Characterization and disruption of cis regulatory elements in cancer

A dissertation presented
by
Rhamy Zeid
to
The Division of Medical Sciences
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Biological and Biomedical Sciences

Harvard University
Cambridge, Massachusetts
April 2016
Characterization and disruption of cis regulatory elements in cancer

Abstract

Enhancers are cis regulatory elements that play key roles in the control of cell-type specific gene expression programs. In cancer, enhancer deregulation plays a key role in maintaining gene regulatory programs that underlie an oncogenic state. This dissertation focuses on understanding and modulating aberrant enhancer activity to identify potential vulnerabilities in human cancers. These studies were empowered by evolving technologies in genome-wide measurements of enhancer factors, computational approaches, and chemical and genetic tools to disrupt enhancer function.

In high-risk pediatric neuroblastoma, the transcription factor MYCN is frequently amplified and treatment options for these patients are largely ineffective thus establishing the need for improved therapeutic options. To identify previously unrecognized dependencies in neuroblastoma, we generated genome-wide maps of the active enhancer gene regulatory landscape leading to the identification of ID1 as an uncharacterized dependency in neuroblastoma. These results outline a strategy to identify alternative therapeutic avenues based on a holistic understanding of aberrant enhancer activity.

While MYCN amplification is the defining feature of high-risk neuroblastoma, a detailed mechanistic understanding of oncogenic transcriptional rewiring has been stalled by a lack of genome-wide binding data.
Here we present the dynamic and temporally resolved landscape of genome-wide MYCN occupancy in neuroblastoma. We find that deregulated MYCN binding at enhancers (termed enhancer invasion) is critical to maintaining the oncogenic station and identify the lineage specific transcription factor TWIST1 as a key collaborator and synthetic lethality of oncogenic MYCN. These data suggest that MYCN enhancer invasion shapes transcriptional amplification in neuroblastoma to promote tumorigenesis.

The development of small molecule inhibitors of the bromodomain and extra-terminal (BET) family of proteins provides a pharmacological strategy to inhibit enhancer activity. The efficacy of BET inhibition in several cancers has prompted efforts to predict and understand mechanisms of resistance to BET inhibition. Here, we use a newly developed class of small molecules to pharmacologically induce targeted degradation of the BET family. In triple negative breast cancer, we demonstrate that targeted BET family degradation effectively overcomes BET inhibitor resistance. These studies suggest BET degradation as a strategy to overcome BET inhibitor resistance and further disrupt and dissect enhancer activity in cancer.
Table of Contents

Abstract................................................................................................................................. iii
List of figures ........................................................................................................................ vii
Acknowledgements ............................................................................................................. ix
Chapter 1: Introduction ........................................................................................................ 1
  Overview of thesis................................................................................................................. 2
  Gene regulation in multicellular organisms ........................................................................ 4
  Classically defined enhancers ............................................................................................. 5
  Current model of enhancer activity ..................................................................................... 7
    Transcription factors bind enhancers to drive target gene expression ......................... 8
    Chromatin structure at enhancers .................................................................................... 12
    Looping model of enhancer activity ................................................................................. 15
  Genome-wide predictions of enhancers ........................................................................... 17
    Predictions based on transcription factor binding sites ................................................. 19
    Predictions based on chromatin accessibility ................................................................. 20
    Predictions based on histone modifications .................................................................... 21
    Predictions based on multiple genome-wide measurements ....................................... 22
  Identification and characterization of super enhancers ..................................................... 23
    Defining super enhancer loci and their target genes ...................................................... 24
    Super enhancers in transcriptional control ...................................................................... 27
    Super enhancers in cancer ............................................................................................... 28
  Functional testing of enhancers genome-wide................................................................. 31
  Enhancer deregulation in cancer ....................................................................................... 33
    MYC family transcription factors .................................................................................. 34
    BET bromodomain family ............................................................................................... 39

Chapter 2: Super enhancer profiling identifies ID1 as a previously unrecognized dependency in neuroblastoma................................................................. 43
  Introduction ....................................................................................................................... 44
  Results ................................................................................................................................. 48
    Enhancer landscape of neuroblastoma reflects MYCN amplification status .................. 48
    Super enhancer profiling identifies key regulatory regions in neuroblastoma .............. 52
    Functional interrogation of super enhancer associated genes identifies ID1 as a dependency in neuroblastoma ................................................................. 57
    Indirectly targeting ID1 for degradation by small molecule inhibition of USP1 ............. 61
  Discussion .......................................................................................................................... 66
  Methods ............................................................................................................................... 68
# Chapter 3: Enhancer invasion shapes MYCN dependent transcriptional amplification in neuroblastoma

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 3: Enhancer invasion shapes MYCN dependent transcriptional amplification in neuroblastoma</td>
<td>76</td>
</tr>
<tr>
<td>Introduction</td>
<td>77</td>
</tr>
<tr>
<td>Results</td>
<td>79</td>
</tr>
<tr>
<td>Hierarchy of deregulated MYCN binding at promoters and enhancers of active genes</td>
<td>79</td>
</tr>
<tr>
<td>MYCN is an amplifier of transcription in neuroblastoma</td>
<td>85</td>
</tr>
<tr>
<td>Enhancer invasion shapes transcriptional sensitivity to MYCN perturbation in neuroblastoma</td>
<td>88</td>
</tr>
<tr>
<td>TWIST1 co-occupies enhancers with MYCN and is a novel dependency in neuroblastoma</td>
<td>96</td>
</tr>
<tr>
<td>Discussion</td>
<td>104</td>
</tr>
<tr>
<td>Methods</td>
<td>108</td>
</tr>
</tbody>
</table>

# Chapter 4: Overcoming BET bromodomain inhibitor resistance with BET degradation in triple negative breast cancer

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 4: Overcoming BET bromodomain inhibitor resistance with BET degradation in triple negative breast cancer</td>
<td>124</td>
</tr>
<tr>
<td>Introduction</td>
<td>125</td>
</tr>
<tr>
<td>Results</td>
<td>131</td>
</tr>
<tr>
<td>Targeted BET family protein degradation in TNBC</td>
<td>131</td>
</tr>
<tr>
<td>BET degradation overcomes BET inhibitor resistance in TNBC</td>
<td>133</td>
</tr>
<tr>
<td>BET degradation results in a global reduction of transcription</td>
<td>136</td>
</tr>
<tr>
<td>BET degradation overcomes intrinsic resistance to BET inhibitors in various cancer cell line models</td>
<td>140</td>
</tr>
<tr>
<td>Discussion</td>
<td>142</td>
</tr>
<tr>
<td>Methods</td>
<td>144</td>
</tr>
</tbody>
</table>

# Chapter 5: Conclusion and future directions

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 5: Conclusion and future directions</td>
<td>150</td>
</tr>
<tr>
<td>Conclusion</td>
<td>151</td>
</tr>
<tr>
<td>Future directions</td>
<td>154</td>
</tr>
<tr>
<td>Functionally defining genome-wide predictions of enhancers</td>
<td>154</td>
</tr>
<tr>
<td>Small molecule DUB inhibitors in neuroblastoma</td>
<td>156</td>
</tr>
<tr>
<td>Therapeutic implications of MYCN as a non-linear amplifier of transcription</td>
<td>157</td>
</tr>
<tr>
<td>Applicability of chemically targeted protein degradation</td>
<td>159</td>
</tr>
<tr>
<td>Pharmacologic inhibition versus pharmacological degradation</td>
<td>160</td>
</tr>
<tr>
<td>Clinical utility of BET degradation</td>
<td>162</td>
</tr>
</tbody>
</table>

# References

References

163
List of figures

Chapter 1: Introduction
Figure 1.1: Transcription factors bind enhancers at TF binding sites ...... 10
Figure 1.2: Histone tail modifications influence chromatin structure at enhancer regulatory elements ........................................... 13
Figure 1.3: Enhancer-promoter interaction is achieved by chromatin ‘looping’ to drive enhancer activity ....................................... 16
Figure 1.4: Defining super enhancer loci from ChIP-seq profiles of enhancer factors ................................................................... 25
Figure 1.5: MYC family TFs heterodimerize with MAX and bind cis regulatory elements at E-box motifs ........................................ 37

Chapter 2: Super enhancer profiling identifies ID1 as a previously unrecognized dependency in neuroblastoma
Figure 2.1: Models of MYCN amplified/non-amplified neuroblastoma ..... 49
Figure 2.2: MYCN-specific enhancer landscapes .................................... 51
Figure 2.3: Super enhancer profiling in neuroblastoma .......................... 53
Figure 2.4: MYCN-specific super enhancer landscapes .......................... 54
Figure 2.5: Super enhancer associated genes are highly expressed and neuroblastoma-specific ...................................................... 56
Figure 2.6: Functional siRNA screening of super enhancer associated genes identifies genes previously implicated in neuroblastoma ..... 58
Figure 2.7: siRNA screening of super enhancer associated genes identifies ID1 as a previously unrecognized dependency .......... 60
Figure 2.8: Small molecule inhibition of USP1 indirectly targets ID1 for degradation ....................................................................... 64
Figure 2.9: Small molecule inhibition of USP1 in MYCN-amplified neuroblastoma ................................................................. 66

Chapter 3: Enhancer invasion shapes MYCN dependent transcriptional amplification in neuroblastoma
Figure 3.1: Genome-wide occupancy of deregulated MYCN in neuroblastoma ............................................................................. 80
Figure 3.2: Hierarchy of deregulated MYCN binding at promoters and enhancers of active genes ..................................................... 83
Figure 3.3: Dynamic chromatin and transcriptional changes of direct MYCN shutdown ................................................................ 86
Figure 3.4: MYCN is an amplifier of transcription in neuroblastoma ...... 89
Figure 3.5: MYCN binding contributions at promoters and enhancers of active genes ............................................................... 91
Figure 3.6: Enhancer invasion shapes transcriptional sensitivity to MYCN perturbation in neuroblastoma ........................................ 94
Figure 3.7: Enhancer invasion defines tumor-specific MYC target gene signatures ........................................................................... 97
Figure 3.8: Enhancer regulatory core transcription factor circuitries ...... 100
Chapter 4: Overcoming BET bromodomain inhibitor resistance with BET degradation in triple negative breast cancer

Figure 4.1: Targeted BET degradation in TNBC ........................................... 132
Figure 4.2: BET degradation overcomes BET inhibitor resistance in TNBC ................................................................. 134
Figure 4.3: BET degradation results in a global reduction in transcription in the SUM149R cell line ................................................................. 137
Figure 4.4: BET degradation overcomes intrinsic resistance to BET inhibition in the NGP cell line ................................................................. 141
Acknowledgements

I would like to take this opportunity to thank, from the bottom of my heart, all of those who have given me the inspiration, support, strength, and guidance along the way. Without a doubt, this work would not have been possible without the expertise, kindness, generosity, and willingness to help from many as I carried out this work.

First, I would like to thank my thesis advisor, Dr. James Bradner. Jay welcomed me into his lab with open arms and instantly provided me with the support and flexibility to pursue any and all scientific opportunities. He taught me the value of having a vision for science and, perhaps more importantly, having the resolve to follow that vision day in and day out. I am also particularly thankful for Jay’s unwavering support in my personal life. My wife and I had two children during my time here and I am grateful for the support, understanding, and advice from one father to another through all of it.

I would like to thank the entire Bradner lab (both past and present) for all of their friendship and support. In my time in the Bradner lab, I have forged friendships for life, and for that, I am forever grateful. In particular, I would like to thank Chris Ott (‘Ott-man’), Michael McKeown (‘Mickey’), Alex Federation (‘Fed’), Charles Lin (‘Linsanity’), Jaime Reyes (‘Jaimes’), Georg Winter (‘Greg’), Matthew Lawlor (‘Lawlinator’) and Michael Erb (‘Merb’) – you made me a better scientist and I am grateful for having the opportunity to learn from you all.

I would like to thank my dissertation advisory committee: Dr. Karen Cichowski, Dr. Steve Elledge, and Dr. Kimberly Stegmaier for going above and
beyond to guide me through my studies. I would like to thank my defense committee: Dr. Karen Cichowski, Dr. Marcia Haigis, Dr. Loren Walensky, and Dr. Angela Koehler for reading this thesis and performing the defense examination.

I am grateful to several folks that have inspired, cultivated, and motivated my interest in pursuing a career in research including but not limited to Dr. Wendy Smith and Dr. Becky Rosengaus. Special thanks to Dr. Tim Heffeman who truly pushed me to new heights, ignited my interest in cancer research, and pushed me to graduate school.

Above all, I want to thank my family: my wife, Jeny, and my daughters Stella and Josephine. From the depths of my heart, I thank you for your support, inspiration, and patience through it all. To Jeny, you are as steady as they come and I couldn’t have done any of this without you. To Stella and Josie, I thank you for constantly reminding me about the benefits of being silly, and showing me what true happiness is. I want to thank my parents, Enas and Ibrahim Zeid for everything that they have done for me. You have, and continue to, support me unconditionally and I am so very grateful for all the sacrifices you have made for me over the years. I would not be here without the opportunities you have provided me. You taught me what it means to be disciplined and, more importantly, what it means to be a good person. I want to thank my brother Sharif and my sister Yasmine. Sharif, thank you for setting me straight when I needed it. Yas, thank you for always being supportive and positive.

Thank you all. I sincerely hope that this body of work makes you proud; this work would not have been possible without you.
Chapter 1: Introduction
Overview of thesis

Enhancers are noncoding DNA regulatory elements that play key roles in the control of cell-type specific gene expression programs. It is becoming increasingly clear that enhancer deregulation plays a prominent role in cancer by altering cell-type specific gene expression programs that underlie tissue-specific malignancies. This work sought to identify, characterize and disrupt cis regulatory elements in tissue-specific malignancies. These studies were powered by the pairing of evolving technologies in genome-wide measurements of chromatin structure with new computational approaches to identify and characterize enhancer regulatory elements in human cell lines and tumor tissues. In turn, this work aimed to disrupt cis regulatory elements both chemically and genetically to investigate the alternative methods of targeting tissue-specific malignancies.

The first chapter provides a background of enhancer biology as it relates to this work. This chapter covers the initial discovery of enhancer regulatory elements, details the known properties of enhancers, describes the use of genome-wide measurements to predict enhancers, and highlights the contribution of enhancer deregulation to cancer.

In chapter 2, we generated genome-wide maps of the active enhancer gene regulatory landscape to identify potential dependencies in neuroblastoma. In a panel of neuroblastoma cell lines and primograft models, we identify large cis regulatory elements (termed super enhancers) adjacent to several genes previously implicated in neuroblastoma. From this observation, we hypothesized that we could identify previously unrecognized dependencies by systematic
functional screening of super enhancer associated genes. siRNA screening of a focused super enhancer associated gene library identifies ID1 as a candidate dependency in both MYCN amplified and non-amplified neuroblastoma. A small molecule inhibitor of the deubiquitinating enzyme (DUB) USP1 indirectly promoted ID1 degradation (secondary effect) and had a potent growth inhibitory effect in a panel of neuroblastoma cell lines. Collectively, these results establish ID1 as a dependency in neuroblastoma and suggest that DUB inhibitors may provide a novel path towards therapeutic applicability.

In chapter 3, we investigate how the MYCN transcription factor occupies the genome and alters transcriptional programs to establish neuroblastoma. While MYCN amplification is the defining prognosticator of high-risk disease, there is no global MYCN occupancy data currently available. We present the first dynamic chromatin and transcriptional landscape of direct MYCN perturbation in neuroblastoma. We find that at oncogenic levels, MYCN associates with E-box (CANNTG) binding motifs in an affinity dependent manner across most active cis regulatory promoters and enhancers. MYCN shutdown results in a global reduction of histone acetylation and transcription, consistent with prior descriptions of MYC proteins as amplifiers of gene expression. However, decreases in gene expression are not uniform and instead correlate strongly with MYCN invasion of proximal enhancers suggesting a role for these tissue specific elements in predating MYCN responsive “target” genes. At these invaded enhancers, we identify the lineage specific bHLH TWIST1 as a key collaborator and synthetic lethality of oncogenic MYCN. These data suggest that MYCN
enhancer invasion helps shape transcriptional amplification in neuroblastoma to promote tumorigenesis.

In the fourth chapter, we turn our attention to triple negative breast cancer to understand mechanisms of resistance to the bromodomain and extra-terminal (BET) family of proteins. The pharmacologic inhibition of BET family proteins provides a strategy to inhibit enhancer activity. The efficacy of disrupting enhancer activity via BET inhibition in several cancers has prompted efforts to predict mechanisms of resistance. We use a newly developed class of small molecules to pharmacologically induce targeted degradation of the BET family in triple negative breast cancer and show that BET family degradation effectively overcomes BET inhibitor resistance. We explore the fundamentally different molecular response of BET degradation versus BET inhibition and subsequently show that BET degradation also overcomes other forms of resistance in different cancer cell line models.

Lastly, the fifth chapter provides concluding remarks on this work and provides perspective on future directions.

**Gene regulation in multicellular organisms**

In multicellular organisms, gene regulation is fundamental to producing a variety of diverse cell types that collectively make up a complex organism. Individual cell types are characterized by unique patterns of gene expression that underscore distinct cell morphologies and function (Davidson and Erwin, 2006). Given that these truly diverse patterns of gene expression are derived from a single genome, proper spatiotemporal gene expression relies on several
integrated layers of regulation. For instance, transcription is coordinated by the presence of cis regulatory elements in the genome. Cis regulatory elements are noncoding DNA regulatory elements such as core promoters and enhancers surrounding gene coding regions. While core promoters sit in close proximity to transcriptional start sites, enhancers are located at greater distances. At the core of transcriptional regulation, RNA polymerase II (Pol II) and its machinery assembles at the transcriptional start site to transcribe genomic DNA into RNA. In addition to the presence of Pol II at core promoters, it has become increasingly clear that enhancers are critical to the control of target gene expression. Since enhancers are often located at greater distances from the core promoter, these cis regulatory elements offer an additional dimension of transcriptional regulation to maximize the complexity from a single genome.

Classically defined enhancers

In 1981, the term ‘enhancer’ was coined following the observation that a DNA element from Simian virus (SV40) could drive exogenous gene expression of a cloned rabbit β-globin and T-antigen reporter gene (Banerji et al., 1981; Moreau et al., 1981). It had previously been observed that transcription of early genes of SV40 were dependent on a tandem 72 basepair (bp) repeat sequence approximately 200 bp upstream of the transcriptional start site (Benoist and Chambon, 1981; Gruss et al., 1981). Schaffner and colleagues went on to show that this 72 bp tandem repeat could enhance the expression of the β-globin gene in HeLa cells (Banerji et al., 1981). This so-called ‘viral enhancer’ was demonstrated to increase transcription at very large distances from the
transcriptional start site and function independently of its orientation both upstream and downstream from the β-globin promoter (Banerji et al., 1981).

Subsequent studies reported the identification of novel enhancer elements in several other animal viruses that shared similar properties (Hansen and Sharp, 1983; Schirm et al., 1985; Spandidos and Wilkie, 1983; Villiers et al., 1982; Weber et al., 1984). These studies again took advantage of a functional readout by linking enhancer sequences to various gene reporter systems and measuring changes in reporter gene expression. It quickly became clear that enhancer activity in these systems were dependent in part on host-cell preferences (Spandidos and Wilkie, 1983; Villiers et al., 1982). For example, enhancers identified in bovine papillomavirus and murine sarcoma virus showed a marked preference for bovine cells and mouse fibroblasts respectively (Spandidos and Wilkie, 1983). These findings suggested that host-cell regulatory factors are a determinant of enhancer activity and provided one of the first lines of evidence for the cell-type specifying nature of enhancers.

Shortly after the discovery of viral enhancers, similar sequences were identified endogenously within mammalian genomes initially in the immunoglobulin heavy chain locus (IgH) (Banerji et al., 1983; Gillies et al., 1983; Neuberger, 1983). The IgH locus consists of several highly homologous sets of genes of different types that purposely rearrange in developing B cells to create a functional Ig heavy chain gene (Cory and Adams, 1980; Hozumi and Tonegawa, 1976; Rabbitts et al., 1980). In myeloma cells, it was observed that the transcriptional outputs of these functionally rearranged classes of genes were
magnitudes higher and that certain regions of DNA were consistently preserved (Davis et al., 1980; Mather and Perry, 1981; Sakano et al., 1981). With the discovery of enhancer elements in viruses, these preserved regions of DNA were suspected and identified as enhancer elements with similar properties (Banerji et al., 1983; Gillies et al., 1983; Neuberger, 1983). After this discovery, several endogenous enhancers were identified and characterized based primarily on functional consequences on target gene expression.

Collectively, these studies initially in viruses and then endogenously established an enhancer as a functionally distinct element from that of promoters capable of enhancing target gene expression. These cis regulatory elements can drive target gene expression from very long distances and function independently of orientation. These studies established the basic properties of enhancer elements that have served as the foundation of our understanding of their role in transcription.

**Current model of enhancer activity**

Since the initial discovery and characterization of enhancer elements, further mechanistic insights have collectively shaped an evolving model of enhancer activity. In particular, our current understanding of enhancer activity is predicated on the critical role of transcription factors (TFs) at enhancer elements. It has become increasingly clear that TFs bind enhancer elements and recruit additional factors to drive target gene expression. In addition, enhancer architecture and activity is also fundamentally correlated with defining properties of chromatin state. Enhancers regions are generally found in regions of open
chromatin accessible by transcription factors and allow a favorable chromatin confirmation (looping) for enhancer activity. The prevailing model of enhancer activity is described in detail below and serves as the basis of understanding for this dissertation.

*Transcription factors bind enhancers to drive target gene expression*

Shortly after the discovery of enhancers, it was shown that specific transcription factors bind to specific short DNA motifs within enhancers to drive their activity (Lee et al., 1987). In a novel finding, Lee and colleagues showed that the transcription factor AP1 binds to an enhancer of the human metallothionein promoter as well as well as the SV40 enhancer that resulted in increased gene expression (Lee et al., 1987). These initial findings established the importance of sequence specificity within these cis regulatory elements as they correspond to transcription factor binding to collectively enable enhancer activity (Shlyueva et al., 2014; Spitz and Furlong, 2012).

Transcription factors typically bind small 6-12 basepair long degenerate DNA sequences with relatively low sequence specificity. Enhancers contain clusters of several different transcription factor binding sites for both activating and repressing TFs (Figure 1.1A). The presence of multiple transcription factor binding motifs suggests that enhancer activity is controlled by the interplay of multiple transcription factors rather than the simple affinity of a single transcription factor for its respective binding motif. Indeed, there are several examples in which the combinatorial binding of transcription factors at enhancers allows for the precise and dynamic control of transcriptional output (Halfon et al.,
For example, in *Drosophila melanogaster* RAS pathway specificity is determined in part by the combinatorial binding of several transcription factors (Mad, dTCF, Tinman, Twist) at an enhancer that drives the transcription factor Even skipped (Eve) (Halfon et al., 2000). The combination of these transcription factors binding to the Eve enhancer is critical for proper transcription. Interestingly, enhancer mediated Eve transcription is a critical determinant of cell fate specification in the somatic mesoderm. Thus, enhancer elements serve as transcription factor binding hubs in which the combinatorial binding of multiple TFs achieves an increasingly complex form of tissue-specific transcriptional regulation.

In addition to the combinatorial binding of TFs at enhancers, TFs also interact directly and indirectly (termed cooperative binding) with each other to further facilitate activity. The mechanisms of TF cooperativity are intrinsically dependent on the location of binding motifs within the enhancer. The positions of transcription factor binding motifs within a given enhancer dictate the spatial arrangement of TF binding and thus the potential for interaction. The importance of the arrangement of transcription factor binding motifs within an enhancer is exemplified in the well-characterized interferon-β (IFN-β) enhancer. The IFN-β is a wonderfully complex example of how the specific overlap of individual TF binding motifs is required to promote cooperative binding and constitute functional enhancer activity (Merika and Thanos, 2001; Panne, 2008; Panne et al., 2007; Thanos and Maniatis, 1995).
The precise organization of the IFN-β enhancer also exemplifies cases of indirect cooperative binding between transcription factors. In some instances, transcription factors may bind their respective motifs within an enhancer and recruit a common cofactor rather than directly interacting (Figure 1.1B). At the IFN-β enhancer, multiple TFs (including p65, IRF1, and ATF2/c-Jun) interact with

**Figure 1.1: Transcription factors bind enhancers at TF binding sites.** (A) Multiple TFs (TF$_A$ and TF$_B$) simultaneously bind their respective transcription factor binding motif within an enhancer (red bar) in open regions of chromatin. (B) TFs bind their respective TF binding motifs and recruit a common cofactor that then recruits additional complex members. This can increase the affinity of each TF for their binding motif and stabilize the enhancer interaction. (C) A single TF (TF$_B$) recruits chromatin remodeling complexes to expose the TF binding motif for another TF (TF$_A$). Figure 1A adapted from Shylyueva et al., 2014, Figure 1B and 1C adapted from Spitz and Furlong, 2012.
the chromatin coactivator CBP/p300 to stabilize TF binding and stabilize the enhancer complex (Merika et al., 1998). Thus, TF binding at enhancers take advantage of the organization of their respective binding motifs to stabilize the affinity to the enhancer region.

Other forms of TF cooperativity include TF recruitment of chromatin remodeling complexes to trigger nucleosome repositioning and uncover binding sites for additional collaborating TFs within an enhancer (Figure 1.1C). For example, in murine mammary epithelial line, the transcription factor AP1 binds its DNA motif and recruits chromatin remodelers to expose a binding site for a collaborating transcription factor GR (Almer et al., 1986). These transcription factors, termed, pioneering factors were first identified in yeast and have since been characterized in several instances (Almer et al., 1986; Biddie et al., 2011; de la Serna et al., 2005; Hoogenkamp et al., 2009; Liber et al., 2010; McManus et al., 2011; Xu et al., 2009). It is important to note that pioneering factors alone are not generally sufficient to form an activating complex at enhancers. This suggests that enhancer activity is purposely linked to the presence of multiple transcription factors to allow for dynamic control of target gene expression.

