PhD degree in Systems Medicine (curriculum in Molecular Oncology)

European School of Molecular Medicine (SEMM),

University of Milan and University of Naples "Federico II"

Settore disciplinare: bio/11

Structure-function analysis of Myc/Max-DNA binding

Paola Pellanda

IIT, Milan

Matricola n. R10314

Supervisor: Dr. Bruno Amati

IEO, Milan

Added Supervisor: Dr. Arianna Sabò

IEO, Milan

Anno accademico 2016-2017

Table of Contents

List of abbreviations

Aa	Amino acid
BAX	BCL2 Associated X Protein
Bcl-2	B cell lymphoma 2
BCL-2A1	Bcl-2-related protein A1
bHLH-LZ	Basic-Helix-Loop-Helix Leucine Zipper
Bim	Bcl-2 interacting mediator of cell death
BrdU	5'-Bromo-2'Deoxyuridine
CDK7	cyclin-dependent kinase 7
CTD	C-terminal domain
DNA	Deoxyribonucleic acid
Doxy	Doxycycline
E-box	Enhancer box
eIF4E	Eukaryotic translation initiation factor 4E
ELISA	Enzyme-linked immunosorbent assay
EMSA	Electrophoretic mobility shift assay
FACS	Fluorescence-activated cell sorting
Fbw7	F-box/WD repeat-containing protein 7
gcPBM	Genomic context protein binding microarray
GSEA	Gene Set Enrichment Analysis
GTF2H1	General Transcription Factor IIH Subunit 1
GTF2H4	General Transcription Factor IIH Subunit 4
HCF-1	Host cell factor 1
HDAC3	Histone deacetylase 3
IRES	Internal ribosomal entry sites
LPS	Lipopolysaccharide
Max	Myc-associated factor X
MbI	Myc box I
MbII	Myc box II
MbIIIa	Myc box IIIa
MbIIIb	Myc box IIIb
MbIV	Myc box IV
Mcm	Mini-chromosome maintenance proteins
Mdm2	Mouse double minute 2 homolog
MDSR	Myc-dependent serum response
Miz1	MYC- interacting zinc finger protein 1
Myc	Myelocytomatosis oncogene
MycER	Myc estrogen receptor

Nmi	N-Myc interactor
OD	Optical Density
OHT	Synthetic steroid 4-hydroxytamoxifen
ORC	Origin Replication Complex
PBM	Protein Binding Microarray
PUMA	p53-upregulated modulator of apoptosis
Pus7	Pseudouridylate Synthase 7
Ras	Rat sarcoma virus oncogene
RBP	RNA-binding protein
Reep6	Receptor Accessory Protein 6
REF	Rat embryonic fibroblasts
Rrp9	Ribosomal RNA Processing 9
SCLC	Small cell lung cancer
SELEX	Systematic Evolution of Ligands by Exponential Enrichment
SHMT	Serine hydroxymethyltransferase
Sin3a	SIN3 Transcription Regulator Family Member A
Sin3b	SIN3 Transcription Regulator Family Member B
Skp2	S-phase kinase-associated protein 2
Smpdl3b	Sphingomyelin Phosphodiesterase Acid Like 3B
SUMO	Small ubiquitin-like modifier
TAD	Transcriptional activation domain
TAF1	TATA-Box Binding Protein Associated Factor 1
TFIIH	Transcription factor II Human
TIAR	T-cell internal antigen-1 (TIA-1)–related protein
TRRAP	Transformation/transcription domain-associated protein
v-*myc*	Viral myelocytomatosis
WDR5	WD repeat-containing protein 5

List of figures

List of tables

Abstract

The c-Myc oncoprotein (or Myc) is a transcription factor of the basic-Helix-Loop-Helix Leucine-zipper (bHLH-LZ) family, whose transcriptional activity depends on dimerization with the bHLH-LZ partner Max and DNA binding, mediated by the basic regions of both proteins. Myc/Max dimers bind preferentially to the hexanucleotide motif CACGTG (known as E-box) and variants thereof. The ability of Myc to bind DNA *in vivo*, however, is not stringently regulated by the presence of the E-box, since many genomic sites targeted by Myc do not contain this motif. Hence, we still need to fully comprehend how Myc recognizes its genomic targets and to what extent sequence-specific DNA binding contributes to this process. Based on the crystal structure of the DNA-bound Myc/Max dimer, we generated a Myc mutant in which two basic region residues engaged in sequence-specific contacts (H359 and E363) were mutated to Alanine (Myc^{HEA}), and compared this with a mutant in which three Arginine residues involved in DNA backbone interactions were mutated to Alanine (Myc^{RA}). While both mutants showed impaired E-box recognition *in vitro*, their over-expression in murine fibroblasts revealed very different genome-interaction profiles, Myc^{RA} showing no detectable DNA binding, and Myc^{HEA} retaining about half of the binding sites seen with Myc^{wt}.

The analysis of the binding intensity of Myc^{wt} and Myc^{HEA} at their binding sites revealed that, as expected, Myc^{wt} bound more strongly the sites containing the E-box, while Myc^{HEA} bound the sites with an E-box as well as the sites without it, confirming that the mutant lost the sequence-specific recognition ability. The interactions retained by the Myc^{HEA} were dramatically reduced with the protein expressed from the endogenous c-*myc* locus, though genome engineering. Thus, unlike Myc^{RA}, the Myc^{HEA} mutant retained non-specific interactions with genomic DNA (detectable at elevated protein levels) but failed to engage more stably through sequence-specific DNA contacts. In spite of this residual DNA-

binding activity, Myc^{HEA} was profoundly impaired in its biological function, undistinguishable from Myc^{RA}: in particular, neither mutant could substitute for wild-type Myc in supporting cell proliferation in murine fibroblasts, whether at normal or supra-physiological levels. While the assessment of transcriptional activities is still ongoing, we conclude that E-box recognition is essential for Myc's biological function.

1. Introduction

1.1 Myc protein

1.1.1 Myc discovery

Myc proteins, encoded by the *myc* proto-oncogenes family, are a group of transcription factors involved in multiple cellular processes, very well conserved across species and often deregulated in tumors. Overexpression of Myc has been estimated to occur in up to 70% of human tumors[1].

Myc deregulation occurs mainly through three events: (1) gene translocation close to transcriptionally active loci[2,3], (2) gene amplification[4,5] or (3) mutations in one of the signaling pathways that lead to increased transcription of the *myc* gene[6].

The first hint of *myc* transforming potential was described in 1911, when Peyton Rous used cell-filtrates from chicken sarcoma to infect susceptible animals[7]. The factor responsible for the infection was isolated only 50 years later: the virus strain MC29, which was able to induce cellular transformation in the hematopoietic compartment, leading either to myelocytomatosis or myelocytomas[8].

In the following years the MC29 virus was demonstrated to transform many different cell types[9–11] and when the genetic element responsible for those disease features was identified, it was named v-*myc* (viral myelocytomatosis)[12,13]. In addition, homologous sequences were found in uninfected vertebrate cells[14,15], confirming the hypothesis of a cellular origin for the viral oncogene[16]. Finally, in 1982 the c-*myc* gene was isolated and characterized in chicken cells[17] and one year later the human gene sequence was identified[18].

1.1.2 Myc protein family

In mammals, the Myc protein family includes three genes: c-*myc*, N-*myc* and L-*myc*. The functions of all the family members are similar, but their expression pattern and oncogenic potential are notably different. Tissue-specific analysis of new born mice reveals c-*myc* expression in all the tissues analyzed, while N-*myc* expression is restricted to brain, kidney, intestine and lungs and L-*myc* is present only in brain, kidney and lungs[19]. Moreover, while c-*myc* is expressed in almost all dividing cells, both N-*myc* and L-*myc* expression are transient and limited mainly to embryo development[19]. In the adult mice N-*myc* is weakly expressed in the heart[19], lungs[19], brain[19,20], some lymphoid organs[19,21] and B-cell precursor[22], while L-*myc* expression is restricted to brain[19,20], lungs[19] and ureter[23,24]. In addition to the tissue specificity, N-*myc* and L-*myc* do not show a homogeneous expression pattern within the same tissue and can be associated to distinct differentiation potential. As an example, it has been reported that during brain development N-*myc* expression is associated with glial commitment, while cells expressing L-*myc* undergo neuronal differentiation[20,24]. Similarly, in the fetal kidney, N-*myc* expression is restricted to the cortical areas[23,25] while L-*myc* protein has been observed the ureter and derived tissues[23,24]. The role of Myc family members during embryonic development has been investigated by generating homozygous mice null for either c-*myc*, N-*myc* or L-*myc*. Both c-*myc* and N-*myc* knock-out mice result in embryonic lethality between day 9.5 and 12.5[26–28]; on the contrary, L-*myc* null mice do not show any congenital defect and the animals' life span is comparable to that of L-*myc* mice[24]. The absence of a phenotype associated with L-*myc* deficiency can be explained by the compensation by c-*myc* and N-*myc* that have been detected in all L-*myc* expressing tissues[24]. A similar compensatory effect has not been reported for c-*myc* and N-*myc* null mice, but substitution of the c-*myc* alleles with the N-*myc* coding region is sufficient to revert almost completely the embryonic lethal phenotype of c-*myc* null mice[29], supporting the idea of a functional redundancy among the Myc family proteins.

Soon after its isolation, c-*myc* was shown to cooperate with the *ras* oncogene to transform rat embryonic fibroblast (REFs)[30]. Later, the same cellular system was used to test the tumorigenic ability of N-*myc* and L-*myc:* while the transforming strength of N-*myc* is comparable to c-*myc*[31], L-*myc* is still able to induce cellular transformation but significantly less than the other family members[32]. *In vivo* though, despite the fact that N-myc is able to induce malignant transformation in REFs as well as c-*myc*, the latter has been reported to have a role in the neoplastic transformation of a wider set of tissues[33,34]. Instead, N-*myc* gene amplification has been reported only in neuroblastomas, where it was originally isolated[35], and more sporadically in other types of neuronal-derived tumors, such as small cell lung cancer, retinoblastoma, glioblastoma and atrocytomas[36]. Coherently with the limited tissue-expression pattern, L-*myc* has been found overexpressed only in small cell lung cancer[37].

1.2   Myc functions

Myc proteins respond to intracellular and extracellular stimuli, such as cytokines, mitogens and growth factors, acting on a variety of cellular processes, either up-regulating or repressing sets of genes. Among the Myc targets, there are genes directly involved in cells cycle regulation and proliferation but also many key components of the metabolic processes necessary to maintain the cellular growth. The main pathways regulated by Myc are schematically summarized in Figure 1[38].
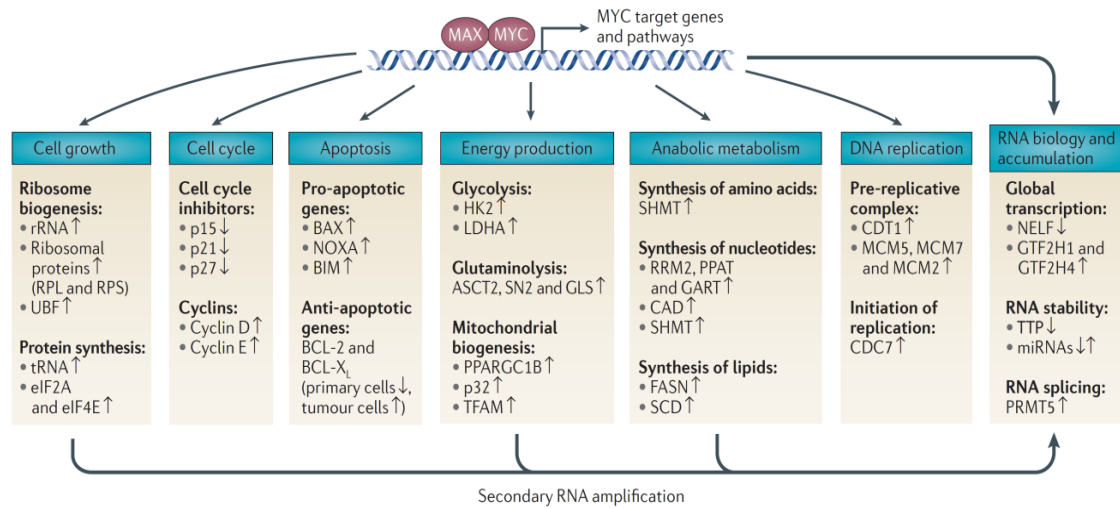
Figure 1. Schematic representation of some of Myc-dependent cellular processes [Modified from [38]].

1.2.1 Proliferation and metabolism

Myc promotes cell division in a dual way: it upregulates genes involved in cell cycle entry, like *cyclin D*[39,40] and *cyclin E*[41,42], and meanwhile it represses cell cycle inhibitors[43–46] and it was shown to bypasses the cell cycle arrest induced by cyclin-dependent kinase inhibitors, such as p27^{KiPl} and p16^{INK4a} [47,48]. Moreover, Myc is known to negatively regulate the genes required for cell growth arrest[49,50] and to accelerate the S-phase[51], as well as to upregulate genes involved in nucleotide synthesis[52,53]. Myc not only pushes the cells into division, but it also plays a role in the biological processes needed to support cell proliferation. Upregulation of genes implicated in different metabolic processes, such as mitochondrial biogenesis[54–56] and glycolysis[57,58], provides the energy required by a cycling cell. A relevant increase in protein content has also been observed in dividing cells and Myc acts on the protein synthesis machinery at multiple levels: it positively regulates the SHMT enzymes involved in the production of the carbon units used for amino acid biosynthesis[59], it activates genes producing rRNAs and tRNAs[60–62], as well as the translation initiation factors eIF4E and eIF2α[63].

1.2.2 Apoptosis

In absence of survival signals, such as growth factors and hormones, high levels of Myc can induce apoptosis in both p53-dependent and p53-independent ways. In the first case, Myc de-regulation determines an increase in ARF protein expression[64] which in turn inhibits the p53 negative regulator Mdm2[65,66]. Activation of p53 increases PUMA and NOXA protein levels that downregulate anti-apoptotic factors, such as Bcl2 and Bcl-X_L[67–70]. p53 also activates the pro-apotic protein Bax, causing the mitochondrial outer membrane permeabilization to induce cell death[71].

Some of these factors are also directly affected by Myc; for example Bax is a transcriptional target of Myc which is upregulated upon Myc overexpression[72], while NOXA promoter is activated by Myc in response to proteasome inhibition[73]. Also Bcl-2 downregulation Myc can occur independently from p53 pathways[68,74] and among the direct Myc targets there is also Bim[75], the major antagonist of Bcl-2.

Myc ability to trigger cell death is believed to provide a safeguard mechanism to prevent uncontrolled cell proliferation as a consequence of Myc deregulation[76,77]. Suppression of Myc-dependent apoptosis is a key feature for tumor onset and requires the loss of a tumor suppressor, such as p53 or ARF, or a second oncogene activation. An example of oncogenic cooperation has been observed between Myc and Bcl-2: Bcl-2 overexpression bypasses Myc-mediated apoptosis but does not affect the proliferative functions of Myc, so that the two oncogene together sustain tumor development[78–80].

1.2.3 Cell adhesion and morphology

The oncogenic activity of Myc is also exerted on cytoskeletal and cell adhesion genes, which play an important role in neoplastic transformation. Myc negatively regulates the expression of many cell surface proteins that interact with the matrix, such as N- and R-cadherins and integrin β1, supporting the anchor-independent growth typical of

transformed cells[81–83]. Moreover, Myc is able to repress many cytoskeletal genes, like *actin*, *cdc42* and *Rho A*[81,83], and determines the morphological alterations typically observed in Myc-overexpressing cells, which acquire a fibroblast-like shape and become more refractile and able to grow at higher density.

1.2.4 DNA and RNA biology

DNA replication is among the cellular process directly regulated by Myc[84]. Different studies described protein-protein associations between Myc and many factors of the pre-replication complex such as the Origin Replication Complex 1 and 2 (ORC1, ORC2)[85,86], the Mcm 2-7 proteins[86,87], cdc6 and cdt1[86]. *Cdt1* gene has also been found as a transcriptional target of Myc[88]. In addition , Myc interacts also with cdc7[89] and cdc45[86], which are essential for the initiation of DNA replication.

Regarding RNA biology, Myc-controls the expression of other transcription factors and co-factors, including the general transcription factors GTF2H1 and GTF2H4[90], AP4[91] and E2F[92]. Moreover, Myc is able to affect mRNA stability by regulating both the expression of proteins involved in mRNA turnover [93] and microRNA expression[94,95].

The role of Myc as transcription factor able to activate and repress specific sets of genes has been recently challenged by a model that described Myc as a transcriptional amplifier that generally upregulates all already active genes. The role of Myc in the cellular transcriptome will be discussed in the section 1.5.3.

1.3 Myc regulation

The threshold between physiological Myc level, fundamental to regulate many different cellular processes, and the pathogenic overexpression that leads to transformation is a critical issue. In order to cope with this, mammalian cells have developed a complex network to strictly regulate Myc expression and activity at any biological step.

- Transcriptional control. The starting point of Myc regulation is the control of its own gene transcription. As already mentioned, *Myc* is very lowly expressed in quiescent cells, but it is an immediate early gene responding to mitogenic signals[96]. Cells controls *Myc* RNA steady-state levels both by reducing the rate of transcriptional initiation and also blocking the nascent mRNA elongation[97–100].

- Post transcriptional control. *Myc* mRNA export to the cytoplasm is mediated by the translation initiation factor eIF4E[101], whose action is controlled by mitogenic stimuli. eIF4E promotes the export of many other mRNAs of genes involved in cell growth recognizing a short sequence in the 5'UTR of the RNA messenger while it is still transcribed, coupling the transcription and export processes. In the cytoplasm, *Myc* transcript half-life is very short, around 10 minutes[102], and it is controlled by a number of miRNAs[103–107] as well as by many RNA binding proteins (RBPs) such as TIAR[108,109], AUF1[109] and HuR[110].

- Translation control. *Myc* mRNA instability is the key mechanism to finely regulate its translation both temporally and quantitatively in physiological conditions. In eukaryotes, the assembly of a ribonucleoprotein complex at the m^7GTP-cap is a fundamental step for initiation of protein synthesis and, according to the classical cap-dependent model, mRNAs with a long and highly structured 5'UTR are impaired in translation initiation[111]. *Myc* 5'UTR is quite long and, in contrast with the majority of the mRNA molecules, is well conserved across species[111]. Several studies reported the effects of Myc 5'UTR on its mRNA translation: both *in vitro* and *in vivo* translation of *c-myc* full length transcript had lower translational efficiency compared to the transcript lacking the exon 1[112,113] and mutation in the 5'UTR region in cell lines derived from multiple myeloma patients was associated with an increase of *Myc* RNA associated with polysomes[114]. The cap-dependent initiation of translation can be bypassed by the presence of a ribosome internal entry site (IRES) in the *Myc* mRNA[115]. IRES mediated cap-independent translation of *Myc* has been shown to be implicated in cancer: increased Myc protein amounts in multiple myeloma

cells were attributed to mutations within the IRES[116,117] as well as the high Myc levels in cell lines derived from Bloom's Syndrome patients, a cancer-prone disorder[118].

Myc protein synthesis can also be blocked as consequence of a stressful stimulus. For example, in response to stress agents which could cause DNA damage and subsequent oncogenic mutations, the TIAR protein have been found associated to the 3' UTR of many key regulators of different cellular processes, including *Myc*, suppressing their translation[108].

-<u>Post translational control</u>. Myc protein undergoes many different post-translational modifications, such as phosphorylation, acetylation, ubiquitinylation and sumoylation, which play a role in Myc stabilization and degradation[119–124]. The protein has a very short half-life, of around 30 minutes[125], and regulation of protein stability critically depends on phosphorylation of two residues in the N-terminal domain: Threonine 58 and Serine 62. Cell growth stimulation leads to Myc stabilization via phosphorylation of Serine 62, which primes the phosphorylation of Threonine 58[119]; this second phosphorylation event, though, triggers the dephosphorylation of the stabilizing phosphate group at Serine 62[126,127]. The phosphorylated T58-Myc protein is recognized by the E3 ubiquitin ligase SCF[Fbw7] and undergoes proteasomal-mediated degradation. Fbw7 is not the only enzyme involved in Myc ubiquitination: Skp2 (S-phase kinase-associated protein 2) has been shown to promote Myc poly-ubiquitination independently from any phosphorylation events[128,129]. In addition, contrarily to Fbw7 and Skp2, which boost Myc degradation, ubiquitination by b-TrCP increases Myc protein stability[120].

Myc can also be acetylated: since both ubiquitination and acetylation occur on Lysine residues, it has been hypothesized that acetylation competes and interferes with the ubiquitination process. Indeed, experimental data showed that acetylation increases Myc protein stability and negatively correlates with ubiquitination[130–132].

Recently, Myc has been also described as a substrate of covalent addition of small ubiquitin-like modifier (SUMO) proteins[122,123,133]. SUMOylation occurs at Lysine residues,

therefore it could also compete with both ubiquitination and acetylation. N-Myc is modified at Lysine 349[122], while mass spectrometry analysis of c-Myc identified a SUMOylation site at Lysine 326[123,133] as well as other nine acceptor Lysines: K52, K148, K157, K317, K323, K389, K392, K398 and K430[123]. Both c-Myc and N-Myc SUMOylation has been reported to play a role in Myc quality control[122,133]; for example multiple SUMO monomers have been found associated to ubiquitin-proteasome pathway[123] and also MYC phosphorylation and dephosphorylation at Serine 62 and Threonine 58 could be a SUMOylation-dependent process[124].

1.4 Myc structure and functional domains

Myc structure resembles that of a typical transcription factor, whose fundamental domains are the transcriptional activation domain (TAD), which lays in Myc N-terminal region, and the DNA binding domain, constituted by the C-terminal portion (Figure 2). The central portion which connected the two terminal domains is instead characterized by many highly conserved motifs.
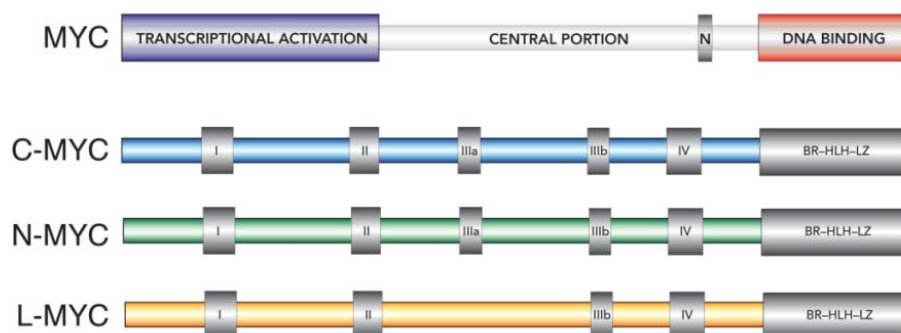
Figure 2. Myc family functional domains[134].

1.4.1 Myc N-terminal region

As mention above, the amino-terminal portion (aa 1-143) of Myc contains the transcriptional activation domain and, when fused to a DNA binding domain, it is

sufficient to trigger gene transcription[135]. The main features of Myc TAD are the two Myc-homology boxes MbI (aa 43-63) and MbII (aa 128-143), which are highly conserved among Myc family members (Figure 2). Myc box I hosts the phosphorylated residues that regulate Myc protein turnover: Serine 62 and Threonine 58 (see above). MbI is also involved in Myc transcriptional activation, as it fundamental for the interaction with p-TEFb[136], the cyclin-CDK complex responsible for RNApolII phosphorylation that stimulates transcription elongation. Myc box II is important for Myc transcriptional (both repression[137] and activation[138]) and transforming[135,139] activities. A key co-factor interacting with MbII is TRRAP (Transformation/transcription domain-associated protein)[140], an adaptor protein found in various complexes containing histone acetyltransferase (HAT) activity; TRRAP is thought to boost Myc-bound gene transcription promoting chromatin opening through histone H4 acetylation [141]. MbII has also a role in Myc degradation, as it is recognized by the E3 ubiquitin-protein ligase complex component Skp2[128].

A schematic representation of some of the known Myc interactors is shown in Figure 3. The interaction with co-factors and other transcription factors[142–144] characterizes Myc N-terminus domain and most likely shapes its transient three-dimensional structure, since in absence of those interactors the domain is highly unstructured and no crystal structures are available.
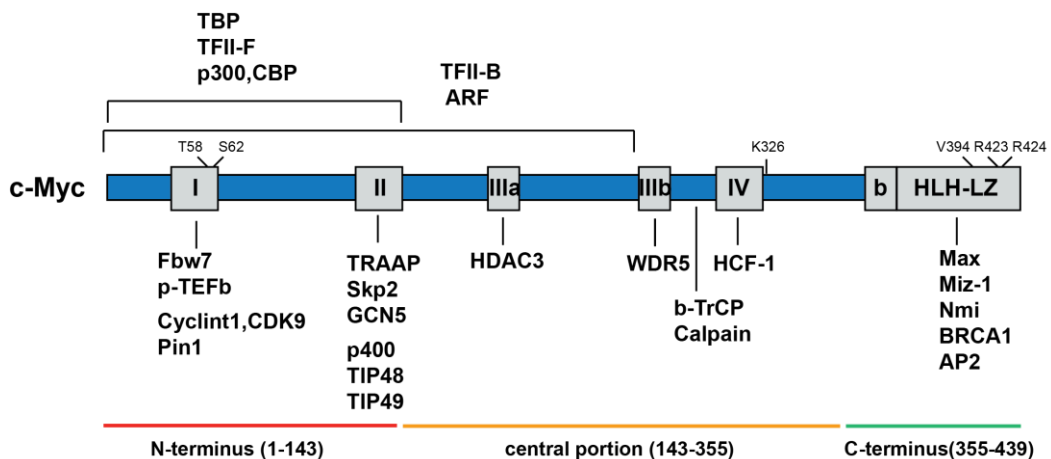
Figure 3. Schematic representation of Myc domains and some of its co-factors.

1.4.2 Myc central region

While both terminal portions of Myc (N- and C-) are well-studied and described, the central portion of the protein is less characterized. The key element of this region is the nuclear localization signal. In c-Myc and N-Myc the signal is composed by two short peptides: the M1 (PAAKRVKLD, aa 320-328) and the M2 (RQRRNELKRSP, aa 364-374); the first one induces complete nuclear localization while the latter determines only a partial nuclear localization[145]. Interestingly, the L-Myc protein lacks the M1 peptide and relies only on the M2 peptide for the translocation into the nucleus[145]. In this region lays also the Lysine 298 residue, which is the substrate of a calpain protease that generates a truncated form of Myc, known as Myc-nick[146].

The other major features of the central portion are Myc homology boxes IIIa, IIIb and IV. Myc box IIIa (aa 180-199) is the only Myc box that is not conserved among all the family members, since it is present in c-Myc and N-Myc but not in L-myc protein (Figure 2). This region is reported to attenuate the pro-apoptotic activity of Myc and therefore it has a role in transformation, both *in vitro* and *in vivo*[147]. It also contains the so called 'D-element', which promotes rapid degradation of ubiquitylated Myc proteins[148] and it is described to mediate gene repression by recruitment of the histone deacetylase HDAC3[147,149]. The homology box IIIb (aa 259-270), despite the fact that it is conserved among all the three protein members, is still poorly understood. Recently a paper showed that Myc box IIIb can directly interact with WD repeat-containing protein 5 (WDR5), which is part of many chromatin remodeling complexes and could facilitate Myc recruitment to target genes[150]. Finally, Myc box IV (aa 304-324) is ambiguously involved in many Myc functions: its deletion impairs Myc-induced apoptosis and partially reduces the transforming potential, but it does not have any effects on cellular proliferation[151]. The SUMO acceptor Lysines lay within this region (N-Myc K323)[122] or immediately outside (c-Myc K326)[123,133]. More recently MbIV has been reported to mediate the interaction with host cell factor 1 (HCF-1),

a cofactor found in many transcriptional and chromatin-modifier complexes, which seems to contribute to Myc tumorigenic ability[152] .

1.4.3 Myc C-terminal region

The carboxy-terminal region of Myc is constituted by its basic helix-loop-helix-leucine zipper motif (bHLH-LZ, aa 355-439), which is common to the bHLH-LZ sub-family of transcription factors. All bHLH proteins bind DNA as obligate dimers[153] and, since Myc homodimers were not detected at physiological concentrations[154–156], great efforts have been spent to identify its dimerization partner. In 1991 Blackwood and Eisenman showed that human c-Myc, as well as N-Myc and L-Myc, interacts with the protein Max[157] and few months later the same interaction was described for the mouse homologous proteins[158]. Max (Myc-associated factor X) belongs to the bHLH-LZ family too and, to date, it is the only known dimerization partner of all the Myc family members. Myc/Max dimerization has been shown to have a fundamental role in DNA binding, transcriptional activation[159–162] and Myc oncogenic activity[163,164].

The bHLH-LZ domain consists of two α-helices connected by a random coil loop (Figure 4). In the crystal structures of Max homodimer and also Myc/Max heterodimer, the first α-helix is constituted by the basic region and the helix H1, which terminates with a Proline (Myc aa 382). Due to its particular structure, Proline cannot be
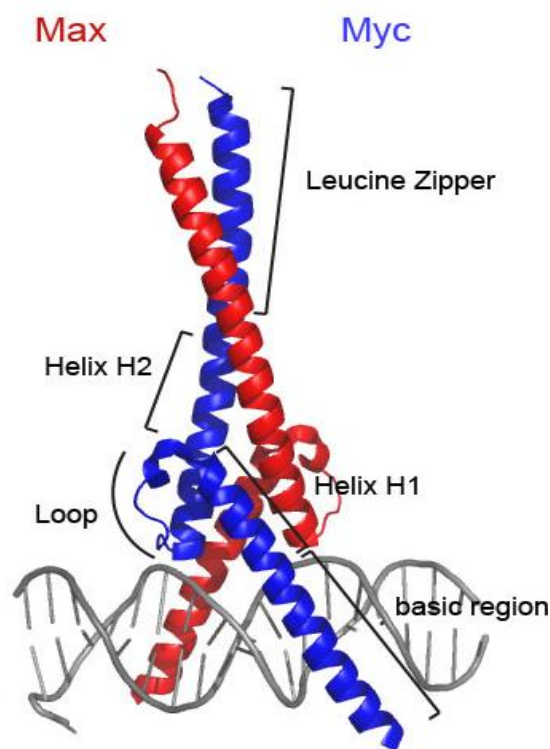
Figure 4. Crystal structure of Myc/Max bHLH-LZ bound to DNA.

fit into a rigid secondary structure and it determines the protein backbone turn that results in the unstructured loop region. The helix H2 and the leucine zipper region compose the second α-helix[165,166]. The HLH-LZ domains mediates the dimerization events[157,159,166,167]. In particular, the leucine zipper region is based on the so called 'heptad repeats motif', a structural motif characterized by the repetition of hydrophobic and polar amino acids and these oppositely charged residues electrostatically interact with the complementary amino acids on the other dimer member[159,168]. Specifically, the crystal structure of Myc-Max dimer shows that for the heterodimer formation the crucial residues of the dimerization interface are Arg423 and Arg424 on Myc and Gln73 and Asn74 on Max [166].

