

# **A MYC INDUCED LONG NON-CODING RNA, DANCR, THAT AFFECTS CELL PROLIFERATION**

by

Yunqi Lu

A dissertation submitted to Johns Hopkins University in conformity with the  
requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

August, 2014

©Yunqi Lu 2014

All rights reserved

# Abstract

Long non-coding RNAs (lncRNAs) are pervasively transcribed in the human genome and are known to be functioning in various cellular processes; however, the exact mechanisms involved are not well understood. The MYC proto-oncogene encodes a transcription factor that regulates cellular functions through the activation of its many RNA targets. In this thesis, I address the role of MYC-regulated lncRNAs in cancer.

I first utilized experimental methods, such as microarray, RNA-seq, and TCGA analysis, to globally explore the regulatory role of MYC on lncRNAs. One lncRNA, DANCR, was identified as one of the most highly MYC-induced targets in human cancers. To further confirm that DANCR is directly regulated by MYC, I demonstrated by CHIP and CHIP-seq that MYC could bind directly to an E-box motif (CACGTG) in the DANCR promoter sequence. Using existing data and my own experimental analysis, DANCR expression was shown to be upregulated in most primary cancers. I next sought to determine whether DANCR is required for MYC-mediated cellular functions in cancer. Abrogation of DANCR expression using siRNA resulted in defective cell proliferation. Cell cycle analysis showed that the defect triggered by DANCR depletion was a consequence of G1 to S phase transition arrest. To assess the expression of genes responsible for this phenotype, I examined the expression profiles of cancer cells treated with DANCR siRNA. Gene Set Enrichment Analysis (GSEA) showed that a large subset of upregulated genes are common to DANCR- and EZH2-depletion.

Moreover, p21, an important G1 phase cell cycle regulator, was found to be among those genes. I later showed that p21 could be suppressed by both DANCER and EZH2. More importantly, a double knock-down experiment of DANCER and p21 resulted in the partial rescue of the proliferation reduction caused by DANCER knock-down alone, providing evidence that p21 is one of the most important targets of DANCER.

I proposed that MYC induces DANCER, which suppresses p21 via the EZH2 pathway, and then provided evidence for this model in a series of IP (immunoprecipitation) experiments to establish interactions between pathway components. A 3'-domain of DANCER was shown to bind EZH2, whereas a 5'-domain of DANCER recognized p21 nascent mRNA via RNA–RNA interactions, which guides DANCER–EZH2 complexes to the p21 locus. Finally, using computational approaches, I also attempted to explore whether my proposed mechanism could be generalized to other DANCER targets, and 8 out of 16 genes tested seemed to conform to this model. My findings elucidate an important missing 'link' in the MYC regulated gene network, showing active roles for lncRNAs in MYC-mediated cancer, which could be important targets for cancer diagnosis and therapy.

**Advisor:** Dr. Chi Dang

**Readers:** Drs. Steve Baylin and Chi Dang

**Other Committee Members:** Drs. John Isaacs and James Herman

# Preface

I would like to thank my mentor, Dr. Chi Dang, who has been an incredible teacher and source of support throughout my graduate school years. His passion and devotion towards science have always inspired me. I also greatly admire his scientific and personal integrity. I would not be able to complete this journey without his guidance.

I have to thank all members of the Dang lab, past and present, for their support and friendship: Karen Zeller, Yan Xiang, Jinsong Xia, Zach Stine, Brian Altman, Ramani Dinavahi, Ping Gao, Peng Sun, Anne Le, Huichun Zhan, Annie Hsieh, Arvin Gowh, and Zandra Walden. They have made this experience such a special chapter in my life. I am indebted to them for the help that I received with all the difficulties both at work and in life. I would also like to acknowledge my collaborators: Cheryl Koh, Dr. Angelo De Marzo, Dr. Karen Sfanos, Jiehuan Sun, Dr. Hongkai Ji, Youyou Zhang, Dr. Lin Zhang, and members of various core facilities on both the Johns Hopkins and University of Pennsylvania campuses for sharing invaluable materials, techniques, and knowledge.

I want to thank the members of my thesis committee, including Dr. John Isaacs, Dr. Steve Baylin, and Dr. James Herman, for their input and advice. I would like to specially thank Dr. Baylin for reading my thesis and sharing his insights. I also would like to thank the Human Genetics program, of which I am privileged to be a part. My appreciation goes to Dr. Dave Valle, Dr. Kirby Smith, and Sandy Muscelli for running such a superb graduate program.



I am grateful to my friends here in the US and also back home who have provided me with so much support all these years. My classmates and basketball teammates are a wonderful group of individuals.

Finally, I would like to thank my family, especially my parents Chunbo Lu and Dongming Li. Even though they live halfway around the world, their endless and unconditional support has given me the courage to overcome any obstacle.

# Table of Contents

Abstract	ii
Preface	iv
Table of Contents	vi
List of Tables	vii
List of Figures	viii
Chapter 1: Introduction	1
Chapter 2: MYC-induced lncRNAs in human cancers	16
Chapter 3: LncRNA DANCR is a MYC target	35
Chapter 4: DANCR in human cancers	45
Chapter 5: Elucidating the functional significance and targets of DANCR	57
Chapter 6: The mechanism of DANCR function	94
References	119
<i>Curriculum vitae</i>	138

# List of Tables

<b>Table 1.</b> MYC-induced genes identified by gene expression array analysis.	20
<b>Table 2.</b> MYC-induced genes identified using RNA-seq.	23
<b>Table 3.</b> Genes that are commonly up-regulated by MYC, DANCR, and EZH2 knock-down.	68
<b>Table 4.</b> Validation of targets regulated by both DANCR and EZH2 in PC3 cells.	69

# List of Figures

<b>Figure 1.</b> MYC-regulated gene clusters.	21
<b>Figure 2.</b> The number of lncRNAs induced by MYC.	22
<b>Figure 3.</b> qRT-PCR validation of the induction of lncRNA expression by MYC	27
<b>Figure 4.</b> qRT-PCR validation of the induction of DANCR by MYC.	28
<b>Figure 5.</b> A heatmap of lncRNAs that are co-expressed with MYC.	29
<b>Figure 6.</b> Correlations between MYC and DANCR expression in breast and prostate cancers.	30
<b>Figure 7.</b> Chromatin immunoprecipitation validation of DANCR as a direct target of MYC.	39
<b>Figure 8.</b> Chromatin immunoprecipitation validation of DANCR as a direct MYC target in human embryonic stem cells.	40
<b>Figure 9.</b> CHIP-seq data-based assessment of MYC–DANCR and MYC–NPM1 binding.	41
<b>Figure 10.</b> MYC binding to the DANCR promoter in six different cell types.	42
<b>Figure 11.</b> DANCR expression profiles of tumor versus normal tissue (from Oncomine.org).	48
<b>Figure 12.</b> MYC copy number alterations in primary tumors.	51
<b>Figure 13.</b> Photomicrographs from 4-week-old Low-MYC mice.	52
<b>Figure 14.</b> MYC and DANCR expression in 20 paired prostate tumor and normal tissues.	54
<b>Figure 15.</b> Kaplan–Meier curves for prostate cancer survival.	55
<b>Figure 16.</b> The effects of DANCR and MYC siRNA-mediated depletion on the proliferation of DU145 prostate cancer cells.	60
<b>Figure 17.</b> The effects of depleting DANCR and MYC on the proliferation of PC3 prostate cancer cells.	61

<b>Figure 18.</b> Cell cycle analysis of DANCR knock-down versus control siRNA in PC3 and DU145 cells.	63
<b>Figure 19.</b> Gene set enrichment analysis (GSEA) against previous expression profiling experiments available in public databases.	66
<b>Figure 20.</b> Illustration of the process for identifying candidate genes affected by DANCR, MYC, and EZH2.	67
<b>Figure 21.</b> Poor survival outcome correlates with p21 expression in prostate cancer patients.	71
<b>Figure 22.</b> Depletion of DANCR by RNAi activates p21.	72
<b>Figure 23.</b> Negative correlation between DANCR and CDKN1A (p21) expression in prostate cancer.	74
<b>Figure 24.</b> Negative correlations between DANCR and CDKN1A (p21) expression in breast cancers.	75
<b>Figure 25.</b> The effects of two vivo-morpholinos that depleted DANCR on the proliferation of P493 cells.	77
<b>Figure 26.</b> Levels of p21 protein increase after 48 h of vivo-morpholino treatment in P493 cells.	79
<b>Figure 27.</b> SnoRA26 RNA not affected by DANCR depletion.	82
<b>Figure 28.</b> No effects of blocking SnoRa26 on the proliferation of PC3 cells.	83
<b>Figure 29.</b> Lack of effects of blocking SnoRa26 on the proliferation of P493 cells.	84
<b>Figure 30.</b> DANCR, MYC, and p21 expression levels in prostate tumor versus normal tissues.	86
<b>Figure 31.</b> The abundance of DANCR, MYC, and p21 RNA in prostate cancer.	87
<b>Figure 32.</b> Double knock-down of DANCR + p21 partially rescues the growth arrest induced by DANCR depletion in PC3 cells.	89
<b>Figure 33.</b> Double knock-down of MYC + p21 cannot rescue MYC depletion-induced growth arrest in PC3 cells.	90

<b>Figure 34.</b> A working model illustrating the potential mechanism of DNACR–EZH2 silencing of p21.	98
<b>Figure 35.</b> Immunoblot of EZH2 and p21 protein levels after knock-down of either EZH2 alone or both EZH2 and p21.	99
<b>Figure 36.</b> EZH2 RIP specifically retrieves DANCR RNA.	100
<b>Figure 37.</b> <i>In vitro</i> transcribed (IVT) biotinylated DANCR retrieves EZH2 protein.	101
<b>Figure 38.</b> The 3'-end of DANCR is required for EZH2 binding.	102
<b>Figure 39.</b> DANCR depletion reduces EZH2 binding at the p21 promoter locus in PC3 cells.	104
<b>Figure 40.</b> DANCR depletion reduces EZH2 binding at the p21 promoter locus in DU145 cells.	105
<b>Figure 41.</b> Possible domain mediating DANCR–p21 mRNA interactions identified by <i>IntaRNA</i> and <i>RNAfold</i> .	108
<b>Figure 42.</b> A schematic illustration of the experimental design for mapping DANCR–p21 mRNA interactions.	109
<b>Figure 43.</b> Full-length DANCR could specifically retrieve p21 mRNA.	110
<b>Figure 44.</b> A DANCR fragment containing the 5'-end could specifically retrieve p21 mRNA.	111
<b>Figure 45.</b> A DANCR fragment containing the 5'-end could specifically retrieve p21 mRNA (continued).	112
<b>Figure 46.</b> A 3'-end DANCR fragment cannot specifically retrieve p21 mRNA.	113
<b>Figure 47.</b> Predicted common interaction domains of DANCR and four of its targets identified by <i>IntaRNA</i> and <i>RNAfold</i> .	115
<b>Figure 48.</b> Predicted common interaction domains of DANCR and another four DANCR targets identified by <i>IntaRNA</i> and <i>RNAfold</i> .	116

# **Chapter 1**

## **Introduction**

The MYC proto-oncogene was first discovered in studies of avian tumors caused by retroviruses. It was identified as a mammalian homolog of the v-myc oncogene of avian myelocytomatosis virus strain 29 (Duesberg and Vogt, 1979; Hu et al., 1979; Sheiness and Bishop, 1979; Vennstrom et al., 1982). The MYC family includes L-MYC, N-MYC, and B-MYC (Brodeur et al., 1984; Kohl et al., 1984; Maris, 2010; Nau et al., 1985). MYC expression is tightly regulated by several mechanisms (Brooks and Hurley, 2010; Hurley et al., 2006; Levens, 2010). Dysregulation of MYC is often thought to be one of the most common oncogenic factors in cancer (Cole and McMahon, 1999). Through its actions on various direct and indirect target genes, the MYC target gene network is involved in essentially all cellular processes (Dang, 1999; Grandori and Eisenman, 1997; Oster et al., 2002; Secombe et al., 2004). Despite the growing number of studies revealing more details of its functions, certain mechanisms whereby MYC regulates its downstream targets are not yet fully understood. MYC has been shown to stimulate transcription from all three nuclear RNA polymerases (I, II, and III), and it was previously shown to regulate the expression of mRNAs, rRNAs, tRNAs, and microRNAs. This thesis presents a new mechanism of MYC regulation, whereby MYC regulation of gene expression can occur through long non-coding RNAs (lncRNAs).



## **Oncogenesis and MYC**

MYC is one of the most important oncogenic transcription factors that is frequently dysregulated in many human and animal cancers (Cole and McMahon, 1999; Dang, 1999; Nesbit et al., 1999). Increased MYC expression is often found in aggressive cancers (Liao and Dickson, 2000). Translocations between MYC on chromosome 8 and immunoglobulin enhancers on chromosomes 2, 14, and 22 are the causes of Burkitt's lymphoma, resulting in constitutive MYC expression and uncontrolled cellular proliferation (Dalla-Favera et al., 1982; Taub et al., 1982). Elevated amplification of MYC expression levels is also a feature of many human cancers (Augenlicht et al., 1997; Beroukhim et al., 2010). Several studies of transgenic mice established that abnormally expressed MYC in different tissues can occur either alone or in combination with additional mutations that lead to tumorigenesis (Adams et al., 1985; Beer et al., 2004; Chesi et al., 2008; Leder et al., 1986; Pelengaris et al., 1999; Secombe et al., 2004). In mouse models, universal promotion of cellular proliferation and, in some instances, reduction of apoptosis appeared to be the main modes of MYC promotion of tumorigenesis (Cappellen et al., 2007; Koh et al., 2011; Wang et al., 2008a). MYC has also been shown, using these models, to play a role in tumor maintenance; some tumors even have "MYC addiction" (Arvanitis and Felsher, 2006). These previous findings made using mouse models emphasize the role of MYC in murine cancers and support its tumorigenic role in human cancers (Dang, 2012).

## **Biological functions of MYC**

In addition to the clear role of MYC in tumorigenesis, it is a pleiotropic protein with various effects on normal cellular functions. The major functions of MYC are the regulation of cell proliferation, growth, apoptosis, and metabolism (Cole and McMahon, 1999; Dang, 1999). MYC is one of the first immediate-early genes to be characterized that is important for facilitating S phase entry from G1 phase (Bravo, 1990). MYC promotes cell cycle progression by activating cell cycle-promoting factors, such as CDK4 and CDK6, and by suppressing cell cycle inhibiting factors, such as p21, Gadd45, and Gas1 (Gartel et al., 2001; Hermeking et al., 2000; Lee et al., 1997; Marhin et al., 1997; Mateyak et al., 1999; Wu et al., 2003). A novel mechanism whereby MYC drives G1–S cell cycle progression will be presented in greater detail in later chapters of this thesis.

Another important function of MYC is the control of cell growth, which increases cell mass and size during normal cell development (Schmidt, 1999; Secombe et al., 2004). Increases in cell size have been observed during B cell development in transgenic mice overexpressing MYC (Iritani and Eisenman, 1999). MYC control of cell size can occur by regulating genes involved in rRNA synthesis, ribosomal protein production, and other bioenergetics pathways for biomass accumulation which lead to cell growth (Guo et al., 2000; O'Connell et al., 2003; Zeller et al., 2001). MYC also regulates cellular processes, such as apoptosis and carbohydrate metabolism (Shim et al., 1998; Yuneva et al., 2007). Studies from my lab demonstrating MYC regulation of cellular glycolytic pathway elements have been previously reported (Kim et al., 2004).

In addition to the above mentioned activities, MYC has also been found to be one of four genes (along with Sox2, Oct4, and KLF4) that together could reprogram differentiated fibroblasts to pluripotent state, indicating the important role of MYC in maintaining cellular differentiation and preventing de-differentiation (Singh and Dalton, 2009; Takahashi et al., 2007; Takahashi and Yamanaka, 2006).

### **Molecular mechanisms of MYC regulation**

MYC is encoded by a gene that has three exons, which can generate several polypeptides through the use of alternative translation initiation sites. In addition to protein initiation at the canonical AUG start codon, two isoforms that are initiated from a CUG codon in exon one and an internal AUG codon are less well-characterized (Blackwood et al., 1994; Hann et al., 1992; Spotts et al., 1997). MYC is a member of the basic helix-loop-helix family of transcription factors. It contains an N-terminal transcriptional regulatory domain which is followed by a nuclear localization signal and a C-terminal domain that includes a DNA-binding domain with a basic region and helix-loop-helix and leucine zipper motifs. MYC can heterodimerize with Max and bind DNA, and then can regulate the expression of target genes (Amati et al., 1993; Amati et al., 1992; Blackwood and Eisenman, 1991; Dang et al., 1989; Grinberg et al., 2004; Kato et al., 1992; Kretzner et al., 1992).

Two highly conserved regions, termed MYC boxes I and II, are followed by MYC boxes III and IV and a nuclear targeting sequence that is located within the N-terminal domain (Cowling et al., 2006; Dang and Lee, 1988; Kato et al., 1990; Pineda-Lucena and Arrowsmith, 2001). The C-terminal domain contains a basic HLH-ZIP domain that can dimerize with Max, a bHLH-ZIP protein, to form a stable heterodimer (Follis et al., 2009; Hu et al., 2005; Mustata et al., 2009). MYC/Max heterodimers preferentially recognize and bind to a consensus sequence, 5'-CACGTG-3', termed E-boxes, and other non-canonical E-box sequences (Adhikary and Eilers, 2005; Blackwell et al., 1993; Blackwell et al., 1990). The MYC N-terminal domain has been shown to recruit many co-factors, such as TRRAP, GCN5, and TBP, to form complexes (Bouchard et al., 2001; Liu et al., 2003; McEwan et al., 1996; McMahon et al., 2000; Vervoorts et al., 2003). These chromatin-modifying complexes facilitate an open chromatin structure, making the DNA accessible to basal transcriptional mechanisms, including RNA polymerase II, as well as RNA polymerases I and III (Felton-Edkins et al., 2003; Gomez-Roman et al., 2003; Grandori et al., 2005; Kenneth et al., 2007), resulting in MYC target gene activation. The main targets of MYC include genes associated with ribosome biogenesis, translation, metabolism, and mitochondrial function (Meyer and Penn, 2008; Sabo et al., 2014; Walz et al., 2014). Among these targets, there could be mRNAs, microRNAs (miRNAs), or long non-coding RNAs (lncRNAs).

### **MYC and its selectivity**

More recently, MYC has been proposed to function as a 'universal' amplifier of transcription (Lin et al., 2012; Nie et al., 2012). This model was laid out in two publications and was based on the observations that MYC can increase cell size and that the expression of all mRNAs are increased by MYC, even after accounting for differences in cell size. The two groups that published this model claimed that MYC appears to non-specifically amplify all transcripts, so they classified its effects on transcription as 'universal'. However, two recent studies showed that MYC is a *bona fide* selective transcription factor, as described previously. The two groups utilized cell lines in which MYC expression could be induced, although cell size was not changed after MYC induction. Only specific genes, and certainly not all genes, were activated by MYC in these cells (Sabo et al., 2014; Walz et al., 2014).

Whether MYC can directly suppress target gene transcription has also been called into question by the proposed amplifier model. The aforementioned two most recent reports have redefined the genes repressed by MYC. Miz-1, a MYC binding partner, was used as an example to demonstrate the role of other transcription factors in selecting MYC targets. MYC–Miz1 heterodimers specifically suppressed Miz-1 target genes (Walz et al., 2014). Additionally, the transcription of many MYC targets requires the binding of other transcription factors, such as E2F, for them to be activated (Li et al., 2003; Zeller et al., 2006). As previous study reported, gene loci with multiple transcription factors bound are more likely to be MYC-responsive in murine embryonic stem cells (Kim et al., 2010a). Therefore, another level of selectivity rendered by these other

transcription factors further strengthened my supposition that MYC regulation is not 'universal' (Elkon et al., 2004).

Moreover, among the thousands of genes bound by MYC, the degree of regulation is almost certainly not correlated with the expression level of MYC. Reasons for this could include that MYC shows varying binding affinities towards different chromatin configurations, and that the effects of other transcription factors that co-regulate MYC target genes will also affect transcription.

### **Transcriptional repression by MYC**

The mechanism whereby MYC represses gene expression is less well characterized. Mad1 is both a bHLH-ZIP transcription factor and a binding partner of Max (Ayer et al., 1993). Mad1/Max heterodimers are able to compete with MYC/Max heterodimers to bind to the consensus E-box (Baudino and Cleveland, 2001; McArthur et al., 1998). Furthermore, Mad1/Max dimers can repress target gene transcription by recruiting chromatin modifying repressor complexes. Histone de-acetylation that results from the activities of these complexes could then prevent transcriptional activation at the E-box sequences (Ayer et al., 1995; Laherty et al., 1997; Schreiber-Agus et al., 1997).

In addition to interactions with Mad1, MYC binding to Miz1 transcription factor is an important mechanism for MYC-mediated repression (Schneider et al., 1997). For example, MYC/Max heterodimers bind to Inr element in p15INK4b gene and repress its transcription by disturbing Miz1 associations with the p300

co-activator (Seoane et al., 2001; Staller et al., 2001). Another study showed that Miz1 could recruit MYC to the p21 promoter. This interaction can disrupt p53 induction of p21, thereby causing apoptosis (Seoane et al., 2002).

MYC can also mediate gene repression by activating a cluster of microRNAs (miRNAs) on human chromosome 13 (Chang et al., 2008; O'Donnell et al., 2005). The miRNA cluster (miR-17-92 cluster) plays important roles in the regulation of the E2F1 and TGF- $\beta$  signaling pathways (Dews et al., 2014; Mestdagh et al., 2010). More recently, a complex network of MYC, polycomb group proteins (PcGs; notably EZH2), and miRNA interactions has been frequently reported to underlie the mechanisms and outcomes of such interactions (Benetatos et al., 2014; Zhao et al., 2013). This thesis proposes a novel MYC-EZH2-mediated gene repression model, which will be explained in detail in subsequent chapters.

Based on the studies described above, there are certain to be different modes whereby MYC can repress the expression of its target genes. LncRNAs have been implicated in facilitating this process.

### **Long non-coding RNAs and function**

LncRNAs are a class of RNA transcripts that are larger than 200 nucleotides (nt) and do not have apparent protein-coding potential. The human genome has ~25,000 protein coding genes, representing less than 2% of the entire genome, whereas nearly 70% of the human genome is transcribed into

RNA transcripts; thus, by far, most transcripts would be non-coding RNAs (2004; Derrien et al., 2012; Djebali et al., 2012). Compared to protein-coding mRNA transcripts, lncRNAs account for a large portion of the human transcriptome. Since the discovery of non-coding RNAs, smaller regulatory non-coding RNAs, such as miRNAs, have been extensively studied. By contrast, far less is known about lncRNAs. Indeed, for a very long time, lncRNAs have been considered to represent 'transcriptional noise'. However, recent studies indicate that lncRNAs are involved in many normal cellular processes, as well as tumor progression (Batista and Chang, 2013; Grandori and Eisenman, 1997; Guttman and Rinn, 2012; Karreth and Pandolfi, 2013; Lee, 2012; Lieberman et al., 2013; Orom and Shiekhattar, 2013; Oster et al., 2002; Prensner and Chinnaiyan, 2011; Ulitsky and Bartel, 2013). Soon after the initial cloning of lncRNAs, which identified H19 and XIST from cDNA libraries, two studies using tiling arrays estimated that the number of lncRNAs is at least similar to that of protein coding genes (Bartolomei et al., 1991; Brown et al., 1991; Kapranov et al., 2002; Rinn et al., 2003; Rinn et al., 2007; Yoo-Warren et al., 1988). Technical advances in array (Gupta et al., 2010; Rinn et al., 2007), chromatin signature analysis (Guttman et al., 2009; Khalil et al., 2009), computational analysis of cDNA libraries (Jia et al., 2010; Maeda et al., 2006), and RNA sequencing technologies (Cabili et al., 2011; Orom et al., 2010; Prensner et al., 2011) have revealed that thousands of lncRNAs are widely expressed and are under tight spatial and temporal regulation. The GENCODE consortium identified and validated 14,880 annotated human lncRNAs originating from 9277 gene loci with experimental evidence (Derrien et



al., 2012; Djebali et al., 2012). These findings indicate that lncRNAs are independent transcriptional units that have been the subject of weak selective pressure throughout evolution. Additionally, similarly to mRNAs, many lncRNAs are primate-specific, subject to histone modifications, and show tissue-specific expression.

lncRNAs have a broad and diverse repertoire of functions. In the regulation of gene expression by lncRNAs, they function at the level of transcription, post-transcriptional processing, and chromatin modification (Mercer et al., 2009; Mercer and Mattick, 2013). One example of a lncRNA regulating transcription would be the repression of cyclin D1 transcription in human cell lines (Wang et al., 2008b). The induction of lncRNAs associated with the cyclin D1 promoter can be triggered by DNA damage. By recruiting the RNA-binding protein TLS, both CREB and p300 activities are inhibited, resulting in the silencing of cyclin D1 expression. Additionally, lncRNAs have been reported to act as co-factors that modulate transcription factor activities or regulate RNA polymerase II activities by interacting with the initiation complex, forming a triplex at the promoter (Feng et al., 2006; Martianov et al., 2007; Ohno et al., 2002). lncRNAs have the ability to recognize complementary sequences, allowing them to act in the post-transcriptional processing of mRNA because of this RNA–RNA specific interaction. This is exemplified by cases in which antisense lncRNAs mask splicing sites in mRNA, resulting alternative splicing (He et al., 2008).

At the epigenetic level, lncRNAs can regulate the transcriptional activity of specific genes or entire chromosomal regions by recruiting polycomb protein

complexes, which are involved in chromatin modifications in either *cis*- or *trans*-acting manners. Xist, a 17 kb X chromosome-specific non-coding transcript, represents one of the earliest characterized and most well-known examples, initiates X chromosome inactivation by coating the X chromosome and recruiting polycomb repressive complexes 2 (PRC2) in *cis* (Brown et al., 1992; Jeon and Lee, 2011; Zhao et al., 2008). The HOTAIR lncRNAs regulate the HoxD cluster of genes in *trans* by serving as a scaffold that enables the RNA-mediated assembly of PRC2 and LSD1, and facilitating binding of this complex to chromatin (Tsai et al., 2010). Various models have been proposed for this specific lncRNA function. lncRNAs contain discernable protein interaction domains that can recruit protein components and act as a scaffold, thereby assisting in the formation of unique functional complexes (Engreitz et al., 2013; Gupta et al., 2010; Huarte et al., 2010; Jeon and Lee, 2011; Simon et al., 2013; Yang et al., 2013b; Yang et al., 2011). RNA–DNA and RNA–RNA interactions can also be included in this process. Alternatively, lncRNAs can act as guides to recruit proteins, namely, the chromatin modification complexes, to target loci. This might occur through RNA–DNA interactions or RNA interactions with DNA-binding proteins (Grote et al., 2013; Huarte et al., 2010; Zhao et al., 2008). Moreover, lncRNAs have been suggested to function as decoys that bind to proteins or miRNAs (Di Ruscio et al., 2013; Hansen et al., 2013; Hung et al., 2011; Memczak et al., 2013; Poliseno et al., 2010; Tay et al., 2011; Tripathi et al., 2010; Tripathi et al., 2013) or act as enhancers that are transcribed from an

enhancer region or a nearby locus to influence gene transcription (Kim et al., 2010b; Lai et al., 2013; Li et al., 2013; Melo et al., 2013; Wang et al., 2011).

## **LncRNAs in cancers**

LncRNA expression profiles differ dramatically between tumors and normal tissues. Previous large-scale RNA-seq analysis of prostate cancer tissues identified a group of 121 lncRNAs that showed distinguishable expression patterns in normal and cancerous tissues (Prensner et al., 2011). Because lncRNA expression patterns tend to be more tissue-specific than those of protein coding transcripts (Cabili et al., 2011; Derrien et al., 2012), lncRNA expression signatures might serve to determine the developmental lineage and tissue origin of human cancers with greater accuracy (Bartolomei et al., 1991; Prensner and Chinnaiyan, 2011; Ulitsky and Bartel, 2013). Additionally, the association of several lncRNAs, including MALAT-1, HOTAIR, PCAT-1, and LET, with cancer metastasis has been established, implying that lncRNAs could also be used as a biomarker for predicting cancer prognosis and survival (Gupta et al., 2010; Ji et al., 2003; Prensner et al., 2011; Yang et al., 2013a). Even though lncRNA research remains in its early stages, it is clear that lncRNAs are involved in regulating cell proliferation (Hung et al., 2011; Tripathi et al., 2013), cell differentiation (Guttman et al., 2009; Guttman et al., 2011; Kretz et al., 2013; Loewer et al., 2010; Ulitsky et al., 2011), cell migration (Gupta et al., 2010; Ling et al., 2013; Orom et al., 2010; Yang et al., 2013a), immune responses

(Carpenter et al., 2013; Gomez et al., 2013), and apoptosis (Huarte et al., 2010). Therefore, it can be hypothesized that the dysregulation of lncRNAs will contribute to tumorigenesis. Many lncRNAs have been shown to function as oncogenes or tumor suppressors. HOTAIR can trigger metastasis by its actions as a scaffold to link PRC2 and LSD1 (Gupta et al., 2010), thereby coordinating their chromatin modifying capacities. ANRIL, an antisense lncRNA of the CDKN2A/2B gene, represses INK4A/B expression by binding to CBX7/PRC1 and SUZ12/PRC2 (Kotake et al., 2011; Pasmant et al., 2007; Yap et al., 2010). Xist deletion results in the development of aggressive myeloproliferative neoplasms and myelodysplastic syndrome in female mice (Yildirim et al., 2013).

### **MYC, lncRNA, and gene regulation**

As previously mentioned, MYC is widely recognized to be one of the most important transcription factors and is responsible for regulating the expression of a large proportion of human genes; however, the exact mechanisms of many of the various modes of MYC regulation remain to be fully characterized. This thesis aims to fill a gap in MYC target gene regulation network by adding a missing 'link', lncRNA.

In chapter 2 of the thesis, I explore the general connection between MYC and lncRNAs using microarray, RNA-seq, and TCGA analyses. After validation, a few interesting lncRNAs are carefully examined, and I focus on DANCR (Differentiation Antagonizing Non-protein Coding RNA) as the one to pursue in

greater detail, and the use to generate a model for the combined role of MYC and lncRNAs in gene regulation. In Chapter 3, using several robust methods, such as ChIP, ChIP-seq, and further TCGA analysis focusing on DANCR, its validation as a *bona fide* MYC target is described. In chapters 4 and 5, further studies of the genomic and biological impacts of DANCR in human cancers are presented. Loss of function experiments revealed its impact on cell growth, and cell cycle analysis showed that growth arrest can be triggered by DANCR depletion that blocks the transition from G1 to S phase. Microarray gene expression analysis followed by GSEA (gene set enrichment analysis) generated a list of downstream genes that might be related to the phenotype I observed; among these genes, p21 was a strong candidate to be a causative gene. Chapter 6 describes the proposed functions of DANCR; I first introduce EZH2 by confirming that it is also a transcriptional regulator of p21. Then, I will propose a MYC-DANCR/EZH2-p21 model and prove it through a series of experiments that establish connections between each component of my proposed model, including a direct link between DANCR and EZH2, and their regulation of p21. A novel mode of lncRNA recognition of target genes through RNA–RNA interactions is also proposed, for which I also provide evidence.

## **Chapter 2**

### **MYC-induced lncRNAs in human cancer**

## Global effects of MYC on lncRNAs

To explore the genome-wide effects of MYC on lncRNAs, I performed microarray and RNA-seq analyses in a human B lymphocyte cell line, P493 cells, in which MYC expression levels can be manipulated. P493 cells are derived from EREB2-5, a human B lymphocyte immortalized with the EBV viral genome and engineered to harbor a tetracycline (tet)-repressible MYC transgene. In the presence of tet, P493 cells exhibit low levels of MYC expression, whereas in the absence of tet, cells have high levels of MYC expression and show a morphology that resembles Burkitt's lymphoma cells (Pajic et al., 2000). Cells arrest in G1 phase after tet treatment, whereas induction of MYC upon the removal of tet allowed for cell cycle progression from G1 to S phase (O'Donnell et al., 2006). P493 cells are tumorigenic in SCID mice and can be used to generate a subcutaneous xenograft model for *in vivo* tumor formation experiments (Gao et al., 2007).

I chose to use gene expression arrays as an initial approach to gain insights into the MYC-regulated gene network. P493 cells were treated with 0.1 µg/ml tetracycline (Sigma) for 48 h, and then harvested alongside untreated control cells. Total RNA was extracted and hybridized to an Affymetrix exon array. The expression profiles of MYC-regulated genes were obtained. The results showed, as expected, that the expression of many genes was altered by increased MYC levels, including genes that are involved in cellular processes such as various metabolic pathways, cell proliferation, and mitochondrial function

(Meyer and Penn, 2008). For the purpose of this project, I decided to focus on those genes whose expression was induced by MYC. A summary of the various genes induced by MYC is shown in Table 1. Interestingly, two lncRNAs, DANCR (previously known as KIAA0114) and LOC286467, seemed to be particularly highly induced by MYC, especially DANCR, which was ranked 11<sup>th</sup> in the list and showed more than a 14-fold change in expression upon MYC induction.

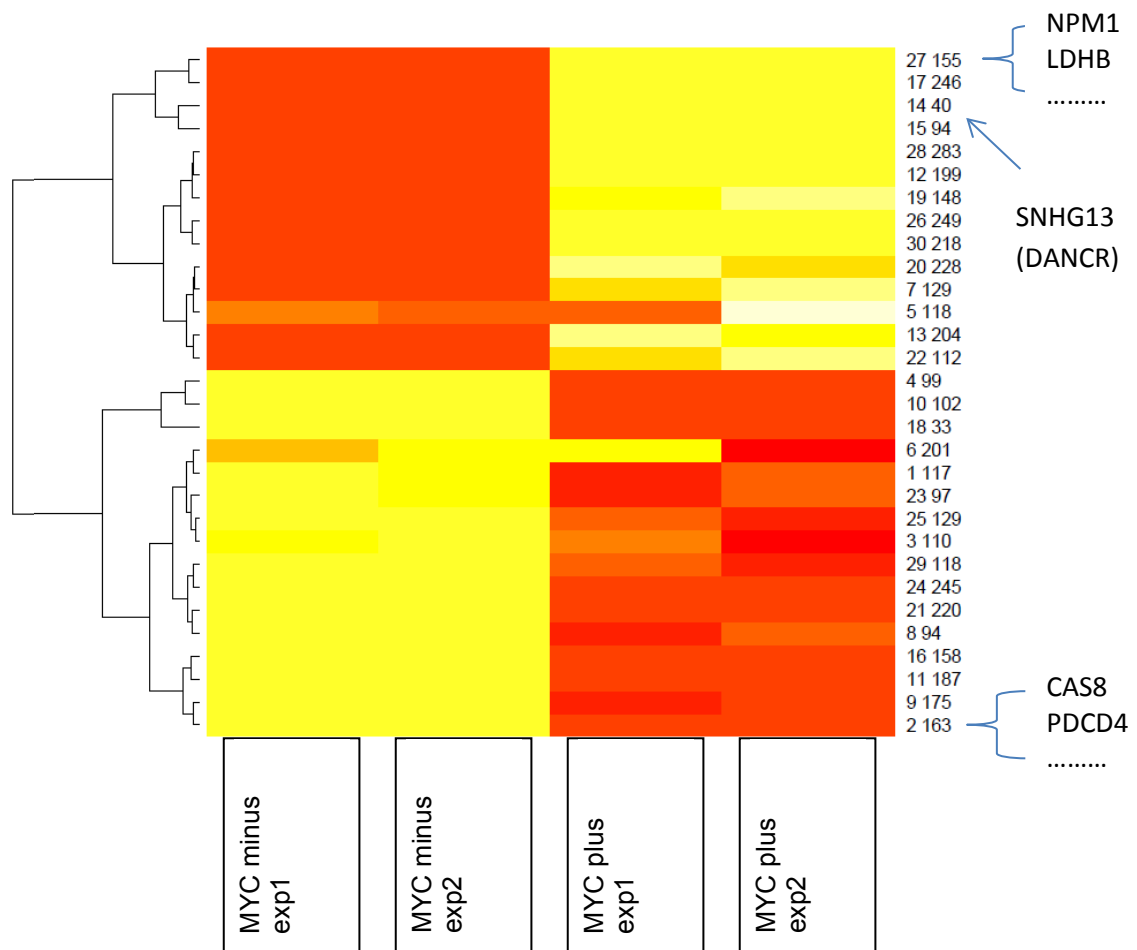
However, one limitation of gene expression arrays is the absence of unannotated transcripts; a large proportion of lncRNAs are still in the process of being annotated, and so they are not included on array chips. I next performed RNA-seq, which is currently the most comprehensive and unbiased technique to globally assess MYC regulation of the human genome. Total RNA samples were extracted from duplicates of high- and low-MYC P493 cells, which I made into libraries using a TruSeq RNA Sample Preparation Kit (Illumina, San Diego, CA, USA) and then sequenced using an Illumina HiSeq 2000 instrument. The RNA-seq reads were subsequently aligned to the human genome using Tophat (Trapnell et al., 2009). Unsupervised hierarchical clustering showed patterned variations in the expression of transcripts that were either induced or repressed by MYC (Figure 1). Again, in this thesis, I focused on the genes that were induced by MYC. By arbitrarily setting a cut-off threshold at a 2-fold expression change after MYC induction, I found that 3582 transcripts remained, including 386 non-coding RNA transcripts, among which 230 were lncRNAs (Figure 2). After a closer examination of the top portion of this list of lncRNAs, I discovered that DANCR (annotated as ‘SNHG13’ in this list) was once again ranked 2<sup>nd</sup> of all



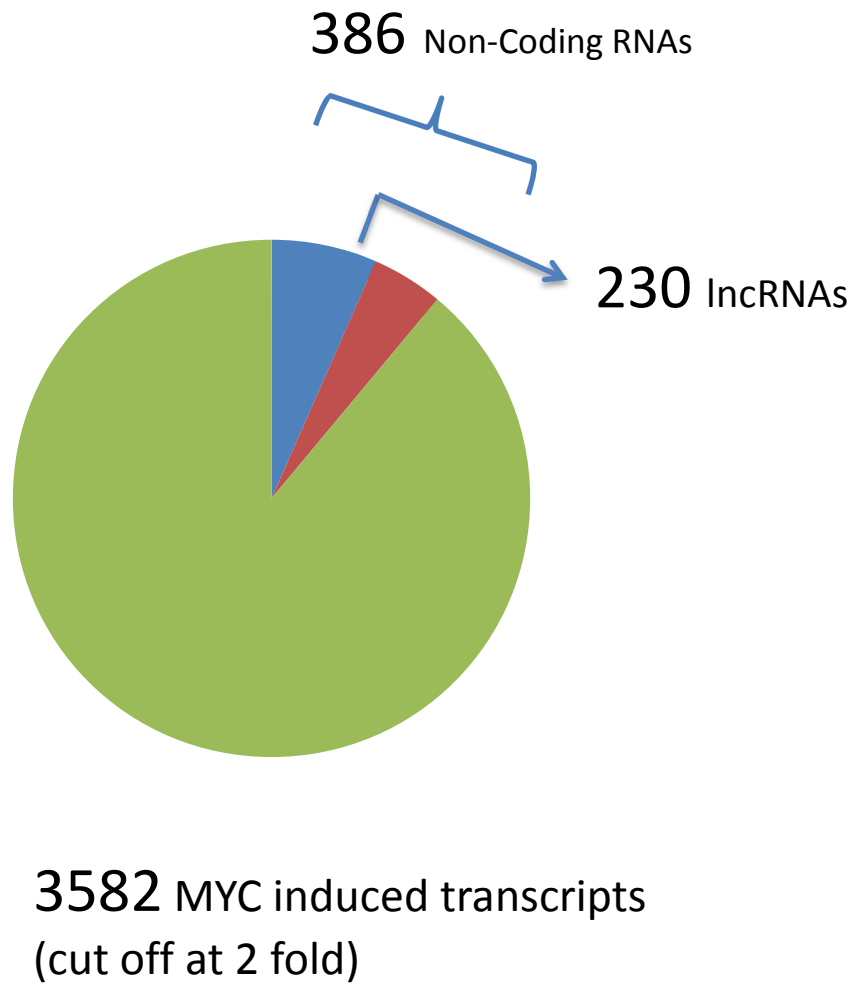
MYC regulated lncRNAs with a 12-fold change, and SNHG4 was the only lncRNA with a higher fold-change for the with vs. without MYC genes (Table 2).