The progression in our understanding of the role of transcription factor binding at enhancers has been critical for our understanding of enhancer activity. Collectively, these insights have informed on the structure and function of these cis regulatory elements in driving target gene expression.
Chromatin structure at enhancers

In order for transcription factors to bind at enhancers and mediate enhancer activity, their respective DNA binding motifs must be accessible. Indeed, active enhancers are generally found in regions of open chromatin that lack nucleosomes and allow for transcription factor accessibility. Thus, chromatin structure provides an additional layer of regulation of enhancer activity by determining the availability of enhancer regulatory regions to transcription factors. Chromatin is a dynamic structure that functionally packs DNA into the cell nucleus. It has become increasingly clear that chromatin state fluidly changes during development and in response to external stimuli (Mikkelsen et al., 2007). Moreover, chromatin states also reflects the particular gene expression programs associated with different cell types (Ernst et al., 2011).

The fundamental unit of chromatin is the nucleosome: a histone octamer containing two copies of four histone proteins (H3, H4, H2A, H2B) in which 147 base pairs of DNA is wrapped around (Kornberg, 1977; Luger et al., 1997). The overall positioning of specific nucleosomes are the primary determinant of DNA accessibility (Han and Grunstein, 1988; Knezetic and Luse, 1986). One key aspect regulating nucleosome positioning and thus overall chromatin state are the presence of covalent modifications on protruding histone tails. There are several types of modifications including acetylation, methylation, phosphorylation, ubiquitylation, and sumoylation. Each of these covalent chemical modifications carry a chemical charge (Figure 1.2A) that can influence nucleosomal contacts and thus changes overall chromatin structure (Fischle et al., 2005;
Krishnamoorthy et al., 2006; Shogren-Knaak et al., 2006). For example, the acetylation of histone tail lysine residues results in the loss of positive charge and thus loosening of DNA bound to chromatin (Aalfs and Kingston, 2000). These various histone tail modifications play a critical role in shaping chromatin and

![Figure 1.2: Histone tail modifications influence chromatin structure at enhancer regulatory elements. (A). Chemical structures of select examples of histone tail modifications. The modification is highlighted and red and labeled below. The net charge (or lack thereof) is denoted. (B) Model depiction of the addition of histone modifications (acetylation and mono-methylation) leading to an active denovo enhancer (active enhancers can also be formed from primed or poised enhancers (not pictured). HATs enzymatically deposit acetyl groups while HDACs enzymatically remove them. Bromodomain containing proteins (depicted in orange) recognize acetyl-lysine groups on chromatin. Figure 1B adapted from Shylyueva et al., 2014.]
have been associated with several respective processes including transcription, DNA replication, and DNA repair. As will be detailed in subsequent sections, genome-wide measurements of these histone modifications have revealed particular patterns of enrichment that correspond to active regions of chromatin (promoters and enhancers) or silent regions of chromatin.

In addition to the effects on nucleosomal contacts, histone tail modifications also serve to recruit chromatin factors with enzymatic activity to further modify chromatin. Chromatin factors termed ‘writers’ deposit histone modifications, while chromatin factors termed ‘erasers’ remove these modifications. In turn, chromatin ‘reader’ proteins contain binding domains that recognize each of these specific histone marks. Collectively, the combination of histone modifications as regulated by these three classes of proteins (writers, erasers, and readers) form what has been termed the ‘histone code’. This histone code influences chromatin state and contributes to a global gene expression program of a cell (Jenuwein and Allis, 2001; Strahl and Allis, 2000). For example and of particular relevance to this dissertation, acetylation of lysine residues on histone H3 play a critical role in the regulation of enhancer elements. Histone acetylation has been particularly well studied and is fundamentally linked to open chromatin and thus gene activation (Aalfs and Kingston, 2000; Shogren-Knaak et al., 2006). Histone acetyltransferases (HATs) deposit acetyl groups, histone deacetylases (HDACs) remove these acetyl groups, and bromodomain containing proteins recognize acetyl groups (Figure 1.2B). Specifically, there have been several key insights into the role of several of these HATs
(CBP/p300), HDACs (HDAC1-8), and bromodomain (BRD) containing proteins at enhancers. It is also important to note that mono-methylation is also found to mark enhancers and can further characterize the state (active, poised, or primed) of a given cis regulatory element (Heintzman et al., 2007).

Collectively, the deposition, removal, and subsequent recruitment to histone modifications play a critical role in enhancer activity. In order for transcription factors to bind their cognate binding sites, the chromatin state must allow for accessibility. Thus, TFs and histone modifications serve as the basis for regulation of enhancer activity and work collaboratively in layers to dictate enhancer activity.

Looping model of enhancer activity

As has been described earlier, transcription factors bind enhancer elements to drive transcription of target genes. Ultimately, productive transcription is predicated on the recruitment and elongation of Pol II from the transcriptional start site. Since enhancer elements can often be located at great distances from its target gene promoter, it is thought that enhancers are brought into close proximity of their target core promoters by looping (Dorsett, 1999; Dorsett and Merkenschlager, 2013). This looping mechanism creates the physical/spatial relationship to allow enhancers to drive their target genes (Figure 1.3). At the heart of this looping interaction, transcription factors and chromatin factors described in the previous sections (driven in part by recruitment via histone modifications) play a key role (Lee and Young, 2013). In addition to recruiting chromatin cofactors, transcription factors also bind to components of
the mediator complex in order to contribute to chromatin looping. For example, in embryonic stem cells, the ELK1 transcription factor binds to a Sur2 subunit of the mediator complex in order to drive gene expression in response to ERK signaling (Stevens et al., 2002). In addition to the central role transcription factors play in the looping, chromatin cofactors such as BRD4 (bromodomain containing protein, 4) that bind to acetylated lysines also bind to the mediator complex to help mediate the promoter-enhancer loop (Wu and Chiang, 2007; Wu et al., 2003). The development of technologies to measure the three-dimensional (3D)
conformation of chromosomes has helped in our understanding of promoter-enhancer looping (Dekker, 2008; Gibcus and Dekker, 2013; Liber et al., 2010). For example, in 2009 a technique known as Hi-C was introduced that adapted chromatin confirmation capture (3C) for genome-wide application (Liber et al., 2010). These techniques are generally carried out by first crosslinking cells and then digesting DNA with appropriate overhanging sequence. Then, digested fragments are ligated together to favor products that were originally in close proximity in their native confirmation. In addition, this loop formed between enhancers and core promoters is further stabilized by key factors including the cohesin complex and associated loading factors such as NIPBL (Nipped-B). The cohesin complex is cylindrically shaped and encircles two nucleosomes to help stabilize the promoter-enhancer interaction (Dorsett and Merkenschlager, 2013; Hadjur et al., 2009; Parelho et al., 2008; Weth and Renkawitz, 2011). Thus, transcription factors, chromatin factors/modifying enzymes, and histone modifications all collectively contribute to the physical promoter-enhancer interaction to enable Pol II to carry out active transcription.

**Genome-wide predictions of enhancers**

The rapidly evolving technologies of next-generation sequencing have transformed our ability to identify and categorize enhancer elements genome-wide in cell lines and tissues. These technologies apply our understanding of enhancers (transcription factor binding, chromatin architecture, histone modifications etc.) described in the previous section to systematically identify cis regulatory elements.
In particular, the development of chromatin immunoprecipitation followed by next-generation sequencing (ChIP-seq) has been instrumental to the genome-wide identification of enhancers and a critical technique utilized in this work. In ChIP-seq, cell lines and/or tissues are crosslinked to preserve a given chromatin state and covalently link chromatin bound factors. The chromatin is then sheared via sonication and an immunoprecipitation using antibodies specific to chromatin bound factors such as TFs and histone modifications are used to pull down corresponding regions of interest. Following immunoprecipitation, chromatin bound complexes are washed, crosslinks reversed, and the corresponding DNA bound regions of the genome are purified. The DNA is then prepared for next-generation sequencing in which immunoprecipitated DNA is subject to adapter ligation, amplification and size selection (collectively termed ChIP-seq library preparation). ChIP-seq libraries are then subject to massively parallel sequencing. Sequencing reads are computationally aligned back to reference genomes and regions of enrichment over background sheared genomic DNA is calculated. Collectively, these computational analyses are used to infer the binding and localization of a given chromatin bound factor of interest genome-wide.

The following sections describe three methodologies that take advantage of next-generation sequencing to predict enhancers genome-wide. The advantages/disadvantages are highlighted below and are of particular relevance to this dissertation.
Predictions based on transcription factor binding sites

Given that transcription factors bind their respective motifs within an enhancer, it is perhaps not surprising that genome-wide measurements of TF localization have been used to predict cis regulatory element loci. The resolution achieved with ChIP-seq studies with transcription factors allows for high degree of specificity in the location of TF binding throughout the genome. Further, the underlying sequence of the enriched TF region can be examined for the presence of the TF binding motif. For example, the combined ChIP-seq profiles of five different TFs in a cardiomyocyte cell line were used to identify cardiac-specific enhancers (He et al., 2011). Given the combinatorial/cooperative nature of TF binding at enhancers, this work overlaps five different transcription factor ChIP-seq profiles in an effort to identify truly functional cis regulatory elements.

While transcription factor ChIP-seq is a reasonable approach to identify enhancers, there are some critical challenges. The use of transcription factor ChIP-seq to identify enhancers is often limited by the technical challenges associated with quality ChIP-grade TF antibodies. Unlike histone marks, transcription factors can often be more difficult to immunoprecipitate due to antibody quality and low abundance. Furthermore, it has become increasingly clear that transcription factor binding in a cis regulatory element does not necessarily delineate the presence of an enhancer (Fisher et al., 2012; Li et al., 2008). This reflects the combinatorial/cooperative mode of transcription factor binding as it relates to enhancer activity. In addition, transcription factors have a
general affinity for DNA and this transient binding can often increase the noise in a given ChIP-seq profile.

**Predictions based on chromatin accessibility**

A number of techniques that take advantage of open chromatin architecture have been coupled with next-generation sequencing to allow for genome-wide predictions of enhancer elements. Since enhancers are generally devoid of nucleosomes, enzymes that nonspecifically cleave DNA can relate a picture of accessible regions within the genome. Here, enzymes such as DNase I and micrococcal nuclease have been coupled to deep sequencing (termed DNase-seq and MNase-seq respectively) to produce high-resolution maps of open chromatin (Boyle et al., 2008; Schones et al., 2008). In comparison to TF binding sites, these techniques provide a much broader body of information but can be especially useful when relevant transcription factors are not known. Since the development of DNase-seq and MNase-seq, other methodologies based on similar principles have been developed including the more recent ATAC-seq (assay for transposase-accessible chromatin) (Buenrostro et al., 2013). ATAC-seq makes use of the hyperactive transposase Tn5 to integrate adapters into open regions of chromatin that can later be amplified. ATAC-seq is performed without fixation on fresh material and has very low input requirements. All of these techniques allow for high-resolution maps and can even delineate transcription factor binding sites based on the protection from cutting/transposing of a TF in a region of open chromatin.
At times, using chromatin accessibility to predict enhancers can be challenging because the measurement only reports generally on open regions of chromatin. It does not directly take advantage of any specific characteristics of enhancers that may delineate them from other cis regulatory elements such as promoters (though promoters do generally have higher signal in these techniques). In addition, these techniques are not particularly well-suited to differentiating functional enhancers and thus it can be difficult to assign relevance to identified cis regulatory elements.

_Predictions based on histone modifications_  

The genome-wide measurements of histone modifications have been instrumental in the identification of enhancers and are the most relevant predictive technique to this dissertation. As was touched on in previous sections, particular histone modifications are enriched and thus mark different chromatin features throughout the genome. For example this histone H3 lysine 9 tri-methylation (H3K9me3) labels transcriptionally silent heterochromatic regions, while histone H3 lysine 27 acetylation (H3K27Ac) typically marks active enhancers and promoters. Active gene promoters are marked with histone H3 lysine 4 tri-methylation (H3K4me3), while repressed regions are marked by histone H3 lysine 27 tri-methylation (H3K27me3) (Bernstein et al., 2006). A combination of histone H3 lysine 4 mono-methylation (H3K4me3) and histone H3K27Ac can be used to identify and classify enhancers (Heintzman et al., 2009; Heintzman et al., 2007). The H3K4me1 modification is used to identify active, poised, and primed enhancers (Heinz et al., 2015). In addition to H3K4me1,
active enhancers are specifically marked with H3K27Ac and thus H3K27Ac is often the mark of choice in genome-wide studies (Creighton et al., 2010; Rada-Iglesias et al., 2011). Collectively, ChIP-seq profiles of these histone marks allow for generation of genome-wide maps from which cis regulatory elements can be inferred. This is perhaps best illustrated with the H3K27Ac mark to identify enhancer elements across a number of tissues and cell lines. As part of a much larger landmark effort, the Encyclopedia of DNA Elements (ENCODE) Consortium mapped enhancer elements across a wide range of cell types (Dunham et al., 2012).

Similar to transcription factor and chromatin accessibility based predictions, enhancer predictions based on histone modifications do not guarantee the presence of a functional enhancer. In addition, histone modifications do not perfectly correlate with enhancer activity and thus there is an element of false-positive identification (Arnold et al., 2013). Moreover, there are continuing disagreements about which histone modifications are best used to predict enhancers (or other regulatory elements).

Predictions based on multiple genome-wide measurements

While each of the techniques described previously can be used to identify enhancers genome-wide, it is becoming increasingly clear that combining multiple measurements can be advantageous. Since each of these methodologies relies on a different aspect of enhancer biology (TF binding, open chromatin, and histone modifications), combining the measurements can allow for a high-confidence prediction that meets more than one criteria for an
enhancer. This is particularly exemplified in the large scale effort of the Roadmap Epigenomics Mapping Consortium that followed after the work of the ENCODE Consortium. In one aspect of this work, a number of histone mark ChIP-seq profiles, chromatin accessibility measurements, and transcription factor ChIP-seq profiles were integrated to identify cis regulatory elements (Ernst et al., 2011). This allows for a more accurate prediction of functional enhancer elements based on the presence of several key components. While enhancer activity is fundamentally measured by the increase in transcriptional output of target genes, these methods provide a predictive framework to identify functional enhancers from genome-wide measurements.

**Identification and characterization of super enhancers**

The ability to make genome-wide predictions of enhancer elements has resulted in the identification of vast numbers of cis regulatory elements in a given cell type. For example, enhancer predictions based on histone modifications can identify on the order of hundreds of thousands of putative cis regulatory elements (Dunham et al., 2012; Lin et al., 2016). Given the predictive nature of this approach, it is highly unlikely that all of these enhancers are functional. As has been discussed in previous sections, a functional enhancer is dependent on an integrated combination of characteristics such as multiple transcription factor binding events. In recent years, it has become increasingly clear that approaches taken to further dissect genome-wide predictions of cis regulatory elements can yield key insights into enhancer mediated transcriptional control. For example, cis regulatory elements have been further characterized into large regions (often
containing multiple enhancers) that drive expression of linked genes, termed locus control regions (LCRs) (Li et al., 2002). Similarly, the identification and characterization of a subset of enhancers with exceptionally high levels of enhancer factors, termed Super Enhancers (SEs) provide a class of enhancers with particularly important functions (Lovén et al., 2013; Whyte et al., 2013). The definition and characterization of SEs are described below and are of particular relevance to this work.

**Defining super enhancer loci and their target genes**

In 2013, a collaborative effort between the Young and Bradner laboratories described a subset of enhancers that had a disproportionately high load of enhancer factors, termed super enhancers (Hnisz et al., 2013; Lovén et al., 2013; Whyte et al., 2013). These SEs were identified by rank ordering ChIP-seq profiles of various enhancer associated factors such as cell-type specific master transcription factors, enhancer-specific chromatin cofactors, and enhancer-specific histone modifications (Whyte et al., 2013). To identify SEs, ChIP-seq signal of a given enhancer associated feature is calculated at all enhancers genome-wide. ChIP-seq signal within these enhancers is then stitched together within a given distance (Figure 1.4A). Initially, this stitching window was fixed at 12.5 kilobases but other approaches using variable stitching distances have since been utilized that allow for the greatest amount of enriched peak space using the fewest discrete regions. After defining a set of stitched enhancer regions, the ChIP-seq signal is calculated in each region and plotted in increasing order (x-axis) versus the ChIP-seq signal (Figure 1.4B). These plots
result in a clear geometrically defined inflection point revealing a subset of regions with exceptionally high signal. SEs are defined as regions above this inflection point (Figure 1.4B). Thus, SEs consists of clusters of enhancers that collectively have exceptionally high signal of a given enhancer-factor. This class of enhancers differs from typical enhancers in size as well as enhancer factor signal density.

Since the initial description of super enhancer loci, a number of enhancer-associated factors have been used to identify super enhancers. All of these factors play a key role in mediating enhancer activity (as described in previous sections). For example, the master transcription factors Oct4, Sox2, and Nanog were used to identify super enhancers in embryonic stem cells (Whyte et al., 2013). Alternatively, enhancer factors such as Mediator complex members (Med1) and BRD4 can also be used to identify super enhancers (Lovén et al.,

**Figure 1.4: Defining super enhancer loci from ChIP-seq profiles of enhancer factors.** (A) Schematic illustrating the stitching of ChIP-seq reads within a given enhancer. The stitching distance can be fixed or variable. Super enhancer locus is shown as a red horizontal bar corresponding the illustrated red ChIP-seq enriched peaks. (B) Exemplary super enhancer plot in which stitched enhancer regions are plotted in an increasing order based on signal (x-axis) versus signal (y-axis). The inflection point used to determine a threshold for being categorized as a super enhancer is shown.
Lastly, histone modifications enriched at enhancers such as H3K27Ac can also be used to identify super enhancers (Chapuy et al., 2013). Since each of these enhancer factors play a key role in enhancer activity, it is perhaps not surprising that their genome-wide enrichment can be used to identify super enhancer loci. In turn, there is a high-degree of overlap in super enhancer loci when different enhancer factors are used particularly at high ranked super enhancers (Lovén et al., 2013; Whyte et al., 2013).

These super enhancer loci drive high expression of their target genes and underlie a given cell's global transcriptional program. A number of different approaches have been taken to predict super enhancer driven target genes genome-wide. Since super enhancers can act at great distances, it can be challenging to assign a target gene based on ChIP-seq enrichment signal. These approaches are predicated on the prediction that super enhancers drive high expression of their target genes. In one such strategy, a target gene window of 50-100 kilobases in either direction of a given super enhancer locus is set and the coding genes within this region are considered. These approaches make use of additional measurements or interpretations to try to infer the transcriptional status of the genes within a given target window. In some instances, studies have relied on paired genome-wide measurements of mRNA transcript levels to predict super enhancer target genes (Lin et al., 2016). Within a given super enhancer target gene window, the gene with the highest transcriptional output is predicted as the primary super enhancer associated gene. Alternatively, super enhancer profiling studies that make use of H3K27Ac ChIP-seq have used
H3K27Ac signal at promoters to infer the transcriptional activity of a given gene (Brown et al., 2014). The predictions made based on these types of predictive approaches have been relatively successful in properly assigning super enhancer associated genes. Indeed, it has been show that super enhancer associated genes on average of higher gene expression presumably due to the presence of a super enhancer (Lovén et al., 2013). Ultimately, functional studies such as the gene reporter activity assays used in the initial discovery of enhancers (previous sections) are required to clearly identify super enhancer target genes. Indeed, recent studies have made use of luciferase reporter assays to evaluate the contribution of different parts of super enhancers in driving target gene expression (Hnisz et al., 2015).

*Super enhancers in transcriptional control*

There is a growing body of evidence that super enhancers are found adjacent to genes involved in cell identity, cell function, and genes implicated in human cancer. Super enhancers have been found to play a key role in cancer by driving the expression of oncogenes and other context-specific gene dependencies. Thus, super enhancer profiling can provide genome-wide level information about the global transcriptional program of a given normal or malignant state and reveal key mechanistic insights.

It has become increasingly clear that SEs drive expression of key regulators of cell identity such as master transcription factors. In embryonic stem cells, it was shown that the master transcription factors Oct4, Sox2, and Nanog collectively occupy super enhancers at key genes involved in control of the
pluripotent state (Whyte et al., 2013). Furthermore, the genes encoding this enhancer bound master transcription factors are themselves driven by super enhancers. The presence of multiple super enhancer driven master transcription factors was recently shown to constitute an enhancer driven core regulatory circuit (Saint-André et al., 2016). The identification of core regulatory circuits can be especially useful in inferring cell of origin in unknown tissues as was recently reported for a subgroup of pediatric medulloblastoma (Lin et al., 2016). Thus, these core regulatory circuits reflect that super enhancers drive master transcription factors that mediate cell identity.

Super enhancers are also found to drive target genes involved in cell-type specific functions. This is particularly well illustrated in a compendium of 86 human cell lines (normal and cancer) in which super enhancers were identified using H3K27Ac ChIP-seq (Hnisz et al., 2013). It was shown that the super enhancer associated genes for each cell type reflected a given cell type function. For example, in hematologic cells super enhancer associated genes were significantly enriched for genes involved in T-cell activation, lymphocyte activation, and leukocyte activation (Hnisz et al., 2013). The fact that super enhancers drive genes associated with cell function reflect that SEs influence the cell-type specific transcriptional program.

Super enhancers in cancer

Super enhancers have also been found at key genes involved in human malignancies across many tumor types (Hnisz et al., 2013). The initial characterization of the role of super enhancers in cancer in part came through
the lens of evaluating the effects of small molecule inhibition of the enhancer cofactor bromodomain containing protein 4 (BRD4). BRD4 is a member of the bromodomain and extra terminal domain (BET) family and contains an acetyl-lysine recognizing bromodomain. BRD4 was first identified as an interaction partner with the murine mediator complex and is a key facilitator of active transcription (Jiang and Veschambre, 1998; LeRoy et al., 2008; Rahman et al., 2011). As has been discussed in the previous section, genome-wide localization measurements of BRD4 can be used to define super enhancers. In Multiple Myeloma cells, BRD4-defined super enhancers were positioned adjacent to key genes involved in the hematologic malignancy (Lovén et al., 2013). Moreover, treatment with the BET family small molecule inhibitor JQ1 caused a preferential loss of BRD4 at super enhancers that in turn resulted in a selective loss of transcription at associated genes. The most striking super enhancer associated was the MYC oncogene. In Multiple Myeloma, the MYC oncogene is translocated into the control of the IgH enhancer, which is categorized as a super enhancer (Shou et al., 2000). In addition, super enhancers also drove target genes previously implicated in Multiple Myeloma such as the plasma transcription factors IRF4 and XBP1. Thus, super enhancers drive key oncogenes as well as other key genes associated with a given malignancy, pointing to their importance to an aberrant transcriptional program.

The observation that super enhancer associated genes played key roles in human cancers led to studies geared at using SE profiling to identify previously unrecognized dependencies. In recent work from our lab, performing super...
enhancer analysis in diffuse large B-cell lymphoma validated this approach. The gene regulatory factor OCA-B was identified as a top ranked super enhancer associated gene and was subsequently shown to be a putative dependency (Chapuy et al., 2013). Thus, super enhancer profiling offers an alternative approach to identifying dependencies that may not be readily identified by looking at gene encoding mutations or gene transcript overexpression.

The importance of super enhancer loci have been further highlighted in recent studies observing rearrangements that result in aberrant SE mediated regulation (Drier et al., 2016; Gröschel et al., 2014; Northcott et al., 2014). In Acute Myeloid Leukemia (AML), it was demonstrated that chromosome 3q rearrangements relocate a GATA2 transcription factor SE into proximity of the proto-oncogene EVI1 to drive aberrant expression (Gröschel et al., 2014). In another such example, the proto-oncogenes GFI1 and GFI2 (Growth Factor Independent) can be translocated into the control of SEs (both distal and proximal) to drive expression in two subgroups of pediatric medulloblastoma (Northcott et al., 2014). Thus, these two examples highlight that super enhancers can play key roles in tumorigenesis by aberrantly driving proto-oncogene expression upon rearrangements of the SE locus or a coding region into proximity leading to misguided enhancer activity.

In addition to rearrangements involving SEs, it has been observed that underlying sequence heterogeneity or mutation can lead to SE deregulation in cancer (Hnisz et al., 2013; Oldridge et al., 2015). For example, a single nucleotide polymorphism (SNP) highly associated with oncogenic addiction to the
LMO1 (Lim domain only 1) transcriptional co-regulator was found to reside in a super enhancer. This SNP is housed in a transcription factor binding motif for the neuroblastoma master transcription factor GATA3 causing differential binding in the LMO1 super enhancer. Like enhancers, SE activity is predicated on the binding of transcription factors and sequence variation leading to differential TF binding can contribute to a malignant phenotype.

Since their initial description, the definition of super enhancers and their associated target genes has been refined as more studies seek to understand the role of SEs in a particular context. In particular, there is a growing body of literature describing the contribution of super enhancers to a malignant phenotype. Super enhancers comprise a subset of enhancers with exceptionally high amounts of transcription factor binding, enhancer cofactors, and corresponding histone modifications. The classical properties of enhancers (as described previously) have helped to guide our understanding of the importance of super enhancers and their role in cancer is of particular relevance to this dissertation.

**Functional testing of enhancers genome-wide**

The use of genome-wide localization measurements to identify on the order of thousands of putative enhancer elements comes with the challenge of delineating functionality. In this sense, genome-wide based identification of enhancers has strayed from the classical definition that was rooted in functional consequences on target gene transcription. A number of techniques have been developed to meet this challenge and assess enhancer functionality on a
genome-wide caliber level. In particular, STARR-seq (self-transcribing active regulatory region sequencing) and the potential of CRISPR-(clustered regulary interspaced short palindromic repeat) Cas9 based approaches to assess enhancer activity are described as two prominent examples.

STARR-seq takes advantage of the ability of enhancers to act irrespective of their orientation or position with respect to the transcriptional start site. Here, candidate enhancer loci are cloned in between a minimal promoter and the reporter gene (Arnold et al., 2013). Thus, active enhancers increase their own transcription that can be measured by the abundance of their transcripts within RNA. This technique is particularly advantageous because measurements do not rely on exogenous barcodes and can handle millions of enhancers of varying sizes.