The central core of Myc/Max dimer is shaped into a small four-helix bundle composed by both the HLH domains, the helix H2 and the leucine zipper together form an extended parallel coiled-coil at the C-terminus of both the proteins, while helix H1 and the adjacent basic region of each protein diverge in opposite directions to form a scissor-like structure which perfectly fits into the major groove of DNA helix[166]. This tertiary structure of the DNA binding domain is common to all the bHLH dimers and reveals how the dimerization event is a fundamental pre-requisite for DNA interaction[165,166,169–172].

Even though it is the obligate Myc-binding partner, Max is not the only protein that interacts with the C-terminal domain of Myc. Examples of Myc-CTD interacting proteins found over the years are Miz1, Nmi, BRCA1 and AP-2. Miz1 was first identified as interactor of Myc in a two-hybrid screening[173] and later the residue implicated in this interaction was identified as the Myc Valine 394: in fact mutation of this residue into Aspartic acid (Myc V394D) disrupts the Myc-Miz1 interaction[174].

The repression of several genes by Myc is mediated by Miz1; an example is *p21Cip1* gene: upon UV irradiation, Miz1 promotes transcription of this gene to trigger the DNA damage-induced cell cycle arrest, Myc binding to Miz1 however negatively regulates *p21Cip1*[174]. In contrast Myc V394D mutant fails in the downregulation of *p21Cip1* and it is not able to switch the cell cycle arrest response to apoptosis[174]. Recently, several genomic studies tried

to shed light on Myc and Miz1 interaction and it has been reported that in physiological conditions Miz1 regulates only few target genes which contains the Miz-1 binding motif in their promoters[82,175–178], but when Myc is overexpressed, as in tumor cells, Miz1 can bind also on new sets of promoters[179]. These observations led to a model in which the transcriptional response correlates with the ratio of Myc and Miz1 at promoters: in case of Myc upregulated genes this ratio is bigger than 1, for Myc-repressed genes instead is close to 1[179,180]. However, an integrative analysis of genomic and transcriptomic data from many cellular and *in vivo* systems have recently revealed that the relative Myc abundance at the promoters is an alternative and more accurate predictor of gene transcriptional outcome, while Myc/Miz1 ratio contribution has been shown to be restricted to some cell lines[181].

The functions of Myc interactor Nmi are not completely clear; it has been identified in yeast in a two hybrid screen for possible interactors with the C-terminal domain of N-Myc[182], but it can interact also with c-Myc and other transcription factors[182]. Later, Nmi has been described as an 'adaptor' molecule that recruits Myc to a trimeric complex composed by Myc, BRCA1 and Nmi itself[183]. A physical association between Myc and BRCA1 was described, both *in vitro* assays and in cellular systems[184]. BRCA1-Myc complex can repress gene expression[183,185] and BRCA1 binding to the C-terminal domain of Myc is shown to repress Myc's transforming ability[184].

AP2 is another factor that negatively regulates Myc activity: AP2 binding to the C-terminal region of Myc does not exclude its dimerization with Max, but impairs DNA binding of the complex[186,187].

1.5 Myc-DNA binding

1.5.1 The E-box sequence

All bHLH proteins bind to a general consensus sequence, CANNTG, called E-box (Enhancer-box)[188] and the family members can be classified in two subclasses according to

the E-box variant preference. Class A bHLH proteins, which contains AP4, MyoD and E12, recognize the CA<u>GC</u>TG hexanucleotide motif [188,189]. Myc and Max belong to class B and bind the core variant CA<u>CG</u>TG[190].

The protein-DNA binding is mediated by the basic region, whose amino acids composition determines the sequence specificity of the two subclasses. In fact, while position 11 and 14 are constant (Glu and Arg, respectively, in red in Figure 5) and position 4 and 12 are highly conserved (Lys or Arg, in orange) among both classes, the residues 7 (His), 8 (Asn or Lys) and 15 (Arg) are well conserved in class B only (in yellow).

Figure 5. Class A and class B bHLH protein basic region.

In the crystal structures of Max/Max[165,191] and Myc/Max[166], the Histidine at position 7 (which correspond to residue 359 on Myc and 28 on Max) and the Glutamate at position 11 (363 on Myc and 32 on Max) make contacts with the G_6 of the E-box and the initial $C_{1'}$-$A_{2'}$ motif on the opposite DNA strand, respectively, as shown in Figure 6. The CA<u>CG</u>TG specificity instead depends on the Arginine at position 15 (R367 and R36), which recognizes the G_4 and $G_{4'}$ in the core sequence[165,166,191,192].
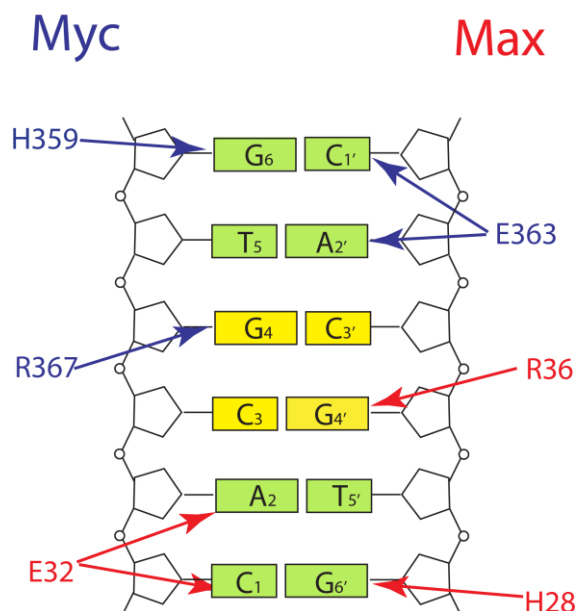
Figure 6. E-box recognition by Myc/Max dimer.

Schematic representation of Myc and Max residues (in blue and red, respectively) which recognize the E-box sequence. The nucleotides of the constant part of the E-box are shown in green, the variable portion instead is shown in yellow.

Several independent *in vitro* studies showed that, beside the high affinity for the canonical CACGTG motif, Myc also binds some core variants: CACGCG, CATGCG, CACGAG, CATGTG[193–195]. These same sequences, named 'non-canonical E-boxes', were later identified also as *in vivo* Myc target sites in ChIP[196] and ChIP-Seq[52] experiments.

In addition to the core variants, also the flanking nucleotides have been shown to play an important role in the protein-DNA binding. Myc/Max binding in yeast was shown to be impaired by a T at position -1 and an A at position +1 of the CACGTG binding sites[197] and similar results were obtained in *in vitro* experiments which led to the identification of an extended consensus binding site of 12 nucleotides, RACCACGTGGTY[198]. Such a strict composition of the ±1 positions was confirmed by ChIP-Seq analysis: the canonical E-box CACGTG strongly prefers C or G at position -1 and, to a lower extent, also A but never T and in the same way A is depleted from position +1[52]. At positions ±2 and ±3 the CACGTG core allows any combinations. The non-canonical sequences instead are more restrictive and allow fewer flanking variants: at position ±1 the only possible bases are C or G, with different preferences among the different core variants; also the combinations

allowed at the other positions are fewer and in some cases the nucleotide at one side influences the one on the other side [52]. In Table 1 a summary of the core E-box motifs with the flanking nucleotide at each position and the consensus sequence determined in a ChIP-seq analysis[52] is shown. More recently, many high resolution *in vitro* approaches such as Selex-seq, PBM (protein binding microarray) and gcPBM (genomic context protein binding microarray), revealed the influence of nucleotide composition of the E-box flanking regions on the tridimensional structure of the DNA binding site[199–202]. In particular, specific symmetries in the DNA sequences surrounding the target binding sites have been reported to significantly affect Myc/Max (and Max/Max) binding specificity[201]. Both the dimers have been tested for their binding strength to DNA sequences containing different kind of symmetries: αNα, αα (where α represents the same base, either A, T, G or C), AT/CG or ATCG. In both cases the PBM analysis revealed a strong preference for the E-box probes flanked by αNα type of symmetric sequences, followed by AT/CG type for Myc and αα type for Max; in general, the recognition of DNA sequence symmetry has been described as an important mechanism by which Max/Max and Myc/Max dimers can increase the strength of the binding to the E- box as well as a new binding mechanism in absence of specific nucleotide recognition[201].

core	CACGTG	CACGCG	CATGCG	CACGAG	CATGTG
Position ±1	**V**CACGTG**B**	**S**CACGCG**G**	**G**CATGCG**Y**	**C**CACGAG**G**	**C**CATGTG**C**
Position ±2	**N**-CACGTG-**N**	**C**-CACGCG-**M** **A**-CACGCG-**C** **T**-CACGCG-**T**	**C**-CATGCG-**A**	**A**-CACGAG-**C**	**W**-CATGTG-**T**
Position ±3	**N**—CACGTG--**N**	**A**--CACGCG--**C**	**G**--CATGCG--**R**	**G**--CACGAG--**C**	**T**--CATGTG--**C** **G**—CATGTG--**T**
Consensus	**NNV**CACGTG**BNN**	**ACS**CACGCG**GMC** **AAS**CACGCG**GCC** **ATS**CACGCG**GTC**	**GCG**CATGCG**YAR**	**GAC**CACGAG**GCC**	**TWC**CATGTG**CTC** **GWC**CATGTG**CTT**

Table 1. Summary of the flanking nucleotides of canonical and non-canonical E-box core variants [modified from[52]].

1.5.2 *In vivo* genome recognition

In eukaryotes, transcription factors binding to their target sequences is restricted by the chromatin context since the DNA wrapped around the nucleosomes forms a strictly packaged chromatin structure that occludes the target sequences. Several studies focused on the link between DNaseI sensitive sites, which define chromatin accessible regions, and transcription factor motif occupancy revealed that the presence of the target sequence in the open chromatin is predictive of the transcription factors binding events[203–206]. Markers of active chromatin include different types of histone post-translational modifications, such as acetylation and H3K4 methylation[207–209] which are recognized by so-called "reader", proteins, such as WDR5 and TAF1, which themselves may play a role in regulating the access of transcription factors to chromatin[209].

In the nucleus transcription does not occur diffusely in a homogenous manner but takes play in the so called "transcription factories". The transcription factories are subnuclear domains composed of active promoters and enhancers together with other regulatory factors and phosphorylated RNA polymerase II[210]. This sub-compartmentalization of transcription was first described in the early 90's[211,212] and the recent development of chromosome conformation capture (3C) method and its variants (4C, 5C and Hi-C) allowed the identification of the genomic loci within the factories, revealing that hundreds of genes, which can be Mb apart, can be associated to the same transcription factory[213,214]. Moreover, biochemical purification of those complexes showed that each factory contains many "core" factors specific for the transcription of that set of genes, but also a number of ribonucleoproteins which are shared with the other factories[215]. Myc/Max binding to DNA is restricted to euchromatin regions[216,217] and in response to stimuli the dimer has been reported to dynamically associate with transcription factories[218].

In physiological conditions, Myc binds preferentially to active promoters; a high percentage of Myc genomic binding sites are located within the CpG islands[219,220] and E-boxes outside an open chromatin context are not bound [221]. Despite the strict sequence

preferences shown *in vitro*, genome-wide analysis revealed that many of these promoter elements do not contain an E-box and the number of the sequence independent-binding events tends to increase when Myc is overexpressed. This phenomenon, named "invasion", occurs both at promoters and enhancers and has been described in many cellular systems and also during tumor progression[38,90,222,223].

Myc invasion can virtually involves all the active regions, probably favored also by the transcriptional machinery, but does not alter the binding hierarchy between high affinity and low affinity sites, with the former bound in the vast majority of cells and the latter only in a small fraction[38,90]. These two types of binding events correspond, respectively, to high and faint signals in ChIP-Seq experiments.

To integrate the traditional view of Myc binding with the emerging genome-wide data, we proposed a model in which Myc is recruited to the chromatin through subsequent steps[224] (Figure 7). The enrollment of Myc/Max dimer to DNA stems from a protein-protein interaction with chromatin-associated proteins without a direct DNA contact (mode 1). After this initial event, the proximity to DNA determines the binding of Myc/Max dimer in a sequence-independent manner (mode 2). This is a low affinity interaction that allows the dimer movement along the DNA until it finds a high affinity site (the E-box) that stabilizes the binding (mode 3).

The fundamental role of protein-protein interactions for Myc target recognition has been recently validated by two independent studies, both focused on H3K4me3-associated proteins. Thomas et al identified a new Myc direct interactor in WDR5, a component of many chromatin remodeling complexes involved in histone methylation[150]. Myc ChIP-Seq signal widely overlaps with WDR5 signal and this co-localization dramatically decreases when the Myc residues involved in the interaction with WDR5 (I262/V264/V265) are mutated. It is noteworthy that these mutations do not impaired WDR5 recruitment to the target sites, but only Myc co-localization implying that is the latter to recruit Myc at the common binding sites [150]. Similarly, Myc has been found to directly interact with the

NURF (ATP-dependent nucleosome-remodelling factor) subunit BPTF and Myc-ChIP peaks distribution and intensity is reduced after BPTF knockdown[225]. These studies confirm the fundamental role of protein-protein interactions (mode 1) in Myc-DNA binding dynamics and the existence of tethering factors which recruit Myc at specific subsets of its own targets, limiting the portion of the genome to be scanned by Myc.

The transition from the sequence-independent binding (mode 2) to the high affinity binding stabilized by sequence recognition (mode 3), instead, is supported by several biophysical studies on bHLH proteins. The conformational changes of human USF and yeast Pho4 in the presence of an E-box or a non-specific DNA sequence have been analyzed using fluorescent spectroscopy[226] and NMR techniques[227], respectively. In both cases the authors described a major rearrangement of the protein structures when any DNA sequence was present, followed by slower changes in the basic regions of the E-box bound samples only, supporting the idea of a two-steps binding reaction. In addition, detailed structural studies have been performed on Max/Max homodimers. The available crystal structures of Max bound to DNA[165,191], show the homodimer perfectly fit into the major groove of the DNA, with the basic region making contacts with specific nucleotides. When Max is not bound to DNA, instead, the basic region is only partially folded but the functional moieties are already primed to interact with the DNA-backbone[228]. These data support the existence of a mechanism that allows Max to discriminate between specific and non-specific DNA sequences: the initial binding event is sequence-independent and led by the natural affinity of the basic region for the DNA backbone. The protein-DNA interaction then triggers conformational changes that allow Max to eventually make contacts with the E-box nucleotides[229]. Considering the high conservation of the basic region among all BHLH proteins this mechanism could be shared by other members of the family.
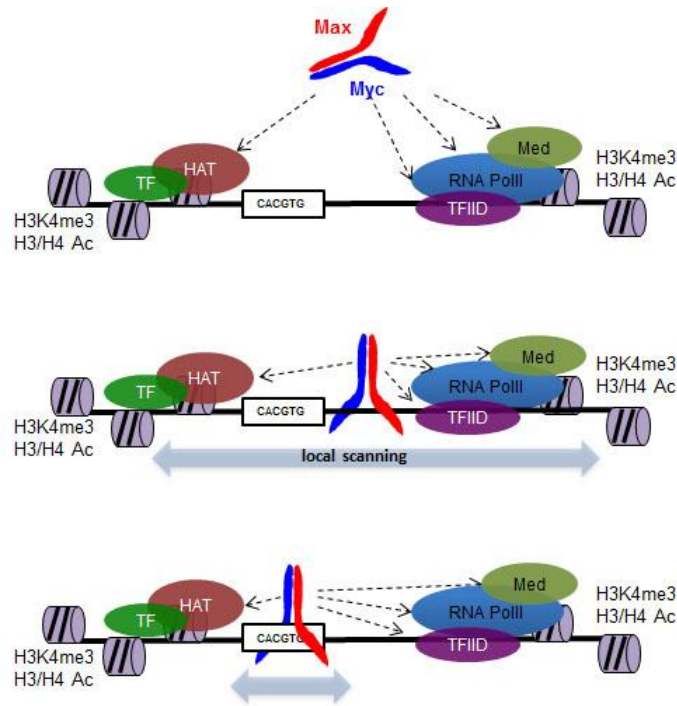
Figure 7. Myc/Max interaction with the DNA [Modified from[224]].

1.5.3 Regulatory models: selective transcription versus general transcriptional amplification

As already described, Myc has a central role in the cell, by actively regulating or repressing many different sets of genes. Traditionally, transcriptional activation by Myc is described as a direct binding event to a target sequence on the promoter DNA, while Myc-dependent repression is thought to occur mainly by indirect DNA binding thought other transcription factors, like Miz1[44,230]. According with this hypothesis, the E-box sequences and its variants have been found significantly underrepresented in the promoters of Myc repressed genes[231,232].

Recent studies pointed out that Myc overexpression is often coupled with an increase in global RNA levels, an event called "RNA amplification". Such phenomenon, together with the spreading of Myc binding to all open chromatin (the so-called "invasion"), has been interpreted as the result of a general transcriptional activation wherever a Myc-DNA interaction occurs[222,223,233]. This model contrasts with the traditional view of Myc as a

transcription factor capable of either activating or repressing genes, but rather describes Myc as a general amplifier which universally upregulates already active genes[222,223]. In this scenario, Myc-dependent gene repression is considered as an artifact due to the normalization procedures used in gene expression analysis: comparing the RNA levels in samples which are characterized by a huge difference in the RNA amount would lead to define as "repressed" genes that are actually up-regulated at a lower level compared to the whole population. According to that model, the few genes which are really downregulated in Myc overexpressing cells can be explained as an indirect consequence: among the genes whose expression is enhanced by Myc there are several transcriptional repressors which in turn can lower the expression of their targets[222,223].

Unfortunately, the transcriptional amplification model failed to take into account a fundamental aspect: Myc activity triggers a series of cellular processes (some of which described in section 1.2) that dramatically impacts on cell physiology. The RNA amplification phenomenon can indeed be explained by the many metabolic changes that rely on Myc and, in turn, impact on global RNA synthesis and turnover[234–237]. As an example, the difference in the total RNA content has been a hallmark to discriminate cycling from quiescent cells for years[238], and the RNA amplification phenomenon is more evident when different Myc levels co-occur with different physiological states, such as normal versus tumor tissues[90,223] or quiescent B-cells versus LPS-activated cells[90,222]. In addition, Myc invasion and RNA amplification are separable events which can occur independently: in serum-stimulated fibroblasts Myc does not completely invade all the open chromatin but still there is an increase in the RNA level when the cells transit from G0/G1 into S-phase[90] . On the other hand, when Myc is overexpressed in already proliferating fibroblasts it invades all the active chromatin without triggering RNA amplification[38,90].

Finally, the idea of Myc as a transcriptional amplifier postulates a cause-to-effect relation between Myc binding and gene regulation, but discriminating Myc productive and non-

productive binding is a challenging issue. In fact, careful analysis of available datasets pointed out that DNA binding is not predictive of actual regulation of gene transcription[38]. Altogether these observations strongly support the long-standing concept that Myc acts as a "traditional" transcription factor, which is able to up- and down-regulate specific sets of genes in response to the environmental stimuli.

1.6 Targeting Myc in cancer

Many studies showed that inhibition of Myc activity triggers tumor regression, according to the idea that cancer cells depend on the oncogene for their sustenance[239–242]. This, as well as the widespread alterations of Myc activity in different tumors, makes targeting Myc one of the most appealing approaches to treat human cancer. However, Myc is not an easily druggable protein since it lacks an enzymatic activity or pre-folded active site, and exerts its function in the nucleus. Over the years, numerous strategies have been attempted but with various degrees of success.

Since the essential pre-requisite for Myc activity is DNA binding and this requires dimerization with Max, an obvious approach to target Myc functions would be the disruption of Myc/Max dimerization. Several *in vitro* and cell-based screens have been performed to develop inhibitors of Myc/Max dimerization[243,244]. However, the size and the structure of the Myc/Max interface make it difficult to design small molecules inhibitors and moreover the dimerization domain lays in the leucine zipper structure, which is well conserved among the bHLH-LZ protein subfamily and could lead to off-target effects.

A second strategy to affect Myc/Max interaction is the Omomyc protein. Omomyc consists of the Myc bHLH-LZ domain in which four amino acids in the leucine zipper domain, E410, E417, R423, R424, have been mutated into T, I, Q, and N, respectively. These substitutions confer the mutant the ability to form homodimers and to dimerize with wild-type Myc[245]. Several *in vivo* studies have shown that Omomyc has a strong anti-tumoral

activity, acting directly in tumor cells to reduce proliferation, increase apoptosis, and interfere with the maintenance of the tumor microenviroment[246–248]. Whether this dominant negative phenotype is due to sequestration of Myc and/or Max in inactive heterodimers[245], or to the occupancy of the Myc/Max target sites by Omomyc homodimers[249], is not fully understood. Whether Omomyc may be efficiently delivered to tumor cells also remains to be addressed. This notwithstanding, Omomyc emerges as a promising candidate for therapy, as it does not interfere with the physiology of normal tissues[246,250]. The inhibition of Myc activity by Omomyc highlights the importance of understanding the structural and functional features of the Myc bHLH-LZ domain and its interaction with the DNA, in particular toward the design of new therapeutic strategies.

1.7   Aim of the project

The dynamics of Myc-DNA interactions and the importance of the E-box recognition for Myc functions are still not completely understood. We have proposed a model in which the first step of Myc-DNA association occurs through the interaction with chromatin-bound proteins and only later Myc engages the DNA. This initial DNA binding event does not require specific sequence recognition but is driven by the general affinity of Myc basic region for the DNA. The sequence-independent binding allows Myc/Max dimer to "scan" the DNA sequence locally, to eventually recognize the E-box. In this study, I addressed this model by generating Myc mutants impaired in their DNA binding capacity. I will present experiments investigating the role of basic region-mediated DNA contacts on a genome-wide level, and evaluating the contribution of E-box recognition to Myc-genome interactions and biology.

2. Materials and methods

2.1 Cell culture

All the cell lines used in this work were cultured in DMEM, supplemented with 10% fetal bovine serum, 2 mM L-glutamine and penicillin/streptomycin. The cb9 MycΔb cell line and the cb9 Myc clones cultured medium was supplemented, if not indicated differently, with 1 µg/ml of doxycycline to activate the tet-myc transgene.

Rat HO15.19 cells and mouse cb9 MycΔb cells were infected with the retroviral vectors pBabe hygro or pQCXIH, respectively, and then selected with 150 µg/ml of hygromycin for 4 days. Mouse 3T9 fibroblasts instead were infected with the retroviral vector pBabe puro and selected with 1.5 ug/ml of puromycin for 2 days; the activation of the MycER fusion protein was achieved added 400 nM of OHT to the culture medium.

293T cells were transfected overnight with 5 µg of the plasmids of interest, mixed in a solution of 240 mM $CaCl_2$ and HBS (25 mM HEPES, pH 7.0, 5 mM KCl, 6mM dextrose, 140 mM NaCl, 0.750 mM NA_2PO_4). The next day the medium was replaced and the cells were collected 48h after transfection.

2.2 Pymol

The three-dimensional visualization and conformational analysis of Myc/Max-DNA structure (PMB 1NKP) was performed with the open-source software Pymol (https://www.pymol.org/).

2.3 Myc-Max co-Immunoprecipitation

293T cells were transfected with plasmids encoding for FLAG Myc^{wt}, FLAG Myc^{HEA}, FLAG Myc^{RA} or FLAG EV (empty vector) and collected 48h after transfection.

After two washes in cold PBS, cells were scraped in 4 ml of cold NHEN buffer (20 mM Hepes pH 7.5, 150 nM NaCl, 0.5% NP-40, 10% glycerol, 1 mM EDTA, protease inhibitors cocktail) and lysed for 20 minutes on a rotating wheel at 4°C.

Complete cell disruption and DNA fragmentation was performed with three cycles of sonication (30 seconds on, 30 seconds off) and the protein concentration was determined by Bradford-based Protein Assay kit (Bio-Rad).

The immunoprecipitation of FLAG Myc was performed incubating 2 mg of cell lysate with 40 µl of Anti-FLAG M2 affinity gel (Sigma-Aldrich) for 3h in agitation at 4°C. In parallel 2.5% of the material used for the IP was collected to be loaded as input.

The beads were then washed five times with 1 ml of wash buffer (20 mM Hepes pH 7.5, 150 nM NaCl, 0.1% Tween, 10% glycerol, 1 mM EDTA), resuspended in 60 µl of loading buffer and boiled at 95°C for 10 minutes. The supernatant was then loaded on a SDS-PAGE gel for immunoblotting analysis.

2.4 Western Blot

Protein extraction was performed resuspending the cells in lysis buffer (300 mM NaCl, 1% NP-40, 50 mM Tris-HCl pH 8.0, 1 mM EDTA, proteases inhibitors) and sonicating them.

Cell extracts were quantified with the Bradford-based Protein Assay kit and separated by SDS-PAGE using 7.5 % polyacrylamide gels.

The proteins were then transfered to a nitrocellulose membranes for 1 h at 0.3 A with a wet transfer apparatus. Membranes were washed in TBS-T (10 mM TrisHCl, 100 mM NaCl, 0.1% Tween at pH 7.4) and blocked with 5% milk in TBS-T for 20 minutes, immunoblotted over-night at 4°C with the indicated primary antibodies, washed three times for 10 minutes with TBS-T and then incubated at room temperature for 1 h with the secondary antibodies. After three washes in TBS-T, chemiluminescent imaging was performed by ChemiDoc MP System (BioRad) using Western ECL reagent (BioRad).

2.5 Antibodies

antibody	company	host
Myc Y69	Abcam (ab32072)	rabbit
Max	Santa Cruz (sc-197)	mouse
Vinculin	Sigma-aldrich (V9264)	mouse
FLAG	Abcam (ab1162)	rabbit
BrdU	Becton Dickinson (347580)	mouse
Myc N262	Santa Cruz (sc-764)	rabbit

Table 2. Primary antibodies.

2.6 Transcriptional Factor Assay Kits: TransAM^{TM} c-Myc

The analysis of Myc^{wt}, Myc^{HEA} and Myc^{RA} binding affinity to the E-box sequence (CACGTG) was performed with the commercially available DNA-binding ELISA TransAM^{TM} c-Myc kit (Active Motif, 43396), following the manufacturer's instructions. We used 10 µg of nuclear extract of rat HO15.19 expressing Myc^{wt}, Myc^{HEA} or Myc^{RA} and the same amount of extract of HO15.19 EV as negative control. 2.5 µg of the provided nuclear extract from Jurkat cells was used as positive control.

2.7 Cycloheximide treatment

Cycloheximide (Sigma-aldrich) was added to the culture medium of rat HO15.19 fibroblasts at a final concentration of 50 µg/ml and incubated at 37°C. At different time-points after cycloheximide administration, cells were washed twice with cold PBS and collected for protein immunoblotting analysis.

2.8 Myc immunofluorescence

HO15.19 and cb9 MycΔb cells were plated on cover slips. The day after the cells were washed twice with PBS, fixed with 4% paraformaldehyde for 10 minutes and then washed twice with PBS.

The cell membranes were permeabilized incubating the cells for 10 minutes in a solution 0.1% Triton in PBS. After two washes in PBS, cells were incubated with the blocking solution (4% BSA and 1% Fish gelatin) for 30 minutes, to prevent unspecific binding of the antibodies. Cells were then incubated with the primary antibody against Myc (abcam Y69, ab37072) diluted to the final concentration of 2ng/µl in the blocking solution for 1h and 30 minutes and then washed twice in PBS and once again with the blocking reagent for 10 minutes. After 45 minutes of incubation with the secondary antibody (anti rabbit Cy3, 1.25 ng/µl) the cells were washed in PBS and the nuclei were stained with DAPI for 2 minutes. After one last wash with water the cover slips were mounted with Moviol on glass slides and ready for microscopic analysis.

2.9 Proliferation assays

For growth curve experiments 70,000 Rat HO15.19 cells were plated in 6-well plates in triplicates and counted every 3 days for 9 days. Similarly, 70,000 3T9 cells were plated in presence or absence of 400 nM OHT, counted every 2 days up to day 6. In the experiment performed with the cb9 MycΔb cells instead 80,000 cells per well were plated in presence of doxycycline for 2 days, then counted and re-plated, with and without doxycycline, every 2 days for 10 days.

For the colony forming assay (CFA) 10,000 cells were plated in 10 cm dishes, let them grow for 6-11 days and then stained with crystal violet (incubation with crystal violet for 10 minutes and then washes with water).

For the S-phase analysis we performed a 5'-Bromo-2'Deoxyuridine (BrdU) staining; the cb9 MycΔb cells expressing Myc^{wt}, Myc^{HEA} or Myc^{RA} (growing in presence or in absence of doxycycline) were incubated for 20 minutes with 30 mmol/L of BrdU (B9285, Sigma); the cells were then washed twice with PBS, collected and fixed with 2 ml of cold ethanol. After one wash in a solution PBS 1%BSA, the cells were incubated at room temperature for 20 minutes in 1 ml of HCl 2N, to denature the DNA; the samples were neutralized with 3 ml of 0.1 M Na$_2$B$_4$O$_7$, pH 8.5.The cells were then washed twice with PBS 1%BSA and then stained with antibody targeting BrdU (to a final concentration of 0.4 µg/ml) for 1 h light protected. The cells were then washed in PBS 1%BSA and stained with the secondary FITC-conjugated donkey-anti- mouse antibody (final concentration 30 µg/ml) for 1 h light protected. After one last wash in PBS 1%BSA the cells were resuspended in 500 µl of PI/RNase solution (2.5 µg/ml of PI and 250 µg/Ml RNaseA in PBS) and stained overnight before the acquisition with MACSQuant® Analyzer.