Gene Symbol	EntrezID	Untreated vs. Tet Lin(FC)	Untreated vs. Tet (p-value)	# of markers
XRCC2	7516	31.2488	7.55E-05	2
PEG10	23089	19.885	7.67E-08	22
WDR43	23160	19.5475	2.26E-05	1
SCD	6319	17.996	1.40E-06	8
SLC7A5	8140	17.5722	7.10E-08	18
MARS2	92935	16.7312	2.76E-06	7
PUS7	54517	16.2558	2.69E-06	26
CAMKV	79012	16.1959	3.13E-08	18
ZC3HAV1L	92092	14.8296	1.45E-06	7
VDAC1	7416	14.4356	0.000137321	1
<b>KIAA0114</b>	<b>57291</b>	<b>14.2</b>	<b>3.36E-06</b>	<b>9</b>
.....	.....	.....	.....	.....
TKT	7086	4.71888	1.44E-05	30
MRPL3	11222	4.71421	2.05E-05	17
HELLS	3070	4.70877	0.000543929	39
URB1	9875	4.6924	9.50E-07	48
CENPH	64946	4.69164	4.35E-05	15
SF3B5	83443	4.68191	0.000699001	3
<b>LOC286467</b>	<b>286467</b>	<b>4.68</b>	<b>0.00093</b>	<b>14</b>
C13orf3	221150	4.67043	0.00082175	16
CDCA7	83879	4.66028	7.04E-06	18
NAT13	80218	4.66219	3.03E-05	11

**Table 1. MYC-induced genes identified by gene expression array analysis.** MYC-induced genes ranked by fold-change (expression difference) between Hi- and Low-Myc conditions in P493 cells, generated using gene expression arrays. The two highlighted 'genes' are the top two lncRNAs identified based on MYC-induced fold-change in expression rankings.



**Figure 1. MYC-regulated gene clusters.** A heat map showing unsupervised hierarchical clustering of expression values of MYC-regulated gene clusters for Hi- vs. Low-MYC states in P493 cells. RNA-seq was performed for RNAs derived from four independent experiments. Each column represents an experiment and each row indicates a group of similarly expressed genes. Transcripts were divided into 30 groups based on MYC responsiveness, which is represented on a red-yellow color scale. On the right side of the figure, the first smaller number in is the group number, and the second number indicates the number of transcripts in that group. Selected genes are also identified on the far right.



**Figure 2. The number of lncRNAs induced by MYC.** A pie chart representing the proportions of MYC-induced non-coding sequences and lncRNAs among all MYC-induced transcripts showing at least a two-fold expression change. The numbers indicate the results from a RNA-seq experiment in P493 cells.

Gene	Transcript	cMycMinus	cMycPlus	FC(cMycMinus/cMycPlus)
SNHG4	NR_036536	6.82	109.36	-16.03
SNHG13	NR_024031	14.36	175.12	-12.19
BZW2	NR_027624	1.9	22.4	-11.8
LDHA	NR_028500	59.91	447.39	-7.47
RCC1	NR_030725	11.77	86.04	-7.31
RCC1	NR_030726	11.76	85.87	-7.3
SNHG3	NR_036473	14.24	100.28	-7.04
SNHG3	NR_002909	14.82	104.17	-7.03
SORD	NR_034039	4.45	30.18	-6.78
NME1-NME2	NR_037149	47.42	310.33	-6.54
PRMT1	NR_033397	22.81	147.34	-6.46
TMEM48	NR_033142	5.07	31.38	-6.18
SRPK1	NR_034069	10.52	57.59	-5.47
DTYMK	NR_033255	7.32	39.79	-5.44
LOC100499227	NR_034160	1.07	5.83	-5.44
MUTED-TXNDC5	NR_037616	6.67	36.1	-5.41
EARS2	NR_003501	3.85	20.48	-5.32
PES1	NR_036550	10.9	55.67	-5.11
NOP56	NR_027700	30.26	152.94	-5.05
MTHFD2	NR_027405	20.46	95.82	-4.68
GGCT	NR_037669	11.23	51.94	-4.63
PDIA5	NR_028444	3.21	14.75	-4.59
CKS1B	NR_024163	3.69	16.13	-4.37
SCML2	NR_033717	1.46	6.27	-4.3
ATL2	NR_024191	5.03	21.56	-4.28
LOC641518	NR_029374	1.36	5.8	-4.27
SNHG8	NR_034010	33.06	141.3	-4.27
SNHG8	NR_003584	26.32	112.28	-4.27
APOO	NR_026545	2.56	10.83	-4.23

**Table 2. MYC-induced genes identified using RNA-seq.** A list of the top MYC-induced lncRNAs ranked by fold-change (expression difference) between Hi- and Low-Myc conditions in P493 cells was generated based on RNA-seq data. Notably, SNHG13 is an alternative name for DANCR. Some of the lncRNA annotations in the table are named as known coding genes, but they are actually lncRNAs transcribed from the same locus as those genes.

## **Validation of microarray and RNA-seq data**

To validate my microarray and RNA-seq data, I performed quantitative real-time PCR analysis. I chose to analyze three of the previously mentioned possible MYC targets (DANCR, SNHG4, and LOC286467) using pre-validated primers. LOC286467 was found to have ~26 fold-changes and SNHG4 had over 40 fold-changes between Hi- and Low-MYC conditions in P493 cells (Figure 3). For DANCR, that fold-change was ~36-fold in P493 cells, and ~6-fold in PC3 cells (Figure 4), a cell line that I will introduce and discuss in detail in later chapters. The 6-fold change was probably a consequence of inefficient MYC siRNA knock-down in PC3 cells compared with the more efficient tetracycline-regulated MYC in P493 cells. Taken together, the expression of all three lncRNAs were validated to be markedly induced by changes in MYC levels.

## **TCGA analysis of lncRNAs in cancer**

To further investigate the role of lncRNAs and their connections with MYC in real life cancers, I collaborated with the lab of Lin Zhang at University of Pennsylvania to analyze The Cancer Genome Atlas (TCGA) data. We retrieved RNA-seq profiles from 4015 specimens across 11 cancer types, along with 350 corresponding normal control specimens across 7 matching tissue types from TCGA. A heatmap was generated for MYC/lncRNAs correlations, and,

interestingly, DANCR was found to show increased expression in the high MYC-expressing primary tumors samples mentioned above (Figure 5).

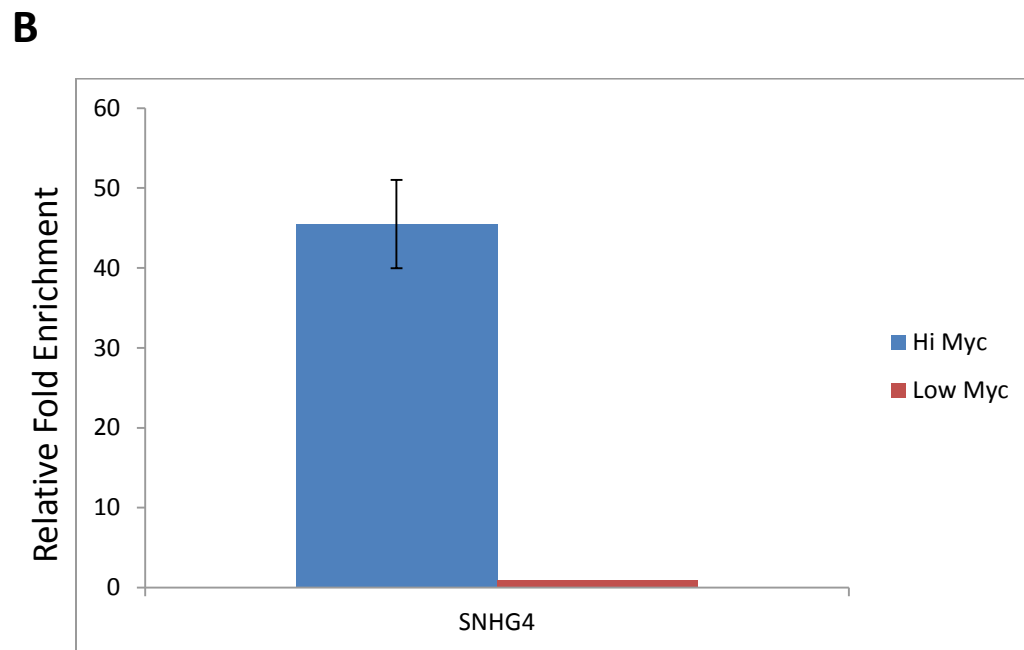
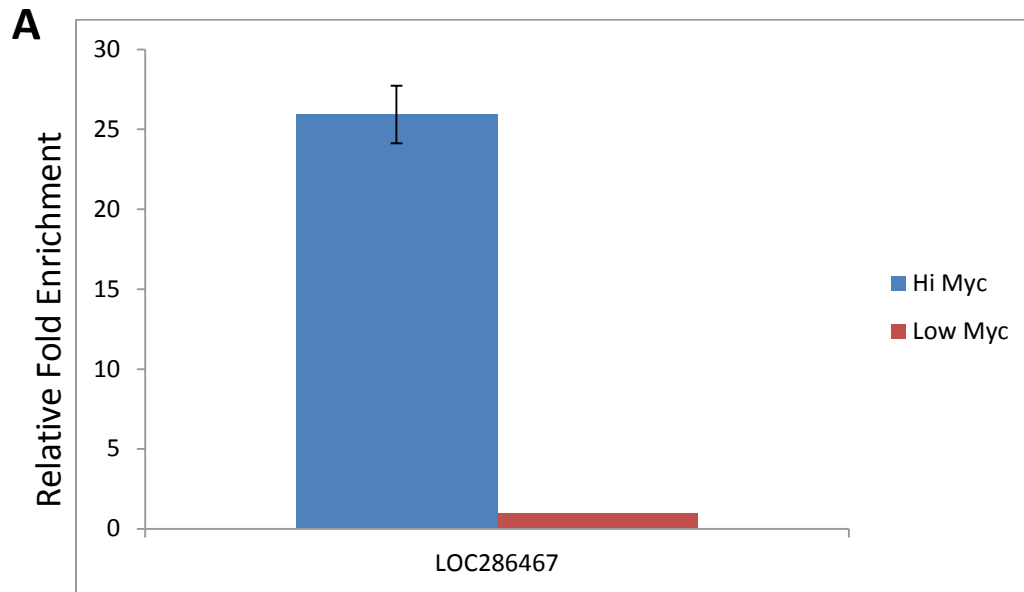
### **MYC/DANCR co-expression analysis in primary cancers**

To probe the role MYC regulation of DANCR in cancer tissues, we used the co-expression analysis method (described in the methods section), which is also known as guilt-by-association (GBA) analysis. GBA analysis for DANCR was performed for the RNA-seq data to identify the most significantly positive and negative co-expressed protein-coding mRNAs in breast and prostate cancers. When I 'zoomed in' on MYC/DANCR correlations, I found that in both breast and prostate cancers, MYC and DANCR are co-expressed with a  $R^2$  value of 0.22 or 0.23 in breast and prostate cancers, respectively, and these correlations were statistically significant ( $P < 0.00001$   $P = 0.00003$ , respectively; Figure 6).

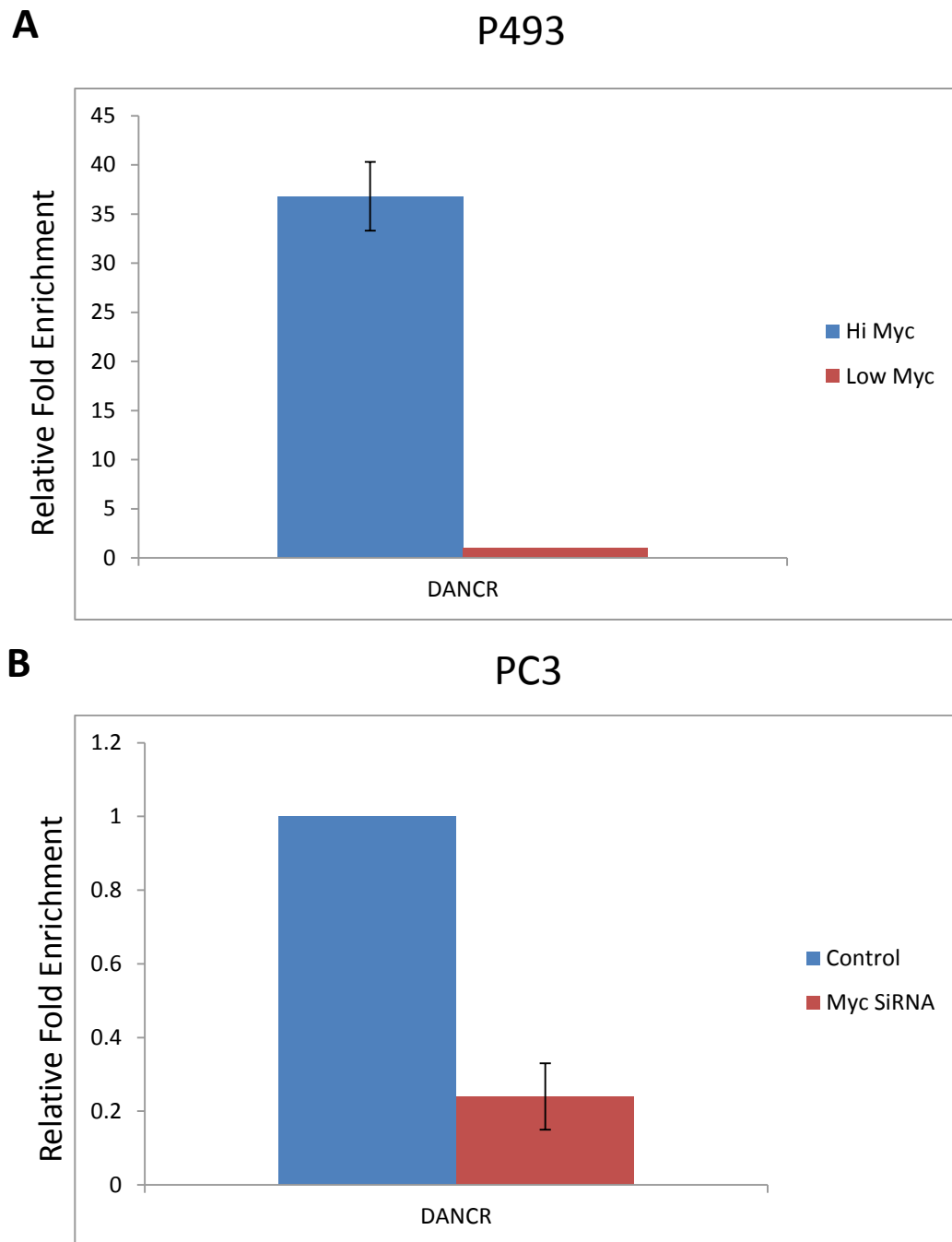
The evidence I have obtained to date all point to DANCR being one of the most highly regulated MYC-dependent transcripts, including both coding and non-coding transcripts. The tight regulation of DANCR by MYC *in vitro* and *ex vivo* in tumors, combined with a recent report in which DANCR was shown to be down-regulated during cell differentiation and required to enforce the undifferentiated cell state (Kretz et al., 2012), suggests that it must confer significant functional importance, which will be characterized in later chapters. Additionally, I carried out preliminary experiments to study the biological

importance of LOC286467 and SNHG4, although I observed no phenotypes associated with these two transcripts (data not shown). Thus, I will only focus on DANCR for the remainder of this thesis.

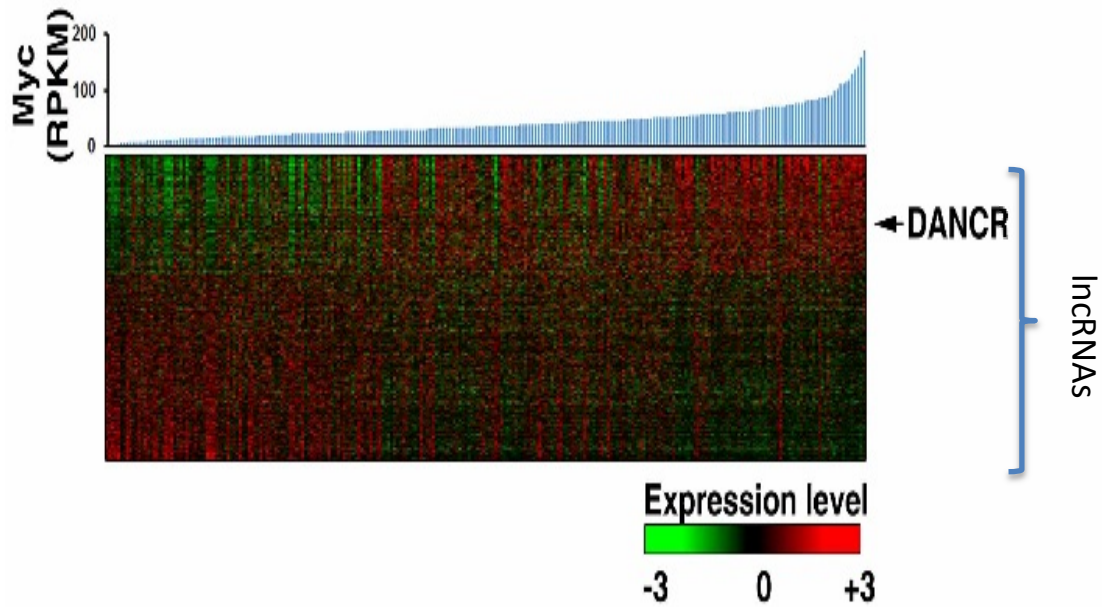




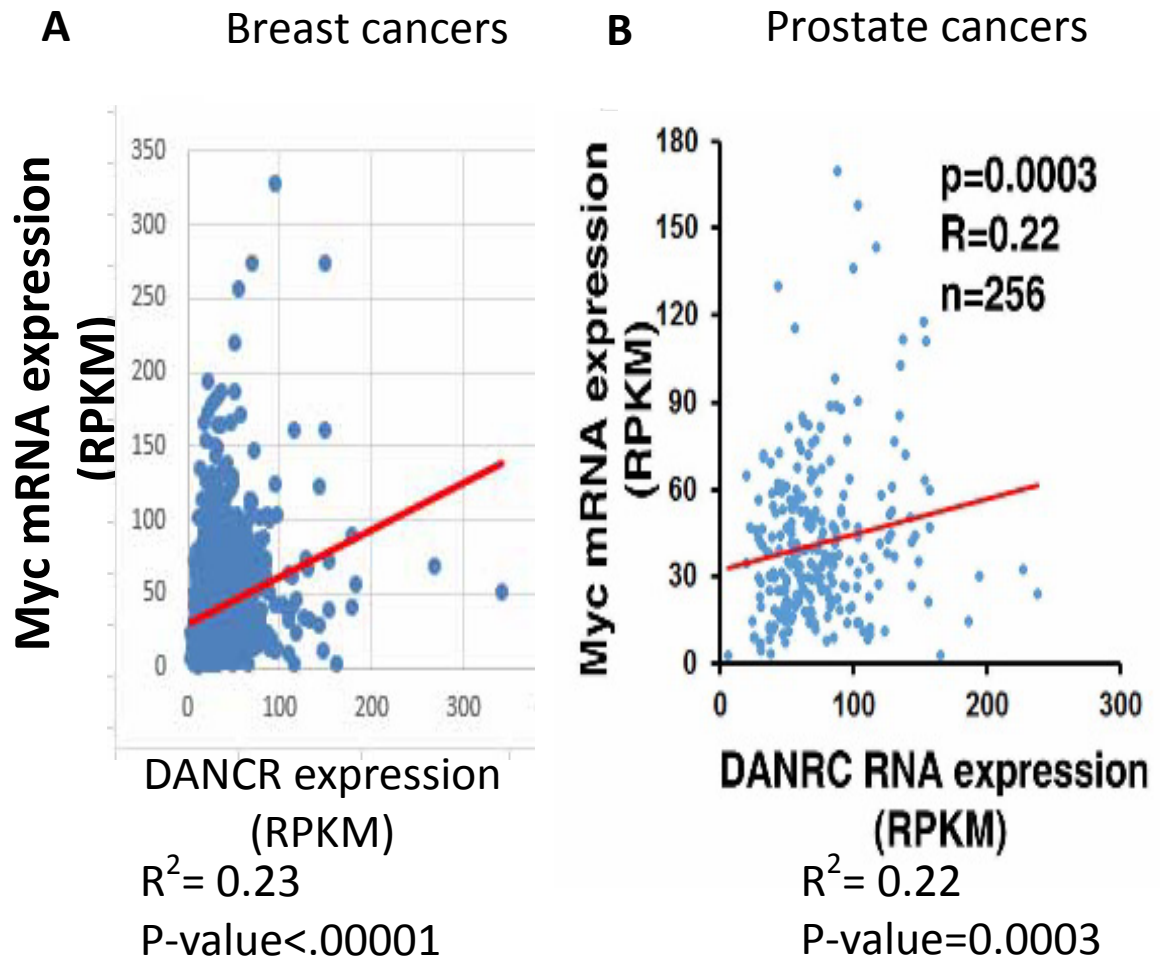
**Figure 3. qRT-PCR validation of the induction of lncRNA expression by MYC.** This experiment was performed in P493 cells. Bar graphs represent the mean mRNA levels from triplicate samples. Error bars represent the standard deviation (n = 3). (A) LOC286467; (B) SNHG4.



**Figure 4. qRT-PCR validation of the induction of DANCR by MYC.** Bar graphs represent the mean mRNA levels from triplicate samples. Error bars represent the standard deviation (n = 3). Experiments were performed in (A) P493 cells of (B) PC-3 prostate cancer cells with the siRNA depletion of MYC.



**Figure 5. A heatmap of IncRNAs that are co-expressed with MYC.** The heatmap generated from an analysis of TCGA RNA-seq data. MYC expression is depicted at the top as reads per kb per million (RPKM). Each column represents a single clinical sample, and each row represents a lncRNA. From top to bottom, lncRNAs are arranged in the order of high to low expression values in a sample, which represented using a red–green color scale. DANCR is one of the top MYC co-expressed lncRNAs in cancers.



**Figure 6. Correlations between MYC and DANCR expression in breast and prostate cancers.** Graphs were generated from TCGA RNA-seq data from breast (n = 1096) and prostate (n = 256) cancer tissues. Guilt by association (GBA) analysis of TCGA RNA-seq data was performed in collaboration with Dr. Lin Zhang. RPKM, reads per kilobase per million. (A) Breast Cancer; (B) Prostate Cancer.

## **Materials and Methods**

### **Cell culture**

P493 cells were a generous gift of Prof. D.Eick at the Institute of Clinical Molecular Biology and Tumors Genetics, GSF-Research Centre, Munich, Germany. The PC3 cell line was obtained from the American Type Culture Collection (ATCC). Cells were cultured in RPMI 1640 with 10% fetal calf serum (FCS) and 1% penicillin/streptomycin. To repress MYC expression, cell cultures were supplemented with 0.1 µg/ml tetracycline (Sigma) for 48 h and harvested.

### **Microarray hybridization and analysis**

At 48 h after tetracycline treatment, samples were processed by the Johns Hopkins Medical Institutions Deep Sequencing & Microarray Core Facility. RNA samples from these cells were used to probe an Affymetrix 1.0 ST exon array. Total RNA was isolated from cells using a RNeasy Plus Mini Kit (Qiagen) according to the manufacturer's instructions. Approximately 5 µg per chip of fragmented and labeled cDNA was hybridized to the array. Fluorescence was detected using an Affymetrix GeneChip Scanner. Analysis was carried out using the Li and Wong method (Li and Wong, 2001).

## **RNA-seq**

Total RNA from cells was extracted using TRIzol (Invitrogen) and a RNeasy Plus Mini Kit (Qiagen). Extracted RNA samples underwent quality control (QC) assessment using an Agilent Bioanalyzer (Agilent, Santa Clara, CA, USA) and all RNA samples that were submitted for sequencing had an RNA Integrity Number (RIN) >9, with a minimum of 1 µg input RNA. Poly-A library preparation and sequencing were performed at the Penn Next Generation Sequencing Core following standard protocols. Briefly, I started with total RNA, and mRNA was purified using poly(A) selection or rRNA depletion before the RNA was chemically fragmented and converted into single-stranded cDNA using random hexamer priming. Next, the second strand was generated to create double-stranded cDNA. Library construction began with the generation of blunt-end DNA fragments from ds-cDNA. Then, A-bases were added to blunt-ends to prepare them for the ligation of sequencing adapters using the TruSeq RNA Sample Preparation Kit (Illumina). Fragments of ~300 to ~400 bp were selected by gel electrophoresis, followed by 15 cycles of PCR amplification. The prepared libraries were then sequenced using an Illumina HiSeq 2000, which generated ~50 to ~100 million unfiltered reads per sample. RNA-seq data were aligned to the hg19 reference genome using Tophat with the default options.

### **Quantitative real-time PCR (RNA quantification)**

Total RNA was extracted from cells using a RNeasy-plus Mini Kit (Qiagen) according to the manufacturer's instructions. RNA concentrations and quality ( $A_{260/280}$  between 1.9 and 2.1) were determined using a Nanodrop. Then, 1–2  $\mu\text{g}$  RNA was reverse transcribed to cDNA using the Taqman-based reverse transcription. Quantitative RT-PCR was carried out using Power SYBR Green PCR Master Mix (Applied Biosystems) with a StepOnePlus™ Real-Time PCR System (ABI). Melting curve analysis was performed at the end of each run and all PCR reactions were performed in triplicate. The relative amount of each gene of interest was calculated using the  $\Delta\Delta\text{Ct}$  method. Sequence-specific primer pairs were designed to cross exon–exon junctions.

### **TCGA analysis**

To analyze expressional aberrations of lncRNA in human cancers, we retrieved RNA sequencing (RNA-seq) profiles of 4015 human tumor specimens across 11 cancer types and 350 corresponding normal control specimens across seven types of matching tissues from TCGA. Initially, we identified ~2000 manually annotated lncRNA genes and calculated the RPKM (reads per kilobase per million mapped reads) for each lncRNA gene. The log-transformed median center of the RPKM values that correspond to a given lncRNA gene was used to represent the normalized expression level of that lncRNA.

Recently, a bioinformatic approach, termed guilt-by-association (GBA) analysis, has been used to gain a global understanding of lncRNAs and protein-coding genes that are significantly co-expressed and, thus, presumably are correlated. Briefly, the GBA approach using gene co-expression analyses can identify protein-coding genes and pathways that are significantly correlated with a given lncRNA. This work was performed in collaboration with the laboratory of Lin Zhang (University of Pennsylvania).



## **Chapter 3**

### **LncRNA DANCR is a MYC target**

As described in previous chapters, MYC regulates cell growth, proliferation, metabolism, and apoptosis, as well as the transcription of a large proportion of all human genes. However, within its complex target gene network, MYC can affect certain target genes more than others. By identifying these critical points or nodes, I can better understand the network whereby MYC can regulate cellular processes, including dysregulated proliferation in cancer. As lncRNAs are emerging as an important mediator of biological functions in cells, herein I attempted to identify key MYC-regulated lncRNAs that are part of this network. Additionally, by exploring the MYC-DANCR relationship, I aim to establish a model of MYC pathways and to identify its critical lncRNA targets.

The large-scale microarray and RNA-seq experiments described in the last chapter showed that DANCR demonstrates great, if not the greatest, MYC-induced responsiveness among all lncRNAs. In this chapter, I sought to determine whether DANCR is directly regulated by MYC by testing MYC binding to the genomic DANCR locus.

### **Chromatin immunoprecipitation to validate DANCR as a direct MYC target**

Although *DANCR* mRNA transcript levels can be positively regulated by MYC, whether DANCR is a direct or indirect MYC target has not yet been fully established. I used chromatin immunoprecipitation (ChIP) to determine where MYC binds within target genes (Zeller et al., 2001). The tendency of MYC to bind to phylogenetically conserved canonical E boxes, 5'-CACGTG-3', in the promoter

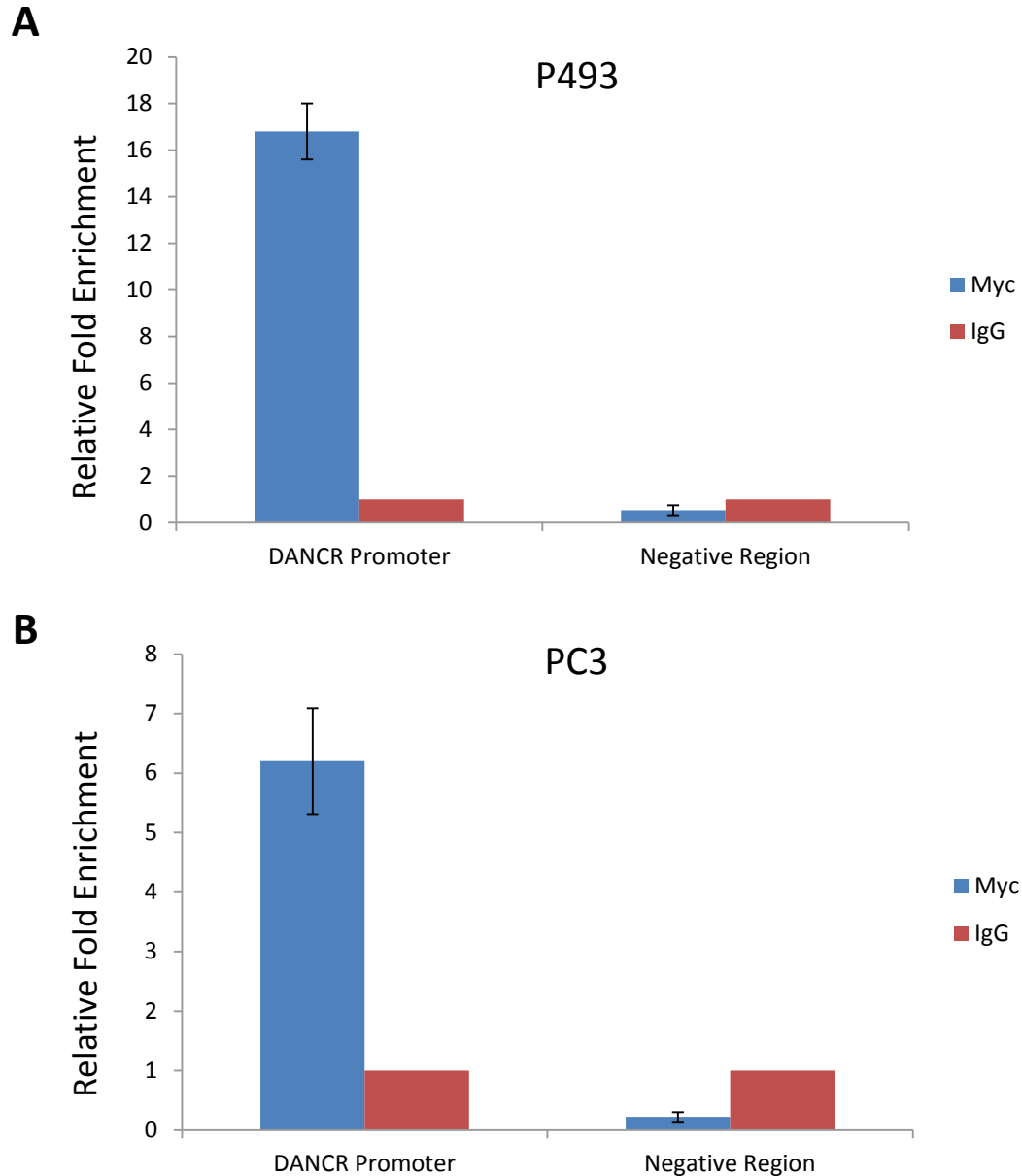
or intron 1 of target genes has been observed previously (Haggerty et al., 2003; Zeller et al., 2003). To extend these findings, I conducted a detailed analysis of DANCR genomic sequences. In this analysis, I localized possible MYC binding sites in the DANCR promoter region, which contains a canonical E box sequence and designed primers to cover that region. As negative controls, I carried out ChIP using a control IgG in place of anti-MYC antibody, and I also amplified an intergenic locus region with no previously known binding to MYC as an additional negative control. ChIP assays were carried out in P493 cells, PC3 cells, and human embryonic stem cells (H1 and H9), followed by quantitative real-time PCR. As expected, DANCR promoter regions showed high levels of MYC binding relative to both IgG controls and to negative control regions in all cell lines examined (Figures 7 and 8).

### **The pattern of MYC binding to the DANCR promoter in various cell types**

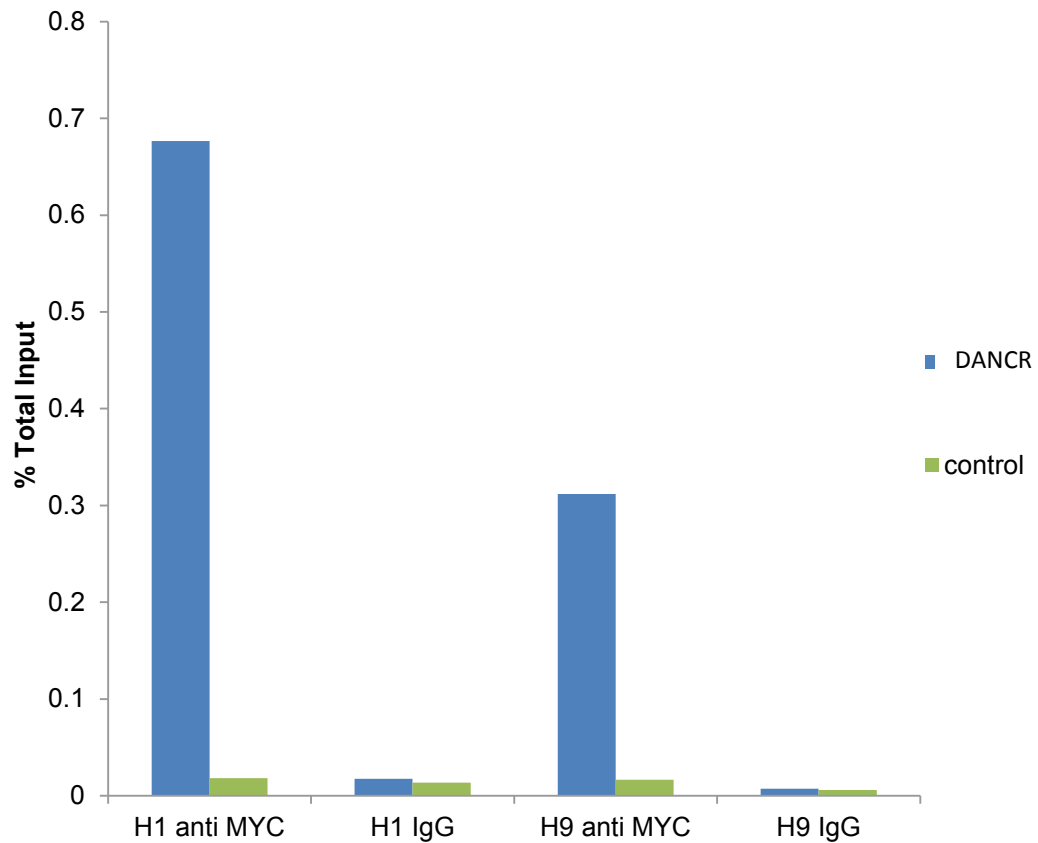
Furthermore, using publicly available Encode ChIP-seq data from UCSC genome browser, I first examined the temporal MYC–DANCR binding patterns in P493 cells. There appeared to be three binding sites for MYC along the DANCR genomic sequences, with the one containing the E-box sequence being the strongest. However, upon closer examination, I observed no MYC binding after tetracycline treatment (to induce MYC depletion), but MYC binding to the E-box-containing site was recovered as early as 1 h after tetracycline removal, and MYC binding to all three sites was recovered 24 h after tet removal (Figure 9). By

contrast, some MYC targets, such as NPM1, exhibit weak MYC binding, even after tet treatment, possibly because of endogenous MYC activity and leakiness of the engineered MYC-responsive unit. These observations suggest that MYC–DANCR binding is more sensitive to MYC levels than some of the other MYC targets. MYC was also found to bind specifically (with definite peaks) in the vicinity of the DANCR promoter region in GM12878, HeLa, Hep02, HUVEC, K562, and MCF7 cells (Figure 10).