The fast-moving CRISPR-Cas9 technology has broad applicability to the genome-wide assessment of enhancer activity. For example, CRISPR-Cas9 allow for targeted genome-editing within a given enhancer with the only requirement being the presence of a protospacer-adjacent motif (PAM) (Hsu et al., 2014; Nishimasu et al., 2014). The application of this technology to understanding enhancer activity was recently illustrated in a study using CRISPR-Cas9 to understand how particular regions within a SE (termed constituents) contribute to overall SE activity (Hnisz et al., 2015). In murine embryonic stem cells, SE constituents were targeted for deletion via the CRISPR-Cas9 system and the effects on target gene expression were measured. The prevalence of genome-wide CRISPR-Cas9 screening strategies...
suggests that genome-wide CRISPR-Cas9 studies targeting all identified enhancers followed up by transcriptional profiling would offer a revolutionary way to assess genome-wide enhancer activity.

**Enhancer deregulation in cancer**

It is perhaps not surprising that enhancer deregulation plays a key role in cancer given the contribution of enhancer activity to maintaining a global transcriptional program. Enhancer deregulation manifests in several ways (some of which have been touched upon in previous sections) including: underlying sequence variation/mutation and aberrant transcription factor binding. These varying manifestations of enhancer deregulation are fundamentally interrelated and often a combination of these events culminates in downstream enhancer deregulation. It is also important to note that these various types of enhancer deregulation can often be part of a series of events with one aberration leading to another. For example, an underlying mutation in a given enhancer can lead to aberrant transcription factor binding and subsequent recruitment of chromatin remodeling enzymes that deposit misplaced histone tail modifications.

Thus, enhancer deregulation can occur at various levels of enhancer activity regulation and drive aberrant expression of target genes. Collectively, human cancers exhibit an aberrant enhancer signature that supports the oncogenic state. The following two sub-sections focus on two specific aspects of enhancer deregulation in cancer of particular relevance to this work: the role of MYC family transcription factors at enhancers and (2) the role of the acetyl-lysine reading BET bromodomain family (BRDT, BRD2, BRD3, BRD4) at enhancers.
**MYC family transcription factors**

Over the past three decades, the MYC family of transcription factors have been intensely studied particularly for its role in driving human malignancies. The importance of deregulated levels of MYC family TFs to cancer cannot be overstated. The MYC family of TFs consists of three members: MYC, MYCN, and MYCL. The MYC family links growth factor stimulation with cell proliferation by activating pro-growth genes downstream of several pathways (Meyer and Penn, 2008). The deregulation of MYC results in uncontrolled growth and ultimately cancer. The importance of MYC deregulation in cancer was highlighted in an analysis of copy number profiles showing that MYC is genetically altered in almost 15% of tumor samples (Beroukhim et al., 2010). There are several examples of MYC driven malignancies including hematologic cancers (such as MYC driven Multiple Myeloma) and solid tumors (such as MYCN driven neuroblastoma). In these MYC family driven malignancies, deregulation results in the elevated expression of the factor. Importantly with both MYC and MYCN, deregulation does not alter the translated protein sequence strongly suggesting that increased and sustained levels of MYC and MYCN wild-type protein drives tumorigenesis. The importance of elevated levels of MYC in driving cancer is illustrated functionally by the systemic expression of a dominant negative form of MYC that triggers rapid tumor regression in a RAS-driven lung cancer model (Soucek et al., 2008). The functional importance of elevated levels of wild type MYC family TFs has led to several studies aimed at understanding how these TFs are regulated in both a normal and malignant state.
MYC family gene expression and subsequent production/degradation of protein product is regulated at nearly every known level of molecular biology. Consequently, MYC family deregulation can occur at any of these levels in cancer. There are several mechanisms of MYC family deregulation including gene amplification and chromosomal translocation. These chromosomal translocations can improperly link cis regulatory elements to drive expression (Meyer and Penn, 2008). In fact, the importance of enhancer regulation in regulating wild type expression of MYC family TFs is evident upon close examination of their adjacent cis regulatory elements. Both MYC and MYCN are located in large gene deserts that harbor multiple large cis regulatory enhancer elements. These enhancers are established by tissue specific transcription factors and act at the termini of multiple signaling cascades to activate MYC or MYCN in different contexts. Amplifications of the MYC and MYCN loci preserve these regulatory elements further suggesting that MYC family deregulation co-opts normal control mechanisms to establish an oncogenic regulatory state. Thus, enhancer elements play a key role in the proper wild type regulation of MYC family TFs and aberrant activity is a major contributor to elevated levels in the malignant state.

In addition to enhancer regulation of MYC family TFs, recent studies have reported key mechanistic insights that have highlighted the importance of MYC family TF binding at enhancers genome-wide in driving aberrant transcriptional programs (Lin et al., 2012; Nie et al., 2012; Sabo et al., 2014; Susanne et al., 2014). MYC family TF binding at enhancers exemplifies the consequences of
aberrant transcription factor binding on enhancer activity. It is becoming increasingly clear that aberrant MYC family TF binding at enhancers is a key mechanism of oncogenesis and of particular relevance to this work. The MYC family functions as a DNA-binding transcriptional activator upon heterodimerization with another basic-helix-loop-helix leucine zipper protein, MAX (Darnell, 2002) (Figure 1.5A). The MYC/MAX dimer binds regions containing an E-box (CACGTG) consensus site (Dang et al., 2006). Upon MYC/MAX dimer binding, MYC family TFs recruit the positive transcription elongation factor (P-TEFb) to drive transcription (Figure 1.5B) (Rahl et al., 2010). P-TEFb is a cyclin dependent kinase that phosphorylates the DRB sensitivity inducing factor (DSIF) and the negative elongation factor (NELF) that liberates paused promoter-proximal Pol II into productive transcriptional elongation (Wada et al., 1998; Yamaguchi et al., 1999). This finding provided key mechanistic details into the consequences of MYC family TF binding at both promoters and enhancers in driving target gene expression. In particular, the recruitment of P-TEFb suggested that MYC family TF binding at promoter and enhancer elements serve a primary function as an activator of transcription.

Shortly after the observation that MYC binding played a role in transcriptional elongation, studies evaluating the genome-wide occupancy of dynamic levels of MYC family TFs provided further mechanistic insights (Lin et al., 2012; Nie et al., 2012; Sabo et al., 2014; Susanne et al., 2014). It is becoming increasingly apparent that deregulated levels of MYC family TFs bind promiscuously at cis regulatory elements (promoter and enhancer elements) to
drive malignant gene expression programs. While these studies have led to
disagreement on some points, they do find consensus with regards to
deregulated MYC family TF binding at enhancers (Wolf et al., 2015). At

Figure 1.5: MYC family TFs heterodimerize with MAX and bind cis regulatory elements at E-box motifs. (A) Structural depiction of the MYC-MAX heterodimer binding DNA. MYC shown as a red ribbon and MAX shown as a white ribbon. PDB: 1NKP. (B) Schematic model of MYC family TF recruitment of P-TEFb to active genes. MYC-MAX dimers bind E-box motifs and recruits P-TEFb to phosphorylate NELF and DSIF resulting in productive transcriptional elongation. (C) Schematic model of MYC family TF enhancer invasion. At low levels, MYC binds high affinity, canonical E-boxes found at promoters. At high (deregulated levels) of MYC, the TF family binds and saturates at promters and ‘invades’ lower affinity, non-canonical E-boxes at enhancers. Figure 1C adapted from Wolf et al., 2015.
physiological levels, both models predict MYC binding to high affinity sites (canonical E-boxes) at the promoters of a small number of genes involved in growth and proliferation. At deregulated levels, MYC proteins, which can only bind to regions of open and accessible chromatin, is so highly expressed that it saturates the cell’s cis regulatory landscape, binding to almost all active promoters and lower affinity sites in enhancers (non-canonical E-boxes) resulting in a complex systemic and pleiotropic transcriptional consequence (Figure 1.5C) (Guccione et al., 2006; Lin et al., 2012). As a result of widespread MYC family TF binding at enhancers, the entire preexisting gene expression program of a given cell is amplified, termed ‘global transcriptional amplification’. Importantly, this global transcriptional amplification is restricted to only active genes, suggesting that MYC family TFs can only bind to regions of active chromatin, devoid of nucleosomes (Lin et al., 2012; Nie et al., 2012). While there is agreement within the field with regards to the global transcriptional amplification phenotype, there is a growing debate as to whether this a direct versus indirect consequence of deregulated MYC TF binding (Wolf et al., 2015). With this debate withstanding, there is consensus that global transcriptional amplification is a key oncogenic component in MYC family driven malignancies.

Collectively, it has become increasingly clear that aberrant enhancer activity is a key component in the deregulated levels of MYC family TFs in cancer. The MYC family TFs themselves are highly regulated by distal enhancers and these enhancers are crucial to drive their expression as evident by the inclusion of enhancer elements in amplifications. Moreover, MYC family TFs are
also a crucial ubiquitous component of enhancer activity in the malignant state. Their promiscuous binding at non-canonical E-box containing enhancers drives a global aberrant transcriptional program. This illustrates the interconnected nature of aberrant enhancer activity in which enhancer deregulation at various levels can contribute to a malignant phenotype.

*BET bromodomain family*

It has become increasingly clear that chromatin cofactors play key roles in maintaining and modulating enhancer activity and consequently can contribute to aberrant enhancer activity in cancer. In the past decade, the acetyl-lysine binding bromodomain and extra terminal domain (BET) family of proteins have been intensely studied for their role in transcription in both a normal and malignant state. The BET family is comprised of four proteins: BRD2, BRD3, BRD4, and BRDT. The BET family proteins contain a pair of bromodomains that bind acetyl-lysine side chains. Bromodomains have a conserved fold consisting of a bundle of four alpha helices ($\alpha_z$, $\alpha_A$, $\alpha_B$, $\alpha_C$), linked by loop regions (ZA and BC loops) of considerable variability. The ZA and BC loops line the deep hydrophobic acetyl lysine binding pocket where the acetylated lysine is anchored by a hydrogen bond to a conserved asparagine residue (Owen et al., 2000). Of all the BET family proteins, BRD4 has been most well characterized and much of this work has focused on its functional role upon binding to acetylated-lysines on histone tails. BET family proteins have been categorized as transcriptional co-activators that help to facilitate transcription upon histone lysine-acetyl binding (LeRoy et al., 2008; Rahman et al., 2011). Similarly to MYC family TFs, BRD4 has emerged
as a crucial transcriptional co-activator by promoting transcriptional elongation via the recruitment of P-TEFb (Wu and Chiang, 2007; Yang et al., 2008; Yang et al., 2005). As has been touched upon in previous sections, BRD4 is often found at enhancers and plays a key role in mediating enhancer activity. In fact, BRD4 was first identified as an interaction partner with the murine Mediator complex and later shown to associate with the enhancer complex in human cells (Dawson et al., 2011; Jiang and Veschambre, 1998; Wu et al., 2003). The current model suggests that BRD4 binding at enhancers via its acetyl-lysine binding bromodomain recruits key factors to facilitate transcription (see schematic in Figure 1.3). Thus, BET family proteins function as a key link between several key components of enhancer activity. BET family proteins interpret the active open nature of chromatin at enhancers (via histone acetylation reading) and recruit several enhancer factors (Mediator, TFs) to help drive enhancer activity.

Given the importance of BRD4 in enhancer activity, it is not surprising that deregulation of the BET family can cause aberrant enhancer activity and contribute to a given malignancy. This is perhaps most directly exemplified in the rare epithelial cancer Nut midline carcinoma (NMC) in which translocation results in fusion proteins between BRD4 and the nuclear protein in testis (NUT). The resulting BRD4-NUT fusion protein is oncogenic and functional studies have shown that genetic knockdown of this fusion protein results in a decrease in cell proliferation and rapid differentiation (French et al., 2001; French et al., 2003; French et al., 2008). This BRD4-NUT fusion driven cancer illustrates how an enhancer cofactor can contribute to a given malignancy.
In addition to direct deregulation of BRD4 itself, BET family proteins have also been explored in cancer due to their key role in facilitating an aberrant enhancer program. It has become increasingly clear that BET family proteins (in particular BRD4) are non-oncogene dependencies in several human cancers (Dawson et al., 2011; Delmore et al., 2011; Mertz et al., 2011; Ott et al., 2012; Zuber et al., 2011). These studies made use of the landmark development of potent and selective small molecule inhibitors of the BET family. In particular, the small molecule JQ1 from our laboratory was the first BET family small molecule inhibitor and set the precedent for the feasibility of bromodomain inhibition (Chung et al., 2011; Filippakopoulos et al., 2010; Nicodeme et al., 2010). These new molecules truly accelerated our understanding of the role of BET family proteins in enhancer mediated transcription and pointed to efficacious nature of targeting enhancer cofactors in cancer. As has been touched upon in previous sections, BRD4 is disproportionately loaded at super enhancers that are critical to driving an oncogenic transcriptional program. The inhibition of BRD4 binding at enhancers has a profound effect in reducing enhancer activity. For example, in Multiple Myeloma, the BET inhibitor JQ1 results in a disproportionate sensitivity of BRD4 load at the translocated IgH-MYC enhancer leading to total abolishment of enhancer activity (Lovén et al., 2013).

The sensitivity of BET family proteins highlights the importance of enhancer cofactor ‘reader’ proteins in driving enhancer activity. Furthermore, the evolution of BET family inhibition highlights the general impact that enhancer activity has in cancer. The sensitivity of a variety of human cancers to BRD4
inhibition exemplifies the functional consequences of both broad and targeted
enhancer inhibition can have on a given malignant transcriptional program. Thus,
it is becoming increasingly clear that enhancer activity is based on the collective
contribution of several components (TFs, histone modifications, enhancer
cofactors, etc.) and pharmacologic inhibition at any one of these components can
have profound effects.
Chapter 2: Super enhancer profiling identifies ID1 as a dependency in neuroblastoma

Rhamy Zeid¹, Alexander J. Federation¹, William Clay Gustafson², Nicole Nasholm², Sarah J. Buhrlage¹, William Weiss² & James E. Bradner¹,³

¹Medical Oncology, Dana Farber Cancer Institute (DFCI), Boston, Massachusetts 02115, ²Department of Pediatrics, UCSF Benioff Children’s Hospital, ³Department of Medicine, Harvard Medical School, Boston, MA.

Author Contributions

RZ: Performed all in vitro experiments, performed computational analysis, wrote the manuscript, and with JEB planned the project.

AJF: Performed computational analysis including identifying super enhancer associated genes.

WCG & NN: Performed in vivo xenograft experiments

SJB: Synthesized and provided all deubiquitinating enzyme small molecule inhibitors.

WW: Supervised in vivo xenograft experiments

JEB: Supervised and developed the project with RZ.
Introduction

Neuroblastoma is the most common extra-cranial solid tumor in children and results in the highest number of cancer-related infant deaths (Brodeur, 2003). The neoplastic cells arise from primitive neural crest cells of the sympathetic nervous system and are likely a result of abnormalities during development. While treatment strategies for low-risk neuroblastoma are relatively successful, the majority of pediatric patients present with high-risk disease and intensive chemotherapy, radiation and surgery are often unsuccessful (Modak and Cheung, 2010). The transcription factor MYCN is frequently amplified and is the single most important prognostic marker of high-risk disease (Schwab et al., 1983; Seeger et al., 1985; Shimada et al., 1995). MYCN amplification is correlated with an overall survival rate of only 15-35% compared to the 80% survival of MYCN non-amplified patients (Seeger et al., 1985). Furthermore, MYCN amplification is correlated with rapid progression and established as a major oncogenic driver (Seeger et al., 1985; Shimada et al., 1995). MYCN down regulation has been shown to lead to growth arrest, apoptosis, and differentiation in various neuroblastoma models (Burkhart et al., 2003; Negroni et al., 1991; Tonelli et al., 2005). The need for a therapeutic strategy to target MYCN in this disease is truly an unmet clinical need, yet the helix-loop-helix structure makes direct small molecule inhibition challenging.

Recently, small molecule inhibitors of the BET (bromodomain and extra-terminal) family of bromodomains have led to a promising strategy to indirectly target the MYC family of transcription factors in several cancers (Chapuy et al.,
2013; Delmore et al., 2011; Shimamura et al., 2013). Many of these studies focused on the consequences of small molecule inhibition against the BET family member BRD4. BRD4 binds acetylated chromatin to facilitate transcriptional activation by recruiting P-TEFb (positive transcription elongation factor) (Bisgrove et al., 2007; Filippakopoulos et al., 2012; Zhang et al., 2012).

In recent work with the Stegmaier lab, we characterized the efficacy of BET inhibition in neuroblastoma (Puissant et al., 2013). Using our first in class small molecule inhibitor JQ1, we showed that BET bromodomain inhibition impaired growth and induced apoptosis in neuroblastoma cell lines. This phenotype was due in part to the downregulation of the MYCN transcriptional program by suppression of MYCN transcript. Specifically, JQ1 treatment led to the displacement of BRD4 from the MYCN promoter as well as a putative enhancer region leading to decreased expression. In a genetically engineered mouse model (GEMM) of MYCN amplified neuroblastoma, JQ1 treatment conferred a significant survival advantage. Like in cell line models, JQ1 treatment led to a decrease in proliferation and induced apoptosis. Collectively, this work has provided a mechanistic understanding of the efficacy of BET inhibition in neuroblastoma. The small molecule inhibitor JQ1 indirectly targets the oncogenic effects of MYCN amplification and provides a rationale for clinical investigation. Indeed, these studies have supported current efforts in the clinical development of BET bromodomain inhibition in children with relapsed or refractory neuroblastoma.
The availability of small molecule BET bromodomain inhibitors have led to new biological insights into the role of these acetyl-lysine reader proteins. Initially, BET inhibition was studied in several blood cancers including multiple myeloma, Burkitt’s lymphoma, acute myeloid leukemia, and acute lymphoblastic leukemia. Like MYCN in neuroblastoma, BET inhibition caused selective repression of the MYC oncogene in these blood cancers (Dawson et al., 2011; Delmore et al., 2011; Mertz et al., 2011; Ott et al., 2012; Zuber et al., 2011). These studies showed that BET inhibitors deplete BRD4 at enhancers that drive MYC expression and thus suggested that BRD4 played an important role at enhancers.

Enhancers are DNA regulatory elements that play key roles in the control of cell-type specific gene expression programs (Calo and Wysocka, 2013; Carey, 1998; Xie and Ren, 2013). Transcription factors bind these distal regulatory elements and recruit multiprotein complexes to ultimately drive target gene expression. These multiprotein complexes are massive and include chromatin remodelers, transcriptional machinery, and stabilizing proteins. At the core of this complex, Mediator anchors enhancer-bound transcription factors with the basal transcriptional machinery on promoters to drive gene expression (Dorsett, 1999; Dorsett and Merkenschlager, 2013; Spitz and Furlong, 2012). The selective depletion of BRD4 from the MYC enhancer was rather unexpected considering BRD4 had been described as a general transcriptional coactivator that localizes to enhancer (and promoter) regions (LeRoy et al., 2008; Rahman et al., 2011).
This observation led us to further investigate the biological role of BRD4 at enhancers in transcriptional control.

In a collaborative effort with Richard Young, we began mapping BRD4 localization genome-wide using chromatin immunoprecipitations coupled to high throughput sequencing (ChIP-seq). We observed a subset of enhancer regions with disproportionately high levels of BRD4 (Lovén et al., 2013). These subsets of enhancers, termed super enhancers differ from normal enhancers in size, master transcription factor density, and sensitivity to perturbation. These large clusters of cis regulatory elements collectively drive high expression levels of key cell-type specific genes and thus play prominent roles in maintaining cell identity (Hnisz et al., 2013; Lovén et al., 2013; Whyte et al., 2013).

Super enhancers have also been found at key oncogenic drivers in many tumor types and this led to hypothesis that super enhancer analysis might identify previously unrecognized tumor dependencies. In our recent work, we validated this approach by performing super enhancer analysis in diffuse large B-cell lymphoma. We identified the gene regulatory factor OCA-B as a super enhancer associated dependency (Chapuy et al., 2013). Thus, these studies provide a rationale to systematically carrying out super enhancer analysis with the goal of identifying unrecognized tumor dependencies.

The efficacy of JQ1 in neuroblastoma establishes the importance of BRD4 to the pediatric cancer. Given the need for alternative therapeutic strategies, we surmised that super enhancer analysis would identify additional neuroblastoma-specific dependencies that may have therapeutic application.
Results

Enhancer landscape of neuroblastoma reflects MYCN amplification status

In an effort to comprehensively map the active enhancer gene regulatory map in neuroblastoma, we studied a series of human cell lines and human primograft tumors (Figure 2.1A). Since neuroblastoma is fundamentally defined by MYCN amplification, these samples were carefully chosen to represent both a MYCN amplified and non-amplified state. We first quantified the amount of MYCN mRNA transcript and protein expression resulting from MYCN amplification in cell lines (Figure 2.1B,C). In these lines, MYCN amplification was previously measured and defined based on copy number analysis using the SNP 6.0 array (Puissant et al., 2013). While non MYCN-amplified lines have no detectable MYCN protein, lines with varying degrees of amplification have drastically increased levels of MYCN transcript and protein.

Using chromatin immunoprecipitation coupled with high throughput sequencing (ChIP-seq), we systematically identified active enhancer regulatory regions. Specifically, the genome-wide localization of histone H3 lysine 27 (H3K27Ac) was used as the primary mark of active enhancers. To define enhancers, H3K27Ac signal for each sample is stitched together to allow for the greatest amount of enriched peak space with the fewest discrete regions. Any H3K27Ac signal within 2.5 kilobases of the promoter were excluded. In total, we identified 51,922 enhancer loci across the six cell lines and five primografts.

We first looked at individual loci of highly active enhancers as measured by the amount of enriched H3K27Ac signal. We identified highly active
enhancers at several genes previously implicated in neuroblastoma including the neural crest transcription factor HAND2 (Figure 2.2A). We observe both proximal and distal enhancers at the HAND2 locus in both cell lines and primografts.

Figure 2.1: Models of MYCN amplified/non-amplified neuroblastoma. (A) Summary schematic of models of MYCN amplified/non-amplified neuroblastoma utilized in this study. (B) Western blot of MYCN protein expression. (C) Cell normalized MYCN mRNA expression levels. mRNA transcript levels are normalized to spike-in RNA standards.
Given the tissue-specific nature of these cis regulatory elements, the presence of neuroblastoma-specific enhancer driven genes suggests that these models offer a faithful recapitulation of the pediatric cancer.

To quantitatively compare the global enhancer landscapes of cell lines and primografts, we performed unsupervised hierarchical clustering of H3K27Ac signal at the 51,922 total enhancers for each sample (Figure 2.2B). Here, the MYCN amplicon is excluded from the analysis to ensure that the abnormally high H3K27Ac signal reflective on MYCN amplification does not artificially drive the clustering. The active enhancer regulatory landscape clustered by MYCN amplification status, suggesting an interplay between MYCN and the enhancer landscape. Interestingly, one non-amplified cell line (SHS-Y5Y) clusters with MYCN amplified lines and primografts. MYC protein levels are relatively high in the SHS-Y5Y line and this may be reflected in the enhancer landscape. It is becoming increasingly clear from recent studies that deregulated MYC family members drive oncogenic transcriptional programs via binding at active enhancer elements (Lin et al., 2012; Nie et al., 2012; Sabo et al., 2014). Given that MYCN amplification is a defining characteristic of high-risk neuroblastoma, we hypothesized that the MYCN-specific enhancer landscape could help provide mechanistic insight into how MYCN drives tumorigenesis and offer alternative therapeutic opportunities.
Figure 2.2: MYCN-specific enhancer landscapes. (A) Gene tracks of H3K27Ac load at the HAND2 locus. MYCN amplified and non-amplified samples are shown in red and black respectively. (B) Unsupervised hierarchical clustering using the median normalized H3K27Ac signal calculated at all enhancers (n=51,922). The H3K27Ac signal within the MYCN amplicon was excluded.
Super enhancer profiling identifies key regulatory regions in neuroblastoma

There is a growing body of literature illustrating that super enhancers drive the expression of genes required in the maintenance of tumor cell identity as well as oncogenes (Chapuy et al., 2013; Hnisz et al., 2013). Thus, we systematically mapped super enhancers to identify cis regulatory elements of potentially particular importance to neuroblastoma. We observed an asymmetric distribution of H3K27Ac signal in each of the lines and primografts and geometrically defined super enhancers as those cis regulatory elements above the inflection point (Figure 2.3A,B). Each of the cell lines had between 300-900 super enhancer loci and the relative dynamic range of signal between each cell line was similar. We again performed unsupervised hierarchical clustering of these super enhancers and observed that super enhancers reflect MYCN amplification status (Figure 2.4). Indeed, clustering of super enhancer loci alone produced very similar clustering to that of all enhancers. This suggests that this very select number of super enhancer regions can effectively distinguish the MYCN amplified cis regulatory landscape and point to the importance of these regions in neuroblastoma.

Since super enhancers are thought to collectively drive high expression of their associated genes, we then looked to accurately predict super enhancer associated genes. We leveraged the H3K27Ac signal at the promoter region of nearby genes. For each super enhancer loci, a target gene window 50 kilobases in both directions was set and the gene with the most amount of H3K27Ac signal at the promoter was assigned. To validate this approach, we leveraged global
Figure 2.3: Super enhancer profiling in neuroblastoma. (A) Distribution of H3K27Ac ChIP-seq signal at enhancers in cell lines (B). Distribution of H3K27Ac ChIP-seq signal at enhancers in primografts. Enhancer regions are plotted in an increasing order based on input normalized signal (x-axis) versus H3K27Ac signal at enhancers (y-axis). Super enhancers are defined as the population of enhancers above the inflection point.
measurements of mRNA transcript levels (measured by microarray) in two representative cell lines. Using the SKNAS and BE2C cell lines, super enhancer associated genes are on average expressed at higher levels than that of typical enhancer associated genes (Figure 2.5A).