2.10 Genome editing: CRISPR/Cas9

The deletion of the c-*myc* basic region as well as the insertion of the HEA mutation in endogenous *c-myc* loci were performed exploiting the type II CRISPR-Cas tool[251].

This system is composed by an endonuclease, the Cas9, and a small RNA molecule called sgRNA (single guide RNA) that leads the Cas9 enzyme to complementary genomic region. The sgRNA is approximately 20 nt long and contains the protospacer adjacent motif (PAM) NGG, when the RNA-DNA hybrid is formed the endonuclease cuts on both DNA filaments close to the PAM site (Figure 8). The non-homologous ends joining machinery will then repair the double strand break causing insertion or deletion of nucleotides.
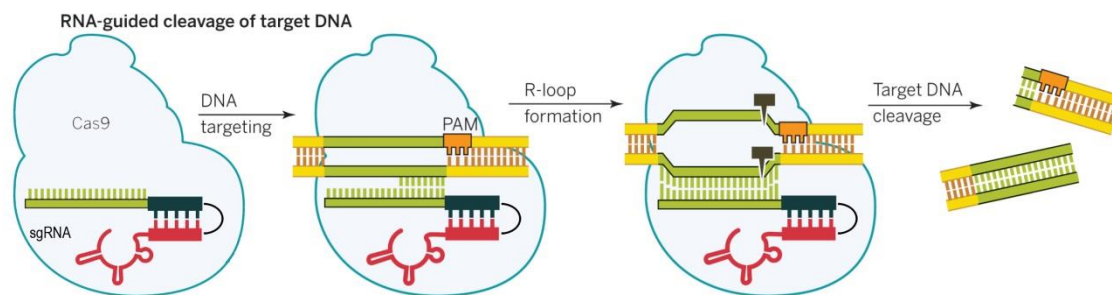
Figure 8. Type II Cas9 [modified from[252]].

The single guide RNA (sgRNA) sequences to target the *c-myc* gene were designed using the online available software CRISPR Design Tool (http://crispr.mit.edu/). We tested 10 different sgRNA sequences cloning them into the pSpCas9 (BB)-2A-GFP (PX458) plasmid, which also encodes for the Cas9 protein and GFP. To determine which sgRNA was more efficient, we transfected 40,000 cb9 tet-Myc cells in 12-well plates using 100 μl of the transfection mixture (0.5 μg plasmid, 2.5 μl Transit 2X (Mirus) in Opti-MEM) in 900 μl of growth medium without antibiotics. After 48h we performed a PCR on genomic DNA (for the primer sequences see Table 3) to amplify the target region and to test for the cutting efficiency in the surveyor nuclease assay[251]. The assay takes advantage of the mismatches introduced by the DNA repair mechanism after the Cas9-mediated cut: the PCR products from transfected cells and untransfected control are mixed and then undergo to a cycle of denaturation and re-annealing, if some heteroduplex DNA are formed the T7 endonuclease recognizes it and cuts the DNA, thereby generating the characteristic digestion pattern (Figure 9).
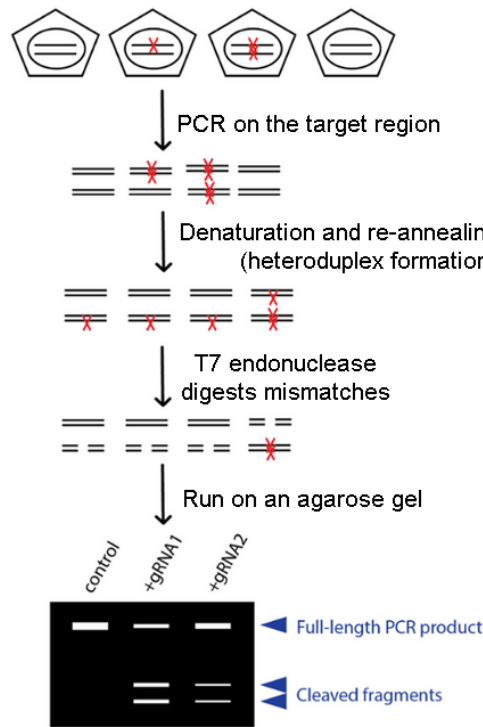
Figure 9. Scheme of the surveyor assay [modified from[253]].

2.10.1 cb9 MycΔb fibroblasts

The cb9 MycΔb cellular clone has been derived from the cb9 tet-Myc cell line, which was produced through the 3T3-immortalization protocol starting from mouse embryonic fibroblasts (E14.5) obtained from Rosa26-rtTA/Tet-Myc mice. We exploited the CRISPR-Cas tool to delete the endogenous c-*myc* basic region in the cb9 tet-Myc cell line.

As already describe, we tested 10 sgRNAs and chose the ones with the highest efficiency, sgRNA8 and sgRNA7 (Table 3), and combine them in a single transfection (0.5 μg sgRNA 8 + 0.5 μg sgRNA7). The scheme of the strategy used to disrupt the basic region and the resulting deleted alleles of the clone we isolated are shown in Figure 10. The cellular clone, named cb9 MycΔb, had both c-*myc* alleles mutated, although in different ways: one allele encoded for a protein missing the basic region, the Helix I and the loop but retained the second Helix and the leucine zipper, while the product of the other allele was a truncated protein which almost completely lacked the C-terminal domain.

A

target of sgRNA8 PAM

Mouse c-myc locus 5'...AGACACGGAGGAA**AACGACAAGAGGCGGACACACAACGTCTTGGAACGTCAGAGGAGG**.........................CCCAAGGTAGTGATCCTCAAA...3'
|||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
3'...TCTGTGCCTCCTT**TTGCTGTTCTCCGCCTGTGTGTTGCAGAACCTTGCAGTCTCCTCC**.........................GGGTTCCATCACTAGGAGTTT...5'

PAM target of sgRNA7

Mouse Myc protein sequence   N D K R R T H N V L E R Q R R . . . . P K V V I L K

Basic Region   **Helix I - Loop**   **Helix II**

B

Mouse wt c-myc locus   5'...AGACACGGAGGAA**AACGACAAGAGGCGGACACACAACGTCTTGGAACGTCAGAGGAGG**.........................CCCAAGGTAGTGATCCTCAAA...3'
Cb9 MycΔb c-myc allele 1 5'...AGACACGGAGGAA**AACGAC**--GTAGTGATCCTCAAA...3'
Cb9 MycΔb c-myc allele 2 5'...AGACACGGAGGAA**AACGACA**--GGTAGTGATCCTCAAA...3'

C

                           Basic Region   **Helix I - Loop**   **Helix II**

Mouse Myc protein sequence   N D K R R T H N V L E R Q R R . . . . P K V V I L K K
Cb9 MycΔb Myc protein_allele 1 N D - V V I L K K
Cb9 MycΔb Myc protein_allele 2 N D R *

Figure 10. Deletion of the basic region of the endogenous c-*myc*.

The cb9 tet-Myc mouse fibroblasts underwent to CRISPR/Cas9 genome editing to delete the basic region DNA sequence of the endogenous c-*myc* loci. (A) Endogenous Myc DNA and protein sequences are shown. In dark green the sequence of the basic region and in black the Helix II. The light green bars represent the sgRNA sequences we have used, with the PAM sequence underlined in red. (B) Alignment of Myc^wt DNA consensus sequence and the sequences of two deleted alleles. (C) Alignment of the protein sequence of Myc^wt and the two deleted forms. The protein derived from the allele 1 almost completely lacked the basic region, the Helix I and the loop region; the protein codified by the allele 2 was a truncated form that lacks the C-terminus domain.

2.10.2 cb9 Myc^HEA clones

The CRISPR/Cas tool was exploited also to insert the HEA mutation into the endogenous *c-myc* alleles of the cb9 tet-Myc cell line. To increase the replacement efficiency, we used a small donor oligo[251,254] (192 nt, the complete sequence can be found in Table 3) whose characteristics are shown in Figure 11. We then followed the same protocol used to produce the cb9 MycΔb, transfecting the donor oligo and the SpCas9 (BB)-2A-GFP plasmid containing the sgRNA8.
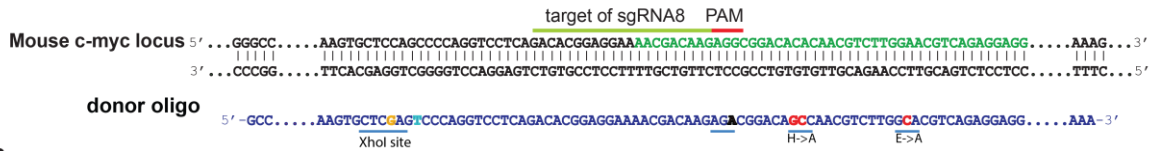
A

B

Figure 11. Generation of cell lines with the endogenous c-*myc* loci mutated into Myc^{HEA}.
The cb9 tet-Myc mouse fibroblasts underwent to CRISPR/Cas9 mediated genome editing to replace the endogenous c-*myc* sequence codifying for the basic region. (A) the sgRNA and donor oligo used to replace the endogenous sequence, which was characterized by the Myc^{HEA} encoding mutation (in red), by three silent mutations to: disrupt the PAM site (in black), to create a XhoI restriction site (in orange, the complete restriction site sequence is underlined) and to disrupt the palindromic sequence flanking the XhoI site (in light blue). (B) The amino acid sequence derived from the substitution of the endogenous c-*myc* alleles, in red the two mutated amino acids.

2.11 RNA extraction and qPCR analysis

Total RNA was purified using Quick-RNA^{TM} Mini prep (Zymo Research) and treated on-column with DNaseI. Complementary DNA (cDNA) was synthetized using the reverse transcriptase ImPromII (Promega) and 10 ng of cDNA were used for quantitative PCR reaction with FAST SYBR Green Master Mix (Applied Biosystems) in a final volume of 20 µl. The primer sequences used are listed in Table 3.

2.12 Chromatin immunoprecipitation (ChIP)

Cells were fixed with formaldehyde 1% in PBS for 10 minutes after which the reaction was stopped by addition of glycine at a final concentration of 0.125 M for 5 minutes. The cells were then rinsed twice with cold PBS and collected in SDS buffer (100 mM NaCl, 50 mM Tris-HCl pH 8.0, 5 mM EDTA, 0.5% SDS, protease inhibitors). After centrifugation the cell pellet was resuspended in 4 ml of cold IP-buffer (100 mM Tris at pH 8.6, 0.3%

SDS, 1.7% Triton X-100, and 5 mM EDTA) and the chromatin was sonicated to an average length of 300-500 bp and used for the immunoprecipitation.

The samples were first precleared for 1 h at 4°C in agitation with 50 μl of protein A sepharose beads (blocked with 0.5 mg/ml *E. coli* tRNA and 0.5 mg/ml BSA). An aliquot of each sample was kept as input; the remaining material was incubated overnight in agitation at 4°C with 10 μg of Myc antibody (N262, Santa Cruz).

The day after, 60 μl of blocked protein A beads were added to each sample and incubated for 2 h in agitation at 4°C. Following that, the beads were washed three times with 1 ml of mixed micelle buffer (150 mM NaCl, 20 mM Tris-HCl pH 8.1, 25 mM EDTA, 0.1% NaN3, 5% Triton-X 100, 1% SDS, 26% sucrose), twice with buffer 500 (0.1% DOC, 1mM EDTA, 50 nM HEPES, 500 mM NaCl, 1% Triton-X 100, 0.2% NaN3), twice with LiCl-detergent buffer (0.5% DOC, 1 mM EDTA, 250 mM LiCl, 0.5% NP-40, 10 mM Tris-HCl pH 8, 0.2% NaN3) and once with TE. To elute the protein-DNA complex and reverse the crosslink, the beads and the input were resuspended in 200 μl of 2% SDS in TE and incubated overnight at 65°C. DNA was then purified by Qiagen columns and quantified using Qubit^{TM} dsDNA Assay kit (Invitrogen). The ChIP quality was checked by qPCR amplification using primer pairs designed on Myc target genes and a non-targeted region as negative control. The reaction was performed using 600 nM primers (listed in Table 3) in a final volume of 20 μl of FAST SYBR Green Master Mix (Applied Biosystems).

1.5-2 ng of ChIP DNA was then used to generate the chromatin immunoprecipitation sequencing (ChIP-Seq) libraries according to the Illumina protocol and then sequenced with HiSeq2000.

2.13 Primers and oligos sequences

	species	Amplicon	Forward sequence	Reverse sequence
Expression	human	c-myc	GATTCTCTGCTCTCCTCGACGG	AGAAGGTGATCCAGACTCTGACC
	mouse	Rplp0	GGCGACCTGGAAGTCCAACT	CCATCAGCACCACAGCCTTC
		c-myc	TTTTTGTCTATTTGGGGACAGTG	CATCGTCGTGGCTGTCTG
		Smpdl3b	GGATGGGGAGATGGTGTATG	GAAGCTGTCGGTATGGTGGT
		Reep6	GTGCAATGTCATCGGATTTG	TTGCCCGCGTAGTAGAAAG
		Pus7	CCCCAAGCATAAAATCAGTGAGG	CCCCGATAAGGAGTAATCTCGAA
		Rrp9	AGAGACCGCACAGGAAAAGA	ACTTCTGCAACCTGCCTCTC
ChIP	mouse	Ncl	GGCGTGGTGACTCCACGT	CGAAATCACCTCTTAAAGCAGCA
		CAD	CGAAGGAGCCCACGTGTGTG	GAACTCAGTAGTGCGCCGC
		D7Wsu128e	GCGCCGCCATGTGGACTAG	CGAAGGAGCCCACGTGTGTG
		Pus7	GCTGCACCGCGTGGAGAC	GGCTGGTGGGATAACCCGT
		AchR	TGCTCATCTCCATCAAGGTCAA	AGGCTCAGCAGGAAGTAGTTGTTG
		C/EMPα	CGCTCTCCTTAGGGTCCTTT	TCTTTTTCATTGCGTCTCCA
DNA_ Surveyor assay	mouse	c-myc basic region	GGTGTCTGTGGAGAAGAGG	AGCGCATCAGTTCTGTCAG

	species	Amplicon	Sequence	
sgRNA	mouse	sgRNA7	ACTCCTAGTGATGGAACCC	
		sgRNA8	ACACGGAGGAAAACGACAAGAGG	
Myc^{HEA} Donor DNA oligo	mouse		GCCAAGTTGGACAGTGGCAGGGTCCTGAAGCAGATCAGCAACAACCGCAAGTGCTCGAGTCCCAGGTCCTCAGACACGGAGGAAAACGACAAGAGACGGACAGCCAACGTCTTGGCACGTCAGAGGAGGAACGAGCTGAAGCGCAGCTTTTTTGCCCTGCGTGACCAGATCCCTGAATTGGAAAACAACGAA	

Table 3. Summary of primers and oligos.

2.14 Computational analysis

2.14.1 Next generation sequencing data filtering and quality evaluation

ChIP-Seq reads were filtered with the fastq_quality_trimmer and fastq_masker tools of the FASTX-Toolkit suite (http://hannonlab.cshl.edu/fastx_toolkit/) and their quality was assessed with the FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/) application. The reads were analyzed with our own pipeline HTS flow[255]. The HTS flow pipeline allows primary analysis, consisting of the quality control of the raw reads and the alignment to a reference genome, and secondary analysis, consisting in the peaks calling.

2.14.2 ChIP-Seq data analysis

The HTS-flow pipeline aligned the ChIP-Seq reads to the mouse reference genome (mm9) through the BWA aligner using default settings[255]. Later, the MACS software[256] was used to call the peaks, setting as cut-off parameter a p-value<1e-5. The reads count within a genomic region was normalized considering the total number of aligned reads in that sample (library size). With the MACS software we also performed the saturation analysis, as control of false negatives, and also estimated the false discovery rate, defined as the proportion of negative (the peaks identified calling on the input using the ChIP as reference) VS positive peaks.

The enrichment of a peak was defined considering the library size-normalized reads of the ChIP falling in the peak region (ChIPw) minus the library size-normalized reads of the input in the same region (inputw) as a logarithmic value, $\log_2$ (ChIPw - inputw).

Peaks were mapped and annotated as promoter, enhancer, intragenic or distal according to the genomic position of the peak midpoint. More in details:

- promoter: the peak is within the genomic region defines by -2Kb and +1Kb from an annotated refgene TSS

- enhancer: all regions, different from -2Kb ad +1 Kb from the TSS, marked with H3K4me1 in 3T9 fibroblasts[90]

- intragenic: the peak is located inside an annotated refgene (in a region different from -2Kb ad +1 Kb from the TSS)

- distal: the peak position does not match any of the former criteria

Myc ChIP peaks were visualized with the UCSC Genome Browser (http://genome.ucsc.edu/.)

Qualitative and quantitative heatmaps of ChIP-seq enrichment were generated using compEpiTools packages, a tool for computational epigenomics[257].

The determination of the 'genomic background' was performed searching for the sequences of interesting with the Biostrings computational packages and the difference between the E-box distribution in Myc^{wt} and Myc^{HEA} samples was tested with the χ^2 test.

Functional annotation analysis to determine the gene set categories was performed using Molecular signature database (MsigDB) of GSEA Broad Institute (http://software.broadinstitute.org/gsea/msigdb/annotate.jsp).

3. Results

3.1 Design of Myc mutants compromised in DNA binding

bHLH proteins binding to DNA is characterized by two types of interactions: the sequence-specific contacts with the E-box nucleotides and the generic affinity for the DNA backbone[165,166,170–172,191,258]. We took advantage of the published Myc/Max-DNA structure (PDB ID:1NKP)[166] to analyze in detail the structural basis for Myc/Max-DNA recognition. In the crystal, Myc residues H359 and E363 recognized the constant part of the consensus E-box (CANNTG, Figure 12, in green), while the R367 contacted with the central G_4 of the CpG core (CACGTG, Figure 12, in yellow), as originally described[166].
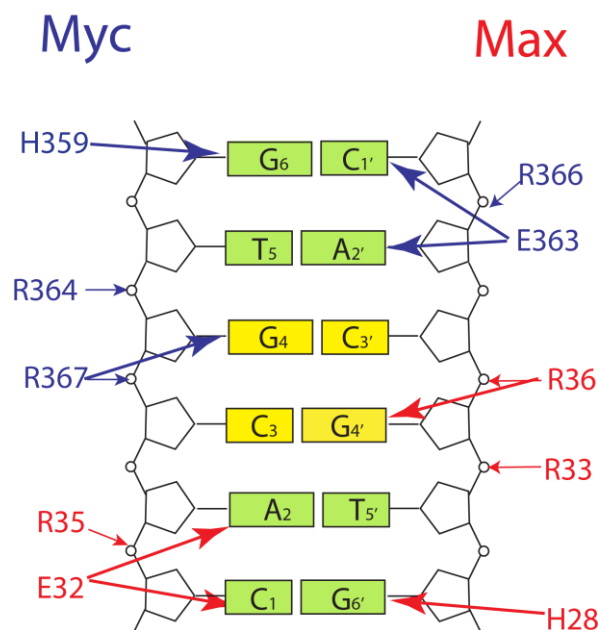

Figure 12. Schematic representation of Myc/Max dimer residues which interact with DNA.
The amino acid residues of Myc (blue) and Max (red) that interact with the DNA backbone and/or the E-box nucleotides are indicated by arrows. The constant part of the E-box is showed in green, in yellow the variable core.

The R367 side chain was also involved in an H-bond with the phosphate group of the DNA backbone. Two other Arginine residues, R364 and R366, mediated similar sequence-independent interactions. The same types of protein-DNA interactions were also observed

on the other half of the binding site: Max residues H28 and E32 recognized the E-box (residues C_1, A_2 and $G_{6'}$, respectively), while R33, R35 and R36 interacted with the backbone.

The residues involved in the backbone recognition are strongly conserved among the bHLH protein family (Figure 13), with the exception of R15 (Myc R367) which is conserved only in the class B bHLH (to which Myc and Max belong) consistently to its role in recognizing the variable core of the E-box. Among the amino acids which contact the constant part of the consensus E-box only the Glutamic acid (E11) has been found conserved among the entire bHLH protein family, while the Histidine (H7) conservation is restricted to the class B. In class A proteins the Histidine can be substituted either by another polar amino acids, like Asparagine, which contacts the E-box nucleotide in a similar way[258] or by the hydrophobic Alanine. In the latter case it has been observed that the Alanine residue is not involved in the DNA interaction, which instead is mediated by the polar amino acid in position 8, which is able to contact the T_5 nucleotide (CANNTG)[170].

Figure 13. bHLH proteins basic region composition.

In the alignment is shown the basic region sequences of some members of the bHLH protein family, divided into the two classes A and B according to the E-box core preference. The residues which recognize the E-box nucleotides are highlighted in green, in light blue the residues which interact with the DNA backbone.

In a work from 1992, the effects of mutations of Myc three Arginines R364, R366 and R367 into Alanines (henceforth, Myc^{RA}) were analyzed; the mutant failed in the transactivation of a reporter gene and also lacked transforming ability consistently with our structural observations. This is not surprising considering the biochemical effects of Arginine to Alanine substitution: the positively charged guanidinium group of the Arginine can protrude from the helix of the Myc basic region to interact with the phosphate group of the DNA backbone while the Alanine, a neutral and small amino acid, would not be able of such interaction. We thus decided to further characterize Myc^{RA}, taking advantage of this mutant to investigate the effects of a general DNA binding impairment on the genomic distribution and biological activity of Myc.

In addition, we designed a new mutant with the aim to inhibit the base-specific contacts between Myc and DNA. In the crystal, the negatively charged side chain of E363 interacts with the E-box nucleotides $C_{1'}$ and $A_{2'}$, while on the other DNA strand, the G_6 is recognized by the H359 residue. If these two residues are mutated into Alanine, the H-bond formation between Myc and the DNA bases should be completed prevented. We thus generated a mutant with Alanine substitutions at those positions, which we named Myc^{HEA}.

3.2 Myc^{HEA} and Myc^{RA} retain normal dimerization with Max

To efficiently bind DNA, Myc needs to dimerize with its binding partner Max. The dimerization domain lays within the Helix-Loop-Helix Leucine zipper domain of both Myc and Max and the mutations we introduced in the basic region should not compromise such interaction. To formally address this, we performed a co-Immunoprecipitation (co-IP) experiment. Human embryonic kidneys (HEK) 293T cells were transfected with plasmids expressing Myc^{wt}, Myc^{HEA} or Myc^{RA} proteins with a FLAG tag at the N-terminus. 48h after transfection, the cells were lysed and subjected to immunoprecipitation with FLAG-agarose conjugated beads. Equivalent amounts of the different FLAG-tagged Myc proteins

co-immunoprecipitated similar amounts of endogenous Max (Figure 14) thus demonstrating that the RA and HEA mutations did not affect Myc/Max dimerization.
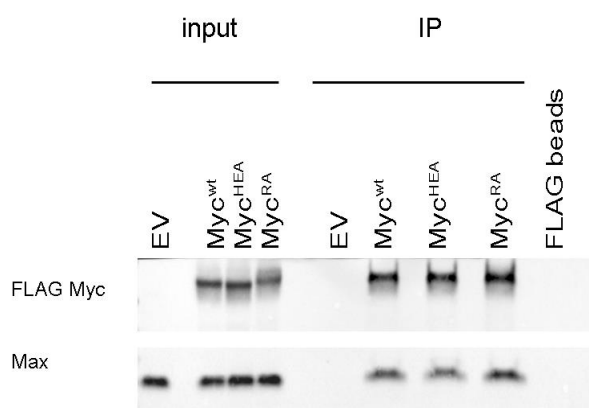

Figure 14. Myc^{HEA} and Myc^{RA} mutants maintained the dimerization ability.
Co-immunoprecipitation of Myc^{wt} or mutant with Max. The blot with the anti-FLAG antibody detected a comparable expression levels among the samples (input) and the immunoprecipitated FLAG-Myc protein (IP); the same samples are tested for the presence of Max.

3.3 Assessment of Myc mutants DNA-binding activities *in vitro*

The Myc protein mutants were first tested for their ability to interact with the DNA in an *in vitro* assay. We used a retroviral vector to express Myc^{wt}, Myc^{HEA} or Myc^{RA} in the c-*myc*-null rat fibroblast cell line HO15.19[259], which lacks the endogenous protein. As assayed by immunoblotting, the various forms of Myc were expressed at comparable levels (Figure 15A). We then used 10 μg of nuclear extracts in an ELISA-based DNA-binding assay (TransAM® c-Myc, as referred in the Materials and Methods section 2.5). As positive control, we used nuclear extract from Jurkat cells (Figure 15B, yellow bar), provided by the manufacturer, while the HO15.19 cells infected with the empty vector (EV) acted as negative control. All three samples expressing Myc showed an increase in the absorbance compared to the EV but the OD value of Myc^{wt} was halved in the Myc^{HEA} sample, confirming the impairment of this mutant in E-box recognition. In turn, the OD value detected in the nuclear extract containing Myc^{RA} was half of the Myc^{HEA} value, according to the idea that the RA mutation was more detrimental for the binding. While still to be

addressed, we speculate that the residual DNA binding activity of the Myc^{RA} mutant was probably due to non-specific interactions between the DNA and other Myc residues; moreover, recognition of half of the palindromic E-box sequence by Max could contribute to the signal observed with either Myc^{HEA} or Myc^{RA}.
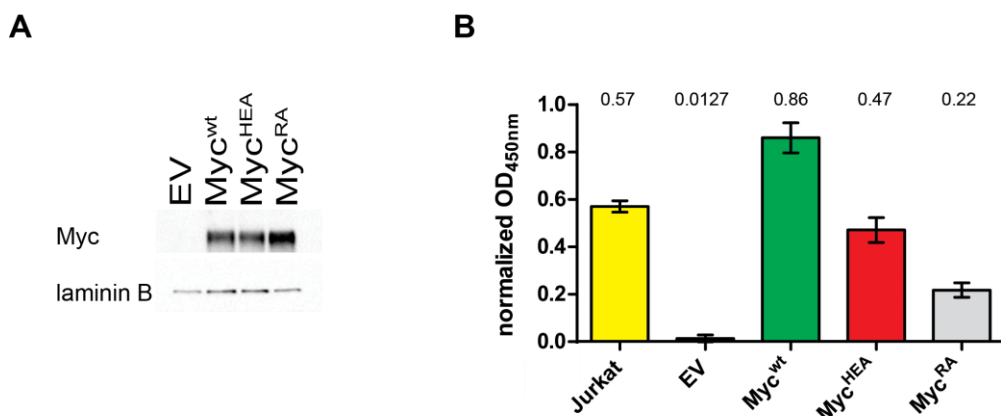
Figure 15. *In vitro* binding ability of Myc^{wt}, Myc^{HEA} or Myc^{RA} protein.
(A) Blot with anti Myc antibody to check the protein levels in the nuclear extracts used for the DNA-binding ELISA. (B) The binding ability of Myc^{wt} and mutants was determined by the TransAM® c-Myc assay (see section 2.1). 10 µg of nuclear extract of rat fibroblast HO15.19 infected with the empty vector (EV) or expressing Myc^{wt}, Myc^{HEA} or Myc^{RA} were used, while 2.5 µg of the positive control (Jurkat cells nuclear extract). On the top of each bar the value of OD detected.

3.4 Re-expression of Myc^{wt} and Myc mutant proteins in rat c-*myc* null fibroblasts

3.4.1 Determination of Myc mutant proteins localization and stability

To investigate the protein stability and the cellular localization of the Myc mutants we took advantage of the aforementioned rat fibroblast line HO15.19, which grows in the absence of the endogenous Myc protein. The levels of Myc^{wt}, Myc^{HEA} and Myc^{RA} expressed from the retroviral vector are shown in Figure 16 together with the endogenous Myc protein in the parental fibroblast cell line, TGR1.
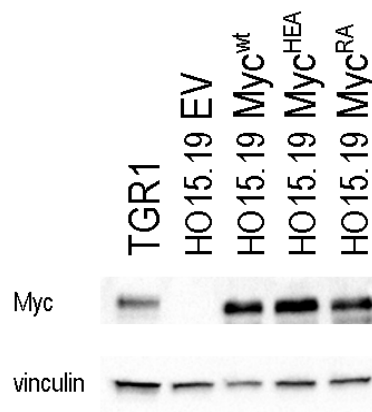
Figure 16. Myc re-expression in rat HO15.19 Myc null fibroblasts.

The expression levels of Myc^wt and mutants in the rat HO15.19 *c-myc* null fibroblasts were tested by western blot. The cells infected with an empty vector (EV) acted as negative control, while the parental cell line from which the HO15.19 has been derived, TGR1, has been used as positive control.

To assess protein stability, cells expressing the different constructs were treated with the translation inhibitor cycloheximide (50 µg/ml) and Myc levels were analyzed at different time-points by western blot, which revealed a similar turnover in all the samples (Figure 17).
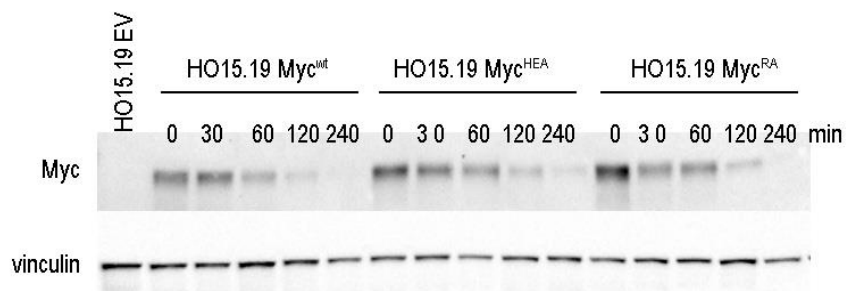
Figure 17. Myc^wt and mutants protein turnover.

Myc protein half-life has been investigated in HO15.19 cells overexpressing Myc^wt and mutants. De novo protein synthesis was blocked by cycloheximide for different time-points and Myc protein levels were analyzed by western blot.

