BRD4 at typical enhancers or super-enhancers was quantified as the fraction of BRD4 occupancy remaining compared to DMSO. The average loss of BRD4 at typical enhancers and super-enhancers is plotted for DMSO and 5nM, 50nM, 500nM JQ1 treatment in Figure 5F. 95% confidence intervals were determined by performing 10,000 random samplings (with replacement) of either super-enhancers or typical enhancers. Sample sizes were identical between typical enhancers and super-enhancers. The
differences in the mean loss of BRD4 between typical enhancers and super-enhancers were also deemed significant (p-value < 1e-8) at 5nM, 50nM, and 500nM JQ1 treatment (Figure 5F) using a Welch's $t$-test.

Quantifying effects of JQ1 treatment on gene expression in MM1.S

Spike-in normalized gene expression levels for all genes were determined as previously described (RNA Extraction and Synthetic RNA Spike-In and Microarray Sample Preparation and Analysis sections). In order to quantify changes in gene expression as a function of JQ1 treatment in a time or dose dependent manner, expression levels were first calculated for either super-enhancer or typical enhancer-associated genes (Figure 6B and 6C). For time dependent changes, the $\log_2$ fold change in expression was calculated relative to untreated control cells. For dose dependent changes, the $\log_2$ fold change in expression was calculated relative to DMSO treated cells. The average change in gene expression at super-enhancer or typical enhancer-associated genes was calculated at each timepoint or dose. In order to determine the significance of changes in the mean between super-enhancer and typical enhancer-associated genes, 95% confidence intervals were determined by performing 10,000 random samplings (with replacement) of the mean change for each set of genes. Sample sizes were identical between typical enhancer and super-enhancer-associated genes. We also determined the significance of the difference between distributions of gene expression values for super-enhancer and typical enhancer-associated genes at each time point or dose. Changes were significant over time at 3 and 6 hours after 500nM JQ1 treatment (p-value < 1e-8, Figure 6B). Changes were significant after 6 hour JQ1 treatments at doses of 50nM, 500nM, and 5,000nM (p-value < 1e-8, Figure 6C).
Identical analysis was performed comparing changes in gene expression between super-enhancer-associated genes with multiple typical enhancers (Figure S3C and S3D). Changes were significant over time at 6 hours after 500nM JQI treatment (p-value = 0.03, Figure S3C). Changes were significant after 6 hour JQI treatments at a dose of 500nM (p-value = 0.052, Figure S3D).

Quantifying effects of JQI treatment on CDK9 and MED1 occupancy at promoters and enhancers

We quantified changes in CDK9 and MED1 occupancy by calculating changes in CDK9 and MED1 ChIP-Seq occupancy at promoters, typical enhancers, and super-enhancers. Since super-enhancers often contain TSS overlapping regions, we used H3K4Me3 promoter definitions (Figure S2B) to exclude TSS overlapping super-enhancers. The change in CDK9 and MED1 occupancy was quantified as the fraction of factor occupancy remaining after 500nM JQI treatment for 6 hours compared to DMSO. The average change in CDK9 and MED1 at promoters, typical enhancers, and super-enhancers is plotted in Figure 6F. To show the significance of changes in the mean, 95% confidence intervals were determined by performing 10,000 random samplings (with replacement) of the mean change at either super-enhancers or typical enhancers.

In order to determine the changes in CDK9 and MED1 occupancy at TSS overlapping and TSS distal portions of super-enhancers, we split super-enhancers into TSS proximal and TSS distal regions (Figure S3F). The average change in CDK9 and MED1 at super-enhancer TSS proximal
and TSS distal regions is plotted in Figure S3F. To show the significance of changes in the mean, 95% confidence intervals were determined by performing 10,000 random samplings (with replacement) of the mean change at either super-enhancers or typical enhancers.

Quantifying effects of JQ1 treatment on transcription elongation

We quantified the effects of JQ1 treatment on transcription elongation of genes by measuring the change in RNA Pol II density in the elongating gene body region of genes. The elongating region of genes was defined as the region beginning 300bp downstream of the TSS and extending 3000bp past the annotated 3’ end of the transcript. RNA Pol II density (rpm/bp) was calculated in this region for all transcribed genes in cells treated with DMSO or 500nM JQ1. The Log2 fold change in RNA Pol II density +/- 500nM JQ1 treatment was calculated for all actively transcribed genes (Figure 6G).

To quantify the effects of JQ1 treatment on RNA Pol II elongation at typical or super-enhancer-associated genes, the Log2 fold change in RNA Pol II density +/- 500nM JQ1 treatment was calculated for typical enhancer or super-enhancer-associated genes. The distributions of changes in RNA Pol II occupancy were statistically different as determined by a two sample Kolmogorov-Smirnov (KS) test (Figure S3G, p-value < 2e-16). The average change in RNA Pol II between the two groups of genes was also examined (Figure 6H) and the significance of change was shown using plotted 95% confidence intervals determined by performing 10,000 random samplings (with replacement) of the mean change for either super-enhancer or typical enhancer-associated genes.
References


Appendix B

Enhancer Decommissioning By LSD1 During Embryonic Stem Cell Differentiation

Warren A. Whyte¹,²*, Steve Bilodeau¹*, David A. Orlando¹, Heather A. Hoke¹,², Garrett M. Frampton¹,², Charles T. Foster³,⁴, Shaun M. Cowley⁴, and Richard A. Young¹,²

¹ Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. ² Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ³ Department of Molecular Biology, Adolf-Butenandt Institut, Ludwig-Maximilians-Universität München, Munich, Germany 80336. ⁴ Department of Biochemistry, University of Leicester, Leicester LE1 9HN, United Kingdom.

*These authors contributed equally to this work

This chapter originally appeared in Nature 2012 vol. 482 pp. 221-225. Corresponding Supplementary Material is included. Supplementary tables and data are available online at http://dx.doi.org/ 10.1038/nature10805.
Personal Contribution to the Project

I performed ChIP-seq of HDAC1 and HDAC2 in mouse ESCs, and assisted with knockdown experiments. I also assisted in preparation of the manuscript.
Summary

Transcription factors and chromatin modifiers play important roles in programming and reprogramming of cellular states during development (Graf and Enver, 2009; Young, 2011). Transcription factors bind to enhancer elements and recruit coactivators and chromatin modifying enzymes to facilitate transcription initiation (Fuda et al., 2009; Li et al., 2007). During differentiation, a subset of these enhancers must be silenced, but the mechanisms underlying enhancer silencing are poorly understood. Here we show that the H3K4/K9 histone demethylase LSD1 (Shi et al., 2004) plays an essential role in decommissioning enhancers during differentiation of embryonic stem cells (ESCs). LSD1 occupies enhancers of active genes critical for control of ESC state. However, LSD1 is not essential for maintenance of ESC identity. Instead, ESCs lacking LSD1 activity fail to fully differentiate and ESC-specific enhancers fail to undergo the histone demethylation events associated with differentiation. At active enhancers, LSD1 is a component of the NuRD complex, which contains additional subunits that are necessary for ESC differentiation. We propose that the LSD1-NuRD complex decommissions enhancers of the pluripotency program upon differentiation, which is essential for complete shutdown of the ESC gene expression program and the transition to new cell states.
The histone H3K4/K9 demethylase LSD1 (Lysine-specific-demethylase-1, KDM1A) is among the chromatin regulators that have been implicated in control of early embryogenesis (Wang et al., 2007; Wang et al., 2009a; Foster et al., 2010). Loss of LSD1 leads to embryonic lethality and ESCs lacking LSD1 function fail to differentiate into embryoid bodies (Wang et al., 2007; Wang et al., 2009a; Foster et al., 2010). These results suggest that LSD1 contributes to changes in chromatin that are critical to differentiation of ESCs, but LSD1’s role in this process is not yet understood. To investigate the function of LSD1 in ESCs, we first identified the sites it occupies in the genome using chromatin immunoprecipitation coupled with massively parallel DNA sequencing (ChIP-Seq; Fig. 1, and Supplementary Fig. 1). The results revealed that LSD1 occupies the enhancers and core promoters of a substantial population of actively transcribed and bivalent genes (Fig. 1a, b, and Supplementary Table 1). Inspection of individual gene tracks showed that LSD1 co-occupies well-characterized enhancer regions with the ESC master transcription factors Oct4, Sox2 and Nanog and the Mediator coactivator (Fig. 1b, and Supplementary Fig. 1). Loci bound by Oct4, Sox2 and Nanog are generally associated with Mediator and p300 coactivators and have enhancer activity (Kagey et al., 2010; Chen et al., 2008). A global view of Oct4, Sox2, Nanog and Mediator -occupied enhancer regions confirmed that 97% of the 3,838 high confidence enhancers were co-occupied by LSD1 (p < 10^{-9}) (Fig. 1c, and Supplementary Table 2). This is consistent with evidence that LSD1 can interact with Oct4 (Pardo et al., 2010; van den Berg et al., 2010). LSD1 signals were also observed at core promoter regions with RNA Polymerase II (Pol II) and TBP (Fig. 1d). The density of LSD1 signals at enhancers was higher than at core promoters (p < 10^{-16}; Supplementary Fig. 1), indicating that LSD1 is associated predominantly with the enhancers of actively transcribed genes in ESCs.

It was striking to find that LSD1 is associated with active genes in ESCs because
Previous studies have shown that LSD1 is not essential for maintenance of ESC state, but is required for normal differentiation (Wang et al., 2007; Wang et al., 2009a; Foster et al., 2010). We used an ESC differentiation assay to further investigate the involvement of LSD1 in cell state transitions (Fig. 2a, b). Prolonged depletion of Oct4 in ZHBTc4 ESCs using doxycycline causes loss of pluripotency and differentiation into trophectoderm (Niwa et al., 2000). As expected, loss of Oct4 expression led to a rapid loss of ESC morphology and a marked reduction in SSEA-1 and alkaline phosphatase, two markers of ESCs (Fig. 2c, and Supplementary Fig. 2). When these ESCs were treated with the LSD1 inhibitor tranylcypromine (TCP) during Oct4 depletion, they failed to undergo the morphological changes associated with differentiation of ESCs (Fig. 2c). Instead, the TCP-treated cells formed small colonies resembling those of untreated ESCs and maintained expression of SSEA-1 and alkaline phosphatase (Fig. 2c, and Supplementary Fig. 3). Very similar results were obtained in LSD1 knockout ESCs (Supplementary Fig. 4, 5), and in cells treated with another LSD1 inhibitor, pargyline (Prg), or an shRNA against LSD1 (Supplementary Fig. 2, 3). LSD1 inhibition also caused an increase in cell death during differentiation, as has been observed with cells lacking LSD1 in other assays (Wang et al., 2009a; Foster et al., 2010). These results suggest that LSD1 may be required for ESCs to completely silence the ESC gene expression program.

Further analysis of ESCs forced to differentiate in the absence of LSD1 activity confirmed that these cells failed to fully transition from the ESC gene expression program; while key genes of the trophectoderm gene expression program were activated, including Cdx2 and Esx1 (Rossant and Cross, 2001), there was incomplete repression of many ESC genes, including Sox2 and Fbox15 (Fig. 2d). A global analysis confirmed that a set of genes neighboring LSD1-occupied enhancers in ESCs are repressed upon differentiation, and that repression of this set of
genes is partially relieved in the presence of TCP (Fig. 2e, and Supplementary Table 3). Similar results were obtained with LSD1 knockout cells (Supplementary Fig. 4, 5), and cells treated with either pargyline or an shRNA against LSD1 (Supplementary Fig. 3). These results indicate that the trophectoderm differentiation program can be induced in cells lacking LSD1 function, but the ESC program is not fully silenced in these cells.