To investigate the relevance of super enhancer associated genes in

![Figure 2.4: MYCN-specific super enhancer landscapes.](figure)

Unsupervised hierarchical clustering using the median normalized H3K27Ac signal calculated at super enhancers (n= 2,506). The H3K27Ac signal within the MYCN amplicon was excluded.
neuroblastoma, we created a gene set of all the super enhancer associated genes in the SKNAS and BE2C lines and compared them to literature gene signature sets (Subramanian et al., 2005). We identified several neuroblastoma-specific gene sets that significantly overlapped with our compiled super enhancer gene set (Figure 2.5B). Furthermore, we also identified MYC family gene sets that significantly overlapped. On an individual gene basis, we observed super enhancers associated to key genes previously implicated in neuroblastoma genesis and maintenance including the master transcription factor GATA2 (Figure 2.5C) (Hoene et al., 2009; Reiff et al., 2010).

Given the previously established importance of the BET bromodomain protein BRD4 in neuroblastoma, we also defined super enhancers by BRD4 ChIP-seq in the SKNAS and BE2C cell line. While there were fewer BRD4-defined super enhancer loci, there was a high degree of overlap with the H3K27Ac defined super enhancers (Figure 2.5D). Interestingly, there appeared to be more similarity with higher ranked super enhancer loci. The decrease in BRD4-defined super enhancers coupled with higher similarity in higher ranked super enhancer loci suggests that BRD4 may define a more discrete set of regions. Ultimately, we concluded that both H3K27Ac and BRD4 could be used to define super enhancer regions with a high degree of similarity. Thus, these super enhancer regions drive increased expression of neuroblastoma-specific genes that we surmised may be critical to the oncogenic state.
Figure 2.5: Super enhancer associated genes are highly expressed and neuroblastoma-specific. (A) Box plots of cell normalized log2 mRNA transcript expression levels. Genes in the top 80% of expressed genes were classified as typical enhancer associated (gray) or super enhancer associated (red). Asterisks indicate a p < .0001, *** or p < .0002, ** obtained using a two-tailed t-test. (B) Select top significant gene sets that overlap with a composite list of super enhancer associated genes in neuroblastoma. Significance is represented by false discovery rate q-value. (C) Gene tracks of H3K27Ac load at the GATA2 locus. MYCN amplified and non-amplified samples are shown in red and black respectively and super enhancer regions are depicted by the black bar. (D) (Left) Distribution of BRD4 ChIP-seq signal at enhancers. Enhancer regions are plotted in an increasing order based on input-normalized signal (x-axis) versus BRD4 signal at enhancers (y-axis). (Right) Rank order comparison of BRD4 versus H3K27Ac defined super enhancer loci.
Functional interrogation of super enhancer associated genes identifies ID1 as a dependency in neuroblastoma

With a high confidence list of super enhancer associated genes, we then sought to functionally screen for super enhancer associated genes required for cell viability. Using the SKNAS and BE2C cell lines as representative models of the MYCN non-amplified and amplified state, we utilized siRNA technology to knockdown super enhancer associated genes and then measured the consequences on cell viability (Figure 2.6A). In an arrayed format, a composite library of 331 BRD4-defined super enhancer associated genes from the SKNAS and BE2C lines were targeted by transfection of four individual siRNA duplexes. In total, there were 29 overlapping super enhancer associated genes between the SKNAS and BE2C cell lines (Figure 2.6B). We assessed cellular viability by surrogate measurements of ATP levels and z-score values were calculated for each gene based on the mean and standard deviation of the population.

Since BRD4 knockdown (as well as inhibition by the BET bromodomain inhibitor JQ1) causes potent cell death, we reasoned that knockdown of candidate genes that produced similar z-scores may represent viable hit candidates. We also benchmarked against z-score values for that of RET kinase. RET kinase has been implicated as a neuroblastoma dependency and its inhibition using the kinase inhibitor vandetanib has been show to impair tumor growth in mouse models (Cazes et al., 2014; Peterson and Bogenmann, 2004). Rank-ordered z-scores in each of the two cell lines showed a spread of effects on viability (Figure 2.6C). Comparing z-scores to siRNA mediated knockdown of
Figure 2.6: Functional siRNA screening of super enhancer associated genes identifies genes previously implicated in neuroblastoma. (A) siRNAs targeting super enhancer associated genes were screened in an arrayed format and assayed for viability in 384-well format. (B) Overlap in super enhancer genes in the SKNAS and BE2C lines. (C) Effects of siRNAs targeted super enhancer associated genes are rank ordered by viability z-score (y-axis) versus siRNA (x-axis). (D) Z-score values of genes previously implicated in neuroblastoma compared to siRNA knockdown of BRD4 (labeled red).
BRD4, a number of previously implicated neuroblastoma specific dependencies were identified including CDC6 and PHOX2B (Feng et al., 2008; Reiff et al., 2010) (Figure 2.6D). The z-scores associated with knockdown of these super enhancer associated genes all ranked very close to the effects observed with siRNA mediated knockdown of BRD4. These data illustrate that like in other malignancies, super enhancers do in fact drive gene expression of key contributors to neuroblastoma.

After identifying previously characterized dependencies, we then looked for previously unrecognized super enhancer associated dependencies. We narrowed down on a focused list of screening hits from each cell line based on several criteria including functional category, super enhancer rank, and transcript expression (Figure 2.7A). Early on, we focused on ID1 (inhibitor of DNA binding protein 1) due its potent effect on viability in both cell lines. ID1 is a helix-loop-helix (HLH) protein that lacks a DNA binding domain and forms heterodimers with members of the basic HLH (bHLH) family of transcription factors (Jonathan et al., 2005; Marianna and Robert, 2003; Sylvia et al., 2004). Consequently, ID1 inhibits bHLH transcription factors from binding DNA and antagonizes their function. There are four ID proteins (ID1, ID2, ID3, ID4) and ID1 in particular has been implicated in several cancers. It is overexpressed in a variety of human cancers (pancreatic, ovarian, prostate, brain) and high expression is often correlated with poor prognosis (Fong et al., 2004; Ruzinova and Benezra, 2003; Sikder et al., 2003). In addition, loss of ID1 and ID3 during murine development leads to premature neural differentiation (Lyden et al., 1999). In neuroblastoma,
there have been limited studies of the ID proteins in experiments working with human and mouse cell lines. In particular, there has been controversy surrounding the role of ID2 with respect to MYCN in neuroblastoma genesis. ID2 has been shown to be a direct transcriptional target of MYCN and it has been suggested that ID2 is a key effector of MYCN (Lasorella et al., 2000; Raetz et al., 2003). However, this work was later disputed and it was claimed that ID2

Figure 2.7: siRNA screening of super enhancer associated genes identifies ID1 as a previously unrecognized dependency. (A) Z-score values of screening hit candidates as compared to siRNA knockdown of BRD4 (red). (B) Cell normalized, spike-in normalized levels of mRNA transcript expression in neuroblastoma cell lines. (C) Effect of siRNA knockdown of ID1 on cell viability as measured by crystal violet staining 96 hours post transfection. (D) Western blot of ID1 protein levels 24 hours post transfection of each individual siRNA duplex. (E) Gene tracks of H3K27Ac ChIP-seq at the ID1 locus. Super enhancer loci are depicted by the black horizontal bar.
expression only correlates with MYCN expression but is dispensable for MYCN mediated oncogenesis (Lasorella et al., 2005; López-Carballo et al., 2002). Given the cell-type specific nature of super enhancer associated genes, we were particularly motivated to understand the role of ID1 with respect to its super enhancer with the hope of taking advantage of a context-specific dependency.

We first looked at cell normalized mRNA transcript levels of ID1 and observed high expression across neuroblastoma cell lines (Figure 2.7B). The ID1 screening hit was then retested in secondary validation in both the SKNAS and BE(2)C cell lines. The four individual siRNA duplexes were retested in triplicate along with the non-targeting control and viability was measured by crystal violet 96 hours post transfection (Figure 2.7C). With respect to cell viability, each siRNA duplex behaved similarly from screening to secondary validation. Furthermore, viability loss with each siRNA duplex was correlated with ID1 protein knockdown (Figure 2.7D). After validating the viability defect seen with siRNA knockdown in the SKNAS and BE(2)C lines, we then looked to see if ID1 was super enhancer associated in additional neuroblastoma lines. As shown, ID1 is super enhancer associated in both MYCN amplified and non-amplified lines (Figure 2.7E). While ID1 is not super enhancer associated in two of the six lines, there is still significant H3K27Ac reflecting potential enhancer activity.

*Indirectly targeting ID1 for degradation by small molecule inhibition of USP1*

In a 2011 study, a group from Genentech showed that the deubiquitinating enzyme (DUB) USP1 promotes ID protein stability in osteosarcoma cells (Williams et al., 2011). USP1 removes polyubiquitin chains from ID1 and rescue
degradation. In turn, selective knockdown of USP1 resulted in the rapid
degradation of ID1 ultimately led to enhanced differentiation. DUBs are key
regulators of the ubiquitin system that cleave ubiquitin moieties from proteins and
thus divert proteins from degradation. DUBs have been divided into five
subclasses and in total about 80 have been identified in the human genome thus
far (Komander et al., 2009; Nijman et al., 2005). The emerging understanding of
this class of enzymes has led to several studies that have implicated DUBs in
several diseases in including cancer, inflammation, and neurodegeneration
(Kondapalli et al., 2012; Liu et al., 2002; Saigoh et al., 1999). Since these
enzymes have potentially druggable active sites, they have also become
increasingly attractive drug targets. In 2013, ubiquitin-rhodamine-based
screening identified a small molecule inhibitor of the DUB USP1 (ubiquitin-
specific protease 1) (Helena et al., 2013). This study went on to show that
chemical inhibition of USP1 by a small molecule SJB3-019A resulted in
degradation of ID1 protein and cytotoxicity in leukemic cells. Since this study,
additional groups have identified additional small molecule inhibitors of USP1
with distinct chemical scaffolds (Chudi and Vickie, 2014; Matthay et al., 2012; Qin
et al., 2014).

As in leukemic cells, we hypothesized that small molecule inhibition of
USP1 would lead to degradation of the ID1 protein in neuroblastoma. As a super
enhancer associated gene dependency, we surmised that USP1 inhibition would
have a viability effect and could represent a downstream therapeutic application.
We screened a tool compound library of DUB inhibitors in a panel of neuroblastoma cell lines. This library included the three published USP1-specific inhibitors as well as several additional small molecule inhibitors targeting other DUBs (Figure 2.8A). We performed cell viability screening in 384-well format in 10-point dose response to generate IC\textsubscript{50} curves. Like siRNA screening, cell viability was measured by surrogate measurements of ATP content. Several of these tool compounds had potent viability effects at 24 hours and the USP1 inhibitors b-AP15 and SJB3-019A were among the most potent across all cell lines (Figure 2.8B). We decided to follow-up on the SJB3-019A compound due to the potency and availability of molecule. At 24 hours, SJB3019A resulted in degradation of ID1 protein in both the SKNAS and BE(2)C lines (Figure 2.8C). This degradation is proteasome dependent as shown by rescue of ID1 degradation when using the proteasome inhibitor epoxomicin (Figure 2.8D).

Since USP1 is predicted to deubiquitinate on the order of several hundred targets, we took a mass-spectrometry proteomics approach to understand how SJB3-019A effects global protein stability. We quantified approximately 7,289 proteins in response to SJB3-019A (compared to DMSO control) at an early and late time point (Figure 2.9A). This experiment was carried out in triplicate and proteins with three spectral counts or below were excluded from the analysis. Since ID1 is such a small protein, it was not quantifiably detectable by this measurement (only having 2 spectral counts). Surprisingly, the most downregulated gene at both 5 and 16 hours of SJB3-019A treatment in the BE2(C) cell line was MYCN. In addition, other super enhancer associated genes
Figure 2.8: Small molecule inhibition of USP1 indirectly targets ID1 for degradation. (A) Chemical structures of three published USP1 small molecule inhibitors. (B) Dose response viability of a chemical library of DUB inhibitors. Normalized viability as measured by ATP concentrations (y-axis) against a 10-point dose of compound (x-axis). SJB3-019A and b-AP15 inhibitors highlighted in red and maroon respectively. (C) Western blot of ID1 levels at 16 hours incubation with SJB3-019A in the SKNAS (top) and BE2C (bottom) cell lines. (D) Western blot of ID1 levels at 16 hours co-incubation with SJB3-019A and Epoxomicin.
were significantly downregulated at both 5 and 16 hours including CDC6 and CCND1. It is important to note that MYCN degradation is observed as early as 5 hours, while ID1 degradation is not observed until 16 hours. Thus, it is unclear whether ID1 degradation in response to USP1 inhibition is a direct effect or an indirect consequence of MYCN degradation. Further, this data suggests that SJB3-019A has off-target effects and/or USP1 in fact also targets MYCN for deubiquitination. Nevertheless, the ability of SJB3-019A to potently degrade MYCN is itself a very promising result and could have a huge impact in MYCN amplified neuroblastoma.

To understand the pharmacokinetic and pharmacodynamics properties of the SJB3-019A compound, we carried out xenograft studies in mice. While there is uncertainty surrounding the mechanistic underpinnings of cell death associated with SJB3-019A (MYCN versus ID1), we were interested in evaluating the in vivo behavior of the compound. Here, SJB3-019A was dosed at 50mg/kg daily by oral gavage and compared to vehicle control in a xenograft model with BE2C cells. There was a limited effect on survival after 18 days of dosing compared to vehicle control (Figure 2.9B). In addition, there did appear to be some toxicity as mice treated with inhibitor had consistently lower weights than animals treated with vehicle control (Figure 2.9C). The effects on animal weight were not drastic enough to require euthanasia throughout the dosing regimen. After 24 hours of the last dose on day 18, tumors were harvested to measure MYCN protein levels. There appeared to be no MYCN degradation in SJB3-019A treated mice (Figure2.9D). Given the solubility challenges with this inhibitor, we hypothesize
that we did not reach critical plasma concentrations to see the desired biological effect. An improved formulation of SJB3-019A may help overcome some of the pharmacological liabilities of this compound and result in better performance \textit{in vivo}.

\textbf{Discussion}

In neuroblastoma, MYCN amplification is the defining feature of high risk disease. We mapped the active cis regulatory landscape in a panel of cell lines

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure2_9.png}
\caption{Small molecule inhibition of USP1 in MYCN-amplified neuroblastoma. (A) Scatter plot of log2 transformed fold change of 7,289 proteins comparing 750 nM SJB3-019A to DMSO treatment at 5 and 16 hours (triplicate analysis) in the BE2C cell line. (B) Kaplan–Meier survival curves comparing SJB3-019A to vehicle treatment in xenograft mouse model of BE2C cells. (C) Scatter plot of change in tumor weight with SJB3-019A versus vehicle treatment over time in BE2C xenograft model. (D) Western blot of MYCN levels 24 hours after the last course of treatment on day 18.}
\end{figure}
and primografts and observed a MYCN amplified-specific enhancer signature. It remains unclear how this distinct enhancer landscape state is achieved. MYC family transcription factors are thought to bind the pre-established active regulatory landscape rather than localizing to inactive genomic regions (Lin et al., 2012; Nie et al., 2012). The MYCN-specific cis-regulatory landscape also suggests that MYCN requires this chromatin state in order to support its oncogenic function. The effects of small molecule inhibition of the acetyl-lysine reading enhancer factor BRD4 further support this (Puissant et al., 2013).

The use of super enhancer analysis in neuroblastoma led us to pursue ID1 as a previously unrecognized dependency. The cell-type specificity of super enhancer associated genes and examples of super enhancers driving oncogenes offers a promising approach to identifying previously unrecognized dependencies in a given malignancy. We established ID1 as super enhancer associated in several human cell lines and established it as a dependency in a representative pair of cell lines (MYCN amplified and non-amplified). In addition to MYCN, we also identified several genes previously implicated in neuroblastoma, suggesting that functional interrogation of super enhancer associated genes is a valuable strategy to identify vulnerabilities. The cell-type specificity of ID1 as well as other screening hit candidates suggests that super enhancers drive expression of neuroblastoma-specific genes. Thus, super enhancer mapping provides a context-specific fingerprint of the transcriptional program of a given malignancy.

We pursued a small molecule inhibitor of USP1 to indirectly target ID1 for degradation. While the SJB3-019A tool compound potently killed neuroblastoma
cells, this effect cannot be solely explained by ID1 degradation because MYCN protein is also degraded. It is important to note that USP1 is predicted to deubiquitinate several targets (on the order of hundreds). In addition, SJB3-019A has significant off-target effects on several other DUBs (Ritorto et al., 2014). Nevertheless, we observed a potent viability loss in neuroblastoma cell lines and identified the first near genome-wide target candidate profile of the SJB3-019A molecule. The lack of efficacy observed in mouse models reflects the poor pharmacological properties of an early stage tool compound. The characterization of both DUB family and small molecule inhibitors is an emerging field. As our understanding of the target profile of USP1 grows, we will be better equipped to understand the effects of inhibition. Moreover, there is a collective effort to improve on the selectivity and potency of small molecule inhibitors of DUBs.

Given the truly dismal prognosis of MYCN-amplified neuroblastoma, these studies provide early evidence that targeting super enhancer associated genes could identify previously unrecognized dependencies. As DUB inhibitors with more favorable pharmacological properties are developed, indirect targeting of ID1 and/or MYCN via USP1 inhibition may prove to be a viable strategy in neuroblastoma.

Methods

Cell Lines and primografts

SKNAS, SHS-Y5Y, NGP, BE2C, KELLY, and LAN1 cells were kindly provided by Dr. Kimberly Stegmaier (Dana Farber Cancer Institute). All cell lines
were cultured in Dulbecco’s Modified Eagle Medium (DMEM) supplemented with 10% FBS. Primograft samples were graciously provided by Dr. Michael Dyer as flash frozen tumor chunks (St. Jude Children’s Research Hospital)(Stewart et al., 2015).

**Immunoblotting**

Cells were lysed using RIPA buffer supplemented with protease inhibitor cocktail (Roche) for 20 minutes. The lysates were spun at 16,000 g for 15 minutes at 4°C and protein concentration was determined using BCA assay (Pierce). Antibodies for western blot were purchased as follows: MYCN (Santa Cruz, sc-56729), ID1 (Santa Cruz, sc-365654), GAPDH (sc-365062), and Tubulin (Santa Cruz, sc-8035). Blots were imaged using the fluorescence-labeled secondary antibodies (LI-COR) on the OdysseyCLX Imager (LI-COR).

**RNA extraction and microarrays**

Prior to RNA extraction, cell numbers were determined using the Countess Automated cell counter (Invitrogen). Total RNA extraction was performed on 500,000 cells using the miVana miRNA total RNA isolation kit (ThermoFisher Scientific, AM1560) according to manufacturers instructions. During isolation, external RNA spike-ins (ERCC, Ambion) were added at the time of cell lysis (Lovén et al., 2013). Microarray expression data were normalized to produce cell count normalized expression values in arbitrary units.

**Chromatin immunoprecipitation**

For cell lines, chromatin immunoprecipitations were performed as described with minor changes(Chapuy et al., 2013). Neuroblastoma cell lines
were grown to 75% confluence in 15 cm plates and cross-linked with 1% formaldehyde (10 minutes) followed by quenching (125 mM glycine). Cells were washed in cold PBS and harvested by cell scraper in cold PBS with protease inhibitors (Roche). Cells were centrifuged at 1650 x g for 5 minutes. Pellets were resuspended in cytosolic then nuclear lysis buffer and DNA was sheared on ice using a waterbath sonicator (Bioruptor, Diagenode) for 25 minutes at high output (30" on, 30" off) in 1mL of sonication buffer supplemented with 0.5% SDS.

Antibodies for ChIP were purchased as follows: H3K27Ac (abcam, ab4729) and BRD4 (Bethyl, A301-985A). Sonicated lysates were cleared by centrifuging at 20,000g for 10 min and incubated overnight end-over-end at 4°C with magnetic beads prebound with antibody. Beads were washed three times with sonication buffer, one time with sonication buffer with 500 mM NaCl added, one time with LiCl wash buffer (20 mM Tris pH 8.0, 1 mM EDTA, 250mMLiCl, 0.5% NP-40, 0.5% Na-deoxycholate) and once with TE. DNA was eluted in elution buffer (50 mM Tris-HCl pH 8, 10mM EDTA, and 1% SDS). Cross-links were reversed overnight at 65°C. RNA and protein were digested with 0.2mg/mL RNase A for two hours followed by 0.2mg/mL Proteinase K for one hr. DNA was purified with phenol chloroform extraction and ethanol precipitation.

Libraries for sequencing were prepared using the Rubicon Thruplex FD library preparation kit. An input of 50 ng of DNA or less were used and following ligation libraries were amplified per manufacturers instructions. Amplified libraries were then size-selected using AMPure beads (Agencourt AMPure XP) per manufacturers instruction. Further size selection was performed using a 2% gel
cassette in the Pippin Prep (SAGE Sciences) set to capture fragments between 200 – 700 base pairs. Libraries were multiplexed at equimolar rations and run together in one lane on the NextSeq (75 base pair, single-end).

For primografs, tumor tissue was minced in 1% formaldehyde and incubated for 10 minutes, followed by quenching with 2.5 M glycine. Samples were homogenized with a Dounce homogenizer using 10 strokes with pestle A and 10 strokes pestle B in cytosolic lysis buffer. All subsequent steps were carried out in the same manner described for cell lines in the previous paragraph.

*ChIP-seq analysis*

All analyses were performed using NCBI37/HG19 annotations. All ChIP-Seq datasets were aligned using Bowtie2 (version 2.2.1). 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. ChIP-Seq reads aligning to the region were extended by 200bp and the density of reads per base pair (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). We used the MACS version 1.4.1 (Model based analysis of ChIP-Seq) 67 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.
Identifying super enhancers

We identified super enhancer loci as described previously with minor adjustments (Chapuy et al., 2013; Lovén et al., 2013). Using H3K27Ac and BRD4 ChIP-seq, we ranked all enhancers by increasing total background subtracted CHIP-seq occupancy and plotted against the total background subtracted CHIP-seq occupancy in units of total reads per million. We geometrically defined the inflection point and used it to establish a cut off for super enhancer delineation. A stitching distance (between 4-20 kilobases) unique to each dataset was used to optimize the greatest amount of enriched peak space using the fewest discrete regions.

To predict super enhancer associated genes, a region 50 kilobases in either direction of a given super enhancer locus was set to define the target gene window. Any gene within this target window that was expressed within the top 80% of all genes was linked to that super enhancer. The gene with the highest gene expression was defined as the top gene target of its respective super enhancer.

Enhancer clustering

We clustered both enhancers and super enhancers as described previously with minor adjustments (Chapuy et al., 2013; Lin et al., 2016). Within a set of samples, all unique regions were identified. In each region, for each sample, background subtracted H3K27Ac was calculated relative to the median enhancer. Pairwise pearson correlations were determined for all patterns of H3K27Ac signal and this correlation score was used to hierarchically cluster...
containing regions together. Patterns of H3K27Ac signal at super-enhancer containing regions were displayed in two-dimensionally clustered heatmaps with each enhancer (or super enhancer) containing region displayed as a row color encoded by H3K27Ac fold signal over the median.

**siRNA screening**

siRNA screening was carried out in 384-well arrayed format with individual siRNA duplexes. The focused custom duplex library was purchased from Dharmacon via the ICCB-Longwood Screening Facility at Harvard University. The 331 gene library was created by taking the BRD4-defined super enhancer associated genes from the SKNAS and BE2C cell lines. Any gene within 50 kilobases in either direction of a super enhancer locus that was expressed in the top 80% of all genes was included in the library. A total of four siRNA duplexes were used per gene in triplicate. Lipofectamine RNAiMAX (Invitrogen) was incubated for 5 minutes with serum-free OptiMEM (Invitrogen) for 5 minutes in a 384-well assay plate. Using an Agilent Bravo, 1.25µl of 1µM siRNA was added to each well and incubated for 20 minutes at room temperature. For both the SKNAS and BE2C cell lines, 1500 cells in 40µl was seeded directly into the transfection mix (reverse transfection) using a multidrop Combi (Thermo). Cell viability was measured 96 hours post transfection.

Cell viability was measured via surrogate measurements of ATP concentration using ATPlite (PerkinElmer) following the manufacturers instructions. A diluted stock of substrate in lysis buffer was made and 25µl was dispensed into each well of the 384-well plate using a BioTek EL406 multi-mode
dispenser. Plates were spun down an allowed to incubate for 10 min before luminescence measurements using a multiplate Envision plate reader (Perkin Elmer).

To ensure data quality, Z-prime values were calculated for each plate within the screen. Z-prime values were calculated based on non-targeting control siRNA and a siRNA targeting BRD4 as a positive control. Screening data was analyzed by first compiling and normalizing raw luminescence reads to the average of the negative control on each respective plate. Normalized values were then rank-ordered based on the lowest average normalized luminescence signal from two of the four duplexes. From there, z-score values were calculated for each gene (based on the two best performing siRNAs) based on the mean and standard deviation of the population. Z-score values were then rank ordered and benchmarked to a relevant biological positive control (siRNA targeting BRD4).

**Compound screening**

Using a BioTek EL406 multi-mode dispenser, 2000 cells in 50µl were plated into a 384-well white bottom assay plate. The DUB inhibitor tool compound set was arrayed in a corresponding 384-well plate. After cell plating, 100 nanoliters of compound was pinned into each well using a Janus workstation. Compound stocks were arrayed in 10 point dose in quadruplicate in DMSO. Cell viability was measured via ATPlite as described for siRNA screening. Nonlinear dose-response curves were fitted to the data using Graphpad Prism software.
**In vivo**

All animals were carried out in the laboratory of Dr. William Weiss at UCSF and the studies were conducted using approved protocols. The BE2C cell line was implanted subcutaneously in the flanks of female NSG mice. After tumor formation, mice were divided into cohorts to be treated with 50 mg/kg of SJB3-019A or vehicle (formulation alone). SJB3-019A was formulated in 10% N-Methyl-2-pyrolidone (NMP), 70% PEG, and 20% propylene glycol. Administration of drug was carried out for a total of 18 days. Tumor volume was determined by caliper measurement and total mouse weight was also monitored.