As already described, the Myc nuclear localization signal is composed by two peptides and the residues we mutated in the basic region were located immediately before (E363) and

within (R364, R366, R367) the M2 peptide (aa 364-374). We thus investigated the cellular localization of Myc^{HEA} and Myc^{RA} by immunofluorescence. As shown in Figure 18, the Myc signal (in red) was localized within the nuclei (DAPI, in blue), as proven in the merged image for all the constructs, excluding that the nuclear localization of the mutant proteins was compromised.
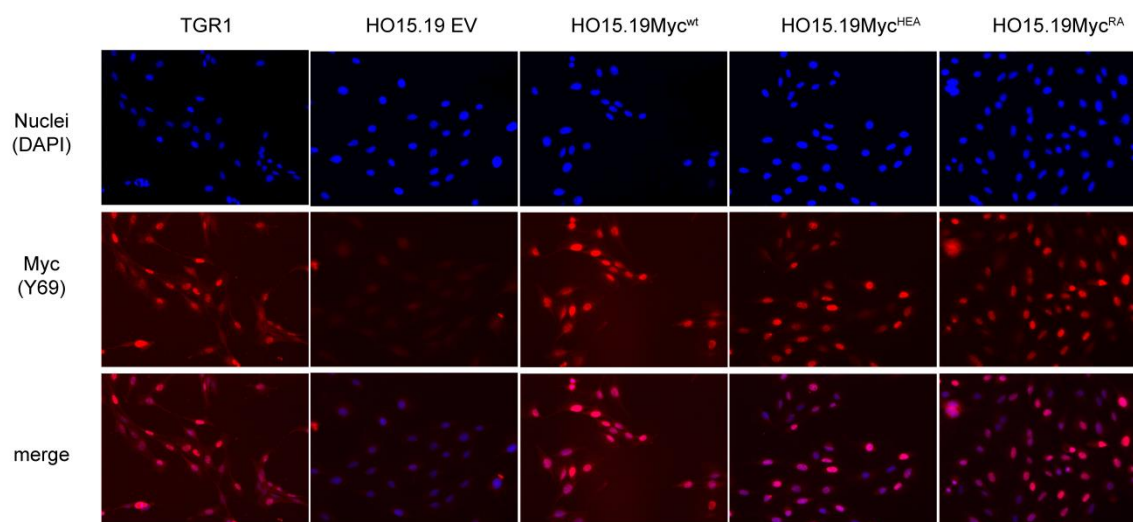
Figure 18. Myc^{wt} and mutants protein cellular localization.
HO15.19 cells infected with Myc^{wt} and Myc mutants were stained for Myc (in red). The nuclei are shown in blue (DAPI) and in the bottom row the two signals are merged. As positive control the parental TGR1 cell line was used, as negative control the HO15.19 Myc null cells infected with an empty vector.

3.4.2 Assessment of the proliferative potential of Myc mutants

To address the capacity of the Myc mutants to promote proliferation we took advantage of the infected HO15.19 cells for a growth curve experiment (Figure 19) and a Colony Forming Assay (CFA, Figure 20). HO15.19 cells infected with the empty vector showed a doubling time of 52 hours, while the parental TGR1 cells had a doubling time of 20 hours. Re-expression of Myc^{wt} in HO15.19 cells reduced the doubling time to 24 hours, while Myc^{HEA} and Myc^{RA} only showed a marginal increase in proliferation relative to EV, with a doubling time around 40 hours.
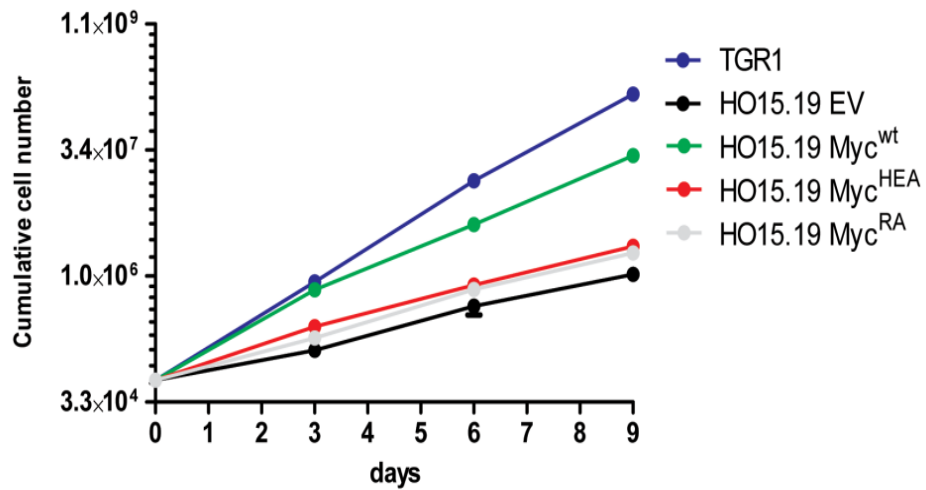
Figure 19. Proliferative ability of rat HO15.19 cells expressing Myc^wt and mutants.
HO15.19 fibroblasts re-expressing Myc^wt, Myc^HEA and Myc^RA have been tested in a growth curve experiment. 70,000 cells were plated at day 0 and subsequently counted and re-plated every three days until day 9. TGR1 and HO15.19 EV cells have been used as positive and negative control, respectively.

This mild proliferative effect of the DNA-binding Myc mutants might be attributable to functions that are not DNA-binding related[146,151,260]. Nevertheless, no rescue at all was observed in the Colony Forming Assay, where only the cells expressing Myc^wt showed an increase in the number of colonies, even if still less than the parental TGR1 cells, while cells expressing Myc^HEA or Myc^RA were comparable to the ones with the EV (Figure 20). Moreover, re-expression of Myc^wt in the HO19.15 cells reconstituted the spindle shape morphology characteristic of the parental fibroblasts TGR1, while Myc^HEA and Myc^RA expressing cells showed a round and flat morphology similar to the cells infected with the EV (Figure 21), suggesting again an impairment of the two Myc mutants.
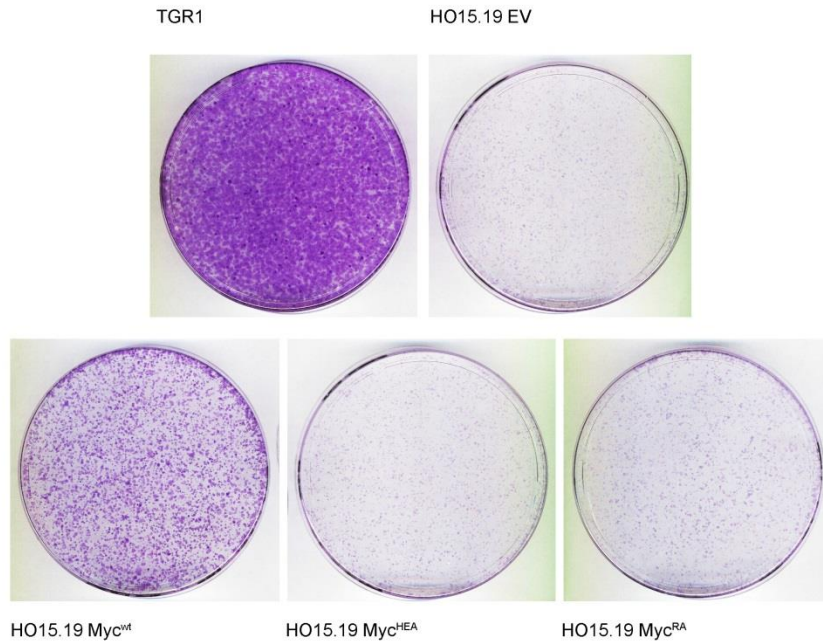
Figure 20. Colony Forming Assay of HO15.19 cells expressing the different Myc mutants.

The ability to promote colonies formation of Myc^{wt} and mutants was tested in a CFA. 10,000 cells were plated at day 0 and after 9 days the cells were stained with crystal violet. The positive control, TGR1, and the negative control, HO15.19 EV, are shown at the top of the figure.
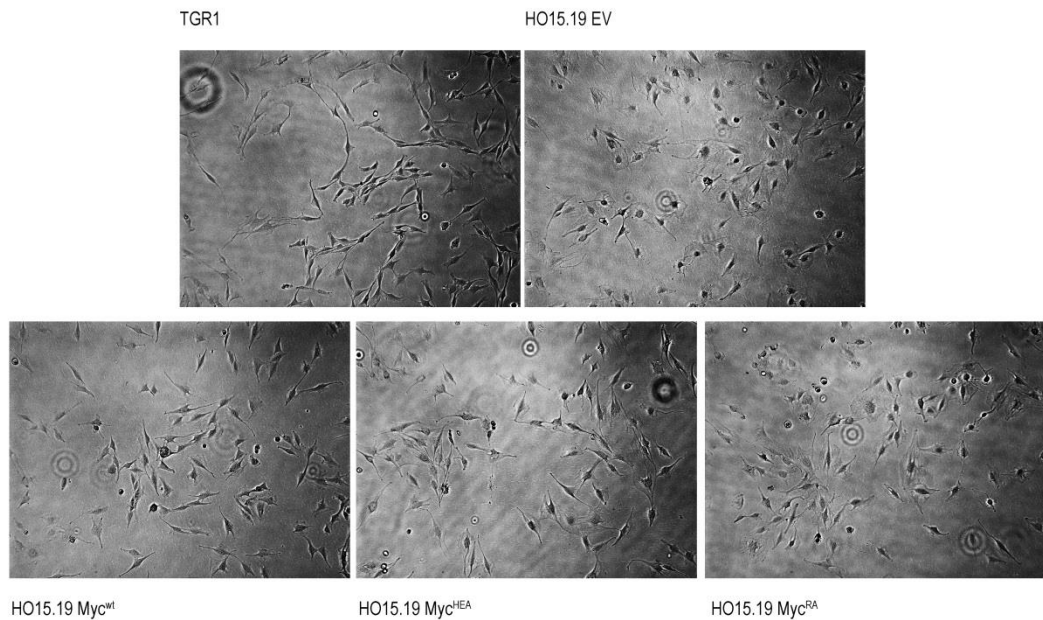
Figure 21. Cell morphology of HO15.19 expressing Myc^{wt} and mutants.

Cellular morphology was visualized by phase-contrast microscopy. In the top panel the positive control, parental TGR1 cells, and the negative control, HO15.19 infected with empty vector.

3.5 Generation of a cellular model for the phenotypic characterization of Myc mutants

To more rigorously assess the activity of our Myc mutants, we decided to generate a cellular system which depended on Myc activity to grow. To this aim, we used an immortalized mouse fibroblast cell line which expressed a tet-myc transgene under the control of a doxycycline-inducible promoter, and targeted the endogenous c-*myc* alleles with CRISPR/Cas9 genome editing. The cellular clone we obtained, named cb9 MycΔb, was characterized by one c-*myc* allele encoding for a truncated protein completely lacking the C-terminal domain while the product of the other allele was a protein deleted of the basic region, the Helix I and the loop (see Materials and Methods, Figure 10). cb9 MycΔb cells were thus functionally knock-out for the endogenous mouse *myc* gene and relied on the expression of the exogenous human tet-Myc transgene for their proliferation as shown by growth curves and colony formation (Figure 22). The doubling time of the cells when the tet-Myc transgene was expressed was around 30 hours, while doxycycline removal led to a complete arrest of the population (Figure 22A) and suppressed colony formation (Figure 22B).
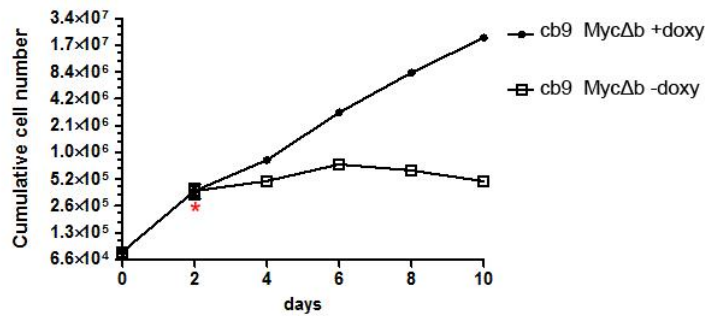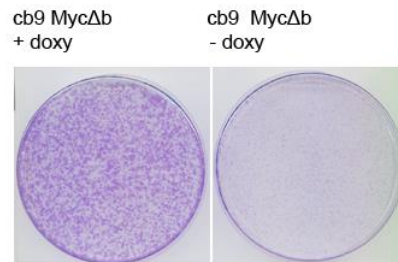
Figure 22. Proliferation ability of cb9 MycΔb fibroblasts.
(A) Growth curve of Cb9 MycΔb cells with and without doxycycline. At day 0 80,000 cells were plated in presence of doxycycline, after two days (red star) the cells were counted and plated with or without doxycycline. The cells were then counted every two days up to day 10. (B) 10,000 cells were plated with or without doxycycline and let grow for 10 days in presence or absence of doxycycline.

We then performed a time-course experiment, collecting cb9 MycΔb cells at different time-points after doxycycline withdrawal to test Myc expression, both at transcript and protein levels. We determined the expression level of the c-*myc* mRNA by qPCR, using primers specific for the human (the tet-Myc transgene) or mouse transcripts (the endogenous transcript, designed in a portion that is maintained after CRISPR/Cas9 deletion) (Figure 23A). Human c-*myc* expression was already suppressed after 8h from doxycycline removal and was maintained silent in all the time-points analyzed. Coherently, the endogenous c-*myc* transcription increased as soon as the tet-Myc was switched off, consistently with the reported auto-regulatory effect[261]. The result was confirmed also by western blot (Figure 23B; note that the double bands visible in the samples without doxycycline most likely correspond to the endogenous mouse Myc). Altogether, cb9

MycΔb fibroblasts were proven as a good Myc-dependent cellular model in which to investigate the functional effects of the mutations in the Myc DNA-binding domain.

A

B
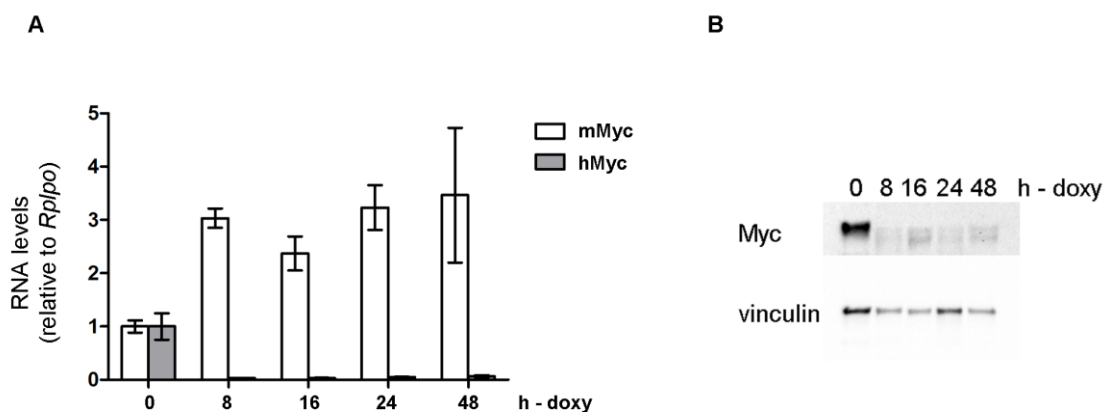
Figure 23. Time-course of tet-Myc transgene expression upon doxycycline withdrawal.
We tested the expression of the tet-Myc transgene at different time-points after doxycycline removal from the growth medium. (A) mRNA levels of the endogenous mouse Myc (white bars) and human tet-Myc transgene (gray bars) at different time-points. (B) Myc protein levels, the antibody used (Y69, abcam) recognizes both human and mouse Myc.

3.5.1 Overexpression of Myc mutants in cb9 MycΔb fibroblasts

We infected the cb9 MycΔb fibroblasts with retroviral vectors expressing either Myc^{wt}, Myc^{HEA} or Myc^{RA} under the control of the CMV promoter. The expression levels of the different constructs were tested 24h after doxycycline removal. Most importantly, the protein levels of the retrovirally expressed Myc proteins were similar among them and lower than the one expressed from the tet-Myc transgene (Figure 24A). The mRNA level of the constitutive Myc^{HEA} was comparable to the Myc^{wt}, while Myc^{RA} transcript level was closer to tet-Myc (Figure 24B). Of notice, overexpression of Myc^{wt} (both the tet-myc and the CMV-driven construct) but not of Myc^{RA} mutant, repressed the endogenous mouse Myc transcription; Myc^{HEA} instead was still able to partially downregulated the transcription of the endogenous loci.
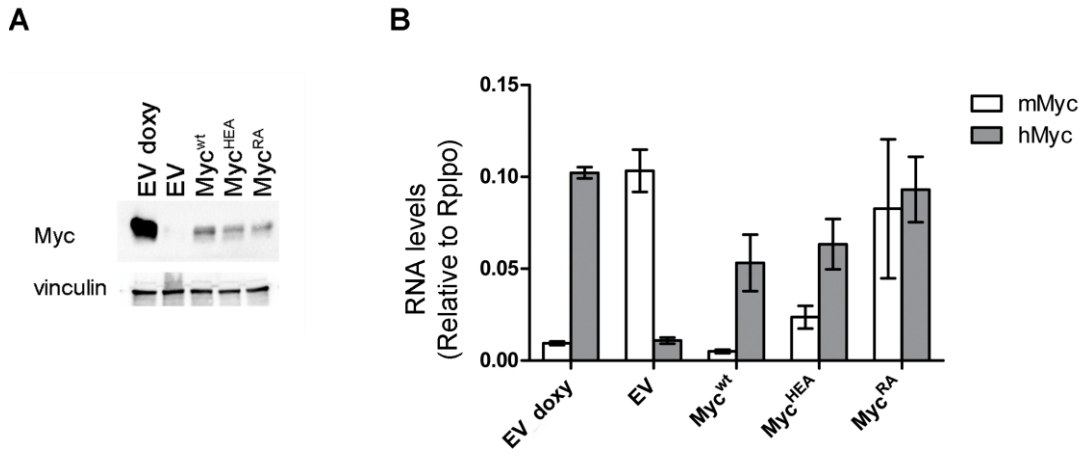
Figure 24. Constitutive Myc expression in cb9 MycΔb cells.

(A) Protein levels of the human Myc proteins constitutively expressed. (B) mRNA levels of the endogenous mouse Myc, white bars, and human Myc transgenes, white bars. Note that the human Myc primers do not discriminate between tet-Myc and constitutive Myc^{wt} and mutants.

The subcellular localization of Myc^{wt} and mutants was checked also in cb9 MycΔb fibroblasts. We performed immunofluorescence experiments on cells growing in presence of doxycycline (expressing both the tet-myc and the CMV-driven constructs, Figure 25A) or 24h after doxycycline removal (expressing only the CMV-driven constructs, Figure 25B): in both cases Myc signal was nuclear, as expected.
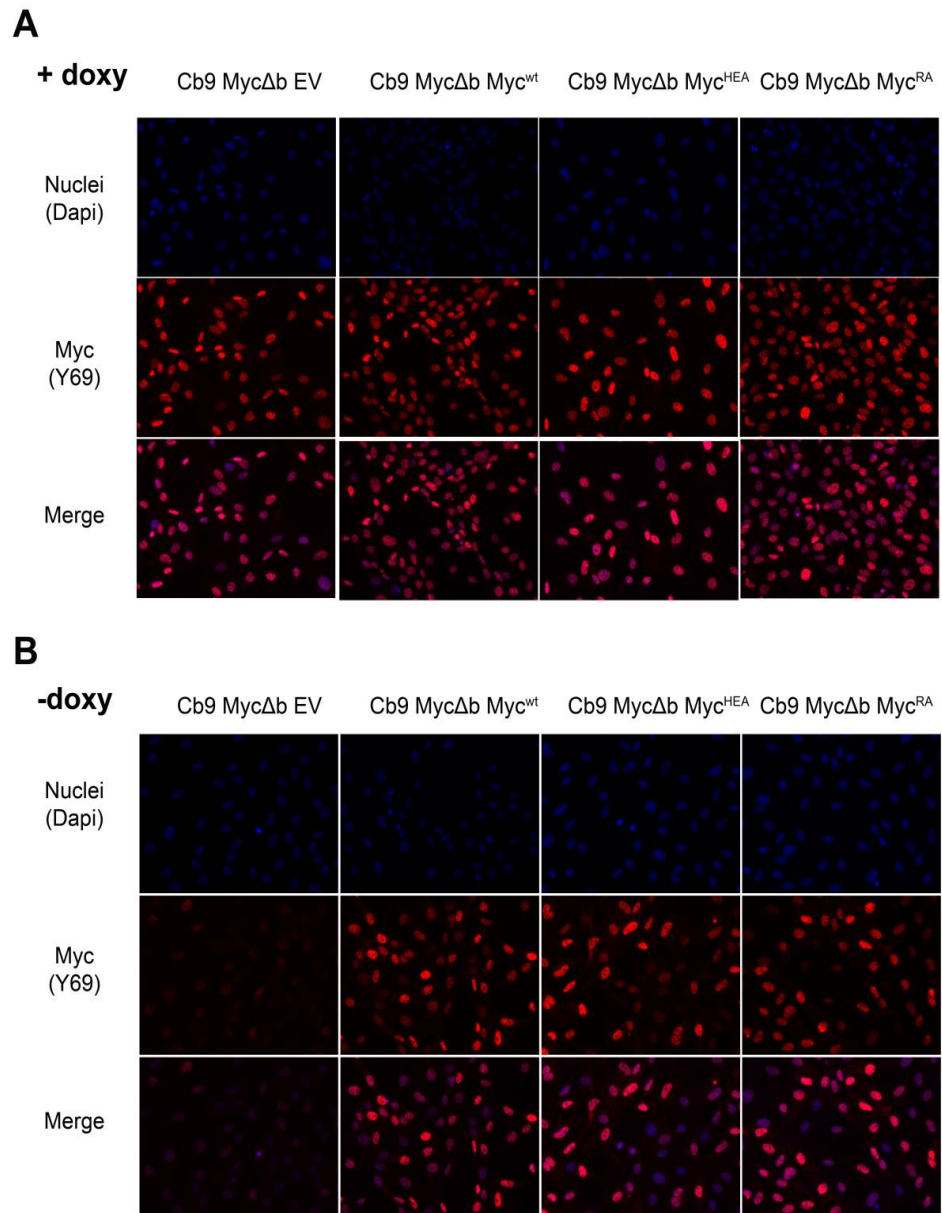
Figure 25. Myc cellular localization in cb9 MycΔb mouse fibroblasts.
Immunofluorescence experiment to test the ability of Myc mutants to localized into the nucleus. The nuclei are shown in blue, Myc signal in red and in the bottom row the two signals are merged. (A) Cells growing in presence of doxycycline were used as positive control. (B) Cells fixed 24h after the doxycycline removal.

3.5.2 Myc mutants show no proliferative activity

To assess the ability of the Myc mutants to sustain cellular proliferation, we determined the doubling time of the cells infected with EV, Myc^{wt}, Myc^{HEA} and Myc^{RA} in presence or in absence of doxycycline (Figure 26). When the tet-Myc transgene was active, all the samples showed a doubling time of around 30h, with the exception of Myc^{HEA}-expressing cells, whose doubling time was delayed to 45 h, suggesting a dominant-negative effect of this mutant (Figure 26A). On the contrary, when tet-Myc was switched off, only the cells expressing Myc^{wt} maintained proliferation (doubling time, ca. 38h), while Myc^{HEA} and Myc^{RA} were both unable to sustain population growth (Figure 26B).
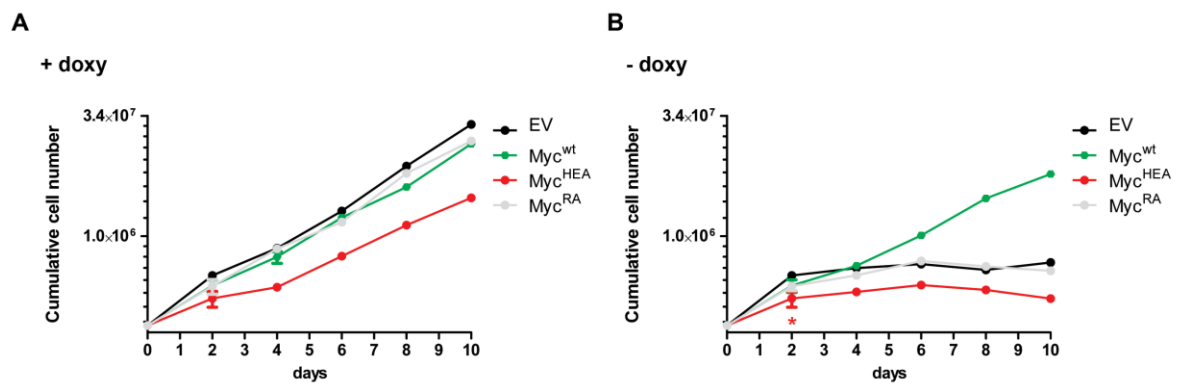
**Figure 26. Proliferative ability of cb9 MycΔb cells expressing Myc^{wt}, Myc^{HEA} or Myc^{RA}.**
(A) Growth curve of cb9 MycΔb cells constitutively expressing Myc^{wt}, Myc^{HEA} or Myc^{RA} in presence of doxycycline. 80,000 cells were plated at day 0, the cells were counted and re-plated every two days up to day 10. (B) Growth curve of the same samples, at day 2 of the growth curve doxycycline was removed from the culture (red star).

We also evaluated the fraction of cells in S-phase with BrdU staining 24h after doxycycline removal (Figure 27). The percentage of BrdU positive cells in the presence of doxycycline was comparable among the samples with a slight increase in cultures expressing the CMV-driven Myc compared to the EV. The removal of doxycycline determined a decrease in the S-phase cells in all the samples except the one expressing the constitutive Myc^{wt}. The impairment of both Myc^{HEA} and Myc^{RA} mutants to sustain cellular

growth was confirmed in a colony forming assay, where only Myc^{wt} expressing cells were able to form colonies in absence of doxycycline (Figure 28).
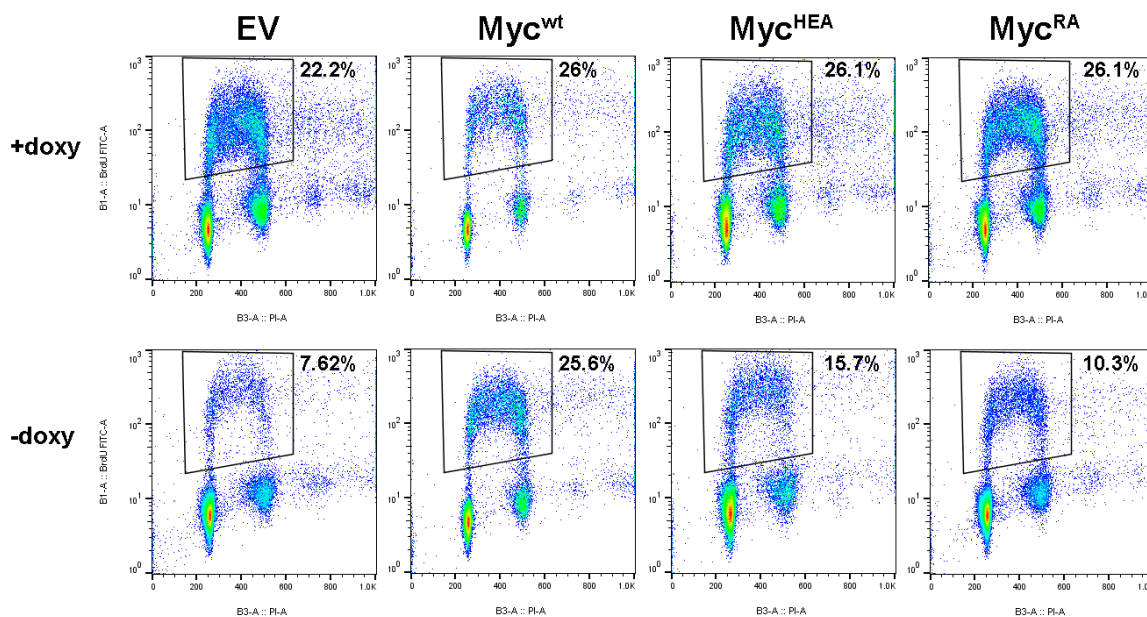
Figure 27. S-phase analysis of cb9 MycΔb cells expressing Myc^{wt}, Myc^{HEA} or Myc^{RA}.
Fluorescence-activated cell sorter (FACS) profiles of cb9 MycΔb cells expressing Myc^{wt}, Myc^{HEA} or Myc^{RA} in presence (upper panel) or in absence (lower panel) of doxycycline in the culture medium along with the percentage of BrdU-positive cells measured.
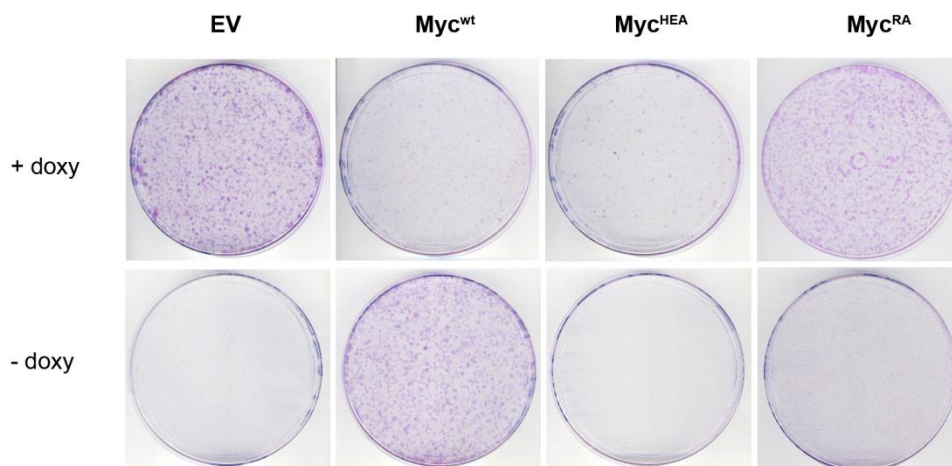
Figure 28. Colony forming potential of cb9 MycΔb cells expressing Myc^{wt}, Myc^{HEA} or Myc^{RA}.
10,000 cells constitutively expressing Myc^{wt}, Myc^{HEA} or Myc^{RA} were plated and stained with crystal violet at day 10. As control all the samples were also grown in presence of doxycycline.

Of notice, cells overexpressing Myc^{HEA} in conditions of the tet-Myc transgene activation showed an impairment in colony formation, consistent with the effect seen in the growth curve, again suggesting a dominant-negative effect. We also observed that Myc^{wt} overexpressing cells form less colonies relative to the EV and Myc^{RA} expressing cells when the tet-myc transgene was on: this may be due to the cumulative expression of Myc, reaching Myc levels above the threshold that induces apoptosis[179,232]. We are currently evaluating this hypothesis by different assays for apoptosis detection.