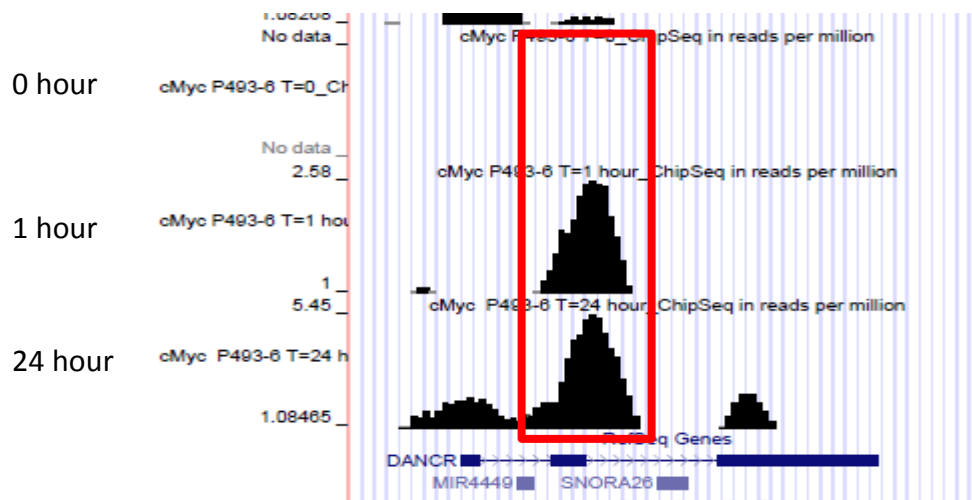
In summary, these *in vitro* data indicate that MYC binds to a canonical E-box sequence in the promoter region of DANCR, and establish that DANCR is a direct MYC target.



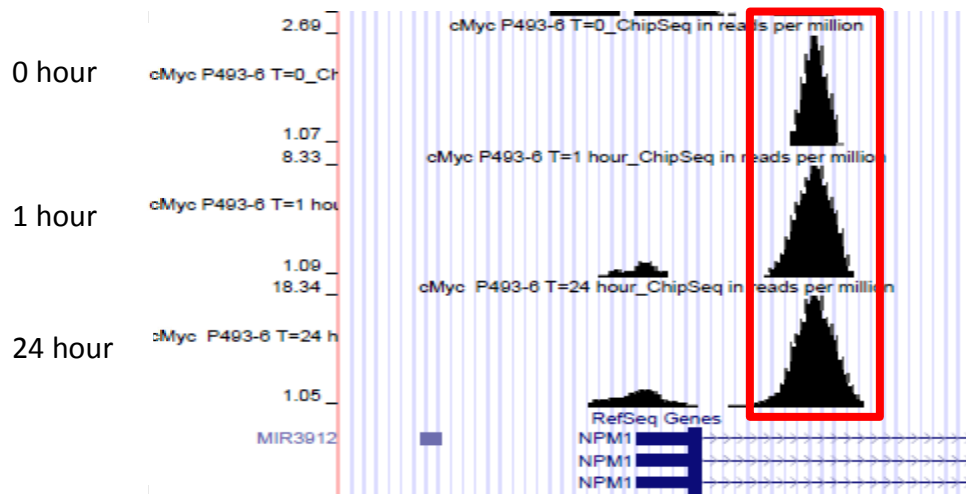
**Figure 7. Chromatin immunoprecipitation validation of DANCR as a direct target of MYC.** CHIP validation of MYC binding to the DANCR promoter region in P493 and PC3 cells. Quantitation of MYC binding by real-time PCR analysis (expressed as the fold-change over IgG), a negative intergenic region that is dozens of kb away from the DANCR promoter region was selected as another negative control. Error bars indicate standard deviation (n = 3). (A) P493 B-cell lymphoma cell line. (B) PC3 prostate cancer cell line.



**Figure 8. Chromatin immunoprecipitation validation of DANCR as a direct MYC target in human embryonic stem cells.** CHIP validation of MYC binding to the DANCR promoter region in hES cells. Quantification of MYC binding by RT-PCR (expressed as the percentage of total input DNA).

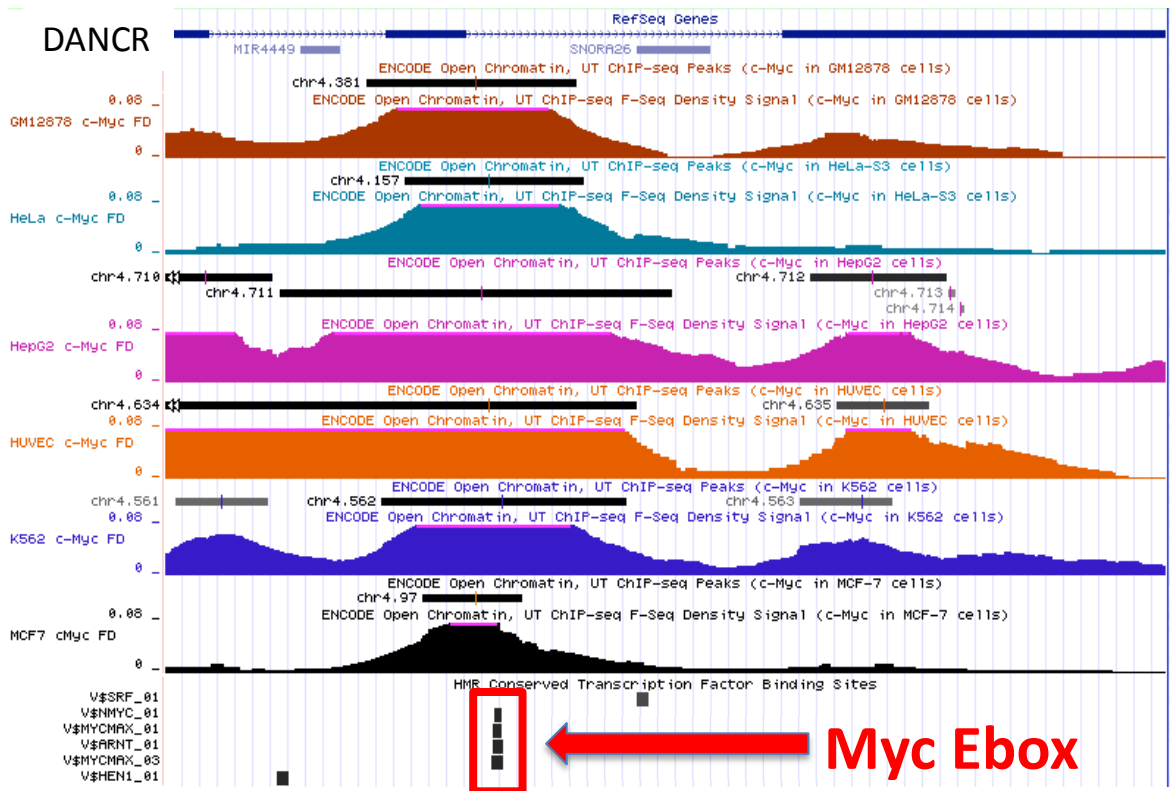


DANCR



NPM1

**Figure 9. CHIP-seq data-based assessment of MYC–DANCR and MYC–NPM1 binding.** Binding of MYC is shown at different time points in P493 cells. MYC binding to the DANCR and NPM1 promoters at 0, 1, and 24 h after tetracycline release is shown. The hg19 genome assembly, obtained from publicly available data at <http://genome.ucsc.edu>, was used.



**Figure 10. MYC binding to the DANCER promoter in six different cell types.** Data shown above were obtained from publicly available data at <http://genome.ucsc.edu>, using the hg19 genome assembly.



## **Materials and Methods**

### **Cell Culture**

P493 and PC3 cells were cultured as described in Chapter 2. The hES cells were maintained by Dr. Karen Zeller in the Dang lab. Briefly, hES cells (H1 and H9 clones) were grown on mouse embryo fibroblasts (MEFs) in culture medium that consisted of 80% DMEM/F12 supplemented with 20% knock-out serum replacement (Invitrogen), 1 mM L-glutamine, 0.1 mM 2-mercaptoethanol, and 1% nonessential amino acids (Sigma–Aldrich).

### **Chromatin Immunoprecipitation**

ChIP was performed using a Imprint Ultra Chromatin Immunoprecipitation Kit (Sigma–Aldrich), following the manufacturer’s protocol. Briefly, cells were cross-linked in 1% formaldehyde for 10 min at room temperature, after which the reaction was stopped adding glycine to a final concentration of 0.125 M. Cells were scraped, collected, and then rinsed twice with PBS. Cells were then lysed in ‘swelling buffer’ and the nuclei were extracted. Nuclear lysis buffer was used to lyse nuclei and sonication (30 sec intervals between each 30 sec sonication at 3 × 10 min) was performed afterwards. Sonicated DNA was blocked with Staph A beads and salmon sperm DNA, followed by overnight immunoprecipitation with anti-MYC antibody (sc-764, Santa Cruz Biotechnology) or IgG control. Staph A

beads were used to 'pull-down' the complexes. Extensive washes were then carried out. The antibody–protein–DNA complexes were eluted from Staph A beads by vortexing. De-crosslinking was performed by adding NaCl and then heating overnight at 65°C, followed by RNase A and Proteinase K treatment. Finally, DNA was eluted and analyzed by quantitative real-time PCR.

### **Quantitative real-time PCR (ChIP validation)**

Quantitation of chromatin IP fragments was performed as described above using Power CYBR Green PCR master Mix (Applied Biosystems) according to manufacturer's instructions. The relative amounts of each ChIP sample were determined using the  $\Delta\Delta$ Ct method and data were presented relative to the IgG signal. Alternatively, known quantities of 10-fold dilutions of total input DNA were used to generate standard curves for each primer pair. The relative amounts of each ChIP sample (expressed as the percentage of total input) were determined in the linear range according to the respective CT values. For each primer set, dissociation curves were used to verify the correct PCR product.

## **Chapter 4**

# **DANCR in Human Cancers**

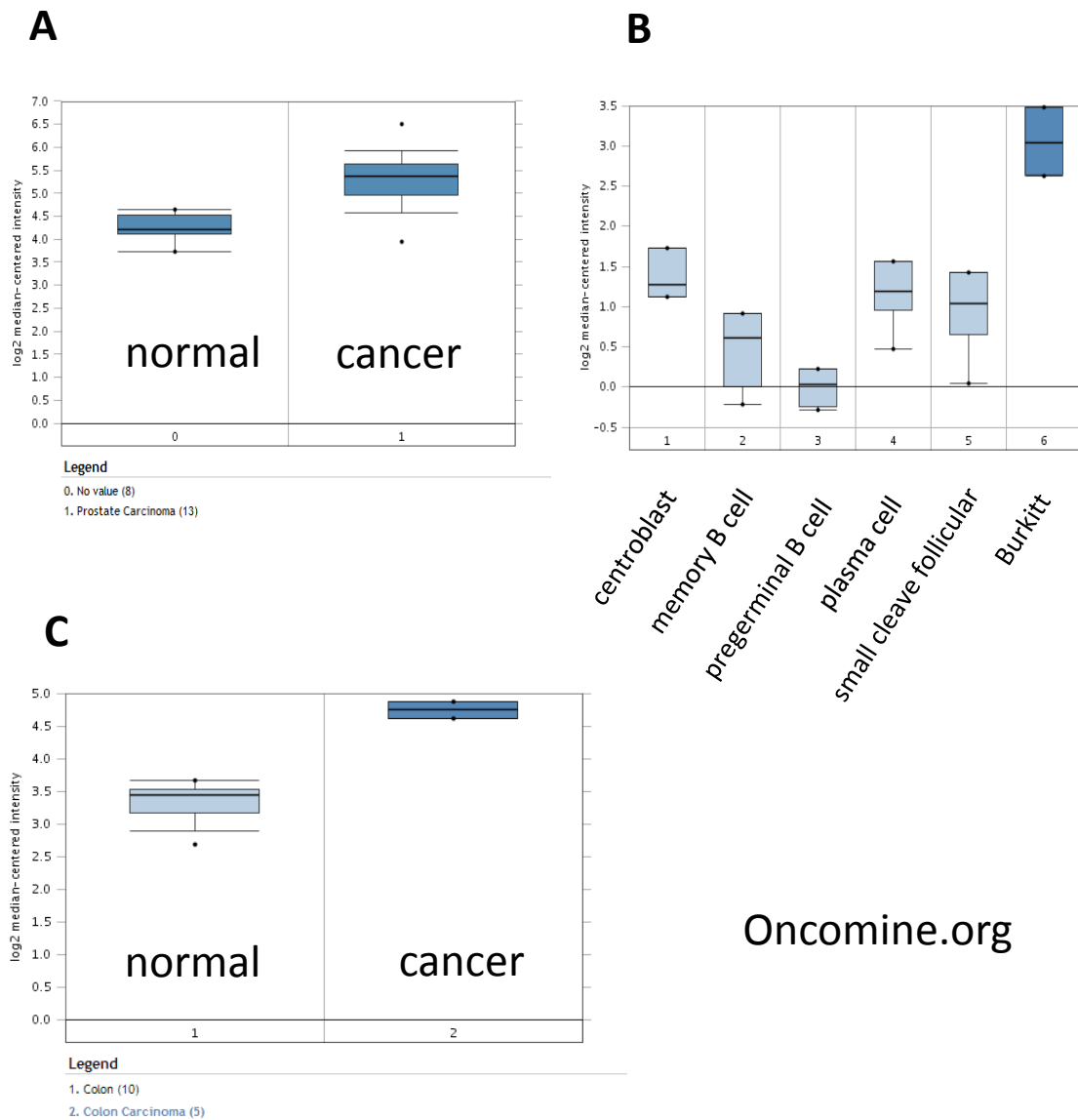
Once I established DANCR to be a direct MYC target and one of the most highly MYC-induced transcripts, I explored whether DANCR plays a role in human cancers and how its expression might be related to MYC expression. Furthermore, to study the biological functional significance of DANCR in cancer, it was important to extend my studies beyond the P493 cell system. Specifically, P493 cells are not amenable for investigation with certain methodologies that the Dang lab frequently utilizes, including RNA interference (RNAi). Thus, I turned my attention to another cancer type, prostate cancer, for most of the following functional studies. I will lay out my rationale for that decision later in this chapter.

Herein, I provide evidence that DANCR is up-regulated in most primary cancers to varying degrees, particularly in prostate cancers.

### **DANCR in human cancers**

To evaluate DANCR expression profiles in cancer tissues *ex vivo*, I first considered the oncomine database, which is an excellent tool for examining gene and transcript activities in a certain type of cancer or across multiple cancer types. Oncomine is a comprehensive expert-curated cancer database that can help determine the expression profiles of selected gene(s) across 41 cancer types, tissues, perturbations, treatments, or other clinical parameters. Therefore, I used it for a quick search of DANCR expression profiles across all available cancer types, and the results obtained are displayed in Figure 11. In brief, in most of the primary human cancers DANCR expression was consistently upregulated in

contrast to normal tissue counterparts. As shown in Figure 11, Burkitt's lymphoma, prostate cancers, and colon cancers were chosen as representative cancers, as they showed significantly higher DANCR expression in tumors.



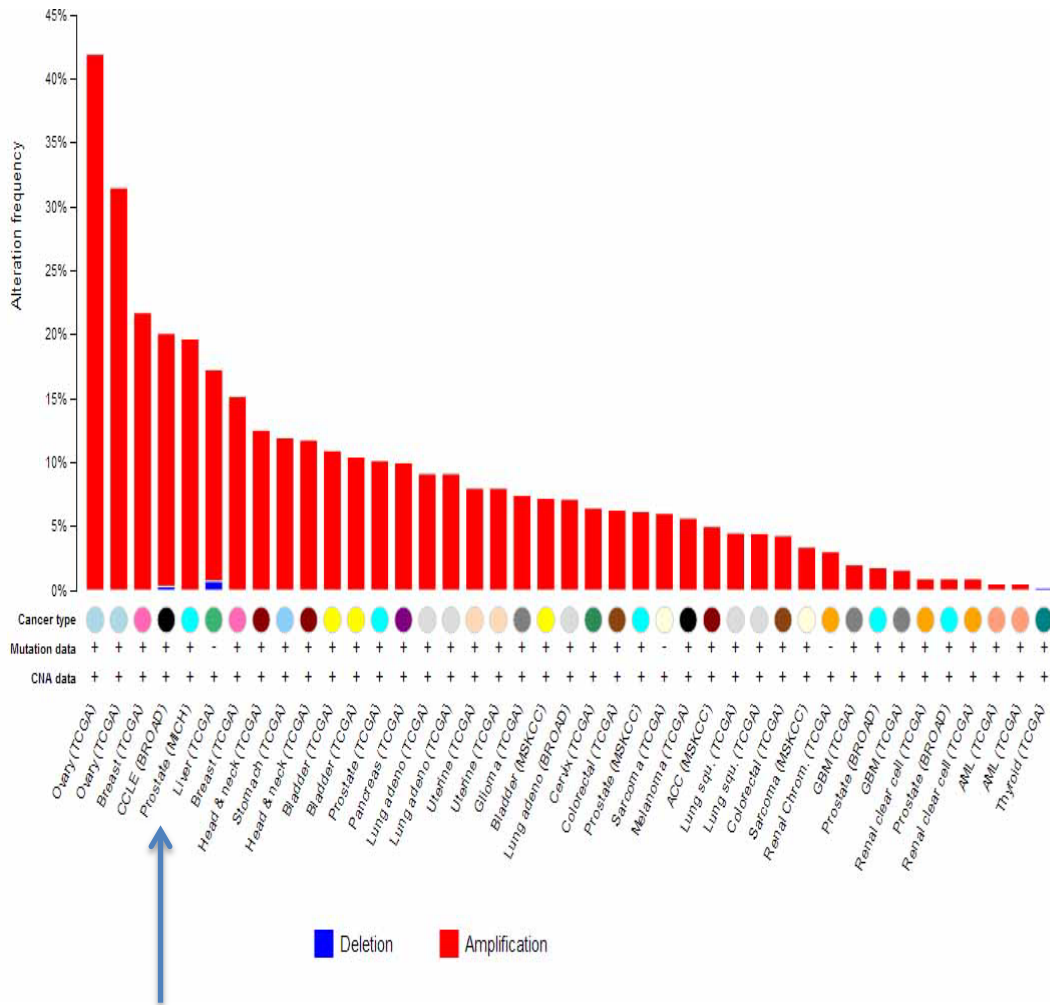
**Figure 11. DANCR expression profiles of tumor versus normal tissue (from Oncomine.org).** Median DANCR expression levels in either tumor or normal samples are represented by bar graphs. (A) Prostate cancer (n=13) vs. normal tissues (n=8); (B) Burkitt's lymphomas (n=5) vs. normal tissues (n=5 each); and (C) colon carcinoma (n=5) vs. normal tissues (n=10).

## **MYC expression in prostate cancer**

Prostate cancer is the most frequently diagnosed male cancer, and the second most common cause of death from cancer in men. In the United States, over 230,000 men will be diagnosed with prostate cancer and ~30,000 will die from it in the year 2014 (Siegel et al., 2014). Activation of the MYC oncogene frequently occurs in aggressive prostate cancers; hence I was interested in the potential role of DANCR in prostate cancer. Elevated expression of MYC in prostate cancer is generally a consequence of gene amplification and increased expression owing to high-risk single nucleotide polymorphisms near the MYC gene on 8q24 (Al Olama et al., 2009; Eeles et al., 2008; Eeles et al., 2013; Hazelett et al., 2014; Yeager et al., 2007). As shown in Figure 12, prostate cancer is one of the top MYC-driven cancers in terms of the level of alterations in MYC copy number (amplifications), as alterations occurred ~20% of the time as shown by TCGA analysis. Furthermore, MYC is activated at the earliest phases of prostate cancer (e.g., in tumor-initiating cells). MYC contributes to disease initiation and progression by stimulating an embryonic stem cell-like signature, which is partially characterized by repressing cell differentiation. Finally, overexpression of MYC alone is sufficient to immortalize primary human prostate epithelial cells *in vitro*, and the overexpression of MYC alone *in vivo* might be sufficient to transform prostatic epithelial cells into PIN cells (Koh et al., 2010) (Figure 13). Considering these findings together, I have demonstrated the important role of MYC in prostate cancer, and show that prostate cancer is a MYC-driven cancer. As prostate cancer is amenable to experimental

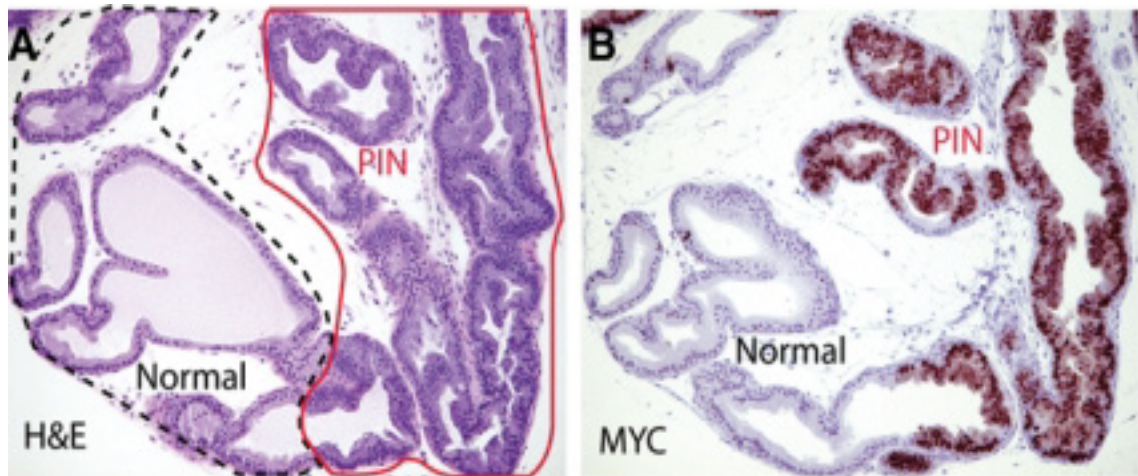
manipulation, it represents a suitable system to study MYC–DANCR interactions in cancer.





Prostate Cancer

**Figure 12. MYC copy number alterations in primary tumors.** Primary tumors are ranked by the frequency of alterations in MYC DNA copy numbers that were obtained from the cBioPortal TCGA database. Bars indicate the frequency of alterations, with amplifications shown in red and deletions shown in blue.



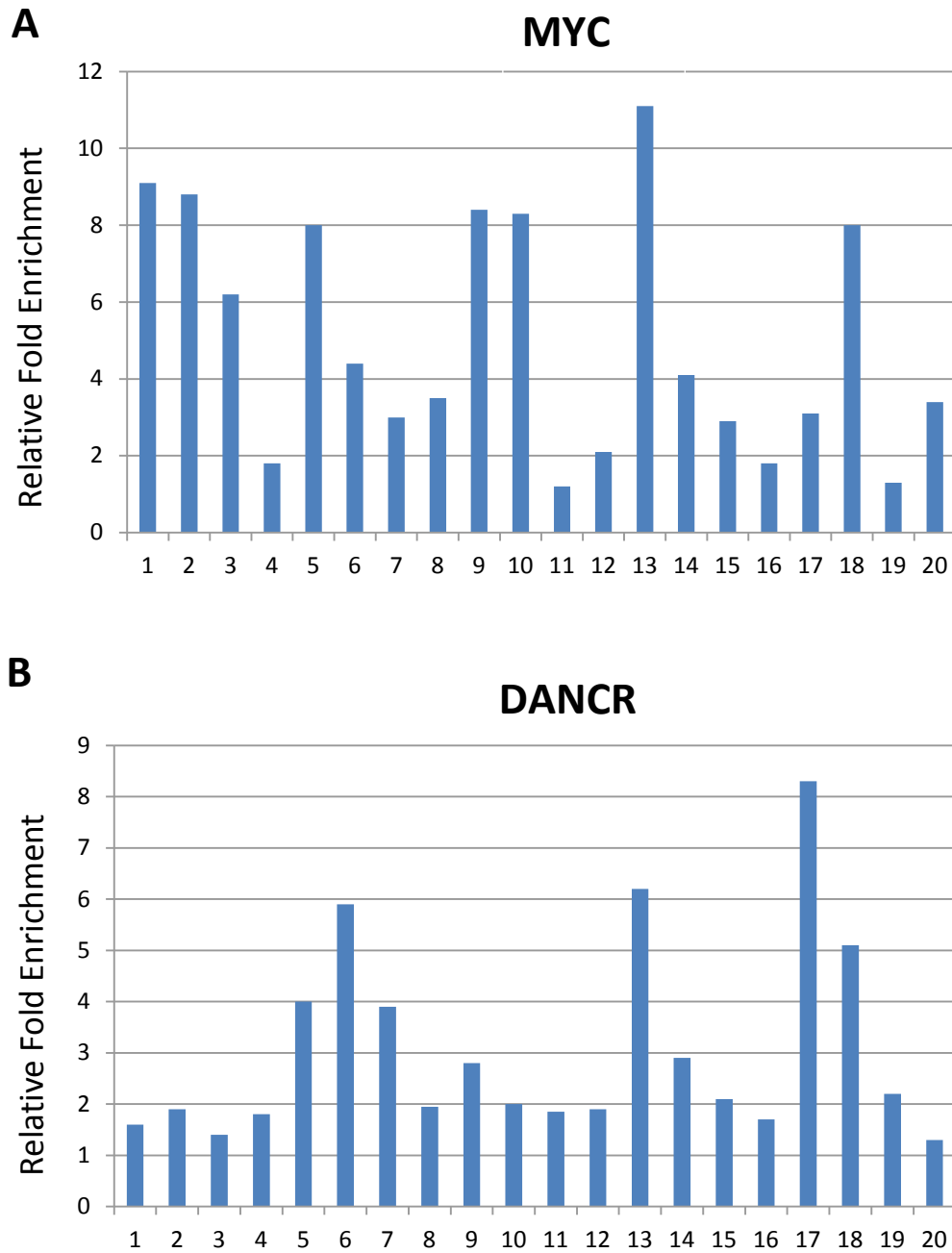
Koh et al., 2010

**Figure 13. Photomicrographs from 4-week-old Low-MYC mice. (A)** H&E showing of two populations of acini. Left, normal histology; right, high-grade PIN. **(B)** Immunohistochemical staining for MYC shows that morphologically transformed cells all stain positively.

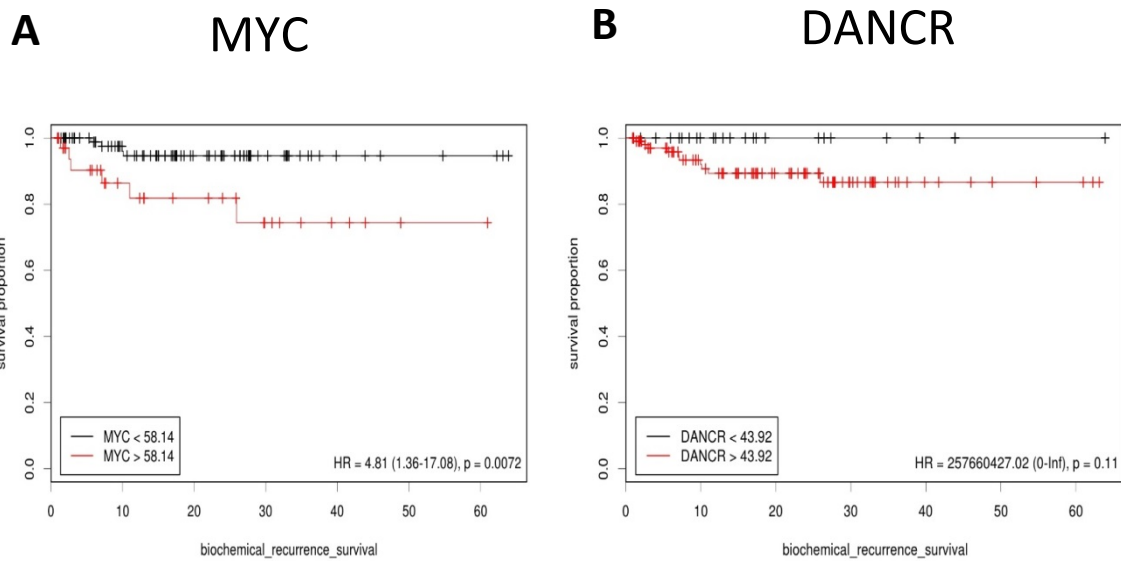
## **DANCR in prostate cancer**

By selecting human prostate cancer as our main system, which is critically driven by MYC activity, I sought to determine whether DANCR is up-regulated in prostate cancer. I compared 20 paired normal versus tumor prostate tissue samples (each pair was collected from the same patient) from Dr. Angelo De Marzo at Johns Hopkins Medical Institutions. Then, I performed quantitative real-time PCR for MYC and DANCR for validation, and I found that MYC and DANCR expression were consistently elevated in every single tumor tissue compared to its accompanying normal paired tissue sample (Figure 14). I also assessed the correlation of elevated levels of expressions between MYC and DANCR for each pair, and observed no concordant pattern within these 20 pairs; this could be attributed to other factors that affect their expressions *in vivo*.

It is also notable that based on our analysis of MYC–DANCR interactions on patient survival (Figure 15), higher MYC expression clearly indicated a poorer outcome in prostate cancer patients ( $P = 0.007$ ). It appears that high DANCR expression showed a trend toward a poorer outcome in prostate cancer (although this did not reach statistical significance,  $P = 0.11$ ). I can tentatively conclude that higher expression of DANCR is likely associated with a poorer outcome in prostate cancer patients, and that DANCR expression could potentially be used as a predictor of cancer prognoses.



**Figure 14. MYC and DANCR expression in 20 paired prostate tumor and normal tissues.** Expression levels in tumor tissues are shown as fold-changes compared to the levels in the paired normal tissue. (A) Bar graphs represent MYC mRNA transcript level (fold-change) in the tumor sample compared to the paired normal tissue. (B) Bar graphs represent DANCR mRNA levels (fold-change) in tumor samples compared to the paired normal tissue. Paired tumor and normal tissues were obtained from Dr. Angelo De Marzo.



**Figure 15. Kaplan–Meier curves for prostate cancer survival.** Poorer survival outcomes and higher MYC and DANCR expression levels correlated in prostate cancer patients (n=256). Data were obtained from TCGA. P = 0.0072 and 0.11 for MYC and DANCR, respectively. (A) MYC. (B) DANCR.

## **Materials and Methods**

### **Survival analysis**

Kaplan–Meier survival curves were generated using Cutoff Finder software (Budczies et al., 2012) with an arbitrarily selected RPKM cut-off of 58.14 or 43.92 for MYC or DANCR, respectively.

**Quantitative real-time PCR** (RNA quantitation) and **TCGA analysis** were performed as previously described in Chapter 2.

## **Chapter 5**

# **Elucidating the functional significance and targets of DANCR**

Having established that DANCR is a direct MYC target that is globally elevated in human cancers, I explored the biological functions of DANCR in cancer and how it fits into the MYC-regulated gene network.

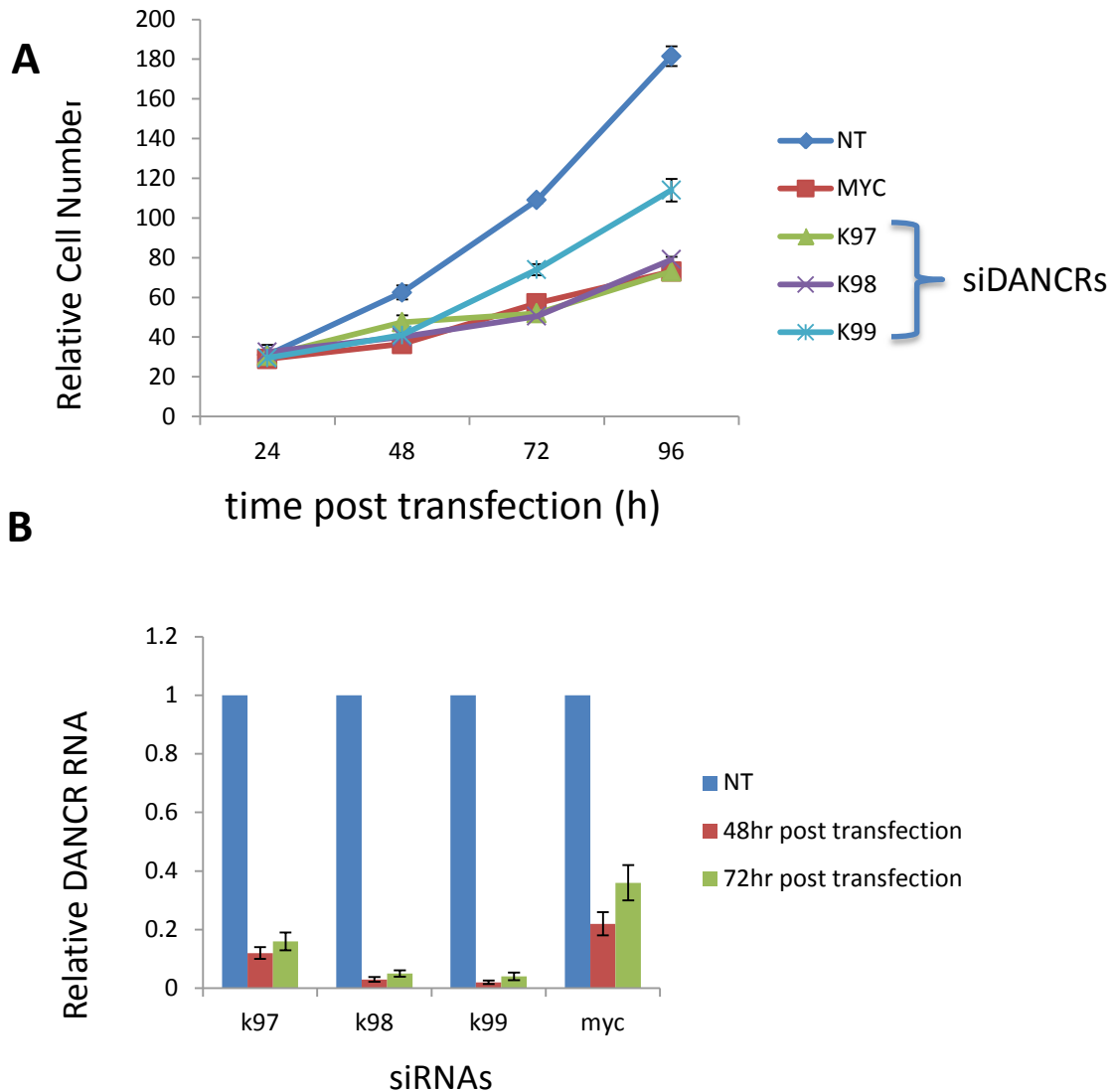
In this chapter, I use methods, such as RNAi, cell cycle analysis, microarray analysis, and morpholino oligos, to study the biological role of DANCR in human cancers. I propose a new mode of gene regulation by MYC, which occurs through lncRNAs or, in this chapter, DANCR.