**Acknowledgements**

We would like to thank the ICCB-Longwood Screening Facility at Harvard University (Jennifer Smith, Sean Johnston, and David Wrobel) for their guidance/expertise/assistance in performing siRNA screening.
Chapter 3: Enhancer invasion shapes MYCN dependent transcriptional amplification in neuroblastoma

Rhamy Zeid¹, Jaime M. Reyes¹, Donald R. Polaski¹, Matthew A. Lawlor¹, Rachel A. Hirsch², Thomas Scott¹, Michael A. Erb¹, Georg E. Winter¹, Charles Y. Lin²*, James E. Bradner¹,³*

¹Medical Oncology, Dana Farber Cancer Institute (DFCI), Boston, MA 02115.
²Baylor College of Medicine, Department of Molecular and Human Genetics, Houston, TX.
³Department of Medicine, Harvard Medical School, Boston, MA.

Author Contributions

RZ: Performed all in vitro experiments not attributed to others. Planned, developed and wrote the manuscript with JEB and CYL.

JMR, DRP, RAH: Performed computational analysis.

MAL, TS: Assisted on ChIP-seq and western blotting.

GEW, MAE: Assisted on CRISPR-Cas9 mutagenesis scanning.

CYL: Supervised/developed the project, performed computational analysis, and wrote the manuscript with RZ and JEB.

JEB: Supervised/developed the project, and wrote the manuscript with RZ and CYL

*These authors contributed equally to this work
Introduction

In neuroblastoma, amplification of the oncogenic basic helix-loop-helix (bHLH) transcription factor (TF) MYCN is the defining prognosticator of high-risk disease, occurs in one-third of neuroblastoma, and drastically reduces overall survival rates (Matthay et al., 2012; Seeger et al., 1985). As a proto-oncogene, targeted MYCN overexpression in peripheral neural crest is sufficient to initiate disease in mouse models (William et al., 1997). In MYCN amplified neuroblastoma, elevated expression of the factor is crucial to maintain tumor stemness (Kang et al., 2006; Wakamatsu et al., 1997) and is associated with increased proliferation and aberrant cell cycle progression, as these tumors lack the ability to arrest in G1 in response to irradiation (Bell et al., 2007; Muth et al., 2010; Tweddle et al., 2001; Yaari et al., 2005). MYCN down-regulation broadly reverses these oncogenic phenotypes in a variety of neuroblastoma models (Burkhart et al., 2003; Negroni et al., 1991; Tonelli et al., 2005) and recent therapeutic strategies to indirectly target MYCN production or protein stability have reduced tumor growth in vivo (Edmond et al., 2014; Puissant et al., 2013; William Clay et al., 2014). These observations motivate an investigation of MYCN binding in MYCN amplified tumors as it remains fundamentally unclear how elevated levels of the factor occupy the genome and alter transcriptional programs in neuroblastoma. Here we present the first dynamic chromatin and transcriptional landscape of direct MYCN perturbation in neuroblastoma. We find that at oncogenic levels, MYCN associates with E-box (CANNTG) binding motifs in an affinity dependent manner across most active cis regulatory promoters and
enhancers. MYCN shutdown globally reduces histone acetylation and transcription, consistent with prior descriptions of MYC proteins as non-linear amplifiers of gene expression. We establish that MYCN load at promoters and proximal enhancers predicts transcriptional responsiveness to MYCN shutdown and that MYCN enhancer binding occurs prominently at the most strongly occupied and down-regulated genes, suggesting a role for these tissue specific elements in predicing MYCN responsive “target” genes. At these invaded enhancers, we identify the lineage specific basic helix-loop-helix (bHLH) TWIST1 as a key collaborator and dependency of oncogenic MYCN. These data suggest that MYCN enhancer invasion helps shape transcriptional amplification of the neuroblastoma gene expression program to promote tumorigenesis.

MYC, MYCN, and MYCL comprise the MYC family of TFs and are the most commonly altered oncogenes in cancer (Beroukhim et al., 2010; Nesbit et al., 1999). They function as the master sensor of cellular signals and mediate a transcriptional response involved in a variety of processes including proliferation, cell growth, differentiation, survival, and pluripotency. While no global MYCN occupancy data exists, several studies have examined the transcriptional consequences of elevated MYC levels, presenting two seemingly conflicting views on the mechanistic nature of MYC (Lin et al., 2012; Nie et al., 2012; Sabo et al., 2014; Susanne et al., 2014). The first suggests that MYC acts in a context specific manner, activating or repressing discrete gene sets (Sabo et al., 2014; Susanne et al., 2014). The second suggests that MYC broadly engages the cell’s existing cis regulatory landscape leading to increased transcription at all active
genes culminating in transcriptional amplification of the cell’s gene expression program (Lin et al., 2012; Nie et al., 2012). At physiological levels, both models predict MYC binding to high affinity canonical (CACGTG) E-box sites at the promoters genes involved in growth and proliferation. At deregulated levels, highly abundant MYC proteins, which can only bind to regions of open and accessible chromatin, saturate the cell’s active cis regulatory landscape, binding to prevalent degenerate non-canonical CANNTG E- and effecting a complex systemic and pleiotropic transcriptional consequence (Guccione et al., 2006). This is further complicated by the numerous functional roles of MYC family proteins, their complex interaction networks, and the difficulty in discriminating direct effects of MYC binding from secondary indirect consequences (Sabo et al., 2014). These observations suggest that MYC blurs the line between a gene-specific and global regulator, with a small number of genes uniquely evolved to rapidly respond to perturbations in MYC levels and other functional responses largely predicated on the cell’s pre-established chromatin landscape and gene expression program (Wolf et al., 2015).

Results

Hierarchy of deregulated MYCN binding at promoters and enhancers of active genes

To understand the genome-wide binding occupancy of deregulated MYCN, we used a panel of human neuroblastoma cell lines with and without MYCN amplification displaying varying levels of MYCN (Figure 3.1A; 3.1B). We
Figure 3.1: Genome-wide occupancy of deregulated MYCN in neuroblastoma. (A) Cell normalized western blot of MYCN protein levels in human neuroblastoma cell lines. (B) Cell normalized steady state mRNA transcript levels of MYCN in human neuroblastoma cell lines. (C) Increasing cell number normalized western blot of MYCN protein levels in the BE(2)-C and SHEP-21N cell lines. (D) Line plots of quantified western blot bands from increasing cell number normalized western blots of MYCN protein levels. Error bars denote ± standard deviation of four replicate blots. (E) Scatter plots of average ranked MYCN occupancy across four cell lines (x-axis) versus ranked MYCN occupancy in each respective cell line (y-axis). Contour lines illustrate the density of correlation of MYCN occupancy and color coded from high density (red) to low density (yellow). (F) Meta track representation of MYCN and H3K27ac ChIP-seq signal (rpm/bp) across four neuroblastoma cell lines at the NPM1 locus. Signal for samples are plotted as a translucent shape and darker regions indicate regions with signal in more samples. An opaque line is plotted and gives the average signal across all samples in the group. ChIP-seq signal of H3K4me3, BRD4, and RNA Pol II at the NPM1 locus are shown in the BE(2)-C cell line. Canonical E-boxes (red) and non-canonical E-boxes (black) are indicated. (G) Pie chart showing the genomic distribution of conserved MYCN binding regions across four neuroblastoma cell lines.
Figure 3.1 (Continued)
also utilized the tet-off MYCN engineered SHEP-21N cell line model (Lutz et al., 1996) to profile MYCN in an on and off state, with levels comparable to BE(2)-C in the “on state” (Figure 3.1C, 3.1D). Across the four MYCN deregulated lines, we used chromatin immunoprecipitation coupled to high-throughput sequencing (ChIP-Seq), to generate a consensus genome-wide map of ~10,000 regions that exhibit strong and consistent MYCN occupancy (Figure 3.1E; supplemental methods).

In a deregulated state, MYC binds to virtually all active promoters and subsequently ‘invades’ promoter-distal enhancer regions (Lin et al., 2012; Nie et al., 2012; Sabo et al., 2014). Like MYC, deregulated MYCN binds strongly to active promoters e.g. at NPM1 (Figure 3.1F) that are corroborated by the presence of active transcription marks/factors (H3K4me3, BRD4, RNA Pol II) in BE(2)-C. Genome-wide conserved MYCN binding regions occur almost exclusively (96%) at active promoters and enhancers and show a strong concordance with active chromatin (H3K27ac) (Figure 3.1G, Figure 3.2A, 3.2B; supplemental methods).

We next aimed to characterize the hierarchy of binding at active promoters and enhancers. We find that MYCN load correlates with H3K27ac load and these regions of increased binding overlap more with active TSS regions versus distal enhancers (Figure 3.2C). Trends in MYCN loading are reflected in the underlying sequence composition of binding sites. Strong promoter associated sites are highly enriched for the canonical (CACGTG) E-Box whereas weaker sites found
at promoters and distal enhancers show higher enrichment of clustered non-
canonical (CANNTG) E-boxes (Figure 3.2C). These trends in E-box occupancy

Figure 3.2: Hierarchy of deregulated MYCN binding at promoters and
enhancers of active genes. (A) Heat map of H3K27ac (blue) and MYCN
(red) signal at promoters in each respective neuroblastoma cell line. Each
row shows the ± 5kb centered on the TSS ranked by average H3K27ac
signal. Color scaled intensities are in units rpm/bp. (B) Heat map of
H3K27ac (blue) and MYCN (red) signal at enhancers in each respective
neuroblastoma cell line. Each row shows the ± 5kb centered on the
enhancer ranked by average H3K27ac signal. Color scaled intensities are in
units rpm/bp. (C) Line plots showing the correlation of ranked MYCN
enriched regions (x-axis) versus (top) average ChIP-seq signal (MYCN: red,
H3K27ac: blue), (middle) percent overlap with a given genomic feature
(TSS: red, Distal enhancer: blue), or (bottom) E-box density (Canonical: red,
Non-canonical: gray) on the y-axis. MYCN regions are broken up into 50
bins on the x-axis and in each bin the average value of the associated
feature is plotted. Error bars represent 95% confidence intervals of the
mean. (D) (Left) De novo motif analysis of conserved MYCN binding regions
across four neuroblastoma lines at promoters and enhancers. (Right) Heat
map of conserved MYCN binding occupancy at E-box sequences. E-box
variants (CANNTG) ranked by MYCN ChIP-seq signal strength. The signal
strength relative to the canonical CACTG E-box is indicated by the shaded
color intensity E-box strength scale. (E) Meta track representation of MYCN
and H3K27ac ChIP-seq signal (rpm/bp) across four neuroblastoma cell lines
at the RPL22 locus. (F) Meta track representation of MYCN and H3K27ac
ChIP-seq signal (rpm/bp) across four neuroblastoma cell lines at an
upstream ID2 enhancer.
Figure 3.2 (Continued)
are re-capitulated by de novo motif finding at promoters and enhancers (Figure 3.2D) and are consistent with MYCN broadly associating with active regulatory elements in an affinity dependent manner, as previously observed for MYC (Lin et al., 2012). We conclude that when deregulated, MYCN binds strongly at canonical E-boxes found at promoters of active genes (e.g. RPL22; Figure 3.2E) and at clustered non-canonical E-boxes found at weaker promoters and distal enhancers (e.g. ID2; Figure 3.2F).

**MYCN is an amplifier of transcription in neuroblastoma**

We next investigated the dynamic chromatin and transcriptional consequences of direct MYCN shutdown. Using the controllable tet-off MYCN SHEP-21N cells, we rapidly and synchronously shut down MYCN with a 96% reduction achieved after only 6 hours (Figure 3.3A, 3.3B). Consequently we chose the 0hr, 2hr, and 24hr time points post shutdown to profile changes in a high, medium, and low MYCN state. In anticipation of global changes to chromatin and transcription that might be masked by relative chromatin enrichment methods, we employed the cell count normalized ChIP-Rx (Orlando et al., 2014) approach to quantify absolute changes in MYCN, H3K27ac, and RNA Pol II occupancy (supplemental methods).

We observe that 2hrs after shutdown, MYCN is already significantly depleted from promoters and enhancers — a result revealed only upon cell-count normalization (Figure 3.3C). A global view of chr21 shows a general loss of MYCN, H3K27ac, RNA Pol II occupancy 24 hours after shutdown (Figure 3.3D). At the exemplary MYC target gene *NPM1*, loss of MYCN from promoter regions
Figure 3.3: Dynamic chromatin and transcriptional changes of direct MYCN shutdown. (A) Western blot of MYCN protein levels upon MYCN shutdown in the SHEP21-N cell line. The percent of MYCN remaining versus 0hr is indicated and calculated based on pixel intensity quantification. (B) Schematic of the inducible MYCN shutdown system in SHEP-21N cells. (C) Boxplots of MYCN (red) and H3K27ac (blue) ChIP-seq signal at active promoters (top) and active enhancers (bottom). MYCN and H3K27ac signal at 0, 2, and 24 hours post MYCN shutdown for standard ChIP-seq, and ChIP-Rx before/after scaling. Significant differences are denoted (welch’s two tailed t-test): *** p-value < 1e-9. ** p-value <1e-6. (D) ChIP-Rx tracks of chromosome 21 at 0, 2, and 24 hours post MYCN shutdown. MYCN (red), H3K27ac (blue), and RNA Pol II (black) in units of scaled reads/basepair. (E) ChIP-Rx tracks of NPM1 locus at 0, 2, and 24 hours post MYCN shutdown. (F) Cell count normalized levels of the NPM1 transcript during MYCN shutdown. Units are fpkm for triplicate biological replicates.
Figure 3.3 (Continued)
initially results in a decrease of elongating RNA Pol II and subsequent loss of mRNA transcript, suggesting that like MYC, MYCN can directly promote transcription elongation (Figure 3.3E, 3.3F). Summarized across regulatory elements, we find that MYCN loss precedes histone acetylation decreases suggesting global chromatin compaction as a direct consequence of MYCN shutdown that is reminiscent of GCN5 mediated MYC induced global hyperacetylation (Knoepfler et al., 2006) (Figure 3.4A - 3.4C). Overall early defects in transcriptional pause release culminate in loss of active histone acetylation, and both initiating and elongating RNA Pol II (Figure 3.4D) resulting in a global dampening of mRNA steady state levels (Figure 3.4E).

*Enhancer invasion shapes transcriptional sensitivity to MYCN perturbation in neuroblastoma*

To better assess the direct contribution of MYCN to transcriptional changes at individual genes, we hypothesized a direct relationship between MYCN loading at proximal promoter/enhancer regions and resulting transcriptional response upon MYCN shutdown. Visual inspection of MYCN load at highly occupied genes revealed a diversity of binding profiles with some genes exhibiting promoter dominant MYCN load (*CDK4*, Figure 3.5A) and others a more mixed array of promoter and enhancer binding (*HAND2*, Figure 3.5B). Globally, increasing MYCN load at regulatory loci proximal to genes correlated with an increased contribution of MYCN enhancer binding suggesting that distal enhancers provide a reservoir for excess MYCN that is only accessed upon deregulation (Figure 3.5C, 3.5D). Indeed, 2 hours after MYCN shutdown, MYCN
Figure 3.4: MYCN is an amplifier of transcription in neuroblastoma.
(A) ChIP-Rx meta tracks of MYCN (red) and H3K27ac (blue) centered at promoters (top) and enhancers (bottom). ChIP-seq signal is centered on the TSS and enhancer (defined by peak H3K27ac signal within an enhancer) and extended 5kb in both directions. (B) Western blot of total histone H3 and H3K27ac levels at 0, 2, and 24 hours post MYCN shutdown. The percent H3K27ac remaining versus 0 hours is indicated. (C) ChIP-Rx meta tracks before scaling of MYCN (red) and H3K27ac (blue) centered at promoters (top) and enhancers (bottom). ChIP-seq signal is centered on the TSS and enhancer extended 5kb in both directions. (D) Pol II meta tracks across all active genes at 0 hours (black), 2 hours (blue), and 24 hours (red) post MYCN shutdown. The TSS and gene body is magnified as indicated and significance denoted (welch's two tailed t-test): *p-value < 1e-3. *** p-value < 1e-9. (Bottom right) Distribution plots of Pol II traveling ratios (TR) for all active genes. Differences in the TR distribution at 0HR and 2HR are significant (welch’s two tailed t-test): *** p-value < 1e-9. **p-value < 1e-6. (E) Box plots of log2 fold changes in active gene expression at the indicated time points versus 0HR post MYCN shutdown. Differences between 2 hours versus 8 hours and 4 hours versus 24 hours are significant (welch’s two tailed t-test): *** p-value < 1e-9. **p-value < 1e-6.
Figure 3.4 (Continued)
Figure 3.5: MYCN binding contributions at promoters and enhancers of active genes. (A) Meta track representation of MYCN and H3K27ac ChIP-seq signal (rpm/bp) across four neuroblastoma cell lines at the CDK4 locus. (B) Meta track representation of MYCN and H3K27ac ChIP-seq signal (rpm/bp) across four neuroblastoma cell lines at the HAND2 locus. (C) (Top) Histogram showing the correlation of ranked proximal MYCN signal (x-axis) versus total proximal MYCN signal at promoters (red) and enhancers (blue) of the top 5,000 genes ranked by proximal MYCN signal in the SHEP-21N line. (Bottom) Line plot showing the correlation of ranked proximal MYCN signal (x-axis) versus the percent enhancer contribution of total proximal MYCN signal (y-axis). (D) (Top) Histogram correlation plots and (bottom) correlative line plots as described in Figure 3.5C in the BE(2)-C cell line. (E) Boxplots of the log2 fold change (versus 0 hours) of MYCN occupancy at 2 and 24 hours post MYCN shutdown in the SHEP-21N system. Genes are grouped according to rank ordered MYCN signal at promoters (red) and enhancers (blue) and the top 1,000 sites respectively are shown. Significant differences are denoted (welch’s two tailed t-test): **p-value <1e-6.
Figure 3.5 (Continued)
binding sites at enhancers are preferentially depleted when compared to canonical E-box enriched promoter binding sites (Figure 3.5E). This preferential loss of MYCN corresponds to increased transcriptional sensitivity as measured by elongating RNA Pol II density and steady state mRNA, and overall we observe a strong dose dependence between MYCN load and transcriptional response (Figure 3.6A – 3.6C).

Enhancer regulated genes are often involved in developmental and tissue specific processes, normally considered outside the auspices of traditional MYC/MYCN regulated growth/proliferation pathways. MYC/MYCN enhancer invasion then could provide an explanatory mechanism for the ability of oncogenic MYC/MYCN to regulate diverse tumor specific processes. To test this hypothesis, we ranked the top 2,000 MYCN loaded genes by their relative promoter/enhancer MYCN contribution and performed leading edge functional analysis (Figure 3.6D, 3.6E). We find that promoter dominant genes statistically enriched for the classically defined core MYC target gene set (Dang et al., 2006; Ji et al., 2011; Kim et al., 2006; Schlosser et al., 2005; Schuhmacher et al., 2001; Zeller et al., 2003), whereas enhancer dominant genes were enriched for key signaling signatures in neuroblastoma including response to Tretinoin (retinoic acid), a pro-differentiating agent used to treat high-risk neuroblastoma (Matthay et al., 1999) (Figure 3.6D).

When this same analysis was performed on two models of MYC driven cancers (IgH/MYC translocated Multiple Myeloma, and MYC amplified small cell lung cancer), we again resolved increased enhancer contribution at highly
Figure 3.6: Enhancer invasion shapes transcriptional sensitivity to MYCN perturbation in neuroblastoma. (A) Log2 fold change (versus 0 hours) of gene expression changes during MYCN shutdown in the SHEP-21N system. Genes are grouped according to rank ordered MYCN proximal load (promoters and enhancers). The top 2,000 proximally MYCN bound genes are shown in red, medium ranked in black, and low in gray. (B) Box plots of the log2 fold change (versus 0 hours) of the amount of RNA Pol II at the TSS (left) and gene body (right) of genes grouped according to rank ordered MYCN proximal load. The top 2,000 proximally MYCN bound genes are shown in red, medium ranked in black, and low in gray. Significance is denoted (welch’s two tailed t-test): *** p-value < 1e-9. **p-value <1e-6. *p-value <1e-4. (C) Log2 fold change (versus 0 hours) of gene expression changes during MYCN shutdown in the SHEP-21N system. Genes are categorized by the MYCN signal at promoters versus enhancers: enhancer dominant (blue), promoter dominant (red), high enhancer and promoter (maroon) and other active genes without proximal MYCN signal (gray). (D) (Top) The MYCN signal contribution in the SHEP-21N line for promoters (red) and enhancers (blue) of associated genes (y-axis) of the top 2,000 proximal MYCN bound regions are shown ranked by relative MYCN enhancer/promoter contribution (x-axis). (Middle) GSEA plots of MYCN bound promoter (red) versus enhancer (blue) bias gene sets defined by leading edge analysis. (Bottom) The normalized enrichment score (NES) of target gene signatures (Molecular Signature Database) are plotted on the x-axis versus the FDR (false discovery rate) on the y-axis. Highly significant gene signatures from promoter (red) and enhancer (blue) bias gene sets are highlighted and tabulated. (E) MYCN signal contribution plots (top), GSEA plots (middle), and GSEA signature enrichment plots as described in Figure 3.5D in the BE(2)-C cell line.
Figure 3.6 (Continued)
occupied genes (Figure 3.7A, 3.7B). Enhancer/promoter analysis revealed the MYC core target gene set at the promoter dominant leading edge. However, in contrast to neuroblastoma, enhancer dominant MYC bound genes enriched for their respective tumor-specific pathways including chemokine and CD40 signaling in multiple myeloma (Aggarwal et al., 2006; Pellat-Deceunynck et al., 1994; Tong et al., 2000; Westendorf et al., 1994) and squamous cell carcinoma identity in small cell lung cancer (Herbst et al., 2008) (Figure 3.7C, 3.7D). These data suggest that enhancer invasion can account for the divergent and often puzzling consequences of oncogenic MYC/MYCN activation and provide a rationale for the tumor specific oncogenic phenotypes observed upon MYCN deregulation.

*TWIST1 co-occupies enhancers with MYCN and is a novel dependency in neuroblastoma*

If MYCN enhancer invasion accounts for the tumor specific consequences of MYCN deregulation, then the TFs that establish tissue specific enhancers likely define MYCN enhancer responsive gene expression programs. In any given cell type, a small set of tissue specific TFs are regulated by large “super-enhancer” regions, and themselves occupy the vast majority of active enhancers and super-enhancers (Hnisz et al., 2013; Whyte et al., 2013). Thus, a reverse analysis of TF motif occupancy at “super-enhancers” allows the inference of TFs that define enhancer landscapes. We applied a recently reported methodology developed by our lab to define the highly interconnected circuitry of enhancer
Figure 3.7: Enhancer invasion defines tumor-specific MYC target gene signatures. (A) (Top) Histogram showing the correlation of ranked proximal MYC signal (x-axis) versus total proximal MYC signal at promoters (red) and enhancers (blue) of the top 5,000 genes ranked by proximal MYC signal in the MM1.S line. (Bottom) Line plot showing the correlation of ranked proximal MYC signal (x-axis) versus the percent enhancer contribution of total proximal MYC signal (y-axis). (B) (Top) Histogram correlation plots and (bottom) correlative line plots as described in Figure 3.7A in the H2171 cell line. (C) (Top) The MYC signal contribution in the MM1.S line for promoters (red) and enhancers (blue) of associated genes (y-axis) of the top 2,000 proximal MYC bound regions are shown ranked by relative MYC enhancer/promoter contribution (x-axis). (Middle) GSEA plots of MYC bound promoter (red) versus enhancer (blue) bias gene sets defined by leading edge analysis. (Bottom) The normalized enrichment score (NES) of target gene signatures (Molecular Signature Database) are plotted on the x-axis versus the FDR (false discovery rate) on the y-axis. Highly significant gene signatures from promoter (red) and enhancer (blue) bias gene sets are highlighted and tabulated. (D) MYC signal contribution plots (top), GSEA plots (middle), and GSEA signature enrichment plots as described in Figure 3.7C in the H2171 cell line.
Figure 3.7 (Continued)
regulating TFs (Lin et al., 2016; Saint-André et al., 2016) (Figure 3.8A – 3.8D). The consensus circuitry (Figure 3.8E, 3.8F) includes several well established regulators of neuroblastoma identity including MEIS1/2, GATA2/3, PHOX2A/B and HAND2 (Geerts et al., 2003; Reiff et al., 2010; Tsarovina et al., 2004; Zha et al., 2014) that are not present in other unrelated tumor circuitries (Figure 3.8G).

From this analysis we also retrieve the TF TWIST1, which like HAND2, is a lineage specific bHLH TF with a well established role in promoting tumorigenesis in several human cancers (Entz-Werlé and Stoetzel, 2005; Kwok et al., 2005; Kyo et al., 2006).

Since bHLH TFs recognize E-box CANNTG motifs, we hypothesized that the clustered non-canonical E-boxes at enhancers invaded by MYCN might be co- or proximally occupied by TWIST1 and HAND2, and that these factors may collaborate to drive oncogenic enhancer transcription. Although TWIST1 has been implicated in MYCN amplified neuroblastoma (Valsesia-Wittmann et al., 2004), the factor has never been mechanistically investigated to fully explain its role as an oncogenic collaborator of MYCN. Performing ChIP-Seq for TWIST1 in the SHEP-21N and BE(2)-C lines, we observe strong overlap of MYCN and TWIST1 at enhancers and show by de novo motif finding that TWIST1 recognizes a CANNTG E-box similar to enhancer bound MYCN (Figure 3.9A).

Inspection of individual MYCN invaded enhancer loci reveal spatially proximal co-localization of MYCN/TWIST1 at clustered non-canonical E-box motifs (Figure 3.9B). Unexpectedly, MYCN shutdown depletes genomic bound TWIST1 with minimal change in TWIST1 protein levels, suggesting an important regulatory link
Figure 3.8: Enhancer regulatory core transcription factor circuitries.