3.5.3 Genome-wide analysis of DNA-binding activities

The investigation of the genomic occupancy of the Myc mutants was performed by chromatin immunoprecipitation coupled to high-throughput sequencing (ChIP-Seq) with a Myc-specific antibody. cb9 MycΔb cells infected with empty vector (EV) or vectors constitutively expressing Myc^{wt}, Myc^{HEA} or Myc^{RA} were plated in absence of doxycycline and fixed with formaldehyde after 24h; as a positive control, we used the cells infected with the empty vector and kept in doxycycline to express the tet-myc transgene (EV doxy). We first checked some Myc-target (*Ncl*, *CAD*, *Pus7* and *D7*) and non-target (*AchR*) promoters by qPCR (Figure 29). While the cells expressing Myc^{wt}, either through the tet-myc transgene (EV doxy) or the retroviral vector (Myc^{wt}), showed Myc binding to all the regions tested, cells expressing the two mutants, showed very low enrichments, comparable to the negative control.
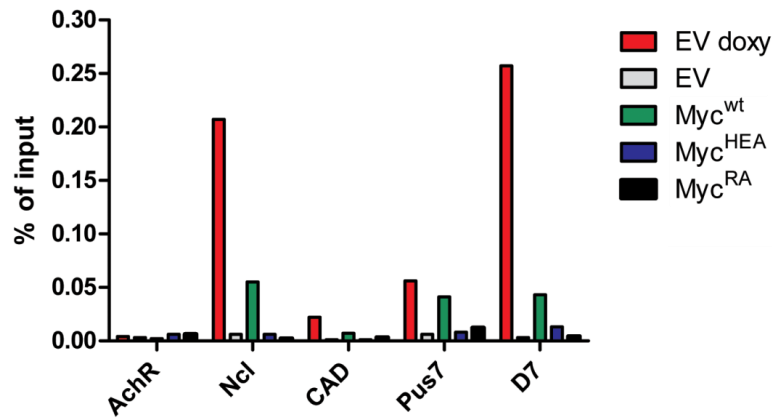
Figure 29. Myc binding to the promoter of some target genes.
ChIP-qPCR at promoters of Myc selected targets (*Ncl*, *CAD*, *Pus7* and *D7*) and negative control region (AchR).

After next-generation sequencing of the immunoprecipitated DNA, the peak calling algorithm retrieved more than 22,000 Myc binding sites in the positive control, only few peaks in the negative, around 16,000 for Myc^{wt}, 8,000 for Myc^{HEA} and only 250 for Myc^{RA} (Figure 30). Hence, at this level already, we could surmise that Myc^{HEA} retained some DNA binding activity, while Myc^{RA} was profoundly impaired.
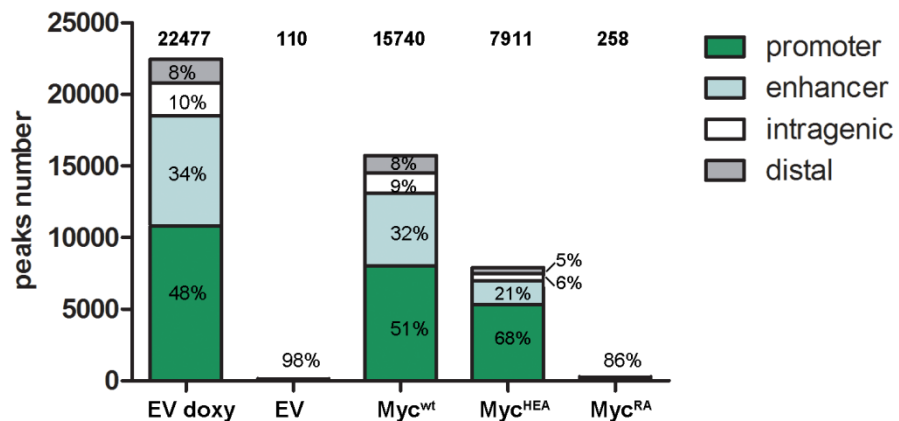
Figure 30. Myc peaks number and distribution.
Number of peaks identified in the ChIP-Seq experiment and their annotation at promoters (-2kb to +1kb from the TSS), enhancers (sites different from promoters and marked with H3K4me1), intragenic (in the gene bodies, more than +1kb away from the TSS) or distal regions (none of the above).

In tet-myc expressing cells, around half of the peaks were located in promoters, while the remaining binding sites were mainly in the enhancer regions (defined as sites marked with H3K4me1 different from -2Kb + 1Kb from an annotated refseq TSS) and only a low percentage inside gene bodies or in distal sites. The distribution of Myc-binding sites in Myc^{wt}-infected cells was similar. The number of peaks was halved in the cells expressing Myc^{HEA}, with a percentage slightly higher on promoter regions (62%).

Almost all the Myc peaks of each sample were a subset of the positive control EV doxy and interestingly the vast majority of the Myc^{HEA} peaks were retrieved also in the Myc^{wt} sample, indicating that the binding sites retained by the Myc^{HEA} mutant were a subset of the Myc^{wt} targets (Figure 31).
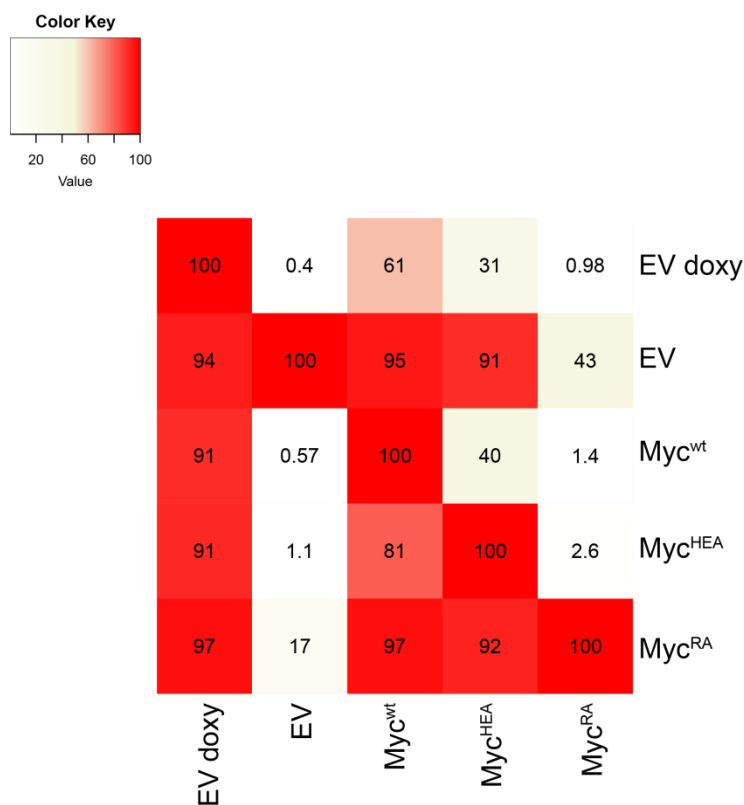
Figure 31. Overlap of Myc-ChIP peaks among the samples.
Percentages of peaks overlapping (by at least 1 bp) among the different samples are reported in each column; cells expressing tet-Myc transgene (EV doxy) and the cells infected with the empty vector (EV) were considered as positive and negative controls, respectively.

The same result could be visualized in a heatmap showing the union of all the peaks on chromosome 1 obtained in the different samples (without expression of the tet-myc

transgene) as shown in Figure 32. The Myc^{RA} sample presented virtually no peaks (similar to the empty vector) and only a very low read density (close to background) in the regions characterized by Myc peaks in the other samples. Myc^{HEA} instead showed many of the Myc^{wt} peaks, but it was characterized by a lower read density.

Of notice, the analysis of the binding intensity (defined as the value of peak enrichment) of Myc^{wt} and Myc^{HEA} proteins to their binding sites revealed that the peaks in common between the two samples, which represented the majority of the Myc^{HEA} peaks, correspond to the Myc high affinity targets (Figure 33).
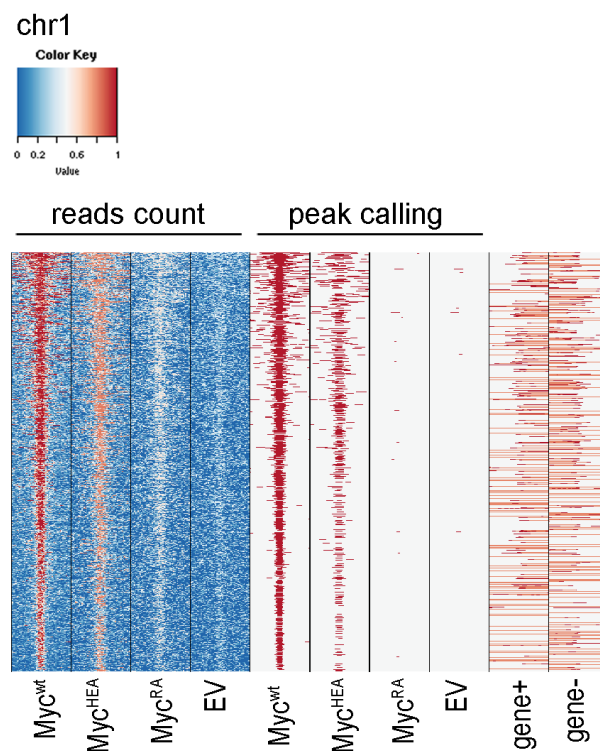
Figure 32. Intensity and distribution of Myc peaks on chromosome 1.
Heatmap showing the library size-normalized ChIP-Seq reads count on chromosome 1 at regions bound by Myc in at least one sample. The first four columns show the reads coverage for Myc^{wt}, Myc^{HEA}, Myc^{RA} and the Empty vector, respectively. In columns five to eight the peaks calling from the same samples is shown. In the last two columns the presence of a gene in the sense (gene +) or antisense (gene -) DNA strand is represented.
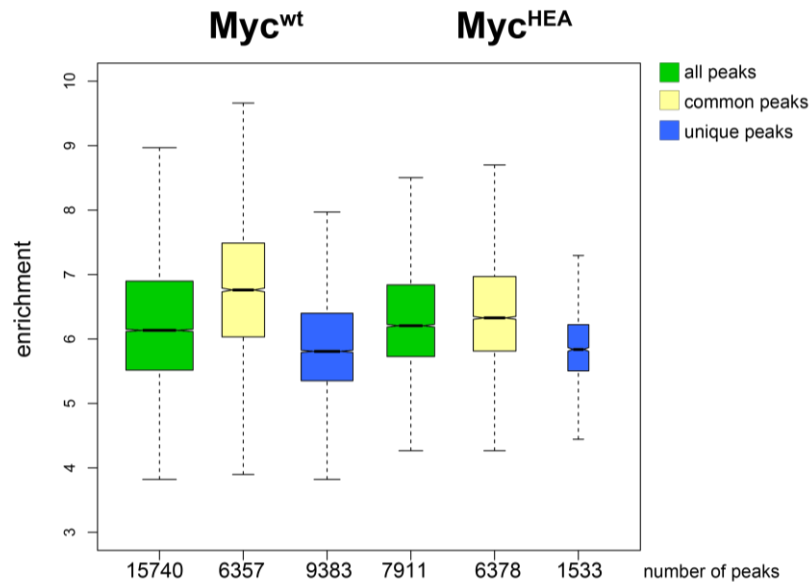
Figure 33. Binding intensity of Myc^{wt} and Myc^{HEA} proteins.

The enrichment of the peaks of each sample: in green the boxplot of all the peaks, in yellow the enrichment of the peaks in common between Myc^{wt} and Myc^{HEA} and in blue the boxplot of the peaks specific for each sample. On the x axis the number of peaks in each category is shown.

3.5.3.1 *E-box analysis*

The primary tracks of the different ChIP-seq experiments in correspondence to some Myc-bound promoters, divided in sites without an E-box in the region ±200 bp from the peak summit (i.e. *Stc35f5*, *Sept2*, *Gigyf2, Gin1*, Figure 34 left panel) or with one or more E-boxes (i.e. *Ncl*, *Pus7, Ubfd1*, also known as *D7*, *Stc2*, Figure 34, right panel), showed a complete impairment in the DNA binding by the Myc^{RA} protein, while Myc^{HEA} seemed to be defective more specifically in the recognition of the target sites containing the E-box. This observation was consistent with the HEA mutation designed, which should disrupt the E-box recognition without interfering with the general DNA binding. We next focused our investigation on Myc^{wt} and Myc^{HEA} samples, analyzing their ability to recognize the E-box.
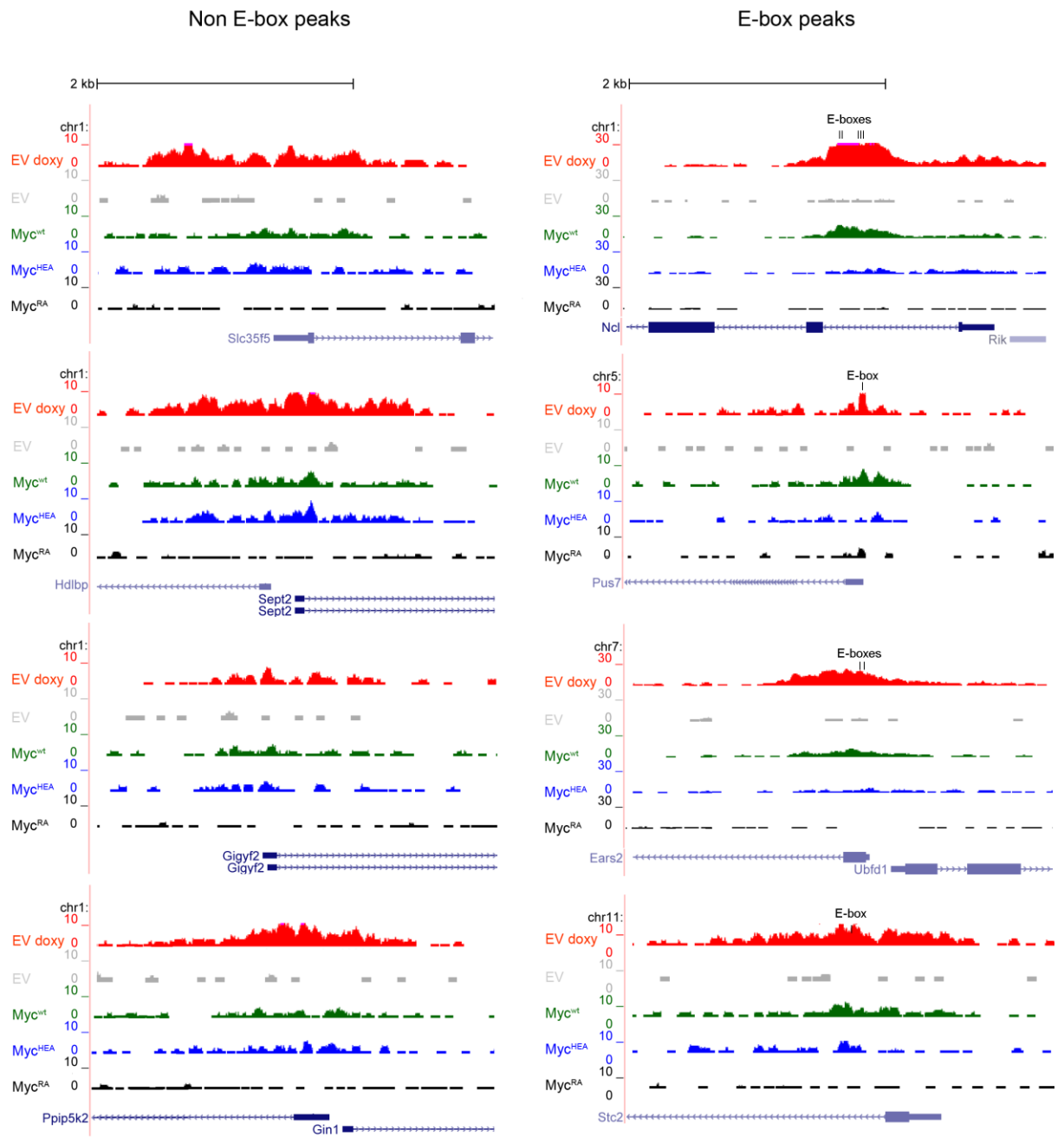
Figure 34. Genome Browser tracks at different Myc binding sites.

Genome Browser tracks of Myc ChIP-seq peaks, divided into non E-box-containing regions (left panels) and E-box-containing targets (right panels). In each screenshot from top to bottom: tet-Myc expressing cells (EV doxy) as positive control, the cells infected with the EV as negative control and the cells expressing Myc^{wt}, Myc^{HEA} and Myc^{RA} are represented .The RefSeq genes tracks are shown at the bottom.

We first determined the proportion of peaks that contained the canonical E-box (CACGTG) or, alternatively, at least one of the non-canonical sequences (CA<u>T</u>GTG, CACG<u>C</u>G, CA<u>T</u>G<u>C</u>G, CACG<u>A</u>G) in Myc^{wt} and Myc^{HEA} samples. As shown in Figure 35, the analysis was performed either on the entire set of peaks, or on peaks localized at promoters, or on those on enhancers. In all three cases, we observed an increase in the fraction of peaks not containing any E-box in the Myc^{HEA} compared to Myc^{wt} sample.
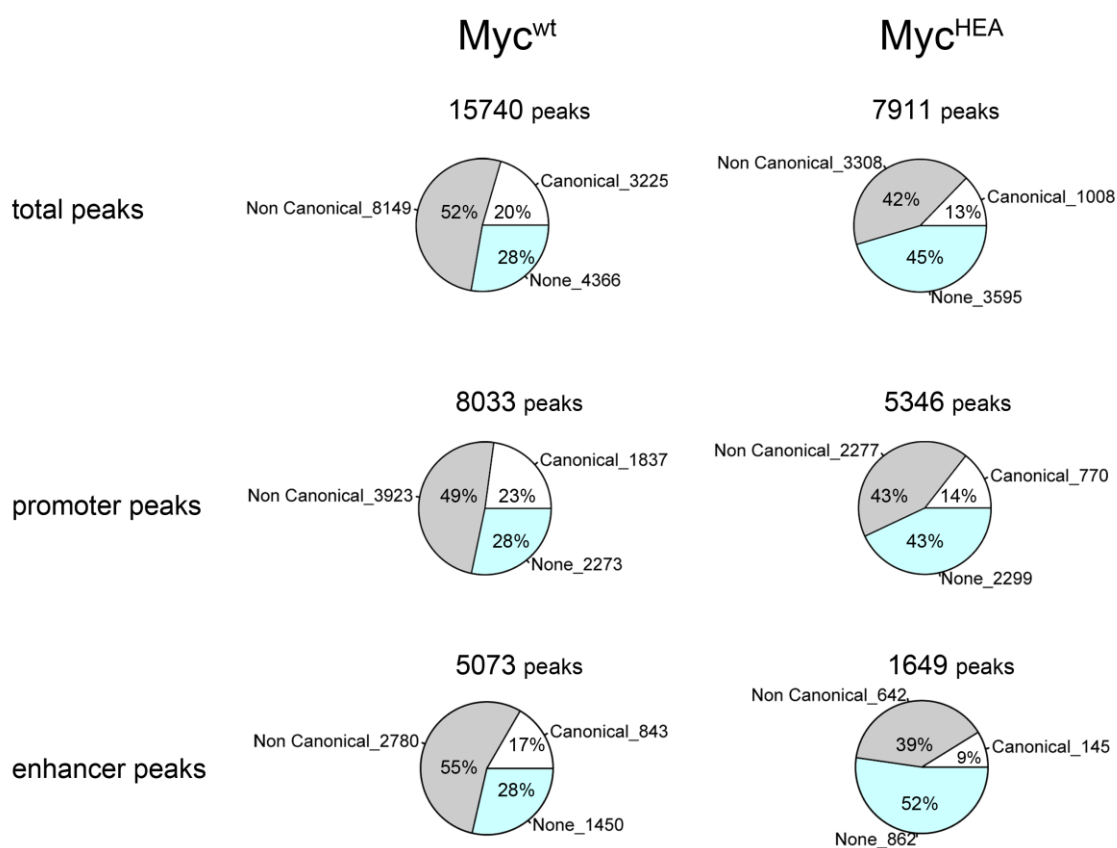
Figure 35. Percentage of Myc binding sites containing the canonical or non-canonical E-boxes.
The pie-charts show the percentages of peaks in Myc^{wt} and Myc^{HEA} ChIP-seq that contain (in the region ±200 from the peak summit) the canonical E-box (CACGTG, in white), or at least one of the four non-canonical E-boxes (CATGTG, CACGCG, CATGCG, CACGAG, gray) or none of them (light blue). The analysis was performed considering all the peaks, only the peaks on the promoters or only the peaks on the enhancers.

We then investigated if the residual fraction of E-box, both canonical and non-canonical, present in the Myc^{HEA} peaks was significantly enriched over the random expectation. We first defined the so called "E-box genomic background" checking the presence of the E-

boxes in a window of 400 nt upstream of all the open TSS (defined based on the DNAseI-hyper-sensitivity data, GEO accession GSM1230377[90]) and then assessed the difference in the E-box distribution among our samples and this background using the χ^2 test. The results of this analysis are shown in Figure 36. The pie-charts at the top represent the canonical E-box in all open promoters, Myc^{wt}- and Myc^{HEA}-bound promoters, respectively, and the pie-charts at the bottom show the same analysis for the non-canonical sequences. Both the canonical and non-canonical E-boxes were enriched in Myc^{wt} and Myc^{HEA}, but with a different degree of significance: for Myc^{wt} sample the p-value for both types of sequences was extremely significant (p<1E-15), while the p-values in the Myc^{HEA} sample revealed a less significant enrichment. Interestingly, the direct comparison between Myc^{wt} and Myc^{HEA} revealed an extremely significant difference in the E-boxes content.
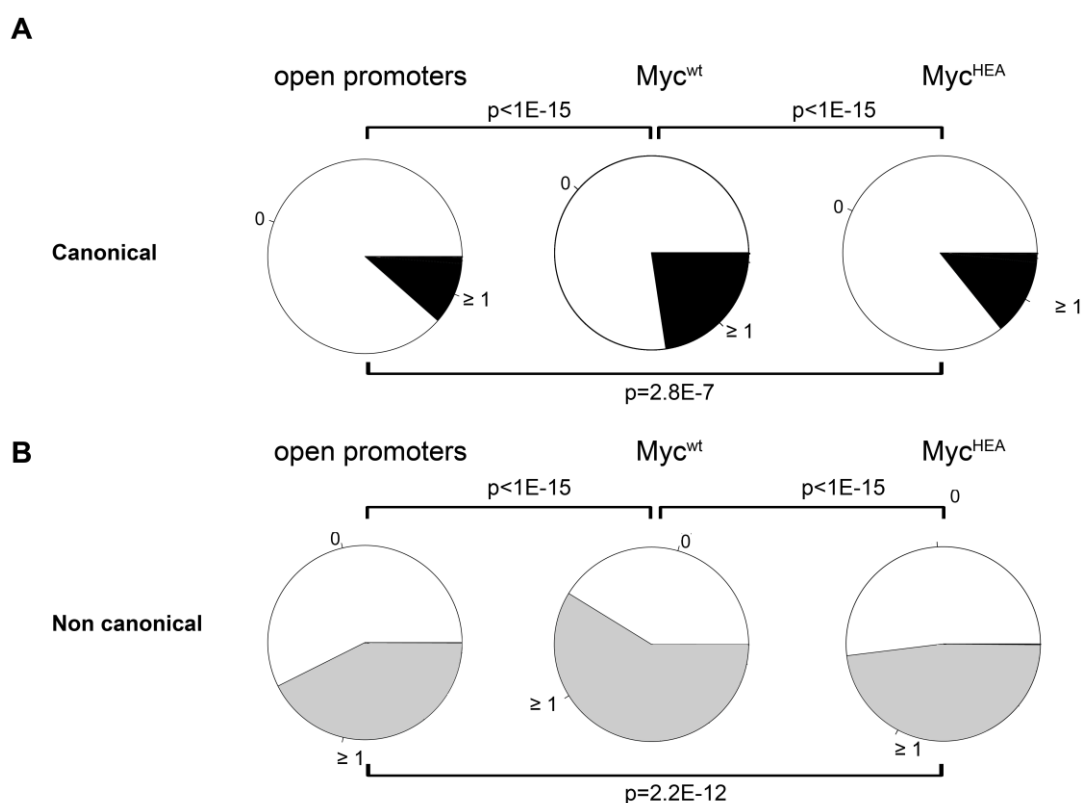
Figure 36. Significance of fraction of E-boxes present in Myc^{wt} and Myc^{HEA} peaks.
Myc^{wt} and Myc^{HEA} bound promoters have been analyzed for the presence or absence, in a window of ±200 nt from the peak summit, of at least one canonical E-box (A) or a non-canonical one (B). The same analysis was performed on all the open promoters, in a genomic range of 400 upstream the TSS of the open promoters. The results are shown as pie-charts and the differences among the samples were tested with the χ^2 test, the p values are shown.

We next analyzed the binding intensity of Myc[wt] and Myc[HEA] at their binding sites stratified according to the presence of the canonical E-box, at least one of the non-canonical E-boxes, or neither of them (Figure 37); for Myc[wt], as expected, the regions containing the canonical E-box had the highest enrichment, followed by the peaks with the non-canonical sequences[18,52,262,263]. This hierarchy was completely lost in Myc[HEA] sample, where the three categories showed comparable enrichments, suggesting that the sequence-recognition ability was lost and the sites containing the E-box were bound as well as the ones without it. This result was confirmed also analyzing the enrichment of only the binding sites that were in common between Myc[wt] and Myc[HEA] both at promoters or enhancers (Figure 38). Altogether, while Myc[HEA] still showed a residual enrichment of the E-box relative to background, which may partly be due to its association with wild-type Max, the data were consistent with impaired sequence recognition by this mutant.
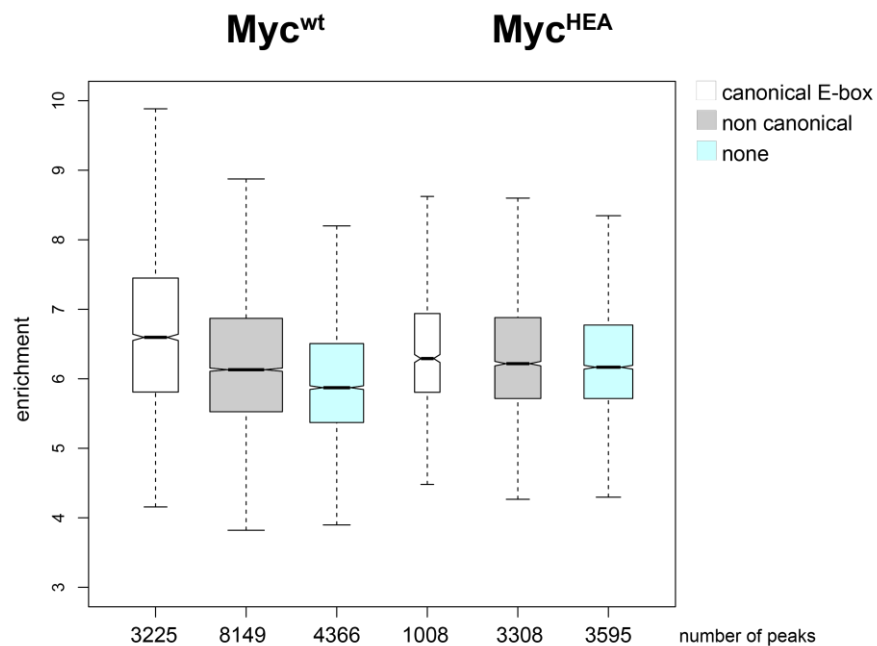
Figure 37. Enrichment values of Myc[wt] and Myc[HEA] peaks.

The enrichment of all the peaks for each sample, categorized according to the presence of the canonical E-box (white), non-canonical E-boxes (gray) or none (light blue), is shown as boxplot. At the bottom of each boxplot the number of peaks belonging to each category is reported.
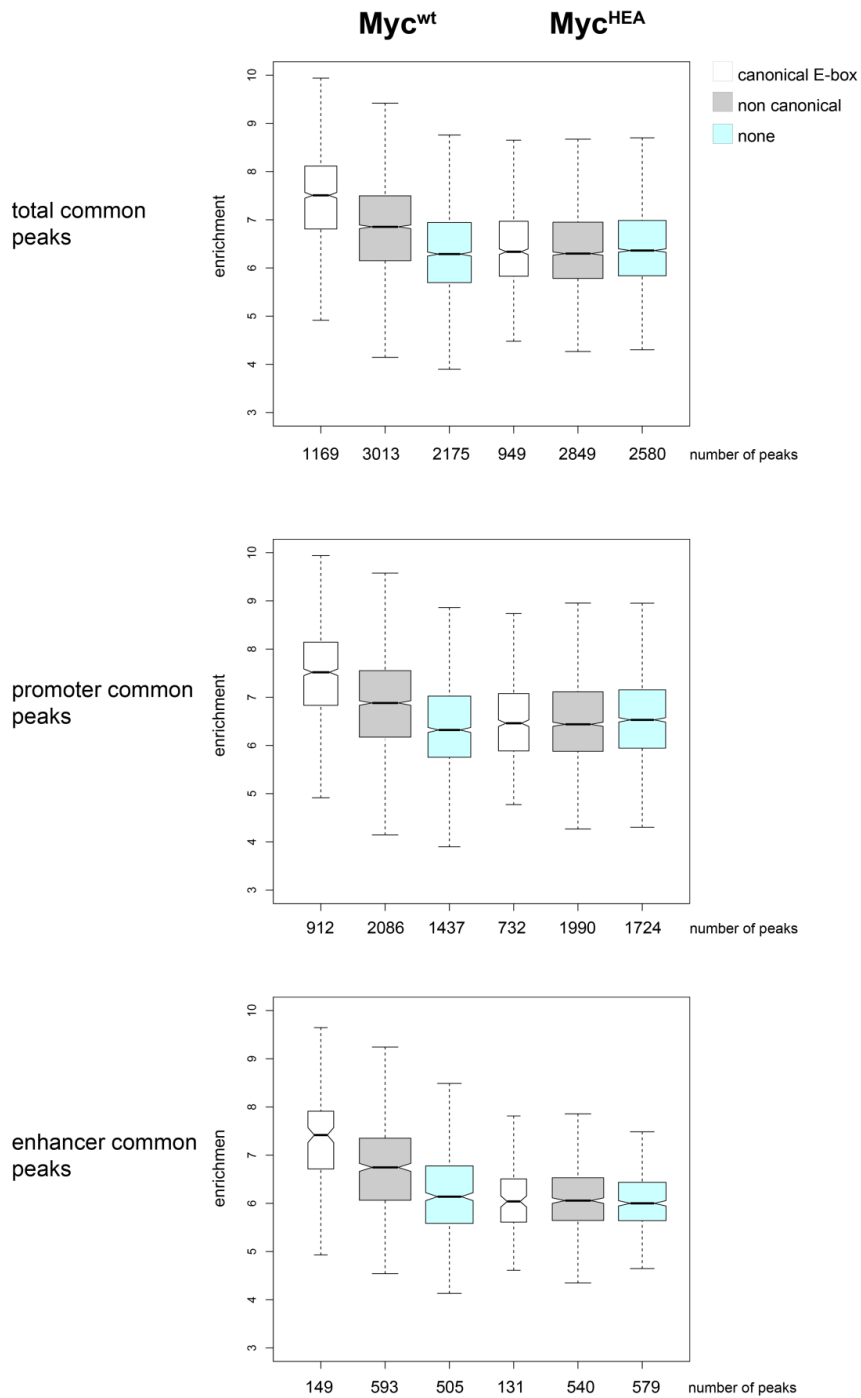
Figure 38. Binding intensity of Myc^{wt} and Myc^{HEA} protein in the common regions.
Boxplot representing the enrichment of Myc peaks in the binding sites shared between Myc[wt] and Myc[HEA], stratified for the presence of the canonical E-boxes (white), non-canonical E-box (gray) or none (light blue). The analysis was performed on all peaks and also dividing the peaks in promoter-specific and enhancer-specific. At the bottom of each boxplot the number of peaks belonging to that category is reported.