To gain further insight into the role of LSD1 in ESC differentiation, we investigated whether LSD1 is associated with previously described complexes, including NuRD (Nucleosome remodeling and histone deacetylase), CoREST (Cofactor of REST) and the AR/ER (Androgen receptor/ Estrogen receptor) complexes (Foster et al., 2010; Metzger et al., 2005; Shi et al., 2005; Wang et al., 2009b). We first studied whether the LSD1 found at Oct4-occupied genes is a component of NuRD because Oct4 and Nanog have been reported to interact with several components of NuRD (Pardo et al., 2010; van den Berg et al., 2010; Liang et al., 2008). ChIP-Seq experiments confirmed that NuRD subunits Mi-2β, HDAC1 and HDAC2 co-occupy sites with LSD1 at enhancers (p < 10^-9; Fig. 3 and Supplementary Table 1). Immunoprecipitation of LSD1 confirmed its association with Mi-2β, HDAC1 and HDAC2 (Fig. 3b, c). We then investigated whether LSD1 is associated with CoREST; ChIP-Seq data revealed that a minor fraction of LSD1 co-occupies sites with CoREST and Rest (2% and 6%, respectively)(Supplementary Fig. 6 and Supplementary Table 1). As expected, LSD1-REST sites were frequently found associated with neuronal genes (Supplementary Fig. 7 and Supplementary Table 4). Immunoprecipitation experiments confirmed that LSD1 is associated with CoREST (Fig. 3b, c). AR and ER are not expressed in ESCs based on the lack of histone H3K79me2 and H3K36me3 (modifications associated with transcriptional elongation) at the genes encoding these proteins (Supplementary Table 1). Further examination of the ChIP-Seq
data revealed that enhancers were significantly more likely to be occupied by the LSD1 and NuRD proteins as compared to REST and CoREST (p < 10^-9) (Fig. 3d and Supplementary Fig. 8). Multiple components of NuRD are dispensable for ESC state but required for normal differentiation (Wang et al., 2007; Dovey et al., 2010; Kaji et al., 2006; Scimone et al., 2010). ESCs with reduced levels of the core NuRD ATPase Mi-2β failed to differentiate properly and partially maintained expression of SSEA-1, alkaline phosphatase and ESC genes (Supplementary Fig. 9), which are the same phenotypes we observed with reduced levels of LSD1. These results indicate that LSD1 at enhancers is associated with a NuRD complex that is essential for normal cell state transitions.

Nucleosomes with histone H3K4me1 are commonly found at enhancers of active genes, and are a substrate for LSD1 (Heintzman et al., 2007; Shi et al., 2004). If LSD1-dependent H3K4me1 demethylase activity is involved in enhancer silencing during ESC differentiation, LSD1 inhibition should cause retention of H3K4me1 levels at active ESC enhancers when differentiation is induced. During trophectoderm differentiation with control ESCs, we found p300 and H3K27ac levels reduced at a set of active ESC enhancers, suggesting these enhancers were being silenced (Supplementary Fig. 10). The levels of H3K4me1 at enhancers were also reduced, as seen for example at Lefty1 (Fig. 4a, and Supplementary Table 5), while the levels of H3K4me1 increased at newly active trophectoderm genes such as Gata2 (Fig. 4b). In contrast, H3K4me1 signals were higher at LSD1-occupied enhancers in differentiating ESCs treated with TCP than in control cells, including Lefty1 and Sox2 (Fig. 4a, c). The majority of enhancers (1,722 of 2,755) that were occupied by LSD1 and that experienced reduced levels of H3K4me1 during differentiation retained H3K4me1 in TCP-treated ESCs compared to untreated control differentiating ESCs (Fig. 4d, e). These results are consistent with the model that LSD1
demethylates H3K4me1 at the enhancers of ESC-specific genes during differentiation, and that this activity is essential to fully repress the genes associated with these enhancers.

Our results indicate that an LSD1-NuRD complex is required for silencing of ESC enhancers during differentiation, which is essential for complete shutdown of the ESC gene expression program and the transition to new cell states. These results, together with those of previous studies on NuRD function (Liang et al., 2008; Scimone et al., 2010; Forneris et al., 2005; Lee et al., 2006), suggest the following model for LSD1-NuRD in enhancer decommissioning. LSD1-NuRD complexes occupy Oct4-regulated active enhancers in ESCs, but do not substantially demethylate histone H3K4 because LSD1's H3K4 demethylase activity is inhibited in the presence of acetylated histones (Forneris et al., 2005; Lee et al., 2006). Enhancers occupied by Oct4, Sox2 and Nanog are co-occupied by the HAT p300 and nucleosomes with acetylated histones (Supplementary Fig. 10 and Chen et al., 2008). Thus, as long as the enhancer-bound transcription factors recruit HATs to enhancers, the net effect of having both HATs and NuRD-associated HDACs present is to have sufficient levels of acetylated histones to suppress LSD1 demethylase activity. During ESC differentiation, the levels of Oct4 and p300 are reduced, thus reducing the level of acetylated histones, which in turn permits demethylation of H3K4 by LSD1. Consistent with this model, we find that the shutdown of Oct4 leads to reduced levels of p300 and histone H3K27ac at enhancers that are occupied by Oct4 and LSD1 (Supplementary Fig. 10, 11), and this is coincident with reduced levels of methylated H3K4 (Fig. 4, and Supplementary Figs. 12, 13). This model would explain why key components of LSD1-NuRD complexes are not essential for maintenance of ESC state, but are essential for normal differentiation, when the active enhancers must be silenced. Additional HATs expressed in ESCs may also contribute to the dynamic balance of nucleosome acetylation.
Future biochemical analysis of HAT, HDAC and demethylase complexes at enhancers will be valuable for testing this model and for further understanding how enhancers are regulated during differentiation.

We conclude that LSD1-NuRD complexes present at active promoters in ESCs are essential for normal differentiation, when the active enhancers must be silenced. Given evidence that LSD1 is required for differentiation of multiple cell types (Wang et al., 2007; Musri et al., 2010; Choi et al., 2010), LSD1 is likely to be generally involved in enhancer silencing during differentiation. The ESC gene expression program can be maintained in the absence of many other chromatin regulators (Young, 2011), and it is possible that some of these also play key roles in the transition from one transcriptional program to another during differentiation.
Methods Summary

ESC Cell Culture Conditions

ESCs were grown on irradiated murine embryonic fibroblasts (MEFs) and passaged as previously described (Kagey et al., 2010). In drug treatment experiments, ESCs were split off MEFs and treated with tranylcypromine (TCP, 1mM) or pargyline (Prg, 3mM) to inhibit LSD1 activity. Lentiviral constructs were purchased from Open Biosystems and produced according to the Trans-lentiviral shRNA Packaging System (TLP4614).

Differentiation assay, Immunofluorescence, and Alkaline Phosphatase Staining

ZHBTc4 ESCs were split off MEFs in ESC media containing 2μg/ml doxycycline to reduce Oct4 expression levels. For Immunofluorescence, ESCs were crosslinked, blocked and permeabilized before incubation with Oct4 (Santa Cruz, sc-9081x; 1:200 dilution) or SSEA1 (mc-480, DHSB, 1:20 dilution) antibodies. Alexa-conjugated secondary antibodies were used for detection. Staining of ESCs for alkaline phosphatase was achieved using the Alkaline Phosphatase Detection Kit (Millipore, SCR004). Cells were harvested at indicated time points for ChIP-Seq, qPCR or expression array analyses.

ChIP-Seq

Chromatin immunoprecipitations (ChIPs) were performed and analyzed as previously described (Kagey et al., 2010). The following antibodies were used: LSD1 (Abcam, ab17721), Mi-2b
(Abcam, ab72418), HDAC1 (Abcam, ab7028), HDAC2 (Abcam, ab7029), REST (Millipore, 07-579), CoREST (Abcam, ab32631), H3K4me1 (Abcam, ab8895), p300 (Santa-Cruz, sc-584) and H3K27Ac (Abcam, ab4729).

For ChIP-Seq analyses, reads were aligned with Bowtie and analyzed as described in Supplemental Information.

Full Methods and any associated references are available in Supplementary Information of the paper at www.nature.com/nature.

Accession Numbers

ChIP-Seq and GeneChip expression data have been deposited in Gene Expression Omnibus with accession number GSE27844.

Acknowledgements

We thank Jakob Lovén, Michael H. Kagey, Jill Dowen, Alan C. Mullen, Alla Sigova, Peter B. Rahl, Tony Lee, and members of Yang Shi’s lab for experimental assistance, reagents and helpful discussions; and Jeong-Ah Kwon, Vidya Dhanapal, Jennifer Love, Sumeet Gupta, Thomas Volkert, Wendy Salmon and Nicki Watson for assistance with ChIP-Seq, expression arrays, and immunofluorescence imaging acquisition. This work was supported by a Canadian Institutes of Health Research Fellowship (SB), a Career Development Award from the Medical
Research Council (SMC), and by NIH grants HG002668 and NS055923 (RY).

**Author Contributions**

The ChIP-Seq, immunofluorescence and expression experiments were designed, conducted and interpreted by W.A.W., S.B., H.A.H., C.T.F., S.M.C and R.A.Y. W.A.W., D.A.O. and G.M.F. performed data analysis. The manuscript was written by S.B., W.A.W. D.A.O., H.A.H., G.M.F. and R.A.Y.
References


FIGURE LEGENDS

Figure 1: LSD1 is associated with enhancer and core promoter regions of active genes in ESCs

a, LSD1 occupies a substantial population of actively transcribed genes in murine ESCs. Pie charts depict active (green), bivalent (yellow) and silent (red) genes, and the proportion (black lines) occupied by either LSD1, Pol II, or the Polycomb protein Suz12 (Supplementary Table 1, and Supplementary Information). The numbers depict the number of genes bound over the total number of genes in each of the active, bivalent, and silent classes. LSD1 ChIP-Seq data is from combined biological replicates using an antibody specific for LSD1 as determined by knockdown experiments (Supplementary Fig. 1). The p-value for each category was determined by a hypergeometric test.

b, LSD1 occupies enhancers and core promoter regions of actively transcribed genes. ChIP-Seq binding profiles (reads/million) for ESC transcription factors (Oct4, Sox2, Nanog), coactivator (Med1), chromatin regulator (LSD1), the transcriptional apparatus (Pol II, TBP) and histone modifications (H3K4me1, H3K4me3, H3K79me2, H3K36me3) at the Oct4 (Pou5f1) and Lefty1 loci in ESCs, with the y-axis floor set to 1. Gene models, and previously described enhancer regions (Okumura-Nakanishi et al., 2005; Yeom et al., 1996; Nakatake et al., 2006) are depicted below the binding profiles.

c, LSD1 occupies enhancer sites. Density map of ChIP-Seq data at Oct4, Sox2, Nanog, and Med1 co-occupied enhancer regions. Data is shown for ESC regulators (Oct4), coactivators (Med1 and p300) and a chromatin regulator (LSD1) in ESCs. Enhancers were defined as Oct4,
Sox2, Nanog and Mediator co-occupied regions. Over 96% of the 3,838 high confidence enhancers were co-occupied by LSD1 (p < 10⁻⁹). Color scale indicates ChIP-seq signal in reads per million.

d, LSD1 occupies core promoter sites. Density map of ChIP-Seq data at transcriptional start sites (TSS) of genes neighboring the 3,838 previously defined enhancers (Fig. 1c). Data is shown for components of the transcription apparatus (Pol II and TBP) and the chromatin regulator LSD1 in ESCs. Core promoters were defined as the closest TSS from each enhancer. Color scale indicates ChIP-Seq signal in reads per million.