### **DANCR is required for cellular proliferation and cell cycle progression in prostate cancer cells**

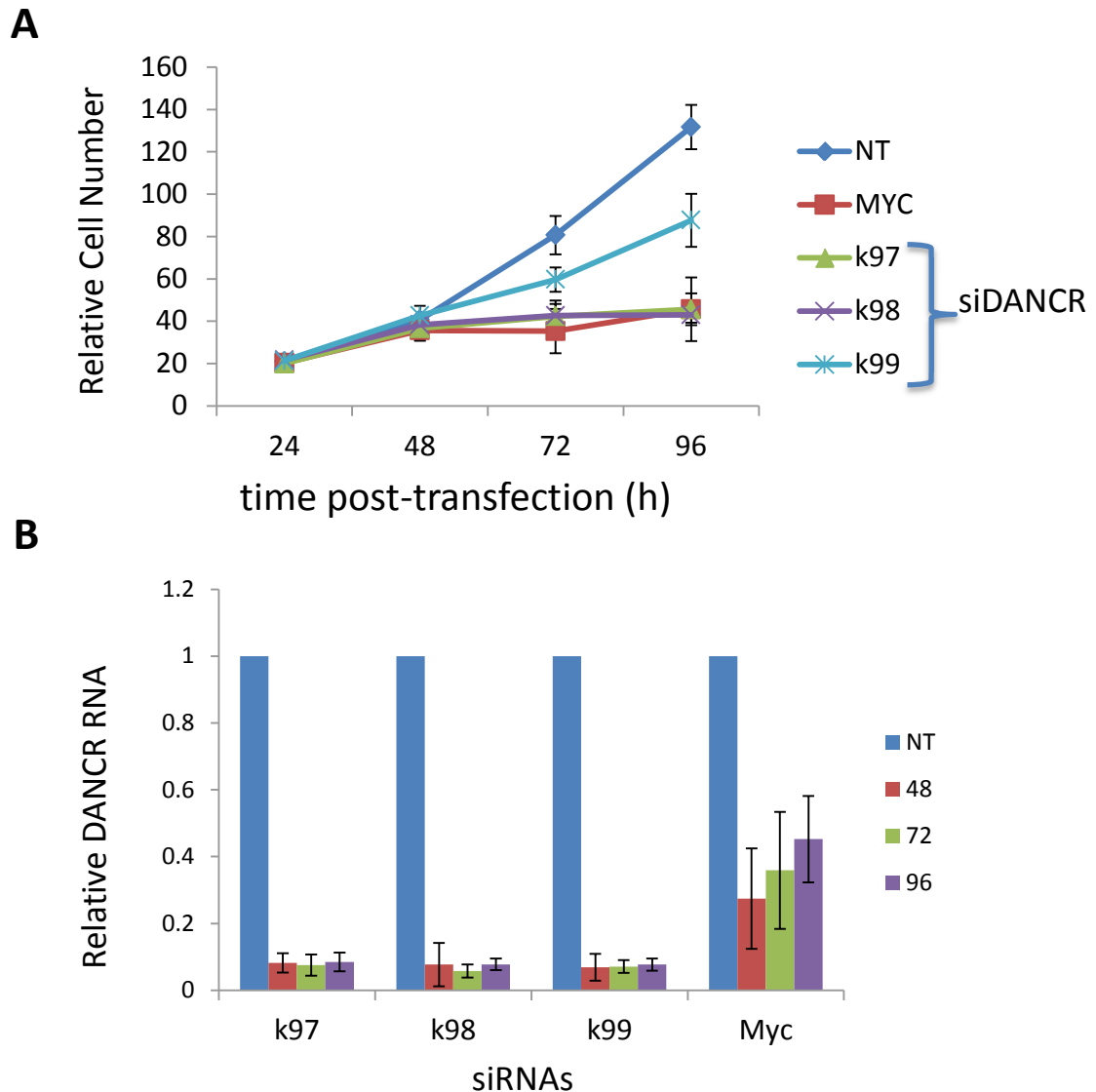
Based on the knowledge that DANCR is a MYC target and that its levels are globally increased in human cancers, especially in prostate cancer, I next sought to directly determine whether DANCR is necessary for MYC-mediated cellular proliferation. I carried out small-interfering RNA (siRNA)-mediated depletion of DANCR in two different human prostate cancer cell lines, DU145 and PC3 cells. I designed siRNA duplexes (three DANCR siRNAs, K97–K99, that targeted different regions of the transcripts and one MYC siRNA) to inhibit the expression of DANCR and MYC (as a positive control), and introduced these siRNAs into cells by Lipofectamine 2000-assisted transfection. To monitor transfection efficiency, DANCR expression was assessed by qRT-PCR. Transfection of all three siRNAs that directly targeted DANCR resulted in ~80–90% reduction of DANCR expression 48 and 72 h post-transfection, whereas siRNA



transfection against MYC resulted in 60–80% reduction of DANCR expression at similar time points (Figures 16B and 17B). Compared to control cells transfected with a non-targeting (NT) pool of ‘scrambled’ siRNAs, DANCR- and MYC-depleted prostate cancer cells exhibited markedly reduced rates of cell proliferation (Figures 16A and 17A).



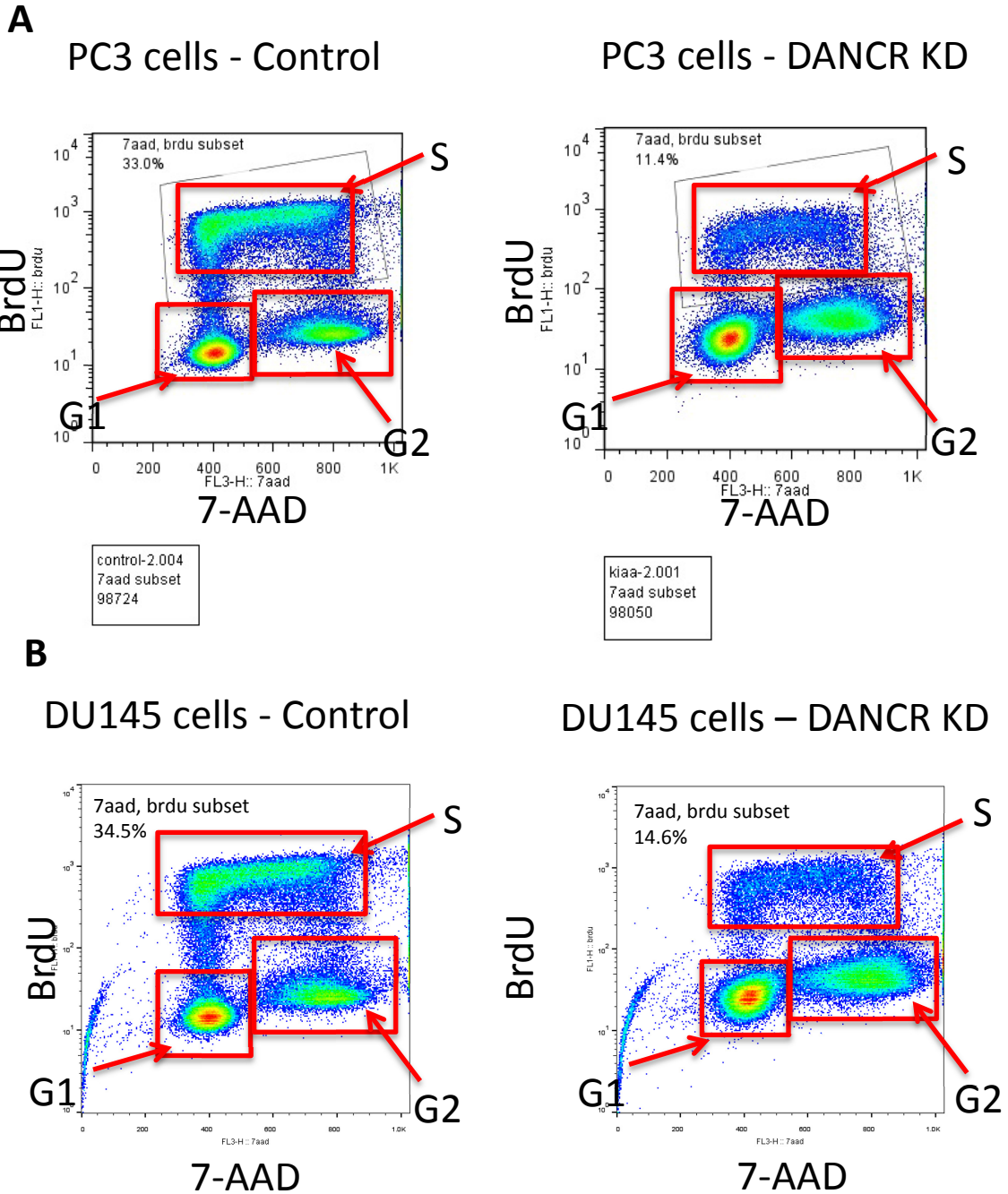
**Figure 16. The effects of DANCR and MYC siRNA-mediated depletion on the proliferation of DU145 prostate cancer cells.** (A) The rates of cellular proliferation were reduced after depletion of DANCR and MYC. Cells were transfected with a non-targeting (NT) siRNA pool as a control. Line graphs represent mean cell numbers, error bars represent standard deviation (n = 3). (B) qRT-PCR RNA quantitation of DANCR for siRNA-mediated DANCR and MYC knock-down; bar graphs represent mean DANCR mRNA expression 48 and 72 h post-transfection relative to NT siRNA-transfected cells; error bars represent standard deviation (n = 3).



**Figure 17. The effects of depleting DANCR and MYC on the proliferation of PC3 prostate cancer cells.** (A) Rates of cellular proliferation were reduced after the depletion of DANCR and MYC. Control cells were transfected with a non-targeting (NT) siRNA pool. Line graphs indicate mean cell numbers; error bars indicate standard deviation ( $n = 3$ ). (B) qRT-PCR RNA quantification of DANCR for siRNA-mediated DANCR- and MYC-knock-downs; bar graphs represent mean DANCR RNA expression 48 and 72 h post-transfection relative to NT siRNA-transfected cells; error bars represent standard deviation ( $n = 3$ ).

Because there was no apparent increase in apoptosis based on morphological observations, I next carried out cell cycle analysis to pinpoint where in the cell cycle DANCR acts to trigger cell proliferation arrest. By double-labeling PC3 and DU145 cells with a synthetic nucleoside, Bromodeoxyuridine (BrdU), and DNA-binding dye, 7-Aminoactinomycin D (7AAD), the cell cycle profile of the population with low DANCR expression was compared to the control cell population (Figure 18A). Depletion of DANCR resulted in significant G1/S phase entry blockade, as the frequency of cells in S phase was 34% of cells among DANCR-positive populations, compared to 11% among DANCR-negative populations. Similar observations were made in DU145 cells (Figure 18B).

As mentioned in Chapter 2, for negative-control purposes, I also silenced two MYC target lncRNAs, SNHG4 and LOC286467; however, successful depletion of both targets did not affect PC3 or DU145 cell proliferation (data not shown). The findings that I presented in this chapter demonstrate that DANCR expression is necessary for progression through the G1/S phase transition and for MYC-mediated cell proliferation.



**Figure 18. Cell cycle analysis of DANCR knock-down versus control siRNA in PC3 and DU145 cells.** Flow cytometry profiles showing DNA content (7-AAD staining) on the x-axis and DNA synthesis (indicated by BrdU incorporation) on the y-axis.

## **DANCR depletion alters downstream gene expressions**

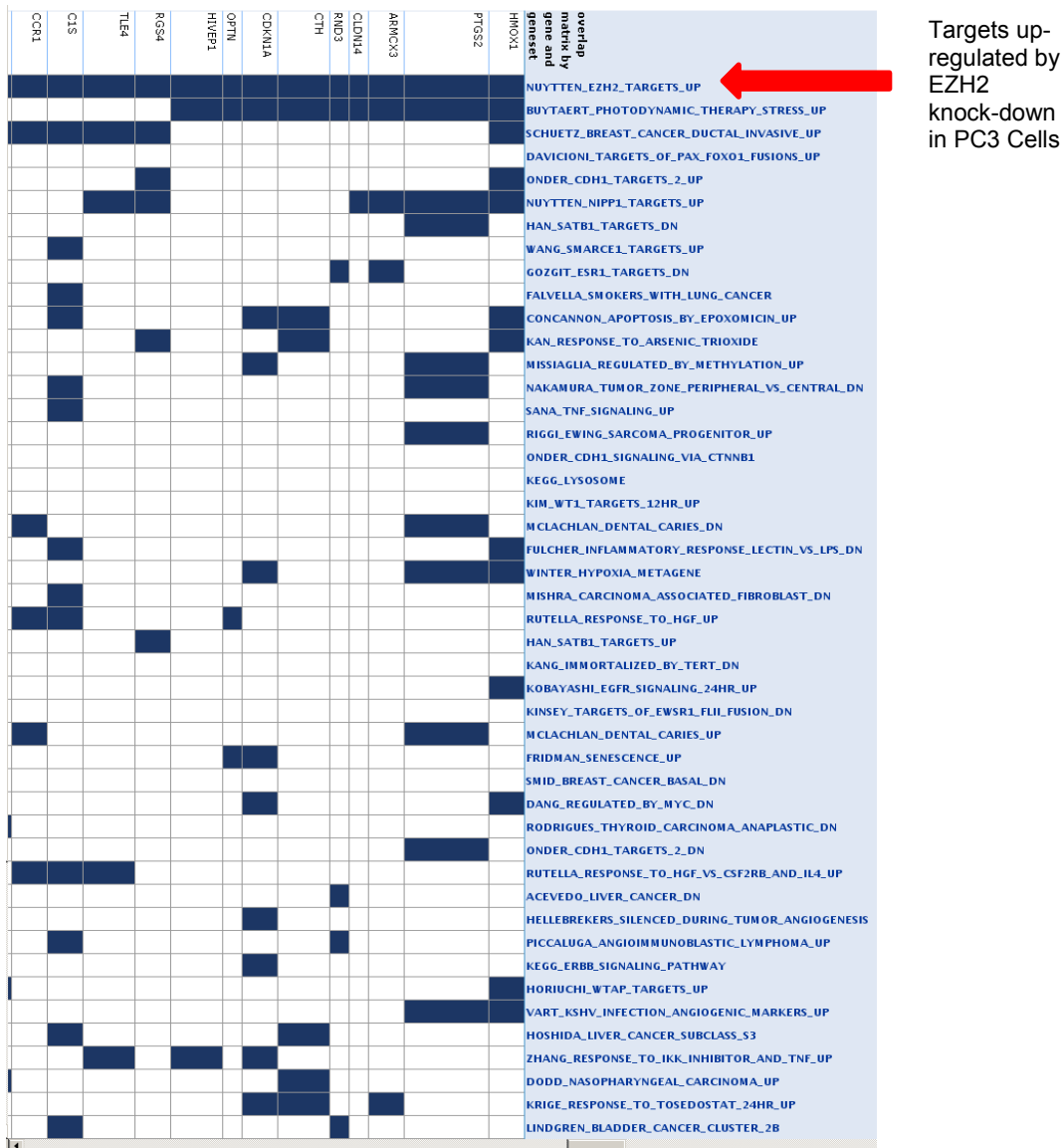
To gain further functional insights into the mechanisms whereby DANCR and MYC affect cancer cell biology, I performed genome-wide mRNA expression profile analysis on PC3 cells after siRNA-mediated depletion of DANCR and MYC. PC3 cells were transfected with DANCR and MYC siRNA, and then were harvested after 72 h. Total RNA was extracted and hybridized to Affymetrix exon 1.0 ST array. Triplicate experiments were performed. As a result, upon DANCR knock-down, my DNA microarray study of Low-DANCR cells indicated that 526 genes showed increased levels of expression with a two-fold cut-off compared to control cells. I selected up-regulated genes to test the hypothesis that DANCR plays a vital role in gene repression. I then applied gene set enrichment analysis (GSEA) on the 526 genes, and found that 112 of these 526 genes overlapped with genes that were reported to be down-regulated by EZH2 by another research group (Nuytten et al.). This list of EZH2 down-regulated genes was based on two previously performed microarray analysis in which EZH2 was knocked-down in the same cell line that I used, PC3 cells. Additionally, the large number of overlapping genes was quite remarkable, so I decided to focus on EZH2. A partial summary of the GSEA results are shown in Figure 19.

As part of the PRC2 polycomb repressive complex, EZH2 is a histone lysine methyltransferase that catalyzes the tri-methylation of histone 3 on lysine 27 (H3K27me3), which has been shown to be involved in chromatin remodeling and subsequent gene silencing (Simon and Lange, 2008; Sparmann and van Lohuizen, 2006). The remarkable number of overlapping genes from GSEA as

well as the important gene regulatory function of EZH2 carries out suggest that it might play a role in MYC–DANCR-mediated gene regulation. Hence, I decided to test the hypothesis that EZH2 is directed by DANCR to suppress specific target genes.

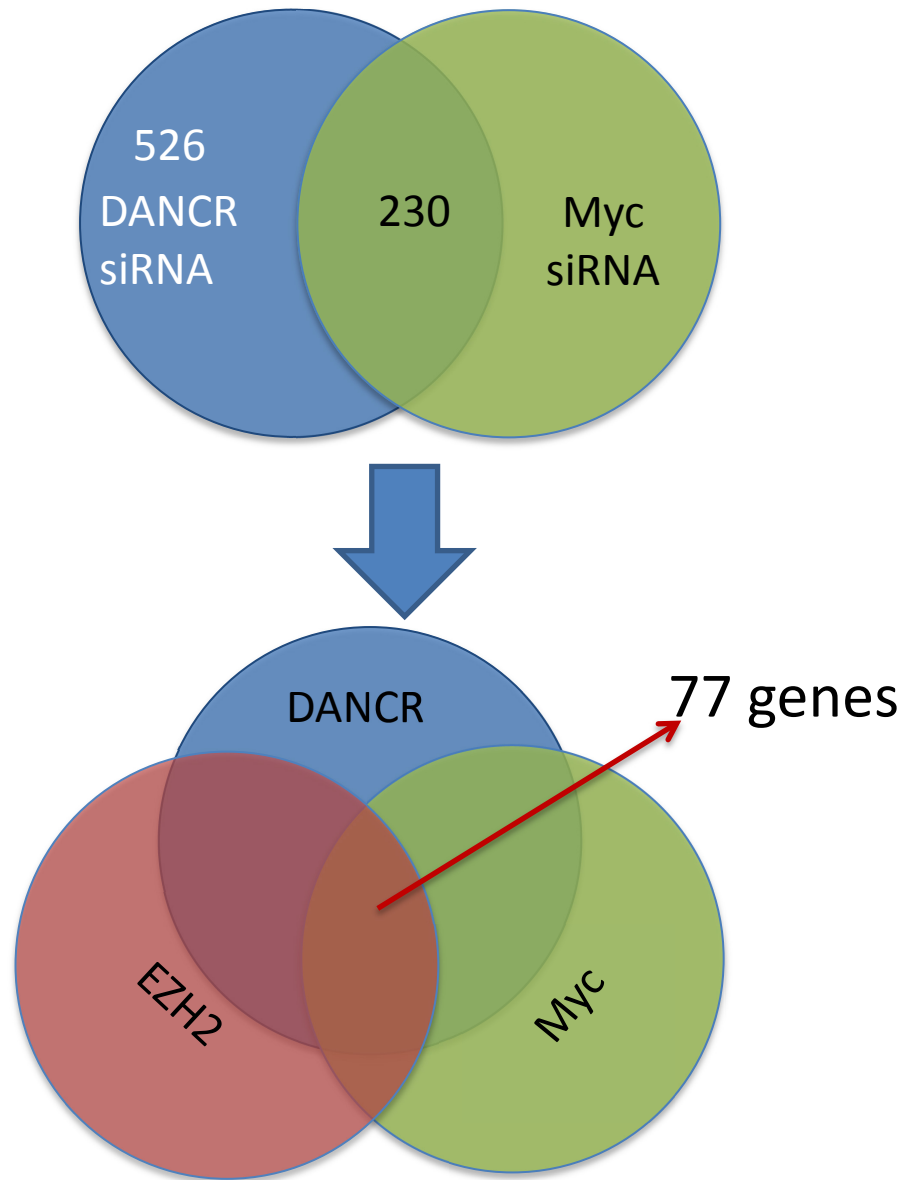
When I also considered the MYC knock-down expression data to rule out false positive targets, 230 out of 526 genes remained, including 61 genes that overlapped with the EZH2 down-regulated gene list. In collaboration with Dr. Hongkai Ji, a biostatistician from Johns Hopkins School of Public Health, I reanalyzed the raw data from these DANCR- and MYC-depletion microarrays, along with the EZH2-depletion array data from the other research group. The process for this analysis is shown in Figure 20. The final analysis yielded a list of 77 commonly up-regulated genes that appeared to be affected by all the depletion of these three regulators, DANCR, MYC, and EZH2 (Table 3).

To validate my microarray and bioinformatics analyses, I selected eight genes from top of the list based on their cell cycle regulatory or cancer related functions. I first performed quantitative real-time PCR analysis for expressional difference validation 72 h after DANCR knock-down using pre-validated primer sets. Then, I carried out EZH2 ChIP-qPCR for the preliminary validation of these genes as *bona fide* EZH2 targets. In summary, I confirmed the up-regulation of all eight genes by DANCR knock-down, which ranged from 2- to 8-fold. Additionally, all eight genes demonstrated fold-enrichments at their promoter regions for EZH2 binding from 2- to 9-fold relative to IgG. I compiled the results of these validations and show them in Table 4.



**Figure 19. Gene set enrichment analysis (GSEA) against previous expression profiling experiments available in public databases.** Top panel, names of the overlapping genes. Right side, a list of previous experiments that showed genes that overlapped with my list, which were ranked by the number of overlapping genes from large to small. Highlighted block, overlapping gene from both lists. This analysis revealed a large overlap between selected genes that were up-regulated upon DANCR knock-down and those that were increased upon EZH2 knock-down.





**Figure 20. Illustration of the process for identifying candidate genes affected by DANCR, MYC, and EZH2.** Venn diagrams represent a comparison of DANCR, MYC, and EZH2 responsive genes.

gene_symbol	gene_id	probeset_id	logFC.myc	logFC.dancr	logFC.ezh2	P.Value.myc	P.Value.dancr	P.Value.ezh2	logFC.kiaa
KRT34	3885	3756911	2.618196	21.03967	2.55E-09	3.15E-05	11.27232	1.980408	
GAS6	2621	3502829	1.908621	17.91575	1.14E-08	4.28E-05	10.14689	1.622981	
C15orf48	84419	3592401	2.33093	13.90522	1.18E-07	0.000147	8.19043	1.51493	
C4orf3	401152	2783473	1.344937	11.62354	5.99E-07	0.00033	6.713334	1.44938	
KRTAP1-5	83895	3756676	1.429181	11.41244	7.06E-07	0.000363	6.559489	1.304027	
CTNS	1497	3706700	1.246062	11.37784	7.25E-07	0.000363	6.533962	1.303564	
<b>TNFAIP3</b>	<b>7128</b>	<b>2927506</b>	<b>1.450799</b>	<b>11.06801</b>	<b>9.28E-07</b>	<b>0.000413</b>	<b>6.301441</b>	<b>1.297282</b>	
<b>SLC16A2</b>	<b>6567</b>	<b>3981959</b>	<b>1.277003</b>	<b>10.19845</b>	<b>1.92E-06</b>	<b>0.000599</b>	<b>5.608588</b>	<b>1.260419</b>	
PDCD6	10016	2798586	1.457571	9.954624	2.37E-06	0.000609	5.402954	1.246328	
<b>SERPINE1</b>	<b>5054</b>	<b>3016148</b>	<b>1.47863</b>	<b>9.944781</b>	<b>2.39E-06</b>	<b>0.000609</b>	<b>5.394543</b>	<b>1.24601</b>	
LOXL4	84171	3302693	1.098424	9.908215	2.47E-06	0.000609	5.36322	1.192332	
NIPSNAP3A	25934	3182957	1.075068	9.855232	2.59E-06	0.000615	5.317624	1.1685	
NIPSNAP3B	55335	3182957	1.075068	9.855232	2.59E-06	0.000615	5.317624	1.167316	
RBMS3	27303	2615060	1.163534	9.772477	2.79E-06	0.000645	5.245898	1.158659	
GHR	2690	2807949	1.198889	9.670203	3.06E-06	0.000696	5.156391	1.158249	
C1S	716	3403168	1.416665	9.532294	3.47E-06	0.000749	5.034163	1.121874	
RUNX2	860	2908762	1.248092	9.515149	3.52E-06	0.000749	5.018843	1.11803	
<b>BMF</b>	<b>90427</b>	<b>3619229</b>	<b>1.44584</b>	<b>9.052038</b>	<b>5.43E-06</b>	<b>0.000867</b>	<b>4.594307</b>	<b>1.09647</b>	
SPOCK1	6695	2876897	1.076418	8.952789	5.97E-06	0.000906	4.500561	1.057089	
TMEM140	55281	3025740	1.512825	8.746306	7.30E-06	0.001044	4.302291	1.055967	
PLAT	5327	3133233	1.283395	8.732089	7.40E-06	0.001051	4.288475	1.055524	
CNIH3	149111	2382467	1.079493	8.668785	7.88E-06	0.00111	4.226703	1.024771	
PDE4DIP	9659	2431886	0.895925	8.462903	9.67E-06	0.001285	4.022843	1.01509	
<b>HMOX1</b>	<b>3162</b>	<b>3944129</b>	<b>1.184129</b>	<b>8.435786</b>	<b>9.94E-06</b>	<b>0.001311</b>	<b>3.99565</b>	<b>1.012775</b>	
TK2	7084	3695107	1.086107	8.273177	1.17E-05	0.001365	3.830885	0.968705	
<b>PIP4K2A</b>	<b>5305</b>	<b>3281068</b>	<b>0.903282</b>	<b>7.868876</b>	<b>1.79E-05</b>	<b>0.001702</b>	<b>3.408241</b>	<b>0.909516</b>	
SLC35F3	148641	2385967	0.877478	7.695785	2.15E-05	0.001819	3.221447	0.898104	
ADM	133	3320123	0.833993	7.540048	2.55E-05	0.001969	3.050284	0.897501	
PGCP	10404	3108226	1.052414	7.432531	2.88E-05	0.00213	2.930372	0.884679	
BDNF	627	3367231	1.165123	7.408399	2.95E-05	0.002163	2.90326	0.882985	
AQP11	282679	3341362	1.141964	7.240579	3.57E-05	0.002369	2.712682	0.87804	
RIT1	6016	2437736	0.874873	7.071204	4.33E-05	0.002695	2.516675	0.864869	
PI3	5266	3886765	1.233718	6.93048	5.10E-05	0.003003	2.350971	0.849657	
CPA4	51200	3023883	1.012539	6.893767	5.32E-05	0.00303	2.30731	0.828596	
KIAA0513	9764	3672059	0.830606	6.883364	5.39E-05	0.00304	2.294905	0.828596	
<b>CTSB</b>	<b>1508</b>	<b>3124537</b>	<b>1.03967</b>	<b>6.679065</b>	<b>6.86E-05</b>	<b>0.003434</b>	<b>2.04834</b>	<b>0.821033</b>	
SLC26A11	284129	3737242	0.758048	6.676366	6.89E-05	0.003434	2.045044	0.821033	
ANKRD46	157567	3146661	0.825258	6.559016	7.93E-05	0.003792	1.900798	0.818223	
DENND3	22898	3118651	0.962493	6.541808	8.10E-05	0.003843	1.879486	0.783078	
CITED4	163732	2408437	1.650338	6.517811	8.34E-05	0.003917	1.849698	0.760282	
SHC3	53358	3213847	0.846711	6.357177	0.000102	0.004417	1.648222	0.751078	
CCDC92	80212	3476330	0.922573	6.331186	0.000105	0.004489	1.615281	0.740932	
SDC2	6383	3108146	0.737372	6.289781	0.00011	0.004667	1.562607	0.726999	
<b>CDKN1A</b>	<b>1026</b>	<b>2905169</b>	<b>0.780852</b>	<b>6.122501</b>	<b>0.000136</b>	<b>0.005364</b>	<b>1.347306</b>	<b>0.714581</b>	
BIRC3	330	3346548	0.652847	5.804543	0.000206	0.006916	0.926905	0.702272	
PTGS2	5743	2448382	0.655707	5.760257	0.000218	0.007151	0.867174	0.702272	

**Table 3. Genes that are commonly up-regulated by MYC, DANCR, and EZH2 knock-down.** Only the top part of the 77 genes list is shown above. Yellow highlighted genes were confirmed by additional validation steps.

	<b>Expression fold change (DANCR kd / nt control)</b>	<b>Fold enrichment (EZH2 ChIP / IgG ChIP)</b>
TNFAIP3	~4-fold	~3-fold
SLC16A2	~3-fold	~2-fold
SERPINE1	~2-fold	~3-fold
BMF	~8-fold	~9-fold
HMOX1	~4-fold	~3-fold
PIP4K2A	~3-fold	~4-fold
CTSB	~3-fold	~2-fold
<b>CDKN1A</b>	<b>~5-fold</b>	<b>~5-fold</b>

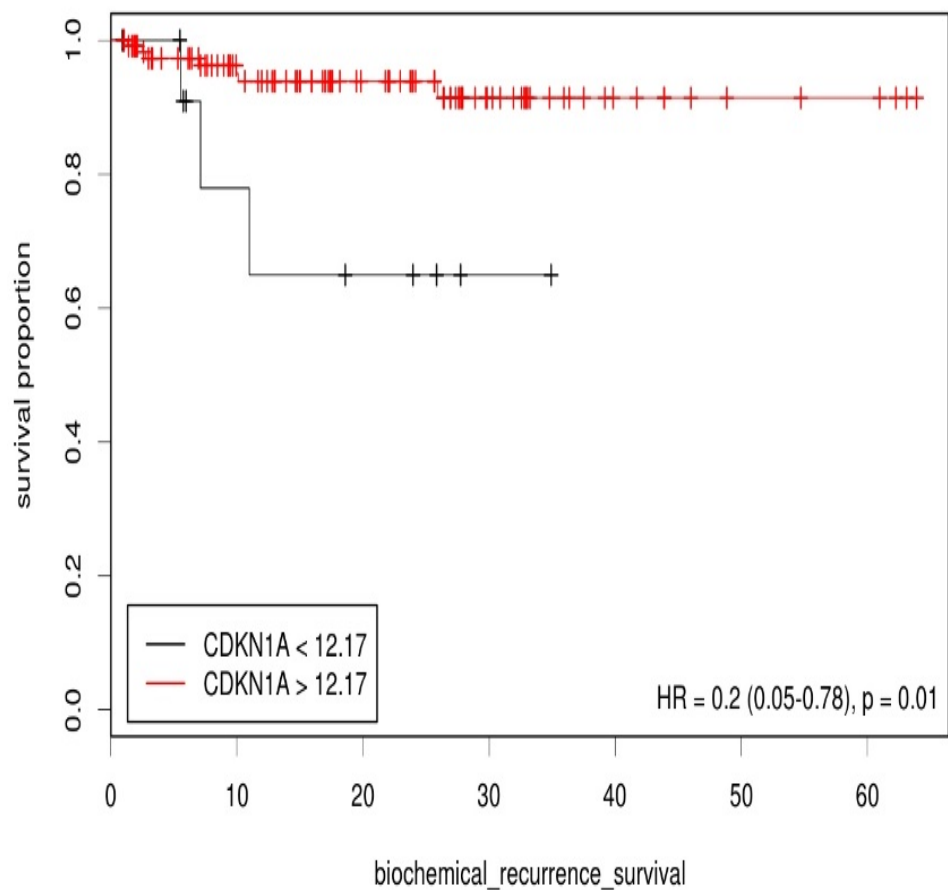
**Table 4. Validation of targets regulated by both DANCR and EZH2 in PC3 cells.** Left column, fold-induction by DANCR depletion for each gene; total mRNA was isolated 72 h after transfection of DANCR siRNA. Right column, fold-enrichment of EZH2 binding compared to IgG for each gene promoter locus.

## **DANCR regulates p21**

Among the eight validated genes, CDKN1A (p21) immediately caught my attention because it is an important G1/S cell cycle regulator. Its regulation by DANCR is consistent with the G1/S phase arrest phenotype that I observed after DANCR depletion.

CDKN1A encodes p21, also known as cyclin-dependent kinase inhibitor 1 or CDK-interacting protein 1, and p21 binds to and inhibits the activity of the CDK2, CDK1, and CDK4/6 complexes. By doing so, it functions as a regulator of cell cycle progression at G1. Additionally, p21 was previously found to mediate p53-dependent cell cycle phase arrest in G1 in response to a variety of stress stimuli, to promote cellular differentiation, and to prevent cell proliferation (el-Deiry et al., 1993; Gartel and Radhakrishnan, 2005; Harper et al., 1993; Xiong et al., 1993). My own TCGA analysis indicated that a poor prognosis is significantly associated with p21 expression in prostate cancer ( $P = 0.01$ ), suggesting its important biological role in human cancers (Figure 21).

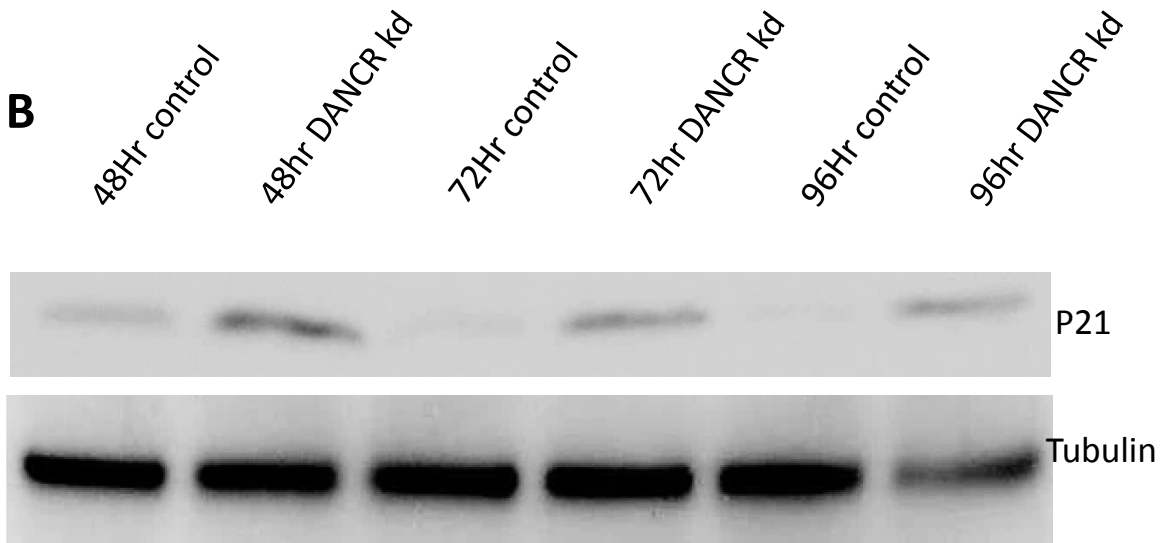
Previous qRT-PCR analysis demonstrated a five-fold up-regulation of p21 mRNA levels in DANCR siRNA-transfected cells 72 h after treatment compared to NT siRNA-transfected control cells. Immunoblot analysis further confirmed the increased levels of p21 protein (Figure 22). These findings confirm that DANCR can suppress p21 in prostate cancer.



**Figure 21. Poor survival outcome correlates with p21 expression in prostate cancer patients.** Data obtained from TCGA (n = 256). A p21 expression cut-off was set at 12.17 RPKM with (P = 0.01). RPKM, reads per kilobase per million.

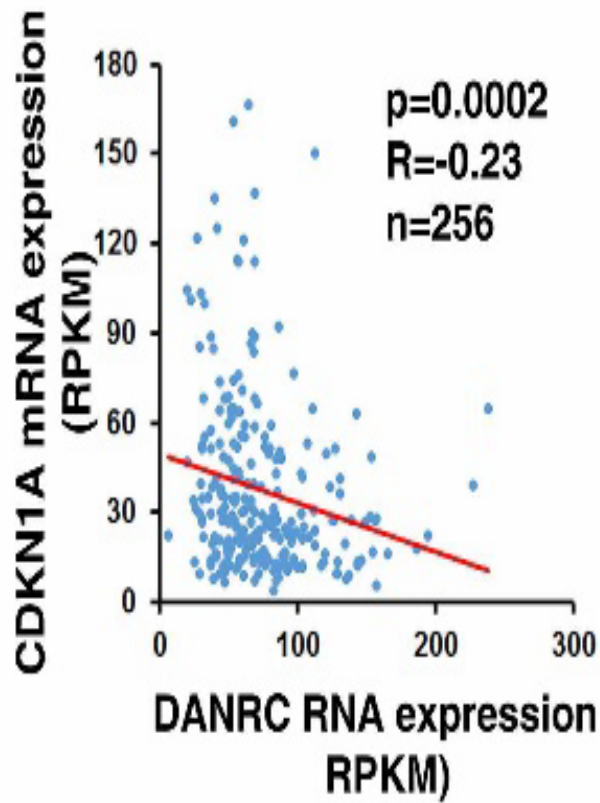
**A**

	Expression fold change (DANCR kd / nt control)
<b>CDKN1A RNA</b>	<b>~ 5 fold</b>

**B**

**Figure 22. Depletion of DANCR by RNAi activates p21.** (A) p21 mRNA transcript fold-changes in expression. (B) Immunoblot of p21 and tubulin (a loading control) protein levels after DANCR depletion. Cell lysates were analyzed by western blotting with antibodies specific for p21 or tubulin.

Lastly, I used the co-expression analysis method with TCGA RNA-seq data to examine the relationship between DANCR and p21 in human cancers. I found that in both prostate and breast cancers, DANCR and p21 were significantly negatively co-expressed with  $R^2$ -values of  $-0.23$  and  $-0.137$ , respectively (Figure 23 and 24).



