Regulation of transcription by MYC
- DNA binding and target genes -

Transkriptionelle Regulation durch MYC
- DNA-Bindung und Zielgene -

Doctoral thesis for a doctoral degree
at the Graduate School of Life Sciences,
Julius-Maximilians-Universität Würzburg,
Section Biomedicine

submitted by

Francesca Lorenzin

from Borgo Valsugana

Würzburg, 2016

Submitted on:

Members of the *Promotionskomitee*:

Chairperson: Prof. Dr. Manfred Gessler

Primary Supervisor: Prof. Dr. Martin Eilers

Second Supervisor: Prof. Dr. Jörg Vogel

Third Supervisor: Bruno Amati, PhD

Date of Public Defence:

Date of Receipt of Certificates:

Table of Contents

Summary

MYC is a transcription factor, whose expression is elevated or deregulated in many human cancers (up to 70%) and is often associated with aggressive and poorly differentiated tumors. Although MYC is extensively studied, discrepancies have emerged about how this transcription factor works. In primary lymphocytes, MYC promotes transcriptional amplification of virtually all genes with an open promoter, whereas in tumor cells MYC regulates specific sets of genes that have significant prognostic value. Furthermore, the set of target genes that distinguish MYC's physiological function from the pathological/oncogenic one, whether it exists or not, has not been fully understood yet.

In this study, it could be shown that MYC protein levels within a cell and promoter affinity (determined by E-box presence or interaction with other proteins) of target genes toward MYC are important factors that influence MYC activity. At low levels, MYC can amplify a certain transcriptional program, which includes high affinity binding sites, whereas at high levels MYC leads to the specific up- and down regulation of genes with low affinity. Moreover, the promoter affinity characterizes different sets of target genes which can be distinguished in the physiological or oncogenic MYC signatures.

MYC-mediated repression requires higher MYC levels than activation and formation of a complex with MIZ1 is necessary for inhibiting expression of a subset of MYC target genes.

Zusammenfassung

MYC ist ein Transkriptionsfaktor, dessen Expression in vielen humanen Tumoren (bis zu 70 %) erhöht oder dereguliert ist. Die Tumore, in denen viel MYC hergestellt wird, zeichnen sich durch einen geringen Differenzierungsgrad aus und verhalten sich sehr aggressiv. Obwohl das biologische Verhalten des MYC Proteins intensiv untersucht wurde, sind unterschiedliche Modelle, wie dieser Transkriptionsfaktor funktioniert, entwickelt worden. In primären Lymphozyten verstärkt MYC die Expression fast aller Gene mit offener Chromatinstruktur, während MYC in Tumorzellen spezifische Gengruppten reguliert, deren Expression mit der Prognose von Patienten korreliert. Es ist also unklar, ob sich die Zielgene der physiologischen Funktion von Myc von den oncogenen/pathophysiologischen Zielgenen unterscheidet und um welche Gene es sich bei letzteren handelt.

In dieser Arbeit konnte gezeigt werden, dass Expressionsniveau von MYC und unterschiedliche Promotoraffinitäten zu MYC (charakterisiert durch den Ebox-Gehalt und Interaktionen zu anderen Proteinen) wichtig für die Aktivität des MYC Proteins sind. So kann Myc bei niedrigen Konzentrationen ein bestimmtes transkriptionelles Programm amplifizieren, das sich aus hochaffinen Promotoren zusammensetzt. Bei hohen Konzentrationen hingegen führt MYC zur transkriptionellen Aktivierung und Repression bestimmter Zielgengruppen, die sich durch niedrige Affinität zu MYC auszeichnen. Somit ist die Promotoraffinität ein Parameter, der physiologische von oncogenen MYC Signaturen trennen kann. Darüberhinaus konnte gezeigt werden, dass MYC-vermittelte Repression höhere MYC Mengen benötigt, als MYC-vermittelte Transaktivierung und die Komplexbildung mit MIZ1 für die Repression einer Gruppe an MYC Zielgenen nötig ist.

Chapter 1:

Introduction

1.1 The proto-oncoprotein MYC

The *MYC* gene was first identified in the late 1970s as a viral oncogene promoting myelocytomatosis in chicken (*v-myc*) (Sheiness and Bishop, 1979). In subsequent years homologous proteins were found in vertebrates forming a protein family of 3 members: c-MYC (hereafter called MYC), N-MYC and L-MYC (Kohl et al., 1983; Nau et al., 1985; Vennstrom et al., 1982). MYC is evolutionary highly conserved and it can be found in *Drosophila melanogaster* (Gallant et al., 1996). MYC regulates several cellular functions mainly involved in cell growth and proliferation. Furthermore, it is necessary for embryonic development since murine embryos bearing homozygous MYC deletion die between day 8.5 and 9.5 (Davis et al., 1993).

1.1.1 Structure of the MYC protein

The three members of the MYC protein family share a significant structural homology. MYC is a 439 amino acids long protein and contains several domains that are important for its function and evolutionary conserve. The basic helix-loop-helix/leucine zipper (bHLH/LZ) domain is present at the C-terminal part of MYC and is needed for binding to DNA and for heterodimerization of MYC with its bHLH/LZ partner MAX (MYC-associated factor X) (Blackwood et al., 1992). Moreover this domain was shown to be responsible for the interaction with MIZ1 (MYC interacting zinc-finger protein 1) in a yeast two-hybrid screen

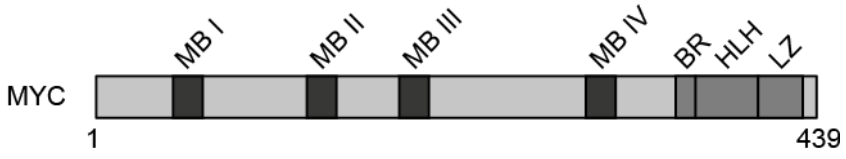


Figure 1. 1: Schematic diagram of the MYC protein. Human c-MYC is a 439 amino acids long protein that bears several conserved domains: MYC boxes I-IV (MB I-IV), a basic region (BR), a helix-loop-helix motif (HLH) and a leucine zipper (LZ).

(Peukert et al., 1997) and p300, an acetyltransferase with a broad range of action (Vervoorts et al., 2003).

The N-terminal region of MYC bears the transactivation domain (TAD) with two highly conserved elements, the so-called MYC boxes I and II. These domains are important for the transforming functions of MYC (Stone et al., 1987), for the interaction with other proteins and for the regulation of MYC stability. Indeed, MYC box I contains two residues, threonine 58 and serine 62, that are recognized by the ubiquitin ligase (FBXW7) upon phosphorylation, leading to proteasomal degradation of MYC (Sears et al., 2000). MYC box II was shown to serve as a binding platform for several proteins involved in chromatin modification. MYC, via MYC box II binds to TRRAP (Transformation/ Transcription Domain-Associated Protein) that in turn recruits proteins with histone acetyltransferase activity such as GCN5 (McMahon et al., 1998, 2000). It is necessary for the *in vivo* interaction with TIP48 and TIP49, which are part of chromatin remodeling complexes and have ATPase/helicase motifs (Wood et al., 2000) and to bind to SKP2, an ubiquitin ligase that beside signaling the MYC turnover is also a potent activator of its transcriptional activity (Kim et al., 2003).

In the central region of MYC other MYC boxes (IIIa, IIIb and IV) are present. They are all important for MYC transforming potential but their functions are less understood. It was shown that this central region is involved in the interaction with SMAD 2 and 3 (Feng et al., 2002), leading to the inhibition of CDKN2B expression probably also via interaction with MIZ1 (Herold et al., 2002). MYC box IIIa interacts with the histone deacetylase (HDAC) HDAC3 (Kurland and Tansey, 2008) and MYC box IV is important for binding to naked DNA (Cowling and Cole, 2006). More recently it was shown that MYC box IIIb directly binds to WDR5, a WD40-repeat protein found for example in H3K4 methyltransferases, driving the broad association of MYC to target genes (Thomas et al., 2015).

The nuclear localization signal (NLS) is localized close to MYC box IV.

1.1.2 MYC binding to chromatin

As a transcription factor MYC directly binds to DNA. The target site for MYC binding is an hexanucleotide sequence called E-box (Enhancer-box) (Blackwell et al., 1990; Blackwood et al., 1992). The canonical sequence for the E-box is CACGTG, but MYC can also bind E-boxes where the two central nucleotides are changed (CANNTG) (Blackwell et al., 1993). As shown by several studies, the *in vitro* affinity for MYC binding to canonical

E-boxes is about 2.5 fold higher compared to that for binding to non-canonical sequences (Hu et al., 2005) and about 200 fold higher compared to any other DNA sequence (Guo et al., 2014).

MYC preferentially binds in the promoter region of genes where histone modifications marking the open and accessible chromatin (e.g. H3K4me3 and H3K27Ac) are present (Chen et al., 2008; Guccione et al., 2006; Martinato et al., 2008; Zeller et al., 2006). Recently, genome wide studies for MYC binding performed using chromatin immunoprecipitation followed by sequencing (ChIP-seq) showed that many MYC peaks are present in regions far from the gene promoters. Part of these peaks reside in regions marked with histone modifications typical of enhancers (high H3K4me1, high H3K27Ac and low H3K4me3) arguing that MYC can bind to active enhancers, too (Lin et al., 2012; Sabò et al., 2014). Moreover, it was noticed that although MYC binding sites are enriched for E-boxes, many of them do not contain any E-box sequence. This observation is supported by other genome-wide studies (Seitz et al., 2011; Zeller et al., 2006) corroborating the idea that some other DNA elements or proteins bound to the DNA are required for MYC binding at these sites (Fernandez et al., 2003; Guo et al., 2014; Lin et al., 2012; Nie et al., 2012). Furthermore Uribesalgo and colleagues showed that the MYC-MAX complex cooperates with RARα (retinoic acid receptor-α) in the repression of genes required for differentiation in an E-box-independent manner (Uribesalgo et al., 2011).

Most of the studies focused on the binding of MYC at RNA polymerase (RNA Pol) II-transcribed genes, however MYC also binds to and regulates genes transcribed by the RNA polymerases I and III.

RNA Pol I transcribes genes encoding the ribosomal (r) RNAs, called rDNA. Poortinga and colleagues showed that MYC influences the expression of rDNA in NIH3T3 fibroblasts by regulating the expression of the upstream binding factor (UBF), which is essential for RNA Pol I transcription (Poortinga et al., 2004). Moreover, ChIP experiments showed that MYC binds to the E-box elements located at the promoters of rDNA in HeLa cells and MYC-induced P493 lymphocytes (Arabi et al., 2005; Grandori et al., 2005). In contrast to vertebrates, the *Drosophila* rDNA locus does not contain any canonical E-boxes, but dMyc is an important regulator of rRNA synthesis as well. dMyc induces transcription of genes encoding factors of the RNA Pol I machinery, thereby leading to an upregulation of the rRNA synthesis (Grewal et al., 2005). This indirect control of rDNA expression by dMyc suggests that the transcriptional control of RNA Pol I is a function of MYC acquired in vertebrates (Grewal et al., 2005).

RNA Pol III transcribes the tRNAs, the 5S rRNA and other small non-coding RNAs (ncRNAs). MYC was shown to be a potent activator of RNA Pol III transcribed-genes and interestingly, these genes do not contain E-boxes. Furthermore, regulation of these genes by MYC does not depend on heterodimerization with MAX as it was shown in *Drosophila melanogaster* (Steiger et al., 2008). Instead, MYC associates with TFIIIB, an essential RNA Pol III factor (Felton-Edkins et al., 2003; Gomez-Roman et al., 2003), and recruits GCN5 via interaction with TRAPP (Kenneth et al., 2007).

1.1.3 The functions of MYC

MYC is part of a network of proteins comprising MAX and the MAD protein family that bind to and regulate overlapping classes of genes. Both MYC and MAD heterodimerize with MAX via the bHLH-LZ region and both heterodimers bind to E-box sites competing with each other. MAX is an ubiquitously expressed protein (Blackwood et al., 1992), whereas expression of MYC and MAD is restricted to certain cellular stages such as proliferation and terminal differentiation, respectively (Dang et al., 2006). While the MYC/MAX complex is mainly acting as an activator of transcription of genes involved in cell growth and division, MAX/MAD dimers repress transcription of target genes via recruitment of the chromatin-modifying complex containing HDAC 1 and 2. Thus during differentiation a switch from MYC/MAX to MAD/MAX binding usually occurs at target genes (Ayer and Eisenman, 1993; Bouchard et al., 2001; Xu et al., 2001).

MYC, in complex with MAX, is the downstream effector of several mitogenic signaling pathways including WNT, NOTCH and receptor tyrosine kinases (e.g. EGFR, IGFR) that activate its expression leading to cell growth and proliferation. Accordingly, MYC was shown to bind to the promoter of genes encoding cyclin D1 and D2, CDK4, and cyclin B1 (Bouchard et al., 2001; Fernandez et al., 2003; Hermeking et al., 2000; Menssen and Hermeking, 2002). Moreover, via interaction with MIZ1, MYC represses expression of CDK inhibitors, such as *CDKN1A* and *CDKN2B*, and proteins involved in cell cycle arrest, such as GADD45 and GAS1 (Dang et al., 2006).

Since cell proliferation requires also an increase in the cellular mass, in parallel to regulate the expression of genes involved in the cell cycle control, MYC also activates several biosynthetic pathway providing ATP and the building blocks for the growing cells. As previously described, MYC activates rRNA and tRNA transcription by RNA Pol I and III (see 1.1.2). MYC regulates RNA Pol II-mediated transcription of spliceosome factors, structural ribosomal protein genes, factors for rRNA processing and ribosome export and

translation initiation factors for both CAP-dependent and independent translation (Koh et al., 2015; van Riggelen et al., 2010). By regulating all three polymerases, MYC orchestrates the balanced expression of factors necessary for protein biosynthesis.

A variety of studies have also linked MYC to the regulation of the cellular energy metabolism. The first evidence was the regulation by MYC of the lactate dehydrogenase A (LDH-A) gene expression (Shim et al., 1997). Subsequently it was shown that MYC regulates glucose uptake, glycolysis, mitochondrial biogenesis and function. Moreover, its ability to coordinately regulate the transcription of the mitochondrial metabolic network is required for rapid cell cycle entry (Morrish et al., 2008; Zhang et al., 2007). Cancer cells indeed, increase the glucose uptake in order to direct it to the glycolytic pathway to produce ATP and the building blocks needed for the growing cells. Accordingly, MYC also induces genes involved in glutamine metabolism: MYC suppresses expression of miRNAs that in turn downregulate expression of glutaminase (GLS) but it also transcriptionally induces expression of glutamine transporters. Glutamine is converted by GLS into glutamate which is oxidized in the TCA cycle providing the building blocks for macromolecular synthesis (Dang, 2013).

MYC indirectly regulates nucleotide biosynthesis by increasing glucose and glutamine uptake that are both needed for this purpose. Furthermore several enzymes involved in nucleotide metabolism are direct targets of MYC and depletion of MYC in melanoma cells decreases deoxyribonucleoside triphosphates (dNTPs) levels and inhibites proliferation (Mannava et al., 2008).

Several MYC target genes encode central enzymes for fatty acid metabolism. For example, *FASN* and *SCD* are both highly responsive to MYC and the encoded proteins catalyze the addition of the two carbon atoms of the acetyl-CoA to the growing fatty acid chain and the introduction of double bonds in long fatty acid chains, respectively (Zeller et al., 2003).

Besides controlling several biosynthetic pathways leading to cell growth and proliferation, MYC regulates other biological activities such as cell adhesion, angiogenesis and apoptosis. MYC represses collagen and integrin genes and enhances anchorage-independent growth that is a hallmark of cancer cells (Barr et al., 1998; Frye et al., 2003; Gebhardt et al., 2006; Yang et al., 1991). It stimulates angiogenesis, both in embryos and in tumors, via regulation of the VEGF (Vascular endothelial growth factor) release in the microenvironment (Baudino et al., 2002) (Shchors et al., 2006). At supra-physiological levels, MYC induces apoptosis as an intrinsic tumor suppression mechanism (Murphy et al., 2009).

Regulating such a plethora of cellular processes, MYC is able to drive quiescent cells into the cell cycle. Additionally, deregulation of MYC expression can result in uncontrolled cell proliferation and tumor development. Therefore, it is not surprising to find activation of the MYC family genes in a wide range of hematological and solid tumors. The most common events driving oncogenic expression of *MYC* are gene amplification in solid tumors and translocation in lymphoma and leukemia, but also point mutations, enhanced translation and protein stability (Vita and Henriksson, 2006) (Dang, 1999).

Its deregulation in about 70% of tumors (Dang, 2012) renders MYC an interesting target for tumor therapy. Unfortunately, MYC was described to be untargetable by any available drugs given its large surface of contact with DNA and its partner protein MAX (Nair and Burley, 2003). Important steps forward in tumor therapy could arise from the understanding of which pathways/proteins act upstream of MYC and regulates its activity. Additionally, identifying its target genes in order to find druggable targets could expand the therapeutic window for treating MYC-driven tumors.

1.2 MIZ1 – The MYC-interacting zinc finger protein 1

1.2.1 Structure of MIZ1

MIZ1 (MYC-interacting zinc finger protein 1) was first identified as a MYC-interacting protein in a yeast two-hybrid screen (Peukert et al., 1997). It is a member of the BTB/POZ zinc finger transcription factors. At the N-terminus a BTB/POZ domain is located that acts as a hydrophobic surface mediating the di- and tetramerization among MIZ1molecules or the interaction with other proteins (Bardwell and Treisman, 1994; Stead et al., 2007). The C-terminal part of MIZ1 bears the DNA binding domain formed by 12 consecutive Cys_2His_2 zinc fingers and one 13^{th} zinc finger separated by an alpha helix region of 80amino acids. The latter is needed for the interaction with MYC (Peukert et al., 1997).

In contrast to other POZ proteins MIZ1 is soluble and mainly found in the nucleoplasm.

Figure 1. 2: Schematic diagram of the MIZ1 protein. MIZ1 consists of 803 amino acids and bears an amino terminal BTB/POZ domain and 13 zinc fingers (ZF)

1.2.2 MIZ1 functions

In mice MIZ1 is an essential protein during development since homozygous MIZ1 null embryos are severely retarded in early embryonic development and die around day 7.5 (Adhikary et al., 2003). Furthermore MIZ1 has a central role in regulating cell cycle exit during hair growth cycle and epithelial morphogenesis (Gebhardt et al., 2007).

MIZ1 is a negative regulator of cell proliferation. It is involved in cell cycle regulation and TGF-β signaling by binding to the core promoter of genes encoding cyclin-dependent kinase inhibitors such as $p15^{INK4b}$ and $p21^{CIP1}$ (Seoane et al., 2001; Staller et al., 2001).

MIZ1 activity is coordinated with cell growth via a ribosomal protein L23-nucleophosmin circuit (both proteins are direct targets of MYC). RPL23 can retain nucleophosmin in the nucleolus avoiding its binding to MIZ1 and thus inhibiting MIZ1 activation. Therefore high levels of L23, that are associated with efficient translation, restrain the G1 arrest induced by MIZ1 (Wanzel et al., 2008).

MIZ1 interacts with BCL6 forming a repressive complex that inhibits expression of *CDKN1A* ($p21^{CIP1}$) and *BCL2* counteracting p53-induced cell cycle arrest and thus controlling proliferation and survival of germinal center B-cells (Phan et al., 2005).

ChIP-seq analysis in neural progenitor cells, which express no or few MYC, showed that MIZ1 binds to the core promoter of about 140 genes that are enriched for regulators of autophagy and proteins involved in vesicular traffic that are required for autophagy. MIZ1 activates the expression of these genes by binding to a non-palindromic DNA sequence present in their core promoters. Moreover, deletion of the protein in the central nervous system leads to a cerebellar neurodegenerative phenotype similar to that obtained with knockout of ATG5, an important mediator of autophagy (Wolf et al., 2013).

MIZ1 interacts with TopBP1 (topoisomerase II binding protein 1), an essential activator of ATR kinase. MIZ1 is required for recruitment of TopBP1 to the chromatin protecting it from proteasomal degradation in unstressed cells and in the early response to UV irradiation. MIZ1 in complex with TopBP1 is implicated in the ATR-dependent signal transduction and constitutes a reservoir from which TopBP1 is recruited to stalled replication forks (Herold et al., 2008). Moreover, MIZ1/TopBP1 is an inactive complex and release of TopBP1 is needed for expression of CDK inhibitors activated by MIZ1 (Herold et al., 2002).

1.3 Regulation of gene expression by MYC and MIZ1

Several genome wide studies have shown that MYC has a broad range of binding and weakly control expression of a large number of genes. As previously described it regulates many cellular processes from proliferation to metabolism to apoptosis by both upregulating and repressing, typically less than two fold, its target genes (Cole and Cowling, 2008). In contrast, MIZ1 strongly regulates expression of fewer target genes involved in cell cycle regulation and autophagy (Seoane et al., 2001; Staller et al., 2001; Wolf et al., 2013).

1.3.1 Transcriptional activation by MYC

Binding of MYC to activated target genes in general leads to the recruitment of co-activators. These include TRRAP (Bouchard et al., 2001) that in turn recruits the histone acetyltransferases (HAT) GCN5 (McMahon et al., 1998, 2000) and TIP60 (Frank et al., 2003) which modify histones 3 and 4, respectively, and p400 E1A-binding protein which do not have HAT activity (Fuchs et al., 2001), the acetyltransferases p300 and CBP (CREB binding protein) (Faiola et al., 2005; Vervoorts et al., 2003) and the SWI/SNF subunit BAF47/SNF5 (Cheng et al., 1999). Moreover, USP22, that deubiquitylates H2B and is part of the SAGA complex, is recruited to MYC target genes and required for their activation (Zhang et al., 2008a). The kinase PIM1 is also recruited by MYC to target genes and is important for the activation of about 20% of MYC target genes through the phosphorylation of serine 10 of histone 3 (Zippo et al., 2007).

The recruitment of these co-regulators is thought to mediate histone modifications and/or remodel the chromatin leading to a more open structure and nucleosome instability, thus enabling transcription of target genes. Indeed, manipulation of MYC levels in the cell leads to changes in histone modification distribution that in turn influences RNA Pol transcription (Guccione et al., 2006; Knoepfler et al., 2006; Martinato et al., 2008).

Besides controlling the chromatin state of target genes, MYC also activates transcription via mechanisms that do not involve chromatin modifications. The Mediator complex interacts with the N-terminus of MYC *in vitro* (Adhikary and Eilers, 2005) and is recruited to MYC target genes *in vivo* (Bouchard et al., 2004). Mediator is a huge complex of proteins that provides a large surface for protein-protein interactions important for the basal transcription and for facilitating the communication between transcription factors bound to regulatory elements and the pre-initiation complex. It is also crucial for the

organization of the genomic DNA in topological domains that enables coordinated regulation of cellular transcription (Allen and Taatjes, 2015).

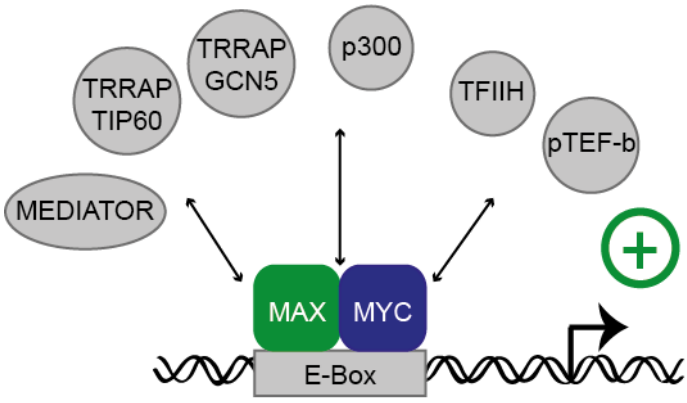


Figure 1. 3: Schematic diagram of transcriptional activation by MYC. MYC activates transcription by interacting and recruiting several coactivator complexes and factors modifying directly RNA Pol state. Modified from Adhikary and Eilers, 2005.

TFIIH is also recruited to target genes by MYC (Bouchard et al., 2004; Cowling and Cole, 2007). It possesses DNA-dependent ATPase, DNA helicase and protein kinase activities and it is involved in the formation of the pre-inititation complex (PIC) with RNA Pol II and other transcription factors. RBP1, the major subunit of RNA Pol II, has a C-terminal domain (CTD) that consists of conserved heptapeptide (YSPTSPS) repeats, 52 in humans. The amino acids in these repeats are subject of several posttranslational modifications that are specific for different steps of the transcription cycle and function as a code for the recruitment and binding of complexes involved in transcription, RNA processing and export, as well as chromatin remodeling (Heidemann et al., 2013; Zhang et al., 2012a). When RNA Pol II is recruited to the DNA in the PIC, the CTD is hypophosphorylated. Transcription initiation is associated with the phosphorylation of serine 5 of the CTD by TFIIH, via its cyclin-dependent kinase 7 subunit (CDK7). This leads to the recruitment of the histone methyltransferase SET1, that trimethylates histone 3 at lysine 4 (H3K4me3), a tag for transcriptional activation, and of the 5′ end capping machinery to the nascent mRNA. CDK7 also phosphorylates serine 7 of the CTD but the role of this modification remains not well understood (Heidemann et al., 2013). In most cases, after transcription initiation, RNA Pol II goes through a pausing phase that involves association with the pausing complexes DSIF (DRB-sensitivity-inducing factor) and NELF (Negative elongation factor) (Adelman and Lis, 2012). It is not clear whether in this phase RNA Pol

II is really stuck 20-50 nucleotides downstream the transcriptional start site (TSS) or whether it is just a slow process that, in ChIP-seq experiments for RNA Pol II, leads to high signal in this region of the gene as the enzyme was blocked (Core et al., 2012). However, in order to continue transcription and start productive elongation, p-TEFb needs to be recruited. P-TEFb is a dimeric protein, formed by cyclin T and CDK9 that triggers pause release by phosphorylating DSIF, NELF and serine 2 of the CTD. Thereby, DSIF is converted into a positive elongation factor that travels with the RNA Pol II till the end of the gene (Peterlin and Price, 2006), NELF dissociates from the polymerase (Peterlin and Price, 2006) and the serine 2 phosphorylated CTD creates a platform for the interaction with RNA processing factors such as splicing and exporting factors and chromatin modifying proteins that facilitate productive RNA synthesis (Adelman and Lis, 2012).
Important evidence shows that MYC is involved in the recruitment of p-TEFb and in promoting transcription elongation in addition to c transcription initiation. It was shown that MYC recruits p-TEFb at the cyclin D2 promoter (Bouchard et al., 2004) and that the MYC-mediated recruitment of p-TEFb at the CAD promoter is needed for stimulating transcription elongation (Eberhardy and Farnham, 2002). Moreover, immunoprecipitation studies using cell extract showed that MYC or the MYC/MAX complex interacts with p-TEFb subunits (Gargano et al., 2007; Kanazawa et al., 2003; Rahl et al., 2010). Most strikingly, treatment of embryonic stem cells with 10058-F4, an inhibitor of MYC/MAX heterodimerization (Yin et al., 2003), caused a reduction of serine 2 phosphorylated RNA Pol II but had no effect on serine 5 phopshorylation (Rahl et al., 2010). Moreover, as determined by ChIP-seq analysis of RNA-Pol II distribution, 10058-F4 treatment decreased the RNA Pol II signal in the gene body and at the termination site but did not at the promoter. The same effect was seen by using an shRNA against MYC or flavopiridol, a molecules that inhibits p-TEFb kinase activity (Chao and Price, 2001), supporting the idea that MYC is needed for transcription elongation (Rahl et al., 2010).

1.3.2 MYC as a general amplifier of gene expression

MYC binds thousand of genomic loci (Eilers and Eisenman, 2008; Guccione et al., 2006) and in embryonic stem cells, manipulation of MYC levels affects 1/3 of the expressed genes (Rahl et al., 2010). Accordingly, two recent studies suggested that MYC is a general amplifier of transcription rather than an on-off specifier of a defined transcriptional program(s) (Lin et al., 2012; Nie et al., 2012). Inducible overexpression of MYC in primary murine B and T cells and in P493-6 B cell model for Burkitt's lymphoma

amplified the output of the existing gene expression program present in these cells. In cells with low levels of MYC, the transcription factor was bound almost exclusively to canonical E-boxes in the promoters of genes found in an open chromatin structure (H3K4me3 and H3K27Ac). At higher levels, MYC occupied both the promoters and the enhancers of all active genes at additional, low affinity E-box sequences. MYC was recruited to these sites according to the amount of RNA Pol II pre-loaded at their promoters and a positive correlation between the expression levels of the genes and the amount of MYC bound was shown. The increased in MYC occupancy led to increase transcription elongation by RNA Pol II and increased levels of transcripts per cell. No direct repression by MYC was observed in this system. Instead repression would result from the induction by MYC of a transcriptional repressor or form a normalization that is not based on the number of cells but on the RNA amount (Lovén et al., 2012; Nie et al., 2012).

Opposite to what described by the Young's and Levens' laboratories, gene expression analyses of tumor samples such as medulloblastoma, neuroblastoma and breast cancer, identified sets of genes that are specifically up- and downregulated by MYC and have significant prognostic value (Brockmann et al., 2013; Horiuchi et al., 2012; Kawauchi et al., 2012; Northcott et al., 2011).

It remains an open question how MYC can act both as general amplifier of gene expression, increasing output of all active promoters and as a specifier able to activate and repress transcription of defined target genes.

1.3.3 MYC-mediated repression

In contrast to MYC-mediated activation, repression by MYC is a less well understood mechanism taking place at genes involved in cell adhesion (Gebhardt et al., 2006; Inghirami et al., 1990) and inhibition of cell cycle progression (Seoane et al., 2001; Staller et al., 2001). It mainly involves MYC binding to transcriptional activators, the displacement of activating co-factors and the recruitment of transcriptional repressors. Nevertheless, MYC binding to MAX and to the DNA is also important for repression. Heterodimerization with MAX is necessary for binding to the *CDKN2B* promoter and E-box elements are found in the promoter of many MYC-repressed genes (Mao et al., 2003).

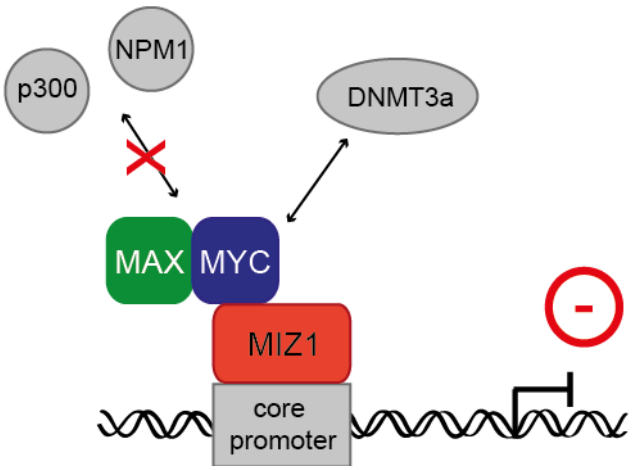


Figure 1. 4: Schematic diagram of the transcriptional repression by the MYC/MIZ1 complex. MIZ1 alone is bound at the core promoter of genes activating their expression. Binding of the MYC/MAX complex to MIZ1 hinder the recruitment by MIZ1 of its coactivators p300 and nucleophosmin (NPM1). Histone modifying enzymes (e.g. DNMT3a) are recruited to MIZ1 target genes by MYC/MAX leading to a close chromatin structure and to repression. Modified from Herkert and Eilers, 2010.

The MYC/MIZ1 complex, beside inhibiting the recruitment of nucleophosmin and p300 by MIZ1 alone (see 1.3.4), interacts with DNA methyltransferases to repress transcription (Brenner et al., 2005; Licchesi et al., 2010). MYC associates with DNMT3a methyltransferase and directs its activity to the *CDKN1A* promoter via binding to MIZ1. DNMT3a methylates the CpG dinucleotides leading to the silencing of the gene (Brenner et al., 2005). MYC might repress *CDKN1A* expression also by binding to the transcriptional activator SP1 (Gartel et al., 2001). SP1 DNA binding sites are present at many MYC-repressed genes indicating that inhibition of SP1 activity might be important for MYC-mediated repression (Herkert and Eilers, 2010). Moreover, N-MYC was shown to recruit HDAC2 to the SP1 site at the cyclin G2 promoter (Marshall et al., 2010).

MYC and MIZ1 were found also in a ternary complex with GFI-1 (growth factor independence-1) repressor, downregulating transcription of *CDKN2B* and *CDKN1A* (Basu et al., 2009; Liu et al., 2010). GFI-1 is a nuclear zinc finger transcriptional repressor with an important role in hematopoiesis and has been implicated in lymphomagenesis.