Figure 2: LSD1 inhibition results in incomplete silencing of ESC genes during differentiation

a, Schematic representation of trophectoderm differentiation assay using doxycycline-inducible Oct4 shutdown murine ESC line ZHBTc4. Treatment with doxycycline for 48 hours leads to depletion of Oct4 and early trophectoderm specification. Cells were treated with DMSO (control) or the LSD1 inhibitor tranylcypromine (TCP) for 6 hours before 2μg/ml doxycycline was added for an additional 24 or 48 hours.

b, Treatment of ZHBTc4 ESCs with doxycycline leads to loss of Oct4 proteins. Oct4 and LSD1 protein levels in nuclear extracts (NE) determined by Western blot (WB) before and after treatment of ZHBTc4 ESCs with 2μg/ml doxycycline. Tubulin served as loading control.

c, Doxycycline-treated cells treated with TCP maintained SSEA-1 cell surface marker expression. Cells were stained for Hoechst (Hoe), Oct4 and SSEA-1. Scale bar = 100μM
d, Expression of selected ESC and trophectodermal genes 48 hours after Oct4 depletion in DMSO- versus TCP-treated cells (black versus grey bars, respectively). Treatment of TCP partially relieved repression of ESC genes, but did not affect upregulation of trophectodermal genes. Error bars reflect standard deviation from biological replicates.

e, Genes neighboring LSD1-occupied enhancers are less downregulated during ESC differentiation following TCP treatment. Mean fold change in expression of the 630 downregulated (at least 1.25 fold; \( p < 0.01 \)) genes nearest LSD1- occupied enhancers (Fig. 1c) during differentiation of TCP-treated and untreated control cells. Alleviation of repression is significantly higher \( (p < 0.005) \) for LSD1 enhancer-bound repressed genes compared to all repressed genes.

**Figure 3: LSD1 is associated with a NuRD complex at active enhancers in ESCs**

a, NuRD components occupy enhancers and core promoter regions of actively transcribed genes. ChIP-Seq binding profiles (reads/million) for transcription factors (Oct4, Sox2, Nanog), coactivator (Med1), and chromatin regulators (LSD1, Mi-2\( \beta \), HDAC1, HDAC2), at the Oct4 (Pou5f1) and Lefty1 loci in ESCs, with the y-axis floor set to 1. Gene models, and previously described enhancer regions (Okumura-Nakanishi et al., 2005; Yeom et al., 1996; Nakatake et al., 2006) are depicted below the binding profiles.

b, LSD1 is associated with NuRD components Mi-2\( \beta \), HDAC1, HDAC2, as well as CoREST. LSD1 and HDAC1 are detected by Western blot (WB) after immunoprecipitation of crosslinked chromatin using LSD1, HDAC1, HDAC2, Mi-2\( \beta \), or CoREST antibodies. IgG is shown as a control.
c, LSD1 and HDAC1 are detected by Western blot (WB) after immunoprecipitation of uncrosslinked nuclear extracts (NE) using LSD1, HDAC1, HDAC2, Mi-2β, or CoREST antibodies. IgG is shown as a control.

d, The occupancy of enhancers by NuRD proteins (Mi-2β, HDAC1 and HDAC2) is significantly greater than the occupancy by CoREST or REST ($p < 10^{-9}$). The height of the bars represents the percentage of the 3,838 enhancers co-occupied by LSD1, NuRD proteins (Mi-2β, and either HDAC1 or HDAC2), CoREST and REST.

**Figure 4: LSD1 is required for H3K4me1 removal at ESC enhancers**

a, H3K4me1 levels are reduced at LSD1-occupied enhancers upon ESC differentiation, and this effect is partially blocked upon TCP treatment.

b, TCP treatment does not affect the increase in H3K4me1 levels at trophectodermal genes during differentiation. ChIP-Seq binding profiles (reads/million) for Oct4 and LSD1 at the Lefty1 and Gata2 loci in ESCs. Below these profiles, histone H3K4me1 levels are shown for ZHTBtc4 control ESCs, cells treated with doxycycline for 48 hours to repress Oct4 and induce differentiation (ESCs +Dox), and ESCs treated with doxycycline and TCP (ESCs +Dox,TCP). For appropriate normalization, ChIP-Seq data for histone H3K4me1 is shown as rank normalized reads/million with the y-axis floor set to 1 (Supplementary Information). Gene models, and previously described enhancer regions (Nakatake et al., 2006; Ray et al., 2009) are depicted below the binding profiles.

c, Sum of the normalized H3K4me1 density +/- 250 nucleotides surrounding LSD1-occupied
enhancer regions before and during trophectoderm differentiation in the presence or absence of TCP. The associated genes were identified based on their proximity to the LSD1-occupied enhancers.

d, Sum of the normalized H3K4me1 density +/- 250 nucleotides surrounding 1,722 LSD1-occupied enhancers before and during differentiation in the presence or absence of TCP. Of the 2,755 LSD1-occupied enhancers having reduced levels of H3K4me1 upon differentiation, 63% (1,722) display higher H3K4me1 levels after TCP treatment (p < 10^{-16}).

e, Heatmap displaying the sum of the normalized H3K4me1 density +/- 250 nucleotides surrounding the 1,722 LSD1-occupied enhancers that retained H3K4me1 in TCP-treated ESCs compared to untreated control differentiating ESCs. Color scale indicates ChIP-Seq signal in normalized reads per million.
LSD1
Active genes 6,844/7,339
P < 10^{-300}
Bivalent genes 2,003/3,094
P = 0.173
Silent genes 1,583/6,976
P = 1

Pol II
Active genes 6,899/7,339
P < 10^{-300}
Bivalent genes 681/3,094
P = 1
Silent genes 349/6,976
P = 1

Suz12
Active genes 147/7,339
P = 1
Bivalent genes 2,413/3,094
P < 10^{-300}
Silent genes 140/6,976
P = 1

Figure 1
Figure 2
Figure 3
Figure 4
SUPPLEMENTARY INFORMATION

Enhancer Decommissioning by LSD1 During Embryonic Stem Cell Differentiation


CONTENTS

Supplementary Tables

Supplementary Data Files

Supplementary Discussion

Supplementary Experimental Procedures

Cell Culture Conditions and Differentiation Assays
Embryonic Stem Cells (ESCs)

Generation of LSD1 Knockout ESC lines (Supplementary Fig. 4, 5)

Generation and Expression of LSD1 Expression Constructs in ESCs (Supplementary Fig. 5)

Differentiation of ESCs (Fig. 2, 4, and Supplementary Fig. 2, 3, 9, and 10)

Inhibition of p300 using HAT inhibitor C-646 (Supplementary Fig. 12)

Lentiviral Production and Infection (Supplementary Fig. 2, 3, 4, 5, and 9)

Immunofluorescence (Fig. 2b and Supplementary Fig. 3, 4, 5, and 9)

Alkaline Phosphatase Staining (Supplementary Fig. 2, 4, 5, and 9)

RNA Extraction, cDNA, and TaqMan Expression Analysis (Supplementary Fig. 1, 3, 5, and 9)

Chromatin Immunoprecipitation (ChIP)

Antibody Specificity (Supplementary Fig. 1, 6, and 9)

ChIP Protocol

Gene Specific ChIPs (Supplementary Fig. 9, 10, 11, 12)

ChIP-Seq Sample Preparation and Analysis

Sample Preparation

Polony Generation and Sequencing
ChIP-Seq Data Analysis

Assigning ChIP-Seq Enriched Regions to Genes (Supplementary Table 1)

Enrichment of LSD1, Pol II, and Suz12 at Genes (Fig. 1)

Definition of Enhancer and Core Promoter (Fig. 1, 4, Supplementary Fig. 8, 13, and Supplementary Table 1)

Determining LSD1 density at Enhancer and Core Promoter (Supplementary Fig. 1)

ChIP-Seq Density Heatmaps (Fig. 1, 4, Supplementary Fig. 8, and 13)

Calculation of the Statistical Significance of the Overlap Between Sets of Genomic Regions (Fig. 1, 3, 4, Supplementary Fig. 8, and 12)

Rank Normalization of H3K4me1 ChIP-Seq Data (Fig. 4, Supplementary Fig. 13, and Supplementary Table 5)

Gene Ontology (GO) Analysis (Supplementary Fig. 7)

Public availability of ChIP-Seq Datasets

Microarray Analysis (Fig. 2)

Cell Culture and RNA isolation
Microarray hybridization and Analysis

Public availability of Microarray Gene Expression Datasets

Protein Extraction and Western Blot Analysis (Fig. 2, Supplementary Fig. 1, 4, 5, 6, and 9)

ChIP-Western and Co-Immunoprecipitation (Fig. 3)

Supplementary References
Supplementary Tables

Supplementary Table 1 – Summary of occupied genes and regions

Supplementary Table 2 – Summary of LSD1 occupancy at active and poised enhancers

Supplementary Table 3 – Gene expression changes 48 hours after Oct4 repression in ZHBTc4 DMSO (control)- and TCP-treated cells

Supplementary Table 4 – Genes co-occupied by LSD1 and REST

Supplementary Table 5 – Enhancer decommissioning 48 hours after Oct4 depletion in ZHBTc4 DMSO (control)- and TCP-treated cells

Supplementary Data File 1 contains ChIP-Seq data in compressed WIG format (WIG.GZ) for upload into the UCSC genome browser (Kent et al., 2002). This file contains data for mES_H3K4me1, DMSO_H3K4me1, DMSO_48HR_H3K4me1, TCP_48HR_H3K4me1,
mES_H3K4me3, mES_H3K79me2, mES_H3K36me3, mES_Oct4, mES_Sox2, mES_Nanog,
mES_Pol II, mES_TBP, mES_LSD1, mES_p300, mES_REST, mES_CoREST, mES_Suz12,
mES_H3K27ac, mES_HDAC1, mES_HDAC2, mES_Mi-2b, mES_DMSO_H3K27ac,
mES_DMSO_H3K4me1, mES_C646_H3K27ac, mES_C646_H3K4me1, WCE_DMSO,
WCE_48HR_DMSO, WCE_48HR_TCP, and WCE_mES.

The first track for each data set contains the ChIP-Seq density across the genome in 25bp bins. The minimum ChIP-Seq density shown in these files is 1.0 reads per million total reads. For DMSO_H3K4me1, DMSO_48HR_H3K4me1, and TCP_48HR_H3K4me1 datasets, the minimum ChIP-Seq density is 1 normalized read per million total reads. Subsequent tracks identify genomic regions identified as enriched at a p-value threshold of $10^{-9}$.