(A) Network depiction of enhancer regulatory TF network in the SHEP-21N cell line. TF nodes are denoted and predicted binding interactions with other enhancer regulated TFs are shown as lines (edges). Enhancer regulated bHLH TFs are highlighted in red. (B) Network depiction of enhancer regulatory TF network in the BE(2)-C cell line. TF nodes are denoted and predicted binding interactions with other enhancer regulated TFs are shown as lines (edges). Enhancer regulated bHLH TFs are highlighted in red. (C) Network depiction of enhancer regulatory TF network in the NGP cell line. TF nodes are denoted and predicted binding interactions with other enhancer regulated TFs are shown as lines (edges). Enhancer regulated bHLH TFs are highlighted in red. (D) Network depiction of enhancer regulatory TF network in the KELLY cell line. TF nodes are denoted and predicted binding interactions with other enhancer regulated TFs are shown as lines (edges). Enhancer regulated bHLH TFs are highlighted in red. (E) Network depiction of conserved enhancer regulatory transcription factor network across four neuroblastoma cell lines. TF nodes are denoted and predicted binding interactions with super enhancers driving other TFs are shown as lines (edges). Enhancer regulated bHLH TFs are highlighted in red. (F) Heat map of enhancer regulated TFs in neuroblastoma cell lines (rows) clustered by similarity of regulatory degree. The bHLH TFs HAND2 and TWIST1 are highlighted in red. (G) Network depiction of enhancer regulatory TF network in the MM1.S cell line. TF nodes are denoted and predicted binding interactions with other enhancer regulated TFs are shown as lines (edges). Enhancer regulated bHLH TFs are highlighted in red.
Figure 3.8 (Continued)
Figure 3.9: TWIST1 co-occupies enhancers with MYCN and delineate sensitivity to MYCN perturbation. (A) (Left) Pie chart showing the genomic distribution of overlapping TWIST1 and MYCN bound sites. (Right) De novo motif analysis of overlapping TWIST1/MYCN enhancer regions for MYCN and TWIST1 binding sites. (B) Meta track representation of MYCN ChIP-seq signal, H3K27ac ChIP-seq signal, and ATAC-seq signal (rpm/bp) at the HAND2 locus. Corresponding ChIP-seq signal of TWIST1 in the BE(2)-C cell line is also shown. (C) Meta track representation of MYCN ChIP-seq signal, H3K27ac ChIP-seq signal, and ATAC-seq signal (rpm/bp) at the HAND2 locus. Below, ChIP-seq signal of TWIST1 at 0, 2, and 24 hours post MYCN shutdown. (D) Boxplot of background subtracted TWIST1 ChIP-seq signal (reads per million) at 0 hours, 2 hours and 24 hours post MYCN shutdown. Significance is denoted (welch's two tailed t-test): *** p-value < 1e-9. (E) Western blot of TWIST1 and MYCN protein levels upon MYCN shutdown in the SHEP21-N cell line. (F) Log2 fold change (versus 0 hours) of gene expression changes during MYCN shutdown in the SHEP-21N system. Genes are grouped by MYCN and/or TWIST1 binding: MYCN and TWIST1 co-bound sites (yellow), MYCN highly bound sites alone (black) and MYCN and TWIST1 lowly co-bound sites (gray).
Figure 3.9 (Continued)
Overall, genes with proximal and strong MYCN/TWIST1 occupied enhancers are more potently down-regulated by MYCN shutdown than genes associated with strong MYCN only (Figure 3.9F).

As a regulator of mesenchymal lineage and de-differentiated cell state in tumors, TWIST1 specifies target genes, but relies on other cues to enforce transcriptional activation or repression. To evaluate TWIST1 as a deregulated MYCN-specific dependency, we used clustered regularly interspaced short palindromic repeat (CRISPR) Cas9 mutagenesis scanning of the TWIST1 locus and its large downstream enhancers with a multiplexed library of 3,351 single guide RNAs (sgRNAs). Deep sequencing revealed a marked depletion of sgRNAs indicative of a decrease in cellular fitness in the 5' region of the TWIST1 locus and at TWIST1 regulatory enhancers that are bound by MYCN, TWIST, and are DNA accessible (ATAC-Seq) (Figure 3.10A, 3.10B). In contrast, CRISPR-Cas9 mutagenesis of the active chromatin regulator NSD1 had no effect on cellular fitness (Figure 3.10C, 3.10D). Importantly, as evidenced by siRNA knockdowns in multiple systems, knockdown of TWIST1 showed a marked decrease in viability in a MYCN deregulated context only (Figure 3.11A – 3.11D).

Discussion

When deregulated, MYCN dominates the active cis regulatory landscape of neuroblastoma to enforce both proliferation through promoter binding and de-differentiation through enhancer invasion. Genes disproportionately bound by MYCN either at promoters, enhancers, or a combination of both are selectively
Figure 3.10: TWIST1 co-occupies enhancers with MYCN and is a dependency in the MYCN-amplified BE(2)-C cell line. (A) CRISPR scan of the TWIST1 locus and its downstream enhancers. (Top) Illumina sequencing readout of log2 fold enrichment/depletion (early versus late time point) of 3,351 sgRNAs. (Bottom) Simple moving average of log2 fold enrichment/depletion is shown. (B) ChIP-seq signal of H3K27ac (blue), ATAC-seq (green), MYCN (red), and TWIST1 (yellow) at the TWIST1 locus and enhancers with respect to CRISPR sgRNAs. Red shaded boxes highlight regions of marked log2 fold depletion. (C) CRISPR scan of the NSD1 locus. (Top) Illumina sequencing readout of log2 fold enrichment/depletion (early versus late time point) of 947 sgRNAs. (Bottom) Simple moving average of log2 fold enrichment/depletion is shown. (D) ChIP-seq signal of H3K27ac (blue), ATAC-seq (green), MYCN (red), and TWIST1 (yellow) at the NSD1 locus with respect to CRISPR sgRNAs.
Figure 3.11: TWIST1 is a MYCN-specific dependency in neuroblastoma. (A) (Left) Western blot of TWIST1, MYCN, and Vinculin protein levels upon siRNA mediated knockdown of TWIST1 in the BE(2)-C cell line. (Right) Corresponding viable cell counts 72 hours post-siRNA transfection. Error bars denote ± standard deviation. (B) (Left) Western blot of TWIST1, MYCN, and Vinculin protein levels upon siRNA mediated knockdown of TWIST1 in the SH-SY5Y cell line. (Right) Corresponding viable cell counts 72 hours post-siRNA transfection. Error bars denote ± standard deviation. (C) (Left) Western blot of TWIST1, MYCN, and Vinculin protein levels upon siRNA mediated knockdown of TWIST1 in the MYCN ‘on’ state in the SHEP21-N cell line. (Right) Corresponding viable cell counts in the MYCN ‘on’ state at 72 hours post-siRNA transfection. Error bars denote ± standard deviation. (D) (Left) Western blot of TWIST1, MYCN, and Vinculin protein levels upon siRNA mediated knockdown of TWIST1 in the MYCN ‘off’ state in the SHEP21-N cell line. (Right) Corresponding viable cell counts in the MYCN ‘off’ state at 72 hours post-siRNA transfection. Error bars denote ± standard deviation.
down-regulated during MYCN shutdown indicative of an oncogenic gene expression program activated only during a MYCN deregulated state.

The observation that MYCN load potentiates MYCN dependent transcriptional responses may potentially resolve the disconnect between models of MYC proteins as transcriptional amplifiers or selective regulators. Instead of focusing only on the overall global transcriptional consequences of MYCN shutdown, or more narrowly at the most differentially regulated target genes, our results prompt a more quantitative consideration of MYCN as a non-linear amplifier of transcription that is shaped by E-box rich cis regulatory elements at promoters and enhancers. Whereas highly MYC/MYCN bound promoters are active in a majority of cell types, the enhancer dominant MYC/MYCN bound genes vary greatly between tumor types, suggesting that enhancers and the factors that form them are responsible for tumor specific MYC/MYCN response. In neuroblastoma, oncogenic collaboration of MYCN and TWIST1 at enhancers demarcates a set of developmental genes important to neuroblastoma tumorigenesis and highly sensitive to MYCN perturbation. In contrast, in the B-cell malignancies multiple myeloma that do not express HAND or TWIST factors, we instead identify TCF3 (E2A) as an E-box binding helix-loop-helix TF that may play an equivalent role (Figure 3.8G). These data implicate tissue specific E-box binding TFs and their co-factors in shaping oncogenic MYC/MYCN enhancer invasion. Importantly, as evidenced by the synthetic lethality of TWIST1 in MYCN deregulated neuroblastoma, they suggest targeting of the tumor enhancer apparatus as a therapeutic strategy to target MYC/MYCN deregulation in cancer.
Methods

Cell lines. SK-N-AS, SH-SY5Y, NGP, BE(2)-C, and KELLY cells were kindly provided by Dr. Kimberly Stegmaier (Dana Farber Cancer Institute) and cultured in Dulbecco’s Modified Eagle Medium (DMEM) supplemented with 10% FBS. SHEP-21N cells were kindly provided by Dr. William Weiss (University of California, San Francisco) and cultured in RPMI supplemented with 10% Tetracycline-Free FBS (Clontech). Tet-Off MYCN shutdown was performed by addition of doxycycline (0.2µg/mL) to growth media.

Immunoblotting. For whole cell lysates, cells were lysed using RIPA buffer supplemented with protease inhibitor cocktail (Roche) for 20 minutes. Lysates were clarified at 16,000 g for 15 minutes at 4°C and protein concentration was determined using BCA assay (Pierce). For cell normalized blots, cell numbers were determined using the Countess automated cell counter (Invitrogen) prior to RIPA lysis. A histone extraction was performed for histone blots on cell normalized pellets by acid extraction. Cells were lysed in Triton Extraction Buffer followed by an overnight acid extraction in 0.2N HCL overnight. Antibodies for western blot were purchased as follows: MYCN (Santa Cruz, sc-56729), Vinculin (Santa cruz, sc-25336), H3K27Ac (abcam, ab4729), H3 (Cell Signaling, 3638S), TWIST1 (abcam, ab50887). Blots were imaged using fluorescence-labeled secondary antibodies (LI-COR) on the OdysseyCLX Imager (LI-COR).
**RNA isolation.** Prior to RNA isolation, cell numbers were determined using the Countess automated cell counter Invitrogen. Total RNA isolation was performed using the miRvana miRNA total RNA isolation kit (ThermoFisher Scientific, AM1560) according to manufacturers instructions. Following isolation, RNA was digested with DNase (Ambion). During isolation, external RNA spike-ins (ERCC, Ambion) were added at the time of cell lysis. Total RNA was subject to polyA selection and adapter ligation in preparation for next-generation sequencing (Illumina stranded mRNA library prep) on Nextseq (75 basepair, single-end).

**siRNA transfection.** Neuroblastoma cells were reverse transfected (Lipofectamine RNAiMAX, Invitrogen) with 40 nM individual siRNA duplexes (Dharmacon). Protein knockdown was assessed 48 hours post transfection and viable cell count measurements were made 72 hours post transfection using the Countess automated cell counter (Invitrogen) in biological triplicate.

**RNA cell count microarray expression analysis.** Biotinylated RNA was prepared according to the standard Affymetrix protocol from 100 nanograms total RNA. Following fragmentation, 12.5 µg of RNA were hybridized at 45°C for 16 hours at 60 RPM on GeneChip Arrays (PrimeView). GeneChips were washed and stained in the Affymetrix Fluidics Station 450 according to the manufacturers instructions, using the buffers provided in the Affymetrix GeneChip hybridization, wash and stain Kit. GeneChips were scanned using the GeneChip Scanner 3000 and images were extracted with Affymetrix GeneChip Expression Console.
Chromatin immunoprecipitation (ChIP).

*Antibodies for ChIP.* Antibodies for ChIP were purchased as follows: MYCN (abcam, ab16898), TWIST1 (abcam, ab50887), H3K27Ac (abcam, ab4729), BRD4 (Bethyl, A301-985A), H3K4me3 (EMD Millipore, 07-473), RNA Pol II (Santa cruz, sc-899).

ChIP. Chromatin immunoprecipitations were performed as described with minor changes (Chapuy et al., 2013). Neuroblastoma cell lines were grown to 75% confluence in 15 cm plates and cross-linked with 1% formaldehyde (10 minutes) followed by quenching (125 mM glycine). Cells were washed in cold PBS and harvested by cell scraper in cold PBS with protease inhibitors (Roche). Cells were centrifuged at 1650 x g for 5 minutes and flash frozen and stored at -80°C. 50E06 cells per pellet. Pellets were resuspended in cytosolic then nuclear lysis buffer and DNA was sheared at 4°C using a waterbath sonicator (Bioruptor, Diagenode) for 25 minutes at high output (30” on, 30” off) in 1mL of sonication buffer supplemented with 0.5% SDS. Sonicated lysates were cleared by centrifuging at 20,000g for 10 min and incubated overnight end-over-end at 4°C with magnetic beads prebound with antibody. Beads were washed three times with sonication buffer, one time with sonication buffer with 500 mM NaCl added, one time with LiCl wash buffer (20 mM Tris pH 8.0, 1 mM EDTA, 250mMLiCl, 0.5% NP-40, 0.5% Na-deoxycholate) and once with TE. DNA was eluted in elution buffer (50 mM Tris-HCl pH 8, 10mM EDTA, and 1% SDS). Cross-links
were reversed overnight at 65°C. RNA and protein were digested with 0.2mg/mL RNase A for two hours followed by 0.2mg/mL Proteinase K for one hr. DNA was purified with phenol chloroform extraction and ethanol precipitation.

ChIP-Rx. ChIP-Rx was performed as described previously substituting mouse embryonic stem cells in place of Drosophila S2 cells (Orlando et al., 2014). Mouse embryonic stem (ES) cells were grown to 75% confluence and cross-linked with 1% formaldehyde followed by quenching (125 mM glycine) as described in the previous section. Cells were washed in cold PBS and harvested by cell scraper in cold PBS with protease inhibitors (Roche). Cells were centrifuged at 1650 x g for 5 minutes and flash frozen and stored at -80°C 10E06 cells per pellet. Fixed mouse ES cell pellets were resuspended in a cytosolic lysis buffer and then spiked directly into the cytosolic lysate of fixed neuroblastoma cells at a ratio of 5% of the total number of cells. The lysate (neuroblastoma cells spiked with mouse ES cells) was then carried through the ChIP protocol.

Library Preparation. Libraries for sequencing were prepared using the Rubicon Thruplex DNA-seq/FD library preparation kit. An input of 50 ng of DNA or less were used and following ligation libraries were amplified per manufacturers instructions. Amplified libraries were then size-selected using AMPure beads (Agencourt AMPure XP) per manufacturers instruction. Further size selection was performed using a 2% gel cassette in the Pippin Prep (SAGE Sciences) set to capture fragments between 200 – 700 base pairs. Libraries were multiplexed
at equimolar ratios and run together either on a HiSeq2000 (40 base pair, single end) or on a NextSeq (75 base pair, single-end).

**Assay for transposase-accessible chromatin (ATAC).** For each cell line, 50,000 cells were lysed for 10 minutes at 4°C in lysis buffer 10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% IGEPAL CA-360). After lysis, the pellets were subject to a transposition reaction (37°C, 60 minutes) using the 2X TD buffer and transposase enzyme (Illumina Nextera DNA preparation kit, FC-121-1030). The transposition mixture was purified using a Qiagen MinElute PCR purification kit. Library amplification was performed using custom Nextera primers and the number of total cycles determined by running a SYBR-dye based qPCR reaction and calculating the cycle number that corresponds to ¼ the maximum. Amplified libraries were purified using a Qiagen PCR purification kit and sequenced on a single lane of an Illumina NextSeq.

**CRISPR-Cas9 mutagenesis scanning.** Cas9 was stably expressed in the BE(2)-C cell line via lentiviral transduction (Addgene #52962). Cells were selected in 10µg of Blasticidin and expression checked by western blot (Cas9 antibody: EMD Millipore, MAC133).

sgRNAs were designed to target all possible PAM sequences on both the plus and minus strand of the TWIST1 and NSD1 regions as part of a larger library (total of 5,337 guides). Any sgRNAs that were predicted to align to more than
one unique region in the genome were excluded. Single stranded oligos were purchased from CustomArray, Inc. in pooled format, PCR amplified, and cloned by Gibson assembly into the U6-sgRNA-EFS-GFP vector (Addgene #65656). The ligation product was transformed into electrocompetent cells and quality of the library was evaluated via Illumina sequencing for proper representation.

For lentiviral packaging, HEK293T cells were transfected with pVSVg, psPAX2, and sgRNA using Lipofectamine 2000 (Invitrogen). Viral supernatants were harvested, filtered and concentrated via Lenti-X Concentrator solution (Clontech, 631232) following manufacturers instructions. Viral titer was calculated by a dilution series with measurements of GFP positivity (Guava Easycyte flow HT, Millipore). We aimed for one sgRNA per cell with a multiplicity of infection (MOI) of approximately 0.3 – 0.4. A total of 24E06 cells were transduced and maintained at all times during the screen to ensure at least 1000x coverage.

The genomic DNA was isolated at day 15 using the Blood and cell culture DNA maxi kit (Qiagen, 13362) following manufacturers instructions. Libraries were constructed as described previously (Shalem et al., 2014). A set of 60 PCR reactions were carried out to amplify the sgRNA cassette, pooled, and subsequently prepared for Illumina sequencing on a NextSeq 500.
The read counts for each individual sgRNA were calculated at an early time point (day 4) and a late time point (day 15). The log2 fold change of the early over the late time point was calculated.

**ChIP-seq data analysis**


*Genomic coordinates and gene annotation.* All coordinates and gene annotations in this study were based on human reference genome assembly hg19, GRCh37 (ncbi.nlm.nih.gov/assembly/2758/) and RefSeq genes.

*ChIP-seq data processing.* All datasets were aligned using Bowtie2 (version 2.2.1) to build version NCBI37/HG19 (Langmead et al., 2009). Alignments were performed using all default parameters except for –N 1. These criteria preserved only reads that mapped uniquely to the genome with one or fewer mismatches.
Calculating read density. We calculated the normalized read density of a ChIP-Seq dataset in any genomic region using the Bamliquidator (version 1.0) read density calculator (https://github.com/BradnerLab/pipeline/wiki/bamliquidator). Briefly, ChIP-Seq reads aligning to the region were extended by 200bp and the density of reads per base pair (bp) was calculated. 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).

Plotting composite representations of ChIP-seq signal at individual loci. To compactly display ChIP-seq signal at individual genomic loci and across MYCN ChIP-seq samples, we developed a simple meta representation as previously described (Lin et al., 2016). For all samples within a group, ChIP-seq signal is smoothed using a simple spline function and plotted as a translucent shape in units of rpm per bp. Darker regions indicate regions with signal in more samples. An opaque line is plotted and gives the average signal across all samples in a group.

Identifying enriched regions. We used the MACS version 1.4.1 (Model based analysis of ChIP-Seq) 67 peak finding algorithm to identify regions of ChIP-Seq enrichment over background (Zhang et al., 2008). A p value threshold of enrichment of 1e-9 was used for all datasets.
Defining conserved MYCN enriched regions. To define high confidence conserved MYCN binding regions across neuroblastoma cell lines, we first took the union of all MYCN bound regions across all cell lines (BE(2)-C, SHEP-21N, Kelly, and NGP). We next determined the MYCN binding rank of each region within a sample, and the average rank across samples. We determined that in any given cell line, MYCN regions in the top ~15,000 by rank appeared to show strong concordance across all cell lines. We established a simple cutoff of rank < 15,000 in 3 out of 4 cell lines to define conserved regions. After filtering out regions in the ENCODE blacklist (https://sites.google.com/site/anshulkundaje/projects/blacklists) and regions with > 20% cumulative content of LINE, Simple and LTR, 9,972 discreet conserved MYCN binding regions remained. These were used in all subsequent analysis.

Defining active genes. Across neuroblastoma cell lines active genes were defined as those with an enriched region in at least one cell line in the +/- 1kb region flanking the transcription start site (TSS). This resulted in 16,410 active genes defined in neuroblastoma.

Mapping typical enhancers and super-enhancers using H3K27ac enhancer definitions. H3K27ac super-enhancers (SEs) and typical enhancers (TEs) in individual neuroblastoma samples were mapped using the ROSE2 software package described in (Brown et al., 2014) and available at https://github.com/BradnerLab/pipeline - ROSE2_main.py. For each dataset, a
stitching parameter was determined that consolidated proximal peaks while optimizing the enriched fraction of stitched peaks. Across neuroblastoma cell lines, we identify a total of 43,869 discreet regions that exhibited enhancer activity in at least one cell line.

Creating heat map and meta representations of ChIP-Seq occupancy. Heat maps and meta plots of ChIP-Seq occupancy for various factors were created as in (Lin et al., 2012). Heat maps were created for all active promoters or all active enhancers as defined above. Each row plots the +/-5kb region flanking the TSS (for promoters) or the enhancer center. Rows are ranked by average occupancy of H3K27ac across all neuroblastoma cell lines for promoters and enhancers. Underlying metas represent the average signal for all rows in each bin (Figure 3.2A, 3.2B; Figure 3.4A, 3.4C).

Correlating MYCN binding with H3K27ac binding and other genomic features. Conserved MYCN regions were ordered by average rank across all samples and binned into 100 bins of size 100. In each bin, the average MYCN and H3K27ac occupancy was calculated and plotted (Figure 3.2C). Average overlap with genomic features (active promoters distal enhancers) was also calculated. Finally, the average density per kb of either canonical CACGTG or non-canonical CANNTG E-boxes was calculated. A best fit line was added using lowess regression. Error bars were calculated as the 95% confidence interval of the
mean as determined by 10,000 random samplings with replacement of values in each bin.

*Identifying MYCN and TWIST1 binding motifs via de novo motif discovery.* We used the DREME (Bailey, 2011) (Discriminative Regular Expression Motif Elicitation) program within the MEME suite to identify enriched sequence motifs in MYCN binding data. To create the MYCN input sequences, we used MACS 1.4.2 to identify the summit within each of the 9,972 conserved MYCN regions and took the flanking +/- 100bp sequence. For TWIST1, we used a similar methodology for all TWST1 regions that were bound in BE(2)-C and SHEP-21N. Default DREME parameters were used. For MYCN and TWIST1, the top scoring motifs are displayed in Figure 3.2D, 3.9A.

*Scaling cell count normalized ChIP-Rx datasets.* Scale factors for each ChIP-Rx dataset were calculated as in (Orlando et al., 2014) with the exception that mouse genomic reads were calculated instead of drosophila genomic reads. ChIP-Rx datasets were scaled by scale factors to create Y-axis arbitrary units scaled to normalize difference in occupancy per cell.

*Quantifying elongating RNA Pol II density in gene body regions.* Elongating RNA Pol II density was defined as the density of RNA Pol II in the +300 bp TSS to +3,000 bp gene end region as has been previously described (Rahl et al., 2010).
Creating meta-gene representations of RNA Pol II ChIP-Seq occupancy. Meta-gene representations of RNA Pol II density (Figure 3.4D) were determined by first binning gene sets into three regions: i) the upstream promoter - from 3kb upstream of the TSS to the TSS (50bp bins), ii) the gene body - from the TSS to the gene end (200 bins), iii) the transcription termination region (TTR) - from the gene end to +3kb downstream (50bp bins). Background subtracted ChIP-Seq density in each region was calculated and the average density in that region was plotted in units of average rpm/bp scaled by ChIP-Rx scale factor. The differences in relative ChIP-Seq density in gene regions were tested for statistical significance using a Welch’s two-tailed t test.

Calculating RNA Pol II traveling ratio. We determined the ratio of background subtracted RNA Pol II ChIP-Seq levels in initiating to elongating regions (termed traveling ratio (Rahl et al., 2010) at each time point in the SHEP-21N system. We defined the initiating region as ±300 bp around the TSS. We defined the elongating region as +300 bp from the TSS to +3,000 bp after the gene end. In order to make higher confidence comparisons, we limited our analysis to genes with detectable signal above noise in the initiating and elongating regions across all samples. The statistical significance of changes in the distribution of traveling ratios was determined using a Welch’s two-tailed t test.

Quantifying contribution of enhancer or promoter MYCN or MYC binding at individual genes. To quantify MYCN binding at either the promoter or enhancer of
individual genes, we first split the 9,972 conserved MYCN binding sites into either TSS proximal (within the +/-1kb region flanking the TSS) and enhancer distal (sites outside of the TSS proximal region). For each gene, the area under curve (AUC) of MYCN binding in the TSS proximal region was quantified as was the AUC of all distal MYCN sites found within 50kb of the promoter. Genes were ranked by their total MYCN contribution (enhancer + promoter) and the MYCN enhancer contribution percentage was calculated using a simple moving average (100 genes per bin with a 50 gene increment). A best fit line was applied by lowess regression (Figure 3.5C, 3.5D). A similar approach was applied for MYC binding (Figure 3.7A, 3.7B).

Leading edge analysis based on promoter/enhancer bias. To determine the functional pathways enriched in MYCN promoter or enhancer dominant genes we first took the top 2,000 genes as quantified by total MYCN contribution (promoter +enhancer). Genes were ranked by the net difference in promoter-enhancer MYCN AUC signal (units of total rpm). This net difference was then used in GSEA leading edge analysis to identify functional pathways enriched on either the promoter or enhancer dominant edge. GSEA was performed using the parameters “-collapse false -mode Max_probe -norm meandiv -nperm 1000 -permute gene_set -rnd_type no_balance -scoring_scheme weighted -metric Diff_of_Classes -sort real -order descending -include_only_symbols true -make_sets true -median false -num 100 -plot_top_x 20 -rnd_seed timestamp”. The MSigDB C2 collection of curated gene sets was used as the primary query.
For each analysis, the negative enrichment score (NES) and normalized false discovery rate (FDR) were plotted for all interrogated datasets to identify gene sets with a strong leading edge enrichment and a FDR < 0.1 (Figure 3.6D, 3.6E). Similar analysis was performed to quantify MYC contribution in other cancer cell lines (Figure 3.7C, 3.7D).