Evidence suggests that at repressed genes MYC recruits histone deacetylases and polycomb proteins (Corvetta et al., 2013; Zhang et al., 2012c). In lymphomas, MYC associates with the histone deacetyltransferase HDAC3 and EZH2, a core protein of the polycomb repressive complex 2 (PRC2). The ternary complex is tethered to the promoter region of miR-29 downregulating its expression via histone deacetylation and

trimethylation, contributing to aggressive clinical outcome of the MYC-associated lymphoma (Zhang et al., 2012c). Recruitment of HDAC3 by MYC was also reported for other repressed target genes such as ID2, GADD153 (Kurland and Tansey, 2008) and miR-15a/16 (Zhang et al., 2012d). Moreover, RNA Pol II is recruited to the GADD153 and GADD45a promoters whether MYC is bound or not, suggesting that MYC might repress transcription of these genes through a post RNA Pol II recruitment mechanism (Barsyte-Lovejoy et al., 2004).

1.3.4 Transcriptional regulation by MIZ1

The activity of MIZ1 was mainly studied as part of a repressive complex formed with MYC that controls transcription of CDK inhibitors such as $p15^{INK4b}$, $p21^{CIP1}$ and $p57^{KIP2}$ (encoded by *CDKN2B*, *CDKN1A* and *CDKN1C*, respectively) (Adhikary et al., 2003; Seoane et al., 2001, 2002; Staller et al., 2001). In contrast to other POZ domain proteins, MIZ1 is a soluble and strong transcriptional activator and the best-studied mechanism for its action is at the promoter of the $p15^{INK4b}$ encoding gene. *CDKN2B* is a target gene of the TBG-β signaling that is activated early in the response and interacts with and inhibits CDK4 and CDK6, the two kinases associated with cyclin D activity. Thereby, p15*INK4b* expression inhibits cell cycle progression at the G1 phase (Hannon and Beach, 1994). MIZ1 binds to the core promoter of the *CDKN2B* gene and activates its expression (Staller et al., 2001). In epithelial cells, TGF-β signaling activation leads to decreased levels of MYC and thus dissociation from MIZ1, and formation of SMAD complex. The SMAD complex binds to the promoter of *CDKN2B* and interacts with MIZ1 leading to the expression of the CKD inhibitor (Seoane et al., 2001). Displacement of MYC from MIZ1 is required for activation of *CDKN2B* so that the histone acetyltansferase p300 or the co-activator nucleophosmin can be recruited by the zinc finger transcription factor (Staller et al., 2001; Wanzel et al., 2008).

MIZ1 -via its POZ domain- heterodimerizes with other POZ domain proteins such as BCL6 and ZBTB4. The latter is a transcriptional repressor that binds to MIZ1 at the *CDKN1A* promoter recruiting the SIN3-histone deacetyltransferase complex, thereby inhibiting cell cycle arrest in response to p53-activation (Weber et al., 2008).

1.4 Aim of the project

The proto-oncogene MYC is one of the most extensively studied transcription factors. Its expression is elevated in many human cancers and this correlates with tumor aggression and poor clinical outcome (Dang, 2012). In normal cells, MYC links growth factor stimulation with cell growth and proliferation, whereas in tumor cells different kinds of events lead to increased MYC levels that uncouple growth factor stimulation and cellular growth and proliferation. How MYC control such a broad spectrum of cellular processes that are sometimes cell type- or context-specific is not clear yet. Moreover, discrepancies exist about whether MYC acts just by enhancing a pre-existing cellular program(s) or whether it regulates specific processes via activation and repression of target genes. Studies in medullobastoma showed that MYC up- and downregulates target genes that are specific and differ from those regulated by N-MYC or pathways that drive other medulloblastoma subgroups (SHH and WNT). Moreover, high MYC expression correlates with a highly aggressive medulloblastoma tumor that carry very poor prognosis (Kawauchi et al., 2012; Northcott et al., 2011). On the other hand, MYC was described by the Levens' and the Young's laboratories as a general transcription factor that, when overexpressed, "invades" all open promoters and enhancers amplifying transcription of all genes having an open chromatin structure (Lin et al., 2012; Nie et al., 2012).

Understanding how MYC works and regulates transcription and how this transcription factor can behave so differently depending on the cellular context was the subject of this work. By using genome wide binding analysis, gene expression analysis and several bioinformatic tools it was investigated what the feature of MYC target genes are and how they react to different MYC levels present in the cells.

Chapter 2:

Materials

2.1 Strains and cell lines

2.1.1 Human cell lines

HCT116	Human colorectal carcinoma cell line (ATCC)
HEK293T	Human embryonic kidney cell line (ATCC)
HeLa	Human cervix carcinoma cell line (ATCC)
HMLE	Human mammary epithelial cell line
IMECs	Immortalized mammary epithelial cells
MCF10A	Human mammary epithelial cell line from fibrocystic disease
U2OS	Human osteosarcoma cell line (ATCC)

2.1.2 Bacterial strains

XL1 blue

Escherichia coli; *recA1, endA1, gyrA96, thi-1, hsdR17, supE44, relA1, lac [F' proAB lacIqZΔM15 Tn10(Tetr)]*; for generation and amplification of plasmids

2.2 Cultivation media and supplements

2.2.1 Media for mammalian cell culture

Dulbecco's Modified Eagle's Medium (DMEM) containing 584mg/ml L-glutamine was purchased by Sigma. Fetal bovine serum (FBS, PAA) was heat-inactivated for 30min at 56°C.

Basal medium: DMEM, 10% FBS, 1% penicillin/streptomycin.

Freezing medium: 90% FBS, 10% DMSO.

Transfection medium: DMEM, 2% FBS.

2.2.2 Antibiotics for mammalian cell culture

Penicillin/streptomycin (100000U/ml, PAA or Sigma) was used to avoid bacterial contaminations. 0.1% Ciprofloxacin (1mg/ml, Sigma) was added to the medium to avoid mycoplasma contaminations.

To select transfected or infected mammalian cells, 2µg/ml puromycin (10mg/ml, InvivoGen) or 2.5µg/ml hygromycin (50mg/ml, Life technologies) was added to the culture medium.

2.2.3 Media and antibiotics for bacterial cell culture

LB medium

10% bacto tryptone

0.5% yeast extract

1% NaCl

LB agar

LB medium

1.2% Bacto-Agar

Autoclaved, cooled to 50°C before adding specific antibiotics, ~10ml poured into 10cm dishes.

Antibiotics

100µg/ml ampicillin was added to the medium to select successfully transformed bacteria.

2.3 Nucleic acids

2.3.1 Primers

DNA primers designed with Primer3 and were synthesized by Sigma (f= forward, r=reverse). Primers for qRT-PCR are all-intron spanning to avoid genomic DNA amplification.

Name	Sequence 5'-3'	Application
NPM1_f	TTCACCGGGAAGCATGG	ChIP-qPCR
NPM1_r	CACGCGAGGTAAGTCTACG	ChIP-qPCR

NCL_f	CTACCACCCTCATCTGAATCC	ChIP-qPCR
NCL_r	TTGTCTCGCTGGGAAAGG	ChIP-qPCR
HSPBAP1_f	ACCACGCAGCTTTGTTTTGA	ChIP-qPCR
HSPBAP1_r	GCTAAGGTCCGGGTTAGGTA	ChIP-qPCR
FBXO32_f	GAGAGGATCTCAAGCGTTGC	ChIP-qPCR
FBXO32_rev	CTCTTCCGGCAACAAAGAGC	ChIP-qPCR
Ctrl_region_ch11_80MB_f	TTTTCTCACATTGCCCCTGT	ChIP-qPCR
Ctrl_region_ch11_80MB_r	TCAATGCTGTACCAGGCAAA	ChIP-qPCR
MYC_f	CACCAGCAGCGACTCTGA	RT-qPCR
MYC_r	GATCCAGACTCTGACCTTTTGC	RT-qPCR
CAMKV_f	TGATTTGGGACAGGTCATCA	RT-qPCR
CAMKV_r	TGGAACTTCTTGCAGGTGTG	RT-qPCR
RGS16_f	CTGCGATACTGGGAGTACTGG	RT-qPCR
RGS16_r	CCACCCCAGCACATCTTC	RT-qPCR
COL5A1_f	GACACCTCCAACTCCTCCAA	RT-qPCR
COL5A1_r	TCTCGTCAAGGTTCCGGATC	RT-qPCR
ALDH3B1_f	AAGCCATCGGAGATTAGCAA	RT-qPCR
ALDH3B1_r	AGCAGCTCTGGTCCACGTAT	RT-qPCR
B2M_f	GTGCTCGCGCTACTCTCTC	RT-qPCR
B2M_r	GTCAACTTCAATGTCGGAT	RT-qPCR

2.3.2 RNA oligonucleotides

Pool of RNA oligonucleotides against c-MYC were purchased from Dharmacon (ON-TARGETplus SMARTpool). As control the siCONTROL (ON-TARGETplus Non-targeting Pool) was used.

2.3.3 Oligonucleotides for shRNA cloning

Name	**Sequence 5'-3'**
shMIZ1_1	CCGGGGTGGACGGTGTTCACTTTCTCGAGAAAGTGAACACCGTCC ACCTTTTTG

2.4 Plasmids

2.4.1 Empty vectors

pLKO: vector for the expression of shRNA in mammalian cells (TRC Consortium)

2.4.2 Expression vectors

The following plasmids were already present in the collection of the group.

pLKO shMIZ1_1: pLKO vector with puromycin resistance and shRNA against MIZ1 mRNA

pLKO shMIZ1_2: pLKO vector with puromycin resistance and shRNA against MIZ1 mRNA

2.4.3 Packaging plasmids for lentivirus production

Packaging plasmids for lentivirus production were obtained from Manfred Gessler.

psPAX2: vector coding for the lentiviral virion packaging system (HIV gag, pol, rev).

pMD2-VsVg: vector encoding for lentiviral envelop (VSV-G) for a higher virus stability.

2.5 Antibodies

2.5.1 Primary antibodies

Protein	Clone	Application	Supplier
MYC	N262	ChIP Immunoflourescence	Santa Cruz
	9E10	Immunoblot	Group Eilers
MIZ1	10E2	ChIP, Immunoblot	Group Eilers
VINCULIN	hVIN-1	Immunoblot	Sigma
ACTIN	AC-15	Immunoblot	Sigma
RNA POL II	N20	ChIP	Santa Cruz
pSer5 POL II	4H8	ChIP	Cavance
pSer2 POL II	ab5095	ChIP	Abcam
BrdU-FITC	B44	FACS	BD Biosciences
panAc H3	06-599	ChIP	Upstate (Millipore)
panAc H4	06-866	ChIP	Upstate (Millipore)
H3K4me1	ab8895	ChIP	Abcam
H3K4me3	ab8580	ChIP	Abcam

H3K27Ac	07-360	ChIP	Upstate (Millipore)
IgG	rabbit	ChIP	GE Healtcare
IgG	mouse	ChIP	Sigma

2.5.2 Secondary antibodies

Name	Clone	Application	Supplier
Anti-rabbit HRP	NA 934	Immunoblot	Amersham
Anti-mouse HRP	NA 931	Immunoblot	Amersham
Anti-rabbit	Alexa 488	Immunofluorescence	Life Technologies
IRDye® 680RD	rabbit	Immunoblot	LI-COR
IRDye® 680RD	mouse	Immunoblot	LI-COR
IRDye® 800CW	rabbit	Immunoblot	LI-COR
IRDye® 800CW	mouse	Immunoblot	LI-COR

2.6 Chemicals

All chemicals were purchased from the following companies without further purification: Sigma, Merck, Roth, Invitrogen and Applichem.

2.7 Enzymes, standards and kits

2.7.1 Enzymes

DNase-free RNase A (Quiagen)

M-MLV reverse transcriptase (Promega)

Proteinase K (Roth)

Restriction endonucleases (Fermentas)

RNAse-free DNase (Quiagen)

RNase A (Roth)

2.7.2 Standards

PageRuler Prestained Protein Ladder (Fermentas)

1 kb DNA Ladder (New England Biolabs)

2.7.3 Kits

RNeasy Kit (Qiagen)

QIAquick PCR Purification Kit (Qiagen)

QIAquick Gel Extraction Kit (Qiagen)

SYBR Green qPCR Master Mix (Thermo Fisher Scientific)

Quant-iT™ PicoGreen® dsDNA assay kit (Life Technologies)

NEBNext ChIP-Seq Library Prep Master Mix Set for Illumina (New England Biolabs)

Sera-Mag Oligo(dT)-Coated Magnetic Particles (Thermo Scientific)

NEBNext® mRNA Library Prep Master Mix Set for Illumina® (New England Biolabs)

NEBNext® Poly(A) mRNA Magnetic Isolation Module (New England Biolabs)

NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (New England Biolabs)

Agencourt® AMPure® XP (Beckman Coulter)

NEBNext® Multiplex Oligos for Illumina® (index Primers Set 1 and 2) (New England Biolabs)

Experion™ RNA analysis kits and chips (HighSens and StdSens) (BIO-RAD)

Experion™ DNA 1K kit and chip (BIO-RAD)

PureLink® HiPure Plasmid Maxiprep Kit (Invitrogen)

2.8 Buffers and solutions

Annealing buffer

10mM Tris pH 7.5

50mM NaCl

1mM EDTA

Blocking solution for PVDF membranes

5% skim milk powder in TBS-T

BSA-PBS

0.5% BSA

In PBS

ChIP lysis buffer 1

5mM PIPES pH 8.0

85mM KCl

0.5% NP-40

ChIP wash buffer 1

20mM Tris pH 8.1

150mM NaCl

2mM EDTA

0.1% SDS

1% Triton-X-100

ChIP wash buffer 2

20mM Tris pH 8.1

500mM NaCl

2mM EDTA

0.1% SDS

1% Triton-X-100

ChIP wash buffer 3

10mM Tris pH 8.1

250mM LiCl

1% NP-40

1% SDS

1mM EDTA

ChIP elution buffer

1% SDS

100mM $NaHCO_3$

Coomassie staining solution

25% isopropanol

10% acetic acid

0.05% Coomassie G250

Coomassie destaining solution

10% acetic acid

20% methanol

DNA loading buffer

40% saccharose (pH 8.0)

0.2% bromophenol blue

0.2% xylene cyanol

10mM EDTA

Doxycycline 1mg/ml

50mg doxycyclin hyclate (Sigma)

to 50ml water

PEI

450µl PEI (10% solution)

150µl HCl (2N)

49.5ml water

Phenol chloroform solution

25ml Phenol

24ml Chloroform

1ml Isoamyl Alcohol

PBS

137mM NaCl

2.7mM KCl

10.1mM Na_2HPO_4

1.76mM KH_2PO_4

autoclaved

Plasmid prep buffer 1

50mM Tris-HCl (pH 8.0)

100mM EDTA

100µg/ml RNase A

Plasmid prep buffer 2

200mM NaOH

1% SDS

Plasmid prep buffer 3

3.1M potassium acetate (pH 5.5)

RIPA buffer

50mM HEPES (pH 7.9)

140mM NaCl

1mM EDTA

1% Triton-X-100

0.1% Sodium deoxycholate

0.1% SDS

SDS sample buffer 6X

1.2g SDS pellet

6mg Bromophenol blue

4.7ml Glycerol 86%

2.1ml water

0.93g DTT

SDS running buffer

25mM Tris Base

250mM Glycine

0.1% SDS

Stripping buffer

62.5% Tris (pH 6.8)

2% SDS

100mM β-mercaptoethanol

Transfer buffer Tris-Glycin system

1.9M glycine

250mM Tris base

0.05% SDS

Adjust pH to 8.0

TAE 50X

2M Tris

5.7% acetic acid

50mM EDTA

adjust pH to 8.0

TBS-T

0.2% Tween-20

25mM Tris

140mM NaCl

Adjust to pH 7.4

TE

10mM Tris

1mM EDTA

Adjust to pH 8.0

Trypsin solution

0.25% Trypsin

5mM EDTA

22.3mM Tris (pH 7.4)

125mM NaCl

200X Reduction agent for Bis-Tris system

1M Sodium bisulfite

3.5X Bis-Tris buffer

1.25M Bis-Tris

Adjust pH 6.7 with HCl

1X Transfer buffer Bis-Tris system

50ml 20X transfer buffer

700ml water
250ml methanol

MES
50mM MES
50mM Tris base
0.1% SDS
1mM EDTA
pH 7.3

20X Transfer buffer Bis-Tris system
25mM Bicine
25mM Bis-Tris
1mM EDTA
pH 7.2

2.9 Consumables and equipment

Consumables such as reaction tubes, cell culture dishes and other plastic products were purchased from Eppendorf, Greiner, Nunc, Apllied Biosystems, Sarsted, Millipore and Kimberley-Clark, B. Braun, Schleicher and Schuell and VWR international.

2.9.1 Equipment

Blotting system
PerfectBlue Tank Electro Blotter Web S (PEQLAB)

Cell culture incubator
BBD 6220 (Heraeus)

Cell counter
CASY cell counter (Innovatis)

Centrifuges
Galaxy MiniStar (VWR Interantional)

Eppendorf 5417 R (Eppendorf)

Eppendorf 542 (Eppendorf)

Multifuge 1S-R

Avanti J-26 XP (Beckman Coulter)

Chemiluminescence imaging

LAS-4000 mini (Fujifilm)

Fluorcytometer

BD FACS Canto II (BD Bioscences)

Fluorescence readers

Odyssey® CLx Infrared Imaging System (LI-COR)

Infinite 200 PRO Microplate Reader (Tecan)

Heating block

Dry Bath System (STARLAB)

Thermomixer® comfort (Eppendorf)

Heat Sealing

ALPS™ 50V (Thermo Fisher Scientific)

Incubator shaker

Model G25 (New Brunswick Scientific)

Microscopes

Axiovert 40CFL (Zeiss)

DMI 6000 B (Leica)

SP5 (Leica)

Nucleic acid analysis

Experion™ Automated Electrophoresis (BIO-RAD)

Mx3000P (Stratagene)

PCR thermal cycler

Mastercycler Pro S (Eppendorf)

Photometers

Ultrospec™ 3100 pro UV/Visible (Amersham Biosciences)

NanoDrop 3000 (Thermo Scientific)

Power supply

PowerPac HC (BIO-RAD)

PVDF transfer membranes

Immobilion P and FL Transfer Membranes (Millipore)

SDS-PAGE system

Mini-PROTEAN Tetra Cell (BIO-RAD)

Sequencing equipment

Illumina Genome Analyzer IIX

Sonifier

Digital Sonifier® W-250 D (Branson)

UV fluorescent table

Maxi UV fluorescent table (PEQLAB)

Vortex mixer

Vortex-Genie 2 (Scientific Industries)

Water bath

ED-5M heating bath (Julabo)

2.10 Software

ApE	M. Wayne Davis
Acrobat Professional	Adobe System, Inc.
BD FACSDiva 6.1.2	BD Biosciences

Bedtools	Quinlan & Hall, 2010
Bowtie	Langmead, 2010
Feature Extraction	Agilent, v10.1.1.1
GraphPad Prism	GraphPad Software, Inc
GSEA	Subramanian et al., 2005
Illustrator	Adobe System, Inc.
Image Studio™ Lite	LI-COR
Integrated Genome Browser	Nicol et al., 2009
Java Tree View	Saldanha, 2004
MACS	Zhang et al., 2008
Mac OS X	Apple Inc.
Microsoft Office 2008	MacMicrosoft Corporation
MultiGauge	Fujifilm Corporation
MxPro qPCR Software	Stratagene
Photoshop	Adobe System, Inc.
R	R foundation
Samtools	Li et al., 2009
Seqminer	Ye et al., 2011
Ubuntu	Canonical Ltd.
Windows 7	Microsoft Corporation

2.11 Online tools and databases

DAVID	http://david.abcc.ncifcrf.gov/
Galaxy	https://main.g2.bx.psu.edu/
GEO	http://www.ncbi.nlm.nih.gov/geo/
MSigDB	http://www.broadinstitute.org/gsea/msigdb/index.jsp
Primer3	http://frodo.wi.mit.edu/
Pubmed	http://www.ncbi.nlm.nih.gov/pubmed
UCSC	https://genome.ucsc.edu/

Chapter 3:

Methods

3.1 Molecular biology methods

3.1.1 Bacterial transformation

Competent cells were thawed on ice and mixed with 1µl of plasmid or the whole ligation mix. After 30min of incubation and 45sec heat shock at 42°C, LB medium was added and incubated at 37°C for 30min. The suspension was then centrifuged and resuspended in 100µl of LB medium and plated on LB-agar dishes containing antibiotics.

3.1.2 Isolation of plasmid DNA from bacteria

For large scale purification of plasmids the PureLink® HiPure Plasmid Maxiprep kit was used following the manufacturer's instructions. 200ml of bacterial suspension grown overnight was pelleted at 8000rpm for 30min at 4°C. Supernatant was discarded and the bacterial pellet was resuspended in R3 buffer (with RNase A). 10ml of L7 lysis solution was added followed by 10ml of neutralization solution N3. The bacterial lysates was centrifuged at 8000rpm for 30min at 4°C and the supernatant was added on specific columns previously equilibrated with 30ml equilibration buffer. The columns were washed twice with wash buffer W8. DNA was eluted with 15ml elution buffer and then precipitated by adding 10ml isopropanol and centrifuging 30min at 800rpm. The DNA pellet was washed twice with 5ml 70% ethanol and resuspended in water. The plasmid solution was digested with sequence-specific endonucleases to verify if the correct plasmid was purified.

3.1.3 Ligation of DNA encoding shRNA into plasmids

The cloning of the shRNAs into the plasmids was performed by other member of the laboratory as follows. The oligonucleotide couple encoding the shRNA was annealed using the annealing buffer and by setting the following program on the thermo cycler:

95°C 2min

to 25°C 1°C/1.5min

dsDNA and plasmids were digested with sequence specific endonucleases and ligated by mixing components as follows:

dsDNA	xµl
plasmid	100ng
T4 DNA Ligase buffer (Fermentas)	1µl
T4 DNA Ligase (Fermentas)	1µl
Total	10µl

The ligation reaction was incubated overnight at 16°C.

3.1.4 Agarose gel electrophoresis

Depending on DNA fragment size, a solution of 1-2% agarose in 1X TAE was prepared. The solution was boiled and poured, with the addition of 0.3µg/ml ethidium bromide, into a gel chamber with combs. DNA loading buffer was mixed with the samples that were then loaded into the wells of the polymerized gel. 10µl of 1kb DNA ladder (NEB) was loaded next to the samples and allowed size determination of the DNA. The gel was run at 120V for one hour, then the DNA was visualized using a UV transilluminator.

3.1.5 DNA extraction and purification from agarose gel

After separation by gel electrophoresis, the fragment DNA of interest was cut out of the gel with a scalpel. The DNA was separated from the agarose gel by using the Gel Extraction kit (Qiagen) following the manufacturer's instructions.

3.1.6 RNA isolation and quantification

For isolation of total RNA from cultured cells the TriFast™ (peqlab) was used. For cells grown on a 10cm dish, 600µl of TriFast was added directly on the dish. The cell suspension was then collected in a 1.5ml eppendorf tube and 200µl of chloroform was added. The mixture was vortexed thoroughly for 1min and then incubated at room temperature for 5min. The tubes were then centrifuged for 5min at 13600rmp and the supernatant containing the RNA was transferred into new tubes. RNA was precipitated by incubating it for 15min with 1volume of isopropanol and 1µl of GlycoBlue Coprecipitant (Life Technologies) and by centrifuging the solution for 15min at 13600rpm at 4°C. The

RNA pellet was washed once with 500µl 70% ethanol and then resuspended in 50µl of water. The sample were stored at -20°C or -80°C and RNA concentration and purity was determined by Nanodrop measurement.

To quantify the amount of total and mRNA in a cell, total RNA was isolated from a fixed number of cells by using the RNeasy Kit (Qiagen) and quantified by NanoDrop 1000 (peqlab). mRNA was isolated from total RNA by using the Sera-Mag Oligo(dT)-Coated Magnetic Particles (Thermo Scientific), was purified by isopropanol precipitation and quantified by NanoDrop 1000 (peqlab).

3.1.7 cDNA synthesis

Total RNA was transcribed into complementary DNA (cDNA). 2µl of random hexanucleotides were added to 2.5µg of RNA diluted in 10µl of water. The mix was heated up at 65°C for 3min and then put directly on ice. Afterwards, the following reaction was set up for each sample:

10µl	5X RT buffer (Promega)
5µl	DTT 0.1M
5µl	DNTPs 2.5M
0.2µl	RiboLock RNasea Inhibitor (40U/l, Life Technologies)
1µl	M-MLV Reverse Transcriptase (200U/l, Promega)
16.8µl	water

Reverse transcription was performed by incubating the samples for 10min at 22°C, 50min at 37°C and 15min at 70°C.

3.1.8 Nucleic acid quantification

For routine analyses, the concentration of DNA and RNA was measured with NanoDrop 1000 (peqlab). The purity of the nucleic acids was determined by assessing the ration of absorbance at 260nm and 280nm. For pure samples the ratio is about 1.8.

3.1.9 Quantitative polymerase chain reaction (qPCR)

cDNA and the recovered DNA were amplified by qPCR to quantify specific mRNA levels and ChIP enrichment, respectively. For each sample technical replicates were performed and the reactions were set up as follows:

1µl DNA

1µl primer f+r mix (10µM)

7µl SYBR Green Mix (Thermo Scientific)

11µl water

The measurements were carried out with the Mx3000P qPCR System (Stratagene) using the following thermal cycling profile:

	95°C	15min
38 cycles	95°C	30sec
	60°C	20sec
	72°C	15sec
1 cycle	95°C	1min
	60°C	30sec
	95°C	30sec

The quantification of the amplified DNA can be determined by fluorescence monitoring in every cycle after the end of the elongation step. Calculation of the relative transcript amount or DNA enrichment was performed using the ΔΔ-CT method (Applied Biosystems User Bulletin 2). For normalization of RNA or ChIP samples, the housekeeping gene β2M or the input sample were used, respectively.

3.1.10 Sample preparation for RNA-sequencing

For RNA-sequencing (RNA-seq), total RNA was isolated by using the RNeasy kit (Qiagen) following the manufacturer's instruction. For cells grown on a 10cm dish, 600µl of Buffer RLT with 1% β-mercaptoethanol were used. Genomic DNA on-column digestion was performed.

The quality and concentration of the isolated RNA was assessed by using the Experion™ RNA analysis kits and chips (StdSens) (BIO-RAD) and the Experion™ Automated Electrophoresis System (BIO-RAD). Good quality samples have a RIQ above 8.

Two different procedures were used to prepare the samples for sequencing.

Protocol 1: isolation of the polyadenylated RNA was performed from 5-10µg of total RNA using the Sera-Mag Oligo(dT)-Coated Magnetic Particles (Thermo Scientific) following the manufacturer's instructions. The removal of the rRNA was assessed by loading 1µl of samples on the Experion™ RNA analysis kits and chips (HighSens) (BIO-RAD). mRNA fragmentation, cDNA sysnthesis, end-repair, dA-tailing, adaptor ligation, size selection and PCR was performed with the NEBNext® mRNA Library Prep Master Mix Set for Illumina® (New England Biolabs), the QIAquick PCR Purification Kit

(Qiagen) and the QIAquick Gel Extraction Kit (Qiagen). When required DNA was purified using QIAquick PCR Purification Kit (Qiagen). Size selection of the adaptor-ligated DNA was performed by loading the samples with 6µl orange sample buffer on a 2% agarose gel with ethidium bromide. The gel was run for 1h at 170V and then a small band of gel corresponding to the height of 200bp DNA was cut out by using the Ultra Cruz™ Disposable Gel Excision tips (6.5x1 mm, Santa Cruz). DNA was isolated from the gel by using the QIAquick Gel Extraction Kit (Qiagen) and amplified with 12-15 PCR cycles using a different index primer for each sample (NEBNext® Multiplex Oligos for Illumina® (index Primers Set 1 and 2) (New England Biolabs)).

Protocol 2: from 1µg total RNA, the polyadenylated RNA was isolated using the NEBNext® Poly(A) mRNA Magnetic Isolation Module (New England Biolabs) following the manufacturer's instructions. Library preparation, consisting of the same steps as protocol 1, was performed with NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (New England Biolabs) following the manufacturer's instructions. DNA purification and size selection were performed using the Agencourt® AMPure® XP (Beckman Coulter) and DNA was amplified with 12 PCR cycles using a different index primer for each sample (NEBNext® Multiplex Oligos for Illumina® (index Primers Set 1 and 2) (New England Biolabs)).

Quality and amount of the generated libraries were assessed using the Experion™ DNA 1K kit and chips (BIO-RAD).

All the samples were mixed together at equimolar concentrations and subjected to sequencing on an Illumina Genome Analalyzer IIx sequencer.

3.1.11 Sample preparation for ChIP-sequencing

For ChIP-seq the same protocol as normal ChIP (see 3.3.9) was used with the following modifications. Chromatin isolated form 50-100x10^6 cells was immunoprecipitated by using 100µl of Dynabeads protein A and G (Life Technologies) in a 1:1 ratio and 10µg of specific antibody.

Quantification of DNA recovered after the immunoprecipitation was performed using the Quant-iT™ PicoGreen® dsDNA assay kit (Life Technologies) following the manufacturer's instructions. Up to 10ng of DNA was then used for library preparation.

Samples were modified for sequencing using the NEBNext ChIP-Seq Library Prep Master Mix Set for Illumina (New England Biolabs) following the manufacturer's instructions.

Briefly, the recovered DNA was end-repaired, dA-tailed, ligated to Illumina adaptors, size-selected on a 2% agarose gel (band corresponding to 175-225bp fragments was cut out using scalpels) and amplified with 18 PCR cycles using a different index primer for each sample (NEBNext® Multiplex Oligos for Illumina® (index Primers Set 1 and 2) (New England Biolabs)). When needed DNA was purified using QIAquick PCR Purification Kit (Qiagen) or QIAquick Gel Extraction Kit (Qiagen).

As for RNA-seq samples, quality and amount of the generated libraries were assessed using the Experion™ DNA 1K kit and chips (BIO-RAD).
All the samples were mixed together at equimolar concentrations and subjected to sequencing on an Illumina Genome Analyzer IIx sequencer.

3.2 Cell biology methods

All cell culture work was performed at sterile workbenches. Cells were grown in incubators at 37°C, 95% relative humidity and 5% CO_2.

3.2.1 Passaging, freezing and thawing cells

For passaging of adherent growing cells, the cultivation medium was removed and cells were washed once with PBS. 1-2ml of trypsin solution was added on the dish and incubated 5min at 37°C. Trypsin activity was stopped by adding fresh medium, the cell suspension was collected in tubes and centrifuged at 1200rpm for 5min. The cell pellet was resuspended in fresh medium and a portion of cell suspension was plated in new dishes with medium. For S1 cell culture, cell were counted with CASY cell counter, for S2 cell number was determined with the Neubauer counting chamber.
For long-term freezing storage, cell pellet were resuspended in 1ml freezing medium, transferred in cryo vials and slowly frozen at -80°C with Mr FROSTY freezing container. After 24h cells were stored in a liquid nitrogen storage tank.
To thaw frozen cells, the cryo vials was quickly heated up at 37°C in a water bath and then the cell suspension was transferred in a dish containing fresh medium.

3.2.2 Transfection of siRNA

For transfection of synthetic siRNAs, 1-1.5x10^6 cells were seeded on a 10cm dish. One hour before trasfection the cell medium was exchanged with 6ml transfection medium. 10µl of siRNA and 10µl Lipofectamine RNAiMAX (Invitrogen) were diluted separately in 700µl Opti-MEM 1 (Invitrogen). The two solutions were mixed, incubated 20min at room temperature and then added to the cells dropwise. 12-16h later, cells were provided with fresh medium and if necessary splitted. Cells were harvested 48-72h after transfection.

3.2.3 Polyethylenimine (PEI) transfection - Lentivirus production

For production of lentivirus PEI transfection was used. The day before transfection 5x10^6 HEK293T cells were seeded on a 10cm dish. The day of the transfection, two sets of 1.5ml tubes were prepared. In one set 700µl of Opti-MEM 1 (Invitrogen) and 30µl of PEI solution were mixed, in the other one 700µl Opti-MEM 1 (Invitrogen), 11.1µg plasmid DNA, 2.8µg pPAX2 and 1.4µg pMD2G were mixed. After having incubated 5min at room temperature, the two solutions were mixed and incubated 20min. Meanwhile, the medium of the HEK293T cells was replaced with 6ml transfection medium. The PEI solution was added to the cells overnight and subsequently the cell medium was replaced with 6ml basal medium. After 24h medium was collected (1st harvesting of the virus) and replaced with 6ml new basal medium. Two more harvesting were done each at 12h interval.
The virus suspension was then filtered with 0.45µm filters, frozen with liquid nitrogen and stored at -80°C.

3.2.4 Infection of cells with lentivirus

Cells were infected with lentivirus to stably integrate plasmid DNA into their genome.
The day before infection, 5x10^5 fast growing cells were seeded in a 10cm dish. Infection of f virus was achieved by adding to the cell a solution containing: 2ml viral suspension, 4ml normal medium and 6µl polybrene (4µg/ml). The infection was repeated twice at 24h interval and then normal medium was added. Appropriate antibiotics were used to select infected cells starting 2day after infection.