**Supplementary Discussion**

To address LSD1 function in ESCs and during ESC differentiation, cells were treated with the monoamine oxidase inhibitors (MAOIs) tranylcypromine (TCP) or pargyline (Prg). Mechanistically, LSD1 is unique relative to other demethylases, as it demethylates lysine residues via a flavin–adenine dinucleotide–dependent reaction (Forneris et al., 2005; Shi et al., 2004). This reaction is inhibited by MAOIs, which are used in the treatment of certain psychiatric and neurological disorders (Lee et al., 2006; Schmidt and McCafferty, 2007; Yang et al., 2007; Mimasu et al., 2010; Liang et al., 2009). Furthermore, TCP inhibits LSD1 at levels comparable to MAO inhibition of clinical mitochondrial MAOI targets (Yang et al., 2007; Mimasu et al., 2010). Finally, the effects of inhibition by TCP and Prg were highly similar in our assays. Therefore, we found TCP and Prg suitable to study LSD1 activity during differentiation of ES cells.
Previous studies demonstrated that LSD1 is required for mouse development beyond e6.5 (Foster et al., 2010). The differentiation defect is recapitulated in vitro in embryoid body (EB) formation assays, where ESCs are plated in suspension in media lacking leukemia inhibitory factor (LIF), causing ESC differentiation. In these assays, deletion of LSD1 leads to considerable cell death and formation of small EBs, with the remaining cells expressing markers of all three germ layers (Foster et al., 2010; Wang et al., 2009a). The differentiation defect is also recapitulated in vitro in the Oct4 depletion assay used in the present study, where ESCs preferentially differentiate into the trophectodermal lineage. In this assay, inhibition of LSD1 with either TCP or Prg leads to considerable cell death, and surviving cells express trophectodermal differentiation markers. Thus, very similar effects are observed in these two assays.

In the Oct4 depletion assay used in the present study, full repression of key ESC regulators such as Nanog is not achieved in the presence of LSD1 inhibitors 48 hours after Oct4 depletion is initiated. Similar results were obtained for Nanog in the assay used by Foster et al. (2010) with LSD1 mutant ESCs 48 hours after LIF removal (Foster et al., 2010). In this latter assay, Nanog levels were ultimately fully reduced in ESCs lacking LSD1. Our interpretation of all these data is that inhibition of LSD1 delays differentiation because there is a delay in reducing levels of key ESC regulators whose enhancers are occupied by LSD1.

Previous studies also report that deletion of LSD1 protein destabilizes Co-REST, which leads to a less active LSD1-CoREST-HDAC complex (Foster et al., 2010). Accordingly, developmental regulators targeted by the LSD1-CoREST-HDAC complex are de-repressed, and this may
explain the defect observed in ESC differentiation. In experiments using small molecule inhibitors of LSD1 activity, we did not observe significant changes in genes associated with LSD1-REST co-occupied sites. One explanation for this discrepancy may be that loss of LSD1 proteins destabilizes the CoREST-HDAC complex, while small molecules inhibiting LSD1 enzymatic activity may not alter the stability of the complex, thereby giving a different transcriptional phenotype. Thus, the different methods (gene deletion versus enzymatic inhibition) of LSD1 inhibition may explain the discrepancy in gene expression.

Differences were observed for LSD1 function in mouse and human ESCs. Although mouse ESCs fail to differentiate into embryoid bodies in absence of LSD1 (Foster et al., 2010; Wang et al., 2009a; Wang et al., 2007), human ESCs differentiate when LSD1 levels are reduced (Adamo et al., 2011). In both mouse and human ESCs, LSD1 is found at Oct4 and Nanog-occupied regions. Similar to our observations in the Oct4 depletion assay, a subpopulation of differentiating human ESCs with reduced LSD1 levels retained the expression of pluripotency markers reduced (Adamo et al., 2011).

Supplementary Experimental Procedures

Cell Culture Conditions and Differentiation Assays

Embryonic Stem Cells
V6.5, ZHBTc4, and E14 murine ESCs were grown on irradiated murine embryonic fibroblasts (MEFs), unless otherwise stated. Cells were grown under standard ESC conditions as described previously (Boyer et al., 2005). Briefly, cells were grown on 0.2% gelatinized (Sigma, G1890) tissue culture plates in ESC media; DMEM-KO (Invitrogen, 10829-018) supplemented with 15% fetal bovine serum (Hyclone, characterized SH3007103), 1000 U/mL LIF (ESGRO, ESG1106), 100 μM nonessential amino acids (Invitrogen, 11140-050), 2 mM L-glutamine (Invitrogen, 25030-081), 100 U/mL penicillin, 100 μg/mL streptomycin (Invitrogen, 15140-122), and 8 nL/mL of 2-mercaptoethanol (Sigma, M7522).

Generation of LSD1 knockout ESC lines (Supplementary Fig. 4, 5)

An E14 ESC line expressing a Cre-estrogen receptor fusion protein from the ROSA26 locus was used to generate cells in which LSD1 can be inactivated conditionally. Briefly, LSD1 Lox/A3 ESCs were generated as previously described (Adamo et al., 2011), in which one LSD1 allele has exon 3 flanked by LoxP sites (floxed) and the second has exon 3 deleted. Induction of Cre activity by addition of 4-hydroxytamoxifen (4-OHT, Sigma, H7904) to the growth medium resulted in complete recombination of the remaining floxed allele (to generate LSD1 Δ3/Δ3) within 6 hours. Cells were grown in the absence of irradiated MEFs under standard ESC conditions as described previously (Boyer et al., 2005).

Generation and Expression of LSD1 Expression Constructs in ESCs (Supplementary Fig. 5)
Two forms of human LSD1 were generated by PCR with primers containing tails of family D vector homology for cloning into pCAGGs expression vectors. An enhanced green fluorescent protein (EGFP) tag was added to the N-terminus of either the full-length LSD1 protein, or the catalytically inactive form of LSD1 with a lysine to alanine mutation introduced at residue 661 by site-directed mutagenesis (Binda et al., 1999).

For expression of LSD1 constructs in ESCs, LSD1 Lox/Δ3 and Δ3/Δ3 E14 ESCs were plated in 6-well culture plates. The following day, cells were transfected with 2μg pCAGGs construct and 2μg of the pMONO-hygro plasmid containing the hygromycin resistance gene (Invivogen, pmonoh-mcs) using 10μl Lipofectamine 2000 (Invitrogen, 11668-019). After 24 hours, the media was removed and replaced with ESC media containing 400μg/ml hygromycin. 24 hours after hygromycin addition, hygromycin concentration was reduced to 200μg/ml for the duration of the experiment.

*Differentiation of ESCs (Fig. 2, 4, and Supplementary Fig. 2, 3, 9, and 10)*

For immunofluorescence, monitoring H3K4me1 levels and ESC expression analysis during differentiation following either DMSO, TCP or Prg treatment, ZHBTc4 ESCs were split off MEFs, placed in a tissue culture dish for 45 minutes to selectively remove the MEFs, and plated on either 6-well or 15cm cell culture plates. The following day, cells were treated with either DMSO, 1mM TCP or 3mM Prg for 6 hours in ESC media. After 6 hours, the media was removed and replaced with ESC media containing 2μg/ml doxycyline and either DMSO, TCP or
Prg. 24 hours after doxycycline addition, the media was replaced with ESC media containing doxycycline and either DMSO, TCP or Prg for another 24 hours (48 hours total).

*Inhibition of p300 using HAT inhibitor C-646 (Supplementary Fig. 12)*

During differentiation, it is presently unclear which acetylated histone residue(s) need to be deacetylated to trigger activation of LSD1 demethylation activity. While p300 and H3K27Ac are very good candidates at enhancers, many other HATs (CBP, GCN5, Tip60, Elp3, Myst3, Myst4) are active in ESCs (Lin et al., 2007; Ura et al., 2011; Miyabayashi et al., 2007). Therefore, regulation of acetylation levels at enhancers is most likely to be the result of a combination of HATs and HDACs complexes.

To test if acetylation of H3K27 prevents demethylation of H3K4 by LSD1, we treated ESCs with the p300-specific acteyltransferase inhibitor C-646 (Bowers et al., 2010). V6.5 ESCs were split off MEFs, placed in a tissue culture dish for 45 minutes to selectively remove the MEFs, and plated in 15cm cell culture plates. The following days, cells were treated with either DMSO (control) or C-646 (Tocris, 4200) for 24 hours. Cells were then crosslinked and collected to generate ChIP-seq datasets for H3K27ac and H3K4me1.

*Lentiviral Production and Infection (Supplementary Fig. 2, 3, 4, 5, and 9)*

Lentivirus was produced according to Open Biosystems *Trans*-lentiviral shRNA Packaging System (TLP4614). The shRNA constructs targeting LSD1, Oct4, CoREST and Mi-2b are listed
below. The shRNAs targeting either GFP (RHS4459) or Luciferase (SHC007) were used as controls.

<table>
<thead>
<tr>
<th>Gene</th>
<th>TRCN Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFP</td>
<td>RHS4459</td>
</tr>
<tr>
<td>Luciferase</td>
<td>SHC007</td>
</tr>
<tr>
<td>LSD1</td>
<td>TRCN0000071373</td>
</tr>
<tr>
<td>Oct4</td>
<td>TRCN000009611</td>
</tr>
<tr>
<td>Mi-2b #1</td>
<td>TRCN000086143</td>
</tr>
<tr>
<td>Mi-2b #2</td>
<td>TRCN000086145</td>
</tr>
<tr>
<td>CoREST</td>
<td>TRCN000071371</td>
</tr>
</tbody>
</table>

For GFP, LSD1 and Mi-2b, ZHBTc4 ESCs were split off MEFs, placed in a tissue culture dish for 45 minutes to selectively remove the MEFs, and plated on either 6-well or 12-well cell culture plates. The following day, cells were infected in ESC media containing 8 µg/ml polybrene (Sigma, H9268-10G). After 24 hours the media was removed and replaced with ESC media containing 3.5 µg/mL puromycin (Sigma, P8833). ESC media with puromycin was changed daily. Three days post infection, the media was removed and replaced with ESC media containing 2µg/ml doxycycline to shutdown Oct4 and induce differentiation. Cells were harvested 48 hours later.

For Luciferase and Oct4, LSD1 Lox/Δ3 or Δ3/Δ3 E14 ESCs were plated on 6-well culture plates. The following day, cells were infected in ESC media containing 8 µg/ml polybrene. Cells
transfected with LSD1 expression constructs were infected in ESC media lacking antibiotics and polybrene. After 16 hours the media was removed and replaced with ESC media containing 3.5 μg/mL puromycin (Sigma, P8833). The media of ESCs transfected with the LSD1 expression constructs also contained hygromycin, as described earlier. ESC media with either puromycin, or puromycin and hygromycin, was changed daily. Cells were harvested 72 hours later.

**Immunofluorescence (Fig. 2b and Supplementary Fig. 3, 4, 5, and 9)**

Following crosslinking, the cells were washed once with PBS, twice with blocking buffer (PBS with 0.25% BSA, Sigma, A3059-10G) and then permeabilized for 15 minutes with 0.2% Triton X-100 (Sigma, T8797-100ml). After two washes with blocking buffer cells were stained overnight at 4 degrees C for either Oct4 (Santa Cruz Biotechnology, sc-9081x; 1:200 dilution) or SSEA1 (mc-480, DHSB, 1:20 dilution) and washed twice with blocking buffer. Cells were incubated for 1 hour at room temperature with either goat anti-rabbit-conjugated Alexa Fluor 488, goat anti-rabbit-conjugated Alexa Fluor 647 or goat anti-mouse conjugated 568 (Invitrogen; 1:1000 dilution) and Hoechst 33342 (Invitrogen; 1:2000 dilution). Finally, cells were washed twice with blocking buffer and twice with PBS before imaging. Images were acquired on either Nikon Inverted TE300 or Zeiss Axiovert 200m Inverted microscopes with a Hamamatsu Orca camera. Openlab (http://www.improvision.com/products/openlab/) was used for image acquisition.