**Defining enhancer core regulatory transcription factor circuitry.** Neuroblastoma core regulatory circuitry analysis was performed using the COLTRON software package as described previously (Lin et al., 2016) and found at: https://pypi.python.org/pypi/coltron. To quantify the interaction network of TF regulation, we calculated the IN and OUT degree of all SE associated TFs. For any given TF (TFi), the IN degree was defined as the number of TFs with an enriched binding motif at the proximal SE of TFi. The OUT degree was defined as the number of TF associated SEs containing an enriched binding site for TFi. To calculate TF binding edge interactions within any given SE, enriched TF binding sites were determined at putative nucleosome free regions (valleys) flanked by high levels of H3K27ac and the presence of an ATAC-seq peak in any neuroblastoma cell line. In these regions, we searched for enriched TF binding sites using the FIMO (Grant et al., 2011) algorithm with TF position weight matrices defined in the TRANSFAC database (Matys et al., 2006). An FDR cutoff of 0.01 was used to identify enriched TF binding sites. Using this approach, we calculated IN and OUT degree for all SE associated TFs for each neuroblastoma cell line. SE associated TFs with a unit normalized total degree (IN + OUT) of
>0.75, indicating that the TF is in the top quartile of regulatory influence, were included in subsequent analysis. For each neuroblastoma cell line, the interactions of all SE associated TFs with a normalized total degree >0.75 are shown in Figure 3.8A – 3.8D. Across neuroblastoma cell lines, the total degree for all SE associated TFs with a normalized total degree >0.75 are hierarchically clustered using a Euclidean distance metric to illustrate groups of TFs with similar regulatory patterns (Figure 3.8F). TFs with a normalized total degree of at least 0.75 in 3/4 cell lines were considered part of the conserved neuroblastoma network. Regulatory interactions between these TFs are displayed in Figure 3.8E.

**ATAC-seq data analysis.** All paired-end datasets were aligned using Bowtie (version 2.2.1) to build version NCBI37/HG19 with the following parameters: --end-to-end --sensitive --no-unal --no-discordant --mm --met-stderr. Aligned bams were filtered and sorted using samtools by removing chrM, filtering against the ENCODE blacklist, and removing discordant reads. Peaks were called using MACS1.4 with p-val =1e-9. For enhancer core regulatory circuitry analysis, the union of all ATAC enriched regions in BE(2)-C, SHEP-21N, NGP, and Kelly was used to define putative TF binding sites.

**RNA-seq data analysis.** Fastq files were aligned to hg19 using HiSat with default parameters. Transcripts were assembled and FPKM values were generated using cuff quant and cuffnorm from the cufflinks pipeline (Trapnell et
al., 2010). FPKM values were then normalized to synthetic ERCC spike-in RNAs. Active transcripts were defined as transcripts with a normalized FPKM value greater than 0.1.

**RNA cell count microarray expression analysis.** Microarray expression data were normalized as in (Brown et al., 2014) to produce cell count normalized expression values in arbitrary units. A CDF provided by Affymetrix, which contained the ERCC probes (PrimeView_withERCC_binary.cdf) was used instead of the standard PrimeView CDF. The data were analyzed with Bioconductor using the MAS5 normalization method.
Chapter 4: Overcoming BET bromodomain inhibitor resistance with BET degradation in triple negative breast cancer

Rhamy Zeid¹, Sarah Vittori¹, Justin M. Roberts¹, Jaime Reyes¹, Georg E. Winter¹, Matthew Lawlor¹, Thomas Scott¹, Dennis L. Buckley¹, Jun Qi¹ & James E. Bradner¹,²

¹Medical Oncology, Dana Farber Cancer Institute (DFCI), Boston, Massachusetts 02115, ²Department of Medicine, Harvard Medical School, Boston, MA.

Author Contributions

RZ: Planned and performed in vitro experiments not attributed to others, performed basic computational analysis, wrote the manuscript, and with JEB planned the project.

SV & TS: Performed/assisted on western blots

JMR: Assisted in cell viability assays

JR: Performed computational analysis for RNA-seq and ChIP-seq datasets.

GEW: Assisted in 72 hour dose response cellular viability

ML: Assisted on ChIP-seq (immunoprecipitations and library prep)

DB: Synthesized and provided all targeted BET degrader compounds

JQ: Synthesized and provided JQ1

JEB: Supervised and developed the project with RZ.
Introduction

Like conventional chemotherapeutics, the clinical efficacy of targeted cancer agents is limited by the rapid emergence of drug resistance. It has become increasingly clear that identifying and characterizing mechanisms of resistance while developing targeted agents are critical to predicting long-term clinical efficacy (Holohan et al., 2013). Moreover, the mechanisms of resistance established for previous targeted therapies can serve as a guiding rationale to understanding mechanisms of resistance to small molecule inhibitors of diverse molecular targets. For example, mechanisms of resistance to small molecule kinase inhibitors have illustrated the distinction between intrinsic and adapted resistance (Longley and Johnston, 2005). The mechanisms of resistance observed in response to the small molecule inhibitor imatinib range from gatekeeper mutations in the BCR-ABL (T315I) ATP-binding pocket (intrinsic) to alternative pathway activation via overexpression of LYN kinase (adapted) (Branford et al., 2002; Donato et al., 2003; Gamas et al., 2009; Lamontanara et al., 2013; Rix et al., 2007). These two distinct mechanisms have dictated parallel strategies to overcome resistance. The intrinsic resistance to imatinib prompted the development of small molecules capable of inhibiting the gatekeeper-mutant form of BCR-ABL (O'Hare et al., 2009). Alternatively, adapted resistance to imatinib can be overcome using later generation BCR-ABL inhibitor such as dasatinib that also inhibit LYN (Bantscheff et al., 2007; Rix et al., 2007). Thus, as new targets are identified and small molecule inhibitors are developed, it will be imperative to proactively understand potential mechanisms of resistance to guide
alternative strategies to overcome molecularly distinct resistant phenotypes when they arise.

In recent years, the discovery and characterization of small molecule inhibitors of the bromodomain and extra terminal protein (BET) family have uncovered a promising new class of targeted therapy (Chung et al., 2011; Filippakopoulos et al., 2010; Nicodeme et al., 2010). The BET family is comprised of four proteins: BRD2, BRD3, BRD4, and BRDT. BET family proteins harbor a pair of bromodomains with deep hydrophobic pockets that mediate acetyl-lysine binding via a conserved asparagine residue (Filippakopoulos et al., 2012; Owen et al., 2000). Small molecule BET inhibitors compete with this acetyl-lysine binding pocket and selectively displace BET proteins from chromatin. BET inhibitors have shown efficacy in several models of cancer ranging from hematologic malignancies to solid tumor malignancies (Bandopadhayay et al., 2014; Dawson et al., 2011; Delmore et al., 2011; Mertz et al., 2011; Ott et al., 2012; Puissant et al., 2013; Shimamura et al., 2013; Zuber et al., 2011). Furthermore, BRD4 has been identified as a putative non-oncogene dependency in Acute Myeloid Leukemia (AML) (Zuber et al., 2011). The efficacy of small molecule inhibition has sparked intense interest in pursuing BET family inhibitors as a novel therapeutic strategy in several cancer types.

The availability of small molecule inhibitors has facilitated a growing mechanistic understanding of the critical role of these acetyl-lysine 'reader' proteins in transcription. The BET family, and BRD4 in particular has been well characterized as a chromatin cofactor that binds acetylated lysines on histone
tails and subsequently recruits core components of the transcriptional apparatus (LeRoy et al., 2008; Rahman et al., 2011). In this sense, BET family proteins such as BRD4 act as a critical bridge in the interplay between regions of open (acetylated) chromatin and active transcription. More specifically, BRD4 recruits the positive elongation factor (P-TEFb) to drive productive transcription (Bisgrove et al., 2007; Wu and Chiang, 2007; Yang et al., 2008; Yang et al., 2005). Several studies have illuminated the critical role of BRD4 at large cell-type specific enhancers (termed super enhancers, SEs) that drive transcription of target genes and can globally underlie a malignant transcriptional program (Chapuy et al., 2013; Lovén et al., 2013). In particular, BET inhibition has been shown to displace BRD4 from super enhancers that drive critical oncogenes resulting in the selective down regulation of transcriptional output. For instance, BET inhibition displaces BRD4 from super enhancers driving oncogenic transcription of the MYC family of transcription factors in various MYC family driven cancers (Lovén et al., 2013; Puissant et al., 2013).

The availability of potent and selective small molecule inhibitors of the BET family combined with a growing mechanistic understanding of the role of BET proteins in cancer has led to rapid translational applicability. Indeed, the first clinical study has reported single-agent activity in hematologic malignancies with several additional clinical trials ongoing (Boi et al., 2015; Odore et al., 2016). While these initial clinical results show great promise, resistance is thought to be inevitable and therefore the identification and characterization of mechanisms of resistance will quickly become critical to maintaining clinical efficacy.
Recently, several studies have characterized *in vitro* and *in vivo* models that demonstrate resistance to BET inhibitors in an effort to predict clinically relevant mechanisms of resistance (Fong et al., 2015; Rathert et al., 2015; Shu et al., 2016). These studies have presented two different mechanisms of BET inhibitor resistance. The first model suggests that in leukemia cells, BET inhibitor resistance is mediated by a downstream increase in Wnt/β-catenin signaling that serves as an alternative path of transcriptional activation. (Fong et al., 2015; Rathert et al., 2015) In primary mouse hematopoietic stem and progenitor cells (HSPCs) ectopically expressing the oncoprotein fusion MLL-AF9, BET inhibitor resistant cells were generated by dose-escalation using I-BET and resistant clones growing in up to 1 micromolar (µM) were isolated (Fong et al., 2015).

These cells were resistant to structurally distinct BET inhibitors such as JQ1 as well as genetic (shRNA mediated) knockdown of BRD4. In the resistant cells, it was noted that Wnt/β-catenin pathway genes were differentially upregulated and appeared to reactive key BRD4 bound genes such as Myc. In a parallel study, it was observed that shRNA mediated knockdown of the Polycomb Repressive Complex 2 (PRC2) in cells derived from an AML mouse model conferred resistance to BET inhibition (Rathert et al., 2015). Specifically, suppression of the PRC2 subunit Suz12 led to a global loss of histone H3 lysine 27 trimethylation (H3K27me3) and promoted resistance to nanomolar (nM) concentrations of the small molecule inhibitor JQ1. Here, it was shown that inactivation of PRC2 led to the upregulation of genes associated with Wnt signaling and that Wnt signaling was responsible for the reactivation of key oncogenes such as Myc. Thus, these
two parallel studies converge on the activation of Wnt signaling as mechanism of adapted resistance to BET inhibitors.

The second model suggests that BET inhibitor resistance is mediated through the bromodomain-independent re-localization of BRD4 to chromatin (Shu et al., 2016). In cell line models of triple negative breast cancer (TNBC), BET inhibitor resistant cells grew in micromolar concentrations of the small molecule inhibitor JQ1 but were still sensitive to shRNA mediated knockdown of BRD4. Interestingly, in these resistant cells BRD4 was not displaced from chromatin while JQ1 was still present at effective doses within cells. It was shown that the re-localization of BRD4 to chromatin in the presence of BET inhibitors was mediated by an increase in phosphorylated BRD4 that in turn resulted in increased binding to mediator complex subunit 1 (MED1). Thus, this allowed BRD4 to relocalize on chromatin in a bromodomain independent manner via MED1.

These studies have identified and characterized two distinct mechanism of BET inhibitor resistance in models of leukemia and triple negative breast cancer. In turn, these two models characterize two distinct mechanisms of resistance that enable the immediate development of strategies to overcome these types of resistance should they arise clinically. Indeed, both studies provide preliminary strategies to overcome BET inhibitor resistance. In leukemia resistance models, the Wnt signaling pathway inhibitor pyrvinium was used in combination with BET inhibition to show an efficacious effect (Fong et al., 2015). In TNBC resistance models, the phenothiazine (PTZ) activator compound was used in combination
with BET inhibition to drive PP2A phosphatase activity resulting in decreased levels of phosphorylated BRD4, resulting in improved efficacy (Shu et al., 2016). These strategies offer preclinical evidence for a combinatorial approach to overcome BET inhibitor resistance based on the detailed characterization of the mechanisms of resistance.

To overcome BET inhibitor resistance with a single agent, we use a newly developed class of small molecules to pharmacologically induce targeted degradation of the BET family. We recently innovated a strategy to induce potent and selective degradation of the BET family proteins via phthalimide conjugation for E3 ubiquitin ligase recruitment (Winter et al., 2015). This all-chemical strategy of targeted protein degradation is similar to natural chemical inducers of E3 ligase heterodimerization such as jasmonate (Thines et al., 2007). We and others have shown that these bifunctional compounds effectively dimerize target ligands with E3 ubiquitin ligases that result in ubiquitination and subsequent proteasomal degradation (Lu et al., 2015; Winter et al., 2015). In our recent study, we developed a bifunctional molecule (termed dBET1) that simultaneously binds the bromodomain of BET proteins and the cereblon (CRBN) E3 ligase (Winter et al., 2015). The enforced proximity of CRBN to BET proteins results in the rapid, potent, and selective degradation of all members of the BET family. Using a chemically improved BET protein degrader (termed dBET6), we show that BET family degradation overcomes BET inhibitor resistance in triple negative breast cancer. Using BET inhibitor resistant TNBC models, we show that BET family degradation with dBET6 has an efficacious effect in in vitro models. We extend
these results to other cell line models that exhibit intrinsic resistance to BET inhibitors and collectively provide a rationale for BET family degradation as a means to overcome emerging BET inhibitor resistance.

**Results**

*Targeted BET family protein degradation in TNBC*

We recently undertook efforts to develop and optimize a chemical series of BET family protein degraders based on our prototype molecule dBET1 resulting in the identification of dBET6 as a highly potent and selective small molecule degrader with improved cellular permeability (Winter et al., 2016). Moreover, dBET6 had much improved efficacy in degrading BET family proteins in several types of human cancer cell lines most notably T-cell acute lymphoblastic leukemia (Winter et al., 2016). dBET6 is a bifunctional molecule consisting of the small molecule inhibitor JQ1 appended to phthalimide-conjugates via a longer linker as compared to dBET1 (Figure 4.1A). To validate a BET degrader chemical series in a panel of triple negative breast cancer cell lines, we first compared the impact of each of these small molecules on cell viability by surrogate measurements of ATP levels at 72 hours (Figure 4.1B). In agreement with our previous characterization, dBET6 had a significantly superior effect on cell viability. In several of the lines, dBET6 had an significantly improved effect on cell viability as compared to the BET family inhibitor JQ1. In the SUM159 line, dBET6 treatment displayed a shift in the half maximal inhibitory concentration (IC$_{50}$) as well an increase in maximum response (E$_{max}$) (Figure 4.1C). The increased potency and maximum response suggests that BET family
Figure 4.1: Targeted BET degradation in TNBC. (A) Chemical structures of JQ1, phthalimide, dBET1, and dBET6. (B) Heatmap of cellular viability as measured by surrogate levels of ATP content. Results of 10-point dose response after 72 hours of incubation are represented by area under curve (AUC). Experiments were performed in quadruplicates and dBET6 is highlighted in red. (C) Dose response curves of JQ1 and dBET6 in the SUM159 line. Normalized viability as measured by ATP concentrations (y-axis) against 10-point dose of compound (x-axis). (D) Western blot of BRD2, BRD3, and BRD4 levels in the SUM159 line after 4 hours of incubation with JQ1 and dBET6.
degradation may provide additional efficacy as compared to BET inhibition in TNBC cell lines. We surveyed the extent of BET family degradation in the SUM159 line and found that dBET6 potently degraded BRD2, BRD3, and BRD4 in dose response pattern (Figure 4.1D). Interestingly, excess concentrations of dBET6 (5µM) fail to degrade BET proteins due to the bifunctional nature of the molecule (termed the ‘hook effect’). In contrast, JQ1 inhibition had no effect on total levels of BET family proteins. Thus, we conclude that BET degradation has an improved efficacious effect as compared to BET inhibition and establish dBET6 as a superior BET family degrader in TNBC.

**BET degradation overcomes BET inhibitor resistance in TNBC**

Given the increased efficacy of BET family degradation as compared to BET family inhibition, we wondered how BET family degradation might affect viability in BET inhibitor resistant cells. We recently established and characterized BET inhibitor resistant clones of the SUM149 and SUM159 TNBC cell line (Shu et al., 2016). These lines (termed SUM149R and SUM159R) reliably grow in micromolar concentrations of the small molecule BET inhibitor JQ1. While these cells are no longer sensitive to BET inhibition due to bromodomain independent recruitment to chromatin, BRD4 remains essential as shown by sensitivity to genetic knockdown (Shu et al., 2016). Thus, we hypothesized that targeted degradation of BET proteins would phenocopy the genetic effects on viability of BRD4 knockdown and could thus overcome...
Figure 4.2: BET degradation overcomes BET inhibitor resistance in TNBC

(A) (Left) Dose response curves of JQ1 and dBET6 in the SUM149R. Normalized viability as measured by ATP concentrations (y-axis) against 10-point dose of compound (x-axis). (Right) Viable cell numbers after JQ1 and dBET6 treatment respectively. (B) Western blot of BRD2, BRD3, and BRD4 levels after 4 hours of incubation with JQ1 and dBET6. (C) (Left) Dose response curves of JQ1 and dBET6 in the SUM159R. Normalized viability as measured by ATP concentrations (y-axis) against 10-point dose of compound (x-axis). (Right) Viable cell numbers after JQ1 and dBET6 treatment respectively. (D) Western blot of BRD2, BRD3, and BRD4 levels after 4 hours of incubation with JQ1 and dBET6.
Figure 4.2 (Continued)
BET inhibitor resistance. In acute 72 hour exposure to dBET6, both the SUM149R and SUM159R lines were significantly more sensitive as compared to JQ1 (Figure 4.2A, 4.2B). Here, dBET6 elicited a shift in both the IC50 as well as the Emax. In long-term cultures, the SUM149R and SUM159R lines were extremely sensitive to varying concentrations of dBET6, while BET inhibition only conferred mino impairments of cellular viability. Interestingly, the SUM149R line appears to grow faster in the presence of JQ1 as we have previously reported (Shu et al., 2016). We again surveyed the extent of BET family degradation in these resistant cells and observed complete degradation of BRD2, BRD3, and BRD4 (Figure 4.2A, 4.2B, left). Taken together, these data show that BET degradation is capable of overcoming BET inhibitor resistance in two resistant TNBC cell lines.

*BET degradation results in a global reduction of transcription*

To mechanistically understand the ability of BET degradation to overcome BET inhibitor resistance, we took an integrative genomic approach combining chromatin localization data with transcriptional profiling. We first profiled the kinetics of BET family degradation at a fixed concentration of dBET6 to identify suitable time points to look for dynamic molecular changes. Within one hour, we had observed significant degradation of BRD4 (Figure 4.3A). We identified time points at two and six hours post dBET6 incubation to profile initial and later stage effects of BET degradation. In the SUM149R resistant line, we profiled the genome-wide localization of BRD4 upon dBET6 treatment after 2 hours of incubation. Here, we compared the basal resistant state of the SUM149R line
Figure 4.3: BET degradation results in global reduction in transcription in the SUM149R cell line (A) Western blot of BET family protein levels with dBET6 (250 nM) in the SUM149R cell line. (B) ChIP-seq tracks of BRD4 levels on chromosome 20 in the SUM149R cell line. The top track (black) shows the basal resistance state in the presence of 10µM of JQ1 and the bottom track (red) reflects 2 hours of incubation with 250 nM dBET6. (C) Boxplot of the global levels of chromatin-bound BRD4 in 10µM of JQ1 versus 250 nM dBET6 for 2 hours. BRD4 ChIP-seq signal is in units reads per million (RPM). (D) Heatmap of log₂ fold changes in gene expression upon treatment with 250 nM of dBET6 at 2 and 6 hours. Expression values were normalized to ERCC spike-ins to allow for cell count normalized quantification. (E) Expression levels of all active genes ranked by their expression in the presence of JQ1 at 2 hours (left) and 6 hours (right) of incubation with dBET6. (F) Western blot of RNA Pol II phosphorylation levels at the indicated time points of incubation with dBET6 at 250 nM.
Figure 4.3 (Continued)
(grown indefinitely in 10µM of JQ1) versus incubation with dBET6 (250 nM).

Upon BRD4 degradation, we observed a global loss of BRD4 from chromatin as exemplified by a high level view of chromosome 20 (Figure 4.3B). We quantified the global loss of BRD4 upon dBET6 treatment and found that BRD4 was displaced from the majority of chromatin (Figure 4.3C). Thus, the rapid degradation of BRD4 after 2 hours is sufficient to eliminate BRD4 localized chromatin. To understand the effects of rapid BRD4 degradation and subsequent loss from chromatin, we made genome-wide measurements of steady state levels of mRNA at two and six hours of treatment with dBET6. At both two and six hours of dBET6 incubation, we observed a widespread decrease in steady state mRNA levels across the majority of active genes, which was more pronounced at the six hour time point (Figure 4.3D 4.3E).

Given the rapid and global effects on transcription, we investigated the effects on RNA Polymerase II (RNA Pol II) upon BET family degradation. It has been previously illustrated that BRD4 plays a key role as a transcriptional co-activator by recruiting the positive elongating factor (P-TEFb) to liberate paused RNA Polymerase II (Wu and Chiang, 2007; Yang et al., 2008; Yang et al., 2005). Thus, we hypothesized that the rapid degradation of BRD4 resulted in a loss of elongated RNA Pol II. Indeed, we observed a decrease in serine two phosphorylated RNA Pol II within four hours of BET degradation, indicative of a loss in elongation (Figure 4.3F). At 24 hours, we observed a more general defect in total Pol II levels as well as serine two and five phosphorylated Pol II (Figure 4.3F). Thus, BET degradation results in the rapid degradation of BRD4-bound
chromatin that in turn causes a global reduction in active gene transcription mediated in part by the loss of initiating and elongated RNA Pol II.

*BET degradation overcomes intrinsic resistance to BET inhibitors in various cancer cell line models*

Thus far we have explored the ability of BET degradation to overcome BET inhibitor resistance mediated by epigenetic (adapted) mechanisms in TNBC cell lines. In the rapidly expanding literature of BET family inhibition, there have also been several reported instances in which various cancer cell types exhibit a basal/intrinsic resistance (Basu et al., 2013). We set out to investigate if BET degradation could be deployed in order to overcome these intrinsic resistance mechanisms. Encouragingly, the estrogen receptor (ER) positive TNBC cell line T47D was rather insensitive to BET inhibition but proved to be sensitive to BET degradation with dBET6 (part of our initial dataset, Figure 1.1B) (Shu et al., 2016). To compliment and extend the validity of BET degradation as a strategy to overcoming intrinsic resistance to BET inhibitors, we profiled BET degradation in a panel of different cancer cell line models. We curated a subset of cell lines from a variety of solid tumor types that have been reported has largely insensitive to BET inhibition (Basu et al., 2013). We then compared the effects of BET inhibition versus BET degradation on cell viability by surrogate measurements of ATP levels at 72 hours (Figure 4.4A). We generally observed a more pronounced effect on cell viability with dBET6 similarly to what we observed in TNBC cell lines.
We focused on the MYCN-amplified neuroblastoma cell line NGP due to the marked difference in viability effect with BET degradation. In dose response, dBET6 exhibited a marked increase in potency versus JQ1 with IC$_{50}$s of 1.8µM

Figure 4.4: BET degradation overcomes intrinsic resistance to BET inhibition in the NGP cell line. (A) Heatmap of cellular viability as measured by surrogate levels of ATP content. Results of 10-point dose response after 72 hours of incubation are represented by area under curve (AUC). Experiments were performed in quadruplicates and NGP line is highlighted in red. (B) Dose response curves of JQ1 and dBET6 in the NGP line. Normalized viability as measured by ATP concentrations (y-axis) against 10-point dose of compound (x-axis). (C) Western blot of BRD2, BRD3, BRD4, and MYCN levels in the NGP line in response to incubation with 250 nM dBET6 at the indicated time points. (D) Relative MYCN mRNA transcript levels in response to incubation with 250 nM dBET6 at the indicated time points. Error bars represent standard deviation of four replicates.
and 187 nM respectively (Figure 4.4B). In addition, dBET6 also increased the maximum response versus JQ1 with $E_{\text{max}}$ values of 58% and 22% respectively. We previously reported the effects of BET inhibition in MYCN-amplified neuroblastoma and showed that BET inhibition resulted in the down regulation of amplified MYCN protein and transcript leading to an efficacious phenotype (Puissant et al., 2013). However, we were surprised to observe that the NGP line in particular was insensitive to BET inhibition and in addition genetic knockdown of BRD4 had no effect on viability. Here, we measured the extent of BET family degradation with dBET6 in the NGP line and observed protein degradation of BRD2, BRD3, and BRD4 within 1 hour of treatment with dBET6 (Figure 4.4C). Furthermore, we also observed an initial reduction in MYCN protein and mRNA levels upon BET family degradation (Figure 4.4C, 4.4D). Interestingly, MYCN protein and mRNA transcript levels were restored at later time points of dBET6 incubation accompanied by a slight increase in BRD2 protein levels. The abrupt loss of MYCN that accompanies BET family degradation may contribute to a mechanism by which BET family degradation reduces cellular viability in the NGP line. Taken together, these data suggest that BET degradation may be capable of overcoming intrinsically mediated BET inhibitor resistance in the NGP cell line.

**Discussion**

The growing body of literature reporting the efficacy of BET inhibition in several cancer types combined with initial success in a clinical setting have motivated the identification and characterization of mechanisms of resistance. It is likely that like other targeted therapies, BET inhibitor resistance will likely
emerge in the clinic and it is critical that we work to understand, predict, and combat forms of resistance. To this end, we recently reported a epigenetic-based mechanism of resistance to BET inhibitors in triple negative breast cancer (Shu et al., 2016). We isolated pairs of sensitive and resistance cell lines in which resistance lines grew in micromolar concentrations of BET inhibitor.