3.2.5 Generation of MYC-inducible cell line

The generation of the MYC-inducible U2OS cell line was performed by Elmar Wolf. Briefly, U2OS cells were stably transfected with a doxyciclin-inducible two vector system

bearing the coding sequencing of MYC (tet-on, Clontec Laboratories). Positively transfected cells were selected with 2.5µg/ml hygromycin (Invitrogen). MYC expression was achieved by adding doxycycline to the medium at a final concentration of 1µg/ml.

3.2.6 Flow cytometry - BrdU/PI FACS

Cells were labeled with 10µM BrdU for 60min at 37°C. Cells were harvested by trypsinization, resuspended in 10ml fresh medium and centrifuged for 5min at 1500rpm. Cells were washed once with 10ml ice cold PBS and then resuspended in 1ml ice cold PBS. The cells-PBS suspension was added drop-wise in 4ml ice cold 100% ethanol while vortexing and incubated overnight at -20°C. Afterwards, the cell suspension was centrifuged 10min at 1500rpm at 4°C, the resulting cell pellet was washed with 5ml PBS and centrifuged for 5min at 1500rpm at 4°C. The pellet was resuspended in 1ml 2M HCl and 0.5% Triton-X-100 solution and incubated 30min at room temperature with gentle mixing. Cells were centrifuged 10min at 1500rpm at 4°C and resuspended in 1ml 0.1$Na_2B_4O_7$ pH 8.5 buffer. The solution was centrifuged 5min at 2000rpm at 4°C and the pellet resuspended in 100µl 1% BSA in PBS-T (0.5% Tween-20 in PBS) and 20µl anti-BrdU-FITC antibody and incubated 30min at room temperature in the dark. After centrifuging 5min at 2000rpm at 4°C, pellet was washed with 200µl 1% BSA in PBS-T, resuspended in about 400µl of buffer containing 38mM sodium citrate, 54µM propidium iodide and 24µg/ml RNase A and incubated 30min at 37°C in the dark. Cells were transferred to a FACS tube, mixed by pipetting and subjected to the FACS measurement (PI: FL2-A LIN, BrdU: FL-H Log).

3.3 Protein biochemistry methods

3.3.1 Generation of protein lysates

To isolate proteins, cells grown on a cell culture dish were fist washed twice with ice cold PBS, scraped in a 1.5ml tube and then pelleted at 1200rpm for 5min at 4°C. The cell pellet was either frozen with liquid nitrogen and stored at -80°C or lysed by adding 50-300µl ice cold RIPA buffer with proteinase and phosphates inhibitors (1:1000) and incubating 20min on ice. Cellular debris were removed by centrifuging the sample at 13600rpm for 5min at 4°C and by transferring the supernatant in a new tube. The protein lysate was frozen with liquid nitrogen and stored at -80°C.

Alternatively, cells were lysed directly on the dish by adding SDS sample buffer and collected in a 1.5ml tube. Samples were stored at -20°C or –80°C.

3.3.2 Protein quantification by Bradford assay

Protein concentration was determined according to (Bradford, 1976).
900µl Bradford Dye reagent and 100µl 1.5M NaCl were mixed with 1.5µl samples in a 1ml cuvette. After vortexing absorbance was measured at 595nm using an appropriate reference. The obtained values were compared to a calibration curve to calculate protein concentration.

3.3.3 Protein quantification by bicinchoninic acid assay (BCA)

After cell lysis, protein concentration was determined by mixing 3µl of protein sample with 200µl bicinchoninc acid and $CuSO_4$ (50:1) solution in a 96-weel plate. The solution was incubated 30min at 37°Cand the absorption was measured at 550 using an appropriate reference. The measured values were compared with a calibration curve to calculate protein concentration.

3.3.4 SDS polyacrylamide gel electrophoresis (SDS-PAGE)

Before being loaded on the SDS-PAGE, samples in sample buffer were boiled for 5min at 95°C. The samples were then loaded in the wells of the gels next to the PageRuler Pre-Stained Protein Ladder (Fermentas) to assess protein size. Gels were run in SDS running buffer or MES, depending on the system used, at 80-110V until the front of the migration was out of the gel.

3.3.4.1 Bis-Tris (Laemmli) gels

The gels with a variable percentage of polyacrylamide were prepared as follows:

For 10ml 10% Running gel: 2.86ml 3.5X Bis-Tris buffer
3.33ml 30% acrylamide/bisacrylamide
3.81ml water
50µl 10% APS
5µ TEMED

For 10ml 4% Stacking gel: 2.86ml 3.5 Bis-Tris buffer
1.33ml 30% acrylamide/bisacrylaminde

5.81ml water
50µl 10% APS
10µl TEMED

3.3.4.2 Tris-Glycine gels

The gels with a variable percentage of polyacrylamide were prepared as follows:

Stacking gel:	4% acrylamide/bis-acrylamide
	125mM Tris HCl pH6.8
	0.1% SDS
	0.1% APS
	0.1% TEMED
Running gel:	10% acrylamide/bisacrylamide
	375mM Tris HCl pH 8.8
	0.1% SDS
	0.1% APS
	0.1% TEMED

3.3.5 Immunoblot

After proteins were separated according to size by SDS-PAGE, they were transferred on a PVDF membrane using a tank blot system. The PVDF membrane of the size of the gel was first incubated for 30sec in methanol and then equilibrated in the transfer buffer. Gel and membrane were layered on top of each other and fixed between Whatman filter paper in a immunoblot transfer chamber filled with transfer buffer (different buffer for Bis-Tris and Tris-glycine system). Transfer was carried out for 3h at 300mA.

The membrane with the immobilized protein was then blocked in blocking solution for at least 30min. The membrane or the membrane pieces were incubated overnight with the primary antibody diluted in the blocking solution, washed three times with TBS-T, incubated with secondary antibody diluted in blocking solution and washed other 3 times with TBS-T. Finally, the protein were visualized via chemiluminescence generated by the horseradish peroxidase coupled to the secondary antibody and the Western Chemiluminescent HRP Substrate (Millipore). Signal was detected with LAS-4000 mini (Fujifilm). Alternatively, proteins were detected with the Odyssey® CLx Infrared Imaging System (LI-COR) where the secondary antibody is coupled with a fluorophore.

3.3.6 Coomassie staining

To visualize recombinant proteins or to verify the complete transfer of protein from the gel to the PVDF membrane, gels were stained with coomassie solution for at least 1 h. To eliminate the excess of dye, the gels were incubated with gentle shaking in destaining solution until a clear background was visible.

3.3.7 Stripping membranes

To remove the antibodies form a PVDF membrane with proteins the membrane was incubate 30min at 60°C in stripping buffer. It was then wash 3 times with TBS-T, blocked in blocking solution and incubated with the primary antibody as previously described.

3.3.8 Indirect immunofluorescence

The cells of interest were grown directly on cover slips. Cells were washed twice with ice cold PBS and fixed by incubating them in 3.7% paraformaldehyde for 15min at room temperature. They were washed twice with 0.1M glycine-PBS solution (3x 10min), permeabilized with 0.1% NP-40 PBS solution (3x10min) and blocked with 5% FBS 0.1%NP-40 PBS (immunofluo-blocking) solution for 45min at 37°C. The cover slips with the cells were moved into a wet chamber and incubated with 40µl primary antibody diluted in immunofluo-blocking solution. After 45 min at 37°C, cells were washed 3 times with immunofluo-blocking solution and then incubated for 45min at 37°C with 40µl secondary antibody in immunofluo-blocking solution and Hoechst nuclear stain (1:5000). After washing 3 times with water, the cover slips were mounted on a glass slide using a small drop of mounting medium. The slides were stored at 4°C in the dark and analyzed with a fluorescence microscope.

3.3.9 Chromatin immunoprecipitation (ChIP)

3.3.9.1 Chromatin preparation

To crosslink proteins to DNA, 1% formaldehyde was added to the cell medium for 10min at room temperature. To stop the cross link 1ml 1M glycine was added for 5min. The medium was removed and the cells washed twice with ice cold PBS. Cells were scraped off the dish in 1ml PBS with proteinase and phosphatase inhibitors (1:1000) and transferred in a 15ml tube (up to 10 dishes in one tube). Cells were centrifuged for 5min at

1200rpm at 4°C, resuspended in 3ml ChIP lysis buffer I containing protease and phosphatase inhibitors (1:1000) and incubated 20min on ice. Afterwards, cells were centrifuged 5min at 1200rpm at 4°C and resuspended in 2ml RIPA buffer with protease and phosphatase inhibitors (1:1000). After 10min incubation on ice cells were sonicated following a specific program depending on the cell type. U2OS cells were sonicated for 20min (10sec on, 30sec off) at 20%amplitude.
After sonication, cell lysates were transferred in a new tube and centrifuged at 13600rpm for 15min at 4°C. The supernatant was moved into new tubes and stored at 4°C for few days or froze with liquid nitrogen and then stored at-80°C.

3.3.9.2 Check fragment size after sonication

25µl of sonified cell lysate was added to 475µl TE buffer with 160mM NaCl and 20µg/ml RNase A. The chromatin was incubated 1h at 37°C and the 6h to overnight at 65°C to revert the crosslink.5mM EDTA and 200µg proteinase K were added to the chromatin and incubated 2h at 45°C. To isolate the DNA 500µl of phenol chloroform solution was added. After vortexing the samples were centrifuged 5min at 13000rpm and the supernatant was transferred in a new 1.5ml tube with 1µl glycoblue, 50µl 3M sodium acetate pH 5.2 and 1ml 100% ethanol. They were incubated 30min at -20°C and then centrifuged for 30min at 13600rpm at 4°C. The pellets were washed with 500µl 70% ethanol, centrifuged for 15min at 13600rpm at 4°C and resuspended in 50µl water. 10µl of sample were mixed with 4µl of DNA loading buffer and loaded on a 2% agarose gel.

3.3.9.3 ChIP

For each immunoprecipitation 30µl of dynabeads protein A and G (Life technologies) 1:1 ratio were used. The beads were washed 3 times with 1ml BSA-PBS (5mg/ml) solution. Afterwards, they were incubated 6h – overnight at 4°C on a rotating wheel with 1ml BSA-PBS and 3µg antibody. Beads were then washed 3 times with BSA-PBS, resuspended in 30µl BSA-PBS and the appropriate amount of chromatin was added. 1% of chromatin was kept as input. After 6h incubation at 4°C on a rotating wheel, the beads with the bound chromatin were washed 3 times with ChIP wash buffer 1, 3 times with ChIP wash buffer 2, 3 times, incubating the beads 5min each time, with ChIP wash buffer 3 and 1 time with TE buffer (beads were also moved into new 1.5ml tubes). The chromatin was eluted twice with 250µl ChIP elution buffer and at the end the two eluates were merged. The reversal of the

crosslink and the purification of the precipitated DNA was performed as previously described (see 3.3.9.2)

3.4 Bioinformatics analysis of ChIP-seq data

3.4.1 Base calling, quality control and filter

The conversion of the pictures taken by the sequencer's camera into text files with quality indication (base calling) was performed with the RTA package form the Illumina Genome Analyzer Data Collection Software (SCS v2.8). Subsequently, fastq files were generated using only high quality reads (PF-clusters) via the Casava software. The quality of the fastq files and therefore of the sequencing run, was verified using the FastQC application (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).
These procedures were performed either by Elmar Wolf or Carsten Ade.

3.4.2 Reads alignment to the genome

To determine the position of the sequencing reads on the genome the program Bowtie was used. Before performing the alignmen,t the reference human genome hg19 and the program were downloaded and installed.
Command: bowtie –t –S –p 14 hg19 file.fastq file.sam
The generated sam file was converted into a binary bam file via the program Samtools.
Command: samtools view –bS –o file.bam file.sam

3.4.3 Peak calling and visualization

The enrichment of reads in certain positions of the genome (peaks) was determined using MACS.
Command: macs14 –t treated_file.bam –c control_file.bam --format BAM --name output_file_name --wig --space 10 -S
The program compares the local enrichment of reads in the samples generated using an antibody against a specific protein (ChIP sample) and that of a control sample (1% input of chromatin).
For MYC and MIZ1 ChIP-seq samples, the --keep-dup option was used (to keep reads that aligns at the same location) and set at 3 and 10, respectively.

MACS generates bed files, where information about the peak localization (start and end), the peak length and summit (highest point of the peak), the p-value, the enrichment over control and the false discovery rate (FDR) are given.
Wiggle files with a resolution of 10bp (--space option) are also produced and can be loaded into the Integrated Genome Browser for visualization.

3.4.4 Peak annotation and overlap

In the Bedtools program, the closestBed function was used to annotate the peaks present in the bed files generated by MACS. closestBed uses the peak region specified in the bed file generated by MACS and assigns it to the nearest transcriptional start site (TSS) present in a reference file. The latter contains also the gene names and was obtained from the UCSC Table Browser (https://genome.ucsc.edu/cgi-bin/hgTables; human genome: hg19, RefSeq or UCSC Genes).
Command: closestBed –a bed_file.txt –b reference_file.txt –t first > output_file.txt
With the output file and the IF function of Microsoft Excel the distance of the peak to the TSS was calculated taking into account the strand orientation.
The intersectBed function was used to determine the peak overlap of two data sets. Two peaks overlap if they have at least 1bp in common (same chromosomal localization for at least 1bp).
Command: intersectBed –a bed_file_a.txt –b bed_file_b.txt –wa –wb > intersect_file.txt
The –wa and –wb options were used to have the information from both input files in the output file.

3.4.5 Tag density calculation and heatmaps

The Seqminer program was used to calculate the density of tags in a specific genomic region/window. A reference file containing the chromosomal coordinates of the regions of interest (TSS or peak summits) is needed as well as bam files containing the sequenced reads from specific ChIP-seq samples. The reference file should be a txt file with a specific column order: 1) chromosome; 2) TSS or peak summit; 3) TSS+1 or peak summit+1; 4) name; 5) name2; 6) strand. The extension of the region to consider is set in the options section and varies according to the purpose. The resolution of the analysis (wiggle step) is adjusted depending on the extension of the region.

For all subsequent analysis the summit of the MYC peaks called in DOX, close to the promoter of a gene (-1.5kb +500bp for TSS) were used.

For calculation of MYC recruitment the extension of the analyzed region was ±100bp from the summits of MYC peaks and the wiggle step was 5bp.

For calculating histone modification and total and Ser5 phosphorylated RNA polymerase II changes, the extension region was ±1000 bp (wiggle step 10bp) from the summit of the MYC peak and -100+300bp (wiggle step 10bp) form the TSS with a close MYC peak, respectively.

For changes of Ser2 phosphorylated RNA polymerase II the extension was ±1000bp (wiggle step 20bp) from the transcription termination site (TES) .

For heatmaps, showing the binding of a protein at the TSS, the extension region was ±5000bp form the TSS and the wiggle step was 50bp.

The output file contains the chromosomal coordinates, the gene name, and the strand from the input files and the number of reads present in each wiggle step in the defined window.

The number of tags of each wiggle step for each condition was summed up to give the occupancy of that protein in that specific region. MYC recruitment and histone and RNA polymerase II changes were calculated as ratio of the occupancies of the proteins in the EtOH and DOX conditions. To avoid 0 tags, 1 tag was added to all occupancies.

For calculation of MYC binding constant, input background signal was subtracted from MYC occupancy in EtOH and DOX, thus 0 could be used as third point. MYC binding constant and maximal occupancy at each target promoter was obtained via non-linear regression analysis based on Michaelis-Menten model in GraphPad Prism.

For heatmaps visualization, the output file was sorted according to the preferred condition (i.e. MYC binding) and loaded into Java Tree View.

3.4.6 Functional analysis of target genes

To functionally annotate genes, the database for annotation, visualization and integrated discovery (DAVID) and Gene Set Enrichment Analysis (GSEA) were used.

For DAVID, the official gene symbol and the default parameters were used.

GSEA was performed with the C2 or C5 gene sets from the MSigDB. The number of permutations was set to 1000, whereas the other parameters were set as default. When the GseaPreranked tool was used, the Enrichment Statistic parameter was set to “classical”.

3.5 Bioinformatics analysis of RNA-seq data

3.5.1 Base calling, quality control and filter and reads alignment

Base calling, quality control and filter was performed by either Elmar Wolf or Carsten Ade as previously described (see 3.4.1).

Reads were aligned to the human reference genome hg19 and bam files were generated as previously described (see 3.4.2).

3.5.2 Identification of differentially expressed genes

The generated bam files were analyzed in R/Bioconductor. The following packages were loaded in R and were needed to identify and annotate differentially expressed genes and to perform statistical analysis: Rsamtools, GenomicFeatures, edger, goseq, biomaRt.

Commands:

```
#E= control sample
#D= treated sample
>txdb=makeTranscriptDbFromUCSC(genome="hg19",tablename="ensGene")
>tx_by_gene=transcriptsBy(txdb,"gene")
>reads_D1=readBamGappedAlignments("../../D1.bam")
>reads_D1=GRanges(seqnames=rname(reads_D1),ranges=IRanges(start=start(r
eads_D1),end=end(reads_D1)),strand=rep("*",length(reads_D1)))
>counts_D1=countOverlaps(tx_by_gene,reads_D1)
>reads_E1=readBamGappedAlignments("../../E1.bam")
>reads_E1=GRanges(seqnames=rname(reads_E1),ranges=IRanges(start=start(r
eads_E1),end=end(reads_E1)),strand=rep("*",length(reads_E1)))
>counts_E1=countOverlaps(tx_by_gene,reads_E1)

>toc1=data.frame(E1=counts_E1,D1=counts_D1,stringsAsFactors=FALSE)

># plot and calculate sample sorrelation
>pdf("../../correlation.pdf")
>par(mfrow=c(4,3))
>plot(log(toc1$E1),log(toc1$D1),pch=20,cex=0.2)
>dev.off()
```

```
>cor(toc1$E1,toc1$D1)
>#
>norm_factors1=calcNormFactors(as.matrix(toc1))
>head(norm_factors1)
>DGE1=DGEList(toc1,lib.size=norm_factors1*colSums(toc1),group=rep(c("E","D")
,c(1,1)))
>disp1=estimateCommonDisp(DGE1)

>pdf("../../disp1_Smearplot.pdf")
>plotSmear(disp1)
>abline(h=c(-1,1),col="blue")
>dev.off()

>disp1_data_frame=as.data.frame(disp1$pseudo.counts)
>head(disp1_data_frame)
>write.csv(disp1_data_frame, file= "../../disp1_data_frame.csv")
>tested1=exactTest(disp1)
># this gives E/D ratio; for D/E ratio write >tested1=exactTest(disp1,pair=2:1)
>tested1_extract=tested1$table
>head(tested1_extract)
>padj=p.adjust(tested1_extract$PValue, method="BH")
>head(padj)
>tested1_extract_padjust=cbind(tested1_extract,padj)
>head(tested1_extract_padjust)
>ensembl=useMart("ensembl")
>ensembl=useMart("ensembl",dataset="hsapiens_gene_ensembl")
>annotated=getBM(attributes=c("ensembl_gene_id","hgnc_symbol","description"),f
ilters="ensembl_gene_id",values=rownames(tested1_extract_padjust),
mart=ensembl)
>head(annotated)
>tested1_extract_padjust_annotated=cbind(tested1_extract_padjust,
rownames(tested1_extract_padjust))
>head(tested1_extract_padjust_annotated)
```

```
>tested1_extract_padjust_annotated=merge(annotated,tested1_extract_padjust_a
nnotated,by.x="ensembl_gene_id",by.y="rownames(tested1_extract_padjust)")
head(tested1_extract_padjust_annotated)
```

3.5.3 Functional analysis of target genes

Functional analysis of gene groups was performed via DAVID using official gene symbol, default parameters and selecting genes that were significantly regulated (threshold depends on the experiment, usually p-adj<0.01).
GSEA was performed using C2 or C5 gene sets form the MSigDB and the library-size-normalized read counts of all genes identified in the RNA-seq.

3.5.4 Heatmaps

The heatmaps, displaying the regulation of genes in samples where MYC levels were manipulated, were done using the gplots package of R. For the siMYC sample and samples with increasing DOX concentration changes in gene expression were calculated relative to the siCTR or EtOH sample, respectively. The selection of genes shown was done based on publically available data sets and based on the expression levels in the RNA-seq experiments (logCPM>0 or rpkm>1).

3.6 Statistics

All statistic analyses were performed in R or Microsoft Excel.
Unless stated differently, data are presented as means with standard deviation as error bars.
To test significant changes in cell size, BrdU-positive cells as well as total and mRNA amount Student's *t*-tests were applied.
For binned plots, genes were sorted and grouped in equally sized bins and the median or mean of each bin is shown in the plot. Linear regression was used to illustrate the data trend and the fitting of the data to the model is given as Pearson's correlation coefficient (r) with the corresponding p-value calculated via a Student's *t*-test. Data were median normalized if applicable.
Data distribution and variance were tested using the Shapiro and the F test, respectively.
Box plots were done using R and the Mann-Whitney test or Student's *t*-test were applied to statistically test the difference of the samples.

Chapter 4:

Results

4.1 MYC activation establishes tumor-cell-specific gene expression profiles

4.1.1 Characterization of MYC-induced U2OS

MYC is an important transcription factor that is deregulated in many tumors (Vita and Henriksson, 2006). Although MYC is extensively studied, discrepancies have emerged about how this transcription factor regulates its target genes and whether it regulates specific processes or acts as a general transcription factor. Indeed, when MYC is manipulated in medulloblastoma tumor models, it both activates and represses target genes that are specific for MYC's transcriptional program and have prognostic value (Kawauchi et al., 2012; Northcott et al., 2011). On contrary, as shown by two studies in primary B cells and in an engineered lymphoma model, MYC overexpression leads to activation of all genes that are found in an open chromatin context defining MYC as an general amplifier of gene expression (Lin et al., 2012; Nie et al., 2012).

In order to understand the discrepancies between these models, gene expression and binding analyses were performed in an osteosarcoma cell line (U2OS). U2OS were chosen since they are a tumor cell line that express lower levels of endogenous MYC than others tumor cell lines, such as HeLa and HCT116, comparable to non-transformed cells, such as some epithelial cell lines (IMECs, HMLE, MCF10A and HEK293) (Fig. 4.1 A).

U2OS cells were engineered to express a doxycycline-inducible allele of MYC (by Elmar Wolf). By adding doxycycline to the culture medium an about 15 fold induction of MYC protein levels and ~100 fold induction of mRNA levels could be reached (Fig. 4.1 B,C). Overexpressed MYC was correctly localized in the nucleus, as shown by immunofluorescence (Fig. 4.1 D). As expected, MYC overexpression led to an increase in the proportion of 5-bromodeoxyuridine (BrdU)-positive cells in a BrdU-PI (propidium iodide) FACS analysis indicating that there are more cells in S-phase in the MYC-induced situation (Fig. 4.1 F).

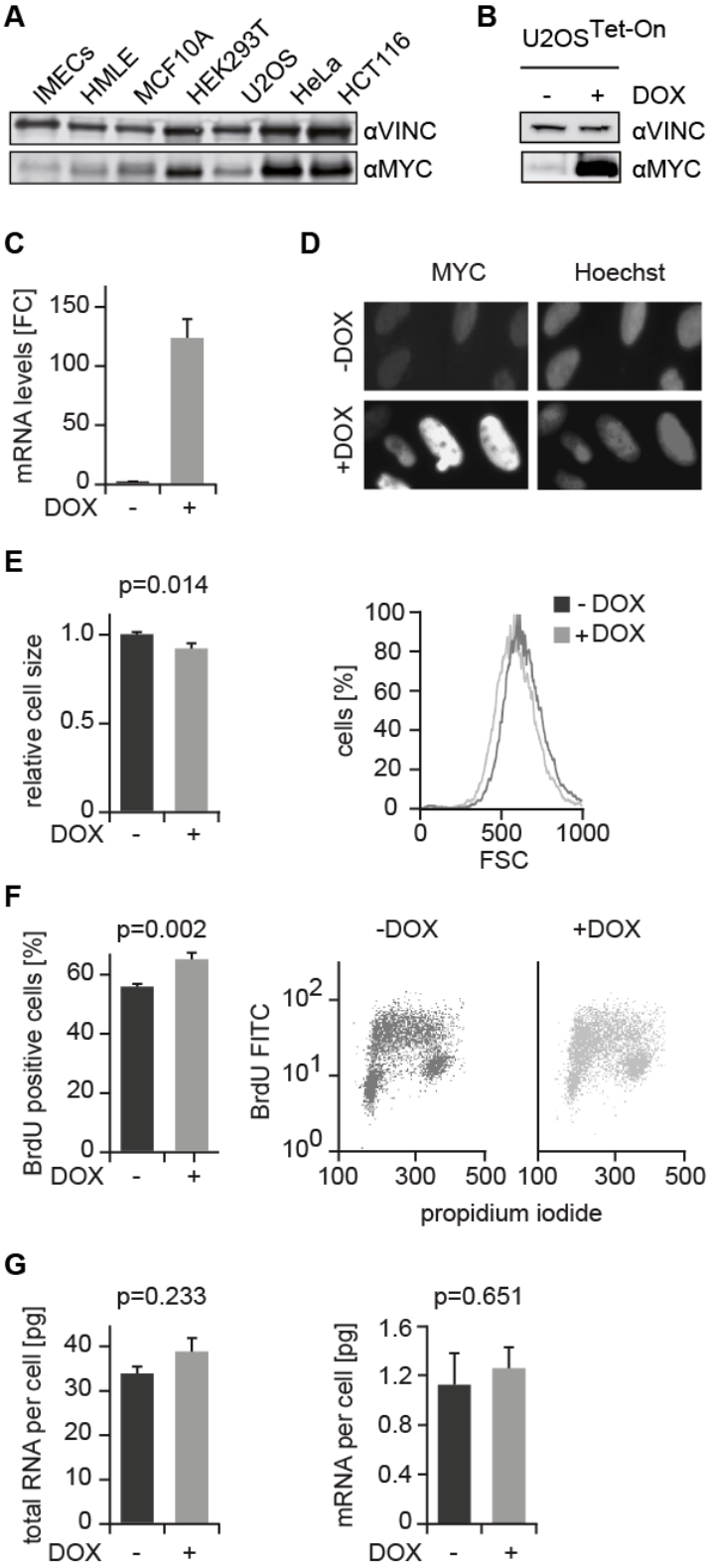


Figure 4. 1: Overexpression of MYC does not increase cell size and total and mRNA amount.[1] **A.** Immunoblot showing MYC protein levels in different cell lines. The same number of cells were loaded in the

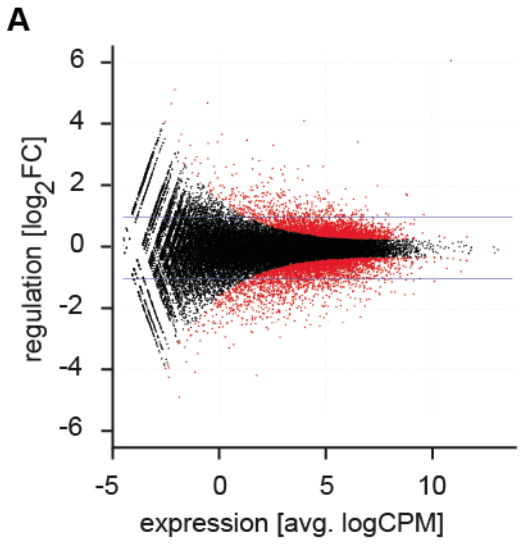
[1] This figure was published in similar form in (Walz et al., 2014) (see also follwing pages).

each well. **B.** Immunoblot of MYC levels in U2OS cells treated with EtOH (-DOX) or 1ng/µl DOX (+DOX) for 30h.Vinculin was used as loading control. **C.** RT-qPCR analysis showing induction of MYC mRNA in U2OS cells treated as in B. **D.** Immunofluorescence using anti-MYC antibody and Hoechst staining in U2OS cells treated as in B. **E.** Cell size of U2OS cells with and without MYC induction measured by FACS using the forward scatter (FSC) in arbitrary units. Error bars show standard deviation (s.d., n=3). **F.** Percentage of BrdU positive cells (right) and FACS analysis (left) of U2OS cells induced or not with DOX. **G.** Total (left) and mRNA (right) amount per cell in U2OS cells before and after MYC induction. Error bars represent standard error of the mean (s.e.m., n=4).

Opposite to what it was observed in B cells (Lin et al., 2012; Nie et al., 2012), MYC overexpression in U2OS cells increases neither cell size, as measured by FACS (Fig. 4.1 E), nor total and mRNA levels within a cell (Fig. 4.1 G).
The absence of changes in cell size, total and mRNA amount allowed the use of RNA-seq to monitor gene expression upon MYC induction.

4.1.2 MYC overexpression induces stereotypic gene expression changes

Total RNA was extracted from U2OS cells treated with EtOH (-DOX) or with 1ng/µl DOX (+DOX) for 30h to monitor gene expression changes via RNA-seq upon MYC induction. To get rid of the rRNA fraction that represents the majority of the RNA present in a cell, magnetic beads covered with oligo d(T) were used. Therefore, only the polyadenylated fraction of RNA was bound by the beads and the mRNAs were enriched in the samples. The isolated mRNAs were fragmented, end-repaired and adaptors were ligated to allow the hybridization of the cDNA obtained from the mRNA to the sequencing flow cell (Illumina). Three replicates for each condition were sequenced and the resulting reads were aligned to the reference human genome hg19 using Bowtie.
By using the edgeR package of R, 1,358 significantly regulated genes (q-value<0.01, $\log_2$FC>1 or <-1) were identified (Fig. 4.2 A). Among these, 462 genes were up-regulated with a $\log_2$FC>1 and 896 down-regulated with a $\log_2$FC<-1 upon MYC-induction.
Gene set enrichment analysis (GSEA) using the normalized reads count of all genes identified in the RNA-seq and the C2 curated gene sets collection, showed that MYC-regulated gene sets in U2OS cells were enriched in genes described as MYC targets in previously published studies (Fig 4.2 B). The "Kim *MYC* amplification targets up" is a gene set that includes genes specifically regulated in lung cancer samples with MYC amplification compared to non-amplified ones and was also enriched (Fig. 4.2 C) validating the U2OS cell system used for the experiment.



B

activated in response to MYC induction

gene set	NES	p-value	q-value
Schuhmacher MYC targets up	2.46	0.000	0.000
Schlosser MYC targets	2.24	0.000	0.000
Coller MYC targets up	2.16	0.000	0.000
Wei MYCN targets with E-box	2.11	0.000	0.002

repressed in response to MYC induction

gene set	NES	p-value	q-value
Dang regulated by MYC down	-2.15	0.000	0.000
Cowling MYCN target	-2.05	0.000	0.000

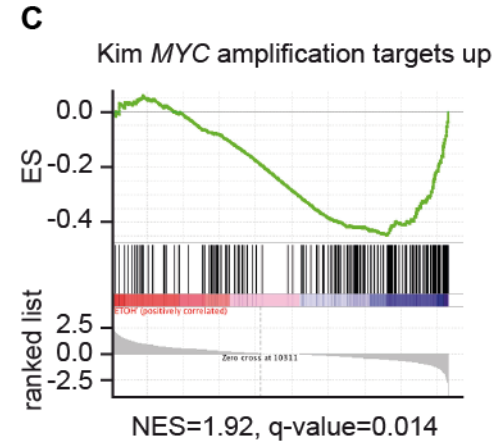


Figure 4. 2: MYC-induced gene expression changes are stereotypic[1]. A. Plot showing the regulation of genes upon MYC induction versus total expression levels. Red dots represent significantly regulated genes (q-value<0.01, n=3). **B.** GSE analysis of genes regulated upon MYC overexpression in the C2 gene sets. **C.** Enrichment plot of one gene set identified in the GSE analysis in B.

4.1.3 MYC binds to thousand of sites in U2OS cells

In parallel to gene expression analysis via RNA-seq, ChIP-seq was performed to map MYC binding sites. DNA and bound proteins were crosslinked using formaldehyde and the crosslinked cells were sonicated to reach nucleosomal size of the DNA fragments. Chromatin from cells before and after MYC induction was precipitated by using an anti-MYC antibody (N262, Santa Cruz) or IgG as control for specific binding. A percentage of the input chromatin (1%) was kept as further control and for normalization (in case of qPCR).

As for RNA-seq, the DNA recovered from the immunoprecipitation was quantified, end-repaired and ligated to adaptors in order to allow hybridization the sequencing surface.

The sequenced reads were aligned to the reference human genome hg19 using Bowtie and significant peaks were identified with MACS software and filtered according to the false discovery rate (FDR<0.1) calculated by the program.