**Alkaline Phosphatase Staining (Supplementary Fig. 2, 4, 5, and 9)**

Staining of ZHBTC4 cells for alkaline phosphatase was achieved using the Alkaline Phosphatase Detection Kit (Millipore, SCR004). Briefly, cells were crosslinked at various timepoints before
addition of Fast Red Violet solution and Napthol AS-BI phosphate solution. Cells were visualized on a Nikon Inverted TE300 with a Hamamatsu Orca camera. Openlab (http://www.improvision.com/products/openlab/) was used for image acquisition.

*RNA Extraction, cDNA, and TaqMan Expression Analysis (Supplementary Fig. 1, 3, 5, and 9)*

RNA utilized for real-time qPCR was extracted with TRIzol according to the manufacturer protocol (Invitrogen, 15596-026). Purified RNA was reverse transcribed using Superscript III (Invitrogen) with oligo dT primed first-strand synthesis following the manufacturer protocol.

Real-time qPCR were carried out on the 7000 ABI Detection System using the following Taqman probes according to the manufacturer protocol (Applied Biosystems).

- Gapdh Mm99999915_g1
- LSD1 Mm01181030_g1
- Mi-2b Mm01190896_m1
- CoREST Mm03053471_s1
- Sox2 Mm00488369_s1
- Nanog Mm02019550_s1
- Lefty1 Mm00438615_m1
<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lefty2</td>
<td>Mm00774547_ml</td>
</tr>
<tr>
<td>Esrrb</td>
<td>Mm00442411_ml</td>
</tr>
<tr>
<td>Trim28</td>
<td>Mm00495594_ml</td>
</tr>
<tr>
<td>Klf4</td>
<td>Mm00516104_g1</td>
</tr>
<tr>
<td>Oct4</td>
<td>Mm03053917_g1</td>
</tr>
<tr>
<td>Apobec1</td>
<td>Mm00482895_ml</td>
</tr>
<tr>
<td>Dppa3</td>
<td>Mm01184198_g1</td>
</tr>
<tr>
<td>Tcl1</td>
<td>Mm00493477_ml</td>
</tr>
</tbody>
</table>

Expression levels were normalized to Gapdh levels. All comparisons were made relative to either DMSO (control)-treated, Luciferase, or GFP-infected cells.

**Chromatin Immunoprecipitation (ChIP)**

*Antibody Specificity (Supplementary Fig. 1, 6, and 9)*

For LSD1 (AOF2/KDM1A)-occupied genomic regions, we performed ChIP-Seq experiments using Abcam ab17721 rabbit polyclonal antibody. The antibody was raised with a synthetic peptide conjugated to KLH derived from within residues 800 to the C-terminus of human LSD1. Antibody specificity was determined by shRNA-mediated knockdown (Open Biosystems) of
LSD1 proteins, followed by Western blot analysis. Knockdowns were carried out using the following shRNAs according to the manufacturer protocol (Open Biosystems). Cells were infected with indicated short hairpins for 5 days, followed by protein extraction and Western blotting.

<table>
<thead>
<tr>
<th>GFP</th>
<th>RHS4459</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSD1</td>
<td>TRCN0000071374</td>
</tr>
</tbody>
</table>

For Mi-2b (CHD4)-occupied genomic regions, we performed ChIP-Seq experiments using Abcam ab72418 rabbit polyclonal antibody. The antibody was raised with a synthetic peptide conjugated to KLH derived from within residues 25 to 75 of human CHD4. Antibody specificity was determined by shRNA-mediated knockdown (Open Biosystems) of Mi-2b proteins, followed by Western blot analysis. Knockdowns were carried out using the following shRNAs according to the manufacturer protocol (Open Biosystems). Cells were infected with indicated short hairpins for 5 days, followed by protein extraction and Western blotting.

<table>
<thead>
<tr>
<th>GFP</th>
<th>RHS4459</th>
</tr>
</thead>
<tbody>
<tr>
<td>shMi-2b #1</td>
<td>TRCN0000086143</td>
</tr>
<tr>
<td>shMi-2b #2</td>
<td>TRCN0000086147</td>
</tr>
</tbody>
</table>
For HDAC1-occupied genomic regions, we performed ChIP-Seq experiments using Abcam ab7028 rabbit polyclonal antibody. The antibody was raised with a synthetic peptide conjugated to KLH derived from within residues 466 to 482 of human HDAC1. Antibody specificity was previously determined (Wang et al., 2009b).

For HDAC2-occupied genomic regions, we performed ChIP-Seq experiments using Abcam ab7029 rabbit polyclonal antibody. The antibody was raised with a synthetic peptide conjugated to KLH derived from within residues 471 to 488 of human HDAC2. Antibody specificity was previously determined (Wang et al., 2009b).

For REST (NRSF)-occupied genomic regions, we performed ChIP-Seq experiments using Millipore 07-579 rabbit polyclonal antibody. The antibody was raised with a GST fusion protein corresponding to residues 801-1097 of human REST. Antibody specificity was previously determined (Singh et al., 2008).

For CoREST-occupied genomic regions, we performed ChIP-Seq experiments using Abcam ab32631 rabbit polyclonal antibody. The antibody was raised with a synthetic peptide conjugated to KLH derived from within residues 450 to the C-terminus of human CoREST. Antibody specificity was determined by shRNA-mediated knockdown (Open Biosystems) of CoREST proteins, followed by Western blot analysis. Knockdowns were carried out using the following shRNAs according to the manufacturer protocol (Open Biosystems). Cells were infected with indicated short hairpins for 5 days, followed by protein extraction and Western blotting.
For H3K4me1-occupied genomic regions, we performed ChIP-Seq experiments using Abcam ab8895 rabbit polyclonal antibody. The antibody was raised with a synthetic peptide conjugated to KLH derived from within residues 1 to 100 of human H3K4me1. Antibody specificity was previously determined (Meissner et al., 2008).

For p300-occupied genomic regions, we performed ChIP-PCR experiments using Santa Cruz sc-584 rabbit polyclonal antibody. The antibody was raised with a synthetic peptide derived from the N-terminus of human p300. Antibody specificity was previously determined (Chen et al., 2008).

For H3K27Ac-occupied genomic regions, we performed ChIP-PCR experiments using Abcam ab4729 rabbit polyclonal antibody. The antibody was raised with a synthetic peptide conjugated to KLH derived from within residues 1-100 of human histone H3, acetylated at K27. Antibody specificity was previously determined (Creyghton et al., 2010).

ChIP Protocol

Protocols describing chromatin immunoprecipitation materials and methods have been previously described (Boyer et al., 2006). v6.5 or ZHBTc4 ESCs were grown to a final count of 198
5-10 x 10^7 cells for each ChIP experiment. Cells were chemically crosslinked by the addition of one-tenth volume of fresh 11% formaldehyde solution for 15 minutes at room temperature. Cells were rinsed twice with 1X PBS and harvested using a silicon scraper and flash frozen in liquid nitrogen. Cells were stored at -80°C prior to use. Cells were resuspended, lysed in lysis buffers and sonicated to solubilize and shear crosslinked DNA. Sonication conditions vary depending on cells, culture conditions, crosslinking and equipment.

For LSD1, Mi-2b, HDAC1, HDAC2, REST, CoREST, p300, H3K27Ac, and H3K4me1 the sonication buffer was 20mM Tris-HCl pH8, 150mM NaCl, 2mM EDTA, 0.1% SDS, 1% Triton X-100. We used a Misonix Sonicator 3000 and sonicated at approximately 24 watts for 10 x 30 second pulses (60 second pause between pulses). Samples were kept on ice at all times. The resulting whole cell extract was incubated overnight at 4 degrees C with 100ul of Dynal Protein G magnetic beads that had been pre-incubated with approximately 10 ug of the appropriate antibody. Beads were washed 1X with the sonication buffer, 1X with 20mM Tris-HCl pH8, 500mM NaCl, 2mM EDTA, 0.1% SDS, 1% Triton X-100, 1X with 10mM Tris-HCl pH8, 250mM LiCl, 2mM EDTA, 1% NP40 and 1X with TE containing 50 mM NaCl.

Bound complexes were eluted from the beads (50 mM Tris-Hcl, pH 8.0, 10 mM EDTA and 1% SDS) by heating at 65 degrees C for 1 hour with occasional vortexing and crosslinking was reversed by overnight incubation at 65°C. Whole cell extract DNA reserved from the sonication step was also treated for crosslink reversal.
Gene Specific ChIPs (Supplementary Fig. 9, 10, and 11)

For gene specific ChIPs carried out in ZHBTc4 ESCs, approximately 5x10⁷ ES cells were grown on two 15cm plates. Plates were treated or not treated with 2ng/ml doxycycline for 12 hours, and crosslinked. SYBR Green real-time qPCR was carried out on the 7000 ABI Detection System according to the manufacturer protocol (Applied Biosystems). Data was normalized to the whole cell extract and control regions. Primer pairs are listed below.

Lefty1

5'-GTAGCCAGCAGACAGGACAA-3'
5'-ATCCCCAATCCACATTCACT-3'

Lefty2

5'-AGGCCTAGCTTTTGATCAC-3'
5'-TCTCCCAGAGTCGATCTTCC-3'

Sall4

5'-GAAATAAACATCTGGGAGAAGGA-3'
5'-GGAAACCCCAGATTGAGAGA-3'
Sox2

5' - TGGCGAGTGTTAAACAGAG - 3'
5' - TAGCGAGAACTAGCCAAGCA - 3'

Trim28

5' - GGTCTGCAATTGAAGGAAGG - 3'
5' - TTAAACAGCAGGGGTAAAGG - 3'

Esrrb

5' - CGAGCTTCAGCTGGCTATTT - 3'
5' - GAGCTCCAGATCCCCTACAC - 3'

Nanog

5' - GGAATTTCTTCCCAGGT - 3'
5' - GGTTGGAACCTAGCTTGTTG - 3'
Ctrl (Olfr460)

5'-AACTGTTATTGTGCCCGTGA -3'

5'-CATTGTCCAAGGCAAGAAA -3'

ChIP-Seq Sample Preparation and Analysis

All protocols for Illumina/Solexa sequence preparation, sequencing and quality control are provided by Illumina (http://www.illumina.com/pages.ilmn?ID=203). A brief summary of the technique and minor protocol modifications are described below.

Sample Preparation

Purified chromatin immunoprecipitated (ChIP) DNA was prepared for sequencing according to a modified version of the Illumina/Solexa Genomic DNA protocol. Approximately 200 ng of ChIP DNA was prepared for ligation of Solexa linkers by repairing the ends and adding a single adenine nucleotide overhang to allow for directional ligation. A 1:200 dilution of the Adaptor Oligo Mix (Illumina) was used in the ligation step. A subsequent PCR step with 18 amplification cycles added additional linker sequence to the fragments to prepare them for annealing to the Genome Analyzer flow-cell. Amplified material was purified by Qiaquick MinElute (Qiagen, Valencia, CA) and a narrow range of fragment sizes was selected by separation on a 2% agarose gel and excision of a band between 150-300bp, representing ChIP fragments between 50 and 200
nt in length and ~100bp of primer sequence. The DNA was purified from the agarose and diluted to 10nM for loading on the flow cell.

For multiplexed samples, libraries were prepared using the Illumina TruSeq adapters (to enable multiplexing) and prepared using Beckman-Coulter's SPRIworks system. For library preparations, ChIP samples were used in their entirety and whole-cell extracts control samples were prepared using 100ng. Adapters were diluted to 1:200. Size selection was 200–400bp before PCR, and the samples were amplified using KAPA Hi-Fi polymerase and 18 cycles of PCR according to manufacturer's cycling recommendations. Amplified material was purified using Agencourt Ampure XP beads using a 0.93 ratio of beads to sample.