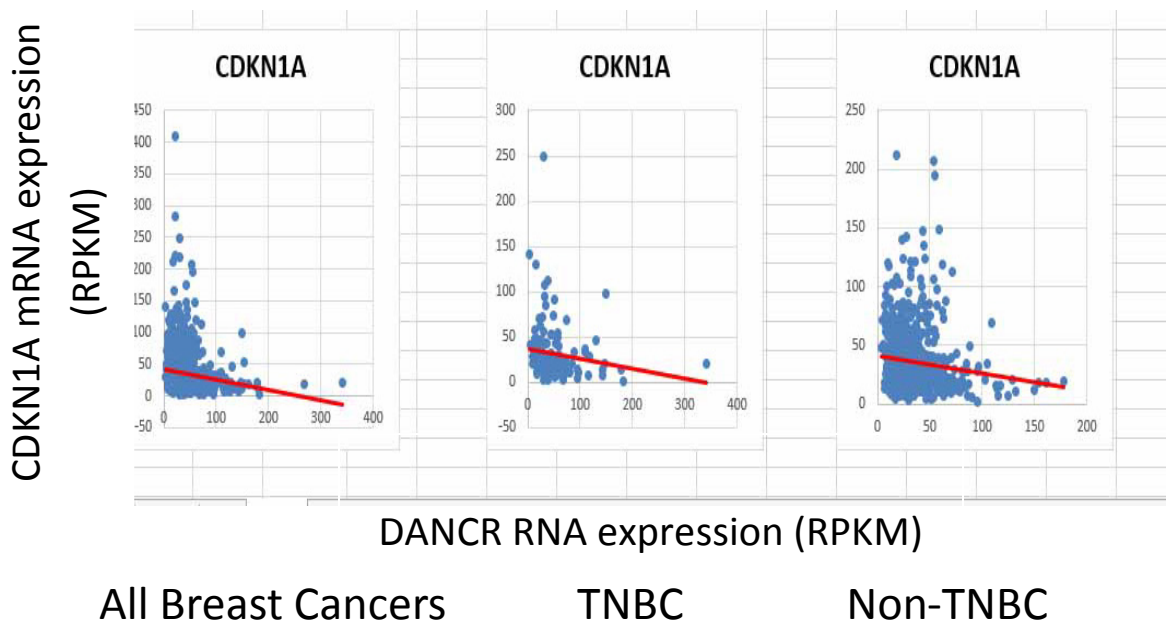
**Figure 23. Negative correlation between DANCR and CDKN1A (p21) expression in prostate cancer.** Levels of mRNA transcripts were compared ( $n = 256$ ;  $R^2 = -0.23$ ;  $P = 0.0002$ ). RPKM, reads per kilobase per million



$R^2 = -0.137$   
 $P < 0.000011$

$R^2 = -0.147$   
 $P < 0.09$

$R^2 = -0.125$   
 $P < 0.0006$



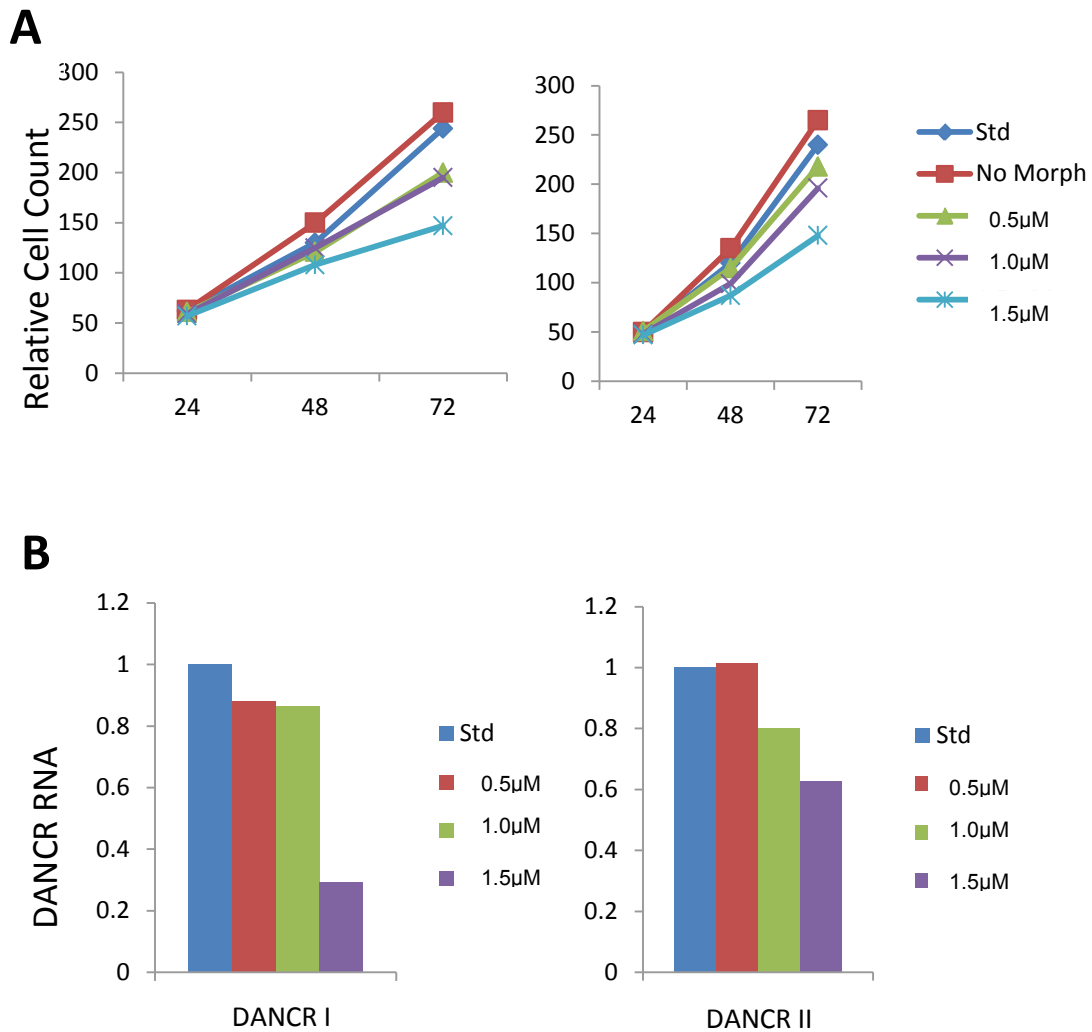
**Figure 24. Negative correlations between DANCR and CDKN1A (p21) expression in breast cancers.** TNBC, Triple-negative breast cancer (n = 1096); n = 136 for TNBC and n = 891 for non-TNBC. RPKM, reads per kilobase per million.

## **DANCR suppresses p21 in other cell types**

As mentioned in Chapter 1, P493 cells are not very amenable to siRNA treatment. However, morpholino oligo technology offers an alternative DANCR knock-down strategy for P493 cells.

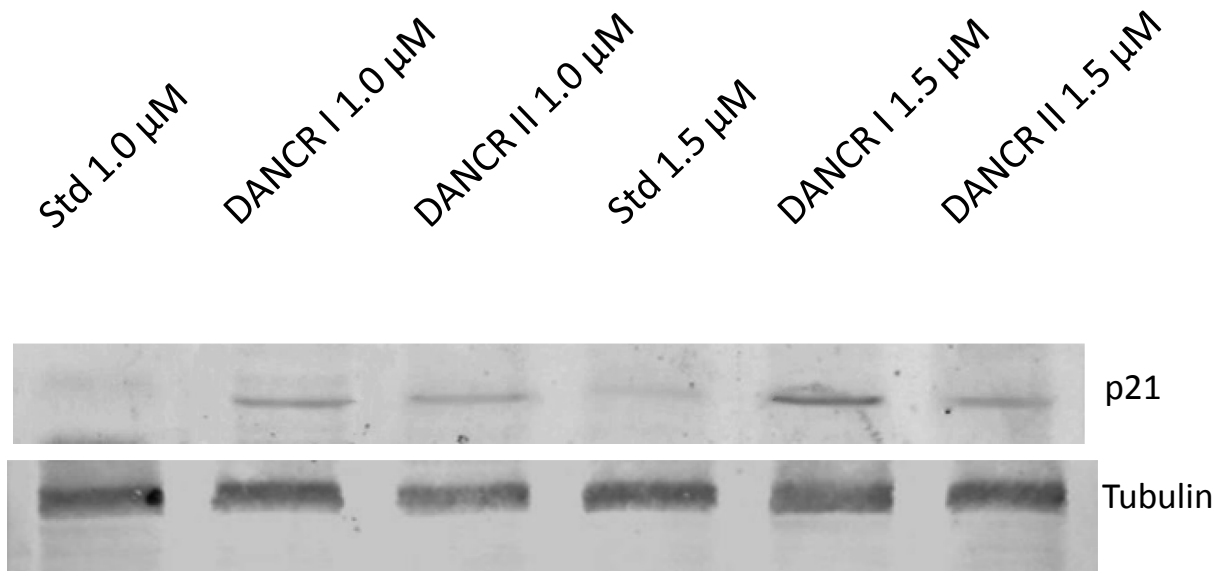
Morpholinos are anti-sense oligos that can block sites on RNA and obstruct cellular processes. A morpholino oligo specifically binds to its selected target site and blocks access by cell components for that target site. A morpholino can be used to block translation, splicing, miRNAs, miRNA targets, or ribozyme activity (Heasman, 2002; Konig et al., 2007; Matter and Konig, 2005).

The DANCR transcript has two introns, I designed two vivo-morpholinos to target the two intron–exon junctions and named them DANCR 1 and DANCR 2 for intron 1 and intron 2 targeting, respectively. As shown in Figure 25A, I tested three different concentrations (0.5, 1.0, and 1.5  $\mu\text{M}$ ) of morpholinos treatments for 48 h, and I found a predictable dose-response, as the severity of the phenotype triggered corresponded to the amount of morpholino that I put in. In summary, I found that all three concentrations of morpholinos caused reductions in cell proliferation. When the morpholino concentration was increased, a greater reduction in cellular proliferation was observed; the 1.5  $\mu\text{M}$  concentration of morpholino yielded the most severe growth reduction and the greatest suppression of DANCR RNA. Additionally, DANCR 1 appeared to more effectively block DANCR expression than did DANCR 2 (Figure 25B).



**Figure 25. The effects of two vivo-morpholinos that depleted DANCR on the proliferation of P493 cells.** (A) Rates of cellular proliferation were reduced after DANCR depletion. A total of three different concentrations of morpholinos were tested (0.5, 1.0, and 1.5  $\mu\text{M}$ ) with standard control morpholino (Std) treatment and untreated samples as controls for 48 h. (B) Bar graphs indicate the DANCR RNA levels after treatment with morpholino at different concentrations. Std, standard control morpholino.

I next examined p21 protein expression in cell lysates after 48 h of morpholino incubation by western blotting analysis, which also revealed increased levels of p21 protein with either DANCR morpholino treatments (Figure 26). These data demonstrate that DANCR is required for normal cell proliferation and can also suppress p21 in P493 cells.



**Figure 26. Levels of p21 protein increase after 48 h of vivo-morpholino treatment in P493 cells.** Doses of 1.0 or 1.5  $\mu\text{M}$  dosage for both morpholinos were used. Std, standard control morpholino. Tubulin levels were assessed as loading controls.

## **Does SnoRA26 confer effects on cell growth?**

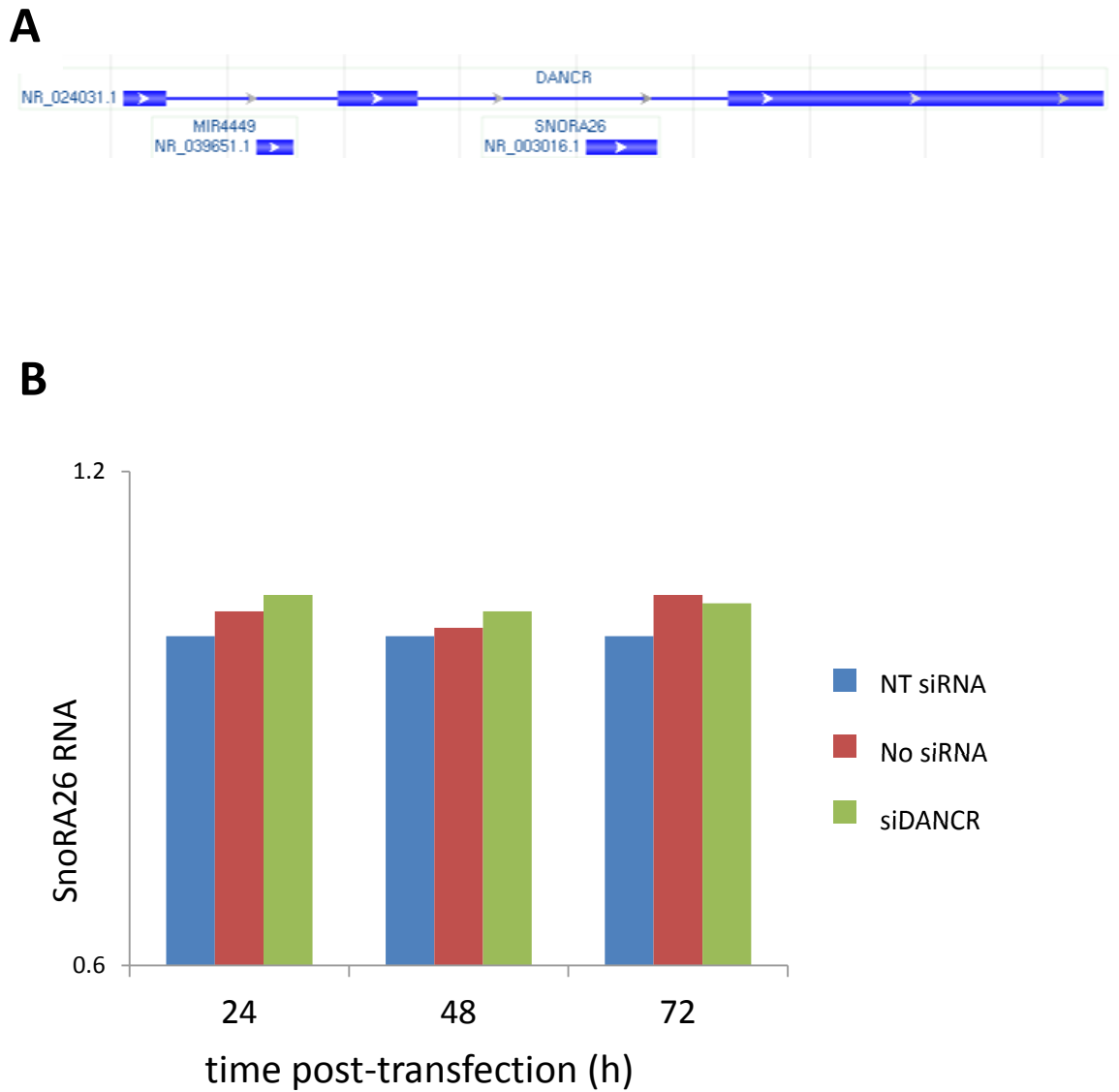
By examining the genomic sequence of DANCR, I noticed a co-transcribed small nucleolar RNA (SnoRNA), SnoRA26, in the second intron of DANCR (Figure 27A). To investigate the potential effects that SnoRA26 might have on the growth phenotypes that I observed in the loss of function studies of DANCR, I carried out the following experiments.

First, I collected total RNA samples from DANCR siRNA-mediated transfection experiments at 24, 48, and 72 h post-transfection in PC3 cells and then performed qRT-PCR with pre-validated primer sets for DANCR and SnoRA26. Successful knock-down of DANCR was achieved; however, I observed no apparent change in SnoRA26 expression with the depletion of DANCR (Figure 27B), suggesting that the reduction in proliferation caused by reduced DANCR RNA levels was likely not associated with SnoRA26.

Second, I designed a vivo-morpholino oligo to specifically target SnoRA26, followed by 72 h incubations with different concentrations of morpholinos in both PC3 and P493 cells. The efficacy of SnoRA26 blocking was assessed by qRT-PCR for samples collected at 48 and 72 h time points post-treatment. Successful blocking was observed at both time points, especially at 48 h, with ~80% depletion of SnoRA26 RNA observed at almost all concentrations tested in both cell lines (Figure 28B and 29B). However, I observed no effects on the proliferation of PC3 cells and a lack of growth effects in P493 cells (Figures 28A and 29A). The only reduction in proliferation that I observed was in P493 cells

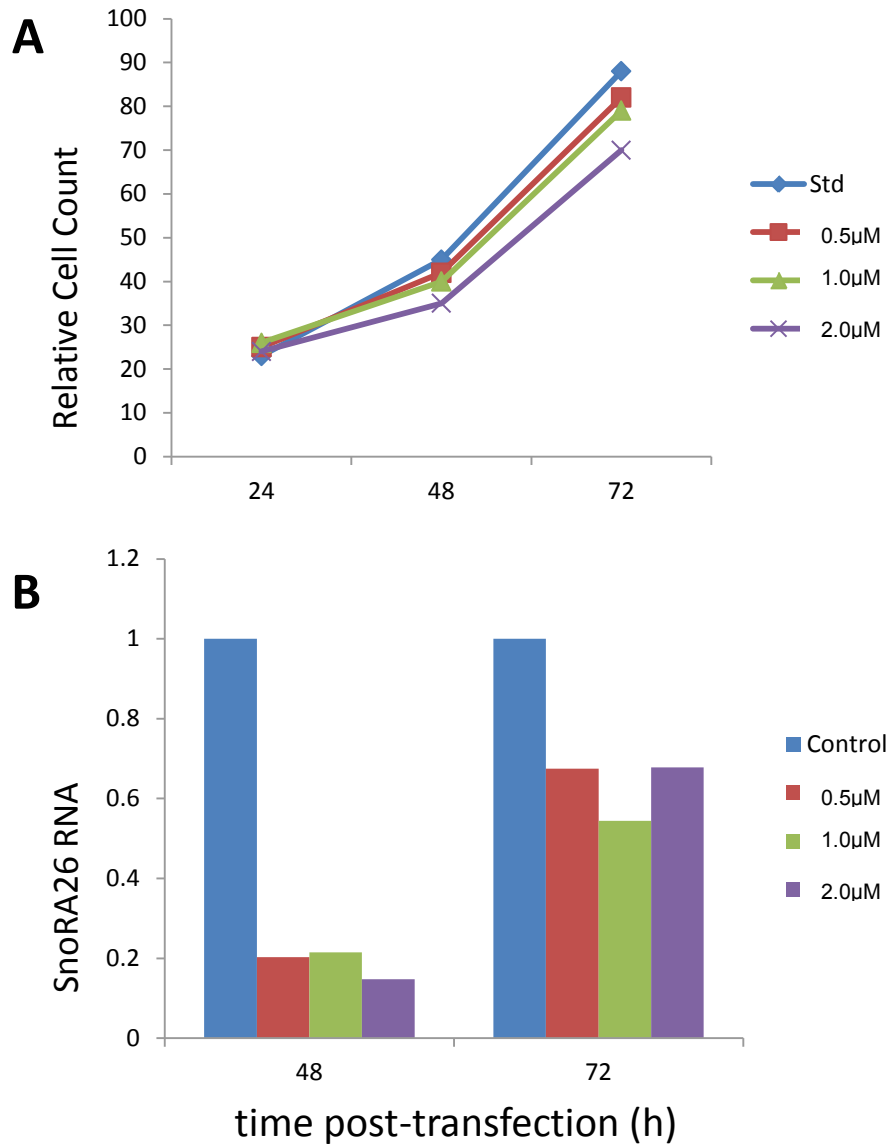
incubated with a higher dose of morpholino, which could be explained by the toxicity of the morpholinos. Note that 0.5 and 1  $\mu$ M concentrations of morpholino-treated P493 cells caused similarly high levels of SnoRA26 blockade (Figure 29B), so the growth difference between the two conditions was unlikely to be a consequence of SnoRA26 blockade.

Based on the experiments described above, I could rule out the involvement of SnoRA26 in any of the previously observed growth phenotypes, indicating that they were solely caused by changes in DANCR expression.

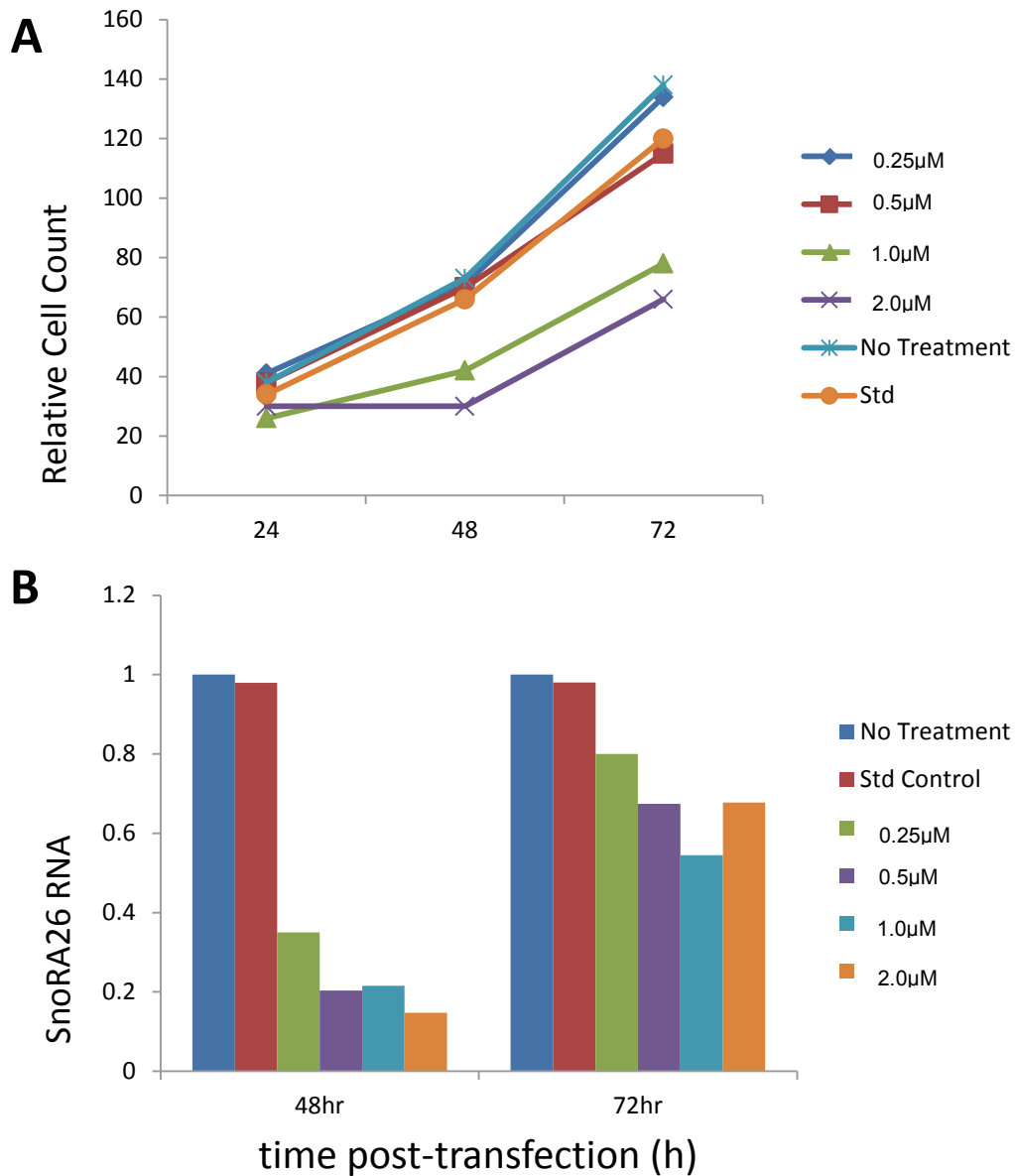


**Figure 27. SnoRA26 RNA not affected by DANCER depletion.** (A) A diagram of DANCER transcript features. (B) A qRT-PCR analysis of RNA expression levels of SnoRA26 after DANCER siRNA or two control siRNA treatments. NT siRNA, non-targeting siRNA.





**Figure 28. No effects of blocking SnoRa26 on the proliferation of PC3 cells.** (A) The rates of cell proliferation were not changed after blocking of SnoRA226; three concentrations of morpholinos were used. (B) qRT-PCR analysis of SnoRA26 RNA levels post-incubation with siRNAs. Std, standard control morpholino.

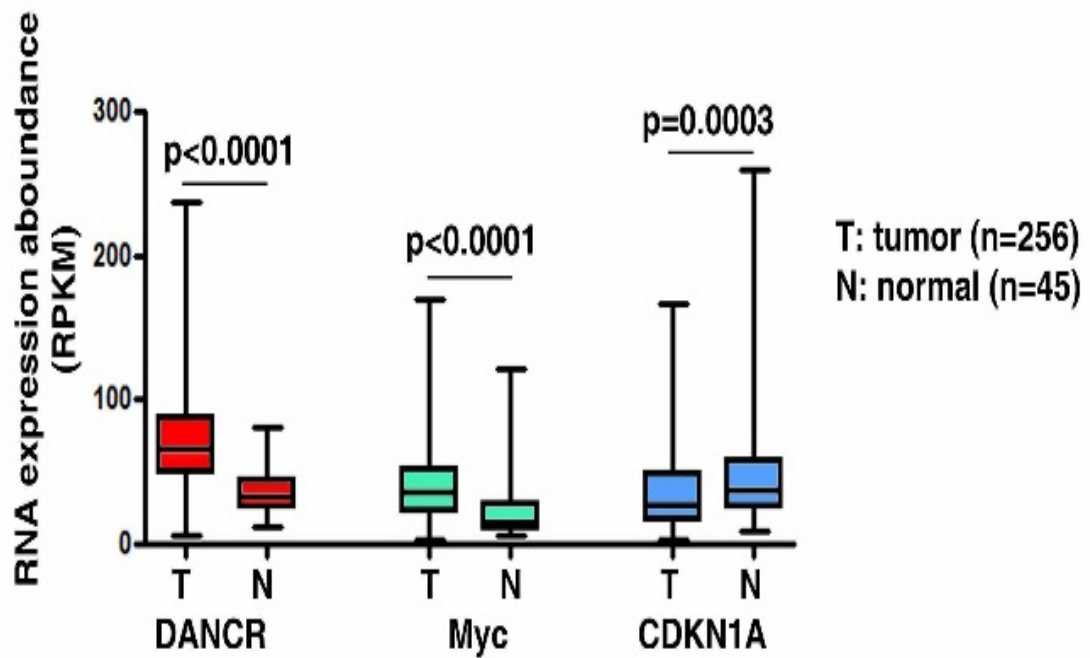


**Figure 29. Lack of effects of blocking SnoRa26 on the proliferation of P493 cells.** (A) The rates of cellular proliferation were not changed after blockade of SnoRA226 with lower concentrations; four concentrations of morpholinos were used. (B) A qRT-PCR analysis of SnoRA26 RNA levels post-morpholino incubation. Std, standard control morpholino.

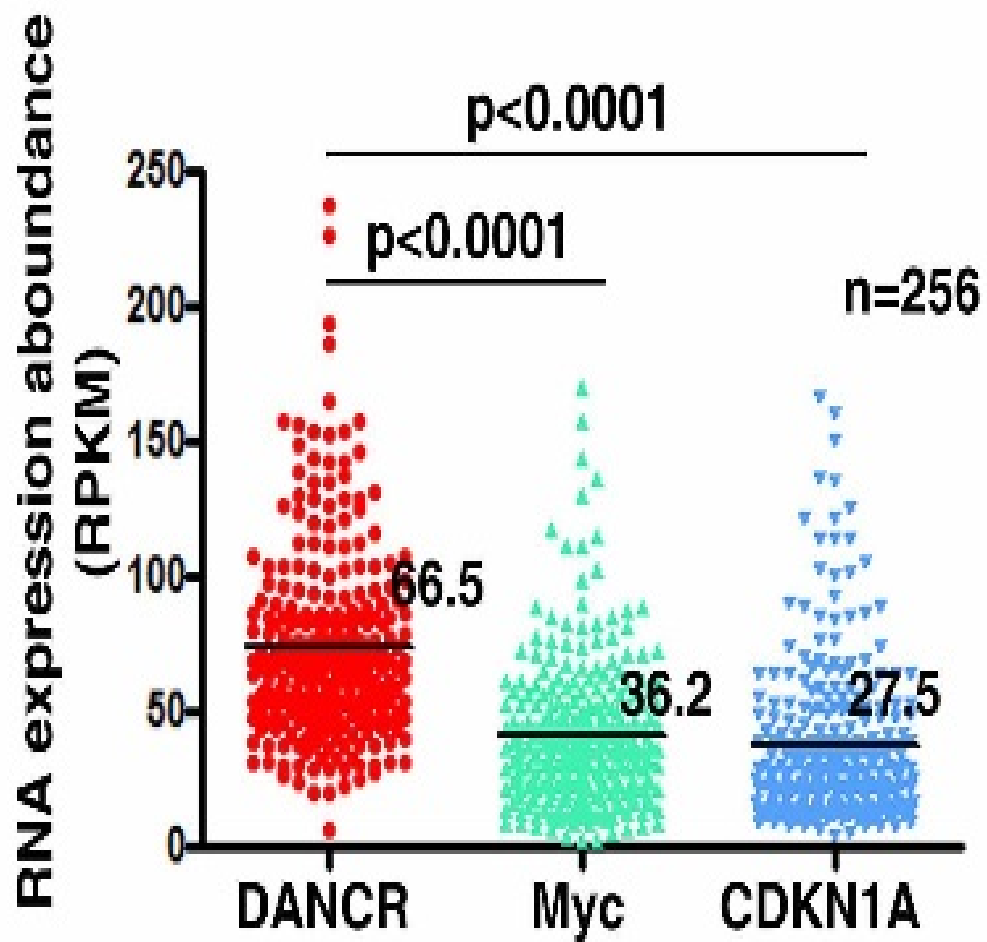
### **The MYC, DANCR, and p21 pathway**

Thus far, I have established that MYC targets DANCR directly in Chapter 3 and that DANCR suppresses p21 earlier in this chapter. Considering these two findings together, I can propose a new pathway, MYC → DANCR ─| p21, adding to diverse modes of gene regulation by MYC.

To validate my model, I re-analyzed the TCGA RNA-seq data, and found that the RNA expression levels of MYC and DANCR were significantly up-regulated in specimens from patients with prostate cancer compared to normal prostate specimens. By contrast, p21 mRNA levels were significantly down-regulated in these paired sample comparisons (Figure 30). These findings are consistent with this newly described pathway. Additionally, it is worth noting that the RNA abundance of DANCR was significantly higher than the mRNA abundance of p21 in specimens from patients with prostate cancer (Figure 31). This finding indirectly supports my hypothesis about the mechanism whereby DANCR can suppress p21, which is described in the next chapter. Notably, the expression levels of most lncRNAs were very low (~100-fold lower than that of DANCR), so the amount of DANCR that was expressed strongly suggests its functional significance.

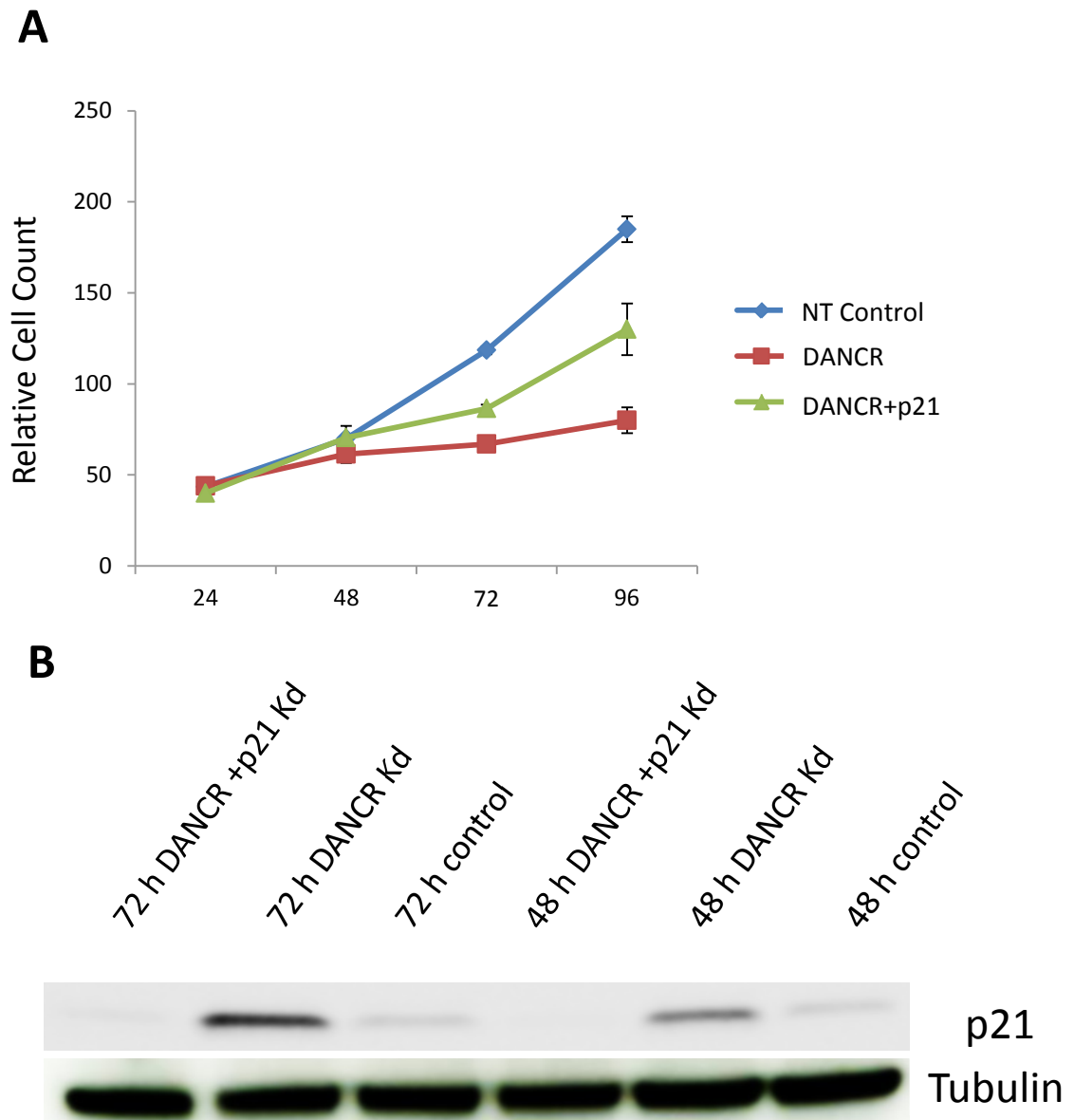


**Figure 30. DANCR, MYC, and p21 expression levels in prostate tumor versus normal samples.** Data were obtained from TCGA prostate cancer RNA-seq data analysis. Bar graphs indicate mean expression levels (n = 256 for tumors, n = 45 for normal samples; P < 0.0001 for DANCR, P < 0.0001 for Myc, and P = 0.0003 for p21 (CDKN1A)). RPKM, reads per kilobase per million.

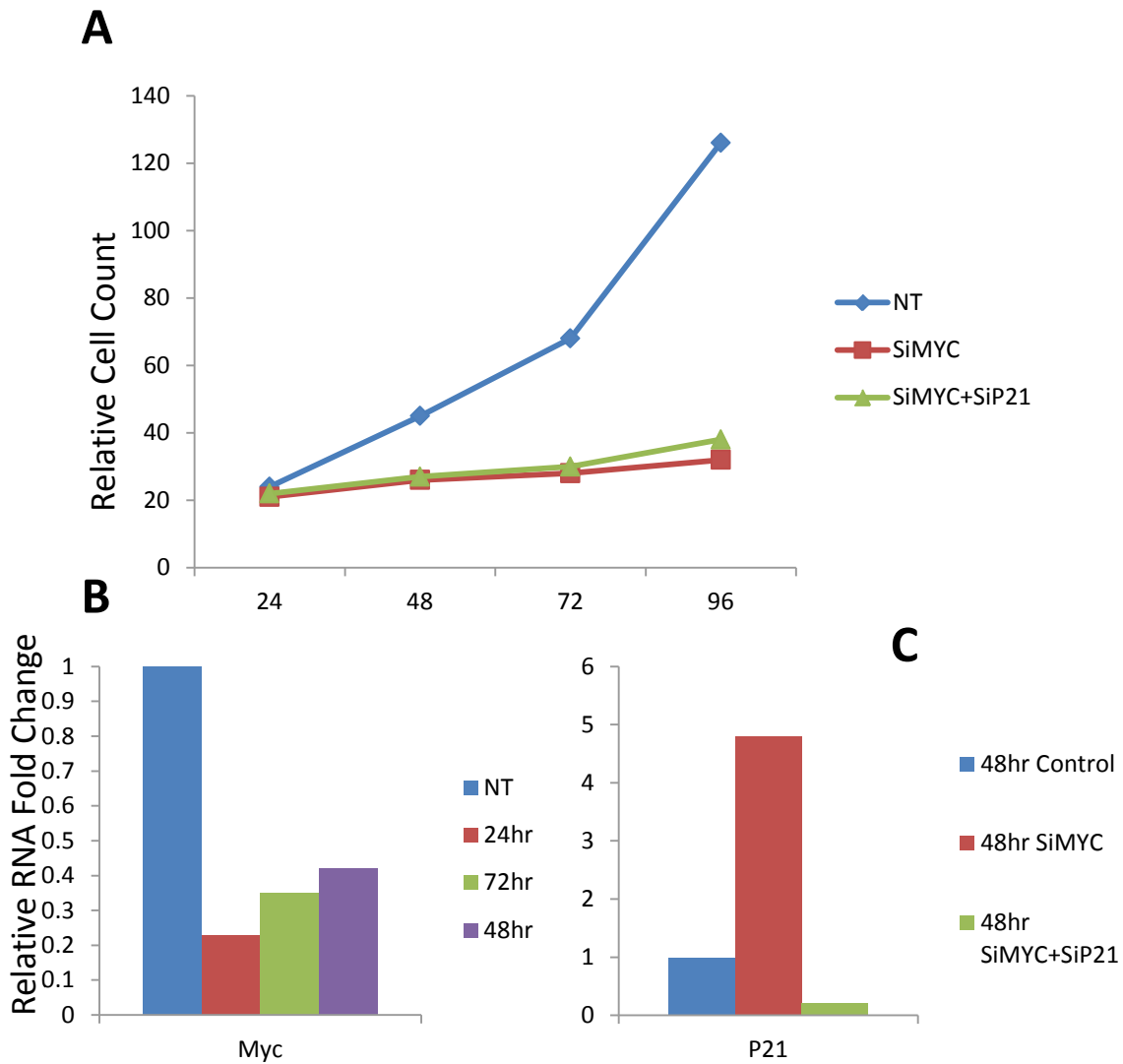


**Figure 31. The abundance of DANCR, MYC, and p21 RNA in prostate cancer.** Data were obtained from TCGA prostate cancer RNA-seq analysis. Dot blots represent the expression level in each sample and the means of those values were calculated (n = 256; P < 0.0001). RPKM, reads per kilobase per million.