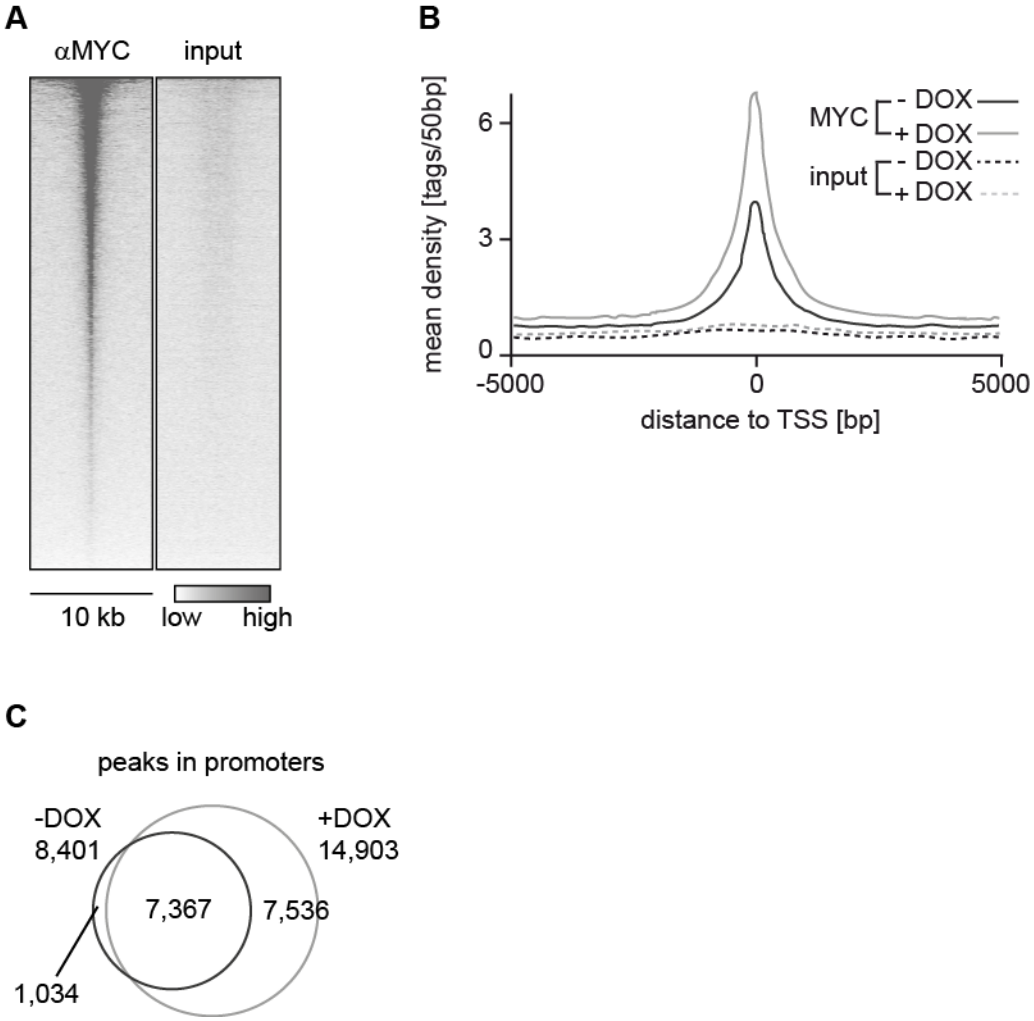


Figure 4. 3: MYC binds to promoter and enhancer regions[1]. A. Heat map of MYC binding in U2OS cells with endogenous MYC levels in a region of ±5000bp from the TSS of all genes present in the UCSC database. **B.** Density plot of MYC binding in the + and –DOX conditions. Input was used as control. The chromosomal location was used for the x-axis. **C.** Venn diagram of MYC-detected peaks in the promoters of genes (-1.5kb + 500bp from the TSS).

20,014 MYC binding sites were identified genome-wide in cells non-induced with DOX. In this context, MYC binding was highly enriched around the TSS of genes (Fig. 4.3 A). When MYC was induced, binding increased genome-wide (Fig. 4.3 B) as well as the number of binding sites (45,645 MYC peaks identified in DOX-treated cells). In promoters, defined as the region between -1.5kb and +500bp from the TSS, the number of detected peaks increased from 8,401 to 14,903 (Fig 4.3 C). The majority of peaks detected

in the –DOX condition overlapped the ones identified in DOX-treated cells suggesting that the biological noise of the experiment was quite low (Fig 4.3 C). Furthermore, inspection of single genes (see Fig. 4.7) showed that the extent of MYC binding and recruitment is not the equal among the overlapping peaks.

4.1.4 Saturation is detected at certain MYC binding sites

In order to analyze the MYC binding data in a genome-wide manner without focusing on a limited number of genes, binned plots were used. For each gene having a MYC peak in the promoter (as defined before) in DOX treated cells the number of MYC tags present in a region of ±100bp around the summit of the peak (occupancy) was calculated using the Seqminer program for the + and –DOX conditions. Furthermore, the relative MYC recruitment for each gene was calculated by dividing MYC occupancy in the induced and non-induced condition. In this way over 9000 genes were included in the analysis.

Genes were sorted according to the MYC recruitment and divided in equally-sized bins and for each of these bins the mean for MYC recruitment and occupancies were calculated and plotted (Fig. 4.4 A).

In the –DOX condition (cells with endogenous MYC levels), MYC occupancy decreases with the increase of MYC recruitment, indicating that genes that are highly occupied by MYC recruit less MYC when overexpressed (MYC recruitment of about 1). Vice versa, genes that have weak MYC binding recruit more MYC when induced with DOX (Fig 4.4 A). When looking at the same plot in the +DOX situation, the MYC occupancy seems not to change with the recruitment. This suggests that when MYC is induced, it binds all genes to the same extent and it tends to fill up former weakly occupied genes. Moreover, the slight difference in MYC occupancy at endogenous and exogenous levels for highly occupied genes suggests that these genes are saturated for binding already at endogenous MYC levels and therefore could not recruit further MYC when overexpressed. To test this hypothesis, MYC and MXD6 (MNT) ChIPs and qPCRs were performed at selected genes. MNT was chosen since it competes with MYC for binding with MAX and for binding to the same DNA sites (Ayer and Eisenman, 1993; Bouchard et al., 2001; Xu et al., 2001). If saturation of MYC binding occurs already at endogenous levels, no MNT signal and no changes in MYC levels in the + and –DOX conditions should be detected at the same site. This was the case of the *NPM1* and *NCL* genes that were both highly bound by MYC. Their MYC binding did not change upon overexpression and MNT signal was at background levels (comparable to IgG and control region signal) in both conditions

indicating the absence of binding (Fig 4.4 B). For other genes, such as *HSPBA1* and *FBXO32*, MYC binding highly increased upon MYC overexpression and concomitantly MNT signal decreased indicating that these genes were not saturated for MYC binding.

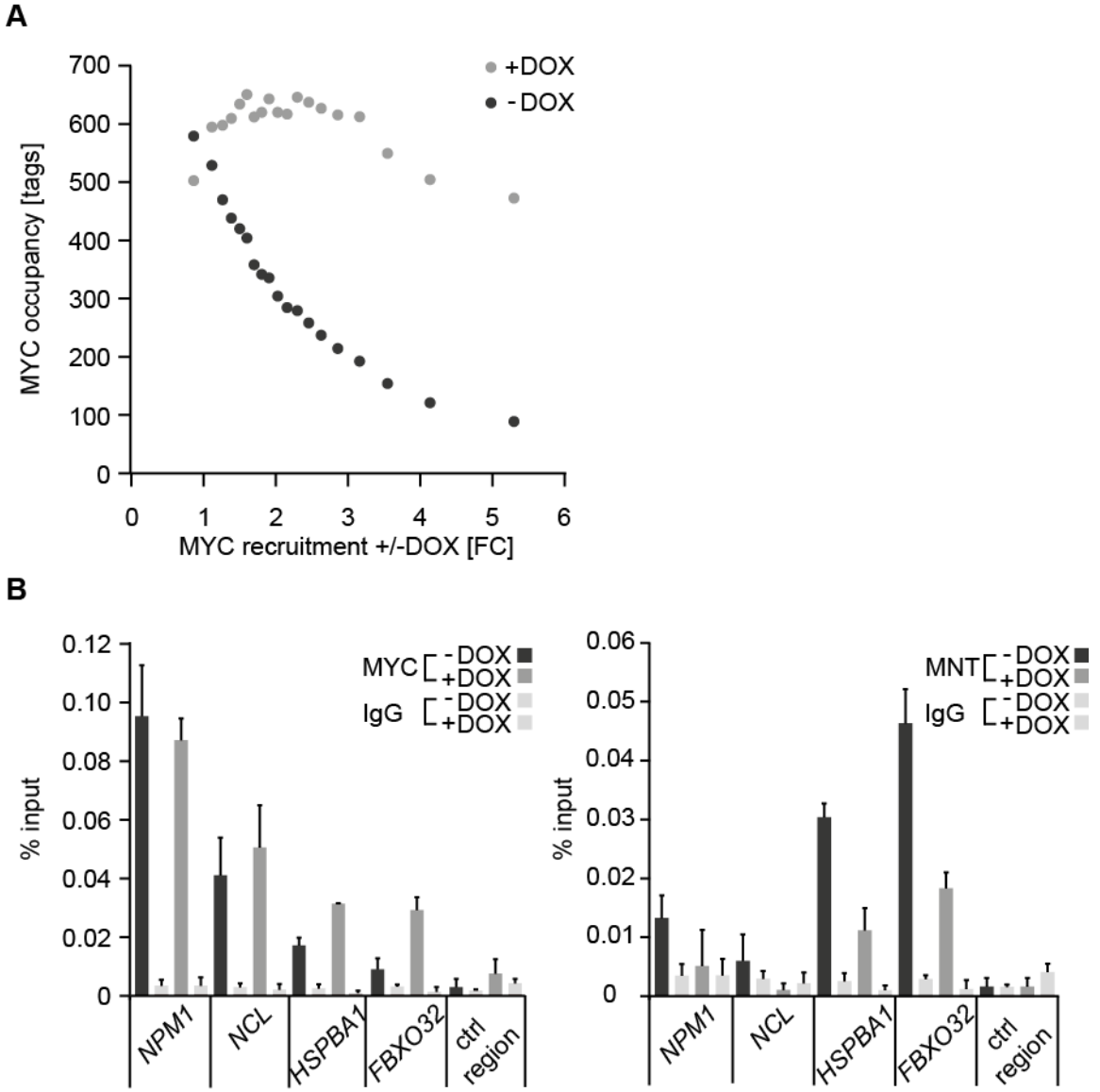
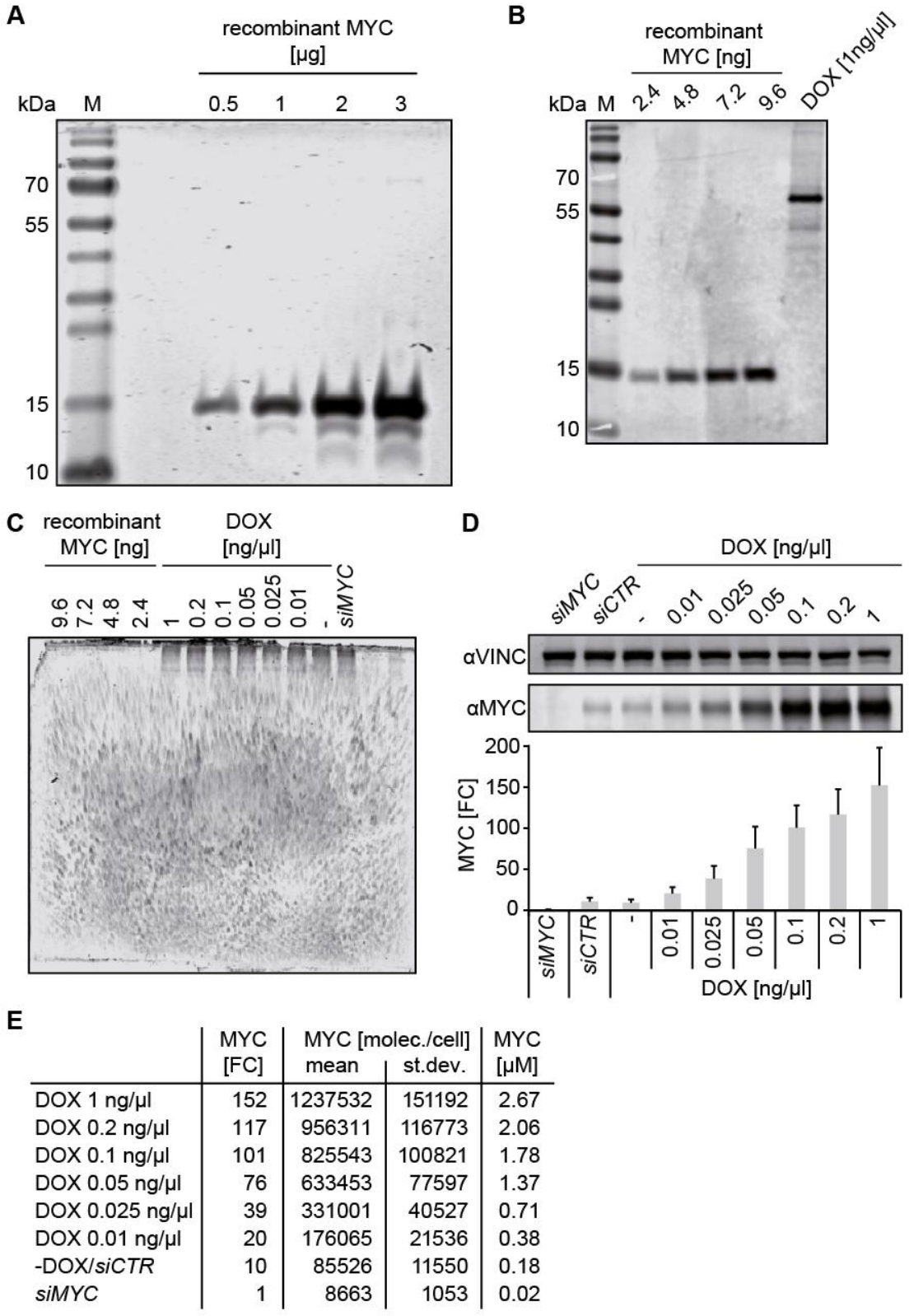


Figure 4. 4: Saturation occurs at certain MYC binding sites[2]. A. Binned plot of MYC recruitment versus MYC occupancy in +DOX and –DOX conditions. Genes were sorted according to recruitment and divided in 20 equally-sized bins. Each dot represents the average value of 422 genes. **B.** Bar plot for MYC and MNT ChIP experiments followed by qPCR. IgG and a control region were used as controls. Data are shown as mean ± standard deviation of technical triplicates.

[2] This figure was published in similar form in (Lorenzin et al., 2016) (see also follwing pages).



E

	MYC [FC]	MYC [molec./cell] mean	MYC [molec./cell] st.dev.	MYC [µM]
DOX 1 ng/µl	152	1237532	151192	2.67
DOX 0.2 ng/µl	117	956311	116773	2.06
DOX 0.1 ng/µl	101	825543	100821	1.78
DOX 0.05 ng/µl	76	633453	77597	1.37
DOX 0.025 ng/µl	39	331001	40527	0.71
DOX 0.01 ng/µl	20	176065	21536	0.38
-DOX/*siCTR*	10	85526	11550	0.18
siMYC	1	8663	1053	0.02

Figure 4. 5: The number of MYC molecules per cell is compatible with the saturation of binding sites[2]. **A.** Coomassie staining of MYC recombinant protein. **B.** Immunoblot of MYC recombinant protein and MYC detected in U2OS cells treated with 1ng/µl DOX. The same number of cells were loaded for each sample. **C.** Coomassie staining of a gel used for immunoblotting after the transfer of proteins to the PVDF membrane. **D.** Immunoblot (upper panel) of MYC in U2OS treated with siMYC, siCTR, EtOH and different DOX

concentrations. Vinculin was used as loading control. After quantification of the immunoblots (lower panel), fold change (FC) of MYC levels was calculated relative to siMYC and values are shown as mean with error bars representing standard deviation of biological triplicates. **E.** Table for MYC protein levels calculated as fold change (FC), number of molecules per cell and concentration in the nucleus.

To establish whether the amount of MYC present in a cell was compatible with the saturation model proposed, the number of MYC molecules was measured in U2OS cells before and after DOX induction. This was achieved by using a recombinant protein consisting only of the C-terminal part of MYC (Fig. 4.5 A, provided by Lisa Jung), which was immunoblotted together with whole cell lysates from U2OS cells where MYC levels were either manipulated or not. To detect the recombinant protein the 9E10 anti-MYC antibody was used. Unfortunately this antibody was not sensitive enough to detect MYC in lysates from cells with endogenous MYC levels. Therefore the protein lysate obtained from U2OS cells treated with the highest DOX concentrations (0.1, 0.2 and 1ng/µl) was blotted together with different amounts of the recombinant protein (Fig 4.5 B). Several lysates coming from U2OS cells treated with either an siRNA against MYC, a control siRNA, EtOH or different DOX concentrations were loaded on an immunoblot and the fold change of MYC levels and the number of MYC molecules per cell were calculated relative to DOX 0.1, 0.2 or 1ng/µl-treated cells (Fig 4.5 D). Since the recombinant and the cellular MYC highly differ in size, to avoid any underestimation of the number of MYC molecules per cells, the polyacrylamide gels used for the immunoblots were checked via Coomassie staining for residual proteins present after the transfer to the PVDF membranes was completed (Fig. 4.5 C). Following this procedure, 85,526 (+/- 11,550) MYC molecules where detected to be present in –DOX/EtOH treated U2OS cells and 1,237,532 (+/- 151,192) in 1ng/µl DOX treated cells, the two conditions used for the ChIP-seq (Fig. 4.5 E). These calculated values were compatible with the saturation model since the number of MYC molecules in EtOH and DOX would be high enough to saturate all the binding sites detected via ChIP-seq in the two conditions. By using the nuclear volume of U2OS cells (Koch et al., 2014), the concentration of MYC in the nucleus was determined (Fig. 4.5 E).

4.1.5 At supraphysiological levels, MYC binds to low affinity sites and weakly expressed genes

As previously shown (Fig 4.5), MYC seems to preferentially occupy different genes according to its protein levels in the cell.

To understand what determines the difference between a site that is occupied already at low MYC levels and another one that becomes occupied only at high and supraphysiological levels some features that could affect MYC function were examined. Among these, it is well established that MYC preferentially binds to specific DNA sites called E-boxes (Blackwell et al., 1990, 1993). Since the E-box CACGTG (canonical) has the highest affinity for MYC binding (Blackwell et al., 1993; Hu et al., 2005; Sauvé et al., 2007), the canonical E-box content was checked in MYC bound genes. Binned plots were used and the number of genes having a canonical E-box (CACGTG) in a region of ±100bp from the summit of the MYC peak were counted in each bin. Before binning, genes were sorted either according to MYC occupancy in -DOX or MYC recruitment.

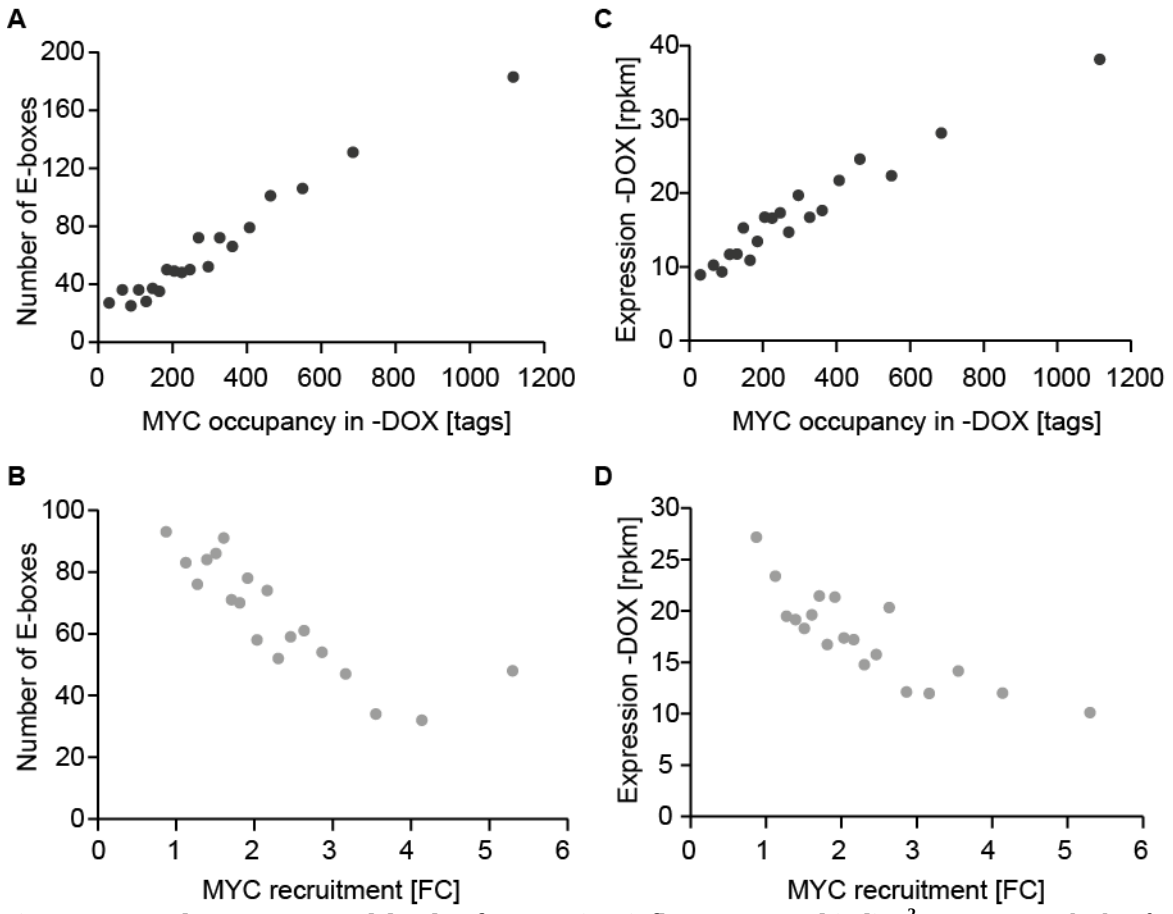


Figure 4. 6: E-box content and levels of expression influence MYC binding[2]. A-B. Binned plot for number of genes in each bin having a canonical E-box (CACGTG) versus MYC occupancy in EtOH (A) or MYC recruitment (B). Each dot represent mean of 422 genes. **C-D.** Binned plot as in A-B, but the mRNA expression of the respective gene (rpkm= reads per kilobases per million mapped reads) is shown.

By sorting according to endogenous MYC occupancy it could be seen that the genes highly bound by MYC have a higher number of canonical E-boxes compared to lowly bound ones

(Fig. 4.6 A). This suggests that at physiological levels MYC preferentially binds to high affinity binding sites. On the other hand, when sorting by MYC recruitment and therefore having a representation of where the overexpressed MYC is mainly bound, the number of canonical E-boxes decreases with the recruitment (Fig. 4.6 B). This indicates that at supraphysiological levels MYC is preferentially recruited to low affinity binding sites. Hence, canonical E-box content is a determinant for MYC binding only at low or endogenous MYC levels whereas at higher MYC levels other factors could be important and necessary to recruit MYC.

One of such factor could be the expression levels of the target genes. Indeed, Nie and colleagues showed that a positive correlation between the extent at which a gene is expressed and the strength of MYC binding exists (Nie et al., 2012). Given that the ChIP-seq and RNA-seq analyses were performed in parallel in U2OS cells, the expression of a specific gene could be correlated to the respective MYC binding. As measurement of the expression of a specific gene, the rpkm (reads per kilobases per million mapped reads) value was used since it takes the number of sequenced reads from the RNA-seq and normalizes them for library size and for the length of the specific transcript. Therefore, the overrepresentation of the longer transcripts is avoided. As for the E-box analysis, a positive correlation was seen between expression of a gene and MYC occupancy in –DOX, whereas a negative correlation was detected between expression and MYC recruitment. This argues that at endogenous levels, MYC is bound to highly expressed genes whereas upon increasing levels, MYC binds to lowly expressed genes. Therefore the level of expression of a certain gene is another feature influencing MYC binding.

4.1.6 MYC recruitment determines gene regulation

To investigate whether the difference in binding properties of MYC upon changes in protein levels translates in the regulation of different target genes, MYC binding data were correlated with the gene expression data obtained from RNA-seq. By dividing genes in activated and repressed, a positive correlation between the strength of the regulation and MYC recruitment was observed for both groups (Fig. 4.7 A). Genes that were already strongly bound by MYC at endogenous levels and did not recruit further MYC (MYC recruitment of about 1) upon overexpression were also weakly regulated by MYC (Fig. 4.7 A and B, *RPL8* as example).

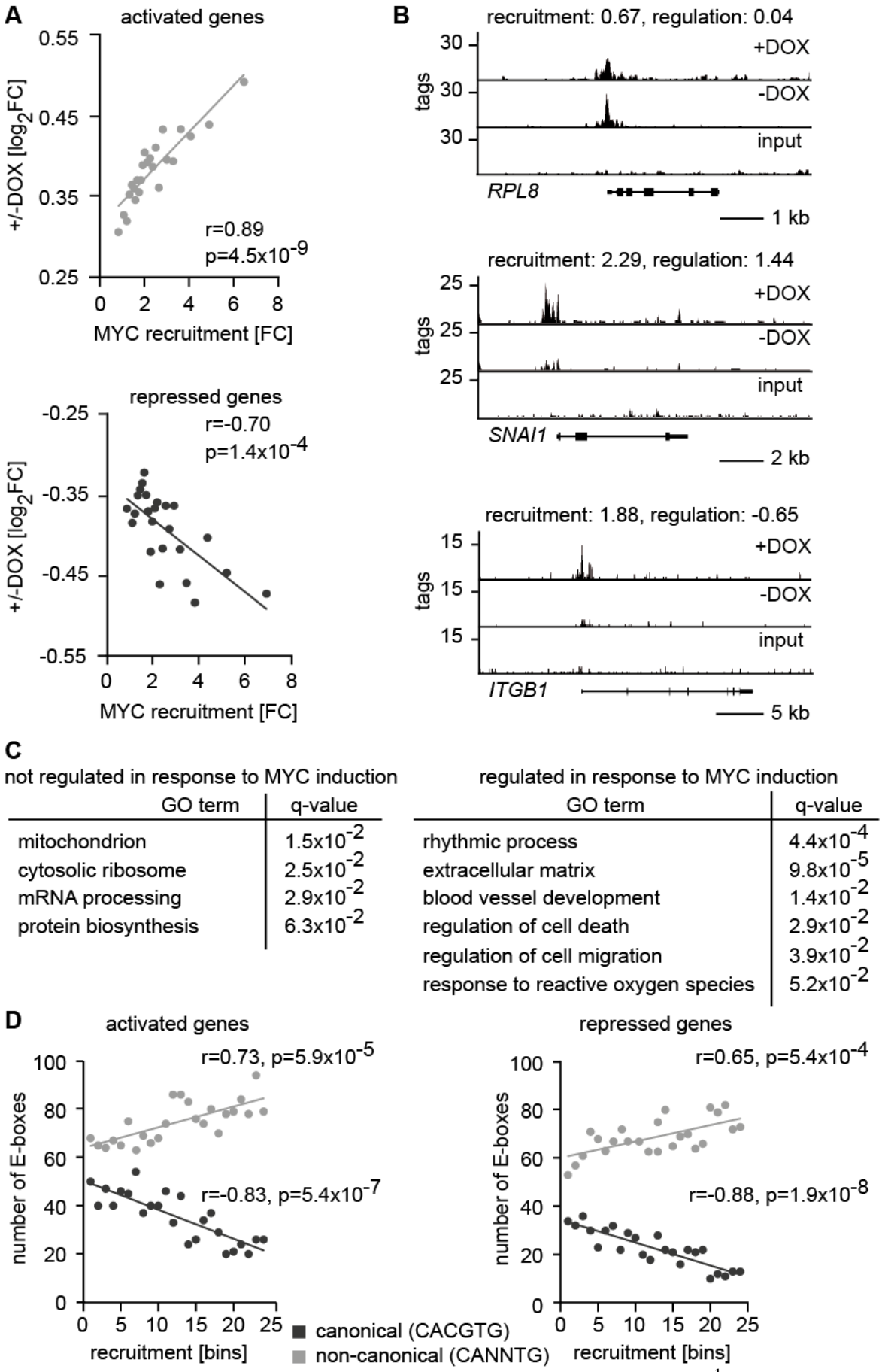


Figure 4. 7: Different MYC levels control functionally different cellular processes[1]. A. Binned plot for

MYC recruitment and gene expression regulation. 5,063 activated genes and 4,485 repressed genes were sorted according to MYC recruitment and divided in 20 equally sized bins. r indicates the Pearson correlation coefficient and p-values were calculated using a Student's t-test. **B.** Example of gene tracks for MYC binding. **C.** DAVID analysis of non regulated (left) and regulated (right) genes in response to MYC induction. **D.** Binned plot for the number of genes containing canonical and non canonical E-boxes versus recruitment. r represent the Pearson correlation coefficient and p-value were calculated using Student's t-test.

On contrary, activated and repressed genes showing a high MYC recruitment were more strongly regulated upon MYC overexpression (Fig. 4.7 A and B, *SNAI1* and *ITGB1* as examples). DAVID analysis indicated that different functional categories were stratified according to MYC recruitment and were therefore differentially regulated by different MYC levels. In fact genes with a recruitment fold change of about 1 and weakly regulated were functionally enriched for genes encoding proteins involved in mitochondrial function, ribosome biosynthesis, RNA processing and protein biosynthesis. On the other hand, genes involved in extracellular matrix metabolism, blood vessel development, regulation of cell migration, ROS metabolism etc were enriched among genes with high MYC recruitment and strongly regulated.

As shown before, the number of genes containing a canonical E-box in the MYC peak present in their promoter decreased with the recruitment whereas the number of genes containing a non-canonical E-box (CANNTG other then CACGTG) increased. This suggests that at low levels MYC binds to and saturates genes with a high affinity binding sites (canonical E-boxes), regulating mainly cell-growth related processes. When MYC levels increase, the high affinity binding sites are already fully occupied and MYC binds to and regulates expression of genes with low affinity sites (such as non canonical E-boxes), controlling processes that could be cell and tumor specific. Moreover, these analyses also argue that, since MYC induced-gene expression activation is thought to be direct and the same behavior (Fig. 4.7 A and D) was observed for activated and repressed genes, MYC also directly represses target genes.

4.1.7 Promoter affinity for MYC binding stratifies functionally distinct processes

Since it is difficult to determine a single factor that accounts for the affinity of a promoter to MYC, a general estimation of the MYC binding affinity to a certain gene was calculated by using the nuclear MYC concentrations and the MYC occupancy obtained via ChIP-seq and subsequent analysis.

The background input signal was subtracted from the occupancy values and thus the 0,0 value could be used as third point in order to perform the fitting of the data, using the non-linear regression method. This led to the calculation of two parameters for every gene bound by MYC in the promoter region: the maximal occupancy and, borrowing a term from pharmacology, the EC_{50} value that represent the MYC concentration needed to have half of the maximal occupancy (Fig. 4.8 A and B). As the Michaelis-Menten constant for enzymes and the dissociation constants obtained via other biochemical methods, the EC_{50} is a rough estimation of the affinity, in this case, of a certain gene to MYC (Fig. 4.8 A and B).

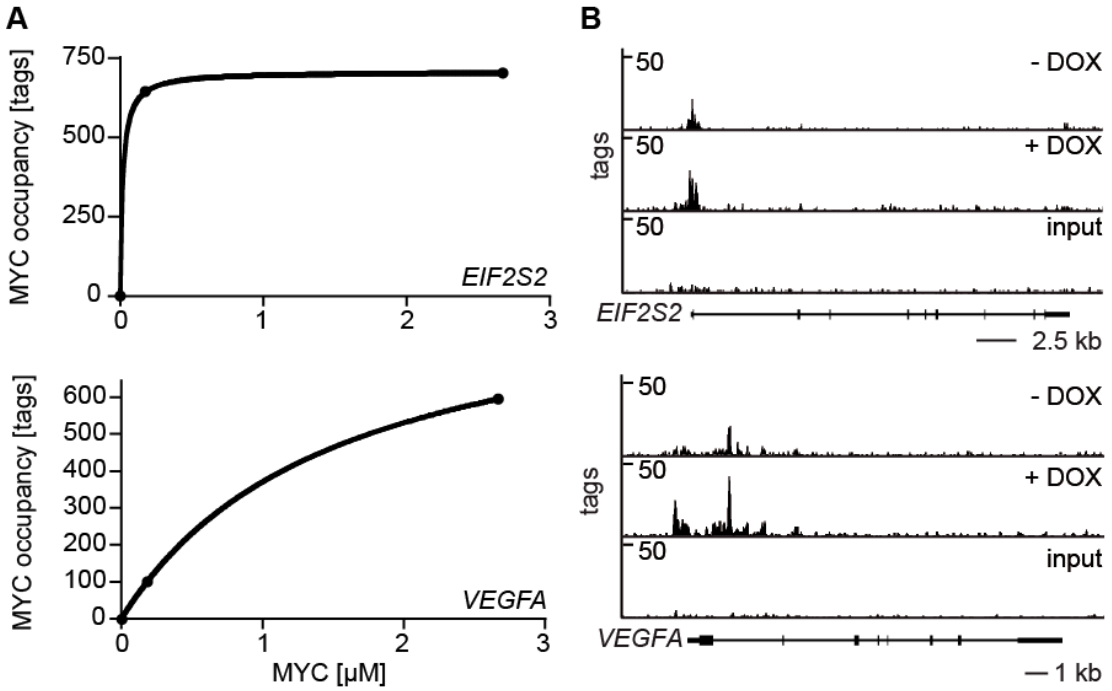


Figure 4. 8: Calculation of promoter binding affinity to MYC[2]. A. Diagram for MYC concentration and MYC occupancy in U2OS cells treated with DOX or EtOH. The line was fitted using non-linear regression. **B.** Gene tracks for MYC binding shown in A. Input is showed as control.