**Polony Generation and Sequencing**

The DNA library (2–5pM) was applied to the flow-cell (8 samples per flow-cell) using the Cluster Station device from Illumina. The concentration of library applied to the flow-cell was calibrated such that polonies generated in the bridge amplification step originate from single strands of DNA. Multiple rounds of amplification reagents were flowed across the cell in the bridge amplification step to generate polonies of approximately 1,000 strands in 1mm diameter spots. Double-stranded polonies were visually checked for density and morphology by staining with a 1:5000 dilution of SYBR Green I (Invitrogen) and visualizing with a microscope under fluorescent illumination. Validated flow-cells were stored at 4°C until sequencing. Flow-cells were removed from storage and subjected to linearization and annealing of sequencing primer on the Cluster Station. Primed flow-cells were loaded into the Illumina Genome Analyzer II or Hi-
seq. After the first base was incorporated in the Sequencing by-Synthesis reaction the process was paused for a quality control checkpoint. A small section of each lane was imaged and the average intensity value for all four bases was compared to minimum thresholds. Flow-cells with low first base intensities were reprimed and if signal was not recovered the flow-cell was aborted. Flow-cells with signal intensities meeting the minimum thresholds were resumed and sequenced for 26 cycles. For multiplexed samples Truseq V2.5 kits were used to cluster them on the cBot and Truseq V2 were used to do a multiplex 40+7 cycle run on the Hi-seq.

**ChIP-Seq Data Analysis**

Images acquired from the Illumina/Solexa sequencer were processed through the bundled Solexa image extraction pipeline which identified polony positions, performed base-calling and generated QC statistics.

ChIP-Seq reads were aligned using the software Bowtie (Langmead et al., 2009) to NCBI build 36 (mm8) of the mouse genome with default settings. Sequences uniquely mapping to the genome with zero or one mismatch were used in further analysis.

When multiple reads mapped to the same genomic position, a maximum of two reads mapping to the same position were used. ChIP-Seq datasets profiling the genomic occupancy of H3K36me2 (Marson et al., 2008), H3K79me2 (Marson et al., 2008), H3K4me3 (Marson et al., 2008), H3K27me3 (Mikkelsen et al., 2007), Oct4 (Marson et al., 2008), Sox2 (Marson et al., 2008),
Nanog (Marson et al., 2008), RNA Polymerase 2 (Seila et al., 2008), TBP (Kagey et al., 2010), Med1 (Kagey et al., 2010), p300 (Chen et al., 2008), H3K4me1 (Meissner et al., 2008), H3K27ac (Creyghton et al., 2010) and Suz12 (Marson et al., 2008) in mouse ESCs were obtained from previous publications. Below is the list of ChIP-Seq datasets used and corresponding GEO Accession numbers.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>GEO Accession Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>H3K27ac</td>
<td>GSM594578, GSM594579</td>
</tr>
<tr>
<td>H3K27me3</td>
<td>GSM307619</td>
</tr>
<tr>
<td>H3K36me3</td>
<td>GSM307152, GSM307153</td>
</tr>
<tr>
<td>H3K4me1</td>
<td>GSM281695</td>
</tr>
<tr>
<td>H3K4me3</td>
<td>GSM307146</td>
</tr>
<tr>
<td>H3K79me2</td>
<td>GSM307150, GSM307151</td>
</tr>
<tr>
<td>Med1</td>
<td>GSM560347, GSM560348</td>
</tr>
<tr>
<td>WCE mES</td>
<td>GSM307154, GSM307155, GSM560357</td>
</tr>
<tr>
<td>WCE mES (matching H3K4me1 mES)</td>
<td>GSM307625</td>
</tr>
<tr>
<td>Nanog</td>
<td>GSM307140, GSM307141</td>
</tr>
<tr>
<td>Oct4</td>
<td>GSM307137</td>
</tr>
<tr>
<td>p300</td>
<td>GSM288359</td>
</tr>
<tr>
<td>RNA Pol II</td>
<td>GSM318444</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Sox2</td>
<td>GSM307138, GSM307139</td>
</tr>
<tr>
<td>Suz12</td>
<td>GSM307144, GSM307145</td>
</tr>
<tr>
<td>TBP</td>
<td>GSM555160, GSM555162</td>
</tr>
<tr>
<td>DMSO H3K4me1</td>
<td>This Paper</td>
</tr>
<tr>
<td>48HR DMSO H3K4me1</td>
<td>This Paper</td>
</tr>
<tr>
<td>48HR TCP H3K4me1</td>
<td>This Paper</td>
</tr>
<tr>
<td>HDAC1</td>
<td>This Paper</td>
</tr>
<tr>
<td>HDAC2</td>
<td>This Paper</td>
</tr>
<tr>
<td>LSD1</td>
<td>This Paper</td>
</tr>
<tr>
<td>Mi-2b</td>
<td>This Paper</td>
</tr>
<tr>
<td>REST</td>
<td>GSM656525, This Paper</td>
</tr>
<tr>
<td>CoREST</td>
<td>This Paper</td>
</tr>
<tr>
<td>WCE – DMSO</td>
<td>This Paper</td>
</tr>
<tr>
<td>WCE – 48HR DMSO</td>
<td>This Paper</td>
</tr>
<tr>
<td>WCE – 48HR TCP</td>
<td>This Paper</td>
</tr>
</tbody>
</table>

All ChIP-Seq datasets, including those obtained elsewhere, were analyzed using the methods described below.
Analysis methods were derived from previously published methods (Marson et al., 2008; Mikkelsen et al., 2007; Guenther et al., 2008; Johnson et al., 2007). Sequence reads from multiple flow cells for each IP target and/or biological replicates were combined. For all datasets, excluding H3K4me1, H3K4me3, H3K79me2, H3K36me3, H3K27ac, and H3K27me3 each read was extended 200bp, towards the interior of the sequenced fragment, based on the strand of the alignment. For H3K4me1, H3K4me3, H3K79me2, H3K36me3, H3K27ac, and H3K27me3 datasets, each read was extended 600bp towards the interior and 400bp towards the exterior of the sequenced fragment, based on the strand of the alignment. Across the genome, in 25 bp bins, the number of extended ChIP-Seq reads was tabulated. The 25bp genomic bins that contained statistically significant ChIP-Seq enrichment were identified by comparison to a Poissonian background model. Assuming background reads are spread randomly throughout the genome, the probability of observing a given number of reads in a genomic bin can be modeled as a Poisson process in which the expectation can be estimated as the number of mapped reads multiplied by the number of bins (8 for all sequences datasets except H3K4me1, H3K4me3, H3K79me2, H3K36me3, H3K27ac, and H3K27me3, which was 40) into which each read maps, divided by the total number of bins available (we estimated 70% of the genome). Enriched bins within 200bp of one another were combined into regions.

The Poissonian background model assumes a random distribution of background reads, however we have observed significant deviations from this expectation. Some of these non-random events can be detected as sites of apparent enrichment in negative control DNA samples and can
create many false positives in ChIP-Seq experiments. To remove these regions, we compared genomic bins and regions that meet the statistical threshold for enrichment to a set of reads obtained from Solexa sequencing of DNA from whole cell extract (WCE) in matched cell samples. We required that enriched bins and enriched regions have five-fold greater ChIP-Seq density in the specific IP sample, compared with the control sample, normalized to the total number of reads in each dataset. This served to filter out genomic regions that are biased to having a greater than expected background density of ChIP-Seq reads. A summary of the bound regions and genes for each antibody is provided (Supplementary Table 1).

Assigning ChIP-Seq Enriched Regions to Genes (Supplementary Table 1)

The complete set of RefSeq genes was downloaded from the UCSC table browser (http://genome.ucsc.edu/cgi-bin/hgTables) on June 1, 2010. For all datasets except H3K4me1, H3K4me3, H3K79me2, H3K36me3, H3K27ac, and H3K27me3, genes with enriched regions within 10kb of their transcription start site were called bound. For H3K4me1, H3K4me3, H3K79me2, H3K36me3, H3K27ac, and H3K27me3 datasets, genes with enriched regions within the gene body were called bound.

Enrichment of LSD1, Pol II, and Suz12 at genes (Fig. 1)

Each gene in the mouse genome was classified into active, bivalent, or silent groups based on the presence of co-occupancy of H3K4me3 (GSE11724), H3K79me2 (GSE11724), and H3K27me3 (GSM307619). See the table below for a description of the number of genes in each category as
well as the classification rules. Using a hypergeometric test, the p-value for enrichment of LSD1, Pol II, and Suz12 binding to the genes in each of these classes was determined.

<table>
<thead>
<tr>
<th>Classification rules</th>
<th>H3K4me3 within +/- 2kb of the TSS</th>
<th>H3K79me2 within first 5kb of gene body</th>
<th>H3K27me3 within +/- 5kb of the TSS</th>
<th>Number of genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>Required</td>
<td>Required</td>
<td>-</td>
<td>7339</td>
</tr>
<tr>
<td>Bivalent</td>
<td>Required</td>
<td>-</td>
<td>Required</td>
<td>3094</td>
</tr>
<tr>
<td>Silent</td>
<td>Absent</td>
<td>Absent</td>
<td>-</td>
<td>6976</td>
</tr>
</tbody>
</table>

*Definition of Enhancer and Core Promoter (Fig. 1, 4, Supplementary Fig. 8, 13, and Supplementary Table 1)*

An enhancer was defined as a Med1 high-confidence enriched region that is overlapped at least 1bp by enriched regions of Oct4, Sox2, and Nanog in mES cells. Using this definition we identified 3,838 enhancers, and the genomic locations of those enhancers are listed in Supplementary Table 1. These enhancers were assigned to the nearest genes to determine the neighboring core promoter regions.
Determining LSD1 density at Enhancer and Core Promoter (Supplementary Fig. 1)

The mean LSD1 density at enhancers (15.91) and core promoters (13.46) was determined by calculating the sum of LSD1 density within 5kb of each enhancer and core promoter, and then taking the average across the 3,838 enhancers and core promoters. A t-test was then performed to obtain a p-value (p < 10^{-16}) indicating that LSD1 density was statistically larger at enhancers compared to core promoters.

ChIP-Seq Density Heatmaps (Fig. 1, 4, Supplementary Fig. 8, and 13)

Selected enriched regions were aligned with each other according to the position of either Med1 (Fig. 1c, 4e, and Supplementary Fig. 5b) or the TSS (Fig. 1d). For each experiment, the ChIP-Seq density profiles were normalized to the density per million total reads. Heatmaps were generated using Java Treeview (http://jtreeview.sourceforge.net/) with color saturation as indicated.

Calculation of the Statistical Significance of the Overlap Between Sets of Genomic Regions (Fig. 1, 3, 4, and Supplementary Fig. 8)

In the manuscript, when assessing the overlap between two sets of genomic regions, we report p-values calculated using a Monte Carlo method. In order to make this type of comparison, a number of different statistical methods could be applied. One could use a chi-square test, and compare the total size of the overlap of two sets of genomic regions to the size of this overlap that would be expected at random based on the total size of the genome and the total size of the
genomic regions in each group. This method accurately models the expected overlap between two datasets if they were truly randomly associated. Unfortunately, since the genome is so large, this method will call the overlap between two datasets statistically significant even if it is not much larger than would be expected at random.