The loss of specific recognition of the E-box by the Myc^{HEA} mutant was also supported by the analysis of the distribution of the motif relative to the peak summits (Figure 39). Indeed, for wild type Myc (both tet-myc and CMV-driven Myc^{wt}) the E-boxes were found exactly under the peak summit, while they were much more delocalized in the Myc^{HEA} sample.

The genomic data we generated suggested that mutations of the Myc residues engaged in the sequence-specific interaction produced a protein still able to interact with DNA and to maintain most of the protein-protein interactions that in part mediate the binding of Myc to the chromatin, but lacking the ability to recognize and get stabilized on the target DNA sequence, i.e. the E-box.
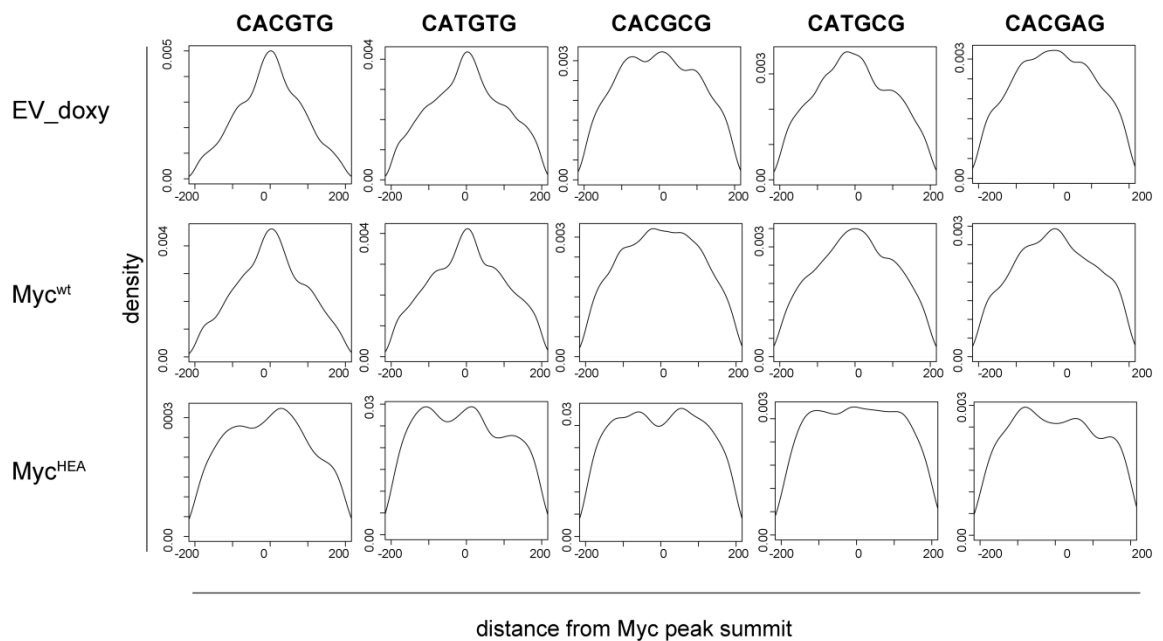
Figure 39. E-box distribution under Myc peaks.
Distribution of the distance from the peak summits of the canonical E-box (CACGTG) and each of the non-canonical sequence (CATGTG, CACGCG, CATGCG and CACGAG) in the different samples (the positive control EV doxy, Myc^{wt} and Myc^{HEA}).

3.5.3.2 Analysis of low-affinity non-E-box motifs

In addition to the E-box and its variants, other sequence preferences were recently found for Myc in high-throughput screenings performed with protein-binding microarrays

(PBMs). Two independent studies pointed out that, apart from the E-box and the non-canonical E-boxes, which are the most significantly bound probes, also the sequences composed by half E-box (either CAC or CAT) were among the top targets [264,265]. Since our Myc[HEA] mutant was not stabilized by the complete E-box hexamer, both canonical and non-canonical, we did not examine our data for such sequences; for the same reason the contribution of the nucleotides flanking the E-box was not investigated, even if they were found to influence the DNA binding strength of Myc in PBM data[201,264,265]. Instead, we evaluated the enrichment of the hexamer AACGTT, which was identified as a motif preferentially bound by Myc/Max both in a PBM experiment and then confirmed with an electrophoretic mobility shift assay (EMSA), and also in ChIP-Seq datasets[265].

As already performed for the E-box, we checked for the presence of the AACGTT sequence in all mouse fibroblast open promoters and defined these counts as the "genomic background". We than applied the χ^2 test to assess the difference in the AACGTT distribution between the Myc[wt] or Myc[HEA] sample and this background. The results of the χ^2 test revealed that the sequence AACGTT was slightly enriched both in Myc[wt] and Myc[HEA] expressing cells, with a p-value of 0.018, but not at the same degree of significance as the E-box sequences. To conclude, beside the new DNA sequences Myc has been found to interact with, the E-box remained the top target and our *in vivo* data demonstrated that when the E-box recognition is impaired the Myc functions are compromised.

3.6 Generation of Myc[HEA] knock-in cell clones

The DNA binding activity of the Myc[HEA] mutant seemed in contrast with the almost null biological activity of the protein in sustaining cell growth. We thus wondered whether this residual binding ability corresponded to a phenomenon of non-productive invasion of the active chromatin in conditions of Myc overexpression. We repeated the characterization of

the Myc^{HEA} mutant in a cellular system where its expression was driven by the endogenous promoter. The HEA mutations were inserted into the endogenous *c-myc* gene loci of cb9 tet-Myc cells by CRISPR/Cas9. We obtained two cell clones (cb9 Myc^{HEA} clones 10 and 33) in which both the c-*myc* alleles were replaced. As control, we isolated two cellular clones from the parental fibroblasts population that still contained the c-*myc* wt sequence (cb9 Myc^{wt} clones 6 and 10). The levels of the endogenous Myc protein differed among the clones and are shown in Figure 40.
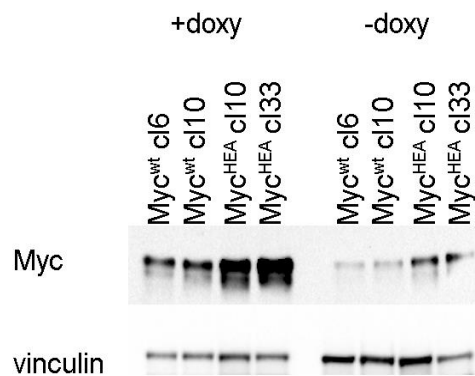
Figure 40. Endogenous Myc protein levels in cb9 clones.
Myc protein levels of the cb9 clones, expressing Myc^{wt} or Myc^{HEA}, with or without doxycycline, were tested in a western blot experiment.

The impairment of the Myc^{HEA} mutant in promoting cellular growth as well as colony formation was confirmed also in this cellular system. The doubling time of the different clones in presence of doxycycline varied between 20 and 25 hours (Figure 41A); when doxycycline was removed from the medium, the growth rate of Myc^{wt} clones decreased to 40 hours, but the cells were still proliferating. On the contrary, Myc^{HEA} clones passed from a doubling time of around 24h to a complete block of the proliferation (Figure 41B). The same phenotype was observed in a CFA (Figure 42): all the samples were able to form colonies in presence of doxycycline, with a variability that is in accord with their clonal nature, but in absence of doxycycline only the Myc^{wt}-expressing clones still formed colonies.
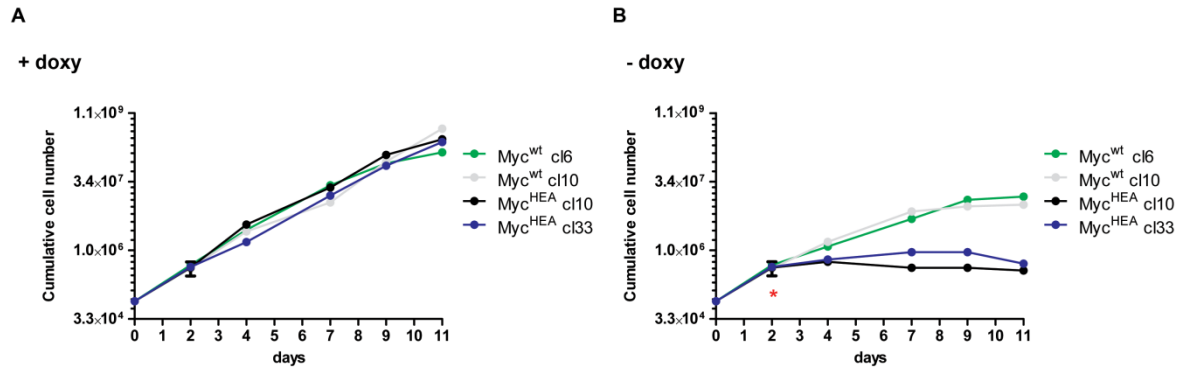
Figure 41. Proliferative ability of cb9 clones expressing Myc^wt or Myc^HEA.
80,000 cells are plated at day 0 and counted every two-three days up to day 11. (A) cb9 clones expressing Myc^wt or Myc^HEA growing in presence of doxycycline. (B) Growth potential of cb9 clones expressing Myc^wt or Myc^HEA in absence of doxycycline, the red star indicates the day when doxycycline was removed from the medium.
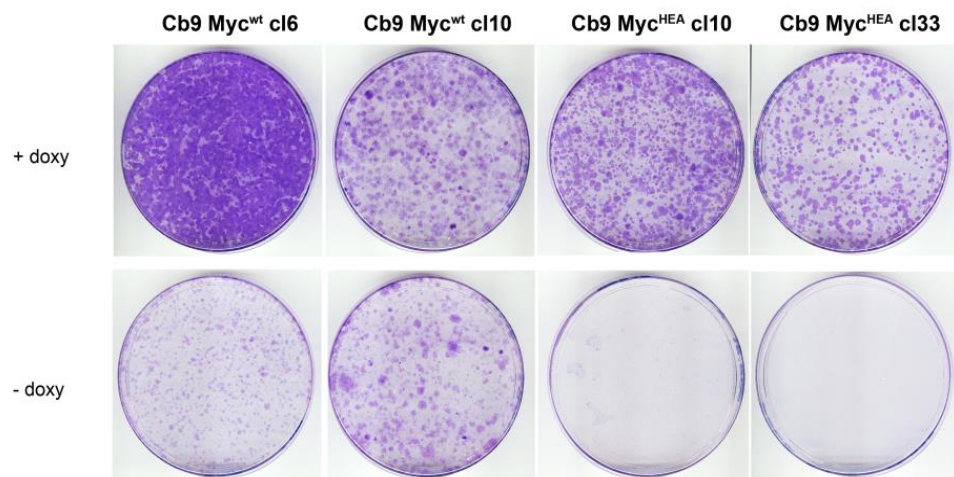
Figure 42. Colony forming assay (CFA) of cb9 clones expressing Myc^wt or Myc^HEA.
The Myc^wt clones and the Myc^HEA clones were plated four days at low concentration, 10,000 cells, with or without doxycycline. After 11 days the samples were stained with crystal violet.

We performed a ChIP-Seq experiment on the different clones both in presence and in absence of doxycycline. We first checked by qPCR two Myc-target promoters (*Ncl* and *CAD*) and a non-target sequence (*C/EMP* α): all the samples grown with doxycycline showed Myc binding to the target regions (Figure 43A). In the absence of doxycycline instead, cells expressing Myc^{wt} showed binding to *Ncl* and *CAD* promoters, while those expressing Myc^{HEA} showed a dramatic decrease in the % of input values (Figure 43B).
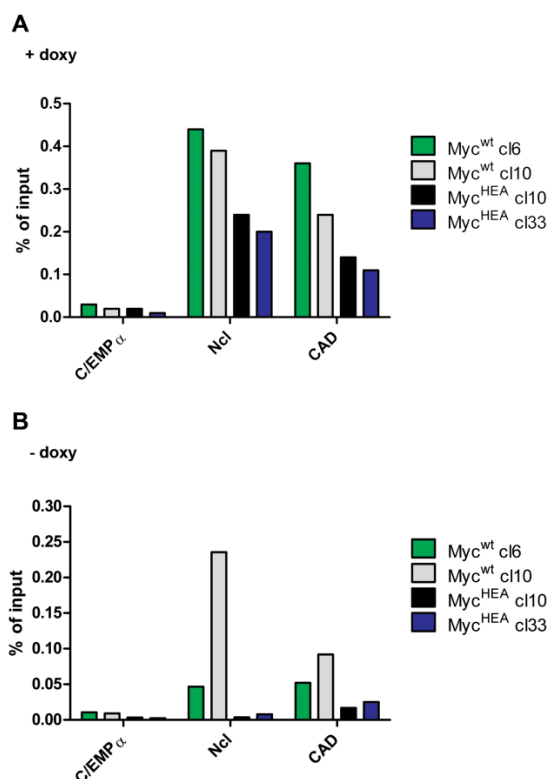
Figure 43. Myc binding to target promoters was impaired in Myc^{HEA}-expressing samples.
ChIP-qPCR of selected promoters bound by Myc (*Ncl* and *CAD*) and a negative control locus (*C/EMPα*), in cb9 Myc clones grown in presence (A) or in absence of doxycycline (B).

All the samples were then subjected to sequencing. Among the samples grown in presence of doxycycline (Figure 44A) we observed a decrease in the number of peaks in the two clones expressing Myc^{HEA} compared to the two Myc^{wt} samples, maybe as consequence of the dominant negative activity the mutant have already shown in the cb9 MycΔb cellular system. Nevertheless the number of binding sites recovered spanned from 17,000 to 30,000, and 24h after doxycycline withdrawal all the samples showed a decrease in peak

numbers: the clones expressing Myc^wt still had around 10,000 peaks, while in the samples expressing Myc^HEA the number of binding sites was dramatically reduced and the few residual peaks were mainly located in promoter regions (Figure 44B).
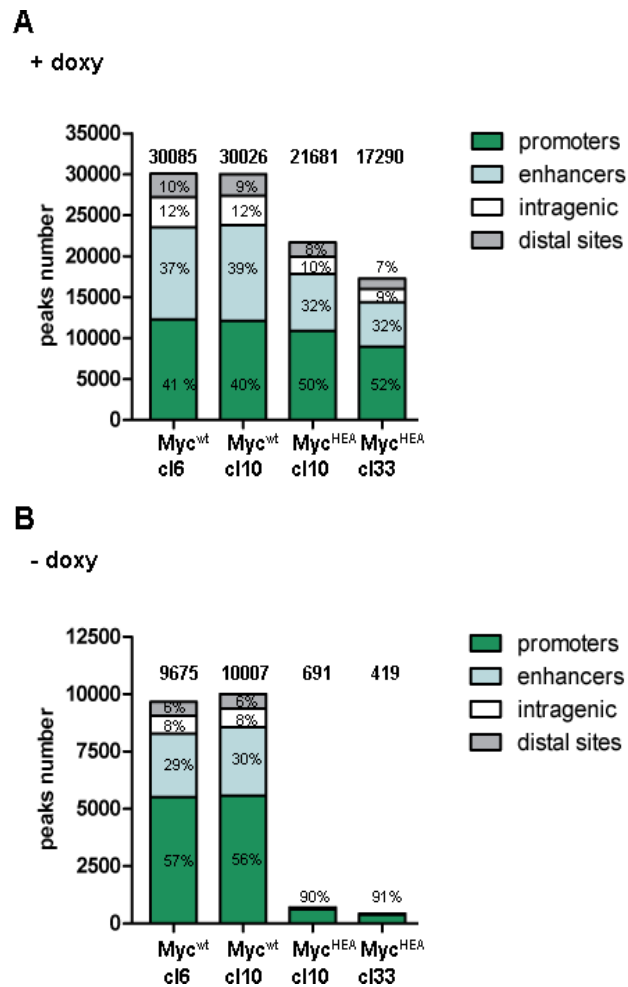
Figure 44. Myc ChIP-Seq peaks number and genomic distribution.
Number of binding sites called in the ChIP-Seq experiment and their annotation at promoters, enhancers, intragenic sites and distal regions. (A) Samples grown in presence of doxycycline were considered as the positive controls. (B) Cells fixed 24h after doxycycline withdrawal.

More than 80% of the peaks retrieved when Myc (wt or HEA) was expressed from the endogenous loci were a subset of the binding sites recovered when it was overexpressed from the CMV promoter, i.e in the cb9 MycΔb dataset (Figure 45). The direct comparison among the clones is shown in the blue box in Figure 45, revealing that the clones expressing Myc^wt contained almost all the peaks of the Myc^HEA samples. Accordingly, the

heatmap representing the union of all the peaks of the different clones (without doxycycline) on chromosome 1 (Figure 46) showed that the samples expressing Myc^{wt} were comparable in the intensity and hierarchy of peaks, while the clones expressing the mutant conserved only very few peaks.
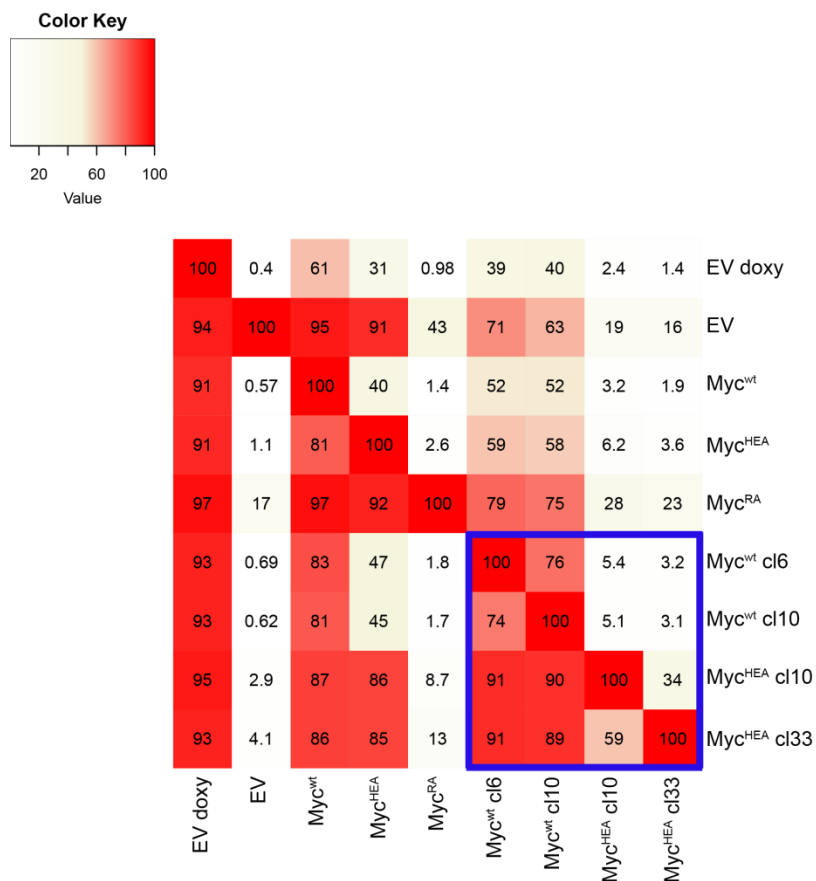
Figure 45. Myc-ChIP peaks overlap.
Percentages of Myc peaks overlapping (at least by 1 bp) among all dataset we generated (cb9 MycΔb CMV-driven Myc expression and the cb9 clones) are shown in each column. The overlap among the four cb9 clones is highlighted in the blue box.
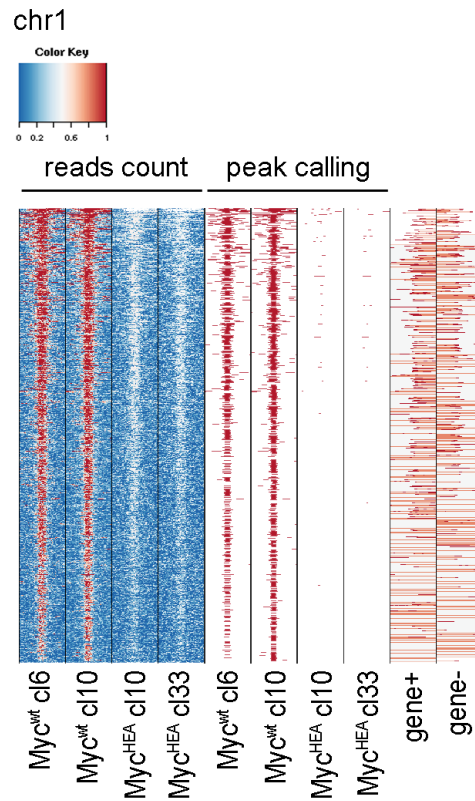
Figure 46. Intensity and distribution of Myc peaks on chromosome 1 in cb9 cellular clones.
Heatmap showing the library size-normalized ChIP-Seq reads count on chromosome 1 at regions bound in at least one clone sample. The first four columns show the reads coverage for the clones expressing Myc^{wt} and Myc^{HEA}, respectively. In columns five to eight the peaks calling from the same samples is shown. In the last two columns the presence of a gene in the sense (gene +) or antisense (gene -) DNA strand.

We then analyzed the fraction of peaks containing E-boxes in each sample (Figure 47). In these settings, the differences between Myc^{wt} and Myc^{HEA} were minimal; we observed a decrease in the non-canonical E-box peaks in favor of sequence-independent binding events in the samples expressing the mutant, while the fraction of peaks with the canonical E-box remained constant or even slightly increased.
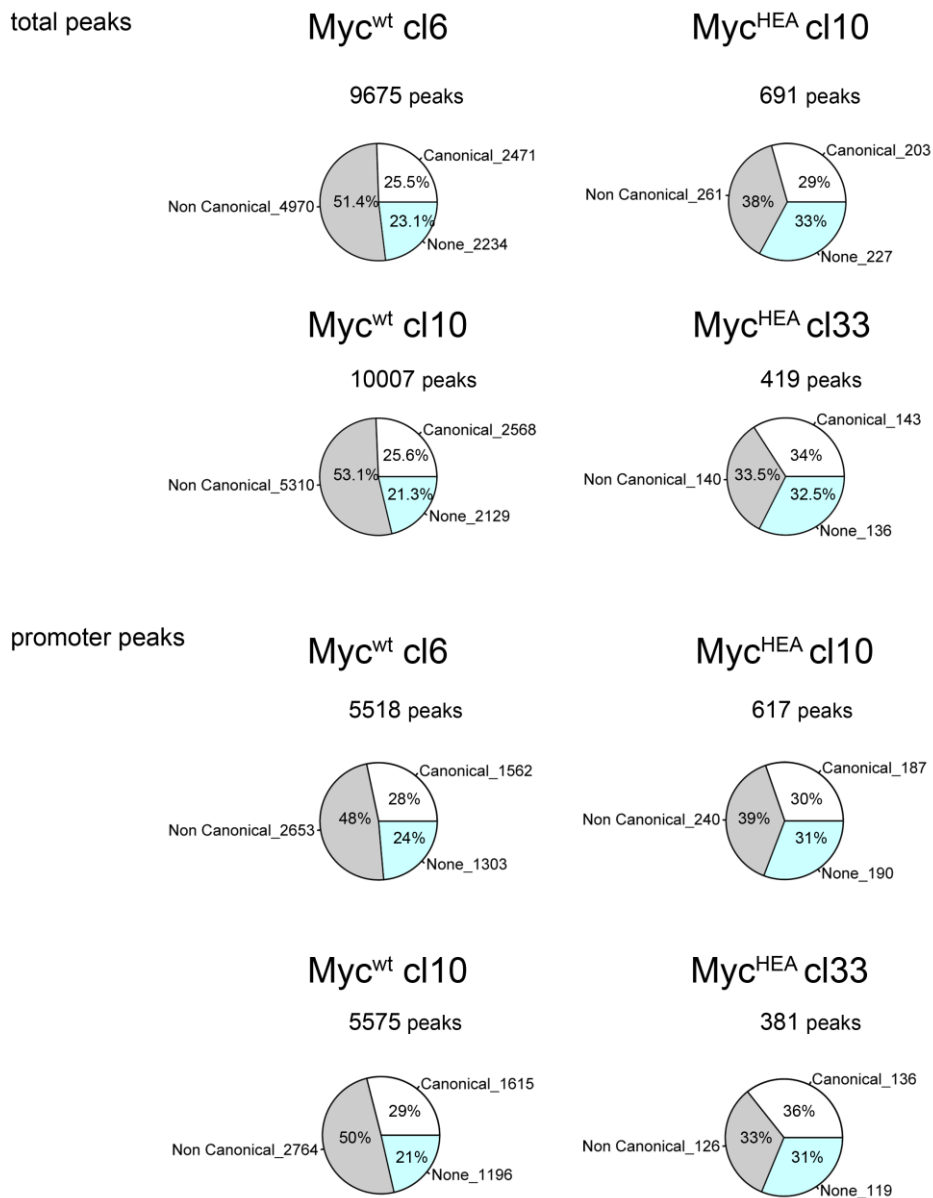
Figure 47. Fractions of Myc binding sites containing the canonical or the non-canonical E-boxes.
The pie-charts show the percentages of total and promoter peaks in cb9 clones expressing either Myc[wt] or Myc[HEA] that contain (in the region ±200 from the summit) the canonical E-box (CACGTG, in white), at least one of the four non-canonical E-boxes (CACGCG, CATGCG, CACGAG, CATGTG, gray) or none of them (light blue).

These results were confirmed by the analysis of the significance of the E-box fraction over the background. As before, we tested the differences among the samples and the genomic background, defined as the open promoters, and in this data set both the Myc[wt]- and the Myc[HEA]-expressing samples showed the same degree of significance in the enrichment for

the canonical E-box (p value<1E-15, Figure 48A), while the enrichment of non-canonical sequences was still higher in the Myc^{wt} samples than in the mutant ones (Figure 48B).
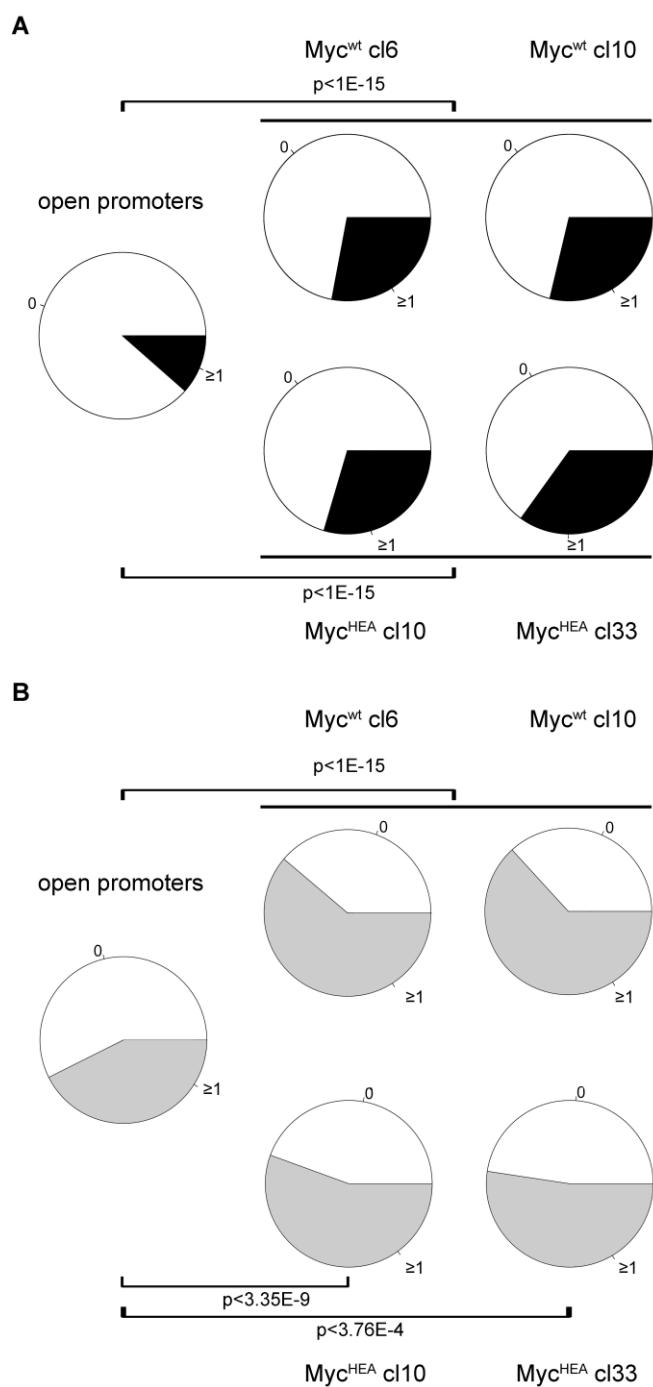
Figure 48. Significance of the E-boxes fractions in the cb9 clones expressing Myc^{wt} or Myc^{HEA}.
The promoter bound by the cb9 expressing Myc^{wt} or Myc^{HEA} have been analyzed for the presence or absence (in the window of ±200 nt from the peak summit), of at least one canonical E-box (A) or a non-canonical sequence (B). The difference between the samples and the fibroblasts active promoters (-400, o from the TSS) was tested with the χ^2 test and the p values are shown.