In an effort to develop strategies to overcome BET inhibitor resistance, we report on the efficacy of the bifunctional small molecule dBET6 that selectively degrade BET family proteins. In TNBC, BET degradation had a marked increased efficacy across several types of TNBC cell lines. Moreover, in models of resistance, nanomolar concentrations of dBET6 drastically reduced viability in BET inhibitor resistant lines. Interestingly, the transcriptional effect in response to BET degradation is markedly different than that of BET inhibition. While BET inhibition has a more selective effect, BET degradation appears to have a more global effect causing widespread reduction in transcription of active genes. The discrepancy in transcriptional response observed may reflect in part the difference in kinetics between BET inhibition and BET degradation. It may also be partially due to the ability of dBET6 to simultaneously degrade all members of the BET family of proteins. dBET6 leads to the rapid degradation of BET family proteins (on the order of hours) and in turn transcriptional effects are observed within 2 hours of incubation.

In addition to BET inhibitor resistance in TNBC, we also suggest that targeted BET degradation may serve as single agent strategy to overcoming other forms of BET inhibitor resistance. In a panel of cell lines that are ntrinsically
resistant to BET inhibition, BET degradation again proved efficacious. In neuroblastoma lines, BET degradation led to a reduction in the MYCN oncogene, while it has been reported that BET inhibition does not affect MYCN levels (Shu et al., 2016). Thus, BET degradation may provide a more general strategy to overcoming BET inhibitor resistance in various contexts. With this in mind, in collaboration we are also pursing BET degradation in two recent proposed models of BET inhibitor resistance in leukemia cells involving an upregulation of Wnt signaling (Dawson et al., 2011; Fong et al., 2015).

Collectively, this study presents evidence supporting the use of targeted BET family degradation as a strategy to overcome BET inhibitor resistance. It has become increasingly clear that targeted degradation of a given target is markedly different from pharmacological inhibition. This is reflected in the kinetics and the nature of the elicited molecular response. It will be critical to assess the utility of targeted BET degradation in an in vivo setting (actively pursued in our lab) to understand the applicability of these findings to potential emerging clinical BET inhibitor resistance.

Methods

Cell Lines

Breast cell lines were kindly provided by Dr. Kornelia Polyak (Dana Farber Cancer Institute) and cultured in Dulbecco’s Modified Eagle Medium (DMEM) supplemented with 10% FBS. The SUM series of cell lines were cultured in a 1:1 mixture of DMEM and Human Mammary Epithelial Cell (HMEC) media (Cell Applications, cat# 815-500) supplemented with 10% FBS. The SUM149R and
SUM159R cell lines were maintained in the presence of 10µM and 20µM of JQ1 respectively. For dBET6 treatment, cells were washed three times in PBS to remove JQ1 and subsequently treated.

**Small molecule inhibitors and degraders**

The dBET series (dBET1 – dBET10) was synthesized by Dr. Dennis Buckley (Dana Farber Cancer Institute). JQ1 was synthesized by Dr. Jun Qi (Dana Farber Cancer Institute).

**Immunoblotting**

Cells were lysed using RIPA buffer supplemented with protease inhibitor cocktail (Roche) for 20 minutes on ice. The lysates were spun at 16,000 g for 15 minutes at 4°C and protein concentration was determined using BCA assay (Pierce). Antibodies for western blot were purchased as follows: BRD4 (Bethyl, A301-985A), BRD3 (abcam, ab56342), BRD2 (Bethyl, A302-582A), Vinculin (Santa Cruz, sc-25336), Actin (Santa Cruz, sc-8432), RNA Pol II (Millipore, 05-623), RNA Pol II Ser2P (Millipore, 04-1571), RNA Pol II Ser5P (abcam, ab5131), and MYCN (Santa Cruz, sc-56729)

**Cell viability**

Using a BioTek EL406 multi-mode dispenser, 2000 cells in 50µl were plated into 384-well white bottom assay plates. Small molecules were arrayed in a corresponding 384-well plate. Immediately following cell plating, 100 nanoliters of compound was pinned into each well using a Janus workstation (Perkin Elmer). Compound stocks were arrayed in 10 point dose in quadruplicate in DMSO. Cell viability was measured at 72 hours via surrogate measurements of
ATP concentration via ATPlite (Perkin Elmer) following manufacturers instructions. A diluted stock of substrate in lysis buffer was made and 25µl was dispensed into each well of the 384-well plate using the BioTek. Plates were spun down and allowed to incubate for 10 minutes before luminescence measurements using a multiplate Envision plate reader (Perkin Elmer).

Nonlinear dose response curves were fitted to the data using Graphpad Prism software. Area under curve (AUC) measurements were also calculated using Graphpad Prism software. Heatmap representation of AUC measurements was performed using GENE-E software (Broad Institute).

RNA isolation

Prior to RNA isolation, cell numbers were determined using the Countess automated cell counter Invitrogen. Total RNA isolation was performed using the miRvana miRNA total RNA isolation kit (ThermoFisher Scientific, AM1560) according to manufacturers instructions. Following isolation, RNA was digested with DNase (Ambion). During isolation, external RNA spike-ins (ERCC, Ambion) were added at the time of cell lysis. Total RNA was subject to polyA selection and adapter ligation in preparation for next-generation sequencing (Illumina stranded mRNA library prep) on Nextseq (75 basepair, single-end). All RNA-seq samples were performed in biological triplicate.

Quantitative PCR (qPCR)

cDNA was synthesized using the Superscript VILO kit (Invitrogen) following the manufacturers instructions. qPCR was performed using SYBR green Master Mix (Applied Biosystems) on a Viia7 in 96-well format. Primer
sequences for mRNA transcript detection of MYCN:
CACAGTGACCACGTCGATTT (forward) and CACAAGGCCCTCAGTACCTC (reverse).

RNA-seq data analysis
All analyses were performed using HG19 annotations. Alignments were performed using TopHatV2.0.11 with default parameters. Transcript abundance were calculated and normalized using the cuffquant module of Cufflinks (Trapnell et al., 2010). Fragment per kilobase of exon per million fragments mapped (FPKM) values were calculated and normalized using cuffnorm. Heat map was generated by clustering the Euclidean distance between samples using the R function hclust calling the “Ward” method.

Chromatin immunoprecipitation
Chromatin immunoprecipitations were performed as described with minor changes (Chapuy et al., 2013). Neuroblastoma cell lines were grown to 75% confluence in 15 cm plates and cross-linked with 1% formaldehyde (10 minutes) followed by quenching (125 mM glycine). Cells were washed in cold PBS and harvested by cell scraper in cold PBS with protease inhibitors (Roche). Cells were centrifuged at 1650 x g for 5 minutes. Pellets were resuspended in cytosolic then nuclear lysis buffer and DNA was sheared on ice using a waterbath sonicator (Bioruptor, Diagenode) for 28 minutes at high output (30” on, 30” off) in 1mL of sonication buffer supplemented with 0.5% SDS. Antibodies for ChIP were purchased as follows: BRD4 (Bethyl, A301-985A). Sonicated lysates were cleared by centrifuging at 20,000g for 10 min and incubated overnight end-over-
end at 4°C with magnetic beads prebound with antibody. Beads were washed three times with sonication buffer, one time with sonication buffer with 500 mM NaCl added, 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 once with TE. DNA was eluted in elution buffer (50 mM Tris-HCl pH 8, 10 mM EDTA, and 1% SDS). Cross-links were reversed overnight at 65°C. RNA and protein were digested with 0.2 mg/mL RNase A for two hours followed by 0.2 mg/mL Proteinase K for one hr. DNA was purified with phenol chloroform extraction and ethanol precipitation.

Libraries for sequencing were prepared using the Rubicon Thruplex FD library preparation kit. An input of 50 ng of DNA or less were used and following ligation libraries were amplified per manufacturers instructions. Amplified libraries were then size-selected using AMPure beads (Agencourt AMPure XP) per manufacturers instruction. Libraries were multiplexed at equimolar rations and run together in one lane on the NextSeq (75 base pair, single-end).

**ChIP-seq analysis**

All analyses were performed using NCBI37/HG19 annotations. All ChIP-Seq datasets were aligned using Bowtie2 (version 2.2.1). 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. ChIP-Seq reads aligning to the region were extended by 200 bp and the density of reads per base pair (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). We used the MACS version 1.4.1 (Model based analysis of ChIP-Seq) 67 peak finding algorithm to identify regions of ChIP-Seq enrichment over background. A p value threshold of enrichment of 1e-6 was used.

Acknowledgements

We would like to thank Shaokun Shu, Doris Tabassum and Kornelia Polyak (Dana Farber Cancer Institute) for past and current collaborative work surrounding BET inhibitor resistance in TNBC.
Chapter 5: Conclusion and future directions
Conclusion

This work focused on understanding and modulating aberrant enhancer activity to identify potential vulnerabilities in cancer. We studied three different aspects of aberrant enhancer activity: enhancer target genes, transcription factor binding at enhancers, and small molecule modulation of enhancer activity. This work has provided insights into the contribution of aberrant enhancer activity to the oncogenic state and in turn identified potential cancer vulnerabilities.

In pediatric neuroblastoma, amplified levels of the oncogenic transcription factor MYCN define high-risk disease and treatment options for these patients are truly an unmet clinical need. Apart from MYCN amplification, neuroblastoma exhibits a relatively low gene coding mutation frequency (Pugh et al., 2013). Thus, we reasoned that the deregulation of noncoding enhancers might present an alternative strategy to identifying novel vulnerabilities. We approached deregulated enhancer activity in neuroblastoma by looking at enhancer target genes as well as the role of MYCN transcription factor binding at these cis regulatory elements.

To identify previously unrecognized dependencies in neuroblastoma, we first generated genome-wide maps of the active enhancer gene regulatory landscape. We identified and catalogued large cis regulatory elements (termed super enhancers) adjacent to several genes previously implicated in neuroblastoma pathogenesis. Interestingly, the global enhancer landscape reflected MYCN amplification status suggesting the importance of the oncogenic transcription factor to enhancer activity (and pursued in chapter three of this
work). We performed systematic functional screening of super enhancer associated genes and identified ID1 as a candidate dependency in both MYCN amplified and non-amplified neuroblastoma. In an effort to identify clinical applicability to this finding, we explored the efficacy of small molecule inhibitor of the deubiquitinating enzyme (DUB) USP1. USP1 inhibition indirectly promoted ID1 degradation (secondary effect) and had a potent growth inhibitory effect in a panel of neuroblastoma cell lines. In addition, USP1 inhibition led to the potent down regulation of MYCN protein contributing to the viability effect observed. Collectively, these results establish ID1 as a dependency in neuroblastoma and suggest that DUB inhibitors may provide a novel path towards therapeutic applicability.

In chapter three, we showed that MYCN binding at enhancers is critically important to the oncogenic state. We generated high quality genome-wide MYCN binding profiles in a series of cell lines to understand how MYCN occupancy on the genome drives its oncogenic program. We took advantage of a tet-off MYCN expression system that allowed for tight control of MYCN protein levels. In these dynamic experiments, we found that MYCN shutdown results in a global reduction of histone acetylation and subsequently transcription. We find that deregulated MYCN ‘invades’ E-box rich proximal enhancer regions and that these regulatory elements dictate expression of MYCN responsive target genes. Given the tissue-specific nature of these enhancers, we hypothesized that lineage specific transcription factors would collaborate at these invaded enhancers to facilitate their activity. Indeed, we identify TWIST1 as a lineage
specific transcription factor that occupies MYCN invaded enhancer regions. Interestingly, MYCN shutdown evicts TWIST1 from these enhancers and TWIST1 knockdown is synthetic lethal with oncogenic MYCN. We confirm the importance of enhancer dependent regulation of TWIST1 by genetically disrupting a downstream TWIST1 super enhancer. Collectively, these data point to the importance of tissue specific transcription factors in establishing active enhancer regulatory regions that oncogenic MYCN ‘invades’ to drive its oncogenic program.

Lastly, in the fourth chapter of this work we explore the pharmacological consequences of enhancer cofactor degradation versus inhibition in triple negative breast cancer. The development of small molecule inhibitors of the bromodomain and extra-terminal (BET) family of proteins has provided a generalizable strategy to target enhancer activity in cancer. The preclinical success of BET inhibition in several cancers as well as encouraging initial clinical results have motivated the identification and characterization of potential resistance mechanisms. We use newly developed bifunctional molecules that hijack E3 ligase activity and induce targeted degradation. We report that targeted BET degradation effectively overcomes both adapted and intrinsic models of BET inhibition via a fundamentally different molecular response.

Together, these studies highlight the contribution of deregulated enhancer activity to the oncogenic state. A functional enhancer is dependent on several factors and thus aberrant enhancer activity can be caused in a variety of ways.
We looked at three distinct facets of enhancer deregulation and used genetic and chemical tools to disrupt their activity and uncover potential vulnerabilities.

**Future directions**

*Functionally defining genome-wide predictions of enhancers*

The rapid evolution of next-generation sequencing has transformed our ability to identify and categorize enhancer elements genome-wide. In neuroblastoma, we made global predictions of enhancer activity based on measurements of enhancer cofactor occupancy via chromatin immunoprecipitation coupled with high throughput sequencing (ChIP-seq). However, it is important to note that this approach is purely correlative and thus does not directly report on the functionality of a putative cis regulatory element. Indeed, these approaches can often identify on the order of thousands of cis regulatory elements and it is unlikely that they are all active (Dunham et al., 2012; Ernst et al., 2011; Lin et al., 2016). In our models of neuroblastoma, it became quite apparent that not all predicted enhancers were functional.

Thus, it will be critical to develop techniques that can rapidly and accurately identify functional enhancers from genome-wide predictions. The development of CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats) technologies offers a genome-wide strategy to functionally annotate enhancers. At a single enhancer driving the TWIST1 transcription factor, we show that targeted editing and subsequent disruption of the enhancer regulatory sequence resulted in a cellular fitness defect. In this case, we used CRISPR-Cas9 to genetically edit the underlying enhancer sequence. In addition, variations
of this technology could be used to disrupt other aspects of enhancer activity such as employing an inactive form of Cas9 (dCas9) to recruit chromatin-modifying enzymes to alter chromatin state. Given the ability of CRISPR-Cas9 technologies to be applied in a genome-wide screening fashion, this technique will be instrumental in establishing truly functional active enhancer regulatory elements from genome-wide predictions. In the same way that genome-wide CRISPR-Cas9 screening is being conducted targeted gene-coding regions, the same could very well be done looking strictly at cis regulatory elements (Rajagopal et al., 2016).

The use of CRISPR-Cas9 to directly interrogate cis regulatory elements also circumvents the need for genome-wide enhancer target gene predictions. As a lab, we have developed several strategies to predict enhancer associated genes integrating several types of data including RNA Polymerase II ChIP-seq and gene expression data. The proper identification of super enhancer associated genes was critical to our work in chapter two in which we functionally screened a focused set of super enhancer associated genes. We took advantage of parallel steady state gene expression data to strengthen our predictions. However, enhancers can act at very long distances and even have multiple target genes. In some instances, it was very difficult to make accurate predictions and for that reason we took a less stringent approach in generating the super enhancer associated gene library. The use of CRISPR-Cas9 to directly functionally disrupt enhancer elements allows direct experimental testing of the
consequences of enhancers on target gene expression that may provide more clarity when evaluating the effects of aberrant enhancer activity.

Small molecule DUB inhibitors in neuroblastoma

The identification of the super enhancer associated gene ID1 as a dependency in neuroblastoma led us to investigate the efficacy of small molecule inhibition of the deubiquitinating enzyme (DUB) USP1. It had previously been reported that pharmacologic USP1 inhibition resulted in degradation of ID1 in leukemic cells (Helena et al., 2013). While USP1 inhibition resulted in ID1 protein degradation, it was confounded by the concurrent degradation of MYCN protein. This is perhaps not surprising given that USP1 is predicted to deubiquitinate several targets and the lack of selectivity with the SJB3-019A tool compound (Ritorto et al., 2014).

In MYCN amplified neuroblastoma, we observed a potent viability loss along with modulation of the high-risk defining oncogene MYCN. The marked down regulation of MYCN protein levels could prove to be a promising avenue of development for this class of inhibitors. Indeed, small molecules that indirectly target MYCN have shown marked reduction in tumor growth both in vitro and in vivo (Gustafson et al., 2014; Puissant et al., 2013). A better understanding of how USP1 inhibition results in such a dramatic viability defect especially with respect to MYCN could point to the validity of developing these inhibitors. In particular, it will be important to identify whether MYCN is a direct substrate of USP1 and more generally establish the global profile of USP1 substrates.
In parallel to understanding the biological role of USP1, the development of more selective USP1 inhibitors would also be critical to understanding the effects of inhibition. DUB enzymes are an attractive drug target due to their enzymatic binding pocket and their biological role in modulating protein stability. In addition to selectivity issues, we observed a lack of efficacy in xenograft models of MYCN amplified neuroblastoma, which were likely due to poor compound solubility. The development of the SJB3-019A compound must also focus on improving the pharmacokinetic/dynamic properties to enable in vivo utility. Regardless of the mechanism underlying sensitivity to USP1 inhibition in neuroblastoma, these small molecule inhibitors are a promising new class with the potential to modulate the MYCN oncoprotein.

*Therapeutic implications of MYCN as a non-linear amplifier of transcription*

As described in chapter three, we find that at oncogenic levels, MYCN ‘invades’ cis regulatory elements resulting in global transcriptional amplification. Moreover, we found that these tissue-specific enhancers were critical to MYCN mediated oncogenesis and potentially help resolve an existing controversy surrounding MYC family transcription factors (TFs) as amplifiers or selective regulators (Lin et al., 2012; Nie et al., 2012; Sabo et al., 2014; Susanne et al., 2014). The non linear nature of MYC family global transcriptional amplification may also provide a general strategy to identify previously unrecognized vulnerabilities in two major ways. First, the tissue-specifying factors that establish dynamic MYCN invaded enhancers represent potential vulnerabilities in MYC family driven malignancies. Second, the target genes of these invaded
enhancers point to genes particularly sensitive to MYC family perturbation and thus may point to non-obvious cancer dependencies.

In neuroblastoma, we identify the bHLH transcription factor TWIST1 as key factor that shapes the enhancer landscape and synthetic lethal partner of MYCN. Thus, these tissue-specific enhancer-specifying factors may represent synthetic lethal partners of MYC family TFs. Given the difficult in direct pharmacological inhibition of MYC family TFs, several studies have identified synthetic lethal partners in an effort to identify vulnerabilities of MYC family driven tumors (Kessler et al., 2012; Toyoshima et al., 2012). The synthetic lethal partners identified in these studies have generally been restricted to particular tumor types, perhaps due to the tissue-specific nature of enhancer invasion. The identification of E-box rich invaded enhancers and the specifying lineage specific transcription factors could provide a more systematic approach to identifying tissue-specific synthetic lethal interactions. While transcription factors are often difficult drug targets, they may point to non-obvious upstream pathways and targets for which therapeutic intervention is feasible.

The set of E-box rich enhancer driven genes that are particularly sensitive to MYCN shutdown could also provide a gene signature predictive of disruption of the MYCN oncogenic program. The use of a predictive gene signature of MYCN shutdown would provide a framework to identify compounds that modulate MYCN function. For example, this strategy could be employed to further develop DUB inhibitors (such as SJB3-019A) in neuroblastoma. The use of a predictive signature of MYCN perturbation would help focus the identification
of DUB inhibitors that modulate MYCN activity. This strategy could also be employed in other MYC family driven tumors each with its own specified MYC family signature based on E-box rich enhancer regulatory elements. We have begun integrating the MYCN enhancer-defined sensitive gene set with small molecule profiling experiments to potentially identify non-obvious small molecules that perturb the oncogenic program (Basu et al., 2013).

**Applicability of chemically targeted protein degradation**

In chapter four, we make use of a newly developed class of molecules that allow for chemically induced targeted protein degradation. We have been intensely focused on expanding this technology to selectively degrade a variety of targets implicated in human cancers. This new class of molecules offers an added dimension into the tractability of small molecule disruption of targets. Unlike small molecule inhibitors, bifunctional degraders only need to bind their target rather than bind and inhibit functionality. Thus, these bifunctional molecules may present opportunities to pursue targets that were otherwise classified as ‘undruggable’.

These bifunctional molecules could provide a chemical avenue to directly modulate transcription factors that have been notoriously difficult to inhibit with small molecules (Wells and McClendon, 2007). For example, there is no doubt that it would be efficacious to directly perturb MYCN function in high-risk neuroblastoma. However, small molecule inhibition of MYC family TFs remains challenging due to the absence of a tractable small molecule ligand-binding domain (Nair and Burley, 2003). MYC family proteins lack enzymatic activity and
their function is predicated on protein-protein interactions and these properties have doomed most drug discovery efforts (McKeown and Bradner, 2014). In spite of these major obstacles, several groups have organized efforts to identify small molecule inhibitors of MYC function focusing mainly on disrupting dimerization between MYC and its partner MAX (Berg et al., 2002; Jeong et al., 2010; Park et al., 2004; Shi et al., 2009; Xu et al., 2006; Yap et al., 2013; Yin et al., 2003). These efforts have yielded limited success with low potency molecules but they have been shown to bind MYC. While these small molecules have not resulted in a true chemical probe, perhaps they could serve as the basis for a bifunctional molecule aimed at targeted degradation. Given the structural similarity between MYC family TFs, it is likely that a bifunctional MYC degrader would target all three members (MYC, MYCN, and MYCL). Thus, targeted degradation offers a new strategy to pursue direct MYC family chemical modulation.

Pharmacological inhibition versus pharmacological degradation

In addition to providing new strategies to approach ‘undruggable’ targets, targeted protein degradation can also be applied to targets with existing small molecule inhibitors. In this work, we explored the increased efficacy of targeted degradation of the BET family versus small molecule inhibition. We were surprised to observe such a marked difference in efficacy as well as molecular response upon BET degradation as compared to BET inhibition. These observed differences point to some of the fundamental differences between inhibition and degradation. It has become apparent that in some instances targeted
degradation features incredibly quick kinetics. We have observed targeted
degradation within one hour of incubation and in turn phenotypic consequences
are often observed earlier than target inhibition. These bifunctional degraders
behave similarly to enzymes with respect to kinetic turnover. While inhibitors
must remain bound to elicit an effect, chemical degraders bind their targets,
induce degradation and then are released. In this way, these bifunctional
molecules are continually being recycled throughout the induced degradation
process.

The fundamental differences between inhibition and degradation provide
alternative therapeutic avenues for a given target. In this work, we took
advantage of targeted BET family degradation to overcome models of BET
inhibitor resistance in triple negative breast cancer. It could prove equally
advantageous to pursue pharmacologically induced degradation of targets with
existing inhibitors in hopes of increasing efficacy. Moreover, targeted degradation
may result in unexpected forms of efficacy not seen with inhibitors. For example,
there are several potent and selective small molecule inhibitors of a variety of
kinases. There are several examples of the efficacy of small molecule inhibition
of deregulated kinase activity in cancer such as ALK inhibition in neuroblastoma
(Chen et al., 2008; Mossé et al., 2008). While kinase inhibition has been
extremely successful, targeted degradation may offer strategies to overcome
observed resistance or result in an unexpected, beneficial molecular response. In
addition, degradation may provide a strategy for targeting non-enzymatic
functions such as scaffolding functions of kinases (Cance et al., 2013; Gogate et al., 2014).

**Clinical utility of BET degradation**

In chapter four, we demonstrated that targeted BET family degradation overcomes models of BET inhibitor resistance in *in vitro* models of triple negative breast cancer. We have also explored the marked efficacy of BET family degradation in hematologic malignancies including acute myeloid leukemia (Winter et al., 2015). Ultimately, the therapeutic applicability of these small molecule tool compounds will depend on their performance in *in vivo* models. We have had initial success in xenograft models of human MV4;11 leukemia cells. In triple negative breast cancer, we are actively conducting *in vivo* experiments testing the efficacy of BET degradation in overcoming resistance. These studies have been critical to identifying pharmacokinetic and pharmacodynamic parameters for improvement. For example, we have developed improved formulations to improve solubility of the dBET6 bifunctional molecule to allow for higher drug exposure. It is also becoming increasingly clear that dosing regimens (dose and frequency) are completely different with targeted degradation versus inhibition. It has become increasingly clear that it is necessary to take into account the enzymatic properties of dBET6 in dosing regimens. It will be critical to test dBET6 in a variety of *in vivo* models to generate a detailed understanding of the pharmacodynamic/pharmacokinetic properties. These preclinical studies will be crucial to guiding the applicability of targeted BET degraders in humans.
References


Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. Nature 478, 529-533.


sequences as an element of the early promoter. Proceedings of the National
Academy of Sciences of the United States of America 78, 943-947.

Guccione, E., Martinato, F., Finocchiaro, G., Luzi, L., Tizzoni, L., Dall’Olio, V.,
recognition in the human genome is determined by chromatin context. Nature cell
biology 8, 764-770.

Gustafson, W., Meyerowitz, J., Nekritz, E.A., Chen, J., Benes, C., Charron, E.,
Simonds, E.F., Seeger, R., Matthay, K.K., Hertz, N.T., et al. (2014). Drugging
MYCN through an Allosteric Transition in Aurora Kinase A. Cancer Cell 26.

Hadjur, S., Williams, L.M., Ryan, N.K., Cobb, B.S., Sexton, T., Fraser, P., Fisher,
interactions at the developmentally regulated IFNG locus. Nature 460, 410-413.

Halfon, M.S., Carmena, A., Gisselbrecht, S., Sackerson, C.M., Jiménez, F.,
Baylies, M.K., and Michelson, A.M. (2000). Ras pathway specificity is determined
by the integration of multiple signal-activated and tissue-restricted transcription


of SV40 promoters. The EMBO journal.

He, A., Kong, S.W., Ma, Q., and Pu, W.T. (2011). Co-occupancy by multiple
cardiac transcription factors identifies transcriptional enhancers active in heart.
Proceedings of the National Academy of Sciences of the United States of
America 108, 5632-5637.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp,
modifications at human enhancers reflect global cell-type-specific gene

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D.,


the National Academy of Sciences of the United States of America 73, 3628-3632.


UAF1 inhibitor links deubiquitination to DNA damage responses. Nature Chemical Biology 10, 298-304.