Another way to assess the statistical significance of the overlap between two sets of genomic regions is to use a Monte Carlo method. In Monte Carlo methods, a tremendous number of simulated datasets are created and the overlaps between the simulated datasets are tabulated. The statistical significant of the actual overlap is then estimated based on the frequency of an overlap at least that large occurring in the simulated datasets. We used a Monte Carlo method to assess the statistical significance of the overlap between two sets of genomic regions in this manuscript. To create the simulated datasets, we shifted each of the genomic regions in one of the datasets a random distance between $-2,000\text{bp}$ and $+2,000\text{bp}$ from its original position. For this manuscript, Monte Carlo simulations were run one billion times.

We noted that, if two sets of genomic regions tend to occur in the same place, then this overlap is statistically significant. This will be true whether the regions occur together 10% of the time, or 100% of the time. Thus, the test for statistical significance is not a very good at describing the degree of overlap between two datasets. Consequently, in each instance that we assessed the statistical significance of the overlap between two sets of genomic regions, we also calculated the fold enrichment of the overlap between the two datasets. This was calculated as the total size of the actual overlap between the two sets of regions and the overlap between the two datasets...
that would be expected at random.

Rank Normalization of H3K4me1 ChIP-Seq Data (Fig. 4, Supplementary Fig. 13, and Supplementary Table 5)

In order to compare the levels of H3K4me1 between three conditions a rank normalization method was used. The three cell types and conditions in which H3K4me1 occupancy was profiled were ZHBTc4 ESCs, ZHBTc4 ESCs treated with doxycycline for 48 hours, and ZHBTc4 ESCs treated with doxycycline for 48 hours and TCP for 54 hours. In each of the three datasets, the genomic bin with the greatest ChIP-Seq signal was identified. The average of these three values was calculated, and the greatest bin in each dataset was assigned this average value. This was repeated for all genomic bins from the greatest signal to the least, assigning each the average ChIP-Seq signal for all bins of that rank across all datasets. Subsequently, the total number of ChIP-Seq reads in the three H3K4me1 datasets was calculated, and this value was used to scale each dataset to the units, rank normalized reads/million.

Using this method, we computed the rank normalized reads/million density in each of the three datasets in 10 bp bins across the genome. The H3K4me1 signal at an enhancer was then calculated as the sum of the observed densities in the 50 bins (+/- 250bp) surrounding each of the 3,838 identified enhancers. This number captures the normalized H3K4me1 signal in dataset, and the value for each enhancer in each of the three datasets is reported in Supplementary Table 5.
Gene Ontology (GO) Analysis (Supplementary Fig. 7)

Gene ontology analysis was performed using the web tool David Bioinformatics Database (http://david.abcc.ncifcrf.gov/home.jsp) (Huang et al., 2009a; Huang et al., 2009b). The complete set of all RefSeq genes was used as a background.

Public availability of ChIP-Seq datasets


Microarray Analysis (Fig. 2)

Cell Culture and RNA isolation

For ESC expression analysis during differentiation following either DMSO or TCP treatment, ZHBTc4 ESCs were split off MEFs, placed in a tissue culture dish for 45 minutes to selectively remove the MEFs and plated in 6-well plates. The following day cells were treated with either DMSO or 1mM TCP for 6 hours in ESC media. After 6 hours a subset of cells from DMSO treated wells was removed for TRIzol RNA isolation, the media was removed and replaced with ESC media containing 2μg/ml doxycycline and either DMSO or 1mM TCP. 24 hours after doxycycline addition, a subset of cells was removed for RNA isolation and the media was replaced with ESC media containing doxycycline and either DMSO or TCP. After another 48 hours, RNA was isolated using TRIzol (Invitrogen, 15596-026) from all time points (at indicated
times), further purified with RNeasy columns (Qiagen, 74104) and DNase treated on column (Qiagen, 79254) following the manufacturer’s protocols.

**Microarray hybridization and Analysis**

For microarray analysis, Cy3 and Cy5 labeled cRNA samples were prepared using Agilent’s QuickAmp sample labeling kit starting with 1ug total RNA. Briefly, double-stranded cDNA was generated using MMLV-RT enzyme and an oligo-dT based primer. *In vitro* transcription was performed using T7 RNA polymerase and either Cy3-CTP or Cy5-CTP, directly incorporating dye into the cRNA.

Agilent mouse 4x44k expression arrays were hybridized according to our laboratory’s standard method, which differs slightly from the standard protocol provided by Agilent. The hybridization cocktail consisted of 825 ng cy-dye labeled cRNA for each sample, Agilent hybridization blocking components, and fragmentation buffer. The hybridization cocktails were fragmented at 60 degrees C for 30 minutes, and then Agilent 2X hybridization buffer was added to the cocktail prior to application to the array. The arrays were hybridized for 16 hours at 60 degrees C in an Agilent rotor oven set to maximum speed. The arrays were treated with Wash Buffer #1 (6X SSPE / 0.005% n-laurylsarcosine) on a shaking platform at room temperature for 2 minutes, and then Wash Buffer #2 (0.06X SSPE) for 2 minutes at room temperature. The arrays were then dipped briefly in acetonitrile before a final 30 second wash in Agilent Wash 3 Stabilization and Drying Solution, using a stir plate and stir bar at room temperature.
Arrays were scanned using an Agilent DNA microarray scanner. Array images were quantified and statistical significance of differential expression for each hybridization was calculated using Agilent's Feature Extraction Image Analysis software with the default two-color gene expression protocol. For each gene in the RefSeq gene list (see ChIP-Seq analysis section), the log10 ratio values (ESC+Dox / ESC, or ESC+DOX+TCP / ESC) and p-value for that gene were determined by averaging log10 ratios and p-values of all Agilent Features annotated to that gene. Genes with no annotated features were reported as NA (Supplementary Table 3). Genes with a log10 ratio of at least 0.0969 (1.25x) and a p-value less than or equal to $10^{-2}$ were determined to have a significant change in expression.

**Public availability of microarray gene expression datasets**


**Protein Extraction and Western Blot Analysis (Fig. 2, Supplementary Fig. 1, 4, 5, 6, and 9)**

ESCs were lysed with CelLytic Reagent (Sigma, C2978-50ml) containing protease inhibitors (Roche). After SDS-PAGE, Western blots were revealed with antibodies against Oct4 (Santa Cruz Biotechnology, sc-5279), LSD1 (Abcam, ab17721), Mi-2b (Abcam, ab72418), CoREST (Abcam, ab32631), p300 (Santa Cruz, sc-584), GAPDH (Abcam, ab9484) or Tubulin (Millipore, 05-661).
ChIP-Western and Co-Immunoprecipitation (Fig. 3)

For ChIP-Western, the same conditions as for ChIP-Seq were used. For co-immunoprecipitation, murine ES cells were harvested in cold PBS and extracted for 30 min at 4 degrees C in TNEN250 (50 mM Tris pH 7.5, 5 mM EDTA, 250 mM NaCl, 0.1% NP-40) with protease inhibitors. After centrifugation, supernatant was mixed to 2 volumes of TNENG (50 mM Tris pH 7.5, 5 mM EDTA, 100 mM NaCl, 0.1% NP-40, 10% glycerol). Protein complexes were immunoprecipitated overnight at 4 degrees C using 5 micrograms of either LSD1 (Abcam, ab17721), HDAC1 (Abcam, ab7028), HDAC2 (Abcam, ab7029), Mi-2b (Abcam, ab72418), CoREST (Abcam, ab32631) or Rabbit IgG (Upstate, 12-370), bound to 50ul of Dynabeads®. Immunoprecipitates were washed three times with TNEN125 (50 mM Tris pH 7.5, 5 mM EDTA, 125 mM NaCl, 0.1% NP-40). For both ChIP-Western and co-immunoprecipitation, beads were boiled for 10 minutes in XT buffer (Biorad) containing 100mM DTT to elute proteins. After SDS-PAGE, Western blots were revealed with antibodies against either LSD1 (Abcam, ab17721), or HDAC1 (Abcam, ab7028).
Supplementary References


Supplementary Figure Legends

Supplementary Fig. 1: Validation and characterization of LSD1 ChIP-Seq dataset

a, Validation of shRNA targeting LSD1. qPCR of LSD1 transcripts in v6.5 ESCs infected with short hairpins (Open Biosystems) against either LSD1 or GFP (control) for 5 days. The data is normalized to Gapdh. The error bars represent the standard deviation of triplicate PCR reactions.

b, LSD1 antibody is specific. Western blot of LSD1 protein levels in GFP or LSD1 knockdown cells, showing a decrease in signal of one band corresponding to LSD1 molecular weight (MW = 92.9 kDa). Tubulin served as loading control.

c, LSD1 ChIP-Seq signals are above experimental noise. ChIP-Seq binding profiles (reads/million) for LSD1 at the Oct4 (Pou5f1) and Lefty1 loci in ESCs, with the y-axis floor set to 1. Gene models, and previously described enhancer regions1-3 are depicted below the binding profiles.

d, ChIP-Seq data, at high resolution, showing coincidental binding of Oct4, Sox2, and Nanog with LSD1. High-resolution ChIP-Seq binding profiles (reads/million) at the Oct4 (Pou5f1) and Lefty1 enhancer loci in ESCs, with y-axis floor at 1.0 reads/million.

e, Metagene showing mean LSD1 density is significantly higher at the 3,838 ESC enhancers compared to 3,838 neighboring core promoters (p<10^-16).

Supplementary Fig. 2: LSD1 is required for doxycycline-induced differentiation of ZHBTc4 ESCs

a, Schematic representation of trophectoderm differentiation assay using doxycycline-inducible Oct4 depletion murine ESC line ZHBTc4. Treatment with doxycycline for 48 hours leads to depletion of Oct4 and early trophectoderm specification. ZHBTC4 cells were treated with DMSO (control) or the LSD1 inhibitors Tranylcypromine (TCP) and Pargyline (Prg) for 6 hours before 2μg/ml doxycycline was added for an additional 48 hours.

b, c, ESCs treated with either
(b) 1mM tranylcypromine (TCP) or (c) 3mM pargyline (Prg) maintained ESC colony morphology and alkaline phosphatase (AP) staining in doxycycline-treated cells despite depletion of Oct4 protein levels. Scale bar = 20μM. d, Schematic representation of trophectoderm differentiation assay using doxycycline-inducible Oct4 depletion murine ESC line ZHBTc4. ESCs were infected with a GFP or LSD1 shRNA for 72 hours before 2μg/ml doxycycline was added for an additional 48 hours. e, ESCs infected with an shRNA targeting LSD1 maintained ESC colony morphology and alkaline phosphatase (AP) staining in doxycycline-treated cells despite depletion of Oct4 protein levels. Scale bar = 20μM.

**Supplementary Fig. 3: LSD1 is required for repression of ESC genes**

a, Treatment of Prg partially maintained SSEA-1 cell surface marker expression in doxycycline-treated cells. Cells were stained for Hoechst (Hoe), Oct4 and SSEA-1. Scale bar = 100μM. b, Treatment of Prg partially relieved repression of ESC genes 48 hours after Oct4 depletion in DMSO- versus Prg-treated cells. Error bars reflect standard deviation of either duplicate or triplicate PCR reactions. c, ESCs with decreased LSD1 levels partially maintained SSEA-1 cell surface marker expression in doxycycline-treated cells. Cells were stained for Hoechst (Hoe), Oct4 and SSEA-1. Scale bar = 20μM. d, Reduced LSD1 levels partially relieved repression of ESC genes 48 hours after Oct4 depletion in GF versus LSD1 shRNA. Error bars reflect standard deviation of either duplicate or triplicate PCR reactions.