In the final part of this chapter, to further substantiate this proposed pathway, I performed a critical experiment in which I tested whether I could rescue the observed growth arrest in response to DANCR depletion in PC3 cells. A double knock-down of DANCR and p21 was performed in PC3 cells. As expected, the double knock-down partially rescued the reduction in proliferation caused by the DANCR knock-down alone (Figure 32). Although DANCR likely regulates the expression of many genes, I suspected that p21 is likely one of its most important targets, which is supported by the partial rescue of this phenotype. I also performed MYC and p21 knock-down in parallel; however, this failed to recapitulate a similar amount of rescue as DANCR/p21 double knock-down (Figure 33). This is possibly because among the many targets of MYC affected by its depletion, simply manipulating p21 alone is not sufficient to alter the phenotype.



**Figure 32. Double knock-down of DANCR + p21 partially rescues the growth arrest induced by DANCR depletion in PC3 cells.** (A) The rates of proliferation partially recovered by DANCR + p21 depletion; error bars indicate the standard deviation ( $n = 3$ ). (B) Western blotting confirmed p21 induction by DANCR knock-down and depletion by the double knock-down. Tubulin was used as a loading control.



**Figure 33. Double knock-down of MYC + p21 cannot rescue MYC depletion-induced growth arrest in PC3 cells.** (A) The rates of proliferation were not recovered by MYC + p21 depletion. (B) Measurements of *MYC* mRNA levels were confirmed by qRT-PCR for different time points. (C) The *p21* mRNA levels after knock-down were measured by qRT-PCR.



## **Materials and methods**

### **Cell culture**

P493 and PC3 cells were grown as described in Chapter 2, and DU145 cells were a generous gift from Dr. Angelo De Marzo at Johns Hopkins Medical Institute. DU145 cells were cultured in RPMI 1640 with 10% Fetal Calf Serum (FCS), and 1% penicillin/streptomycin.

### **RNA interference**

The siRNA duplexes directed against DANCER and a non-targeting controls were purchased from Ambion®. An on-target plus smart pool siRNA against MYC was purchased from Dharmacon Inc. All siRNAs were used according to manufacturer's instructions. Briefly, cells were transfected using Lipofactamine 2000 (Invitrogen), and 200,000 cells in 1ml serum-free media were transfected per well in six-well plates. A total of 3  $\mu$ L Lipofactamine 2000 was used per well. The siRNAs were transfected at a final concentration of 50 nM. Then, 1 ml media containing 20% FBS was added 4–6 h following transfection. For larger cell numbers, reaction volumes were increased accordingly. Cells were counted, collected, and assayed daily for gene function every ~24 h after transfection for 4 days.

### **Cell proliferation assays**

Cells were stained with 0.4% trypan blue, and the number of viable cells was counted using a hemocytometer.

### **Cell cycle analysis**

Cell cycle analysis was carried out using the FITC BrdU Flow Kit purchased from BD Pharmingen™ following manufacturer's protocol. Cells were filtered and then analyzed by flow cytometry using a Becton Dickinson FACScan or FACS Calibur. Data analysis was carried out using FlowJo software.

### **Microarray hybridization and analysis**

PC3 cells were transfected with DANCR, MYC, and NT control siRNAs and then were harvested after 72 h. The microarray experiment was carried out as described in Chapter 2.

### **Western blotting**

For immunoblot analysis, cells were lysed in incomplete Laemmli buffer at 95°C. Concentrations of protein samples were determined using a bicinchoninic acid (BCA) kit (Pierce). Proteins were electrophoresed using precast Tris-HCl polyacrylamide gels (Bio-Rad) and transferred to nitrocellulose membranes with

a SNAP i.d.<sup>™</sup> System (Millipore) for immunoblotting. Membranes were blocked and probed with primary and secondary antibodies. Signals were visualized either with ECL Western Blotting Detection Reagent (Amersham) and Hyperfilm ECL (Amersham) or, more recently, with an Odyssey Imaging System (LI-COR) and the accompanying v3.0 software.

### **Morpholinos**

Vivo-morpholinos were designed and custom synthesized by GENE TOOLS, LLC. Cells were incubated with fixed concentrations of morpholinos for 48–72 h; notably, vivo-morpholinos do not require endo-porters for *in vitro* delivery.

**Quantitative real-time PCR (RNA quantitation and ChIP validation), survival analysis and TCGA analysis.** These assays were performed as previously described.

## **Chapter 6**

### **The mechanism of DANCR function**

In previous chapters, I proposed the pathway MYC  $\rightarrow$  DANCR  $\dashv$  p21, and established that DANCR is directly induced by MYC. The major question that still remained was the exact mechanism whereby DANCR suppresses p21. Recently, several reports of functional characterizations of lncRNAs have been published. For example, HOTAIR can bind to histone modifying enzymes (PRC2 and LSD1), thereby mediating target gene silencing through chromatin state modifications; and other reports showed that lncRNAs can promote cancer metastasis have emerged (Gupta et al., 2010; Khalil et al., 2009; Tsai et al., 2010). These reports, combined with my finding that p21 is one of many overlapping genes that can be co-suppressed by DANCR and EZH2, prompted me to develop a model for DANCR and EZH2 targeting p21 (Figure 34). DANCR can bind to EZH2, and guides it to the p21 gene locus, where EZH2 induces H3K27me3 and changes the chromatin structure to a closed state, eventually shutting off p21 expression.

In this chapter, I provide evidence for a DANCR–EZH2 partnership that can trigger a target gene silencing mechanism and propose a lncRNA gene targeting mode that occurs via RNA–RNA interactions. Several steps to confirm the interactions between components in my model will be described in this chapter.

### **Further confirmation that p21 is an EZH2 target**

The first step to validate my model was to further confirm that EZH2 directly regulates p21 in prostate cancer. Though there have been reports of EZH2 targeting p21 in some cell lines, I want to confirm this in prostate cancer cell lines too. I have previously shown that EZH2 binds to the p21 promoter by ChIP and validated this finding by qRT-PCR, showing over 5-fold enrichment of the EZH2 signal relative to IgG controls at p21 locus. Similar to previous experiments that showed DANCR suppression of p21, I performed one more experiment on EZH2–p21 using a loss-of-function approach using siRNA-mediated double knock-down of EZH2 and p21 in PC3 and DU145 cells. After EZH2 depletion, p21 protein levels were significantly increased in both cell lines (Figure 35). This result further confirms that p21 is directly suppressed by EZH2.

### **EZH2 directly binds to DANCR**

Next, I sought to establish the direct connection between EZH2 and DANCR through a series of binding experiments. RNA immunoprecipitation (RIP) of EZH2 from PC3 cells specifically recovered endogenous DANCR RNA. Conversely, IP for IgG did not retrieve DANCR, and neither EZH2 nor IgG IP retrieved U1 RNA, a nuclear ncRNA that served as a negative control (Figure 36). Thus, I confirmed that DANCR can be ‘pulled down’ by EZH2 IP. To be thorough and prove that this direct interaction is valid, I also performed a RNA pull-down assay. Biotin-labeled DANCR, antisense DANCR, and green fluorescent protein

(GFP) were *in vitro* transcribed and purified. RNAs were then mixed with PC3 nuclear extracts and incubated at room temperature. Streptavidin-agarose beads were added, incubated with the mixture, washed, and boiled. The retrieved protein was detected by standard western blotting. Only purified biotinylated DANCR RNA, but not GFP RNA or an antisense DANCR RNA fragment, specifically retrieved EZH2 from PC3 cell nuclear extracts (Figure 37).

To identify the specific interaction domain required for DANCR/EZH2 interaction, I used a series of DANCR deletion mutants, and different lengths of DANCR fragments were *in vitro* transcribed (Figure 38). The EZH2-binding activity was mapped to the 3'-end of DANCR transcripts, as only full-length DANCR was able to 'pull-down' EZH2, but not the fragments without the 3'-end sequences. Thus, the 3'-end of DANCR is a required for binding of EZH2.

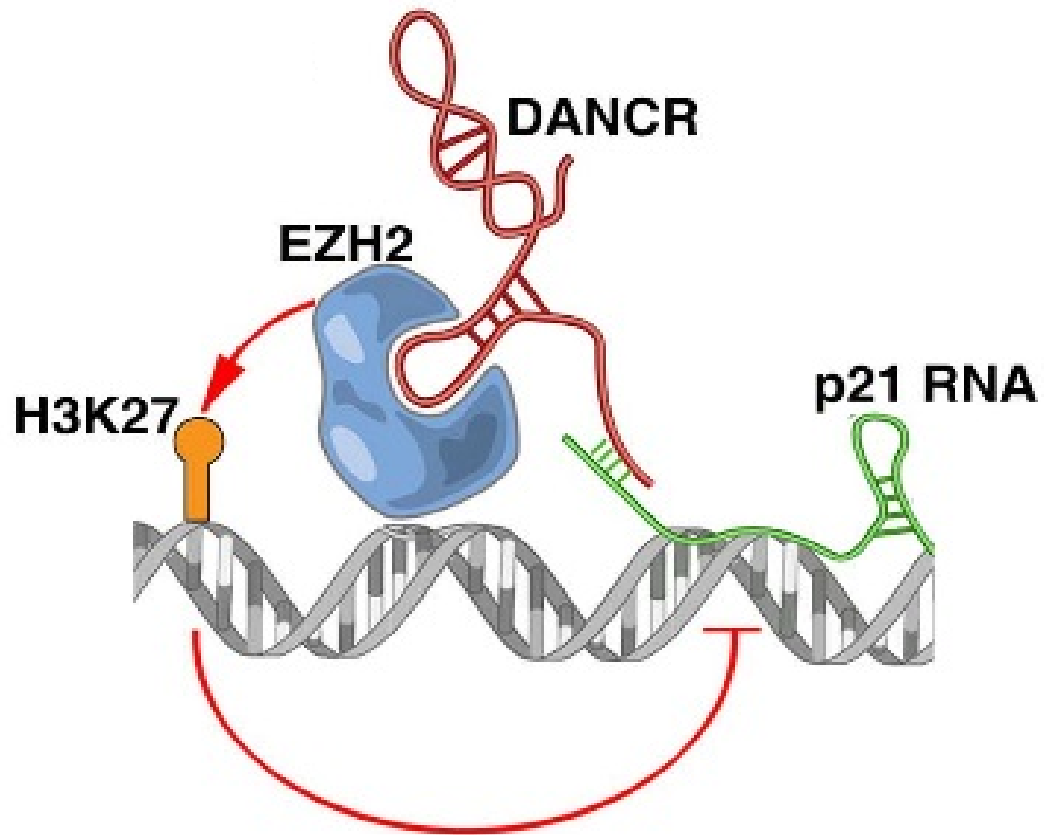
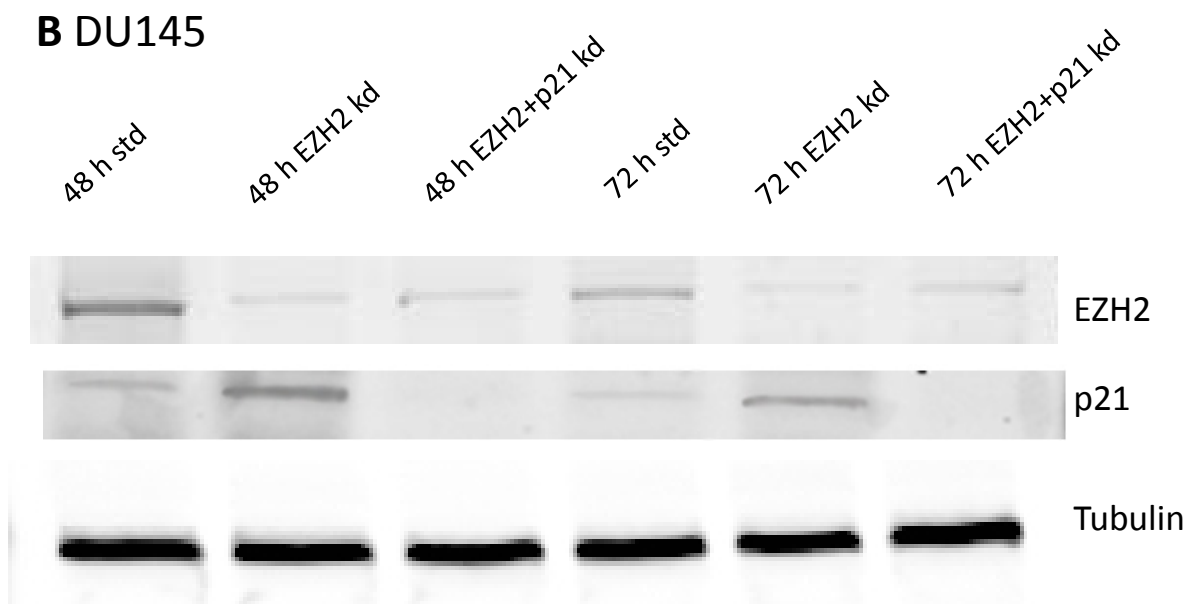
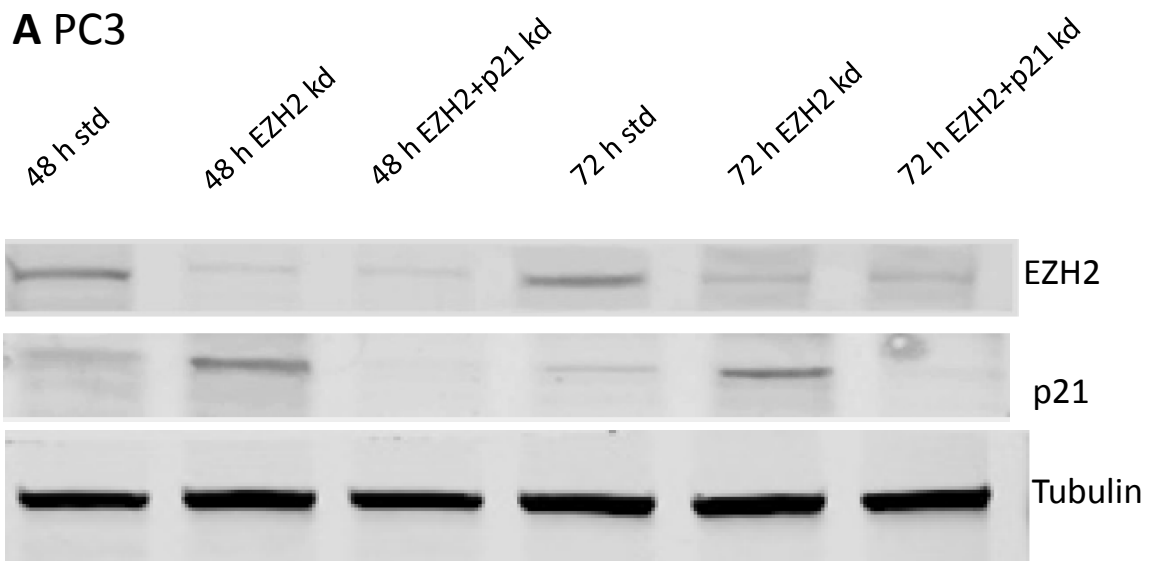
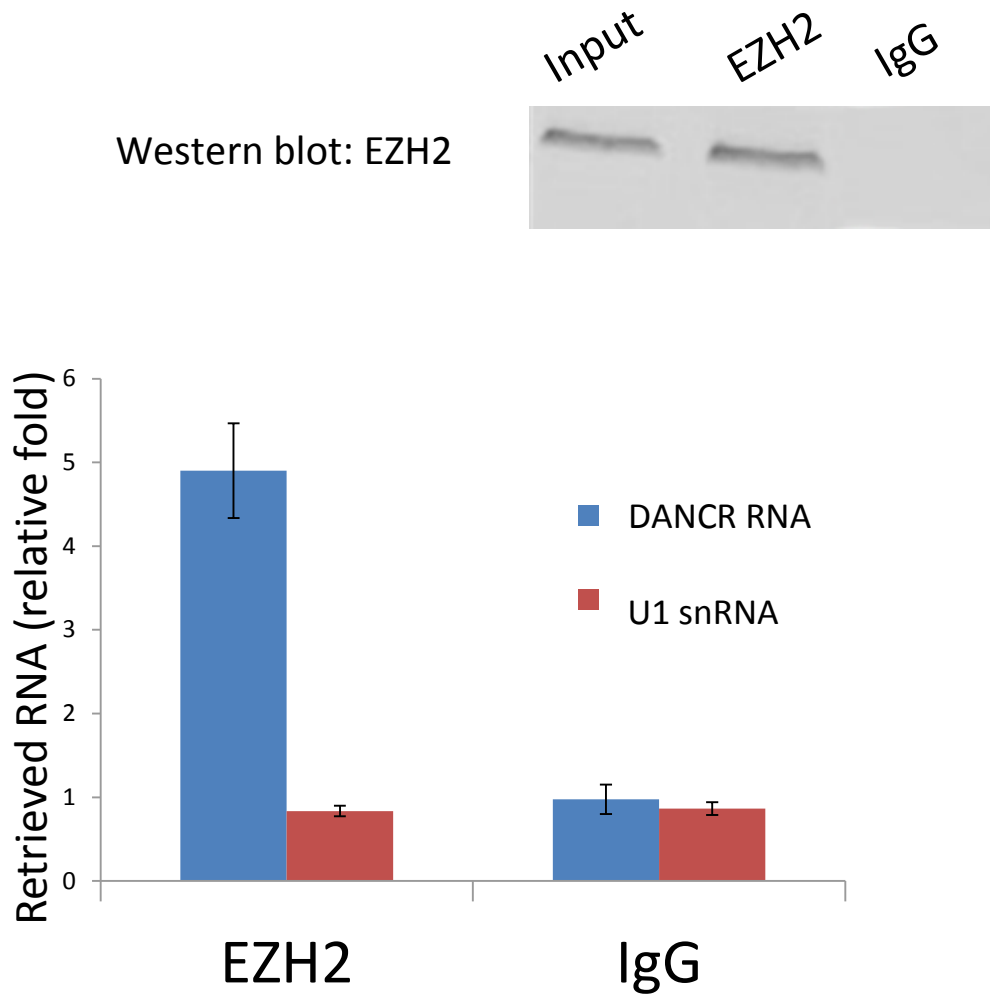


Figure 34. A working model illustrating the potential mechanism of DNACR–EZH2 silencing of p21.

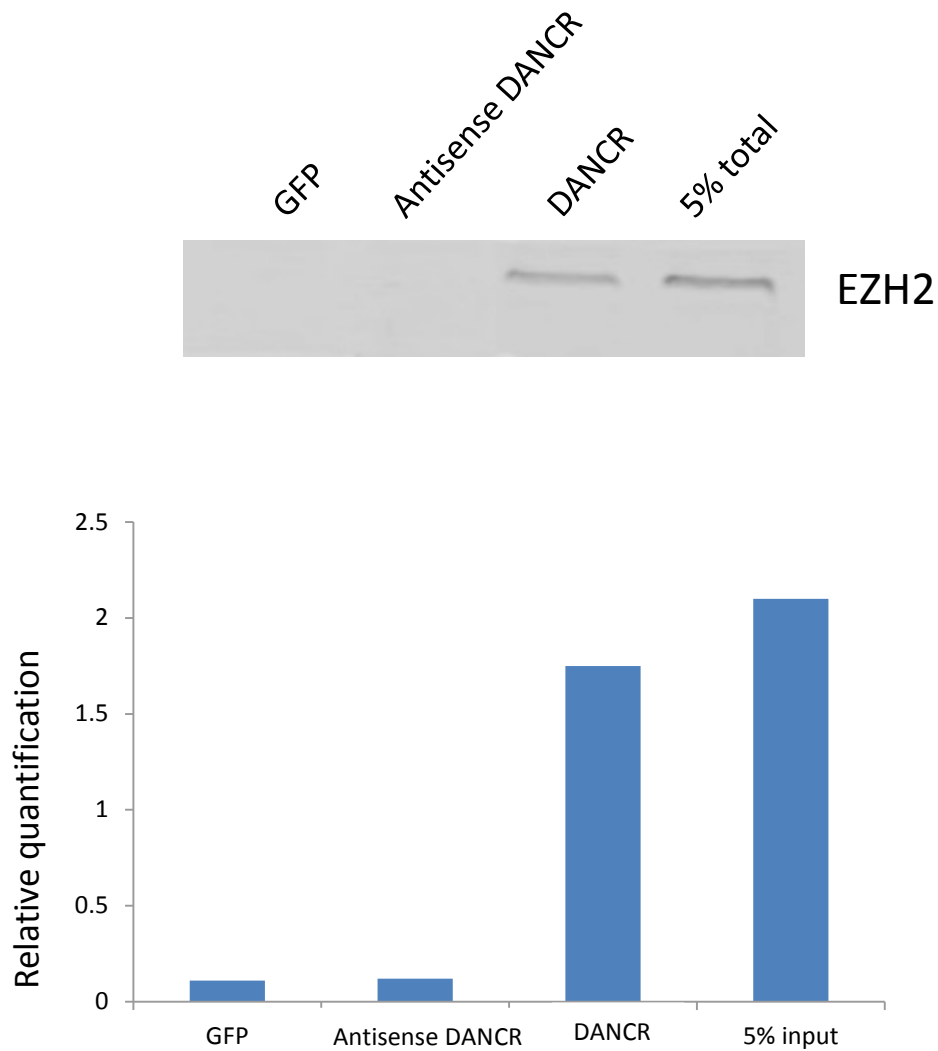




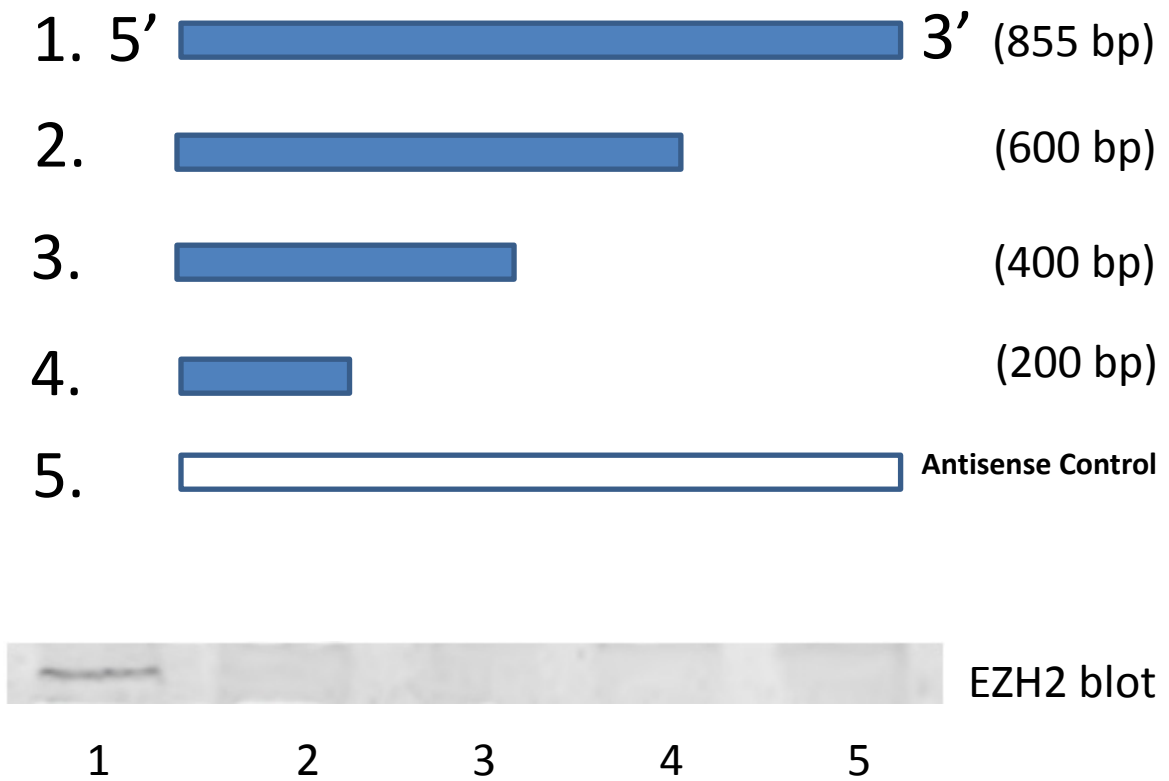
**Figure 35. Immunoblot of EZH2 and p21 protein levels after knock-down of either EZH2 alone or both EZH2 and p21.** Protein samples from 48 and 72 h after transfection were measured by western blotting. Tubulin was used as a loading control. Std, non-targeting siRNA (A) in PC3 cells, or (B) in DU145 cells.



**Figure 36. EZH2 RIP specifically retrieves DANCR RNA.** Bar graphs indicate retrieved RNA fold-changes compared to U1 snRNA. Data is normalized to mock-IP (immunoglobulin G). Error bars indicate standard deviation (n = 3). Retrieved EZH2 proteins were detected by immunoblotting



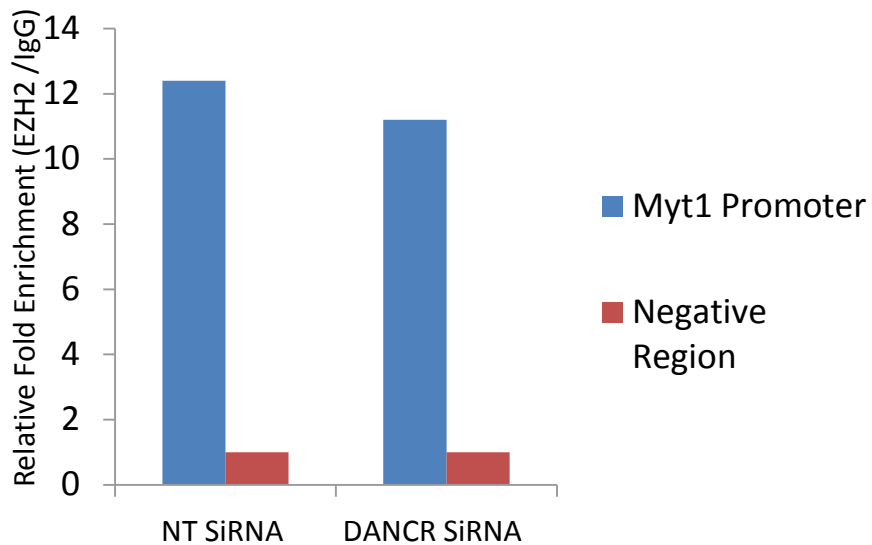
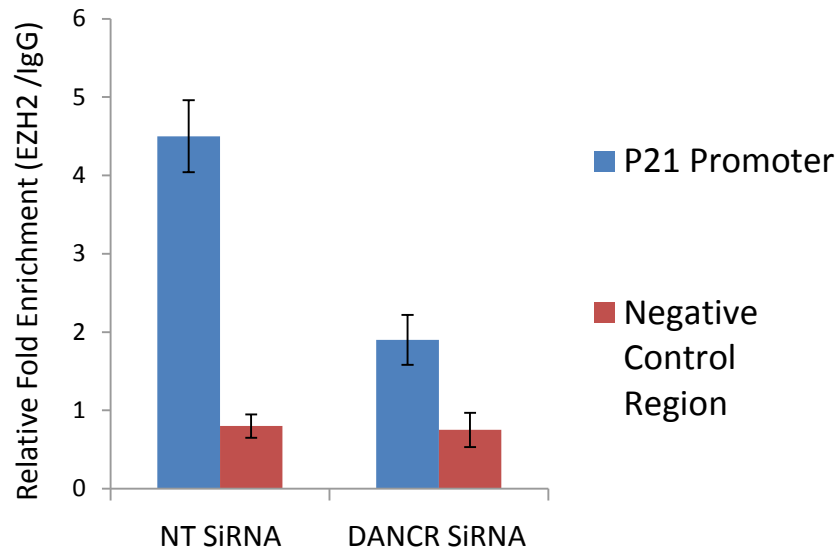
**Figure 37. *In vitro* transcribed (IVT) biotinylated DANCR retrieves EZH2 protein.** A total of 5% total input was included as a positive control, and both GFP and anti-sense DANCR were used as negative controls. EZH2 signals on the blot were quantified using Odyssey V3.0 software.



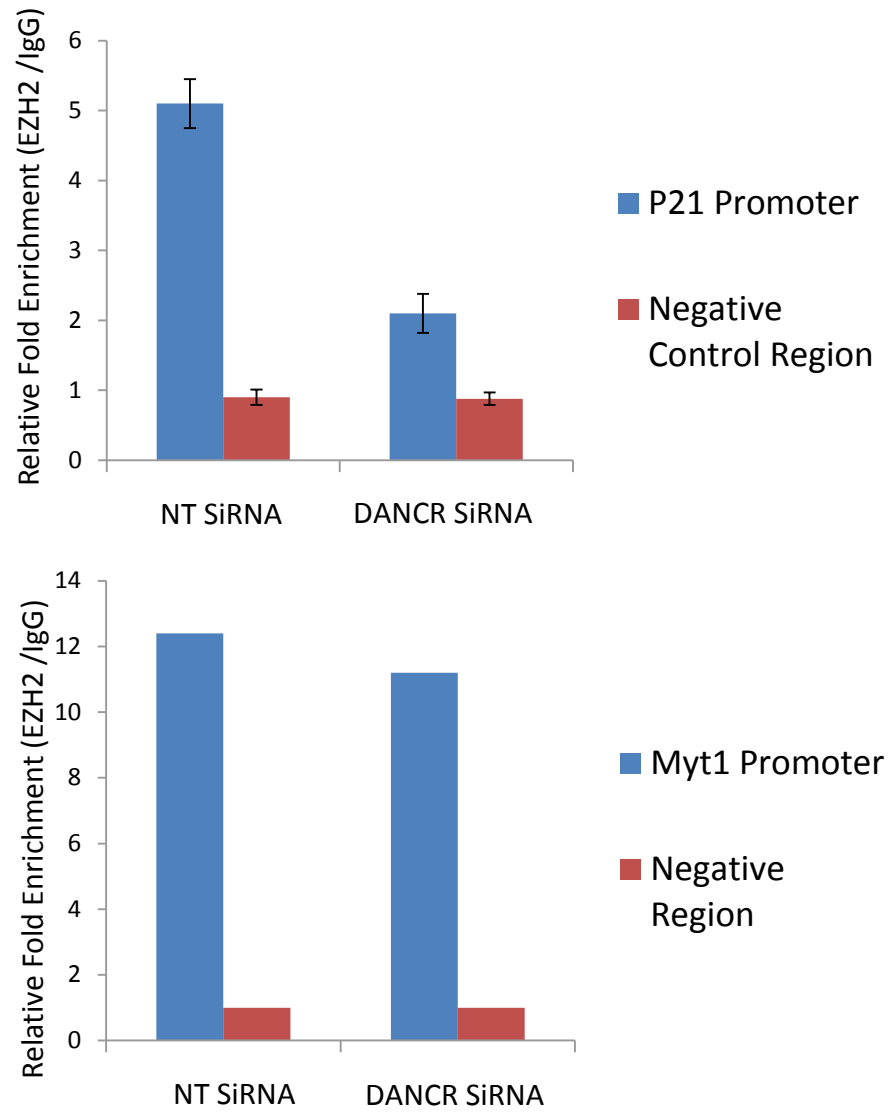
**Figure 38. The 3'-end of DANCR is required for EZH2 binding.** Lanes 1–5 on the western blot correspond to the pull down blot from the fragments numbered 1–5 above. Number 5 is an anti-sense DANCR control colored in white.

## **EZH2 binding to the p21 locus requires DANCR**

To determine whether DANCR is necessary for EZH2 binding to p21 and suppression of its expression, I designed a CHIP-qPCR experiment to test whether siRNA-mediated DANCR depletion would affect EZH2 binding to the p21 locus. A locus nearby that has no previously reported EZH2 binding activity was selected as a negative control. Additionally, MYT1, a known EZH2 target for which expression did not seem to be affected by DANCR manipulation in my analysis, was also chosen as another negative control. I found that DANCR knock-down significantly reduced EZH2 binding at the p21 promoter region compared to the negative regions in both PC3 and DU145 cells, and MYT1 binding by EZH2 did not appear to change after siRNA knock-down of DANCR (Figures 39 and 40). This finding suggests that DANCR is required for EZH2 targeting of the p21 promoter.



**Figure 39. DANCR depletion reduces EZH2 binding at the p21 promoter locus in PC3 cells.** Bar graphs represent the fold-change of the EZH2 ChIP signal over IgG for each locus. Error bars indicate standard deviation (n = 3).



**Figure 40. DANCR depletion reduces EZH2 binding at the p21 promoter locus in DU145 cells.** Bar graphs show the fold-change of EZH2 ChIP signals over IgG for each locus. Error bars indicate standard deviation (n = 3).

## DANCR targets p21 via RNA–RNA interactions

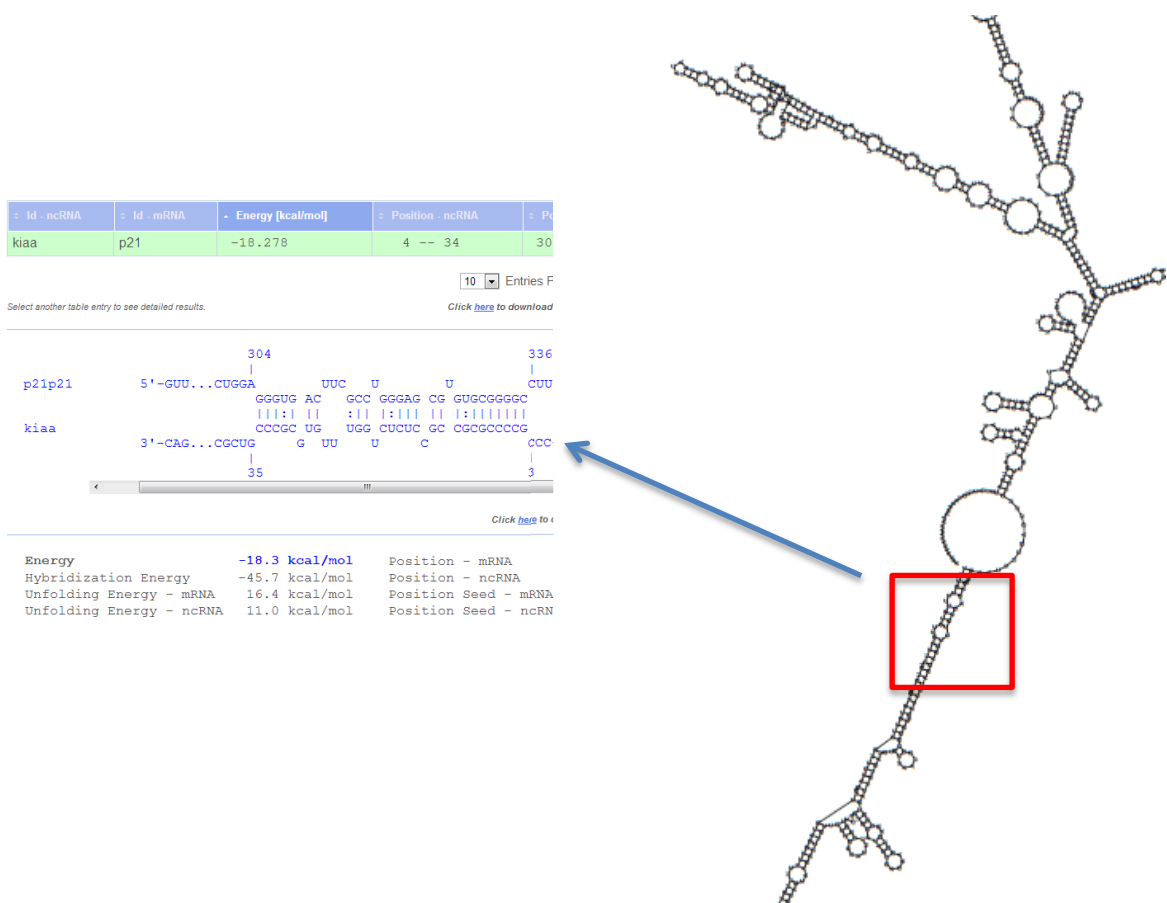
The question of how DANCR targets p21 still remained, so I sought to address question of the exact mechanism whereby DANCR guides EZH2 to the p21 locus. Since the early 1980's, there have been reports identifying ncRNA interactions with pre-mRNAs through sequence complementary, which were followed by genetics and *in vitro* affinity binding experiments for validation (Lerner et al., 1980; Zhuang and Weiner, 1986). I suspected that DANCR might recognize p21 mRNA transcripts through complementary sequences via RNA–RNA interactions. However, mapping of RNA–RNA interactions remains a challenge. Until recently, *in silico* approaches have been utilized to investigate RNA-RNA interactions (Dieterich and Stadler, 2013). I first attempted to predict DANCR–p21 mRNA interaction sites and to simulate the possible folding of the two using publically available software, *IntaRNA* (Busch et al., 2008; Wright et al., 2014) and *RNAfold* webserver (<http://rna.tbi.univie.ac.at/>). A 5'-domain of DANCR was computationally predicted to be most likely to interact with p21 mRNA, consistent with my earlier finding that the 3'-domain of DANCR is a binding site for EZH2 (Figure 41).