For example *EIF2S2* is a gene equally bound by MYC both in + and –DOX conditions and thus the calculated value is much lower than *VEGFA*, whose MYC binding increased considerably upon induction.

Analysis of the distribution of the EC_{50} values via a density plot showed that the majority of the genes had EC_{50} values between the endogenous MYC concentration in U2OS cells (-DOX) and the MYC concentration present when MYC was depleted (siMYC). This suggests that these genes are close to saturation at low MYC levels (Fig. 4.9). However,

many genes have an EC_{50} higher than the -DOX MYC concentration indicating that these genes need higher MYC levels to be highly bound and saturated by MYC (Fig. 4.9).

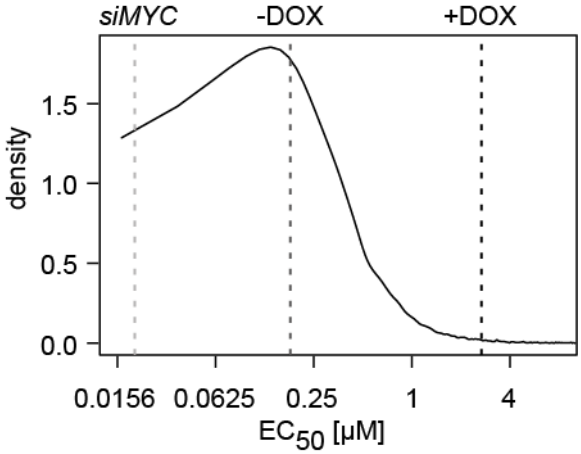


Figure 4. 9: Distribution of EC_{50} values[2]. Density plot of all the EC_{50} values calculated by using the MYC concentrations and occupancies in + and -DOX treated U2OS cells. The dashed lines indicate the nuclear MYC concentration calculated in the indicated conditions in U2OS cells. The x-axis is shown in a logarithmic scale.

All MYC bound genes whose fitting could be performed by the Prism program were sorted according to the EC_{50} values and functional analysis was performed. By using the GseaPreranked tool the C5 gene set collection was investigated. Genes with low EC_{50} and thus high binding affinity to MYC were enriched in growth-related processes encoding structural constituents of ribosome, proteins involved in translation, RNA binding and other cellular biosynthetic processes (Fig 4.10 A and B left panel). On the other side, genes with high EC_{50} and thus low affinity for MYC, were enriched for genes encoding for proteins involved in processes that could be cancer-related such as activity related to G protein-coupled receptors, specific transporters, TGF-β signaling and the response to hypoxia (Fig. 4.10 A and B right panel). Taken together these results indicate that promoter affinity to MYC stratifies/differentiates different cellular processes.

A

low EC_{50}

	gene set	NES	p-value	q-value
1	structural constituent of ribosome	-3.46	0.000	0.000
2	translation	-3.07	0.000	0.000
3	RNA binding	-2.95	0.000	0.000
4	cellular biosynthetic process	-2.92	0.000	0.000

high EC_{50}

	gene set	NES	p-value	q-value
7	G protein coupled receptor activity	2.48	0.000	0.008
9	substrate specific transporter activity	2.37	0.000	0.014
11	transforming growth factor beta receptor signaling pathway	2.21	0.002	0.036
16	response to hypoxia	2.92	0.008	0.095

B

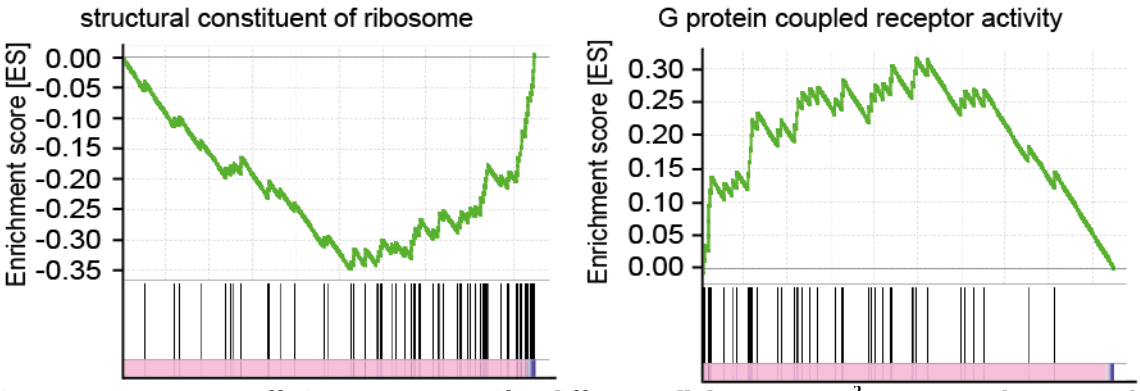


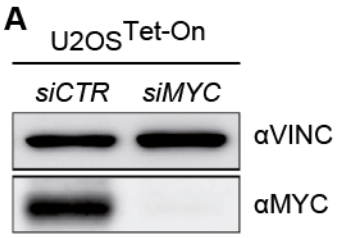
Figure 4. 10: Promoter affinity to MYC stratifies different cellular processes[2]. A. GSE analysis using the Preranked tool of 9,500 genes sorted according to the EC_{50} values. The C5 gene set collection was used. **B.** Example of enrichment plots form the GSE analysis in A.

4.1.8 Promoter affinity for MYC correlates with differential regulation of the corresponding gene

Genes encoding proteins belonging to functionally different classes are characterized by different MYC binding affinity. To show that genes having different EC_{50} and thus different affinity to MYC are actually differentially regulated by changes in MYC levels, RNA-seq was performed in U2OS cells at different MYC levels.

To identify the genes that would respond to endogenous or low MYC concentrations, U2OS cells were transfected with a pool of siRNA against MYC or control siRNAs. MYC depletion was very efficient (Fig. 4.11 A). RNA-seq was performed and the C5 gene set collection containing GO terms was investigated via GSE analysis. Among the genes that were most strongly regulated upon MYC depletion, genes encoding for proteins involved in ribosome biogenesis, RNA binding and processing and in mitochondrial metabolism were enriched (Fig. 4.11 B). Moreover, these genes (structural constituent of ribosome was

used as example) were strongly regulated only upon MYC depletion and not upon MYC overexpression (Fig. 4.11 C).



B

	gene set	NES	p-value	q-value
1	structural constituent of ribosome	2.64	0.000	0.000
2	nucleolus	2.43	0.000	0.000
3	ribonucleoprotein complex	2.38	0.000	0.000
4	ribonucleoprotein complex biogenesis and assembly	2.28	0.000	0.000
5	RNA binding	2.16	0.000	0.000
6	RNA processing	2.14	0.000	0.000
7	mitochondrial matrix	2.12	0.000	0.000

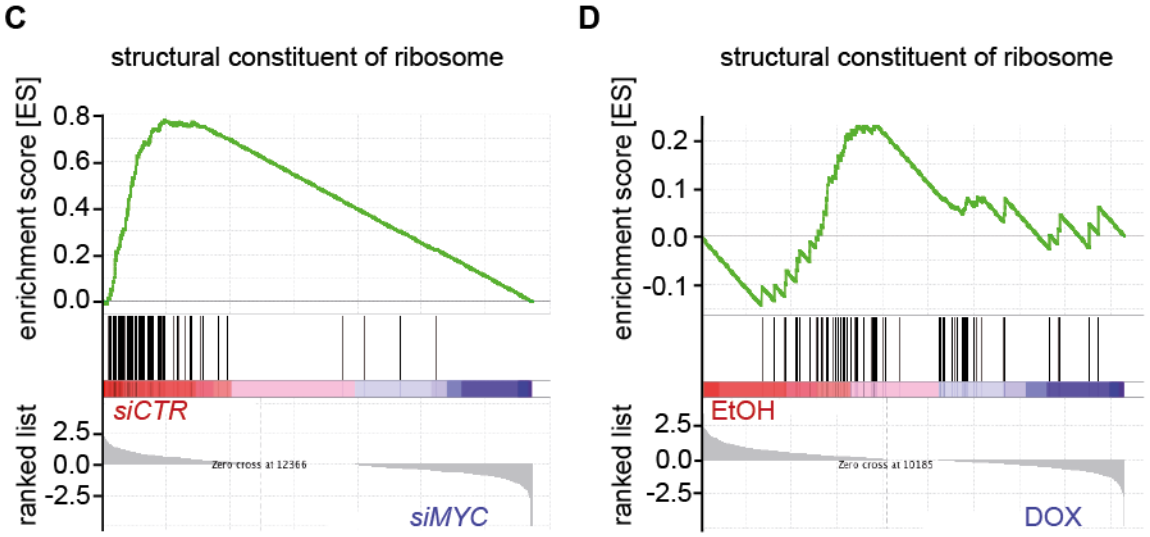


Figure 4. 11: MYC depletion regulates gene sets involved in cellular growth[2]. A. Immunoblot for MYC in U2OS cells transfected with siRNA against MYC or control siRNA. VINCULIN was used as loading control. **B.** GSE analysis using the C5 collection of genes identified via RNA-seq of cells used in A. **C-D.** Enrichment plots of one gene set from GSE analysis in B and GSEA from U2OS cells treated with EtOH, as control, or DOX to induce MYC.

The promoter affinity of the gene sets regulated by MYC depletion was compared with the one of genes regulated upon MYC overexpression (Walz et al., 2014). Furthermore a set containing genes regulated by MYC knockout in 3T9 fibroblasts was also used (Perna et al., 2012). Interestingly, also in this system the induced deletion of endogenous MYC (CreER-loxP sites system) led to the regulation of cellular-growth-related processes such as nucleotide metabolism, ribosome biogenesis and translation, RNA/rRNA/tRNA processing etc. (Perna et al., 2012).

This analysis showed that the EC_{50} values of the genes regulated by MYC depletion in U2OS (Fig. 4.12 dark grey) was comparable to the one of the fibroblast's gene set (Fig.

4.12 black) and lower than the gene sets regulated by MYC overexpression (Fig. 4.12 light grey). This argues that the genes regulated by low or physiological MYC levels, mimicked in this case by MYC depletion, have higher affinity for MYC than genes regulated only when MYC levels increase and that the high affinity genes are enriched for genes encoding growth-related proteins.

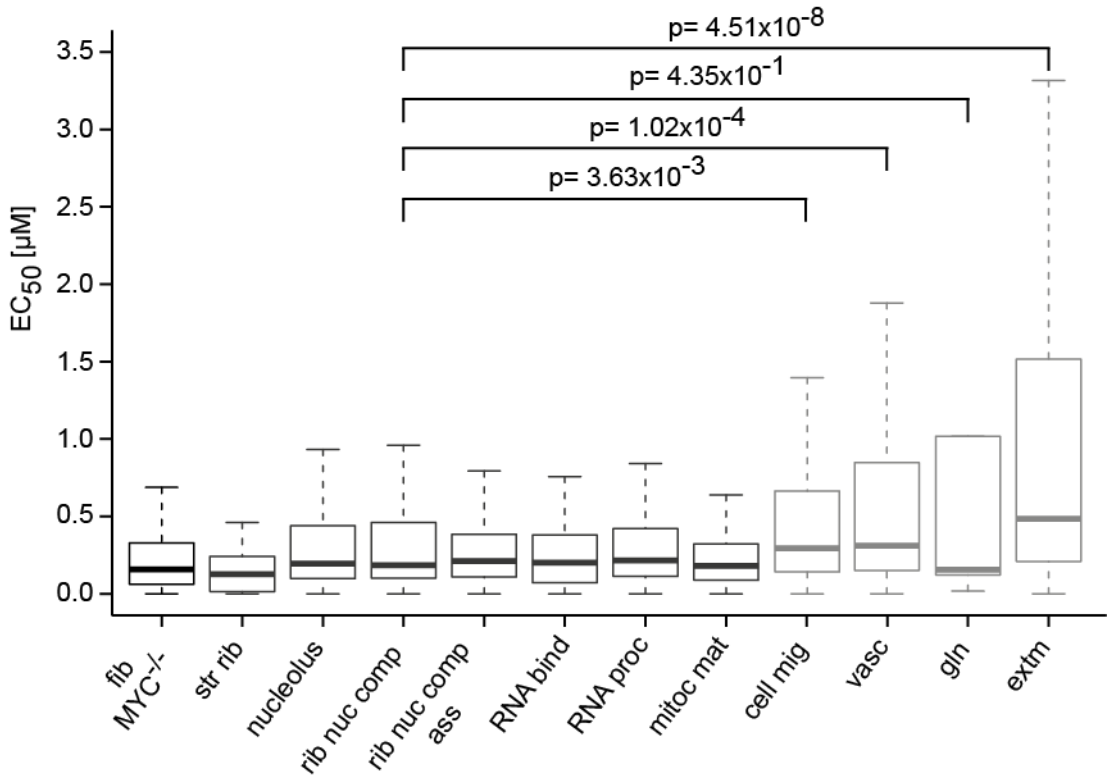


Figure 4. 12: Growth-related processes have the lowest affinity for MYC binding[2]. Boxplot summarizing the EC50 value distribution for genes sets regulated by MYC in different conditions. In black, the gene set obtained by MYC knockout in 3T9 fibroblasts is shown (Perna et al., 2012), in dark grey depicts gene sets from GSEA in Fig. 4.11 B and light grey indicates genes from DAVID analysis in Fig. 4.7 C. p-values were calculated with the Mann-Whitney-Wilcoxon test. Comparisons were performed using the gene set with the broader distribution (i.e. rib nuc comp).

To further demonstrate that MYC binding affinity determines regulation by different MYC levels, RNA-seq was performed in U2OS cells where MYC levels were titrated. U2OS cells were treated with different DOX amounts to induce increasing MYC levels (Fig. 4.13 A). Total and subsequently mRNA was isolated and RNA-seq was performed in biological duplicates. Heat maps with the same contrast were used to check the change in expression of genes belonging to the gene sets previously identified to have different EC_{50} and thus different MYC binding affinity in the Preranked GSE analysis (Fig. 4.10 A). As an example the structural constituent of ribosome, the RNA binding and the substrate specific

transporter activity gene sets were chosen. The genes contained in the latter group were strongly regulated also by MYC overexpression (Fig. 4.13 B), whereas the other two gene sets were not (Fig. 4.13 B). This further validates the hypothesis that MYC regulates different cellular processes according to its cellular levels and that promoter affinity determines this regulation.

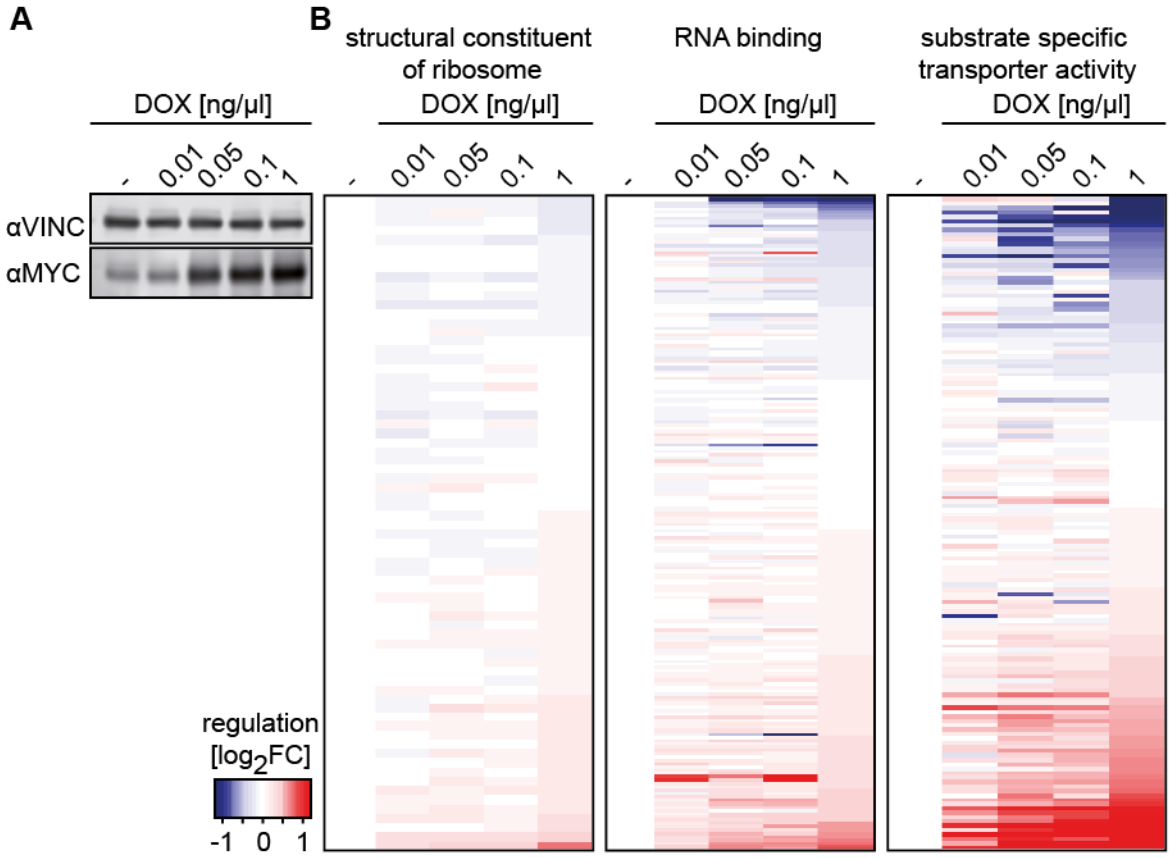


Figure 4. 13: Increasing MYC levels in U2OS cells regulate only genes with low binding affinity[2]. A. Immunoblot for MYC in U2OS cells treated with increasing DOX concentration. VINCULIN was used as loading control. **B.** Heat maps for changes in genes expression induced by titration of MYC levels.

4.2 Regulation of repressed genes by MYC partially depends on MIZ1

4.2.1 Repressed genes require high MYC levels to be regulated

To date, a clear mechanism to describe how MYC-dependent repression of gene expression is achieved is still missing.

The MYC titration RNA-seq experiments performed previously (4.1) suggested that high levels of MYC were needed to mediate repression of target genes.

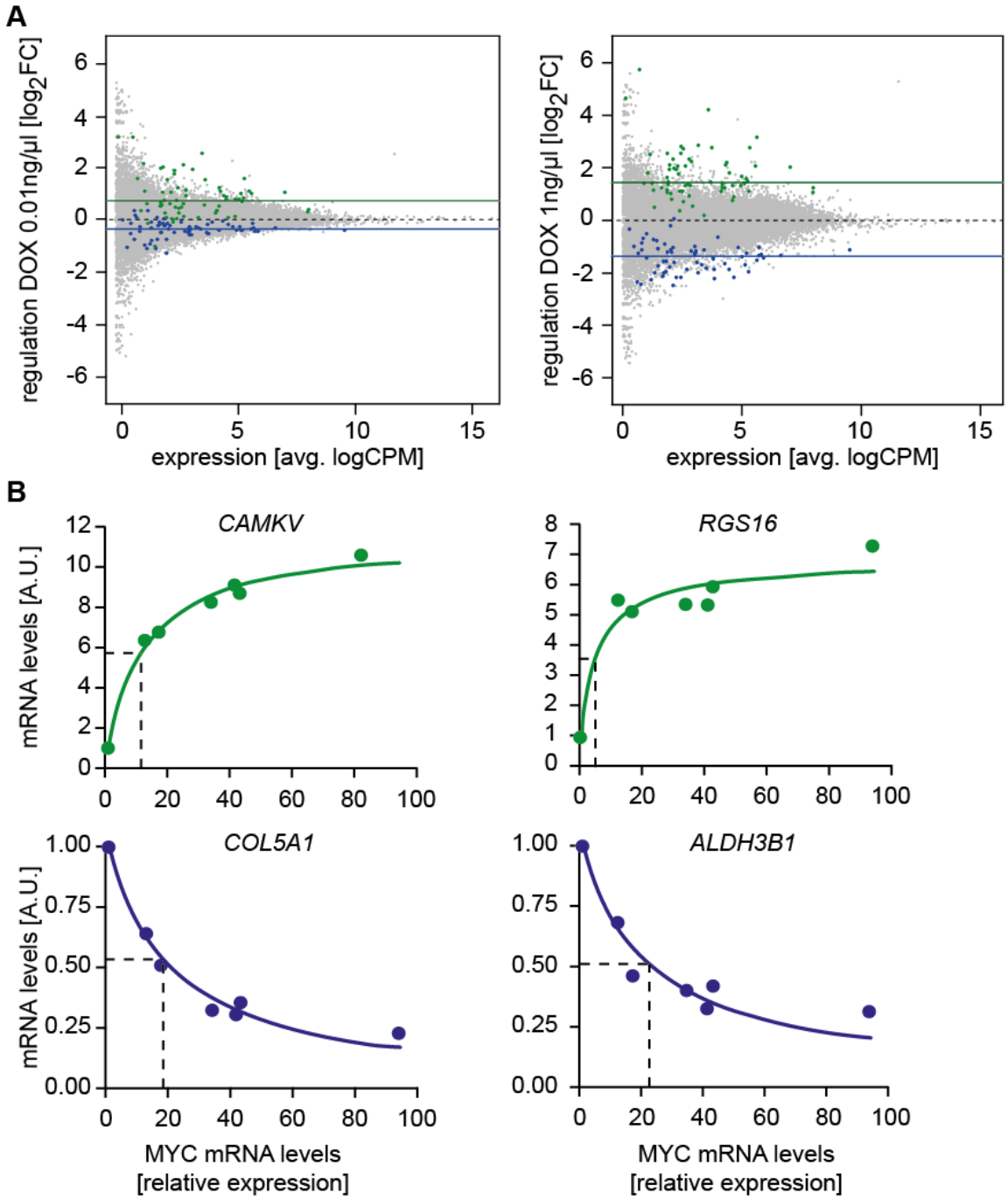


Figure 4. 14: Higher levels of MYC are needed for repression[2]. A. Plots for MYC induced changes in gene expression (log_2FC) versus absolute expression of the respective mRNAs (avg. logCPM). Green indicates activated genes (bound by MYC, $log_2FC>1.5$ and p-adj<0.01 in DOX 1ng/µl from RNA-seq of Fig. 4.2 A) and blue repressed genes (bound by MYC, $log_2FC<-1.5$ and p-adj<0.01 in DOX 1ng/µl from RNA-seq of Fig. 4.2 A). The colored lines represent median values of activated and repressed genes. **B.** RT-qPCR of selected activated or repressed genes. Dashed line represents putative EC_{50} value.

Genes activated and repressed to the same extent (note that the dashed lines in the plots represent the median of the regulation of the group of genes, Fig. 4.14 A) by the highest DOX concentration, thus the highest MYC levels, were instead regulated to a different

degree when a lower DOX concentration/MYC levels were present in the cell. Indeed, the median value for the activated genes was 0.72 whereas the one of the repressed genes was -0.36 indicating that activated genes were more strongly regulated than repressed ones at low overexpressed MYC levels. RT-qPCR of few selected activated or repressed genes (Fig. 4.14 B) reinforced the hypothesis that repression of gene expression by MYC is achieved at higher levels compared to activation. For *CAMKV* and *RGS16* almost full activation is reached already at the first DOX concentration (=0.01ng/µl), whereas for *COL5A1* and *ALDH3B1* full repression is achieved only with higher DOX concentrations (>0.05ng/µl).

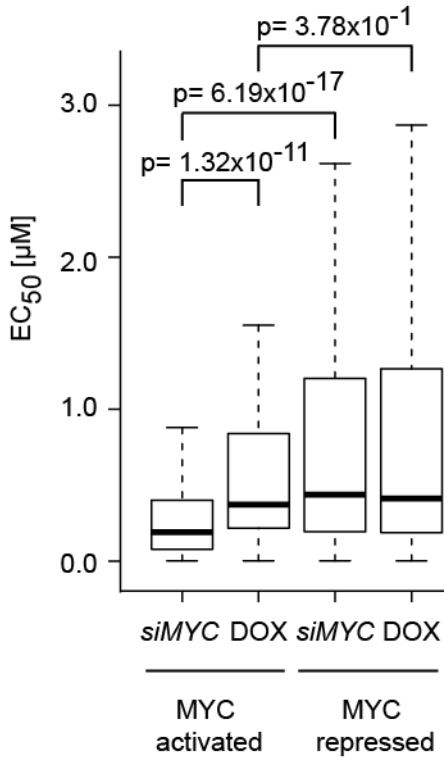


Figure 4. 15: EC_{50} of genes activated by MYC is lower than that of repressed genes[2]. Box plots of EC_{50} values of genes regulated by MYC overexpression ($log_2FC>1$ or <-1, p-value<0.05 and logCPM>0) or depletion ($log_2FC>1$ or <-1, p-value<0.05 and logCPM>0). p-values were calculated using the Mann-Whitney-Wilcoxon test.

Investigation of the EC_{50} of genes repressed or activated by MYC in U2OS cells were MYC was depleted or overexpressed showed that genes repressed by MYC depletion (*siMYC*, MYC activated) have higher affinity for MYC binding (lower EC_{50}), than genes activated by MYC overexpression (DOX, MYC activated) and repressed by MYC (MYC repressed) (Fig. 4.15). This argues that MYC-dependent repression happens at low affinity

binding sites and that again genes responding to MYC depletion have higher affinity to MYC.

MIZ1 is one of the proteins shown to be important for MYC-mediated repression. The formation of a complex with MIZ1 was shown to be necessary for the repression of the cell cycle inhibitors *CDKN2B* and *CDKN1A* by MYC (Seoane et al., 2001; Staller et al., 2001). To examine whether MIZ1 is involved in MYC-mediated repression in a more general manner, depletion of MIZ1 was induced in U2OS cells together with MYC overexpresssion and gene expression was analyzed via RNA-seq. MIZ1 depletion was achieved by infecting the cells with lentivirus containing pLKO plasmids encoding shRNAs against the MIZ1 mRNA. As control the empty vector was used. MYC was induced by adding 0.05 ng/µl DOX to the culture medium, concentration lower than that used before but high enough to achieve full repression (Fig. 1.14 B). This was done to be closer to a stoichiometric concentration of MYC and MIZ1 even in the overexpression condition.

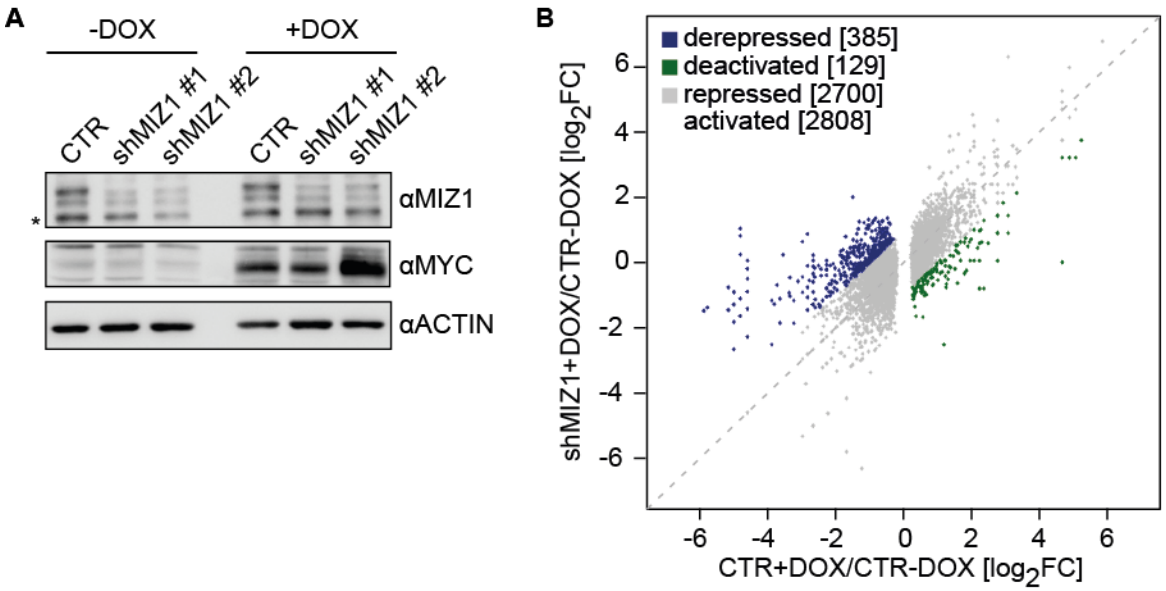


Figure 4. 16: MIZ1 depletion affects a group of MYC-repressed genes. A. Immunoblot of MIZ1 and MYC in U2OS cells treated with EtOH or DOX (0.05ng/µl) infected with lentivirus containing empty vector (CTR) or 2 different shRNA against MIZ1. ACTIN was used as loading control. * indicates an unspecific band. **B.** Plot of changes in gene expression induced by MYC overexpression (x-axis) and by MYC overexpression in absence of MIZ1 (y-axis). In square brackets are the number of genes belonging to each category.

MIZ1 depletion and MYC overexpression were efficient as shown by immunoblotting (Fig. 4.16 A).

To find out which genes are regulated by MYC and dependent on MIZ1, the changes in gene expression of cells overexpressing MYC (CTR+DOX) versus cells with endogenous MYC levels (CTR-DOX) were compared with those identified in cells with MYC overexpression and MIZ1 depletion (shMIZ1+DOX) relative to cells with endogenous levels of MYC and MIZ1 (CTR-DOX) (Fig. 4.16 B). In a plot, genes regulated to the same degree by MYC overexpression with or without MIZ1 depletion would appeared around the diagonal representing correlation of 1 between the two conditions (dashed grey line in the plot of Fig. 4.16 B). Both, genes activated with a $log_2FC>0$ and p-value<0.05 and repressed with $log_2FC<0$ and p-value<0.05 were examined and the majority of them behaves in this manner (grey dots Fig. 4.16 B), indicating that their regulation is not MIZ1-dependent. However, a group of repressed genes, containing about 3 times more genes than the one of activated genes, was less strongly regulated in the absence of MIZ1 (Fig. 4.16 B). Precisely, 385/2700 (blue, Fig. 4.16 B) genes were more than 2 fold less strongly repressed in absence of MIZ1 and only 129/2808 (green, Fig 4.16 B) genes were more than 2 fold less activated. This argues that MIZ1 is necessary for the repression of a certain group of MYC target genes but not for MYC-dependent activation.

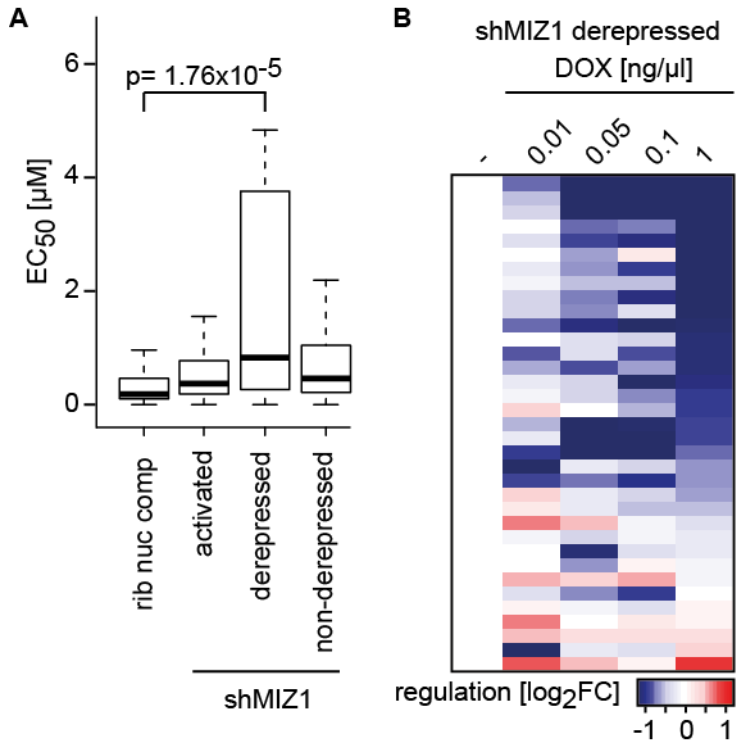


Figure 4. 17: MYC/MIZ1-dependent genes have low affinity for MYC[2]. A. Box plots of EC_{50} of genes regulated by MYC overexpression and MIZ1 depletion. The ribonucleoprotein gene set was used as reference for a gene set with affinity for MYC binding. **B.** Heat map of genes repressed by MYC in a MIZ1-dependent manner in the MYC titration experiment. The regulation scale is the same as Fig. 4.13 B.