Nevertheless, by plotting the enrichment of the different categories of peaks (containing a canonical E-box, a non-canonical E-box or none), we noticed again that while the Myc^{wt} bound was stronger in the regions containing a canonical E-box or, to a lesser extent, a non-canonical sequence, Myc^{HEA} binding intensity was completely sequence-independent (Figure 49), suggesting the E-box found under the Myc^{HEA} peaks were not actually bound by the protein but may be merely present in the CpG islands Myc usually is associated to.
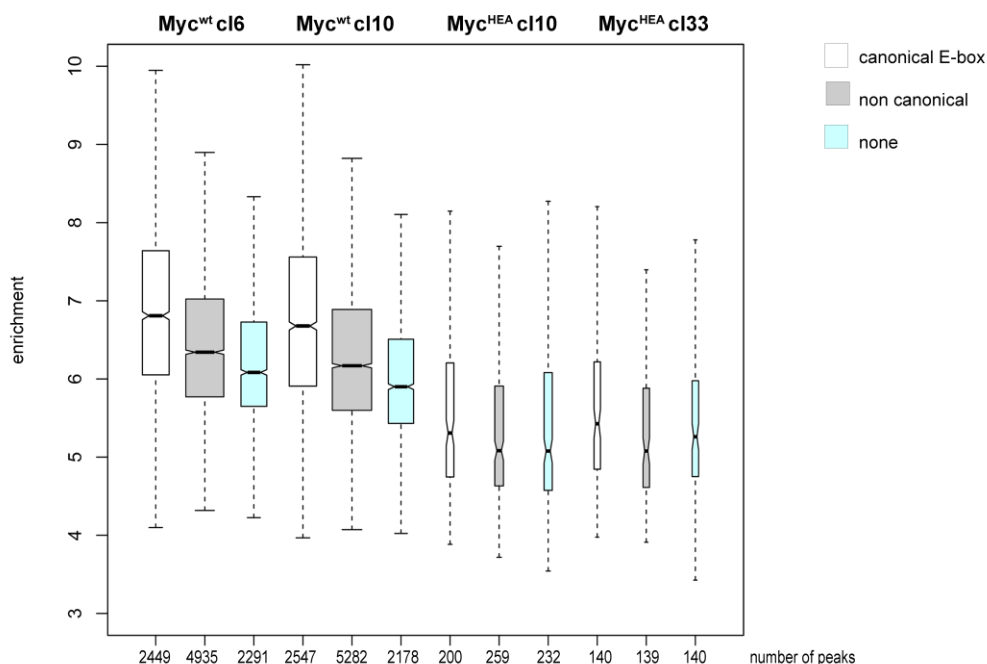
Figure 49. Enrichment analysis of Myc-ChIP peaks in cb9 clones expressing Myc^{wt} or Myc^{HEA}.
The enrichment of all the peaks of each cellular clone is shown as boxplot. At the bottom of each boxplot the number of peaks is reported. The peaks are categorized according to the presence of the canonical E-box (white), non-canonical E-boxes (gray) or none (light blue).

Moreover, analyzing the distribution of the E-boxes under the summit of the peaks we found that, similarly to the experiments in which we immunoprecipitated the overexpressed proteins, the E-boxes present under the Myc^{HEA} peaks were not localized under the summit (Figure 50). Altogether, our data suggest that the Myc^{HEA} mutant when expressed at endogenous levels is almost completely impaired in DNA binding, while when overexpressed it can still interact with a large number of open regions, most probably in non-productive manner.
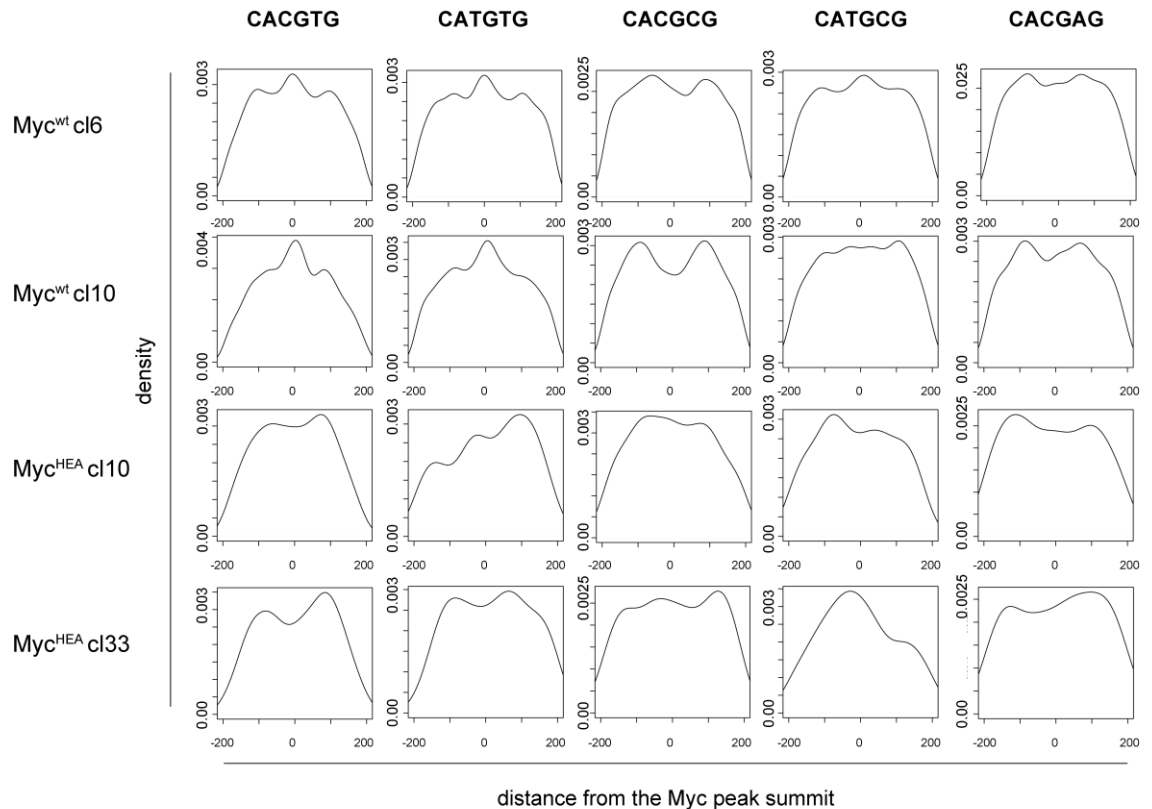
Figure 50. E-box distribution under Myc peaks in the cb9 clones.
Distribution of the distance from the peak summits of the canonical E-box (CACGTG) and each of the non-canonical sequence (CATGTG, CACGCG, CATGCG and CACGAG) in the Myc^wt and Myc^HEA clones.

3.7 Myc^HEA is impaired in activating gene expression

To investigate the transcriptional activity of our Myc mutants, we selected several known Myc-activated genes with promoters containing either canonical E-boxex (*smpdl3b* and *pus7)* or non-canonical E-boxes (CATGTG, *reep6*) and analyzed their transcriptional response in cb9 MycΔb cells overexpressing Myc^wt, Myc^HEA or Myc^RA (Figure 51). The expression of all loci was induced by Myc^wt (both tet-Myc and CMV-driven Myc) and dramatically decreased in the negative EV control. Myc^HEA instead, was not able to activate the expression of those genes, even if it was physically present on their promoter, as shown by the genome browser tracks of the ChIP-Seq experiment (Figure 51, right); a similar impairment in triggering gene expression was observed also in the Myc^RA sample but in this case its occupancy on the genomic regions was minimal.
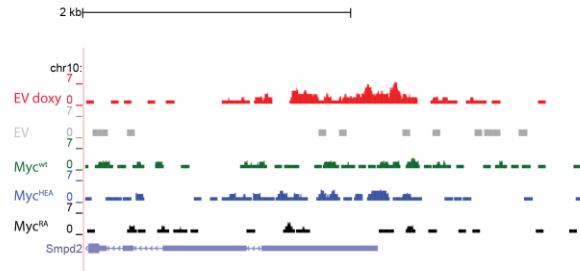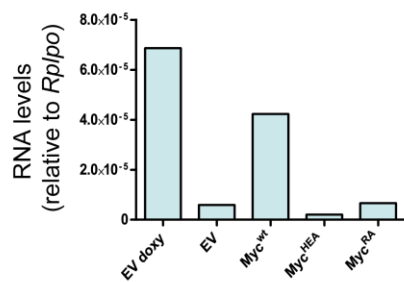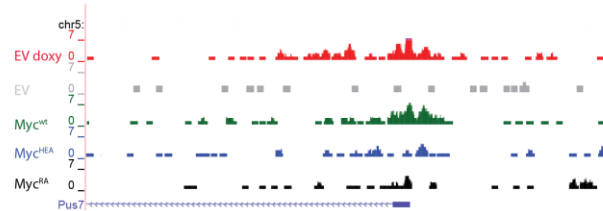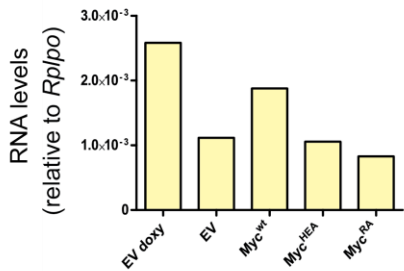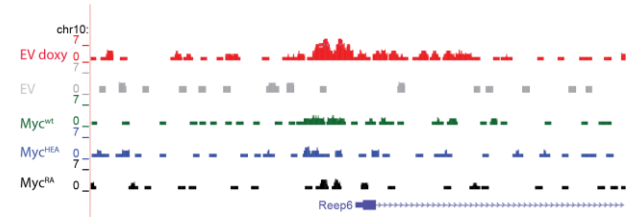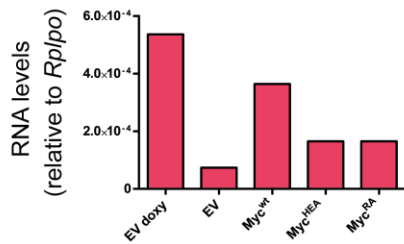
Figure 51. Expression analysis of some Myc target genes in cb9 MycΔb cells constitutively expressing Myc^wt or mutants.

On the left RNA expression levels of two Myc-target promoters containing the canonical E-box (*smpdl3b* and *pus7*) and one containing the non-canonical E-box CATGCG (*reep6*) are shown. On the right the Genome Browser tracks of the same targets in which are represented, in order from the top to the bottom: tet-Myc expressing cells (EV doxy), the cells infected with the empty vector (EV) and the cells expressing Myc^wt, Myc^HEA and Myc^RA.

If this failure in gene regulation was extended to all Myc^HEA bound promoters, it would justify the growth impairment of the Myc^HEA-expressing cells. An alternative, not mutually exclusive, explanation could be that among the genes mostly affected by the Myc^HEA DNA-binding defect there were key factors for the cell cycle progression. We already showed that the binding to the E-box-containing promoters was the most compromised in

the Myc[HEA] mutant, so we applied a gene set enrichment analysis (GSEA) to this class of genes. The sets of genes identified are listed according to their p-value in Figure 52 and contains fundamental biological processes, such as metabolism of amino acids and proteins, RNA metabolism and translation, nucleotides and carbohydrates metabolism (Figure 52, blue stars). Other interesting categories were G2M_CHECKPOINT and CELL_CYCLE (Figure 52, red stars), which included *cdk4*, *cdk7* and *cyclin D1* and *E2* genes. Failure in the regulation of one or more of these gene categories would led to cell cycle arrest and could explained the biological impairment of the Myc[HEA] expressing cells.
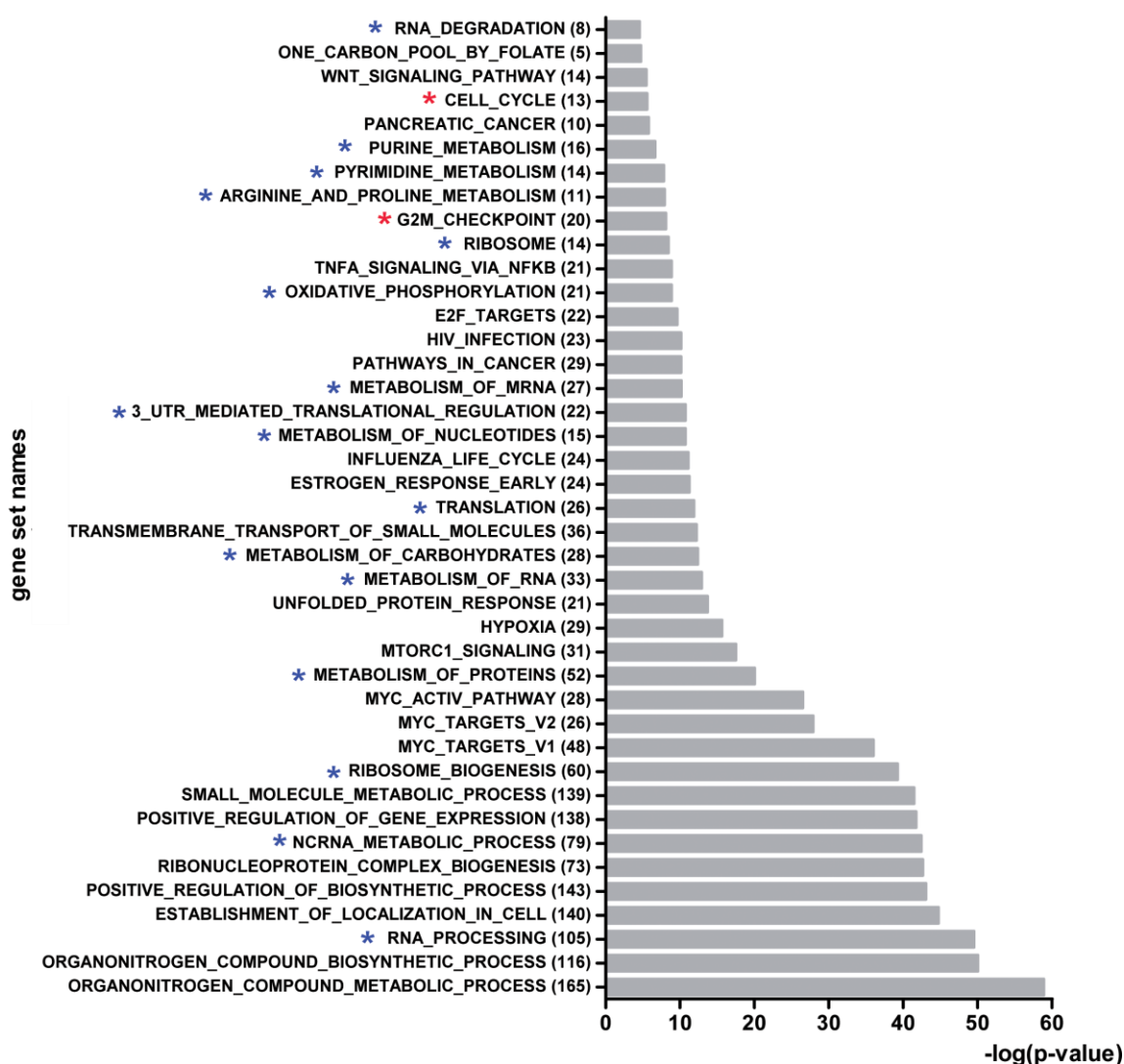
Figure 52. Gene set categories enriched in Myc^wt and Myc^HEA common E-box containing peaks.
The gene set categories are listed according to the -(log₁₀(p-value)), for each category the number of genes overlapping between our samples and the gene sets is shown in brackets. The blue stars highlight the metabolic processes of protein, nucleotides and carbohydrates, in red instead the cell cycle-related processes.

In order to complete the picture of the effects of the impairment in the DNA binding ability, we are currently investigating by RNA-Seq the transcriptomes of cb9 MycΔb fibroblasts overexpressing Myc^{wt}, Myc^{HEA} or Myc^{RA}. This analysis will give us the opportunity to better understand the controversial connection between Myc genome binding and its transcriptional activity.

Moreover, we are planning to address the transcriptional responses to the Myc mutants in another cellular system: the 3T9 mouse fibroblasts infected with retroviral vectors expressing the protein MycER^{wt}, MycER^{HEA} or MycER^{RA} (Figure 53). MycER is a well characterized fusion protein between Myc and the ligand-binding domain of a mutant estrogen receptor (ER)[266]. The ER domain lacks a transcriptional activity but responds to the synthetic steroid 4-hydroxytamoxifen (OHT) translocating the MycER protein from the cytoplasm into the nucleus.
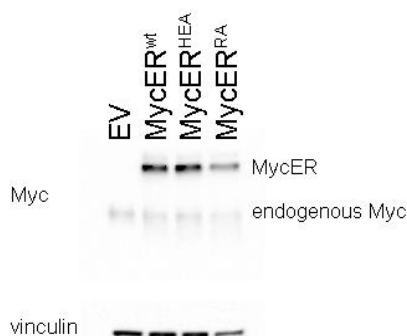
Figure 53. MycER^{wt} and mutants protein levels in 3T9 fibroblasts.
Protein levels of the fusion protein MycER^{wt} and mutants and of the endogenous mouse Myc expressed in the 3T9 fibroblasts.

An advantage of this setting is that the growth of the cells does not depend on exogenous Myc; as shown in the growth curve experiment and the CFA (Figure 54 and Figure55) all the samples grew in a comparable way in absence of OHT (of notice, also in this cellular system the overexpression of Myc^{HEA} in presence of Myc^{wt} is detrimental for the cells).

This will allow the characterization of the short-term response to the MycER activation and also could obviate transcriptional differences due to secondary effects.
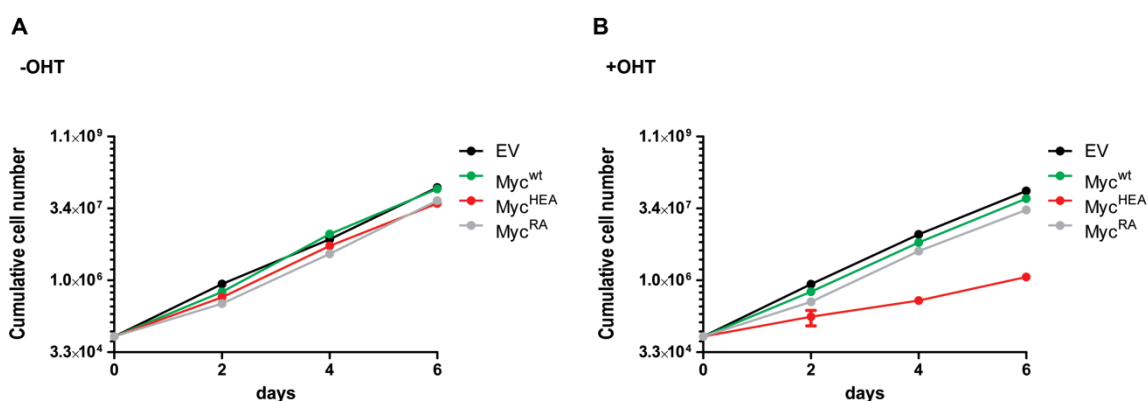
Figure 54. Proliferative ability of 3T9 cells expressing MycER^{wt}, MycER^{HEA} or MycER^{RA}.
70,000 3T9 cells were plated at day 0 in absence (A) or in presence (B) of 400 nM OHT. The cells were kept in culture and counted every two days until day 6.
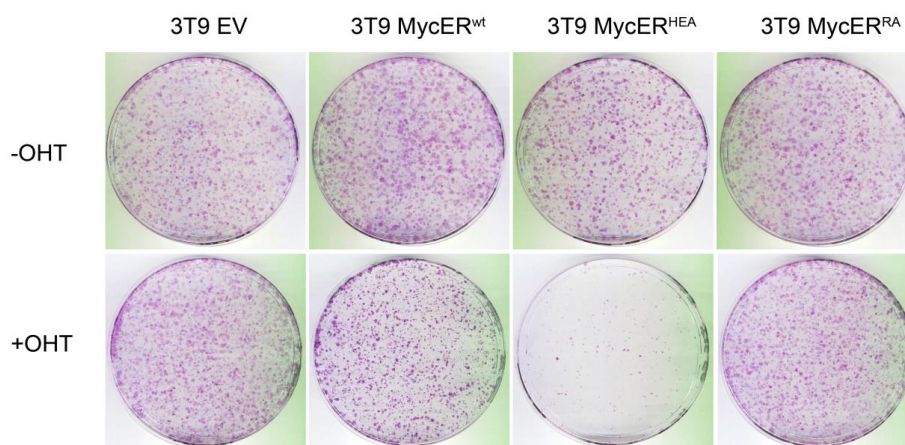
Figure 55. Colony forming potential of 3T9 cells expressing MycER^{wt} or mutants.
We plated 10,000 cells in presence or absence of OHT and the cells were stained with crystal violet ad day 6.

At the moment, we have checked the responses of some Myc target genes with promoters containing the E-box (CACGTG *smpdl3b,* CATGTG *reep6* and *rrp9,* Figure 56*)*. Myc^{HEA} completely failed in triggering the activation of *smpdl3b* and *rrp9,* similarly to the Myc^{RA} sample, while it was able to induce *reep6* expression, showing a residual transcriptional activity, even though less pronounced that the wt protein.
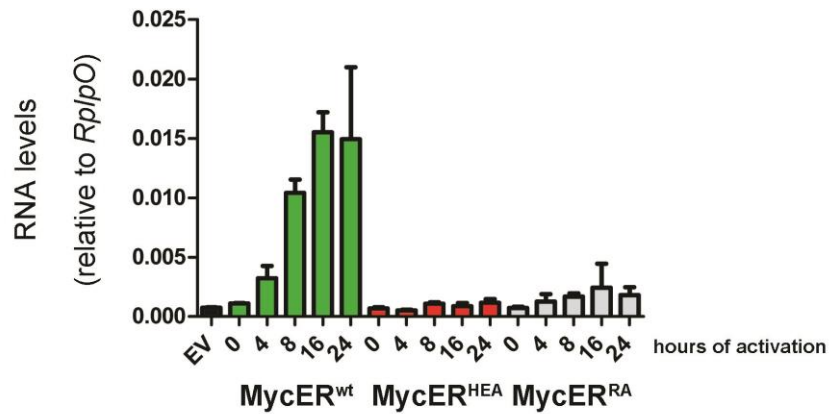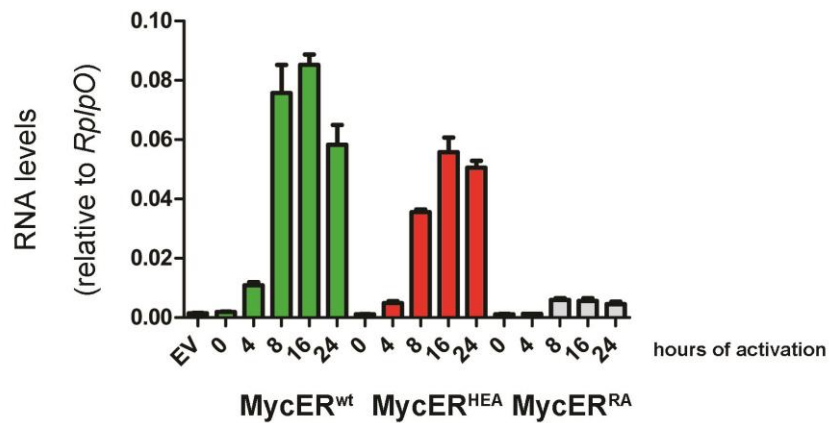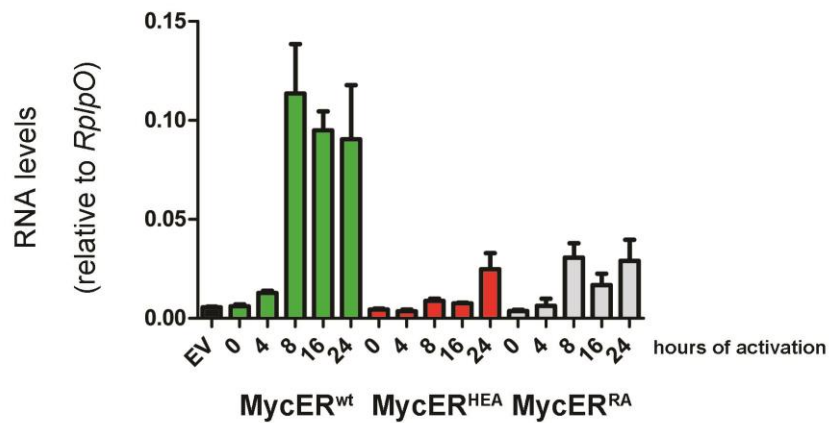
Figure 56. Activation of some Myc-dependent genes upon OHT addition in 3T9 MycER cells.
Some Myc target genes were analyzed for their expression at 0h, 4h, 8h, 16h and 24h after MycER[wt] (in green), MycER[HEA] (in red) or MycER[RA] (in gray) activation. In black the expression level of the 3T9 cells infected with the EV.

Altogether the data presented in this thesis consistently show the fundamental role of the E-box recognition for Myc biological activity. In all systems analyzed the cells expressing Myc^{HEA} were unable to sustain the cellular growth, similarly to the Myc^{RA} mutant.

Genomic analysis revealed that the biological impairment was not coupled to the complete absence of the Myc^{HEA} mutant protein from the chromatin, but the binding to a conspicuous subset of Myc^{wt} targets in the cb9 MycΔb cells was not enough to ensure cellular proliferation. Our hypothesis that the Myc^{HEA} binding events were not coupled to the transcriptional regulation of the bound targets was partially confirmed by the expression analysis by qPCR of few Myc targets, proving the necessity of a deeper characterization of the transcriptional events depending on Myc^{HEA} and Myc^{RA}, which is currently ongoing.

4. Discussion

4.1 Mutations in the Myc DNA binding domain impair its ability to sustain cellular growth

Myc, as all the members of the bHLH protein family, cannot bind DNA as a monomer but needs to dimerize with its partner Max to form a tertiary structure that interacts with the DNA helix. This binding is believed to occur in two steps: a sequence-independent contact with the DNA backbone and a specific interaction relying on the E-box sequence (CACGTG or variants thereof). Based on the crystal structure of the Myc/Max dimer, Myc residues involved in the recognition of the constant part of the consensus E-box (<u>CA</u>NN<u>TG</u>) are Histidine 359 and Glutamic acid 363, while Arginine 367 interacts with the G of the central dinucleotide CpG[166]. Our re-analysis of the structure (PDB ID:1NKP)[166] confirmed that both the Histidine and the Glutamic residues make contacts with the E-box sequence, while Arginine 367, beside interacting with the E-box core, could mediate the generic binding to the DNA backbone, together with two other Arginine residues (R366 and R367). Our observations were in agreement with the knowledge that Arginine to Alanine substitutions of Myc R364, R366 and R367 residues completely abolished Myc-induced transformation and its ability to transactivate a reporter gene [159]. We took advantage of this already published mutant (Myc^{RA}) to investigate the effects of the general DNA binding disruption on Myc functions. Instead, to evaluate the impact of the impairment in the other mode of DNA interaction, we generated the Myc^{HEA} mutant, in which the residues involved in sequence specific recognition, H359 and E363, were mutated into Alanine.

In an *in vitro* assay both Myc^{HEA} and Myc^{RA} mutants showed an impairment in DNA binding, at different extent according to their original design. Since both the mutants

retained a residual binding ability when compared to a negative control, we are planning to assess the specificity of Myc^{HEA} and Myc^{RA} binding to the E-box probe by competition binding assays.

These mutant proteins were used to further investigate the role of the two types of DNA interactions in Myc biology and functions.

We characterized the ability of the Myc mutants to sustain and/or promote cellular proliferation in different cellular models: rat fibroblasts knock out for the *c-myc* gene (rat HO15.19) and mouse fibroblasts in which we engineered the endogenous c-*myc* alleles in order to produce a Myc protein devoid of the basic region (cb9 MycΔb).

In the first system we used, the rat HO15.19 cells, overexpression of both Myc^{RA} and Myc^{HEA} mutants produced a slight advantage in proliferation (as assessed by growth curves) if compared to the cells infected with the empty vector. These cells were produced by targeted homologous recombination that disrupted both the endogenous c-*myc* loci and are characterized by a lengthened cell cycle and growth impairment[259]. In 2007 Cowling and Cole used this cellular system to investigate Myc biological activity which did not depend on the direct DNA binding[260] by taking advantage of two Myc mutants, one completely lacking the C-terminal bHLH-LZ domain and the other with the residues 364-367 of the basic region, RQRR, mutated into ADAA. Of notice, this second mutant is quite similar to our Myc^{RA} mutant and its expression in the HO15.19 cells partially rescued the proliferative defects of those cells, as we also observed for the Myc^{RA} and Myc^{HEA} mutants. Surprisingly, also the mutant devoid of the entire C-terminal domain (and thus completely unable to bind the DNA) showed similar rescue ability. Given this data, we interpreted the growth advantage of the HO15.19 cells expressing Myc^{RA} or Myc^{HEA} as an effect of DNA binding-unrelated functions of Myc in a cellular system which is adapted to grow without Myc at all.

Indeed, in the mouse cb9 MycΔb cells, whose growth was dependent on the presence of Myc, both Myc^{HEA} and Myc^{RA} were totally unable to sustain cellular proliferation when

wild-type Myc was inactivated, demonstrating that both kinds of Myc-DNA interactions (general backbone recognition and sequence-driven binding) are fundamental for Myc functions.