The experimental design for validating that binding included first treating PC3 cells with Bortezomib to artificially increase the expression levels of p21 mRNA, as the endogenous level of p21 in cancer cells was too low for detect any potential interactions (Figure 42B). A series of biotinylated DANCR fragments of different lengths, along with negative control antisense DANCR transcripts, were *in vitro* transcribed and incubated with total RNA purified from Bortezomib-treated

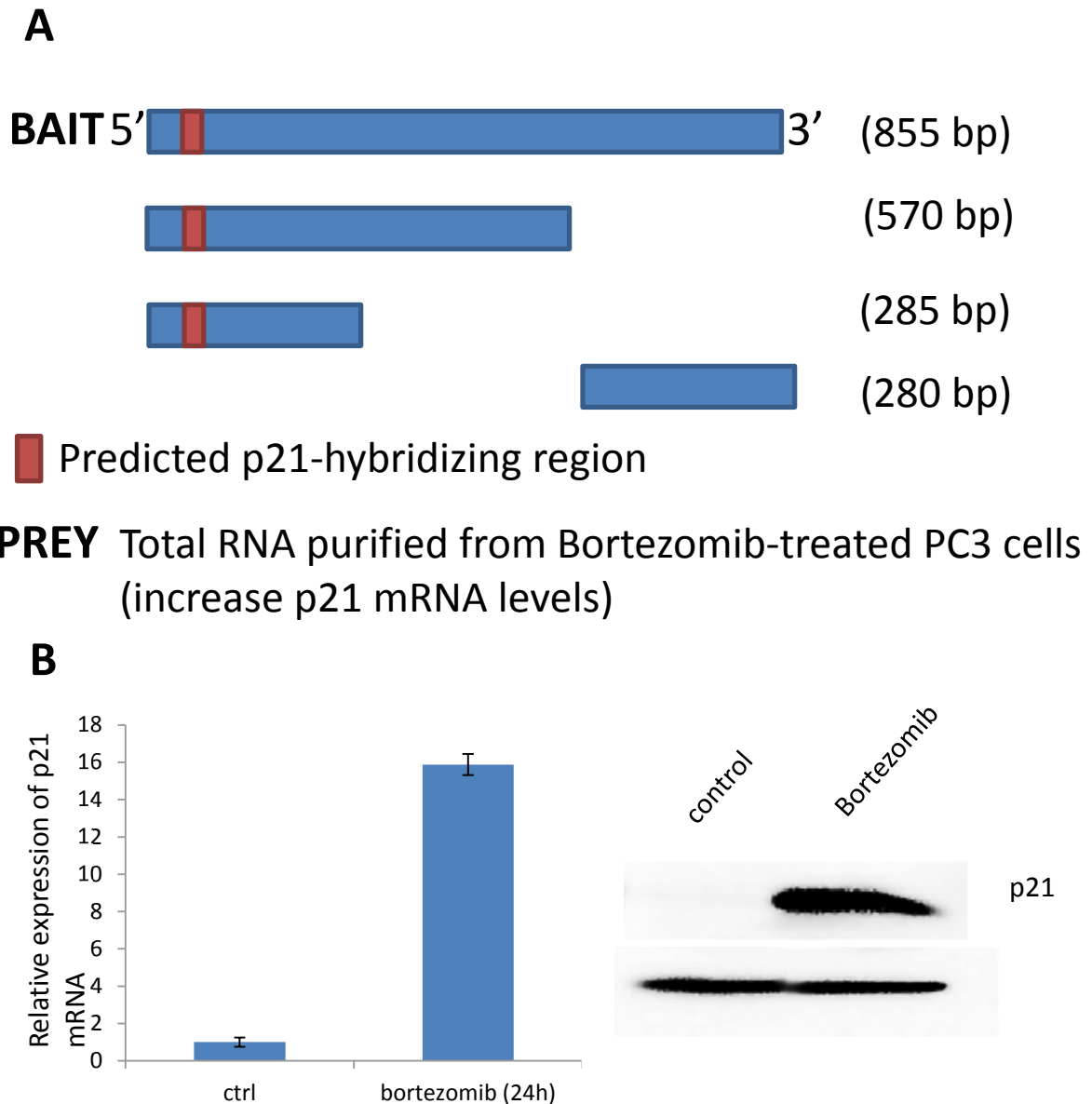


PC3 cells (Figure 42A). This mixture was cross-linked with 4'-aminomethyltrioxalen (AMT), a psoralen-derivative cross-linker that only generates cross-links between uridine bases in RNA, but does not react to proteins (Calvet and Pederson, 1979). Pull-down were then achieved by streptavidin agarose beads. mRNA levels for p21 and two negative controls (MYT1 and HOXC8, genes for which expression levels appeared not to be regulated by DANCR) were detected using qRT-PCR. The data shown in Figures 43–46 confirmed my hypothesis that 5'-end domain of DANCR interacts with p21 mRNA.

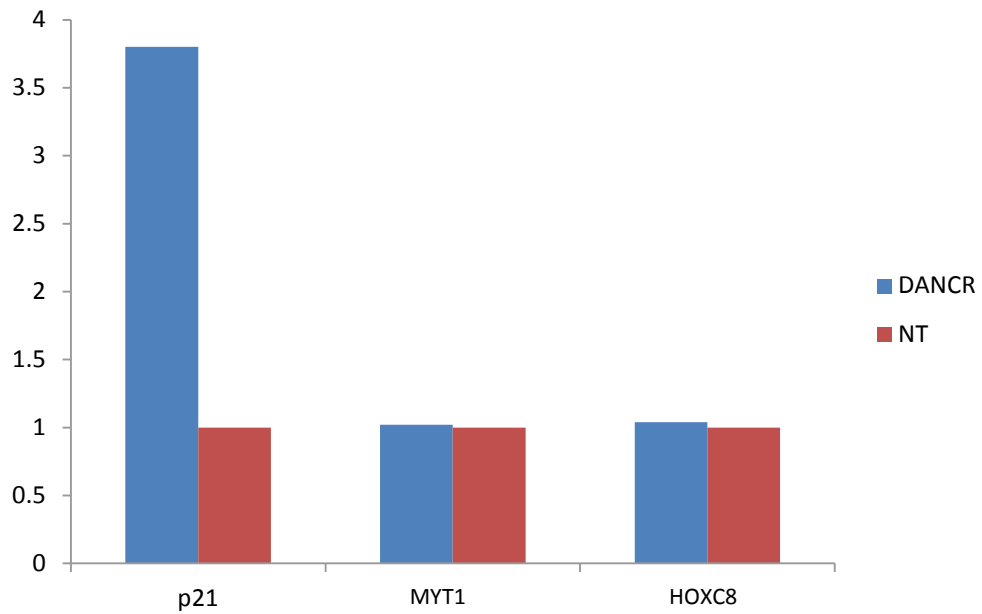
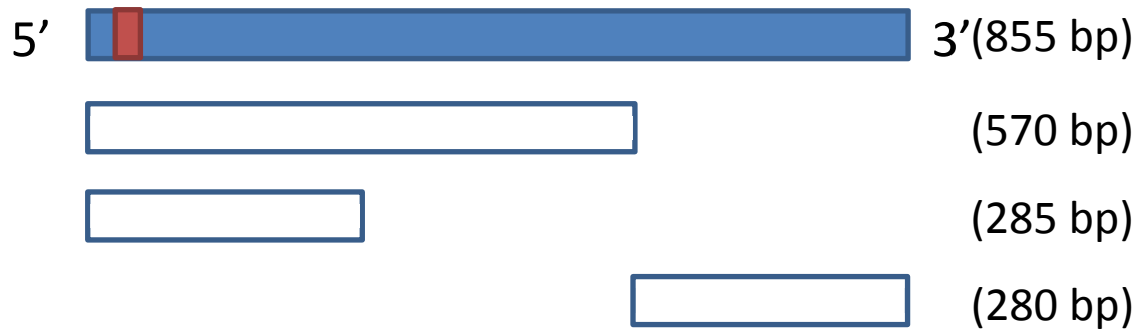
Together, these findings confirm my proposed mechanistic model for DANCR–EZH2-specific silencing of p21 by guiding this complex to the locus via RNA–RNA interactions.



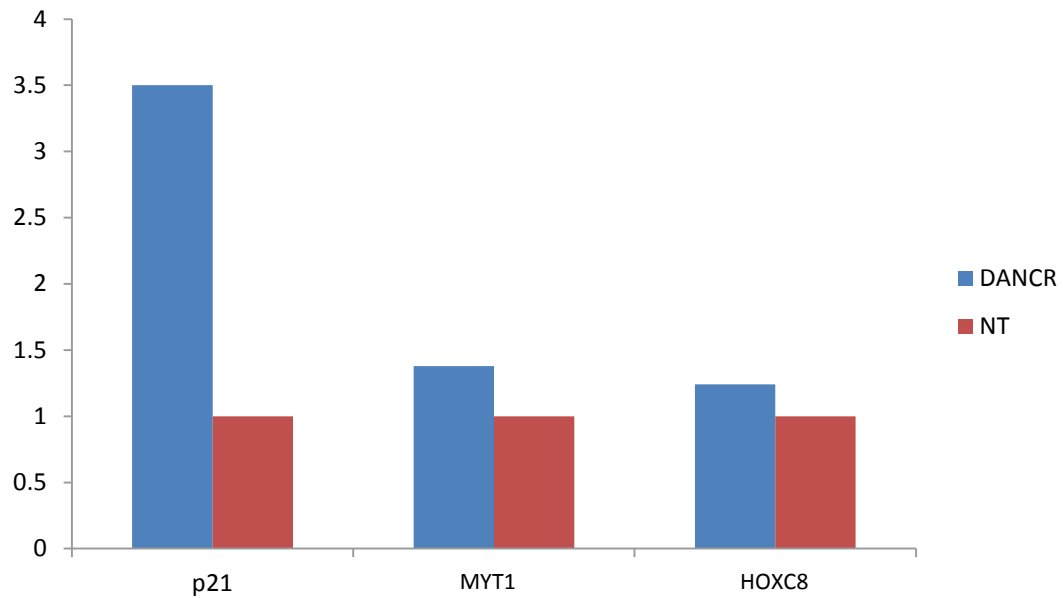
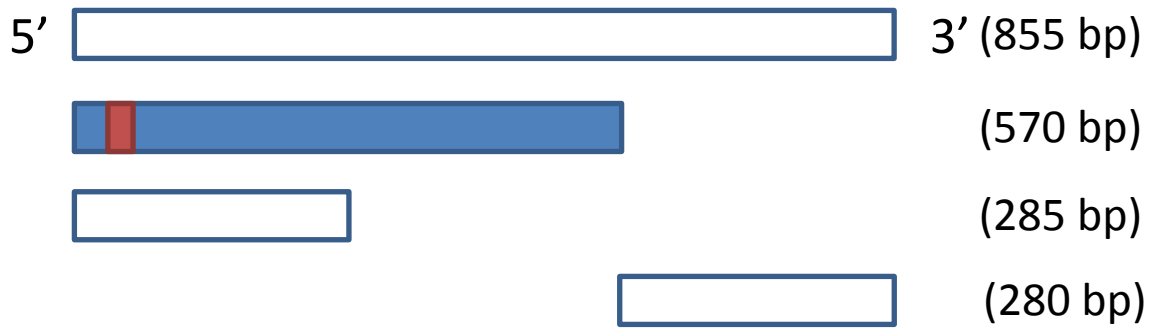
**Figure 41. Possible domain mediating DANCR–p21 mRNA interactions identified by IntaRNA and RNAfold.** Simulated interaction was consistent with a minimum free energy model. The sequence and location of a possible binding site is shown on the left



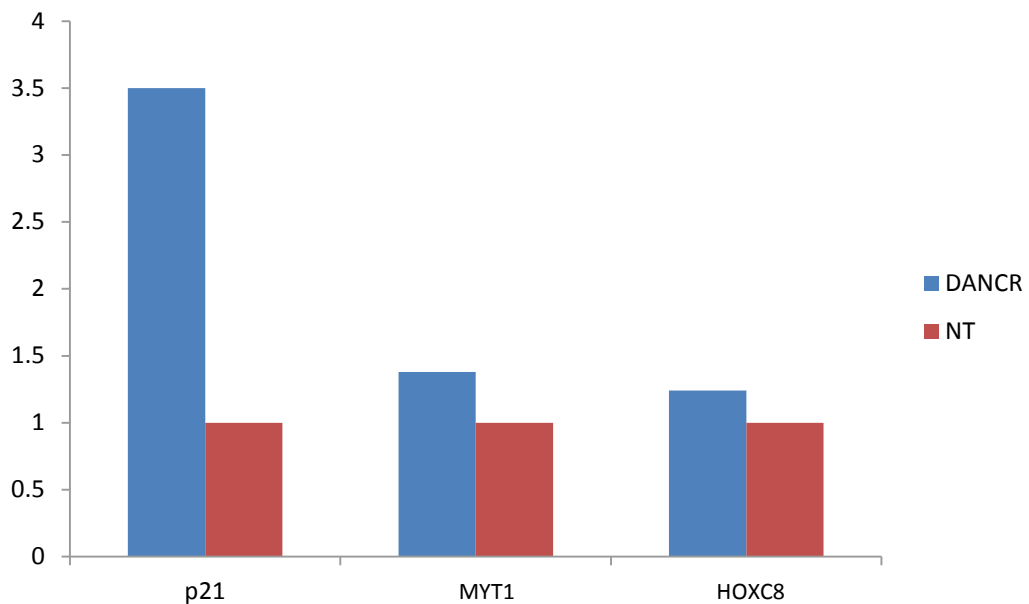
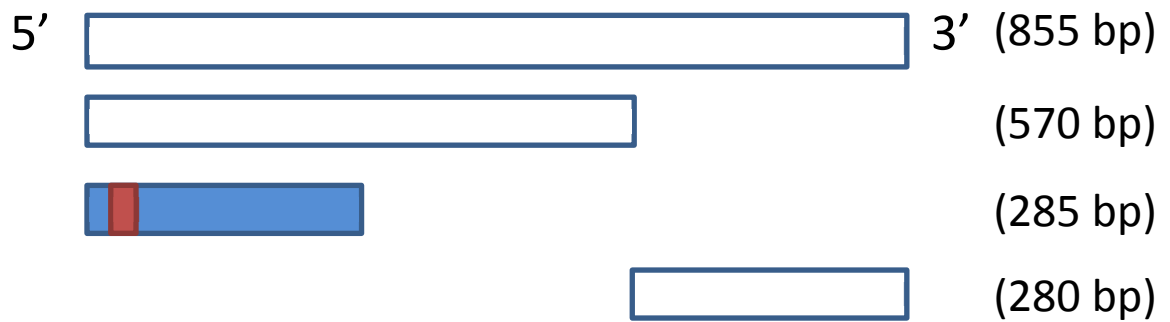
**Figure 42. A schematic illustration of the experimental design for mapping DANCR–p21 mRNA interactions.** (A) Experimental design: Blue bars represent in vitro transcribed DANCR fragments as baits, red square is the predicted p21 hybridizing site. (B) qRT-PCR analysis and immunoblotting showing increased p21 levels.



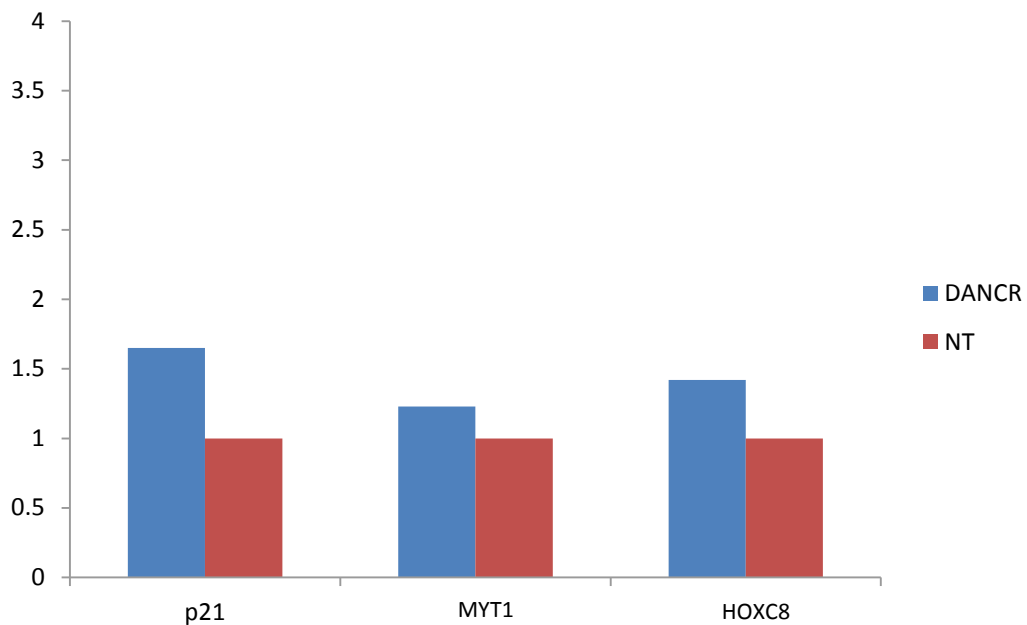
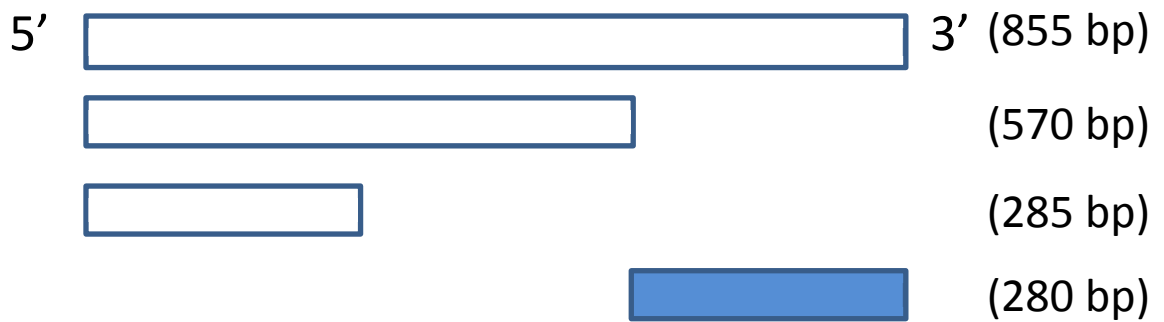
**Figure 43. Full-length DANCR could specifically retrieve p21 mRNA.** The fragment shown in blue was used to retrieve associated RNA molecules. NT is an antisense and was used as a negative control. MYT1 and HOXC8 were also included as negative controls.



**Figure 44. A DANCER fragment containing the 5'-end could specifically retrieve p21 mRNA.** The fragment shown in blue was used to retrieve associated RNA molecules. NT indicates an antisense DANCER fragment used as a negative control. MYT1 and HOXC8 were also included as negative controls.



**Figure 45. A DANCER fragment containing the 5'-end could specifically retrieve p21 mRNA (continued).** The fragment shown in blue was used to retrieve associated RNA molecules. NT is an antisense DANCER fragment that was used as a negative control. MYT1 and HOXC8 were also included as negative controls.



**Figure 46. A 3'-end DANCR fragment cannot specifically retrieve p21 mRNA.** The fragment shown in blue was used to retrieve associated RNA molecules. NT is an antisense DANCR fragment that was used as a negative control. MYT1 and HOXC8 were also included as negative controls.

## **DANCR targets genes via RNA–RNA interactions as a general mechanism**

To this point, I validated the p21 gene silencing by DANCR–EZH2 partnership model with experimental evidence. Whether this model can be applied to a larger group of genes targeted by this duo as a general mechanism remains an open question. In the final part of this thesis, I attempt to expand my proposed model using exploratory *in silico* methods.

I selected 16 genes that included my validated DANCR target genes and the top DANCR target genes that were identified from the microarray analysis (Table 3). Similar to my approach for p21, the mRNA sequences of these 16 transcripts were subjected to computational folding and binding simulations with DANCR by *IntaRNA* and *RNAfold*. I found that 8 of the mRNA transcripts for the 16 genes have predicted DANCR-interacting domains that are exact matches or largely overlap with the DANCR–p21 RNA–RNA interaction domain at the 5'-end of DANCR. More specifically, for CTSB, PDCD6, PIP4K2A, and SLC16A2, the putative binding sites are located on the 5'-end of the respective pre-mRNA transcripts, and the sites on GAS6, GHR, KRT34, and LOXL4 are further away from the 5'-ends, but remain within the first half of the pre-mRNA molecules (Figures 47 and 48). These findings suggest that the DANCR guiding EZH2 to targets via RNA–RNA recognition model proposed by us is not an isolated event that only applies to p21, but instead suggests that it could possibly account for the mechanism whereby many EZH2 regulated genes are targeted.



## CTSB

Id - ncRNA	Id - mRNA	- Energy [kcal/mol]	Position - ncRNA	Position - mRNA
kliaa	ctsb-1	-16.128	1 -- 18	1 -- 18

10 Entries Per Page Page 1 of 1

Select another table entry to see detailed results. [Click here to download the current ordered table.](#)

```

ctsb-1ctsb-1b-1      1           19
                    |           |
                    5'-      G C A UACUU...CCU-3'
                    |           |
                    GGGCGG GC GGG GGG
                    ||||| || ||| |||
                    CUCGCGC GC CCC CCC
kliaa                3'-CAG...GGUUCU      G C G -5'
                    |           |
                    19          1
    
```

[Click here to download this interaction.](#)

Energy	-16.1 kcal/mol	Position - mRNA	1 -- 18
Hybridization Energy	-31.7 kcal/mol	Position - ncRNA	1 -- 18
Unfolding Energy - mRNA	9.5 kcal/mol	Position Seed - mRNA	1 -- 7
Unfolding Energy - ncRNA	6.1 kcal/mol	Position Seed - ncRNA	12 -- 18



## GAS6

Id - ncRNA  Records 1-1 of 1 [reset](#)

Id - ncRNA	Id - mRNA	- Energy [kcal/mol]	Position - ncRNA	Position - mRNA
kliaa	gas6	-24.262	3 -- 44	732 -- 783

10 Entries Per Page Page 1 of 1

Select another table entry to see detailed results. [Click here to download the current ordered table.](#)

```

gas6gas66          731
                   |
                   5'-CCG...GGCAG CAAGACAUA A C
                   |           |           |
                   GACCUGC GAGC GUGGCAGACU G GAGGC
                   ||||| ||| ||| ||| |||
                   CUGGAGC CUGC CGGUGUUGG C CUCGG
kliaa              3'-CAG...GGAGC      C      U
                   |           |
                   45          1
    
```

[Click here to download this interaction.](#)

Energy	-24.3 kcal/mol	Position - mRNA	732 -- 783
Hybridization Energy	-61.6 kcal/mol	Position - ncRNA	3 -- 44
Unfolding Energy - mRNA	21.9 kcal/mol	Position Seed - mRNA	756 -- 762
Unfolding Energy - ncRNA	15.5 kcal/mol	Position Seed - ncRNA	23 -- 29



## GHR

Id - ncRNA	Id - mRNA	- Energy [kcal/mol]	Position - ncRNA	Position - mRNA
kliaa	ghr-2	-16.191	8 -- 30	1599 -- 1622

10 Entries Per Page Page 1 of 1

Select another table entry to see detailed results. [Click here to download the current ordered table.](#)

```

ghr-2ghr-2-2      1598          1623
                   |           |
                   5'-GAG...CUGGC UGC AUGUU...CCA-3'
                   |           |           |
                   GC CAGSCUAGAG GGGGSCGC
                   || ||| ||| ||| ||| |||
                   GC GUGUGGUCUC CCGCCGSC
kliaa              3'-CAG...GCCCC U      CCGCCCC-5'
                   |           |
                   31          7
    
```

[Click here to download this interaction.](#)

Energy	-16.2 kcal/mol	Position - mRNA	1599 -- 1622
Hybridization Energy	-37.4 kcal/mol	Position - ncRNA	8 -- 30
Unfolding Energy - mRNA	10.3 kcal/mol	Position Seed - mRNA	1616 -- 1622
Unfolding Energy - ncRNA	10.9 kcal/mol	Position Seed - ncRNA	8 -- 14



## KRT34

Id - ncRNA	Id - mRNA	- Energy [kcal/mol]	Position - ncRNA	Position - mRNA
kliaa	krt34	-17.619	16 -- 63	1109 -- 1152

10 Entries Per Page Page 1 of 1

Select another table entry to see detailed results. [Click here to download the current ordered table.](#)

```

krt34krt3434      1108
                   |
                   5'-GGA...GUCCU UC G GC A
                   |           |           |
                   AGCUGGCAGAGA C CUGUGACCCU GA GGC GA
                   ||||| ||| ||| ||| ||| |||
                   UCGACUGUUCU G GACGUGGAC CU CCG GU
kliaa              3'-CAG...UGAGG      G GC CCGU
                   |           |
                   64          1
    
```

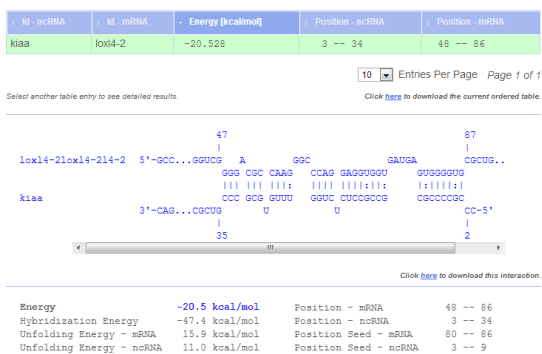
[Click here to download this interaction.](#)

Energy	-17.6 kcal/mol	Position - mRNA	1109 -- 1152
Hybridization Energy	-56.5 kcal/mol	Position - ncRNA	16 -- 63
Unfolding Energy - mRNA	15.5 kcal/mol	Position Seed - mRNA	1143 -- 1149
Unfolding Energy - ncRNA	23.4 kcal/mol	Position Seed - ncRNA	20 -- 26

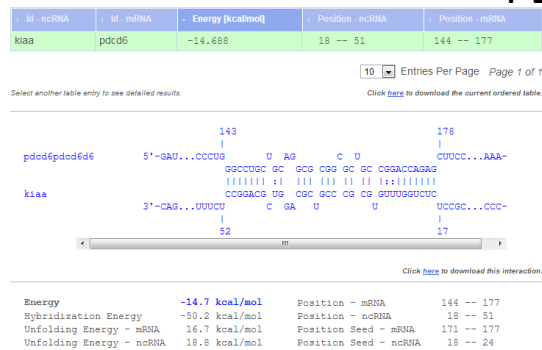


**Figure 47. Predicted common interaction domains of DANCR and four of its targets identified by *IntaRNA* and *RNAfold*.**

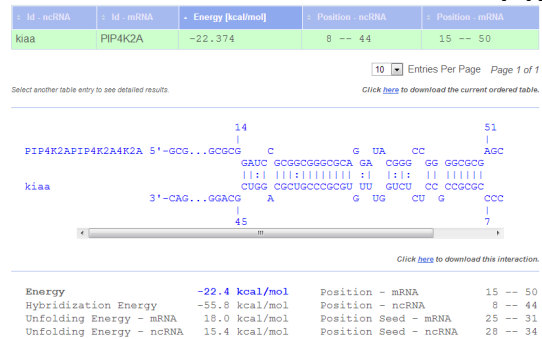
## LOXL4



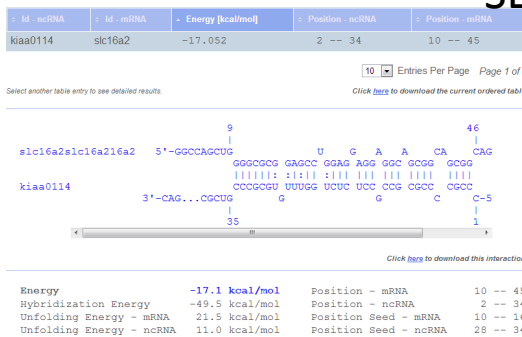
## PDCD6



## PIP4K2A



## SLC16A2



**Figure 48. Predicted common interaction domains of DANCR and another four DANCR targets identified by *IntaRNA* and *RNAfold*.**

## **Materials and Methods**

### **RNA immunoprecipitation**

The RIP experiment was carried out using a Magna RIP™ RNA-Binding Protein Immunoprecipitation Kit (Millipore) following the manufacturer's protocol.

### **RNA pull-down assay**

The RNA pull-down assay was carried out as previously described (Tsai et al., 2010). Briefly, biotin-labeled RNAs were *in vitro* transcribed using Biotin-RNA Labeling Mix (Roche) and T7 RNA polymerase (Promega), treated with RNase-free DNase I (Promega), and then purified with a RNeasy Mini Kit (QIAGEN). Biotinylated RNA was heated to 90°C for 2 min, incubated on ice for 2 min, supplied with RNA structure buffer, and then heated to room temperature for 20 min to allow proper secondary structures to form. PC3 cell pellets were resuspended in PBS, nuclear isolation buffer, and incubated on ice for 20 min (with frequent mixing). Nuclei were pelleted by centrifugation at 2500×g for 15 min. Nuclear pellets were resuspended in RIP buffer. Nuclear membranes and debris were pelleted by centrifugation at 13,000 RPM for 10 min. Folded RNA was then mixed with PC3 nuclear extracts in RIP buffer and incubated at room temperature for 1 h. Washed streptavidin-agarose beads (Invitrogen) were added to each binding reaction, which were further incubated at room temperature for 1 h. Beads were washed and boiled in SDS buffer, and the retrieved protein was detected by standard western blotting.

### **RNA–RNA interaction predictions**

The DANCR sequence was queried against target gene mRNA using *intaRNA* to predict the interaction domains between the two transcripts. Additionally, the RNA–RNA interaction MFE (minimum free energy) secondary structure was predicted using the *RNAfold* webserver.

**Cell culture, RNA interference, western blotting, quantitative real-time PCR, and chromatin immunoprecipitation.** These experiments were performed as described previously.

## REFERENCES

- Adams, J.M., Harris, A.W., Pinkert, C.A., Corcoran, L.M., Alexander, W.S., Cory, S., Palmiter, R.D., and Brinster, R.L. (1985). The c-myc oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice. *Nature* *318*, 533-538.
- Adhikary, S., and Eilers, M. (2005). Transcriptional regulation and transformation by Myc proteins. *Nature reviews Molecular cell biology* *6*, 635-645.
- Al Olama, A.A., Kote-Jarai, Z., Giles, G.G., Guy, M., Morrison, J., Severi, G., Leongamornlert, D.A., Tymrakiewicz, M., Jhavar, S., Saunders, E., *et al.* (2009). Multiple loci on 8q24 associated with prostate cancer susceptibility. *Nature genetics* *41*, 1058-1060.
- Amati, B., Brooks, M.W., Levy, N., Littlewood, T.D., Evan, G.I., and Land, H. (1993). Oncogenic activity of the c-Myc protein requires dimerization with Max. *Cell* *72*, 233-245.
- Amati, B., Dalton, S., Brooks, M.W., Littlewood, T.D., Evan, G.I., and Land, H. (1992). Transcriptional activation by the human c-Myc oncoprotein in yeast requires interaction with Max. *Nature* *359*, 423-426.
- Arvanitis, C., and Felsher, D.W. (2006). Conditional transgenic models define how MYC initiates and maintains tumorigenesis. *Seminars in cancer biology* *16*, 313-317.
- Augenlicht, L.H., Wadler, S., Corner, G., Richards, C., Ryan, L., Multani, A.S., Pathak, S., Benson, A., Haller, D., and Heerdt, B.G. (1997). Low-level c-myc amplification in human colonic carcinoma cell lines and tumors: a frequent, p53-independent mutation associated with improved outcome in a randomized multi-institutional trial. *Cancer research* *57*, 1769-1775.
- Ayer, D.E., Kretzner, L., and Eisenman, R.N. (1993). Mad: a heterodimeric partner for Max that antagonizes Myc transcriptional activity. *Cell* *72*, 211-222.
- Ayer, D.E., Lawrence, Q.A., and Eisenman, R.N. (1995). Mad-Max transcriptional repression is mediated by ternary complex formation with mammalian homologs of yeast repressor Sin3. *Cell* *80*, 767-776.
- Bartolomei, M.S., Zemel, S., and Tilghman, S.M. (1991). Parental imprinting of the mouse H19 gene. *Nature* *351*, 153-155.
- Batista, P.J., and Chang, H.Y. (2013). Long noncoding RNAs: cellular address codes in development and disease. *Cell* *152*, 1298-1307.

- Baudino, T.A., and Cleveland, J.L. (2001). The Max network gone mad. *Molecular and cellular biology* 21, 691-702.
- Beer, S., Zetterberg, A., Ihrle, R.A., McTaggart, R.A., Yang, Q., Bradon, N., Arvanitis, C., Attardi, L.D., Feng, S., Ruebner, B., *et al.* (2004). Developmental context determines latency of MYC-induced tumorigenesis. *PLoS biology* 2, e332.
- Benetatos, L., Vartholomatos, G., and Hatzimichael, E. (2014). Polycomb group proteins and MYC: the cancer connection. *Cellular and molecular life sciences : CMLS* 71, 257-269.
- Beroukhim, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M., *et al.* (2010). The landscape of somatic copy-number alteration across human cancers. *Nature* 463, 899-905.
- Blackwell, T.K., Huang, J., Ma, A., Kretzner, L., Alt, F.W., Eisenman, R.N., and Weintraub, H. (1993). Binding of myc proteins to canonical and noncanonical DNA sequences. *Molecular and cellular biology* 13, 5216-5224.
- Blackwell, T.K., Kretzner, L., Blackwood, E.M., Eisenman, R.N., and Weintraub, H. (1990). Sequence-specific DNA binding by the c-Myc protein. *Science* 250, 1149-1151.
- Blackwood, E.M., and Eisenman, R.N. (1991). Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. *Science* 251, 1211-1217.
- Blackwood, E.M., Lugo, T.G., Kretzner, L., King, M.W., Street, A.J., Witte, O.N., and Eisenman, R.N. (1994). Functional analysis of the AUG- and CUG-initiated forms of the c-Myc protein. *Molecular biology of the cell* 5, 597-609.
- Bouchard, C., Dittrich, O., Kiermaier, A., Dohmann, K., Menkel, A., Eilers, M., and Luscher, B. (2001). Regulation of cyclin D2 gene expression by the Myc/Max/Mad network: Myc-dependent TRRAP recruitment and histone acetylation at the cyclin D2 promoter. *Genes & development* 15, 2042-2047.
- Bravo, R. (1990). Genes induced during the G0/G1 transition in mouse fibroblasts. *Seminars in cancer biology* 1, 37-46.
- Brodeur, G.M., Seeger, R.C., Schwab, M., Varmus, H.E., and Bishop, J.M. (1984). Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage. *Science* 224, 1121-1124.
- Brooks, T.A., and Hurley, L.H. (2010). Targeting MYC Expression through G-Quadruplexes. *Genes & cancer* 1, 641-649.

Brown, C.J., Ballabio, A., Rupert, J.L., Lafreniere, R.G., Grompe, M., Tonlorenzi, R., and Willard, H.F. (1991). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* 349, 38-44.

Brown, C.J., Hendrich, B.D., Rupert, J.L., Lafreniere, R.G., Xing, Y., Lawrence, J., and Willard, H.F. (1992). The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* 71, 527-542.

Budczies, J., Klauschen, F., Sinn, B.V., Gyorffy, B., Schmitt, W.D., Darb-Esfahani, S., and Denkert, C. (2012). Cutoff Finder: a comprehensive and straightforward Web application enabling rapid biomarker cutoff optimization. *PLoS one* 7, e51862.

Busch, A., Richter, A.S., and Backofen, R. (2008). IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics* 24, 2849-2856.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes & development* 25, 1915-1927.

Calvet, J.P., and Pederson, T. (1979). Heterogeneous nuclear RNA double-stranded regions probed in living HeLa cells by crosslinking with the psoralen derivative aminomethyltrioxsalen. *Proceedings of the National Academy of Sciences of the United States of America* 76, 755-759.

Cappellen, D., Schlange, T., Bauer, M., Maurer, F., and Hynes, N.E. (2007). Novel c-MYC target genes mediate differential effects on cell proliferation and migration. *EMBO reports* 8, 70-76.

Carpenter, S., Aiello, D., Atianand, M.K., Ricci, E.P., Gandhi, P., Hall, L.L., Byron, M., Monks, B., Henry-Bezy, M., Lawrence, J.B., *et al.* (2013). A long noncoding RNA mediates both activation and repression of immune response genes. *Science* 341, 789-792.

Chang, T.C., Yu, D., Lee, Y.S., Wentzel, E.A., Arking, D.E., West, K.M., Dang, C.V., Thomas-Tikhonenko, A., and Mendell, J.T. (2008). Widespread microRNA repression by Myc contributes to tumorigenesis. *Nature genetics* 40, 43-50.

Chesi, M., Robbiani, D.F., Sebag, M., Chng, W.J., Affer, M., Tiedemann, R., Valdez, R., Palmer, S.E., Haas, S.S., Stewart, A.K., *et al.* (2008). AID-dependent activation of a MYC transgene induces multiple myeloma in a conditional mouse model of post-germinal center malignancies. *Cancer cell* 13, 167-180.

Cole, M.D., and McMahon, S.B. (1999). The Myc oncoprotein: a critical evaluation of transactivation and target gene regulation. *Oncogene* 18, 2916-2924.

Cowling, V.H., Chandriani, S., Whitfield, M.L., and Cole, M.D. (2006). A conserved Myc protein domain, MBIV, regulates DNA binding, apoptosis, transformation, and G2 arrest. *Molecular and cellular biology* 26, 4226-4239.

Dalla-Favera, R., Bregni, M., Erikson, J., Patterson, D., Gallo, R.C., and Croce, C.M. (1982). Human c-myc onc gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells. *Proceedings of the National Academy of Sciences of the United States of America* 79, 7824-7827.

Dang, C.V. (1999). c-Myc target genes involved in cell growth, apoptosis, and metabolism. *Molecular and cellular biology* 19, 1-11.

Dang, C.V. (2012). MYC on the path to cancer. *Cell* 149, 22-35.

Dang, C.V., and Lee, W.M. (1988). Identification of the human c-myc protein nuclear translocation signal. *Molecular and cellular biology* 8, 4048-4054.

Dang, C.V., McGuire, M., Buckmire, M., and Lee, W.M. (1989). Involvement of the 'leucine zipper' region in the oligomerization and transforming activity of human c-myc protein. *Nature* 337, 664-666.

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., *et al.* (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome research* 22, 1775-1789.