MIZ1 was shown to be mainly an activator of genes involved in autophagy in neuronal progenitor cells where very few or no MYC is present (Wolf et al., 2013). It is possible that the interaction between MYC and MIZ1 and thus the formation of the complex that leads to transcriptional repression occurs only when MYC levels are high. An indirect evidence for this was that genes which were less repressed by MYC in the absence of MIZ1 (derepressed) had an EC_{50} higher than activated genes and genes found in the ribonucleoprotein complex set used as control for genes with a high affinity for MYC binding (Fig. 4.17 A). Moreover these genes reacted to MYC overexpression (Fig. 4.17 B), suggesting that they are not saturated by endogenous MYC levels in U2OS cells and thus, are targets of MYC at high levels.

Taken together these results indicate that MIZ1 is more important for MYC-dependent repression than for MYC-dependent activation and that the MYC/MIZ1 target genes have low affinity binding sites.

4.3 MYC influences RNA polymerase II phosphorylation and distribution as well as histone modifications

4.3.1 Recruitment and pause-release of RNA polymerase II are controlled by MYC

Transcription of a gene follows several discrete steps. First of all, the promoter is bound by general transcription factors that in order recruit the RNA Pol II. This enzyme is bound to the preinitiation complex (PIC) in a hypophosphorylated form. TFIIH, via its kinase subunit CKD7, phosphorylates the CTD of the RNA Pol II at serine 5 and favors the start of the transcription with the release of abortive small mRNAs. At this stage RNA Pol II is also bound by two complexes that facilitate the pausing of the enzyme: DSIF and NELF. In order to start the productive elongation of the mRNAs, pTEF-b is recruited. pTEFb has a kinase subunit (CDK9) that phosphorylates the two pausing complexes, leading to the release of NELF and to the conversion of DSIF into a positive elongation factor, and the CTD of RNA Pol II at serine 2 residues. In this hyperphosphorylated state RNA Pol II starts the efficient elongation of the mRNAs till the 3' of the gene where termination takes place.

As just mentioned, the steps of transcription are characterized by the different phosphorylation states of the RNA Pol II and specific antibodies can recognize the enzyme

phosphorylated at different residues. Thus, ChIP-seq experiments were performed to check which step of transcription is affected by MYC. RNA Pol II distribution and phosphorylation was monitored via ChIP-seq in U2OS cells treated with EtOH (-DOX), as control, and with 1µg/ml DOX to induce MYC overexpression. For this analysis an antibody recognizing the N-terminal part of the RBP1 subunit of RNA Pol II was used to estimate RNA Pol II distribution and recruitment. An antibody binding to serine 5 phosphorylated RNA Pol II was employed to analyze transcription initiation. An antibody recognizing the serine 2-phosphorylated RNA Pol II was used to monitor the transcription elongation efficiency. As for MYC ChIP-seq, an IgG serum and 1% input of the precipitated chromatin were also sequenced and used as controls.

After sequencing, Seqminer was used to measure the changes in RNA Pol II induced by MYC. The ratio between the occupancy of total RNA Pol II in the promoter region of ±DOX-treated cells (recruitment) was calculated for all MYC-bound genes. The same was done for the occupancy of serine 5 phosphorylated RNA Pol II in the two conditions to check if MYC influences TFIIH in the phosphorylation of RNA Pol II and thus transcription initiation. To examine whether MYC controls transcription elongation, the ratio of the occupancy of serine 2 phopshorylated RNA Pol II in ±DOX-treated cells for each gene was calculated around the termination site (±1000bp from the UCSC annotated termination site). The calculated ratios were correlated with changes in gene expression induced by MYC at target genes and therefore, as shown before (Fig. 4.7 A), indirectly with MYC recruitment. Specifically, genes were sorted according to gene expression regulation (measured previously via RNA-seq) and divided in 20 equally-sized bins. For each bin, median change in gene expression induced by MYC overexpression and median change in RNA Pol II occupancy or phosphorylation was calculated and plotted (Fig. 4.18 B).

A positive correlation with gene expression regulation by MYC (Fig. 4.18 B) was observed for all the modifications analyzed. At activated genes, MYC led to an increase in RNA Pol II recruitment at the promoter and to higher phosphorylation of RNA Pol II at serine 5 (Fig. 4.18 A and B). In contrast, at repressed genes MYC overexpression decreased the amount of total and serine 5 phosphorylaed RNA Pol II present at the promoter (Fig. 4.18 A and B) indicating that MYC suppressed RNA Pol II recruitment and transcription initiation at downregulated genes.

Since the slope of the line fitting the data was comparable between the two modifications (Fig. 4.18 B, $m_{RNA\ Pol\ II}$=0.2089 and $m_{pSer5RNA\ Pol\ II}$=0.2260), MYC regulated mainly RNA Pol II recruitment and consequently this influenced transcription initiation.

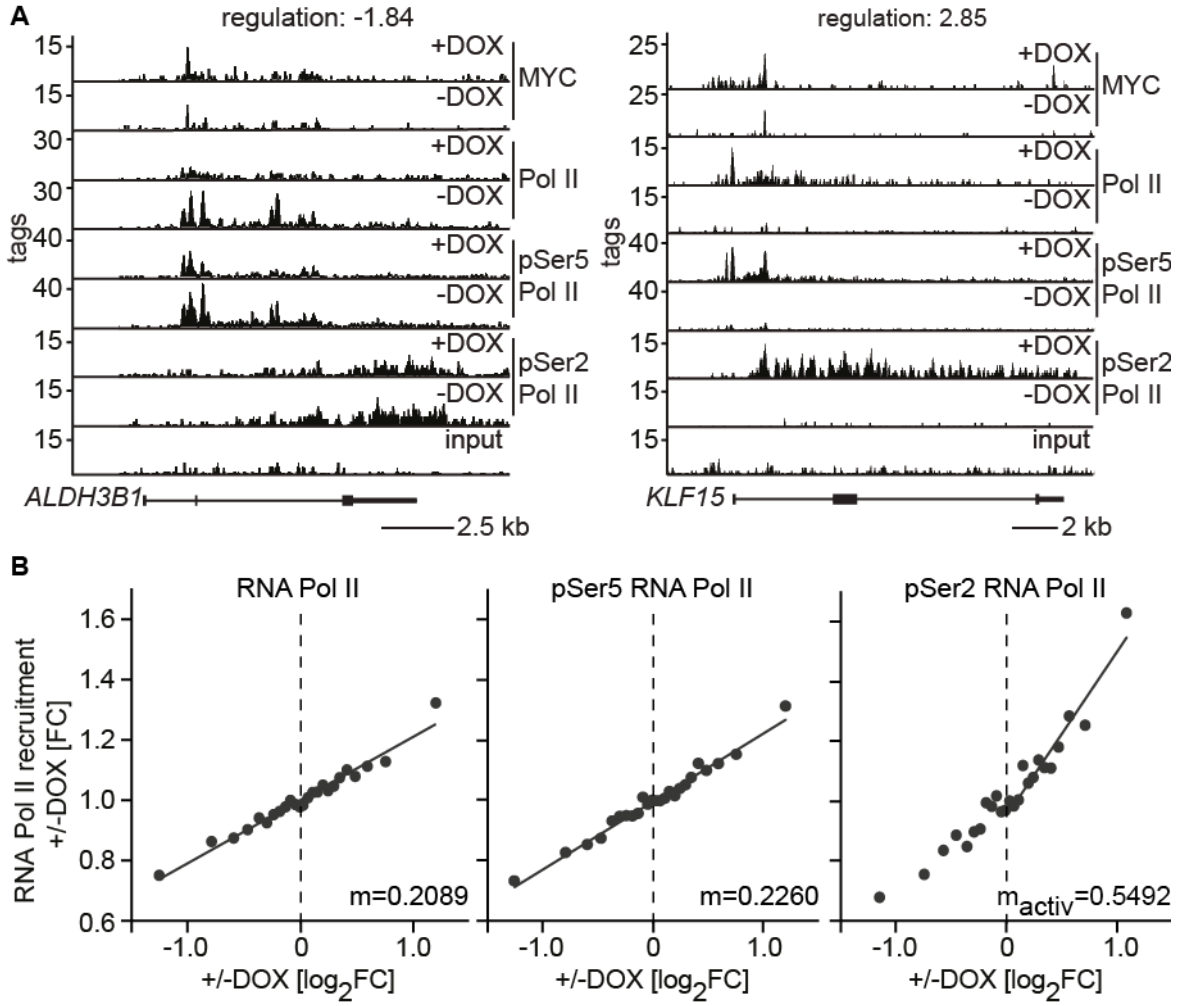


Figure 4. 18: MYC controls RNA Pol II recruitment and phosphorylation[1]. A. Gene tracks of MYC binding, total RNA Pol II (Pol II), phosphorylated serine 5 RNA Pol II (pSer5 Pol II) and phosphorylated serine 2 RNA Pol II (pSer2 Pol II) distribution in U2OS cells treated with EtOH (-DOX) as control, or 1ng/µl DOX to induce MYC overexpression. Two examples of MYC target genes are shown. Input was used as control. **B.** Binned plots of MYC-induced changes in gene expression (x-axis) versus changes in total RNA Pol II and pSer5 RNA Pol II at the promoter and pSer2 RNA Pol II around the termination site at each MYC target. m indicates the slope of the lines fitting the data. Each dot represents the median of 365 genes.

Looking at the serine 2 phosphorylated RNA Pol II around the termination site and having an estimation of the transcription elongation efficiency (Fig. 4.18 A and B), it can be seen that on repressed genes the extent of the regulation was highly comparable with the one seen for total RNA Pol II (Fig. 4.18 B). This indicates that at repressed genes MYC only influences RNA Pol II recruitment. On the other hand, at activated genes the degree of

change of serine 2 phosphorylated RNA Pol II occupancy induced by MYC was higher than that induced on total RNA Pol II (Fig. 4.18 B), suggesting that at these genes, beside inducing recruitment, MYC also stimulates transcription elongation. The effect that MYC exerted on the transcription elongation of the activated genes was additive to the one it had on RNA Pol II recruitment. A possible interpretation of these results suggests that MYC might control mainly RNA Pol II recruitment at all target genes and the changes in phosphorylation at serine 5 (and serine 2 for repressed genes) might be a consequence of this regulation. Only at activated genes MYC additionally controls RNA Pol II phosphorylation at serine 2 and thus transcription elongation.

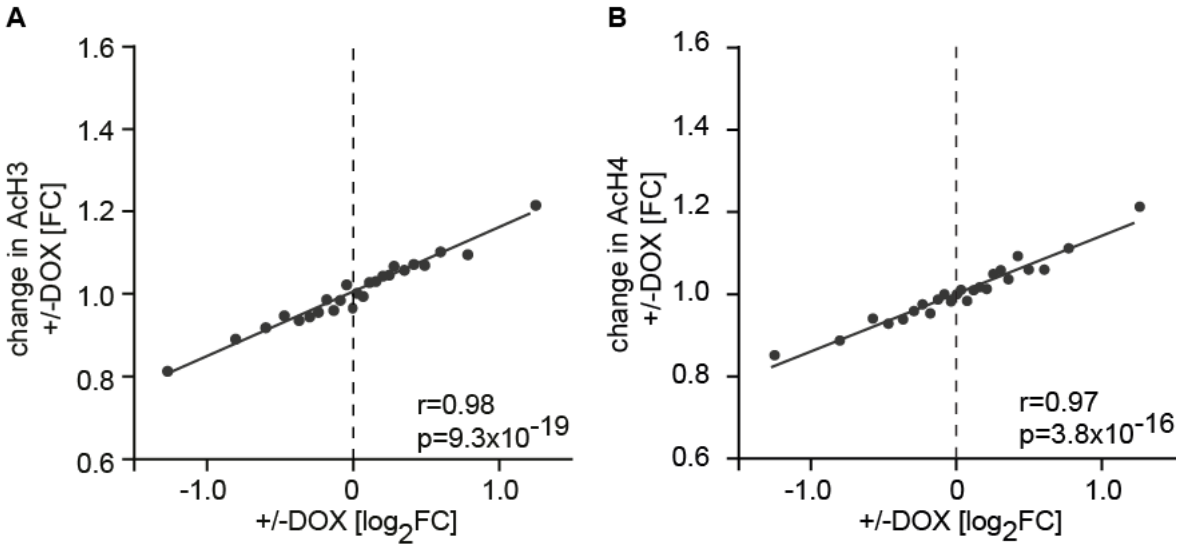
4.3.2 MYC controls histone modification deposition at target genes

A possible mechanism by which MYC controls RNA Pol II recruitment, and additionally transcription elongation, could be the recruitment/interaction of MYC with histone modifying enzymes that specifically modify histones and alter the chromatin state of MYC target genes. In general high acetylation of histones 3 and 4 is associated with transcription activation whereas low acetylation of the same comes along with inactive transcription (Eberharter and Becker, 2002). Furthermore, previous studies have shown that MYC influences the chromatin state of target genes by indirectly altering the distribution of histone modifications (Guccione et al., 2006; Knoepfler et al., 2006; Martinato et al., 2008).

To investigate the changes in histone modifications induced by MYC in U2OS cells and whether they correlate with the MYC-dependent regulation of gene expression, pan-acetylation of histones 3 (pan-AcH3) and 4 (pan-AcH4) was checked. ChIP-seq was performed in U2OS cells treated with EtOH (-DOX) or 1ng/µl DOX using antibodies that recognize pan-AcH3 and pan-AcH4. IgG ChIP-seq and 1% input samples were included as controls. As for RNA Pol II analysis, Seqminer was used to calculate the amount of acetylation around (±1000bp) the summit of the MYC peaks in the two conditions. Subsequently, the ratio between the two occupancies was calculated for each MYC-bound gene, genes were sorted according to regulation and divided in equally sized-groups. For each group the median of gene expression regulation and the median change in acetylation was calculated and plotted.

Both for pan-AcH3 and pan-AcH4, there was a positive correlation between gene expression regulation and changes in histone acetylation. At MYC-activated genes, the increase in gene expression was accompanied by an increase in acetylation of histone 3 and

4 in the promoter region. Vice versa, at repressed genes the extent of the acetylation decreased with the intensity of the regulation. In agreement with previous studies, this suggests that regulation of gene expression by MYC is associated with changes in histone acetylation. Furthermore, since gene expression regulation correlated with MYC



recruitment, changes in histone modifications indirectly correlated with MYC recruitment, arguing that recruitment of MYC and not occupancy influences acetylation of H3 and H4.

Figure 4. 19: MYC-dependent gene regulation is associated with changes in histone acetylation[1]. A. Binned plot for MYC-induced changes in gene expression (x-axis) versus changes induced by MYC in acetylation of histone 3 (y-axis). 9614 genes were divided in 20 equally-sized-bins. Each dot represents median of the bin. r is the Pearson correlation coefficient of the line fitting the data and the p-value for r was calculated using a Student's *t*-test. **B.** Binned plot as in **A** but acetylation of histone 4 is shown.

Histone modifications can also be used to identify enhancers. Particularly, enhancers are defined as regions with high monomethylation of the lysine 4 of histone 3 (H3K4me1), high acetylation of lysine 27 of histone 3 (H3K27Ac) and low tri-methylation of lysine 4 of histone 3 (H3K4me3) (Calo and Wysocka, 2013; Creyghton et al., 2010; Lin et al., 2012). Moreover, enhancer activation and usage are very cell type specific and data sets generated from one cell type or tissue cannot usually be used for another one (Heinz et al., 2015).

Previous papers have shown that MYC binds to enhancers in primary B and T cells and in a human lymphoma model, reinforcing its role as a general amplifier of gene expression (Lin et al., 2012; Nie et al., 2012).

To investigate whether this is the case also in U2OS cells, ChIP-seqs for histone modifications to identify enhancers were performed in parallel to MYC binding analysis.
As for promoter regions, MYC binding was present at enhancers and increased with MYC overexpression (Fig 4.20). Further analysis on this data set was performed by Susanne Walz and showed no correlation between MYC recruitment at enhancers and the regulation of gene expression induced by MYC (Walz* and Lorenzin* et al., 2014). This indicates that, although MYC binds to enhancers, the changes in MYC binding at enhancers did not influence expression of the neighboring genes in U2OS cells. New more accurate methods to define enhancers, which have been developing in the last years (Dekker et al., 2002; Wit and Laat, 2012), might lead to different results.

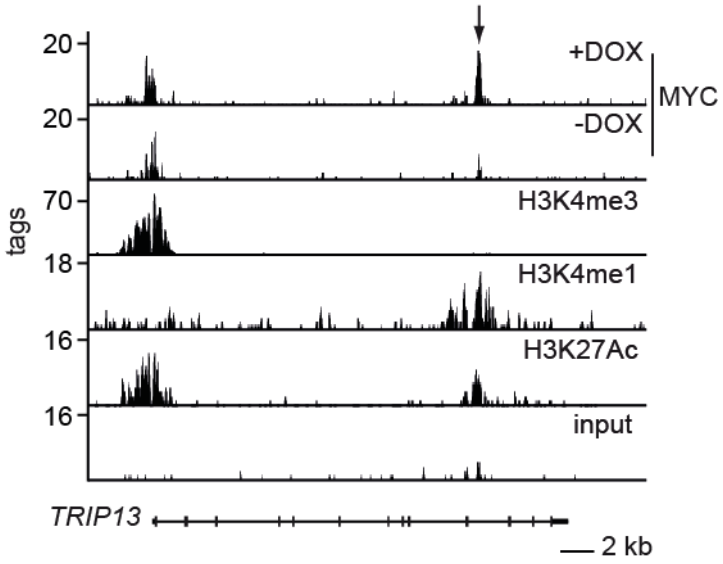


Figure 4. 20: ChIP-seq traces for MYC binding and for histone modifications that were used to define enhancers (high H3K4me1 and H3K27Ac, low H3K4me3)[1]. The arrow indicates the putative enhancer of the *TRIP13* gene.

Chapter 5:

Discussion

MYC expression is elevated or deregulated in up to 70% of all human cancers (up to 70%) and is often associated with aggressive and poorly differentiated tumors. Several studies in transgenic mouse models of human tumors showed that deregulated expression of MYC proteins promotes tumorigenesis and that established tumors depends on elevated MYC levels (Dang, 2012; Gabay et al., 2014; Soucek et al., 2013).

Genetically targeting MYC has been revealed to be a successful way of treating tumors in mice but therapies targeting MYC proteins are still far away from entering the clinics (Soucek et al., 2008, 2013). It has been difficult to target MYC activity directly with small molecules given the large surface the protein has to interact with the DNA and with its partner protein MAX (Nair and Burley, 2003; Yin et al., 2003). However, proteins acting upstream of MYC, that regulate its stability or important for regulating expression of its target genes, have been already targeted by several developed drugs leading to tumor regression in mouse models (Brockmann et al., 2013; Delmore et al., 2011; Mertz et al., 2011; Peter et al., 2014).

Since MYC exerts its functions via the regulation of transcription (Eilers and Eisenman, 2008), new ways of targeting tumors could come from the identification of genes that are cancer specific and regulated by MYC only at oncogenic levels. This could lead to the design of drugs that target proteins expressed or activated only in tumors and thus affect cancer but not normal cells, opening a therapeutic window to treat tumors with deregulated MYC expression.

With the development of the microarray and next generation sequencing technologies many studies have identified MYC target genes (Coller et al., 2000; Mao et al., 2003; Menssen and Hermeking, 2002; Perna et al., 2012; Schuhmacher et al., 2001; Zeller et al., 2003, 2006). These studies identified a core set of processes that are regulated by MYC including ribosome biogenesis, protein translation, several biosynthetic metabolic pathways, cell adhesion and cytoskeleton. Although thousands of MYC responsive genes have been identified, only a minority of target genes are regulated or bound by MYC in all studies with the differences coming from the cellular system (cell type or species) and the

kind of technology used. Furthermore, the set of target genes that distinguish MYC physiological function from the pathological/oncogenic one, whether it exists or not, has not been fully understood yet. Indeed, deregulation of MYC could, not only activate and repress MYC physiological targets, but also regulate expression of genes that are not its targets under physiological conditions (Dang et al., 2006).

During the recent years, discrepancies have emerged about how MYC recognizes its target genes (Blackwell et al., 1993; Guo et al., 2014) and whether it regulates a pre-existing cellular program(s) amplifying expression of all genes with an open promoters (Lin et al., 2012; Nie et al., 2012) or it activates and represses expression of specific sets of genes (Brockmann et al., 2013; Horiuchi et al., 2012; Kawauchi et al., 2012; Northcott et al., 2011) ("amplifier" vs. "specifier" model).

In order to shed light onto these topics genome-wide binding analysis and gene expression analysis in U2OS cells having endogenous/physiological or supraphysiological MYC levels were performed.

The osteosarcoma U2OS cell line was chosen as cellular system for this work since it is a tumorigenic cell line that has relatively low levels of endogenous MYC, as compared with other tumor cell lines and with un-transformed epithelial cell lines (Fig. 4.1 A). Analysis of MYC targets in this setting could reveal the role of MYC at physiological levels, although in an oncogenic environment. Supporting this view, depletion of MYC by siRNA affected the same processes as knockout of MYC in serum stimulated fibroblasts (Fig. 4.11 B and Perna et al., 2012). Expression of genes encoding for proteins involved in ribosome biogenesis, RNA processing, nucleotide metabolism and mitochondrial matrix was dowregulated both in U2OS and fibroblast upon MYC depletion/knockout.

In order to investigate oncogenic MYC function, U2OS were engineered to overexpress MYC in an inducible manner. Prolonged overexpression of MYC in this system induced apoptosis, indicating the expression of a functional protein (data not shown) (Walz* and Lorenzin* et al., 2014). Moreover, GSE analysis showed that among the gene sets regulated by MYC overexpression, the "Kim MYC amplification targets up" gene set was enriched (Fig. 4.2 C). This indicated that by overexpressing MYC in the U2OS tumor cell line, that have physiological levels of MYC, there is the activation of a oncogenic MYC signature and validated further the cellular system and set up used for the experiments.

Estimation of the number of MYC molecules expressed by U2OS cells showed that approximately 85,500 molecules of endogenous MYC are present in these cells and this number rises to about $1x10^6$ upon treatment with 1µg/ml doxycyline (Fig. 4.5). Previous

estimates found that three different human tumor cell lines express between 110,000 and 880,000 molecules of MYC per cell (Lin et al., 2012), confirming that U20S cells express relatively low levels of endogenous MYC and that MYC levels reached upon doxycycline induction can be found in human tumor cells.

5.1 Several factors shape transcriptional amplification by MYC

B-cells proliferate slowly in absence of MYC and the activation of its expression (via LPS stimulation of naïve cells or overexpression) induces massive cell growth and finally cell division (Lin et al., 2012; Nie et al., 2012; Sabò et al., 2014). Moreover, activation of B-cells from the naïve state increases synthesis of all macromolecules that are required to satisfy such rapid growing cells, in particular the transcription machinery and other components that are needed to activate gene expression. In this context, MYC invades promoters and enhancers, even with low binding affinity, and directly amplifies transcription of all open promoters but not to the same extent. Enhancers and promoters differ in their affinity for MYC and this dictates differences in the response (Wolf et al., 2014). Moreover, in B cells MYC activation is accompanied by an increase of 2-3 folds in total and mRNA levels. By using experimental and data normalization methods for gene expression analysis that do not account for the increase in RNA content of one sample compare to the others, genes that are strongly amplified would appear as activated, whereas genes that are left behind (weakly regulated) would appear as repressed by MYC. In order to avoid this problem, normalization based on the number of cell used and/or on spike-in standards should be used (Lin et al., 2012; Lovén et al., 2012; Nie et al., 2012; Sabò et al., 2014).

Beside direct transcription amplification induced by MYC binding to promoters and enhancers, MYC can also indirectly amplify transcription by inducing GCN5, that acetylates histones causing a global opening of the chromatin, and PRPS2 (phosphoribosyl-pyrophosphate synthetase 2), which promotes the already enhanced nucleotide biosynthesis (Cunningham et al., 2014; Knoepfler et al., 2006; McMahon et al., 2000). Therefore, in B cells MYC can clearly induce amplification of gene expression as result of direct MYC binding to promoters and enhancers and indirect activation of chromatin remodeler genes.

U2OS cells differ from B-cells. They did not increase cell size in response to MYC overexpression, rather it decreased (Fig. 4.1 E), and their total and mRNA amount did not change upon MYC overexpression (Fig. 4.1 F and G). Therefore, the same amount of RNA from the different samples could be used to perform gene expression analysis via RNA-seq. This led to the identification of a set of genes that is activated by MYC and a set that is repressed (Fig. 4.2 A), confirming that in this cellular system MYC does not amplifies all open promoters. Parallel binding analysis via ChIP-seq showed that about half of the genes activated and ¼ of the genes repressed were also directly bound by MYC. Moreover, there was a correlation between MYC recruitment and gene expression changes for both activated and repressed genes, further supporting a direct role of MYC in mediating activation but also repression of transcription.

As shown by analyzing the U2OS data, saturation for MYC binding occurred at several genes (Fig. 4.4 and 4.8) and promoter affinity stratified cellular processes that are regulated by different MYC levels (Fig. 4.7 C, 4.10 A, 4.12 and 4.13). Moreover, repressed genes seemed to require higher levels to be regulated and accordingly, had also a higher EC_{50} and thus lower affinity for MYC binding (Fig. 4.14 and 4.15).

This might suggest that in a cellular system that goes from resting/quiescent state, with very low levels of MYC, to proliferation, the function of MYC is to activate transcription and amplify expression of genes necessary for cellular growth, such as those involved in ribosome biogenesis, translation, RNA processing etc. (Schuhmacher et al., 2001). These are the genes with the highest affinity for MYC, highly enriched for canonical E-box in their promoters and thus are the first to be occupied. This would be consistent with the role of MYC in *Drosophila melanogaster*. In flies, MYC binds to E-boxes (CACGTG) and activate transcription of target genes encoding proteins involved in RNA and protein biosynthesis (Gallant et al., 1996; Johnston et al., 1999; Orian et al., 2003). Consistently, dMYC stimulates cell growth, but not proliferation. Moreover, MYC binding and gene expression analysis after weak MYC overexpression in breast epithelial cells, that have low levels of endogenous MYC (Fig. 4.1 A), showed that MYC only activates target genes involved in cellular growth and no direct MYC-mediated repression was detected (von Eyss et al., 2015; Jaenicke et al., 2015).

In other biological conditions, such as embryonic development, regeneration of compromised tissues, wound healing, or at specific stages of tumor development, MYC levels could increase above those needed for cellular growth. If this happens, the high affinity binding sites would already be saturated and MYC would spill over to other free

sites that have low affinity. These are present in the promoter of genes involved in the regulation of angiogenesis, cell migration, extracellular matrix and in the response to hypoxia (Fig. 4.7), processes that may be needed only in these particular conditions. Consistently, MYC repressed genes have low affinity sites in their promoters (Fig. 4.7 and4.14) and are enriched for genes involved in cell adhesion and extracellular matrix (data not shown and Gebhardt et al., 2006; Inghirami et al., 1990).
MYC could have evolved from a transcription factor that only activates and amplifies expression of genes involved in cellular growth, as in flies and in cells that have low levels, to one that via interaction with low affinity sites, and probably via low affinity protein-protein interactions, both activates and represses transcription.

Different factors could explain how MYC can amplify gene expression in one system and specifically activate and repress transcription in another one.
The amount of MYC protein present in the cells, beside a pre-established transcriptional program, is important to determine the MYC induced-transcriptional output. Indeed, breast epithelial cells and naïve B-cells have low levels of MYC compared to U2OS (for breast epithelial cells see Fig. 4.1 A, naïve B-cells have 13,000 MYC molecules per cells (Lin et al., 2012) compared to 85,526 of U2OS).

Protein-protein interactions are crucial for MYC function. It was shown that MYC, mainly via MYC box II but also via its C-terminal part and other MYC boxes, recruits many co-activators. Moreover, MYC was shown to have thousand of binding sites in the genome and to be recruited at genes where histone marks for open chromatin are present. This would be consistent with a general role of MYC in activating transcription and amplifying expression of a preexisting transcriptional program present in the cell, which is marked and recognized by MYC via specific histone modifications. This is probably the case for B-cells, where stimulation via LPS or MYC overexpression leads to a global opening of the chromatin with amplification of expression of all genes that have an open promoter (Lin et al., 2012; Nie et al., 2012; Sabò et al., 2014). In U2OS cells instead, although thousands of MYC binding sites are detected (Fig. 4.3 C) and a correlation between MYC binding sites and histone modifications, such as H3K4me3, is found (data not shown and Walz et al., 2014), no amplification of transcription but specific up- and downregulation of selected target genes was detected. This could be explained by saturation of binding sites (see above) and by the interaction of MYC with MIZ1.

MIZ1 alone is a transcriptional activator that in neuronal progenitor cells binds few hundreds of promoters found in genes involved in autophagy and in the control of the vesicular trafficking required for autophagy (Wolf et al., 2013). Moreover, MIZ1 is a negative regulator of cell cycle progression and participates in the TGF-β signaling by binding and activating expression of CDK inhibitors (CDKN2B, CDKN1A and CDKN1C) (Adhikary et al., 2003; Seoane et al., 2001, 2002; Staller et al., 2001). At these genes, the formation of the ternary complex MAX/MYC/MIZ1 represents one of the best studied examples of MYC mediated-repression. MYC can be recruited with MAX, to the promoter of target genes by interaction with MIZ1 (Herold et al., 2002; Mao et al., 2003). This hinders the recruitment by MIZ1 of other transcriptional co-activators, such nucleophosmin and p300 acetyltransferase (Staller et al., 2001; Wanzel et al., 2008) and thus inhibits gene expression. Furthermore, MYC directs DNMT3a methyltransferase activity to MIZ1 target genes (Brenner et al., 2005).

Neuronal progenitor cells do not express MYC and in these cells MIZ1 only occupies few binding sites, which contain the direct MIZ1-binding sequence (Wolf et al., 2013). ChIP-seq analysis for MYC and MIZ1 in tumor cells, such as HeLa cells, T-cell lymphoma and pancreatic tumor, showed that MYC heavily influences MIZ1 binding to DNA, and increasing levels of MYC proteins correspond to increase amount of MIZ1 bound to the DNA (Walz et al., 2014). Furthermore, inspection of MYC and MIZ1 binding at wide genomic regions showed that the binding sites of the two proteins highly correlates, suggesting a broader role of MIZ1 in MYC-mediated regulation of transcription, than anticipated. At these sites, MYC and MIZ1 binds in a ternary complex with MAX (Walz et al., 2014).

In U2OS cells MIZ1 depletion by shRNA affects MYC-mediated repression but not activation. Indeed, a part of genes that were repressed after MYC overexpression, were de-repressed by two fold or more in absence of MIZ1 (Fig. 4.16 B) indicating that MIZ1 is required for the repression of a part of, but not all, MYC target genes. Overexpression in U2OS cells of a mutant form of MYC, MYC(V394D), led to the same results (Walz et al., 2014). MYC(V394D) is a single point mutant of MYC where the valine residue at position 394 in the helix-loop-helix domain is mutated to aspartic acid. It was previously shown that this region of MYC is important for interaction with MIZ1 (Peukert et al., 1997) and indeed substitution of one amino acid with a negative charge blocked interaction with MIZ1 but retained the ability to bind to MAX and to activate transcription (Herold et al., 2002). Accordingly, for all MYC-repressed genes detected after MYC overexpression in

U2OS cells, the extent of the de-repression by depletion of MIZ1 correlated with their enhanced expression in cells overexpressing MYC(V394D) instead of MYC (Walz et al., 2014). Transcriptional activation by MYC(V394D) was mostly unaffected. These supports further the idea that the formation of a MYC/MIZ1 complex can alter MYC function and shape/restrict the general amplifier activity of MYC. The *in vivo* relevance of the complex formation between MYC and MIZ1 was shown by van Riggelen and colleagues. They revealed that forming a MYC/MIZ1 repressive complex is important for MYC to induce development of and maintain T-cell lymphoma by antagonizing the ability of TGF-β pathway to suppress proliferation and induce senescence (Van Riggelen et al., 2010).

The formation of a MYC/MIZ1 repressive complex is thought to happen when supraphysiological MYC levels are present and, probably, the transcriptome amplification induced by MYC needs to be limited. Consistently, as previously described, MYC and MIZ1 broadly associate on chromatin in tumor cells with high levels of MYC (Walz et al., 2014). The genes repressed by this complex have higher EC_{50}, and thus lower affinity for MYC binding, than MYC-activated genes or genes involved in ribosome biogenesis (Fig. 4.17) indicating that higher levels of MYC needs to be present in order for the transcription factor to strongly bind these genes.