**Supplementary Fig. 4: Characterization of LSD1 knockout ESCs**
a, Beginning with an E14 ESC line (WT), cells were engineered to express an inducible Cre-ER fusion protein from the endogenous ROSA26 locus. Sequential gene targeting produced cells in which one LSD1 allele has exon 3 flanked by LoxP sites (floxed) and the second has exon 3 deleted (Lox/Δ3). Induction of Cre activity by addition of 4-hydroxytamoxifen (4-OHT) to the growth medium resulted in complete recombination of the remaining floxed LSD1 allele, generating LSD1 homozygous null Δ3/Δ3) cells. Loss of exon 3 disrupts the open reading frame of LSD1 such that a premature stop codon is introduced into exon 4, resulting in the progressive loss of LSD1 protein by 48 hours, as detected by Western blot (WB). Gapdh served as a loading control. b, LSD1 heterozygous (Lox/Δ3) and homozygous null (Δ3/Δ3) ESCs maintain Oct4 and SSEA-1 cell surface marker expression compared to E14 wild type (WT) cells. Cells were stained for Hoechst (Hoe), Oct4 and SSEA-1. Scale bar = 100μM. c, LSD1 heterozygous (Lox/Δ3) or homozygous null (Δ3/Δ3) ESCs maintain ESC morphology and alkaline phosphatase (AP) staining compared to E14 wild type (WT) ESCs. Scale bar = 100μM. LSD1 Lox/Δ3 ESCs are used as a control throughout the remainder of the study. d, Schematic representation of trophectoderm differentiation assay. LSD1 Lox/Δ3 cells were treated for 48 hours with either ethanol (control) or 4-OHT to generate LSD1 Δ3/Δ3 ESCs before being infected with a GFP (shRNAOct4, control) or Oct4 (shRNAOct4) shRNA to knockdown Oct4 and induce differentiation. e, LSD1 Δ3/Δ3 ESCs partially maintained SSEA-1 cell surface marker expression when Oct4 levels are decreased. Cells were stained for Hoechst (Hoe), Oct4 and SSEA-1. Scale bar =100μM. f, LSD1 Δ3/Δ3 ESCs partially maintained ESC colony morphology and alkaline phosphatase (AP) staining when Oct4 levels are decreased.

Supplementary Fig. 5: Catalytic activity of LSD1 is required for repression of ESC genes
a. Overexpression of LSD1 wildtype and catalytic mutant in LSD1 homozygous null (Δ3/Δ3) ESCs. Homozygous null (Δ3/Δ3) cells were transfected with LSD1 wildtype (EGFP-LSD1wt) and a catalytic mutant (EGFP-LSD1mut) fused to EGFP. LSD1 expression in Δ3/Δ3 ESCs are relative to signal in LSD1 heterozygous (Lox/Δ3) cells. The data is normalized to Gapdh. The error bars represent the standard deviation of triplicate PCR reactions. b. Western blot (WB) showing homozygous null (Δ3/Δ3) ESCs transfected with LSD1 transgenes express high levels of LSD1 protein. c. Homozygous null (Δ3/Δ3) and homozygous null (Δ3/Δ3) ESCs expressing the LSD1 transgenes maintain Oct4 and SSEA-1 cell surface marker expression compared to LSD1 heterozygous (Lox/Δ3) ESCs. Cells were stained for Hoechst (Hoe), Oct4 and SSEA-1. Scale bar = 100μM. d. Homozygous null (Δ3/Δ3) ESCs expressing LSD1 transgenes maintain ESC morphology and alkaline phosphatase (AP) staining compared to LSD1 heterozygous (Lox/Δ3) ESCs. Scale bar = 100μM. e. Schematic representation of trophectoderm differentiation assay. LSD1 Lox/Δ3 cells were treated for 48 hours with either ethanol (control) or 4-OHT to generate LSD1 Δ3/Δ3 ESCs before transfected with either an LSD1 wildtype or catalytic mutant (K661A) transgene fused to EGFP (EGFP-LSD1wt or EGFP-LSD1mut). Cells were transduced with an shRNA targeting either Luciferase (control) or Oct4 (shRNAOct4) to knockdown Oct4 and induce differentiation. f. Validation of shRNA targeting Oct4. qPCR of Oct4 transcripts in LSD1 heterozygous (Lox/Δ3) ESCs infected with a short hairpin (Open Biosystems) against either Oct4 or Luciferase (control) for 72 hours. The data is normalized to Gapdh. The error bars represent the standard deviation of triplicate PCR reactions. g. Western blot of Oct4 protein levels in Luciferase (control) or Oct4 knockdown cells, showing a decrease in signal of one band corresponding to Oct4 molecular weight (MW = 38 kDa). Tubulin served as loading control. h. Expression of ESC genes was downregulated during differentiation when wildtype, but not the
catalytic mutant, LSD1 was introduced into homozygous null (A3/A3) cells. Expression of ESC genes was maintained in differentiating LSD1 null cells compared to control cells. Overexpression of wildtype LSD1, but not the catalytic mutant, rescued this phenotype. Error bars reflect standard deviation of triplicate PCR reactions.

**Supplementary Fig. 6: Validation of CoREST antibody**

a, Validation of shRNA targeting CoREST. qPCR of CoREST transcripts in v6.5 ESCs infected with short hairpins (Open Biosystems) against either CoREST or GFP (control) for 5 days. The data is normalized to Gapdh. The error bars represent the standard deviation of triplicate PCR reactions. b, CoREST antibody is specific. Western blot of CoREST protein levels in GFP or CoREST knockdown cells showing a decrease in signal of one band corresponding to CoREST molecular weight (MW = 70 kDa). Gapdh served as loading control.

**Supplementary Fig. 7: LSD1 occupies a subset of genomic sites with REST**

a, LSD1 occupies a subset of genomic sites with REST. ChIP-Seq binding profiles (reads/million) for transcription factors (Oct4, Sox2, Nanog, REST), chromatin regulator (LSD1), the transcriptional apparatus (Pol II, TBP) and histone modifications (H3K4me1, H3K4me3, H3K79me2, H3K36me3) at the NeuroD4 and VGF loci in ESCs. Gene models are depicted below the binding profiles. b, GO analysis on the top 500 (based on peak height) REST bound genes co-occupied by LSD1 reveals significant enrichment for genes involved in neurogenesis.
Supplementary Fig. 8: LSD1 occupies ESC enhancers with NuRD

LSD1 occupies ESC enhancer sites with Mi-2b and either HDAC1 or HDAC2. Density map of ChIP-Seq data at Oct4, Sox2, Nanog, and Med1 co-occupied enhancer regions. Data is shown for chromatin regulators (LSD1, Mi-2b, and either HDAC1 or HDAC2) in ESCs. Over 70% of the 3,838 high confidence enhancers were co-occupied by LSD1 and NuRD (p < 10^-9). Color scale indicates ChIP-seq signal in reads per million.

Supplementary Fig. 9: NuRD is required for doxycycline-induced differentiation of ZHBTc4 ESCs

a, Validation of shRNAs targeting Mi-2b. qPCR of Mi-2b transcripts in v6.5 ESCs infected with short hairpins (Open Biosystems) against either Mi-2b or GFP (control) for 5 days. The data is normalized to Gapdh. The error bars represent the standard deviation of triplicate PCR reactions. b, Mi-2b antibody is specific. Western blot of Mi-2b protein levels in GFP or Mi-2b knockdown cells, showing a decrease in signal of two bands corresponding to Mi-2b molecular weight (MW = 280 and 218 kDa). Gapdh served as loading control. c, ESCs infected with shRNAs targeting Mi-2b partially maintained ES cell colony morphology and alkaline phosphatase (AP) staining in doxycycline-treated cells despite depletion of Oct4 protein levels. Scale bar = 20μM. d, ESCs with decreased Mi-2b levels partially maintained SSEA-1 cell surface marker expression in doxycyclinetreated cells. ESCs were infected with Mi-2b shRNA as represented in Supplementary Fig. 2d and were stained for Hoechst (Hoe), Oct4 and SSEA-1. Scale bar = 20μM. e, Mi-2b is required for downregulation of ESC genes. Expression of ESC genes was
maintained 48 hours after Oct4 depletion in GFP versus Mi-2b shRNA treated cells. Error bars reflect standard deviation of either duplicate or triplicate PCR reactions.

Supplementary Fig. 10: Reduced p300 and H3K27Ac levels at ESC enhancers during differentiation

a, b, Reduced levels of transcriptional coactivator p300 and H3K27Ac at enhancers during ESC differentiation. ChIP–PCR analysis of p300 and H3K27Ac enrichment at enhancer loci in doxycycline-treated (+dox) and untreated (-dox) ZHBTc4 cells. Error bars reflect standard deviation of either duplicate or triplicate PCR reactions.

Supplementary Fig. 11: LSD1 occupies ESC enhancers during differentiation

a, b, LSD1 is approximately 0-5% reduced at 12 hours and 30-50% at 48 hours following Oct4 depletion and differentiation of ESCs. ChIP–PCR analysis of LSD1 enrichment at enhancer loci in doxycycline-treated (+dox) and untreated (-dox) ZHBTc4 cells. Error bars reflect standard deviation from biological replicates.

Supplementary Figure 12: p300 inhibition results in reduced H3K27ac and H3K4me1 levels at ESC enhancers

a, Reduced levels of H3K27ac and H3K4me1 at enhancers of ESCs treated with p300 inhibitor C-646. ChIP-Seq binding profiles (reads/million) for chromatin regulators (p300, LSD1) at the
Lefty1 locus in ESCs. Below these profiles, histone H3K27ac and H3K4me1 levels are shown for DMSO treated (control) ESCs, and cells treated with p300 inhibitor C-646 for 24 hours. For appropriate normalization, ChIP-Seq data for histone H3K27ac and H3K4me1 is shown as rank normalized reads/million with the y-axis floor set to 1 (Supplementary Information). Gene models, and previously described enhancer regions3 are depicted below the binding profiles. b, Mean of the normalized H3K27ac and H3K4me1 density +/- 250 nucleotides surrounding LSD1-occupied enhancer regions in the presence or absence of C-646. The associated genes were identified based on their proximity to the LSD1-occupied enhancers (Supplementary Information). c, Mean of the normalized H3K27ac and H3K4me1 density +/- 250 nucleotides surrounding 1,234 LSD1-occupied enhancers showing reduced levels of H3K27ac in the presence of C-646. Of the 1,234 LSD1-occupied enhancers having reduced levels of H3K27ac upon C-646 treatment, 63% (773) display reduced H3K4me1 levels after C-646 treatment (p < 10-3).

Supplementary Fig. 13: LSD1 is required for H3K4me1 removal at ESC enhancers

Heatmap of normalized H3K4me1 density +/- 250 nucleotides surrounding the 2,755 LSD1 occupied enhancers having reduced levels of H3K4me1 upon differentiation. Of the 2,755 LSD1 occupied enhancers having reduced levels of H3K4me1 upon differentiation, 63% (1,722) display higher H3K4me1 levels after TCP treatment (p < 10-16). Color scale indicates ChIP-Seq signal in normalized reads per million.
Figure S1
Figure S2
Figure S3
Figure S4
Figure S5
Figure S6
Figure S7
Figure S8
Figure S9
Figure S10
Figure S11
Figure S12
Figure S13