Dews, M., Tan, G.S., Hultine, S., Raman, P., Choi, J., Duperret, E.K., Lawler, J., Bass, A., and Thomas-Tikhonenko, A. (2014). Masking epistasis between MYC and TGF-beta pathways in antiangiogenesis-mediated colon cancer suppression. *Journal of the National Cancer Institute* 106, dju043.

Di Ruscio, A., Ebralidze, A.K., Benoukraf, T., Amabile, G., Goff, L.A., Terragni, J., Figueroa, M.E., De Figueiredo Pontes, L.L., Alberich-Jorda, M., Zhang, P., *et al.* (2013). DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* 503, 371-376.

Dieterich, C., and Stadler, P.F. (2013). Computational biology of RNA interactions. *Wiley interdisciplinary reviews RNA* 4, 107-120.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., *et al.* (2012). Landscape of transcription in human cells. *Nature* 489, 101-108.



Duesberg, P.H., and Vogt, P.K. (1979). Avian acute leukemia viruses MC29 and MH2 share specific RNA sequences: evidence for a second class of transforming genes. *Proceedings of the National Academy of Sciences of the United States of America* 76, 1633-1637.

Eeles, R.A., Kote-Jarai, Z., Giles, G.G., Olama, A.A., Guy, M., Jugurnauth, S.K., Mulholland, S., Leongamornlert, D.A., Edwards, S.M., Morrison, J., *et al.* (2008). Multiple newly identified loci associated with prostate cancer susceptibility. *Nature genetics* 40, 316-321.

Eeles, R.A., Olama, A.A., Benlloch, S., Saunders, E.J., Leongamornlert, D.A., Tymrakiewicz, M., Ghossaini, M., Luccarini, C., Dennis, J., Jugurnauth-Little, S., *et al.* (2013). Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nature genetics* 45, 385-391, 391e381-382.

el-Deiry, W.S., Tokino, T., Velculescu, V.E., Levy, D.B., Parsons, R., Trent, J.M., Lin, D., Mercer, W.E., Kinzler, K.W., and Vogelstein, B. (1993). WAF1, a potential mediator of p53 tumor suppression. *Cell* 75, 817-825.

Elkon, R., Zeller, K.I., Linhart, C., Dang, C.V., Shamir, R., and Shiloh, Y. (2004). In silico identification of transcriptional regulators associated with c-Myc. *Nucleic acids research* 32, 4955-4961.

Engreitz, J.M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., Kadri, S., Xing, J., Goren, A., Lander, E.S., *et al.* (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* 341, 1237973.

Felton-Edkins, Z.A., Kenneth, N.S., Brown, T.R., Daly, N.L., Gomez-Roman, N., Grandori, C., Eisenman, R.N., and White, R.J. (2003). Direct regulation of RNA polymerase III transcription by RB, p53 and c-Myc. *Cell Cycle* 2, 181-184.

Feng, J., Bi, C., Clark, B.S., Mady, R., Shah, P., and Kohtz, J.D. (2006). The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. *Genes & development* 20, 1470-1484.

Follis, A.V., Hammoudeh, D.I., Daab, A.T., and Metallo, S.J. (2009). Small-molecule perturbation of competing interactions between c-Myc and Max. *Bioorganic & medicinal chemistry letters* 19, 807-810.

Gao, P., Zhang, H., Dinavahi, R., Li, F., Xiang, Y., Raman, V., Bhujwala, Z.M., Felsher, D.W., Cheng, L., Pevsner, J., *et al.* (2007). HIF-dependent antitumorigenic effect of antioxidants in vivo. *Cancer cell* 12, 230-238.

Gartel, A.L., and Radhakrishnan, S.K. (2005). Lost in transcription: p21 repression, mechanisms, and consequences. *Cancer research* 65, 3980-3985.

Gartel, A.L., Ye, X., Goufman, E., Shianov, P., Hay, N., Najmabadi, F., and Tyner, A.L. (2001). Myc represses the p21(WAF1/CIP1) promoter and interacts with Sp1/Sp3. *Proceedings of the National Academy of Sciences of the United States of America* 98, 4510-4515.

Gomez-Roman, N., Grandori, C., Eisenman, R.N., and White, R.J. (2003). Direct activation of RNA polymerase III transcription by c-Myc. *Nature* 421, 290-294.

Gomez, J.A., Wapinski, O.L., Yang, Y.W., Bureau, J.F., Gopinath, S., Monack, D.M., Chang, H.Y., Brahic, M., and Kirkegaard, K. (2013). The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon-gamma locus. *Cell* 152, 743-754.

Grandori, C., and Eisenman, R.N. (1997). Myc target genes. *Trends in biochemical sciences* 22, 177-181.

Grandori, C., Gomez-Roman, N., Felton-Edkins, Z.A., Ngouenet, C., Galloway, D.A., Eisenman, R.N., and White, R.J. (2005). c-Myc binds to human ribosomal DNA and stimulates transcription of rRNA genes by RNA polymerase I. *Nature cell biology* 7, 311-318.

Grinberg, A.V., Hu, C.D., and Kerppola, T.K. (2004). Visualization of Myc/Max/Mad family dimers and the competition for dimerization in living cells. *Molecular and cellular biology* 24, 4294-4308.

Grote, P., Wittler, L., Hendrix, D., Koch, F., Wahrlich, S., Beisaw, A., Macura, K., Blass, G., Kellis, M., Werber, M., *et al.* (2013). The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Developmental cell* 24, 206-214.

Guo, Q.M., Malek, R.L., Kim, S., Chiao, C., He, M., Ruffy, M., Sanka, K., Lee, N.H., Dang, C.V., and Liu, E.T. (2000). Identification of c-myc responsive genes using rat cDNA microarray. *Cancer research* 60, 5922-5928.

Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M.C., Hung, T., Argani, P., Rinn, J.L., *et al.* (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464, 1071-1076.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., *et al.* (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458, 223-227.

Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G., Young, G., Lucas, A.B., Ach, R., Bruhn, L., *et al.* (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* 477, 295-300.

Guttman, M., and Rinn, J.L. (2012). Modular regulatory principles of large non-coding RNAs. *Nature* 482, 339-346.

Haggerty, T.J., Zeller, K.I., Osthus, R.C., Wonsey, D.R., and Dang, C.V. (2003). A strategy for identifying transcription factor binding sites reveals two classes of genomic c-Myc target sites. *Proceedings of the National Academy of Sciences of the United States of America* 100, 5313-5318.

Hann, S.R., Sloan-Brown, K., and Spotts, G.D. (1992). Translational activation of the non-AUG-initiated c-myc 1 protein at high cell densities due to methionine deprivation. *Genes & development* 6, 1229-1240.

Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K., and Kjems, J. (2013). Natural RNA circles function as efficient microRNA sponges. *Nature* 495, 384-388.

Harper, J.W., Adami, G.R., Wei, N., Keyomarsi, K., and Elledge, S.J. (1993). The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. *Cell* 75, 805-816.

Hazelett, D.J., Rhie, S.K., Gaddis, M., Yan, C., Lakeland, D.L., Coetzee, S.G., Henderson, B.E., Noushmehr, H., Cozen, W., Kote-Jarai, Z., *et al.* (2014). Comprehensive functional annotation of 77 prostate cancer risk loci. *PLoS genetics* 10, e1004102.

He, Y., Vogelstein, B., Velculescu, V.E., Papadopoulos, N., and Kinzler, K.W. (2008). The antisense transcriptomes of human cells. *Science* 322, 1855-1857.

Heasman, J. (2002). Morpholino oligos: making sense of antisense? *Developmental biology* 243, 209-214.

Hermeking, H., Rago, C., Schuhmacher, M., Li, Q., Barrett, J.F., Obaya, A.J., O'Connell, B.C., Mateyak, M.K., Tam, W., Kohlhuber, F., *et al.* (2000). Identification of CDK4 as a target of c-MYC. *Proceedings of the National Academy of Sciences of the United States of America* 97, 2229-2234.

Hu, J., Banerjee, A., and Goss, D.J. (2005). Assembly of b/HLH/z proteins c-Myc, Max, and Mad1 with cognate DNA: importance of protein-protein and protein-DNA interactions. *Biochemistry* 44, 11855-11863.

Hu, S.S., Lai, M.M., and Vogt, P.K. (1979). Genome of avian myelocytomatosis virus MC29: analysis by heteroduplex mapping. *Proceedings of the National Academy of Sciences of the United States of America* 76, 1265-1268.

Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M.J., Kenzelmann-Broz, D., Khalil, A.M., Zuk, O., Amit, I., Rabani, M., *et al.* (2010). A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* 142, 409-419.

Hung, T., Wang, Y., Lin, M.F., Koegel, A.K., Kotake, Y., Grant, G.D., Horlings, H.M., Shah, N., Umbricht, C., Wang, P., *et al.* (2011). Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nature genetics* 43, 621-629.

Hurley, L.H., Von Hoff, D.D., Siddiqui-Jain, A., and Yang, D. (2006). Drug targeting of the c-MYC promoter to repress gene expression via a G-quadruplex silencer element. *Seminars in oncology* 33, 498-512.

Iritani, B.M., and Eisenman, R.N. (1999). c-Myc enhances protein synthesis and cell size during B lymphocyte development. *Proceedings of the National Academy of Sciences of the United States of America* 96, 13180-13185.

Jeon, Y., and Lee, J.T. (2011). YY1 tethers Xist RNA to the inactive X nucleation center. *Cell* 146, 119-133.

Ji, P., Diederichs, S., Wang, W., Boing, S., Metzger, R., Schneider, P.M., Tidow, N., Brandt, B., Buerger, H., Bulk, E., *et al.* (2003). MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene* 22, 8031-8041.

Jia, H., Osak, M., Bogu, G.K., Stanton, L.W., Johnson, R., and Lipovich, L. (2010). Genome-wide computational identification and manual annotation of human long noncoding RNA genes. *RNA* 16, 1478-1487.

Kapranov, P., Cawley, S.E., Drenkow, J., Bekiranov, S., Strausberg, R.L., Fodor, S.P., and Gingeras, T.R. (2002). Large-scale transcriptional activity in chromosomes 21 and 22. *Science* 296, 916-919.

Karreth, F.A., and Pandolfi, P.P. (2013). ceRNA cross-talk in cancer: when ce-bling rivalries go awry. *Cancer discovery* 3, 1113-1121.

Kato, G.J., Barrett, J., Villa-Garcia, M., and Dang, C.V. (1990). An amino-terminal c-myc domain required for neoplastic transformation activates transcription. *Molecular and cellular biology* 10, 5914-5920.

- Kato, G.J., Lee, W.M., Chen, L.L., and Dang, C.V. (1992). Max: functional domains and interaction with c-Myc. *Genes & development* 6, 81-92.
- Kenneth, N.S., Ramsbottom, B.A., Gomez-Roman, N., Marshall, L., Cole, P.A., and White, R.J. (2007). TRRAP and GCN5 are used by c-Myc to activate RNA polymerase III transcription. *Proceedings of the National Academy of Sciences of the United States of America* 104, 14917-14922.
- Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B.E., van Oudenaarden, A., *et al.* (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proceedings of the National Academy of Sciences of the United States of America* 106, 11667-11672.
- Kim, J., Woo, A.J., Chu, J., Snow, J.W., Fujiwara, Y., Kim, C.G., Cantor, A.B., and Orkin, S.H. (2010a). A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. *Cell* 143, 313-324.
- Kim, J.W., Zeller, K.I., Wang, Y., Jegga, A.G., Aronow, B.J., O'Donnell, K.A., and Dang, C.V. (2004). Evaluation of myc E-box phylogenetic footprints in glycolytic genes by chromatin immunoprecipitation assays. *Molecular and cellular biology* 24, 5923-5936.
- Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., *et al.* (2010b). Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465, 182-187.
- Koh, C.M., Bieberich, C.J., Dang, C.V., Nelson, W.G., Yegnasubramanian, S., and De Marzo, A.M. (2010). MYC and Prostate Cancer. *Genes & cancer* 1, 617-628.
- Koh, C.M., Iwata, T., Zheng, Q., Bethel, C., Yegnasubramanian, S., and De Marzo, A.M. (2011). Myc enforces overexpression of EZH2 in early prostatic neoplasia via transcriptional and post-transcriptional mechanisms. *Oncotarget* 2, 669-683.
- Kohl, N.E., Gee, C.E., and Alt, F.W. (1984). Activated expression of the N-myc gene in human neuroblastomas and related tumors. *Science* 226, 1335-1337.
- Konig, H., Matter, N., Bader, R., Thiele, W., and Muller, F. (2007). Splicing segregation: the minor spliceosome acts outside the nucleus and controls cell proliferation. *Cell* 131, 718-729.
- Kotake, Y., Nakagawa, T., Kitagawa, K., Suzuki, S., Liu, N., Kitagawa, M., and Xiong, Y. (2011). Long non-coding RNA ANRIL is required for the PRC2

recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene* 30, 1956-1962.

Kretz, M., Siprashvili, Z., Chu, C., Webster, D.E., Zehnder, A., Qu, K., Lee, C.S., Flockhart, R.J., Groff, A.F., Chow, J., *et al.* (2013). Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nature* 493, 231-235.

Kretz, M., Webster, D.E., Flockhart, R.J., Lee, C.S., Zehnder, A., Lopez-Pajares, V., Qu, K., Zheng, G.X., Chow, J., Kim, G.E., *et al.* (2012). Suppression of progenitor differentiation requires the long noncoding RNA ANCR. *Genes & development* 26, 338-343.

Kretzner, L., Blackwood, E.M., and Eisenman, R.N. (1992). Myc and Max proteins possess distinct transcriptional activities. *Nature* 359, 426-429.

Laherty, C.D., Yang, W.M., Sun, J.M., Davie, J.R., Seto, E., and Eisenman, R.N. (1997). Histone deacetylases associated with the mSin3 corepressor mediate mad transcriptional repression. *Cell* 89, 349-356.

Lai, F., Orom, U.A., Cesaroni, M., Beringer, M., Taatjes, D.J., Blobel, G.A., and Shiekhata, R. (2013). Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* 494, 497-501.

Leder, A., Pattengale, P.K., Kuo, A., Stewart, T.A., and Leder, P. (1986). Consequences of widespread deregulation of the c-myc gene in transgenic mice: multiple neoplasms and normal development. *Cell* 45, 485-495.

Lee, J.T. (2012). Epigenetic regulation by long noncoding RNAs. *Science* 338, 1435-1439.

Lee, T.C., Li, L., Philipson, L., and Ziff, E.B. (1997). Myc represses transcription of the growth arrest gene gas1. *Proceedings of the National Academy of Sciences of the United States of America* 94, 12886-12891.

Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L., and Steitz, J.A. (1980). Are snRNPs involved in splicing? *Nature* 283, 220-224.

Levens, D. (2010). You Don't Muck with MYC. *Genes & cancer* 1, 547-554.

Li, C., and Wong, W.H. (2001). Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proceedings of the National Academy of Sciences of the United States of America* 98, 31-36.

Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X., *et al.* (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* 498, 516-520.

- Li, Z., Van Calcar, S., Qu, C., Cavenee, W.K., Zhang, M.Q., and Ren, B. (2003). A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells. *Proceedings of the National Academy of Sciences of the United States of America* 100, 8164-8169.
- Liao, D.J., and Dickson, R.B. (2000). c-Myc in breast cancer. *Endocrine-related cancer* 7, 143-164.
- Lieberman, J., Slack, F., Pandolfi, P.P., Chinnaiyan, A., Agami, R., and Mendell, J.T. (2013). Noncoding RNAs and cancer. *Cell* 153, 9-10.
- Lin, C.Y., Loven, J., Rahl, P.B., Paranal, R.M., Burge, C.B., Bradner, J.E., Lee, T.I., and Young, R.A. (2012). Transcriptional amplification in tumor cells with elevated c-Myc. *Cell* 151, 56-67.
- Ling, H., Spizzo, R., Atlasi, Y., Nicoloso, M., Shimizu, M., Redis, R.S., Nishida, N., Gafa, R., Song, J., Guo, Z., *et al.* (2013). CCAT2, a novel noncoding RNA mapping to 8q24, underlies metastatic progression and chromosomal instability in colon cancer. *Genome research* 23, 1446-1461.
- Liu, X., Tesfai, J., Evrard, Y.A., Dent, S.Y., and Martinez, E. (2003). c-Myc transformation domain recruits the human STAGA complex and requires TRRAP and GCN5 acetylase activity for transcription activation. *The Journal of biological chemistry* 278, 20405-20412.
- Loewer, S., Cabili, M.N., Guttman, M., Loh, Y.H., Thomas, K., Park, I.H., Garber, M., Curran, M., Onder, T., Agarwal, S., *et al.* (2010). Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nature genetics* 42, 1113-1117.
- Maeda, N., Kasukawa, T., Oyama, R., Gough, J., Frith, M., Engstrom, P.G., Lenhard, B., Aturaliya, R.N., Batalov, S., Beisel, K.W., *et al.* (2006). Transcript annotation in FANTOM3: mouse gene catalog based on physical cDNAs. *PLoS genetics* 2, e62.
- Marhin, W.W., Chen, S., Facchini, L.M., Fornace, A.J., Jr., and Penn, L.Z. (1997). Myc represses the growth arrest gene gadd45. *Oncogene* 14, 2825-2834.
- Maris, J.M. (2010). Recent advances in neuroblastoma. *The New England journal of medicine* 362, 2202-2211.
- Martianov, I., Ramadass, A., Serra Barros, A., Chow, N., and Akoulitchev, A. (2007). Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* 445, 666-670.

Mateyak, M.K., Obaya, A.J., and Sedivy, J.M. (1999). c-Myc regulates cyclin D-Cdk4 and -Cdk6 activity but affects cell cycle progression at multiple independent points. *Molecular and cellular biology* 19, 4672-4683.

Matter, N., and Konig, H. (2005). Targeted 'knockdown' of spliceosome function in mammalian cells. *Nucleic acids research* 33, e41.

McArthur, G.A., Laherty, C.D., Queva, C., Hurlin, P.J., Loo, L., James, L., Grandori, C., Gallant, P., Shiio, Y., Hokanson, W.C., *et al.* (1998). The Mad protein family links transcriptional repression to cell differentiation. *Cold Spring Harbor symposia on quantitative biology* 63, 423-433.

McEwan, I.J., Dahlman-Wright, K., Ford, J., and Wright, A.P. (1996). Functional interaction of the c-Myc transactivation domain with the TATA binding protein: evidence for an induced fit model of transactivation domain folding. *Biochemistry* 35, 9584-9593.

McMahon, S.B., Wood, M.A., and Cole, M.D. (2000). The essential cofactor TRRAP recruits the histone acetyltransferase hGCN5 to c-Myc. *Molecular and cellular biology* 20, 556-562.

Melo, C.A., Drost, J., Wijchers, P.J., van de Werken, H., de Wit, E., Oude Vrielink, J.A., Elkon, R., Melo, S.A., Leveille, N., Kalluri, R., *et al.* (2013). eRNAs are required for p53-dependent enhancer activity and gene transcription. *Molecular cell* 49, 524-535.

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M., *et al.* (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 495, 333-338.

Mercer, T.R., Dinger, M.E., and Mattick, J.S. (2009). Long non-coding RNAs: insights into functions. *Nature reviews Genetics* 10, 155-159.

Mercer, T.R., and Mattick, J.S. (2013). Structure and function of long noncoding RNAs in epigenetic regulation. *Nature structural & molecular biology* 20, 300-307.

Mestdagh, P., Bostrom, A.K., Impens, F., Fredlund, E., Van Peer, G., De Antonellis, P., von Stedingk, K., Ghesquiere, B., Schulte, S., Dewes, M., *et al.* (2010). The miR-17-92 microRNA cluster regulates multiple components of the TGF-beta pathway in neuroblastoma. *Molecular cell* 40, 762-773.

Meyer, N., and Penn, L.Z. (2008). Reflecting on 25 years with MYC. *Nature reviews Cancer* 8, 976-990.

Mustata, G., Follis, A.V., Hammoudeh, D.I., Metallo, S.J., Wang, H., Prochownik, E.V., Lazo, J.S., and Bahar, I. (2009). Discovery of novel Myc-Max heterodimer



disruptors with a three-dimensional pharmacophore model. *Journal of medicinal chemistry* 52, 1247-1250.

Nau, M.M., Brooks, B.J., Battey, J., Sausville, E., Gazdar, A.F., Kirsch, I.R., McBride, O.W., Bertness, V., Hollis, G.F., and Minna, J.D. (1985). L-myc, a new myc-related gene amplified and expressed in human small cell lung cancer. *Nature* 318, 69-73.

Nesbit, C.E., Tersak, J.M., and Prochownik, E.V. (1999). MYC oncogenes and human neoplastic disease. *Oncogene* 18, 3004-3016.

Nie, Z., Hu, G., Wei, G., Cui, K., Yamane, A., Resch, W., Wang, R., Green, D.R., Tessarollo, L., Casellas, R., *et al.* (2012). c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells. *Cell* 151, 68-79.

O'Connell, B.C., Cheung, A.F., Simkevich, C.P., Tam, W., Ren, X., Mateyak, M.K., and Sedivy, J.M. (2003). A large scale genetic analysis of c-Myc-regulated gene expression patterns. *The Journal of biological chemistry* 278, 12563-12573.

O'Donnell, K.A., Wentzel, E.A., Zeller, K.I., Dang, C.V., and Mendell, J.T. (2005). c-Myc-regulated microRNAs modulate E2F1 expression. *Nature* 435, 839-843.

O'Donnell, K.A., Yu, D., Zeller, K.I., Kim, J.W., Racke, F., Thomas-Tikhonenko, A., and Dang, C.V. (2006). Activation of transferrin receptor 1 by c-Myc enhances cellular proliferation and tumorigenesis. *Molecular and cellular biology* 26, 2373-2386.

Ohno, M., Fukagawa, T., Lee, J.S., and Ikemura, T. (2002). Triplex-forming DNAs in the human interphase nucleus visualized in situ by polypurine/polypyrimidine DNA probes and antitriplex antibodies. *Chromosoma* 111, 201-213.

Orom, U.A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytznicki, M., Notredame, C., Huang, Q., *et al.* (2010). Long noncoding RNAs with enhancer-like function in human cells. *Cell* 143, 46-58.

Orom, U.A., and Shiekhattar, R. (2013). Long noncoding RNAs usher in a new era in the biology of enhancers. *Cell* 154, 1190-1193.

Oster, S.K., Ho, C.S., Soucie, E.L., and Penn, L.Z. (2002). The myc oncogene: Marvelously Complex. *Advances in cancer research* 84, 81-154.

Pajic, A., Spitkovsky, D., Christoph, B., Kempkes, B., Schuhmacher, M., Staeger, M.S., Brielmeier, M., Ellwart, J., Kohlhuber, F., Bornkamm, G.W., *et al.* (2000). Cell cycle activation by c-myc in a burkitt lymphoma model cell line. *International journal of cancer Journal international du cancer* 87, 787-793.

Pasmant, E., Laurendeau, I., Heron, D., Vidaud, M., Vidaud, D., and Bieche, I. (2007). Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF. *Cancer research* 67, 3963-3969.

Pelengaris, S., Littlewood, T., Khan, M., Elia, G., and Evan, G. (1999). Reversible activation of c-Myc in skin: induction of a complex neoplastic phenotype by a single oncogenic lesion. *Molecular cell* 3, 565-577.

Pineda-Lucena, A., and Arrowsmith, C.H. (2001). <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N resonance assignments and secondary structure of the c-Myc binding domain (MBD) and the SH3 domain of the tumor suppressor Bin1. *Journal of biomolecular NMR* 19, 191-192.

Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W.J., and Pandolfi, P.P. (2010). A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* 465, 1033-1038.

Prensner, J.R., and Chinnaiyan, A.M. (2011). The emergence of lncRNAs in cancer biology. *Cancer discovery* 1, 391-407.

Prensner, J.R., Iyer, M.K., Balbin, O.A., Dhanasekaran, S.M., Cao, Q., Brenner, J.C., Laxman, B., Asangani, I.A., Grasso, C.S., Kominsky, H.D., *et al.* (2011). Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nature biotechnology* 29, 742-749.

Rinn, J.L., Euskirchen, G., Bertone, P., Martone, R., Luscombe, N.M., Hartman, S., Harrison, P.M., Nelson, F.K., Miller, P., Gerstein, M., *et al.* (2003). The transcriptional activity of human Chromosome 22. *Genes & development* 17, 529-540.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., *et al.* (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129, 1311-1323.

Sabo, A., Kress, T.R., Pelizzola, M., de Pretis, S., Gorski, M.M., Tesi, A., Morelli, M.J., Bora, P., Doni, M., Verrecchia, A., *et al.* (2014). Selective transcriptional regulation by Myc in cellular growth control and lymphomagenesis. *Nature* 511, 488-492.

Schmidt, E.V. (1999). The role of c-myc in cellular growth control. *Oncogene* 18, 2988-2996.

Schneider, A., Peukert, K., Eilers, M., and Hanel, F. (1997). Association of Myc with the zinc-finger protein Miz-1 defines a novel pathway for gene regulation by Myc. *Current topics in microbiology and immunology* 224, 137-146.

Schreiber-Agus, N., Alland, L., Muhle, R., Goltz, J., Chen, K., Stevens, L., Stein, D., and DePinho, R.A. (1997). A biochemical and biological analysis of Myc superfamily interactions. *Current topics in microbiology and immunology* 224, 159-168.

Secombe, J., Pierce, S.B., and Eisenman, R.N. (2004). Myc: a weapon of mass destruction. *Cell* 117, 153-156.

Seoane, J., Le, H.V., and Massague, J. (2002). Myc suppression of the p21(Cip1) Cdk inhibitor influences the outcome of the p53 response to DNA damage. *Nature* 419, 729-734.

Seoane, J., Pouponnot, C., Staller, P., Schader, M., Eilers, M., and Massague, J. (2001). TGFbeta influences Myc, Miz-1 and Smad to control the CDK inhibitor p15INK4b. *Nature cell biology* 3, 400-408.

Sheiness, D., and Bishop, J.M. (1979). DNA and RNA from uninfected vertebrate cells contain nucleotide sequences related to the putative transforming gene of avian myelocytomatosis virus. *Journal of virology* 31, 514-521.

Shim, H., Chun, Y.S., Lewis, B.C., and Dang, C.V. (1998). A unique glucose-dependent apoptotic pathway induced by c-Myc. *Proceedings of the National Academy of Sciences of the United States of America* 95, 1511-1516.

Siegel, R., Ma, J., Zou, Z., and Jemal, A. (2014). Cancer statistics, 2014. *CA: a cancer journal for clinicians* 64, 9-29.

Simon, J.A., and Lange, C.A. (2008). Roles of the EZH2 histone methyltransferase in cancer epigenetics. *Mutation research* 647, 21-29.

Simon, M.D., Pinter, S.F., Fang, R., Sarma, K., Rutenberg-Schoenberg, M., Bowman, S.K., Kesner, B.A., Maier, V.K., Kingston, R.E., and Lee, J.T. (2013). High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature* 504, 465-469.

Singh, A.M., and Dalton, S. (2009). The cell cycle and Myc intersect with mechanisms that regulate pluripotency and reprogramming. *Cell stem cell* 5, 141-149.

Sparmann, A., and van Lohuizen, M. (2006). Polycomb silencers control cell fate, development and cancer. *Nature reviews Cancer* 6, 846-856.

- Spotts, G.D., Patel, S.V., Xiao, Q., and Hann, S.R. (1997). Identification of downstream-initiated c-Myc proteins which are dominant-negative inhibitors of transactivation by full-length c-Myc proteins. *Molecular and cellular biology* 17, 1459-1468.
- Staller, P., Peukert, K., Kiermaier, A., Seoane, J., Lukas, J., Karsunky, H., Moroy, T., Bartek, J., Massague, J., Hanel, F., *et al.* (2001). Repression of p15INK4b expression by Myc through association with Miz-1. *Nature cell biology* 3, 392-399.
- Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131, 861-872.
- Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126, 663-676.
- Taub, R., Kirsch, I., Morton, C., Lenoir, G., Swan, D., Tronick, S., Aaronson, S., and Leder, P. (1982). Translocation of the c-myc gene into the immunoglobulin heavy chain locus in human Burkitt lymphoma and murine plasmacytoma cells. *Proceedings of the National Academy of Sciences of the United States of America* 79, 7837-7841.
- Tay, Y., Kats, L., Salmena, L., Weiss, D., Tan, S.M., Ala, U., Karreth, F., Poliseno, L., Provero, P., Di Cunto, F., *et al.* (2011). Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *Cell* 147, 344-357.
- Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105-1111.
- Tripathi, V., Ellis, J.D., Shen, Z., Song, D.Y., Pan, Q., Watt, A.T., Freier, S.M., Bennett, C.F., Sharma, A., Bubulya, P.A., *et al.* (2010). The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Molecular cell* 39, 925-938.
- Tripathi, V., Shen, Z., Chakraborty, A., Giri, S., Freier, S.M., Wu, X., Zhang, Y., Gorospe, M., Prasanth, S.G., Lal, A., *et al.* (2013). Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB. *PLoS genetics* 9, e1003368.
- Tsai, M.C., Manor, O., Wan, Y., Mosammamarast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 329, 689-693.
- Ulitsky, I., and Bartel, D.P. (2013). lincRNAs: genomics, evolution, and mechanisms. *Cell* 154, 26-46.

- Ulitsky, I., Shkumatava, A., Jan, C.H., Sive, H., and Bartel, D.P. (2011). Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* 147, 1537-1550.
- Vennstrom, B., Sheiness, D., Zabielski, J., and Bishop, J.M. (1982). Isolation and characterization of c-myc, a cellular homolog of the oncogene (v-myc) of avian myelocytomatosis virus strain 29. *Journal of virology* 42, 773-779.
- Vervoorts, J., Luscher-Firzlaff, J.M., Rottmann, S., Lilischkis, R., Walsemann, G., Dohmann, K., Austen, M., and Luscher, B. (2003). Stimulation of c-MYC transcriptional activity and acetylation by recruitment of the cofactor CBP. *EMBO reports* 4, 484-490.
- Walz, S., Lorenzin, F., Morton, J., Wiese, K.E., von Eyss, B., Herold, S., Rycak, L., Dumay-Odelot, H., Karim, S., Bartkuhn, M., *et al.* (2014). Activation and repression by oncogenic MYC shape tumour-specific gene expression profiles. *Nature* 511, 483-487.
- Wang, D., Garcia-Bassets, I., Benner, C., Li, W., Su, X., Zhou, Y., Qiu, J., Liu, W., Kaikkonen, M.U., Ohgi, K.A., *et al.* (2011). Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* 474, 390-394.
- Wang, H., Mannava, S., Grachtchouk, V., Zhuang, D., Soengas, M.S., Gudkov, A.V., Prochownik, E.V., and Nikiforov, M.A. (2008a). c-Myc depletion inhibits proliferation of human tumor cells at various stages of the cell cycle. *Oncogene* 27, 1905-1915.
- Wang, X., Arai, S., Song, X., Reichart, D., Du, K., Pascual, G., Tempst, P., Rosenfeld, M.G., Glass, C.K., and Kurokawa, R. (2008b). Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature* 454, 126-130.
- Wright, P.R., Georg, J., Mann, M., Sorescu, D.A., Richter, A.S., Lott, S., Kleinkauf, R., Hess, W.R., and Backofen, R. (2014). CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. *Nucleic acids research* 42, W119-123.
- Wu, S., Cetinkaya, C., Munoz-Alonso, M.J., von der Lehr, N., Bahram, F., Beuger, V., Eilers, M., Leon, J., and Larsson, L.G. (2003). Myc represses differentiation-induced p21<sup>CIP1</sup> expression via Miz-1-dependent interaction with the p21 core promoter. *Oncogene* 22, 351-360.
- Xiong, Y., Hannon, G.J., Zhang, H., Casso, D., Kobayashi, R., and Beach, D. (1993). p21 is a universal inhibitor of cyclin kinases. *Nature* 366, 701-704.

Yang, F., Huo, X.S., Yuan, S.X., Zhang, L., Zhou, W.P., Wang, F., and Sun, S.H. (2013a). Repression of the long noncoding RNA-LET by histone deacetylase 3 contributes to hypoxia-mediated metastasis. *Molecular cell* 49, 1083-1096.

Yang, L., Lin, C., Jin, C., Yang, J.C., Tanasa, B., Li, W., Merkurjev, D., Ohgi, K.A., Meng, D., Zhang, J., *et al.* (2013b). lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. *Nature* 500, 598-602.

Yang, L., Lin, C., Liu, W., Zhang, J., Ohgi, K.A., Grinstead, J.D., Dorrestein, P.C., and Rosenfeld, M.G. (2011). ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. *Cell* 147, 773-788.

Yap, K.L., Li, S., Munoz-Cabello, A.M., Raguz, S., Zeng, L., Mujtaba, S., Gil, J., Walsh, M.J., and Zhou, M.M. (2010). Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Molecular cell* 38, 662-674.

Yeager, M., Orr, N., Hayes, R.B., Jacobs, K.B., Kraft, P., Wacholder, S., Minichiello, M.J., Fearnhead, P., Yu, K., Chatterjee, N., *et al.* (2007). Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nature genetics* 39, 645-649.

Yildirim, E., Kirby, J.E., Brown, D.E., Mercier, F.E., Sadreyev, R.I., Scadden, D.T., and Lee, J.T. (2013). Xist RNA is a potent suppressor of hematologic cancer in mice. *Cell* 152, 727-742.

Yoo-Warren, H., Pachnis, V., Ingram, R.S., and Tilghman, S.M. (1988). Two regulatory domains flank the mouse H19 gene. *Molecular and cellular biology* 8, 4707-4715.

Yuneva, M., Zamboni, N., Oefner, P., Sachidanandam, R., and Lazebnik, Y. (2007). Deficiency in glutamine but not glucose induces MYC-dependent apoptosis in human cells. *The Journal of cell biology* 178, 93-105.

Zeller, K.I., Haggerty, T.J., Barrett, J.F., Guo, Q., Wonsey, D.R., and Dang, C.V. (2001). Characterization of nucleophosmin (B23) as a Myc target by scanning chromatin immunoprecipitation. *The Journal of biological chemistry* 276, 48285-48291.

Zeller, K.I., Jegga, A.G., Aronow, B.J., O'Donnell, K.A., and Dang, C.V. (2003). An integrated database of genes responsive to the Myc oncogenic transcription factor: identification of direct genomic targets. *Genome biology* 4, R69.

Zeller, K.I., Zhao, X., Lee, C.W., Chiu, K.P., Yao, F., Yustein, J.T., Ooi, H.S., Orlov, Y.L., Shahab, A., Yong, H.C., *et al.* (2006). Global mapping of c-Myc

binding sites and target gene networks in human B cells. *Proceedings of the National Academy of Sciences of the United States of America* 103, 17834-17839.

Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J., and Lee, J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* 322, 750-756.

Zhao, X., Lwin, T., Zhang, X., Huang, A., Wang, J., Marquez, V.E., Chen-Kiang, S., Dalton, W.S., Sotomayor, E., and Tao, J. (2013). Disruption of the MYC-miRNA-EZH2 loop to suppress aggressive B-cell lymphoma survival and clonogenicity. *Leukemia* 27, 2341-2350.

Zhuang, Y., and Weiner, A.M. (1986). A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. *Cell* 46, 827-835.

# Curriculum Vitae for Ph.D. CANDIDATES

## The Johns Hopkins University School of Medicine

Yunqi Lu

December 2014

### Educational History

Ph.D. expected	2014	Graduate Program in Human Genetics Johns Hopkins University, School of Medicine Mentor: Chi V Dang, MD, PhD
B.S.	2008	Biology University of Washington GPA: 3.9/4.0
B.S.	2006	Biology Sichuan University GPA: 3.9/4.0

### Other Professional Experience

Teaching Assistant	2010	Molecular Biology & Genomics Course Johns Hopkins University, School of Medicine
Undergrad Researcher	2006-2008	Lab of Dr. Arnold Bendich, University of Washington
Undergrad Researcher	2004-2008	Lab of Dr. Song Mingjing Genetic Center of Chinese Academy of Medical Sciences, China

### Academic and other honors

2006-2008	Dean's List (every quarter) University of Washington
2004-2006	Principal Scholarships (Highest Honor) Sichuan University

### Publications

Blocking lactate export by inhibiting the Myc target MCT1 Disables glycolysis and glutathione synthesis. Doherty JR, Yang C, Scott KE, Cameron MD, Fallahi M, Li W, Hall MA, Amelio AL, Mishra JK, Li F, Tortosa M, Genau HM, Rounbehler RJ, **Lu Y**, Dang CV, Kumar KG, Butler AA, Bannister TD, Hooper AT, Unsal-Kacmaz K, Roush WR, Cleveland JL. *Cancer Res.* 2014 Feb 1;74(3):908-20. doi: 10.1158/0008-5472.CAN-13-2034. Epub 2013 Nov 27.



Huang CR, Schneider AM, **Lu Y**, Niranjana T, Shen P, Robinson MA, Steranka JP, Valle D, Civin CI, Wang T, Wheelan SJ, Ji H, Boeke JD, Burns KH: Mobile interspersed repeats are major structural variants in the human genome. **Cell**. 2010 Jun 25;141 (7):1171-82.

Xu L, Song M, Zhang F, Xu Y, Gao Q, Li C, **Lu Y**: Isolation and characterization of microsatellite markers in the striped hamster (*Cricetulus barabensis*). *Mol Ecol Resour*. 2008 Nov; 8(6):1500-2. doi: 10.1111/j.1755-0998.2008.02224.x.

### **Posters**

2013

University of Pennsylvania Cancer Research Symposium

2012

University of Pennsylvania Oncology Department Annual Retreat, New York

### **Service and Leadership**

2012-2013

Vice President, University of Pennsylvania Chinese Students and Scholars Association

2010-2011

Social Committee Chair, Johns Hopkins Chinese Students and Scholars Association