The direction of the transcriptional response to MYC and/or MIZ1 binding depends on the ratio of the two proteins bound at a given promoter (Walz et al., 2014). Analysis of MYC and MIZ1 binding in HeLa cells and in MYC-driven T-cell lymphoma model showed that the MYC/MIZ1 ratio present at each promoter determines activation or repression of the gene: genes with high MYC/MIZ1 ratio – highly bound by MYC – are MYC-activated target genes. Genes with low MYC/MIZ1 ratio – highly bound by MIZ1 – are MIZ1-activated genes. Genes with a ratio of about 1 are MYC/MIZ1 repressed genes (Walz et al., 2014).

Beside MYC levels, other proteins affect MYC/MIZ1 complex formation.

The ubiquitin ligase HUWE1 was shown to associates both with MYC and with MIZ1 and is required for growth of colorectal cancer cells in culture and in xenograft models. Inhibition of HUWE1 via small molecules inhibits MYC-dependent transactivation by stabilization of MIZ1. MIZ1 accumulates at MYC bound promoters, blunts activation and enhances repression (Inoue et al., 2013; Peter et al., 2014).

The expression of the ARF tumor suppressor protein is induced by supraphysiological levels of MYC and the encoded protein interacts with MYC and inhibits MYC-dependent

transactivation (Qi et al., 2004; Zindy et al., 1998). Furthermore, ARF also binds to and inhibits HUWE1 promoting association of MIZ1 with MYC (Chen et al., 2005; Herkert et al., 2010).

Determination of whether transcriptional amplification by MYC occurs in tumors still needs further investigation. Although one of the two studies that identified MYC as a general amplifier of gene expression focused the analysis on a human lymphoma model, the evidence to support this model came mostly from the transition of B-cells from a resting to a proliferating state in response to stimulation (Lin et al., 2012; Nie et al., 2012). B-cells reacted in an exaggerated way to their induction increasing size and RNA content. It is hard to envision that in a solid tumor, a cell could allow itself to increase heavily cell size and to direct all the available energetic resources to transcription amplification, considering the limited nutrients present in the tumor microenvironment and the tissue homeostasis. Moreover, to show that tumor cells with MYC amplification bear elevated RNA content compared to non MYC-amplified cells, small cell lung cancer cells were analyzed coming from two different patients, having therefore a different genetic background (Lin et al., 2012). On the other hand, the analysis performed in U2OS cells, where no MYC induced-amplification was observed, relies on an induction of MYC expression that might lead to secondary effects, among which transcriptional repression.

5.2 MYC binding to DNA

MYC as a helix-loop-helix transcription factor, heterodimerizes with MAX and binds to specific DNA sequences called E-boxes. CACGTG is the canonical E-box sequence bound by the MYC/MAX complex, but the two central nucleotide can also change (non canonical E-boxes) retaining the ability to bind the complex (Blackwell et al., 1990, 1993; Blackwood and Eisenman, 1991).

Although the dissociation constant (K_D) values for MYC binding to canonical E-boxes change a lot depending on the experimental setup used (from 0.1 to 2.12 to 90.5nM) (Fieber et al., 2001; Guo et al., 2014; Hu et al., 2005), several *in vitro* studies showed that MYC binding affinity to canonical E-boxes is higher than that to non canonical E-boxes or other DNA sequences. Indeed, Hu and colleagues showed that the MYC/MAX complex has 2.5fold more binding affinity for the canonical E-box sequence CACGTG than to the non canonical sequence CAGGTG (Hu et al., 2005). Instead, MYC/MAX binds with

200fold more affinity the canonical E-box than other DNA, represented by the sequence ATCTAG (Guo et al., 2014). These observations, together with a large number of studies focusing on the effect of MYC on specific genes, led to the prevailing model that MYC, in conjunction with MAX, binds to E-box sequences and regulates transcription of RNA Pol II and I (Dang, 2012; Eilers and Eisenman, 2008; Lüscher and Vervoorts, 2012). However, the discovery that MYC acts globally regulating all expressed genes hint that MYC recruitment might be less sequence-dependent.

Canonical E-boxes are also enriched in the MYC binding sites in the U2OS system (Fig. 4.6 A and 4.7 D). However, only one third of all genes bound by MYC bear in their promoter a canonical E-box. Furthermore, mathematical modeling of MYC/MAX binding suggested that the sole binding to the DNA backbone is not sufficient to account for the wide chromatin binding of MYC at sites where no E-boxes are present (Uwe Benary and Jana Wolf personal communication). This indicates that other factors could contribute to recruitment of MYC.

Non canonical E-box sequences, with low affinity binding for MYC, play a role in recruiting MYC at high levels, at which canonical and high affinity binding sites are already occupied and saturated (Fig. 4.7 D). These sequences are also enriched in the enhancers bound by MYC in P-493 cells (Lin et al., 2012).

Specific histone modification could also play a role in recruiting MYC. MYC binding indeed highly correlates with histone modifications that are present at open promoters and are associated with active transcription. Specifically, MYC is present at promoters of genes marked with H3K4me3, H3K27Ac, and it has been proposed that histone 3 lysine 4 and 79 methylation is required for MYC to engage target genes (Guccione et al., 2006; Martinato et al., 2008; Nie et al., 2012; Sabò et al., 2014; Zeller et al., 2006). Enhancers that have, beside high H3K4me1 and low H3K4me3, acetylation of lysine 27 of histone 3, a marker for active enhancers, are also bound by MYC. Moreover, MYC binding correlates with pan histone 3 and 4 acetylation (Fig. 4.19 and Martinato et al., 2008; Nie et al., 2012). Since MYC binding correlates also with RNA Pol II distribution and the expression level of genes (Fig. 4.6 C and Guo et al., 2014; Nie et al., 2012), is unknown whether the histone marks required for MYC or present at its bound sites are recognized by specific epigenetic readers or are just associated with the accessibility of the DNA in modified nucleosomes. Nevertheless, MYC does not have domains to interact directly with the histone modifications, but via contact with its numerous interactors it could be recruited to these

sites. For example a recent study showed that MYC interacts with WDR5 which promotes target gene recognition and tumorigenesis by MYC (Thomas et al., 2015). WDR5 is a WD40-repeat-containing protein that assembles into chromatin regulatory complexes including MLL/SET methyltransferases (methylates H3K4) and MOF/NLS histone acetyltransferases (H4Ac) (Thomas et al., 2015).

MYC can also interact directly with the basal transcription machinery and thus could be recruited to the DNA independently of E-boxes. As previously shown MYC interacts with TFIIH and pTEF-b (Bouchard et al., 2004; Cowling and Cole, 2007; Eberhardy and Farnham, 2002; Gargano et al., 2007; Kanazawa et al., 2003; Rahl et al., 2010), important proteins for promoting transcription initiation and elongation. Moreover, MYC interacts with TFIIIB and is recruited to RNA Pol III genes via this interaction (Felton-Edkins et al., 2003; Gomez-Roman et al., 2003). rDNA genes that are transcribed by RNA Pol III do not have E-boxes.
A possible role for E-boxes at MYC target sites could be to stabilize MYC binding once the transcription factor is recruited by other means.

5.3 Consequences of MYC binding

Upon binding at target genes, MYC controls their expression in several ways.
As shown earlier by several studies and by ChIP-seq of histone modifications upon induction of MYC in U2OS cells, MYC controls acetylation of histones (Guccione et al., 2006; Martinato et al., 2008). At repressed genes acetylation of histone 3 and 4 decreases, whereas at activated genes it increases (Fig. 4.19). MYC does this by recruiting histone modifying enzymes (see 1.3.1 and 1.3.3). The augmented acetylation, besides leading to a more open chromatin, could also recruit other co-activators that favor transcription. For example, the BET protein family employs tandem bromo domains to recognize specific acetylated lysines in the N-terminal of histones (Wu and Chiang, 2007). One member of this family is BRD4, which plays an important role in regulating expression of growth-related genes by recruiting p-TEFb (Moon et al., 2005; Yang et al., 2005, 2008). BRD4-mediated recruitment of p-TEFb was shown to be important for expression of the *MYC* gene itself, and inhibition of BRD4 via JQ1 suppresses expression of MYC-dependent genes (Delmore et al., 2011; Filippakopoulos et al., 2010; Venkataraman et al., 2014).

Although no evidence showed that MYC binding associates with BRD4 recruitment, MYC could induce acetylation that in turn enhances the recruitment of p-TEFb via BRD4, providing an indirect and druggable mechanism by which MYC induces gene expression. Yet, MYC was shown to directly interact with p-TEFb leading to increase phopshorylation of serine 2 residues of the CTD of RNA Pol II enhancing transcription elongation (Bouchard et al., 2004; Eberhardy and Farnham, 2002; Gargano et al., 2007; Kanazawa et al., 2003; Rahl et al., 2010). ChIP-seq of RNA Pol II phosphorylated at serine 2 in U2OS cells also showed the presence of an increase amount of serine 2 phosphorylated RNA Pol II in the gene body in cells overexpressing MYC compared to cells with endogenous MYC levels (Fig. 4.18). Two scenarios could lead to this result: on one hand, MYC binding could enhance the activity of the RNA Pol II that is ready and fully-equipped for the efficient transcription of the nascent mRNA. On the other hand, MYC recruitment could prematurely push RNA Pol II into elongation of the mRNA getting the enzyme stuck as transcription proceeds. Indeed, the recruitment of splicing factors and proteins for the export of the mRNA is necessary to remove the mRNA from the template DNA and to ensure its correct elongation. Depletion in human cells of THOC1, a subunit of the THO/TREX complex which mediates the export of the mRNA to the cytoplasm (Rondón et al., 2010), reduces transcription elongation, RNA export and leads to formation of RNA-DNA hybrids (R-loops) which cause genome instability (Domínguez-Sánchez et al., 2011). Interestingly, an shRNA screen looking for genes that are required for MYC function identified *THOC1* as gene whose knockdown prevent cells to undergo MYC-induced apoptosis (Popov et al., 2007). This indicates that in order for MYC to fulfill its function, the transcription apparatus and its co-factors need to be available and in proximity of the transcription site.

Accumulation of the RNA Pol II in the gene body could also be due to the depletion of the nucleotide pool needed for the transcription. Although MYC enhances the nucleotide biosynthetic pathways (Liu et al., 2008; Mannava et al., 2008), it cannot be excluded that the increase in transcription of several hundred genes upon MYC overexpression consumes the available nucleotides causing the stalling of the RNA Pol II while it is transcribing.

To discriminate whether the RNA Pol II accumulated in the gene body of MYC target genes is actually efficiently transcribing or it is stalled because of missing substrates or/and co-factors, techniques that can determine whether RNA Pol II is only DNA-bound or effectively engaged in transcription should be employed. For example the global run-on-sequencing (GRO-seq) assay enables to map and quantify transcriptionally engaged

polymerase genome wide (Core et al., 2008). This assay is based on a nuclear run-on assay, where nascent RNAs that are associated with transcriptionally engaged RNA Pol are elongated in conditions where new initiation is prohibited. To specifically recognize newly elongated RNA, the ribonucleotide analog 5-bromouridine 5′-triphosphate (BrUTP) and an antibody against it are used. The BrU-tag nascent RNAs are then sequenced and mapped to the corresponding genome allowing the identification of genes that are actively transcribed at a specific time or in a defined condition (Core et al., 2008, 2012).

Recent studies showed that high levels of MYC are associated with invasion by MYC of all open promoter and enhancers (Lin et al., 2012; Sabò et al., 2014). The analysis performed in U2OS cells also detected MYC binding at enhancer regions, albeit no correlation with gene expression regulation was present (Walz et al., 2014). The detection of MYC binding at enhancers could be considered either as an off target effect of the ChIP-seq cross-linking procedure or could have a functional meaning. Since enhancers are usually placed at sites far away from the target gene(s), the role of MYC could be that, via its numerous binding partners, it helps to organize and bring together enhancers with the corresponding genes. Indeed, preliminary results in the lab showed that MYC (mainly N-MYC) interacts and colocalizes on chromatin with proteins that are involved in the organization of topologically associated domains (TADs; Anne Carstensen personal communication). TADs are linear fragments of chromatin that fold as three-dimensional structures favoring internal chromatin interactions and joining genes with their regulatory elements (Ciabrelli and Cavalli, 2015).

5.4 Model for MYC-mediated regulation of transcription

The use of the U2OS cell line as model system provided a good strategy to study MYC diverse cellular effects and to determine which factors could account for the different role of MYC at physiological and supraphysiological levels (Fig. 5.1).

At low and physiological levels, MYC binds to and regulates sets of genes that are highly involved in regulating cell growth, such as ribosome biogenesis, protein translation, mitochondrial functions etc. The genes belonging to these sets have high affinity binding sites for MYC in their promoter. Sorting of the genes bound by MYC according to their EC_{50} values for MYC binding, although calculated using only two experimental points, and functional annotation analysis showed that cellular growth-related processes have indeed

low EC_{50} and thus high affinity for MYC (Fig. 4.10). E-box sequences are enriched in most promoters of these genes.

Surprisingly, a specific class of genes with high affinity promoters does not have E-boxes in their promoters. Bioinformatics analysis of the promoter region of genes encoding for the structural components of the ribosomes (RPL and RPS proteins of the large and small, respectively, ribosomal subunit) showed that they do not contain any E-box sequences (Elmar Wolf and Susanne Walz personal communication). However, these sites are highly occupied by MYC and are among the most enriched gene sets that react to MYC depletion. Binding motifs for SP1, GABP and YY1 were enriched in the promoter of the ribosomal genes indicating that protein-protein interactions could be relevant for recruiting MYC at these sites (Elmar Wolf and Susanne Walz personal communication).

At physiological levels of MYC, cellular growth-related genes are fully occupied ("saturated") by MYC, whereas low affinity binding sites are almost completely unoccupied. When MYC levels increase, the high affinity binding sites cannot recruit more MYC and therefore MYC "spill over" to the low affinity binding sites that are free to be occupied. These low affinity binding sites are not enriched for processes that directly control cellular growth, but for processes that allowed the cells to react at conditions that could compromise their survival and might be considered to be more cancer-related. Among these are angiogenesis or blood vessel development, cell death, response to hypoxia and substrate specific transporter activity (Fig. 4.10 and 4.12). The binned plot for occupancy of MYC at endogenous and supraphysiological levels showed that MYC occupies high and low affinity site to the same extent (Fig. 4.4 A). However, MYC ChIP followed by qPCR found that at certain genes MYC bind less even when it is overexpressed (Fig. 4.4 B). This does not compromise the analysis but change the interpretation of the plot. High and low affinity binding sites are not occupied equally but the gap in MYC binding between them is reduced indicating that overexpressed MYC is mainly recruited to low affinity sites.

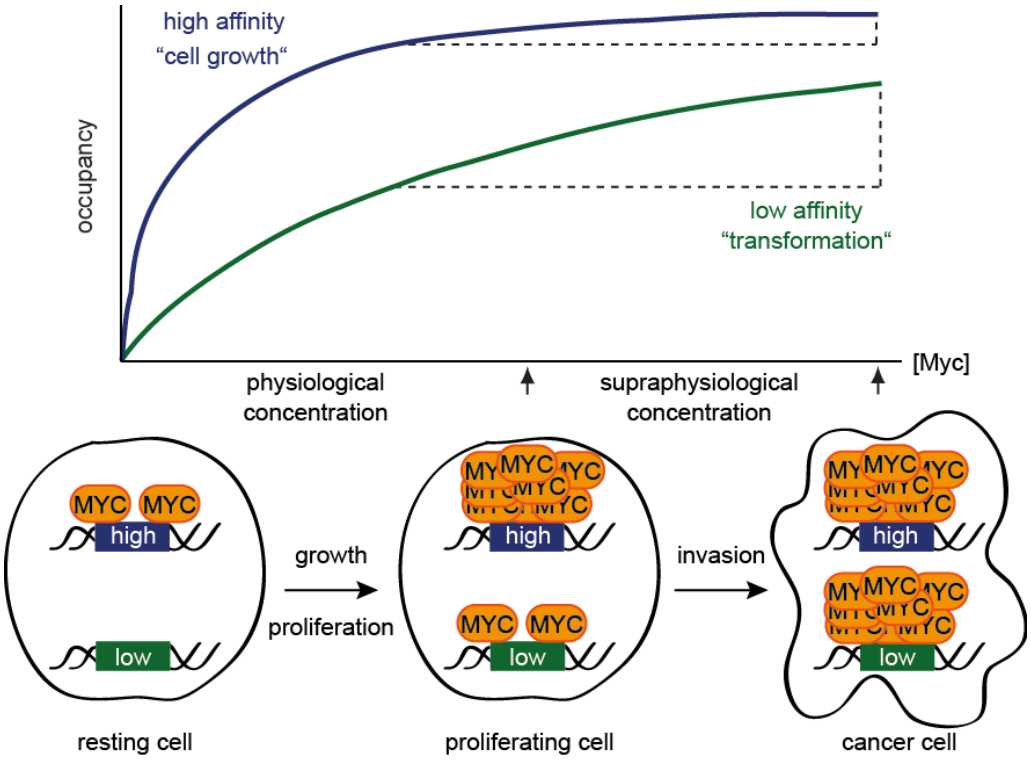


Figure 5.1: Model for MYC function in normal and cancer cells.

Determination of whether the levels of MYC, responsible for the saturation of high affinity binding sites and therefore for the specific regulation of cellular processes that do not belong to the core set of processes regulated by MYC at physiological levels, are reached *in vivo* still need further investigation. Analysis of MYC binding and related changes in gene expression via ChIP-seq and RNA-seq during embryonic development, the wound healing process or the tissue regeneration could provide important evidence. These are all processes that require spatial and temporal coordination between cellular growth, proliferation, migration and signaling. By tuning MYC protein levels, the cells could differentially or uniformly regulate these cellular processes.

The gene expression and the MYC binding analyses performed in U2OS suggested that MYC recruitment and not MYC occupancy is important for the regulation of gene expression when high MYC levels are present in the cell. Indeed, comparison between MYC recruitment and the fold change in expression induced by MYC overexpression showed a significant correlation between these two parameters (Fig. 4.7 A). Furthermore, other mechanisms could be important for the MYC-dependent regulation of transcription. For example, ubiquitination of MYC and the control of its turnover were shown to influence MYC transcriptional activity. Many ubiquitin ligases have been shown to modify

MYC, in most of the cases, at the MYC transcriptional activation domain, underlining the importance of MYC ubiquitination in the control of transcription by MYC (Muratani and Tansey, 2003). SCF(SKP2) ubiquitin ligase ubiquitinates MYC at MYC box II, thereby promoting its turnover and stimulating expression of several target genes (Kim et al., 2003; Von Der Lehr et al., 2003; Zhang et al., 2012b). Moreover, HUWE1 and FBXO28 ubiquiting ligases promotes MYC ubiquitination and its transcriptional function (Adhikary et al., 2005; Cepeda et al., 2013). Several others ligases modify MYC and have a negative impact on MYC activity, however it was recently shown that ubiquitin-dependent turnover of MYC at the promoter of target genes is required to drive productive transcription elongation by RNA Pol II (Jaenicke et al., 2015).

Other modifications could impact on MYC activity at the promoter of target genes, such as phosphorylation and acetylation. MYC contains in the MYC box I two amino acids, T58 and S62 that can be phosphorylated. Phosphorylation of both residues is recognized by the SCF(FBW7) ubiquitin ligase and lead to proteasomal degradation of MYC (Welcker et al., 2004; Yada et al., 2004; Yeh et al., 2004). Phosphorylation of S62 alone, induces the PIN1-mediated isomerization of MYC, which stimulates the recruitment of p300, GCN5 and p-TEFb (Farrell et al., 2013).

Several residues in the MYC protein are also the substrates of enzymes with acetyltransferase activity (Faiola et al., 2005; Patel et al., 2004; Vervoorts et al., 2003). GCN5 and p300 were shown to acetylates MYC increasing its protein stability (Faiola et al., 2005; Patel et al., 2004; Vervoorts et al., 2003), but the acetylated lysines could serve also as a docking site for other proteins (Vervoorts et al., 2006).

5.5 Conclusions

This work could show that promoter affinity and MYC protein levels are important factors that influence MYC activity in the regulation of functionally distinct groups of genes.

At low levels, MYC binds preferentially to genes with high binding affinity. These genes encode for ribosomal components, factors involved in ribosome biogenesis and in general for proteins implicated in cellular growth. They are not regulated at oncogenic MYC levels because they are already fully saturated. On contrary, genes with low affinity binding sites are regulated when MYC in further expressed in proliferating cells. In tumor cells that express high and oncogenic MYC levels, processes with low affinity for MYC binding are for example angiogenesis, cell migration and specific substrate uptake.

Signaling pathways, which control MYC protein levels within the cell, can tune MYC activity in order to achieve regulation of specific genes having different promoter affinity (Fig. 5.2).

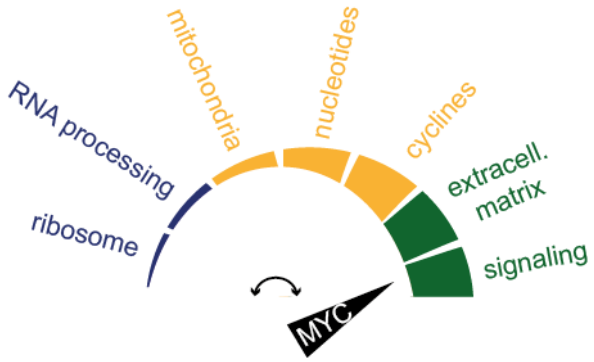


Figure 5. 2: Model for MYC-regulated processes.

The identification of processes that have low affinity binding sites for MYC and are regulated only in cells with oncogenic MYC levels could open a therapeutic window to treat tumors. Drugs targeting these processes could be design so that only cancer but not normal cells are affected. Following this direction, a dominant negative allele of MYC, OMOMYC, have been developed (Soucek et al., 1998). The use of OMOMYC for treatment of tumor mouse models has been successful (Soucek et al., 2004, 2013) and recent finding showed that OMOMYC competes with MYC at low affinity sites, exclusively inhibiting MYC tumor specific gene expression profile while preserving its physiological functions (Lisa Jung personal communication).

Bibliography

Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat. Rev. Genet. *13*, 720–731.

Adhikary, S., and Eilers, M. (2005). Transcriptional regulation and transformation by Myc proteins. Nat. Rev. Mol. Cell Biol. *6*, 635–645.

Adhikary, S., Peukert, K., Karsunky, H., Beuger, V., Lutz, W., Elsa, H., and Mo, T. (2003). Miz1 Is Required for Early Embryonic Development during Gastrulation. *23*, 7648–7657.

Adhikary, S., Marinoni, F., Hock, A., Hulleman, E., Popov, N., Beier, R., Bernard, S., Quarto, M., Capra, M., Goettig, S., et al. (2005). The ubiquitin ligase HectH9 regulates transcriptional activation by Myc and is essential for tumor cell proliferation. Cell *123*, 409–421.

Allen, B.L., and Taatjes, D.J. (2015). The Mediator complex: a central integrator of transcription. Nat. Rev. Mol. Cell Biol. *16*, 155–166.

Arabi, A., Wu, S., Ridderstråle, K., Bierhoff, H., Shiue, C., Fatyol, K., Fahlén, S., Hydbring, P., Söderberg, O., Grummt, I., et al. (2005). c-Myc associates with ribosomal DNA and activates RNA polymerase I transcription. Nat. Cell Biol. *7*, 303–310.

Ayer, D.E., and Eisenman, R.N. (1993). A switch from Myc: Max to Mad: Max heterocomplexes accompanies monocyte/macrophage differentiation. Genes Dev. *7*, 2110–2119.

Bardwell, V.J., and Treisman, R. (1994). The POZ domain : A conserved protein- protein interaction motif. Genes Dev. *8*, 1664–1677.

Barr, L.F., Campbell, S.E., Bochner, B.S., Lung, C., Cells, C., and Dang, C. V (1998). Association of the Decreased Expression of α 3 β 1 Integrin with the Altered Cell : Environmental Interactions and Enhanced Soft Agar Cloning Ability of c- myc - overexpressing Small Cell Lung Cancer Cells Association of the Decreased Expression of Â « 3 /.

Barsyte-Lovejoy, D., Mao, D.Y.L., and Penn, L.Z. (2004). c-Myc represses the proximal promoters of GADD45a and GADD153 by a post-RNA polymerase II recruitment mechanism. Oncogene *23*, 3481–3486.

Basu, S., Liu, Q., Qiu, Y., and Dong, F. (2009). Gfi-1 represses CDKN2B encoding p15 INK4B through interaction with Miz-1. *106*, 1433–1438.

Baudino, T. a., McKay, C., Pendeville-Samain, H., Nilsson, J. a., Maclean, K.H., White, E.L., Davis, A.C., Ihle, J.N., and Cleveland, J.L. (2002). c-Myc is essential for

vasculogenesis and angiogenesis during development and tumor progression. Genes Dev. *16*, 2530–2543.

Blackwell, T.K., Kretzner, L.E.O., Blackwood, E.M., Eisenman, R.N., and Weintraub, H. (1990). Sequence-Specific DNA Binding by the c-Myc Protein. Science (80-.). *250*, 1149–1151.

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. Mol. Cell. Biol. *13*, 5216–5224.

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., Luscher, B., and Eisenman, R.N. (1992). Myc and Max associate in vivo. Genes Dev. *6*, 71–80.

Bouchard, C., Dittrich, O., Kiermaier, A., Dohmann, K., Menkel, A., Eilers, M., and Lüscher, 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 service Regulation of cyclin D2 gene expression by the Myc / Max / Mad network : Myc-dependent TRRAP. 2042–2047.

Bouchard, C., Marquardt, J., Brás, A., Medema, R.H., and Eilers, M. (2004). Myc-induced proliferation and transformation require Akt-mediated phosphorylation of FoxO proteins. EMBO J. *23*, 2830–2840.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. *72*, 248–254.

Brenner, C., Deplus, R., Didelot, C., Loriot, A., Viré, E., De Smet, C., Gutierrez, A., Danovi, D., Bernard, D., Boon, T., et al. (2005). Myc represses transcription through recruitment of DNA methyltransferase corepressor. EMBO J. *24*, 336–346.

Brockmann, M., Poon, E., Berry, T., Carstensen, A., Deubzer, H.E., Rycak, L., Jamin, Y., Thway, K., Robinson, S.P., Roels, F., et al. (2013). Small Molecule Inhibitors of Aurora-A Induce Proteasomal Degradation of N-Myc in Childhood Neuroblastoma. Cancer Cell *24*, 75–89.

Calo, E., and Wysocka, J. (2013). Modification of Enhancer Chromatin: What, How, and Why? Mol. Cell *49*, 825–837.

Cepeda, D., Ng, H.-F., Sharifi, H.R., Mahmoudi, S., Cerrato, V.S., Fredlund, E., Magnusson, K., Nilsson, H., Malyukova, A., Rantala, J., et al. (2013). CDK-mediated activation of the SCF(FBXO) (28) ubiquitin ligase promotes MYC-driven transcription and tumourigenesis and predicts poor survival in breast cancer. EMBO Mol. Med. *5*, 999–1018.

Chao, S.H., and Price, D.H. (2001). Flavopiridol Inactivates P-TEFb and Blocks Most RNA Polymerase II Transcription in Vivo. J. Biol. Chem. *276*, 31793–31799.

Chen, D., Kon, N., Li, M., Zhang, W., Qin, J., and Gu, W. (2005). ARF-BP1/mule is a critical mediator of the ARF tumor suppressor. Cell *121*, 1071–1083.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008). Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. Cell *133*, 1106–1117.

Cheng, S.W., Davies, K.P., Yung, E., Beltran, R.J., Yu, J., and Kalpana, G. V (1999). c-MYC interacts with INI1/hSNF5 and requires the SWI/SNF complex for transactivation function. Nat. Genet. *22*, 102–105.

Ciabrelli, F., and Cavalli, G. (2015). Chromatin-Driven Behavior of Topologically Associating Domains. J. Mol. Biol. *427*, 608–625.

Cole, M.D., and Cowling, V.H. (2008). Transcription-independent functions of MYC: regulation of translation and DNA replication. Nat. Rev. Mol. Cell Biol. *9*, 810–815.

Coller, H. a, Grandori, C., Tamayo, P., Colbert, T., Lander, E.S., Eisenman, R.N., and Golub, T.R. (2000). Expression analysis with oligonucleotide microarrays reveals that MYC regulates genes involved in growth, cell cycle, signaling, and adhesion. Proc. Natl. Acad. Sci. U. S. A. *97*, 3260–3265.

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

Core, L.J., Waterfall, J.J., Gilchrist, D. a, Fargo, D.C., Kwak, H., Adelman, K., and Lis, J.T. (2012). Defining the status of RNA polymerase at promoters. Cell Rep. *2*, 1025–1035.

Corvetta, D., Chayka, O., Gherardi, S., D'Acunto, C.W., Cantilena, S., Valli, E., Piotrowska, I., Perini, G., and Sala, A. (2013). Physical interaction between MYCN oncogene and polycomb repressive complex 2 (PRC2) in neuroblastoma: Functional and therapeutic implications. J. Biol. Chem. *288*, 8332–8341.

Cowling, V.H., and Cole, M.D. (2006). Mechanism of transcriptional activation by the Myc oncoproteins. Semin. Cancer Biol. *16*, 242–252.

Cowling, V.H., and Cole, M.D. (2007). The Myc transactivation domain promotes global phosphorylation of the RNA polymerase II carboxy-terminal domain independently of direct DNA binding. Mol. Cell. Biol. *27*, 2059–2073.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M. a, Frampton, G.M., Sharp, P. a, et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. U. S. A. *107*, 21931–21936.

Cunningham, J.T., Moreno, M. V., Lodi, A., Ronen, S.M., and Ruggero, D. (2014). Protein and nucleotide biosynthesis are coupled by a single rate-limiting enzyme, PRPS2, to drive cancer. Cell *157*, 1088–1103.

Dang, C. V (1999). c-Myc Target Genes Involved in Cell Growth , Apoptosis , and Metabolism. *19*.

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

Dang, C. V. (2013). MYC, metabolism, cell growth, and tumorigenesis. Cold Spring Harb. Perspect. Biol. *5*.

Dang, C. V., O'Donnell, K. a., Zeller, K.I., Nguyen, T., Osthus, R.C., and Li, F. (2006). The c-Myc target gene network. Semin. Cancer Biol. *16*, 253–264.

Davis, a. C., Wims, M., Spotts, G.D., Hann, S.R., and Bradley, a. (1993). A null c-myc mutation causes lethality before 10.5 days of gestation in homozygotes and reduced fertility in heterozygous female mice. Genes Dev. *7*, 671–682.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306–1311.

Delmore, J.E., Issa, G.C., Lemieux, M.E., Rahl, P.B., Shi, J., Jacobs, H.M., Kastritis, E., Gilpatrick, T., Paranal, R.M., Qi, J., et al. (2011). BET bromodomain inhibition as a therapeutic strategy to target c-Myc. Cell *146*, 904–917.

Domínguez-Sánchez, M.S., Barroso, S., Gómez-González, B., Luna, R., and Aguilera, A. (2011). Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. PLoS Genet. *7*, e1002386.

Eberhardy, S.R., and Farnham, P.J. (2002). Myc recruits P-TEFb to mediate the final step in the transcriptional activation of the cad promoter. J. Biol. Chem. *277*, 40156–40162.

Eberharter, A., and Becker, P.B. (2002). Histone acetylation: A switch between repressive and permissive chromatin. Second in review on chromatin dynamics. EMBO Rep. *3*, 224–229.

Eilers, M., and Eisenman, R.N. (2008). Myc's broad reach. Genes Dev. *22*, 2755–2766.

Von Eyss, B., Jaenicke, L.A., Kortlever, R.M., Royla, N., Wiese, K.E., Letschert, L., McDuffus, L.A., Sauer, M., Rosenwald, A., Evan, G.I., et al. (2015). A MYC-driven change in mitochondrial dynamics limits YAP/TAZ function in mammary epithelial cells and breast cancer. Cancer Cell.

Faiola, F., Liu, X., Lo, S., Pan, S., Zhang, K., Lymar, E., Farina, A., and Martinez, E. (2005). Dual regulation of c-Myc by p300 via acetylation-dependent control of Myc protein turnover and coactivation of Myc-induced transcription. Mol. Cell. Biol. *25*, 10220–10234.

Farrell, A.S., Pelz, C., Wang, X., Daniel, C.J., Wang, Z., Su, Y., Janghorban, M., Zhang, X., Morgan, C., Impey, S., et al. (2013). Pin1 regulates the dynamics of c-Myc DNA binding to facilitate target gene regulation and oncogenesis. Mol. Cell. Biol. *33*, 2930–2949.

Felton-Edkins, Z. a., Kenneth, N.S., P. Brown, T.R., and Daly, N.L. (2003). Direct Regulation of RNA Polymerase III Transcription by RB, p53 and c-Myc. Cell Cycle *2*, 180–183.

Feng, X.H., Liang, Y.Y., Liang, M., Zhai, W., and Lin, X. (2002). Direct interaction of c-Myc with Smad2 and Smad3 to inhibit TGF-??-mediated induction of the CDK inhibitor p15Ink4B. Mol. Cell *9*, 133–143.

Fernandez, P.C., Frank, S.R., Wang, L., Schroeder, M., Liu, S., Greene, J., Cocito, A., and Amati, B. (2003). Genomic targets of the human c-Myc protein. Genes Dev. *17*, 1115–1129.

Fieber, W., Schneider, M.L., Matt, T., Kräutler, B., Konrat, R., and Bister, K. (2001). Structure, function, and dynamics of the dimerization and DNA-binding domain of oncogenic transcription factor v-Myc. J. Mol. Biol. *307*, 1395–1410.

Filippakopoulos, P., Qi, J., Picaud, S., Shen, Y., Smith, W.B., Fedorov, O., Morse, E.M., Keates, T., Hickman, T.T., Felletar, I., et al. (2010). Selective inhibition of BET bromodomains. Nature *468*, 1067–1073.

Frank, S.R., Parisi, T., Taubert, S., Fernandez, P., Fuchs, M., Chan, H.-M., Livingston, D.M., and Amati, B. (2003). MYC recruits the TIP60 histone acetyltransferase complex to chromatin. EMBO Rep. *4*, 575–580.

Frye, M., Gardner, C., Li, E.R., Arnold, I., and Watt, F.M. (2003). Evidence that Myc activation depletes the epidermal stem cell compartment by modulating adhesive interactions with the local microenvironment. Development *130*, 2793–2808.

Fuchs, M., Gerber, J., Drapkin, R., Sif, S., Ikura, T., Ogryzko, V., Lane, W.S., Nakatani, Y., and Livingston, D.M. (2001). The p400 complex is an essential E1A transformation target. Cell *106*, 297–307.

Gabay, M., Li, Y., and Felsher, D.W. (2014). MYC activation is a hallmark of cancer initiation and maintenance. Cold Spring Harb. Perspect. Med. *4*, 1–14.

Gallant, P., Shiio, Y., Cheng, P.F., Parkhurst, S.M., and Eisenman, R.N. (1996). Myc and Max homologs in Drosophila. Science (80-.). *274*, 1523–1527.

Gargano, B., Amente, S., Majello, B., and Lania, L. (2007). P-TEFb is a crucial co-factor for Myc transactivation. Cell Cycle *6*, 2031–2037.

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. Proc. Natl. Acad. Sci. U. S. A. *98*, 4510–4515.

Gebhardt, A., Frye, M., Herold, S., Benitah, S.A., Braun, K., Samans, B., Watt, F.M., Elsässer, H.P., and Eilers, M. (2006). Myc regulates keratinocyte adhesion and differentiation via complex formation with Miz1. J. Cell Biol. *172*, 139–149.

Gebhardt, A., Kosan, C., Herkert, B., Möröy, T., Lutz, W., Eilers, M., and Elsässer, H.-P. (2007). Miz1 is required for hair follicle structure and hair morphogenesis. J. Cell Sci. *120*, 2586–2593.

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*, 1698–1701.

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. Nat. Cell Biol. *7*, 311–318.

Grewal, S.S., Li, L., Orian, A., Eisenman, R.N., and Edgar, B. a (2005). Myc-dependent regulation of ribosomal RNA synthesis during Drosophila development. Nat. Cell Biol. *7*, 295–302.

Guccione, E., Martinato, F., Finocchiaro, G., Luzi, L., Tizzoni, L., Dall' Olio, V., Zardo, G., Nervi, C., Bernard, L., and Amati, B. (2006). Myc-binding-site recognition in the human genome is determined by chromatin context. Nat. Cell Biol. *8*, 764–770.

Guo, J., Li, T., Schipper, J., Fordjour, F.K., Nilson, K. a, Jeffrey, J., Gordân, R., and Price, D.H. (2014). Sequence specificity incompletely defines the genome-wide occupancy of Myc. 1–14.

Hannon, G.J., and Beach, D. (1994). p15INK4B is a potential effector of TGF-beta-induced cell cycle arrest. Nature *371*, 257–261.

Heidemann, M., Hintermair, C., Voß, K., and Eick, D. (2013). Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. Biochim. Biophys. Acta - Gene Regul. Mech. *1829*, 55–62.

Heinz, S., Romanoski, C.E., Benner, C., and Glass, C.K. (2015). The selection and function of cell type-specific enhancers. Nat Rev Mol Cell Biol *16*, 144–154.

Herkert, B., and Eilers, M. (2010). Transcriptional repression: the dark side of myc. Genes Cancer *1*, 580–586.

Herkert, B., Dwertmann, A., Herold, S., Abed, M., Naud, J.F., Finkernagel, F., Harms, G.S., Orian, A., Wanzel, M., and Eilers, M. (2010). The Arf tumor suppressor protein inhibits Miz1 to suppress cell adhesion and induce apoptosis. J. Cell Biol. *188*, 905–918.

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. Proc. Natl. Acad. Sci. U. S. A. *97*, 2229–2234.

Herold, S., Wanzel, M., Beuger, V., Frohme, C., Beul, D., Hillukkala, T., Syvaoja, J., Saluz, H.P., Haenel, F., and Eilers, M. (2002). Negative regulation of the mammalian UV response by Myc through association with Miz-1. Mol. Cell *10*, 509–521.

Herold, S., Hock, A., Herkert, B., Berns, K., Mullenders, J., Beijersbergen, R., Bernards, R., and Eilers, M. (2008). Miz1 and HectH9 regulate the stability of the checkpoint protein, TopBP1. EMBO J. *27*, 2851–2861.

Horiuchi, D., Kusdra, L., Huskey, N.E., Chandriani, S., Lenburg, M.E., Gonzalez-Angulo, a. M., Creasman, K.J., Bazarov, a. V., Smyth, J.W., Davis, S.E., et al. (2012). MYC pathway activation in triple-negative breast cancer is synthetic lethal with CDK inhibition. J. Exp. Med. *209*, 679–696.

Hu, J., Banerjee, A., and Goss, D.J. (2005). Assembly of b/HLH/z proteins cMyc, Max and Mad1 with Cognate DNA: Importance of Protein-Protein and Protein-DNA Interactions. Biochemistry *44*, 11855–11863.

Inghirami, G., Grignani, F., Sternas, L., Lombardi, L., Knowles, D.M., and Dalla-Favera, R. (1990). Down-regulation of LFA-1 adhesion receptors by C-myc oncogene in human B lymphoblastoid cells. Science *250*, 682–686.

Inoue, S., Hao, Z., Elia, A.J., Cescon, D., Zhou, L., Silvester, J., Snow, B., Harris, I.S., Sasaki, M., Li, W.Y., et al. (2013). Mule/Huwe1/Arf-BP1 suppresses Ras-driven tumorigenesis by preventing c-Myc/Miz1-mediated down-regulation of p21 and p15. Genes Dev. *27*, 1101–1114.

Jaenicke, L.A., von Eyss, B., Carstensen, A., Wolf, E., Xu, W., Greifenberg, A., Geyer, M., Eilers, M., and Popov, N. (2015). Ubiquitin-dependent turnover of MYC antagonizes MYC/PAF1C complex accumulation to drive transcriptional elongation. Mol. Cell.

Johnston, L. a., Prober, D. a., Edgar, B. a., Eisenman, R.N., and Gallant, P. (1999). Drosophila myc regulates cellular growth during development. Cell *98*, 779–790.

Kanazawa, S., Soucek, L., Evan, G., Okamoto, T., and Peterlin, B.M. (2003). c-Myc recruits P-TEFb for transcription, cellular proliferation and apoptosis. Oncogene *22*, 5707–5711.

Kawauchi, D., Robinson, G., Uziel, T., Gibson, P., Rehg, J., Finkelstein, D., Qu, C., Pounds, S., Ellison, D.W., and Richard, J. (2012). A mouse model of the most aggressive subgroup of human medulloblastoma. Cancer Cell *21*, 168–180.

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. Proc. Natl. Acad. Sci. U. S. A. *104*, 14917–14922.

Kim, S.Y., Herbst, A., Tworkowski, K. a, Salghetti, S.E., Tansey, W.P., Brook, S., and York, N. (2003). Skp2 Regulates Myc Protein Stability and Activity. *11*, 1177–1188.

Knoepfler, P.S., Zhang, X., Cheng, P.F., Gafken, P.R., McMahon, S.B., and Eisenman, R.N. (2006). Myc influences global chromatin structure. EMBO J. *25*, 2723–2734.

Koch, B., Sanchez, S., Schmidt, C.K., Swiersy, A., Jackson, S.P., and Schmidt, O.G. (2014). Confinement and Deformation of Single Cells and Their Nuclei Inside Size-Adapted Microtubes. Adv. Healthc. Mater. 1–6.

Koh, C.M., Bezzi, M., Low, D.H.P., Ang, W.X., Teo, S.X., Gay, F.P.H., Al-Haddawi, M., Tan, S.Y., Osato, M., Sabò, A., et al. (2015). MYC regulates the core pre-mRNA splicing machinery as an essential step in lymphomagenesis. Nature *523*, 96–100.

Kohl, N.E., Kanda, N., Schreck, R.R., Bruns, G., Latt, S. a, Gilbert, F., and Alt, F.W. (1983). Transposition and amplification of oncogene-related sequences in human neuroblastomas. Cell *35*, 359–367.

Kurland, J.F., and Tansey, W.P. (2008). Myc-mediated transcriptional repression by recruitment of histone deacetylase. Cancer Res. *68*, 3624–3629.

Langmead, B. (2010). Alignment with Bowtie. 1–24.

Von Der Lehr, N., Johansson, S., Wu, S., Bahram, F., Castell, A., Cetinkaya, C., Hydbring, P., Weidung, I., Nakayama, K., Nakayama, K.I., et al. (2003). The F-box protein Skp2 participates in c-Myc proteosomal degradation and acts as a cofactor for c-Myc-regulated transcription. Mol. Cell *11*, 1189–1200.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Licchesi, J.D.F., Neste, L. Van, Tiwari, V.K., Cope, L., Lin, X., Baylin, S.B., and Herman, J.G. (2010). Transcriptional regulation of Wnt inhibitory factor-1 by Miz-1 / c-Myc. Oncogene *29*, 5923–5934.

Lin, C.Y., Lovén, 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.

Liu, Q., Basu, S., Qiu, Y., Tang, F., and Dong, F. (2010). A role of Miz-1 in Gfi-1-mediated transcriptional repression of CDKN1A. Oncogene *29*, 2843–2852.

Liu, Y.C., Li, F., Handler, J., Huang, C.R.L., Xiang, Y., Neretti, N., Sedivy, J.M., Zeller, K.I., and Dang, C. V. (2008). Global regulation of nucleotide biosynthetic genes by c-myc. PLoS One *3*.

Lorenzin, F., Benary, U., Baluapuri, A., Walz, S., Jung, L.A., von Eyss, B., Kisker, C., Wolf, J., Eilers, M., and Wolf, E. (2016). Different promoter affinities account for specificity in MYC-dependent gene regulation. Elife *5*, 1–35.

Lovén, J., Orlando, D. a., Sigova, A. a., Lin, C.Y., Rahl, P.B., Burge, C.B., Levens, D.L., Lee, T.I., and Young, R. a. (2012). Revisiting global gene expression analysis. Cell *151*, 476–482.

Lüscher, B., and Vervoorts, J. (2012). Regulation of gene transcription by the oncoprotein MYC. Gene *494*, 145–160.

Mannava, S., Grachtchouk, V., Wheeler, L.J., Im, M., Zhuang, D., Slavina, E.G., Mathews, C.K., Shewach, D.S., and Nikiforov, M. a. (2008). Direct role of nucleotide metabolism in C-MYC-dependent proliferation of melanoma cells. Cell Cycle *7*, 2392–2400.

Mao, D.Y.L., Watson, J.D., Yan, P.S., Barsyte-Lovejoy, D., Khosravi, F., Wong, W.W.-L., Franham, P.J., Huang, T.H.-M., and Penn, L.Z. (2003). Analysis of Myc Bound Loci Identified by CpG Island Arrays Shows that Max Is Essential for Myc-Dependent Repression. Curr. Biol. *13*, 882–886.

Marshall, G.M., Gherardi, S., Xu, N., Neiron, Z., Trahair, T., Scarlett, C.J., Chang, D.K., Liu, P.Y., Jankowski, K., Iraci, N., et al. (2010). Transcriptional upregulation of histone deacetylase 2 promotes Myc-induced oncogenic effects. 5957–5968.

Martinato, F., Cesaroni, M., Amati, B., and Guccione, E. (2008). Analysis of myc-induced histone modifications on target chromatin. PLoS One *3*.

McMahon, S.B., Van Buskirk, H. a., Dugan, K. a., Copeland, T.D., and Cole, M.D. (1998). The novel ATM-related protein TRRAP is an essential cofactor for the c- Myc and E2F oncoproteins. Cell *94*, 363–374.

McMahon, S.B., Wood, M. a, and Cole, M.D. (2000). The essential cofactor TRRAP recruits the histone acetyltransferase hGCN5 to c-Myc. Mol. Cell. Biol. *20*, 556–562.

Menssen, A., and Hermeking, H. (2002). Characterization of the c-MYC-regulated transcriptome by SAGE: identification and analysis of c-MYC target genes. Proc. Natl. Acad. Sci. U. S. A. *99*, 6274–6279.

Mertz, J. a, Conery, A.R., Bryant, B.M., Sandy, P., Balasubramanian, S., Mele, D. a, Bergeron, L., and Sims, R.J. (2011). Targeting MYC dependence in cancer by inhibiting BET bromodomains. Proc. Natl. Acad. Sci. U. S. A. *108*, 16669–16674.

Moon, K.J., Mochizuki, K., Zhou, M., Jeong, H.S., Brady, J.N., and Ozato, K. (2005). The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. Mol. Cell *19*, 523–534.

Morrish, F., Neretti, N., Sedivy, J.M., and Hockenbery, D.M. (2008). The oncogene c-Myc coordinates regulation of metabolic networks to enable rapid cell cycle entry. Cell Cycle *7*, 1054–1066.

Muratani, M., and Tansey, W.P. (2003). How the ubiquitin-proteasome system controls transcription. Nat. Rev. Mol. Cell Biol. *4*, 192–201.

Murphy, J.D., Junttila, M.R., Pouyet, L., Karnezis, A., Bui, D. a, Brown-swigart, L., Johnson, L., and Gerard, I. (2009). Distinct thresholds govern Myc's biological output in vivo. *14*, 447–457.

Nair, S.K., and Burley, S.K. (2003). X-ray structures of Myc-Max and Mad-Max recognizing DNA: Molecular bases of regulation by proto-oncogenic transcription factors. Cell *112*, 193–205.

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.

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.

Northcott, P. a., Korshunov, A., Witt, H., Hielscher, T., Eberhart, C.G., Mack, S., Bouffet, E., Clifford, S.C., Hawkins, C.E., French, P., et al. (2011). Medulloblastoma comprises four distinct molecular variants. J. Clin. Oncol. *29*, 1408–1414.

Orian, A., Steensel, B. Van, Delrow, J., Bussemaker, H.J., Li, L., Sawado, T., Williams, E., Loo, L.W.M., Cowley, S.M., Yost, C., et al. (2003). Genomic binding by the. Genes Dev. 1101–1114.

Patel, J.H., Du, Y., Ard, P.G., Carella, B., Chen, C., Rakowski, C., Chatterjee, C., Lieberman, P.M., Lane, W.S., Blobel, G. a, et al. (2004). The c-MYC Oncoprotein Is a Substrate of the Acetyltransferases hGCN5 / PCAF and The c-MYC Oncoprotein Is a Substrate of the Acetyltransferases hGCN5 / PCAF and TIP60. *24*, 10826–10834.

Perna, D., Fagà, G., Verrecchia, a, Gorski, M.M., Barozzi, I., Narang, V., Khng, J., Lim, K.C., Sung, W.-K., Sanges, R., et al. (2012). Genome-wide mapping of Myc binding and gene regulation in serum-stimulated fibroblasts. Oncogene *31*, 1695–1709.

Peter, S., Bultinck, J., Myant, K., Jaenicke, L. a, Walz, S., Muller, J., Gmachl, M., Treu, M., Boehmelt, G., Ade, C.P., et al. (2014). Tumor cell-specific inhibition of MYC function using small molecule inhibitors of the HUWE1 ubiquitin ligase. EMBO Mol. Med. *6*, 1525–1541.

Peterlin, B.M., and Price, D.H. (2006). Controlling the elongation phase of transcription with P-TEFb. Mol. Cell *23*, 297–305.

Peukert, K., Staller, P., Schneider, A., Carmichael, G., and Eilers, M. (1997). An alternative pathway for gene regulation by Myc. *16*, 5672–5686.

Phan, R.T., Saito, M., Basso, K., Niu, H., and Dalla-Favera, R. (2005). BCL6 interacts with the transcription factor Miz-1 to suppress the cyclin-dependent kinase inhibitor p21 and cell cycle arrest in germinal center B cells. Nat. Immunol. *6*, 1054–1060.

Poortinga, G., Hannan, K.M., Snelling, H., Walkley, C.R., Jenkins, A., Sharkey, K., Wall, M., Brandenburger, Y., Palatsides, M., Pearson, R.B., et al. (2004). MAD1 and c-MYC regulate UBF and rDNA transcription during granulocyte differentiation. EMBO J. *23*, 3325–3335.

Popov, N., Wanzel, M., Madiredjo, M., Zhang, D., Beijersbergen, R., Bernards, R., Moll, R., Elledge, S.J., and Eilers, M. (2007). The ubiquitin-specific protease USP28 is required for MYC stability. Nat. Cell Biol. *9*, 765–774.

Qi, Y., Gregory, M. a, Li, Z., Brousal, J.P., West, K., and Hann, S.R. (2004). p19ARF directly and differentially controls the functions of c-Myc independently of p53. Nature *431*, 712–717.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R. a, McCuine, S., Burge, C.B., Sharp, P. a, and Young, R. a (2010). c-Myc regulates transcriptional pause release. Cell *141*, 432–445.

Van Riggelen, J., Yetil, A., and Felsher, D.W. (2010). MYC as a regulator of ribosome biogenesis and protein synthesis. Nat. Rev. Cancer *10*, 301–309.

Van Riggelen, J., Müller, J., Otto, T., Beuger, V., Yetil, A., Choi, P.S., Kosan, C., Möröy, T., Felsher, D.W., and Eilers, M. (2010). The interaction between Myc and Miz1 is required to antagonize TGFβ-dependent autocrine signaling during lymphoma formation and maintenance. Genes Dev. *24*, 1281–1294.

Rondón, A.G., Jimeno, S., and Aguilera, A. (2010). The interface between transcription and mRNP export: from THO to THSC/TREX-2. Biochim. Biophys. Acta *1799*, 533–538.

Sabò, 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.

Sauvé, S., Naud, J.F., and Lavigne, P. (2007). The Mechanism of Discrimination between Cognate and Non-Specific DNA by Dimeric b/HLH/LZ Transcription Factors. J. Mol. Biol. *365*, 1163–1175.

Schuhmacher, M., Kohlhuber, F., Hölzel, M., Kaiser, C., Burtscher, H., Jarsch, M., Bornkamm, G.W., Laux, G., Polack, a, Weidle, U.H., et al. (2001). The transcriptional program of a human B cell line in response to Myc. Nucleic Acids Res. *29*, 397–406.

Sears, R., Nuckolls, F., Haura, E., Taya, Y., Tamai, K., and Nevins, J.R. (2000). Multiple Ras-dependent phosphorylation pathways regulate Myc protein stability. Genes Dev. *14*, 2501–2514.

Seitz, V., Butzhammer, P., Hirsch, B., Hecht, J., Gütgemann, I., Ehlers, A., Lenze, D., Oker, E., Sommerfeld, A., von der Wall, E., et al. (2011). Deep sequencing of MYC DNA-Binding sites in Burkitt lymphoma. PLoS One *6*.

Seoane, J., Pouponnot, C., Staller, P., Schader, M., Eilers, M., and Massagué, J. (2001). TGF β influences Myc , Miz-1 and Smad to control the CDK inhibitor. *3*.

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.

Shchors, K., Shchors, E., Rostker, F., Lawlor, E.R., Brown-Swigart, L., and Evan, G.I. (2006). The Myc-dependent angiogenic switch in tumors is mediated by interleukin 1β. Genes Dev. *20*, 2527–2538.

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. J. Virol. *31*, 514–521.

Shim, H., Dolde, C., Lewis, B.C., Wu, C.S., Dang, G., Jungmann, R. a, Dalla-Favera, R., and Dang, C. V (1997). c-Myc transactivation of LDH-A: implications for tumor metabolism and growth. Proc. Natl. Acad. Sci. U. S. A. *94*, 6658–6663.

Soucek, L., Helmer-Citterich, M., Sacco, a, Jucker, R., Cesareni, G., and Nasi, S. (1998). Design and properties of a Myc derivative that efficiently homodimerizes. Oncogene *17*, 2463–2472.

Soucek, L., Nasi, S., and Evan, G.I. (2004). Omomyc expression in skin prevents Myc-induced papillomatosis. Cell Death Differ. *11*, 1038–1045.

Soucek, L., Whitfield, J., Martins, C.P., Finch, A.J., Murphy, D.J., Sodir, N.M., Karnezis, A.N., Swigart, L.B., Nasi, S., and Evan, G.I. (2008). Modelling Myc inhibition as a cancer therapy. Nature *455*, 679–683.

Soucek, L., Whitfield, J.R., Sodir, N.M., Massó-Vallés, D., Serrano, E., Karnezis, A.N., Swigart, L.B., and Evan, G.I. (2013). Inhibition of Myc family proteins eradicates KRas-driven lung cancer in mice. Genes Dev. *27*, 504–513.

Staller, P., Peukert, K., Kiermaier, A., Seoane, J., Lukas, J., Karsunky, H., Bartek, J., Massagué, J., Hänel, F., and Eilers, M. (2001). Repression of p15 INK4b expression by Myc through association with Miz-1. *3*.

Stead, M.A., Trinh, C.H., Garnett, J.A., Carr, S.B., Baron, A.J., Edwards, T.A., and Wright, S.C. (2007). C OMMUNICATION A Beta-Sheet Interaction Interface Directs the Tetramerisation of the Miz-1 POZ Domain. 820–826.

Steiger, D., Furrer, M., Schwinkendorf, D., and Gallant, P. (2008). Max-independent functions of Myc in Drosophila melanogaster. Nat. Genet. *40*, 1084–1091.

Stone, J., de Lange, T., Ramsay, G., Jakobovits, E., Bishop, J.M., Varmus, H., and Lee, W. (1987). Definition of regions in human c-myc that are involved in transformation and nuclear localization. Mol. Cell. Biol. *7*, 1697–1709.

Thomas, L.R., Wang, Q., Fesik, S.W., Tansey, W.P., Thomas, L.R., Wang, Q., Grieb, B.C., Phan, J., Foshage, A.M., Sun, Q., et al. (2015). Interaction with WDR5 Promotes Target Gene Recognition and Tumorigenesis by MYC Article Interaction with WDR5 Promotes Target Gene Recognition and Tumorigenesis by MYC. 1–13.

Uribesalgo, I., Buschbeck, M., Gutiérrez, A., Teichmann, S., Demajo, S., Kuebler, B., Nomdedéu, J.F., Martín-Caballero, J., Roma, G., Benitah, S.A., et al. (2011). E-box-independent regulation of transcription and differentiation by MYC. Nat. Cell Biol. *13*, 1443–1449.

Venkataraman, S., Alimova, I., Balakrishnan, I., Harris, P., Birks, D.K., Griesinger, A., Amani, V., Cristiano, B., Remke, M., Taylor, M.D., et al. (2014). Inhibition of BRD4 attenuates tumor cell self-renewal and suppresses stem cell signaling in MYC driven medulloblastoma. Oncotarget *5*, 2355–2371.

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. J. Virol. *42*, 773–779.

Vervoorts, J., Lüscher-Firzlaff, J.M., Rottmann, S., Lilischkis, R., Walsemann, G., Dohmann, K., Austen, M., and Lüscher, B. (2003). Stimulation of c-MYC transcriptional activity and acetylation by recruitment of the cofactor CBP. EMBO Rep. *4*, 484–490.

Vervoorts, J., Lüscher-Firzlaff, J., and Lüscher, B. (2006). The ins and outs of MYC regulation by posttranslational mechanisms. J. Biol. Chem. *281*, 34725–34729.

Vita, M., and Henriksson, M. (2006). The Myc oncoprotein as a therapeutic target for human cancer. Semin. Cancer Biol. *16*, 318–330.

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.

Wanzel, M., Russ, A.C., Kleine-Kohlbrecher, D., Colombo, E., Pelicci, P.-G., and Eilers, M. (2008). A ribosomal protein L23-nucleophosmin circuit coordinates Miz1 function with cell growth. Nat. Cell Biol. *10*, 1051–1061.

Weber, A., Marquardt, J., Elzi, D., Forster, N., Starke, S., Glaum, A., Yamada, D., Defossez, P.-A., Delrow, J., Eisenman, R.N., et al. (2008). Zbtb4 represses transcription of P21CIP1 and controls the cellular response to p53 activation. EMBO J. *27*, 1563–1574.

Welcker, M., Orian, A., Jin, J., Grim, J.E., Harper, J.W., Eisenman, R.N., and Clurman, B.E. (2004). The Fbw7 tumor suppressor regulates glycogen synthase kinase 3 phosphorylation-dependent c-Myc protein degradation. Proc. Natl. Acad. Sci. U. S. A. *101*, 9085–9090.

Wit, E. De, and Laat, W. De (2012). A decade of 3C technologies : insights into nuclear organization. 11–24.

Wolf, E., Gebhardt, A., Kawauchi, D., Walz, S., von Eyss, B., Wagner, N., Renninger, C., Krohne, G., Asan, E., Roussel, M.F., et al. (2013). Miz1 is required to maintain autophagic flux. Nat. Commun. *4*, 2535.

Wolf, E., Lin, C.Y., Eilers, M., and Levens, D.L. (2014). Taming of the beast: shaping Myc-dependent amplification. Trends Cell Biol. 1–8.

Wood, M. a, McMahon, S.B., and Cole, M.D. (2000). An ATPase/helicase complex is an essential cofactor for oncogenic transformation by c-Myc. Mol. Cell *5*, 321–330.

Wu, S.Y., and Chiang, C.M. (2007). The double bromodomain-containing chromatin adaptor Brd4 and transcriptional regulation. J. Biol. Chem. *282*, 13141–13145.

Xu, D., Popov, N., Hou, M., Wang, Q., Björkholm, M., Gruber, a, Menkel, a R., and Henriksson, M. (2001). Switch from Myc/Max to Mad1/Max binding and decrease in histone acetylation at the telomerase reverse transcriptase promoter during differentiation of HL60 cells. Proc. Natl. Acad. Sci. U. S. A. *98*, 3826–3831.

Yada, M., Hatakeyama, S., Kamura, T., Nishiyama, M., Tsunematsu, R., Imaki, H., Ishida, N., Okumura, F., Nakayama, K., and Nakayama, K.I. (2004). Phosphorylation-dependent degradation of c-Myc is mediated by the F-box protein Fbw7. EMBO J. *23*, 2116–2125.

Yang, B.S., Geddes, T.J., Pogulis, R.J., de Crombrugghe, B., and Freytag, S.O. (1991). Transcriptional suppression of cellular gene expression by c-Myc. Mol. Cell. Biol. *11*, 2291–2295.

Yang, Z., Yik, J.H.N., Chen, R., He, N., Moon, K.J., Ozato, K., and Zhou, Q. (2005). Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. Mol. Cell *19*, 535–545.

Yang, Z., He, N., and Zhou, Q. (2008). Brd4 recruits P-TEFb to chromosomes at late mitosis to promote G1 gene expression and cell cycle progression. Mol. Cell. Biol. *28*, 967–976.

Yeh, E., Cunningham, M., Arnold, H., Chasse, D., Monteith, T., Ivaldi, G., Hahn, W.C., Stukenberg, P.T., Shenolikar, S., Uchida, T., et al. (2004). A signalling pathway controlling c-Myc degradation that impacts oncogenic transformation of human cells. Nat. Cell Biol. *6*, 308–318.

Yin, X., Giap, C., Lazo, J.S., and Prochownik, E. V (2003). Low molecular weight inhibitors of Myc-Max interaction and function. Oncogene *22*, 6151–6159.

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 Biol. *4*, R69.

Zeller, K.I., Zhao, X., Lee, C.W., Chiu, K., 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. Proc. Natl. Acad. Sci. U. S. A. *103*, 17834–17839.

Zhang, D.W., Rodríguez-Molina, J.B., Tietjen, J.R., Nemec, C.M., and Ansari, A.Z. (2012a). Emerging Views on the CTD Code. Genet. Res. Int. *2012*, 347214.

Zhang, H., Gao, P., Fukuda, R., Kumar, G., Krishnamachary, B., Zeller, K.I., Dang, C.V., and Semenza, G.L. (2007). HIF-1 Inhibits Mitochondrial Biogenesis and Cellular Respiration in VHL-Deficient Renal Cell Carcinoma by Repression of C-MYC Activity. Cancer Cell *11*, 407–420.

Zhang, Q., Spears, E., Boone, D.N., Li, Z., Gregory, M. a, and Hann, S.R. (2012b). Domain-specific c-MYC ubiquitylation controls c-MYC transcriptional and Apoptotic Activity. Proc. Natl Acad. Sci. USA.

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

Zhang, X., Chen, X., Lin, J., Lwin, T., Wright, G., Moscinski, L.C., Dalton, W.S., Seto, E., Wright, K., Sotomayor, E., et al. (2012d). Myc represses miR-15a/miR-16-1 expression through recruitment of HDAC3 in mantle cell and other non-Hodgkin B-cell lymphomas. Oncogene *31*, 3002–3008.

Zhang, X.-Y., Varthi, M., Sykes, S.M., Philips, C., Warzecha, C., Zhu, W., Wyce, A., Thorne, A.W., Berger, S.L., and McMahon, S.B. (2008a). The putative cancer stem cell marker USP22 is a subunit of the human SAGA complex required for activator-driven transcription and cell cycle progression. Mol. Cell *29*, 102–111.

Zhang, Y., Liu, T., Meyer, C. a, Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008b). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zindy, F., Eischen, C.M., Randle, D.H., Kamijo, T., Cleveland, J.L., Sherr, C.J., and Roussel, M.F. (1998). Myc signaling via the ARF tumor suppressor regulates p53-dependent apoptosis and immortalization. Genes Dev. *12*, 2424–2433.

Zippo, A., De Robertis, A., Serafini, R., and Oliviero, S. (2007). PIM1-dependent phosphorylation of histone H3 at serine 10 is required for MYC-dependent transcriptional activation and oncogenic transformation. Nat. Cell Biol. *9*, 932–944.

Publication List

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.
* These authors contributed equally to this work

Lorenzin, F., Benary, U., Baluapuri, A., Walz, S., Jung, L.A., von Eyss, B., Kisker, C., Wolf, J., Eilers, M., and Wolf, E. (2016). Different promoter affinities account for specificity in MYC-dependent gene regulation. Elife *5*, 1–35.

Curriculum Vitae

Personal data

Francesca Lorenzin
Via F. Maccani 13/a
38050 Castelnuovo (TN)
Italy

Born May 18th 1987 in Borgo Valsugana (TN), Italy

Education

09/2001 – 07/2006	High School Degree in "Technical expert, biological specialization", Istituto Paritario Ivo de Carneri, Civezzano, Italy
10/2006 – 07/2009	Bachelor Degree in Molecular Biology, Università degli Studi di Padova, Padova, Italy
10/2009 – 07/2011	Master Degree in Sanitary Biology, Università degli Studi di Padova, Padova, Italy
09/2011 – 07/2015	PhD, laboratory of Prof. Dr. Martin Eilers, Department of Biochemistry and Molecular Biology, Universität Würzburg, Würzburg, Germany
12/2015- at present	Research fellow, laboratory of Francesca Demichelis, CIBIO, Università di Trento, Trento, Italy

Acknowledgements

I would like to thank my supervisor, Prof. Dr. Martin Eilers, for the encouragement, the advice and the guidance he provided throughout my time as a PhD student. His supervision helped, and will help, me in all the time of my scientific career.

I would like to express my sincere gratitude to Dr. Elmar Wolf, for the continuous support, his motivation and his encyclopedic knowledge. He inspired me constantly with such enthusiasm in regard to research.

I also thank Prof. Dr. Jöerg Vogel and Dr. Bruno Amati for their insightful comments as my thesis committee.

I thank my fellow lab mates: all the members of the Eilers, Gallant and Murphy group who assisted me with technical help and countless discussions.
Especially, I would like to thank Jiajia, Anne, Eva, Francesca D. for the wonderful girl's trips; Giacomo for sharing so many morning coffees/fruits with me; Silvia for the brief appearance. I would not have made it without you all!

I want to express my gratitude towards the Graduate School of Life Science for their support and for providing soft skill trainings.

Last but not the least, I would like to thank Lorenzo and my whole family for supporting me spiritually throughout my PhD time and my life in general.

Affidavit

I hereby confirm that my thesis entitled “Regulation of transcription by MYC - DNA binding and target genes -” is the result of my own work. I did not receive any help or support from commercial consultants. All sources and / or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor similar form.

Place, Date Signature