4.2 E-box recognition is required for stabilization of Myc binding to DNA

Despite a similar phenotype, Myc[HEA] and Myc[RA] mutants were dramatically different in their genomic distribution: the number of Myc binding sites retrieved after a ChIP-Seq experiment in cells overexpressing Myc[HEA] was around 8,000 (half respect to Myc[wt]), while Myc[RA] showed less than 300 peaks, consistent with the fact that mutations of the residues involved in the interaction with the DNA backbone completely disrupted the binding ability of the protein. In depth analysis of Myc[HEA] binding sites revealed that half of them still contained an E-box (canonical or non-canonical), but those sites were not bound at higher enrichment relative to regions devoid of binding motifs, as instead observed for Myc[wt], in our own experiments as well as in several others[52,90,219,222,223]. These data suggested that the E-boxes found under the Myc[HEA] peaks were actually not specifically recognized by the mutant protein. This hypothesis was confirmed by the fact that while the E-box was exactly positioned under the peak summits in the Myc[wt] ChIP-seq, it was more de-localized in the Myc[HEA] sample. We thus concluded that the Myc[HEA] protein was not able to discriminate and preferentially bind the E-box compared to any other sequence, validating the rationale followed for the generation of this mutant. The residual enrichment of the E-boxes observed in the Myc[HEA] bound sites may be partially due to the protein association with wild-type Max that can recognize half of the palindromic sequence. Even more important, the E-box sequence is a CpG-containing motif and CpG islands were one of the first genomic features to be described as major determinant for Myc binding *in vivo*[219,220].

The above results were confirmed in another cellular system, the cb9 clones, in which the Myc[HEA] mutant was not overexpressed, but transcribed from the endogenous loci. In this case, the number of ChIP-seq peaks retained by the Myc[HEA] samples was drastically reduced compared to the Myc[wt] samples: indeed, only few hundreds high affinity peaks were maintained.

Of the two datasets we generated for the analysis of the Myc[HEA] mutant, the first one was characterized by the overexpression of the protein (cb9 MycΔb cells), while the second one allowed the investigation of the DNA binding capacity of the endogenously expressed protein (cb9 clones). Even if the phenotypic characterization in both cellular systems showed a similar impairment of the Myc[HEA] mutant in sustaining cellular growth, at the genomic level the differences were striking: when Myc[HEA] was expressed at endogenous levels, it retained hundred peaks, which corresponded to the top targets of the Myc[wt] samples; when it was overexpressed instead, it still bound half of the Myc[wt] binding sites. We interpreted the difference in the DNA binding profiles as the effect of the so called chromatin invasion capacity of Myc: when overexpressed, Myc can be crosslinked to virtually all active chromatin[38,90,222,223], most probably thanks to the non-specific DNA binding events and the protein/protein interactions that the Myc[HEA] protein still retained. The sequence-recognition impairment of the Myc[HEA] mutant led to a unique situation in which on one hand the DNA binding was widespread along the open chromatin, but on the other hand the binding hierarchy was completely lost, as the high affinity sites containing the E-box were bound with the same strength as the sites without any target sequences.

4.3   Myc binding to chromatin is not predictive of gene regulation

Given the ability of the overexpressed Myc[HEA] mutant to invade the open chromatin we had to face another issue: why did the HEA and RA mutations have a similar effect on the capacity of Myc to sustain cellular proliferation, given the differences observed in their

DNA binding capacity at the genome-wide level? In fact, while the Myc[RA] mutant was almost completely impaired as DNA binding protein, the Myc[HEA] protein, at least when overexpressed, still had around 8,000 binding sites, many of them localized in promoter regions. The most conservative explanation would be that, even if bound to DNA, this mutant was actually transcriptionally inactive. Indeed, we verified that this was the case, at least at selected Myc target genes, and we are now planning transcriptomic analysis at the genome-wide level by RNA-Seq to obtain a more complete picture. If this would be confirmed, it will reinforce the notion that Myc DNA binding, even at promoters, is not predictive of gene regulation, which implies that, although often co-occurring, Myc invasion and RNA amplification are functionally independent phenomena[38,52,90,267].

4.4 Myc genome recognition *in vivo*

As already described, when Myc is overexpressed it invades all the open chromatin regions, maintaining nevertheless the binding hierarchy between high affinity and low affinity sites. Recently, from the analysis of genome-wide data, a debate emerged regarding the recruitment of Myc to chromatin and the actual relevance of sequence-dependent binding. Our group proposed a stepwise model for Myc/Max recruitment to its high affinity targets, illustrated in Figure 57[224]: we hypothesized an initial protein-protein interaction with chromatin-associated factors, which does not require any direct DNA contacts and would already restrict Myc binding to a subset of the potential targets in the whole genome. This first event would be followed by sequence-independent engagement of Myc/Max onto DNA, allowing sliding of the dimer along the DNA until encountering a high affinity site (E-box) to which it binds in a more stable way. This stepwise binding model found a confirmation in the analysis of the genome-wide profiles of Myc[HEA] and Myc[RA] mutants. The few peaks retained by the Myc[RA] protein in our ChIP-seq experiment were consistent with the notion that this mutant was designed to be completely impaired in

DNA binding, so it could undertake only the first mode of interaction (Figure 57B); since the crosslink with formaldehyde does not efficiently stabilize the protein-protein interactions, the number of binding sites obtained was comparable to the negative control. The Myc^{HEA} mutant instead was designed to be able to interact with the DNA, but to fail in the third mode of interaction, the sequence-specific binding (Figure 57 A). Myc^{HEA} protein was recruited to the chromatin by protein-protein interactions, as Myc^{RA}, and once in proximity to the DNA it engaged a non-specific interaction which allowed the dimer sliding along the DNA in search of high affinity targets. Of notice, even in those Myc^{HEA} peaks which actually contained an E-box, or a non-canonical E-box, the target sequences were not found under the peak summit (as in the Myc^{wt} sample) but were widespread around, indicating that binding of the Myc^{HEA} protein was not stabilized by the E-box.

The differences in the DNA binding mechanisms of the Myc^{HEA} and Myc^{RA} mutants led also to an interesting condition: Myc^{HEA}, when overexpressed in presence of Myc^{wt} (both endogenous or expressed from a transgene) acted as dominant negative protein and we hypothesized that this phenotype, which was not observed with the Myc^{RA} mutant, may be intrinsic of the Myc^{HEA} ability to bind DNA. It is known that Myc/Max dimers are stabilized through the interaction with DNA, a feature preserved in Myc^{HEA}/Max but not in Myc^{RA}/Max dimers: consequently only Myc^{HEA} would efficiently sequester Max in transcriptionally inactive complexes.

Interestingly, substitutions of the E residue in human bHLH proteins Twist1 and Twist2 have been found associated with different craniofacial disorders and the characterization of such mutations in *Caenorhabditis elegans* revealed that both the mutated proteins retained DNA-binding ability but acted as dominant negative, thus affecting gene expression[268–270]. As all the bHLH proteins, Twist1 and Twist2 have to bind the DNA as obligate dimers and the authors proposed a model in which the mutated proteins titrate the binding partner, similarly to how we hypothesized Myc^{HEA} could sequester Max.
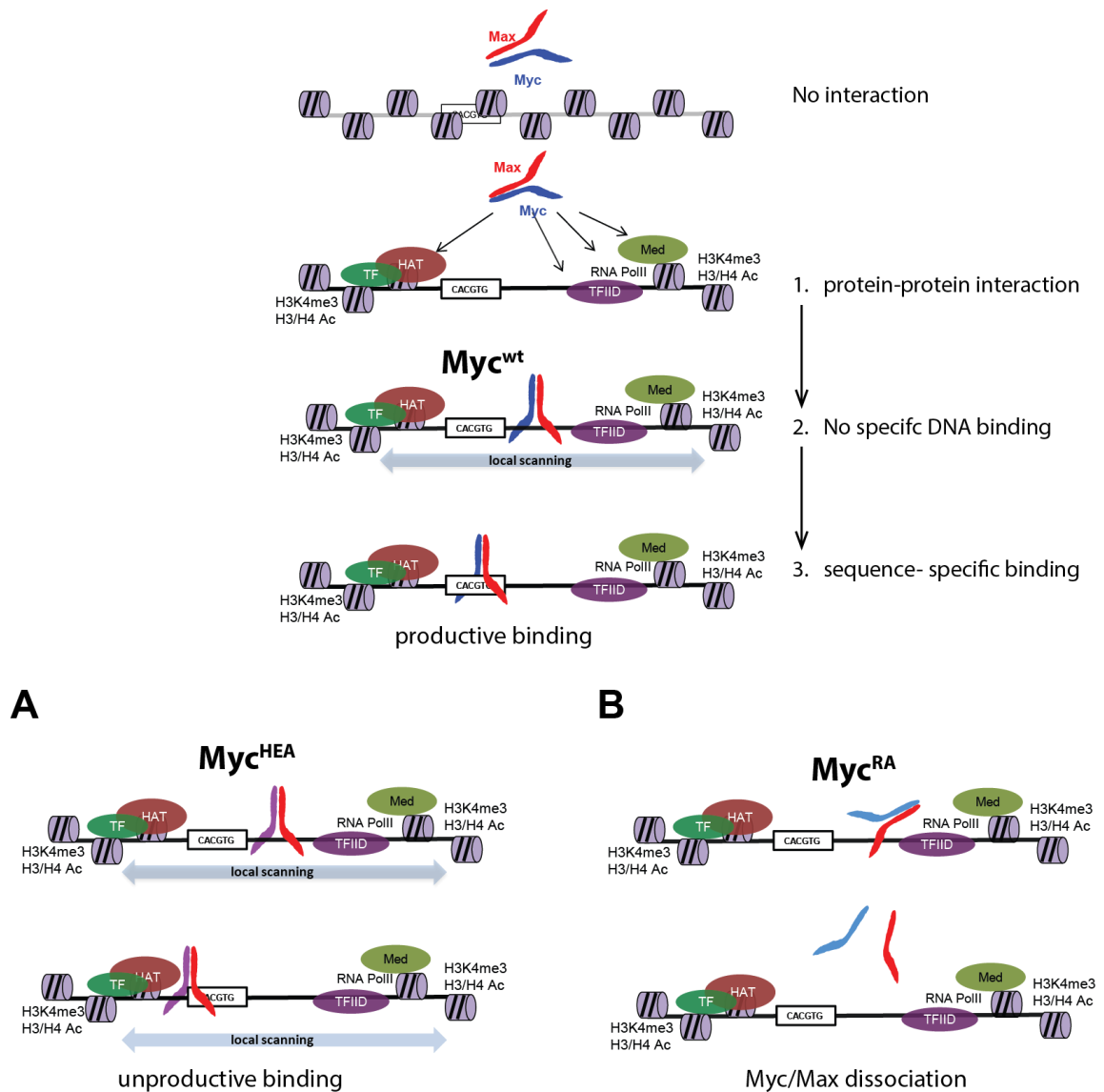
Figure 57. *In vivo* **Myc/Max DNA binding model.**

This model described Myc/Max binding to DNA as a stepwise process; first the dimer is recruited to chromatin by protein-protein interactions, without a direct DNA contact (mode1). This initial event is followed by the direct interaction with the DNA in a sequence independent way (mode 2) and the low affinity interaction between Myc/Max basic regions and the DNA backbone allows the dimer to move along the DNA scanning for an E-box, giving rise at last to a transcriptional productive binding (mode 3).

The two mutants are impaired in different modes of interaction. (A) Myc[HEA], in purple, can bind the DNA and moving along it but is not able to recognize the E-box sequence (transition from mode 2 to 3); nevertheless the Myc/Max dimer is stabilized by the generic interaction with the DNA. (B) Instead Myc[RA], in light blue, is unable to go from mode 1 to mode 2 and this impairment leads to the dissociation of the dimer.

Altogether, the data presented in this thesis not only confirmed our model of Myc-chromatin interactions, but also suggested that Myc unspecific binding to DNA is not sufficient for the transcriptional regulation since the sequence-specific DNA binding is fundamental to trigger the gene expression.

References

1. Dang, C. MYC on the Path to Cancer Chi. *Cell* **149,** 22–35 (2013).
2. Dalla-Favera, R. *et al.* Human c-myc one gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells. *PNAS* **79,** 7824–7827 (1982).
3. Subar, M., Neri, A., Inghirami, G., Knowles, D. & Dalia-Favera, R. Frequent c-myc Oncogene Activation and Infrequent Presence of Epstein-Barr Virus Genome in AIDS-Associated Lymphomas. *Blood* **72,** 667–671 (1988).
4. Little, C., Nau, M., Carney, D., Gazdar, A. & Minna, J. Amplification and expression of c-myc oncogene in human lung cancer cell lines. *Nature* **306,** 194–196 (1983).
5. Berns, E. M. *et al.* TP53 and MYC gene alterations independently predict poor prognosis in breast cancer patients. *Genes. Chromosomes Cancer* **16,** 170–179 (1996).
6. Amati, B., Alevizopoulos, K. & Vlach, J. Myc and the cell cycle. *Front. Biosci.* **3,** 250–268 (1998).
7. Peyton, R. A Sarcoma of the Fowl Transmissible by an Agent. *J. Exp. Med.* **13,** 397–411 (1911).
8. Ivanov, X., Mladenov, Z., Nedyalkov, S., Todorov, T. G. & Yakimov, M. Experimental investigations into avian leucoses. V. Transmission, haematology and morphology of avian myelocytomatosis. *Bull. Inst. Pathol. Comp. Anim. Acad. Bulg. Sci* **10,** 5–38 (1964).
9. Bolognesi, D. P., Langlois, A. J., Sverak, L., Bonar, R. A. & Beard, J. W. In vitro chick embryo cell response to strain MC29 avian leukosis virus. *J. Virol.* **2,** 576–586 (1968).
10. Langlois, A. J. *et al.* Response of Bone Marrow to MC29 Avian Leukosis Virus in Vitro. *Cancer Res.* **29,** 2056–2074 (1969).
11. Graf, T. Two types of target cells for transformation with avian myelocytomatosis virus. *Virology* **54,** 398–413 (1973).
12. Sheiness, D., Fanshier, L. & Bishop, J. Identification of Nucleotide Sequences Which May Encode the Oncogenic Capacity of Avian Retrovirus MC29. *J. Virol.* **28,** 600–610 (1978).
13. Mellon, P., Pawsontf, A., Bister, K., Martint, S. S. & Duesberg, P. H. Specific RNA sequences and gene products of MC29 avian acute leukemia virus. *PNAS* **75,** 5874–5878 (1978).
14. Roussel, M. *et al.* Three new types of viral oncogene of cellular origin specific for haematopoietic cell transformation. *Nature* **281,** 452–455 (1979).
15. Sheiness, D. & Bishop, J. M. 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 (1979).
16. Stehelin, D., Varmus, H. E., Bishop, J. M. & Vogt, P. K. DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. *Nature* **260,** 643–645 (1976).
17. Vennstrom, B., Sheiness, D., Zabielski, J. & Bishop, J. M. 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 (1982).
18. Colby, W., Chen, E., Smith, D. & Levinson, A. Identification and nucleotide sequence of a human homologous to the v-myc oncogene of avian myelocytomatosis virus MC29. *Nature* **301,** 722–725 (1983).
19. Zimmerman, K. A. *et al.* Differential expression of myc family genes during murine development. *Nature* **319,** 780–783 (1986).
20. Bernard, O., Drago, J. & Sheng, H. L-myc and N-myc influence lineage determination in the central nervous system. *Neuron* **9,** 1217–1224 (1992).
21. Strieder, V. & Lutz, W. Regulation of N-myc expression in development and disease. *Cancer Lett.* **180,** 107–119 (2002).
22. Morrow, M. A., Lee, G., Gillis, S., Yancopoulos, G. D. & Alt, F. W. Interleukin-7 induces N-myc and c-myc expression in normal precursor B lymphocytes. *Genes Dev.* **6,** 61–70 (1992).
23. Mugrauer, G. & Ekblom, P. Contrasting Expression Patterns of Three Members of the myc

Family of Protooncogenes in the Developing and Adult Mouse Kidney. *J. Cell Biol.* **112,** 13–25 (1991).

24. Hatton, K. S. *et al.* Expression and activity of L-Myc in normal mouse development. *Mol. Cell. Biol.* **16,** 1794–804 (1996).

25. Hirvonen, H. *et al.* The N-myc proto-oncogene and IGF-II growth factor mRNAs are expressed by distinct cells in human fetal kidney and brain. *J. Cell Biol.* **108,** 1093–1104 (1989).

26. Stanton, B. R., Perkins, A. S., Tessarollo, L., Sassoon, D. A. & Parada, L. F. Loss of N-myc function results in embryonic lethality and failure of the epithelial component of the embryo to develop. *Genes Dev.* **6,** 2235–2247 (1992).

27. Charron, J. *et al.* Embryonic Lethality in Mice Homozygous for a Targeted Disruption of the N-Myc Gene. *Genes Dev.* **6,** 2248–2257 (1992).

28. Davis, A. C., Wims, M., Spotts, G. D., Hann, S. R. & Bradley, A. 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 (1993).

29. Malynn, B. A. *et al.* N- myc can functionally replace c- myc in murine development , cellular growth , and differentiation. *Genes Dev.* **14,** 1390–1399 (2000).

30. Land, H., Parada, L. F. & Weinberg, R. A. Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes. *Nature* **304,** 596–602 (1983).

31. Yancopoulos, G. D. *et al.* N-myc can cooperate with ras to transform normal cells in culture. *PNAS* **82,** 5455–5459 (1985).

32. DePinho, R. A., Hatton, K. S., Tesfaye, A., Yancopoulos, G. D. & Alt, F. W. The human myc gene family: structure and activity of L-myc and an L-myc pseudogene. *Genes Dev.* **1,** 1311–1326 (1987).

33. Nesbit, C. E., Tersak, J. M. & Prochownik, E. W. MYC oncogenes and human neoplastic disease. *Oncongene* **18,** 3004–3016 (1999).

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

35. Schwab, M., Alitalo, K. & Klemphauer, K. Amplified DNA with limited homology to myc cellular oncogene is shared by human neuroblastoma tumor. *Nature* **305,** 245–248 (1983).

36. Schwab, M. MYCN in neuronal tumours. *Cancer Lett.* **204,** 179–187 (2004).

37. Nau, M. M. *et al.* L-myc, a new myc-related gene amplified and expressed in human small cell lung cancer. *Nature* **318,** 69–73 (1985).

38. Kress, T. R., Sabò, A. & Amati, B. MYC: connecting selective transcriptional control to global RNA production. *Nat. Rev. Cancer* **15,** 593–607 (2015).

39. Perez-Roger, I., Kim, S., Griffith, B., Sewing, A. & Land, H. Cyclins D1 and D2 mediate Myc-induced proliferation via sequestration of p27(Kip1) and p21(Cip1). *EMBO J.* **18,** 5310–5320 (1999).

40. Yu, Q., Ciemerych, M. A. & Sicinski, P. Ras and Myc can drive oncogenic cell proliferation through individual D-cyclins. *Oncogene* **24,** 7114–7119 (2005).

41. Beier, R. *et al.* Induction of cyclin E-cdk2 kinase activity, E2F-dependent transcription and cell growth by Myc are genetically separable events. *EMBO J.* **19,** 5813–5823 (2000).

42. Pérez-Roger, S. D., Sewing, A. & Land, H. Myc activation of cyclin E/Cdk2 kinase involves induction of cyclin E gene transcription and inhibition of p27Kip1 binding to newly formed complexes. *Oncogene* **14,** 2373 (1997).

43. Seoane, J. *et al.* TGFβ influences Myc, Miz-1 and Smad to control the CDK inhibitor p15INK4b. *Nat. Cell Biol.* **3,** 400–408 (2001).

44. Staller, P. *et al.* Repression of p15INK4b expression by Myc through association with Miz-1. *Nat. Cell Biol.* **3,** 392–399 (2001).

45. Yang, W. *et al.* Repression of transcription of the p27(Kip1) cyclin-dependent kinase inhibitor gene by c-Myc. *Oncogene* **20,** 1688–1702 (2001).

46. Gartel, A. L. *et al.* Myc represses the p21(WAF1/CIP1) promoter and interacts with Sp1/Sp3. *PNAS* **98,** 4510–4515 (2001).

47. Hennecke, S. & Amati, B. Cyclin E and c-Myc promote cell proliferation in the presence of p16 INK4a and of hypophosphorylated retinoblastoma family proteins. *EMBO J.* **16,** 5322–5333 (1997).

48. Vlach, J., Hennecke, S., Alevizopoulos, K., Conti, D. & Amati, B. Growth arrest by the cyclin-dependent kinase inhibitor p27Kip1 is abrogate by c-Myc. *EMBO J.* **15,** 6595–6604

(1996).

49. Marhin, W. W., Chen, S., Facchini, L. M., Fornace, A. J. & Penn, L. Z. Myc represses the growth arrest gene gadd45. *Oncogene* **14,** 2825–2834 (1997).

50. Lee, T. C., Li, L., Philipson, L. & Ziff, E. B. Myc represses transcription of the growth arrest gene gas1. *PNAS* **94,** 12886–12891 (1997).

51. Robinson, K., Asawachaicharn, N., Galloway, D. A. & Grandori, C. c-Myc accelerates S-phase and requires WRN to avoid replication stress. *PLoS One* **4,** (2009).

52. Perna, D. *et al.* Genome-wide mapping of Myc binding and gene regulation in serum-stimulated fibroblasts. *Oncogene* **31,** 1695–709 (2012).

53. Miltenberger, R. J., Sukow, K. A. & Farnham, P. J. An E-box-mediated increase in cad transcription at the G1/S-phase boundary is suppressed by inhibitory c-Myc mutants. *Mol Cell Biol* **15,** 2527–2535 (1995).

54. Fogal, V. *et al.* Mitochondrial p32 is upregulated in Myc expressing brain cancers and mediates glutamine addiction. *Oncotarget* **6,** 1157–1170 (2014).

55. Li, F. *et al.* Myc Stimulates Nuclearly Encoded Mitochondrial Genes and Mitochondrial Biogenesis. *Mol. Cell. Biol.* **25,** 6225–6234 (2005).

56. Zhang, H. *et al.* 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 (2007).

57. Kim, J., Gao, P., Liu, Y. C., Semenza, G. L. & Dang, C. V. Hypoxia-inducible factor 1 and dysregulated c-Myc cooperatively induce vascular endothelial growth factor and metabolic switches hexokinase 2 and pyruvate dehydrogenase kinase 1. *Mol. Cell. Biol.* **27,** 7381–93 (2007).

58. Shim, H. *et al.* c-Myc transactivation of LDH-A: implications for tumor metabolism and growth. *PNAS* **94,** 6658–63 (1997).

59. Nikiforov, M. A. *et al.* A functional screen for Myc-responsive genes reveals serine hydroxymethyltransferase, a major source of the one-carbon unit for cell metabolism. *Mol. Cell. Biol.* **22,** 5793–800 (2002).

60. Grandori, C. *et al.* c-Myc binds to human ribosomal DNA and stimulates transcription of rRNA genes by RNA polymerase I. *Nat. Cell Biol.* **7,** 311–318 (2005).

61. Arabi, A. *et al.* c-Myc associates with ribosomal DNA and activates RNA polymerase I transcription. *Nat. Cell Biol.* **7,** 303–310 (2005).

62. Gomez-Roman, N., Grandori, C., Eisenman, R. N. & White, R. J. Direct activation of RNA polymerase III transcription by c-Myc. *Nature* **421,** 1698–1701 (2003).

63. Jones, R. M. *et al.* An essential E box in the promoter of the gene encoding the mRNA cap-binding protein (eukaryotic initiation factor 4E) is a target for activation by c-myc. *Mol. Cell. Biol.* **16,** 4754–64 (1996).

64. Zindy, F. *et al.* Myc signaling via the ARF tumor suppressor regulates p53-dependent apoptosis and immortalization. *Genes Dev.* **12,** 2424–2433 (1998).

65. Weber, J., Taylor, L., Roussel, M., Sherr, C. & Bar-Sagi, D. Nucleolar Arf sequesters Mdm2 and activates p53. *Nat. Cell Biol.* **1,** 10–26 (1999).

66. Tao, W. & Levine, A. P19 ARF stabilizes p53 by blocking nucleo-cytoplasmic shuttling of Mdm2. *PNAS* **96,** 6937–6941 (1999).

67. Oda, E. *et al.* Candidate Mediator of p53-Induced Apoptosis Noxa, a BH3-Only Member of the Bcl-2 Family and. *Science* **288,** 1053–1505 (2000).

68. Eischen, C. M., Woo, D., Roussel, M. F. & Cleveland, J. L. Apoptosis Triggered by Myc-Induced Suppression of Bcl-X L or Bcl-2 Is Bypassed during Lymphomagenesis Apoptosis Triggered by Myc-Induced Suppression of Bcl-X L or Bcl-2 Is Bypassed during Lymphomagenesis. *Mol. Cell. Biol.* **21,** 5063–5070 (2001).

69. Maclean, K. H., Keller, U. B., Rodriguez-Galindo, C., Nilsson, J. a & Cleveland, J. L. c-Myc augments gamma irradiation-induced apoptosis by suppressing Bcl-XL. *Mol. Cell. Biol.* **23,** 7256–7270 (2003).

70. Chipuk, J., Bouchier-Hayes, L., Kuwana, T., Newmeyer, D. & Green, D. PUMA Couples the Nuclear and Cytoplasmic Proapoptotic Function of p53. *Science* **309,** 1732–1735 (2005).

71. Chipuk, J. E. *et al.* Direct Activation of Bax by p53 Mediates Mitochondrial Membrane Permeabilization and Apoptosis. *Science* **303,** 1010–1014 (2004).

72. Mitchell, K. O. *et al.* Bax is a transcriptional target and mediator of c-Myc-induced

apoptosis. *Cancer Res.* **60,** 6318–6325 (2000).

73. Nikiforov, M. A. *et al.* Tumor cell-selective regulation of NOXA by c-MYC in response to proteasome inhibition. *PNAS* **104,** 19488–93 (2007).

74. Eischen, C. M. *et al.* Bcl-2 is an apoptotic target suppressed by both c-Myc and E2F-1. *Oncogene* **20,** 6983–6993 (2001).

75. Egle, A., Harris, A. W., Bouillet, P. & Cory, S. Bim is a suppressor of Myc-induced mouse B cell leukemia. *PNAS* **101,** 6164–9 (2004).

76. Evan, G. *et al.* Induction of apoptosis in fibroblasts by c-myc protein. *Cell* **69,** 119–128 (1992).

77. Askew, D., Ashmun, R., Simmons, B. & Cleveland, J. Constitutive c-myc expression in an IL-3-dependent myeloid cell line suppresses cell cycle arrest and accelerates apoptosis. *Oncogene* **6,** 1915–1922 (1991).

78. Strasser, A., Harris, A., Bath, M. & Cory, S. Novel primitive lymphoid tumours induced in transgenic mice by cooperation between myc and bcl-2. *Nature* **348,** 331–333 (1990).

79. Fanidi, A., Harrington, E. A. & Evan, G. I. Cooperative interaction between c-myc and bcl-2 proto-oncogenes. *Nature* **355,** 242–244 (1992).

80. Bissonnette, R. P., Echeverri, F., Mahboubi, A. & Green, D. R. Apoptotic cell death induced by c-myc is inhibited by bcl-2. *Nature* **355,** 242–244 (1992).

81. Shiio, Y. *et al.* Quantitative proteomic analysis of Myc oncoprotein function. *EMBO J.* **21,** 5088–5096 (2002).

82. Gebhardt, A. *et al.* Myc regulates keratinocyte adhesion and differentiation via complex formation with Miz1. *J. Cell Biol. Cell Biol.* **172,** 139–149 (2006).

83. Lee, L. A. & Dang, C. V. Myc target transcriptomes. *Curr. Top. Microbiol. Immunol.* **302,** 145–167 (2006).

84. Dominguez-sola, D. & Gautier, J. MYC and the Control of DNA Replication. *Cold Spring Harb. Perspect. Med.* **4,** 1–20 (2014).

85. Takayama, M. A., Taira, T., Tamai, K., Iguchi-Ariga, S. M. M. & Ariga, H. ORC1 interacts withc-Myc to inhibit E-box-dependent transcription by abrogating c-Myc-SNF5/INI1 interaction. *Genes to cells* **5,** 481–490 (2000).

86. Dominguez-Sola, D. *et al.* Non-transcriptional control of DNA replication by c-Myc. *Nature* **448,** 445–451 (2007).

87. Koch, H. B. *et al.* Large-scale identification of c-MYC-associated proteins using a combined TAP/MudPIT approach. *Cell Cycle* **6,** 205–217 (2007).

88. Valovka, T. *et al.* Transcriptional control of DNA replication licensing by Myc. *Sci. Rep.* **3,** 3444 (2013).

89. Kwan, K. Y., Shen, J. & Corey, D. P. C-MYC transcriptionally amplifies SOX2 target genes to regulate self-renewal in multipotent otic progenitor cells. *Stem Cell Reports* **4,** 47–60 (2015).

90. Sabò, A. *et al.* Selective transcriptional regulation by Myc in cellular growth control and lymphomagenesis. *Nature* **511,** 488–492 (2014).

91. Chou, C. *et al.* c-Myc-induced transcription factor AP4 is required for CD8(+) T cell-mediated host protection. *Nat. Immunol.* **15,** 884–893 (2014).

92. Leone, G. *et al.* Myc requires distinct E2F activities to induce S phase and apoptosis. *Mol. Cell* **8,** 105–113 (2001).

93. Rounbehler, R. J. *et al.* Tristetraprolin impairs Myc-induced lymphoma and abolishes the malignant state. *Cell* **150,** 563–574 (2012).

94. Chang, T.-C. *et al.* Widespread microRNA repression by Myc contributes to tumorigenesis. *Nat. Genet.* **40,** 43–50 (2008).

95. Tao, J., Tao, J. & Zhao, X. C-MYC-miRNA circuitry: A central regulator of aggressive B-cell malignancies. *Cell Cycle* **13,** 191–198 (2014).

96. Kelly, K., Cochran, B., Stiles, C. & Leder, P. Cell-specific regulation of the c-myc gene by lymphocyte mitogens and platelet-derived growth factor. *Cell* **35,** 603–610 (1983).

97. Bentley, D. & Groudine, M. A block to elongation is largely responsible for decreased transcription of c-myc in differentiated HL60 cells. *Nature* **321,** 702–706 (1986).

98. Eick, D. & Bornkamm, G. Transcriptional arrest within the first exon is a fast control mechanism in c-myc gene expression. *Nucleic Acids Res.* **14,** 8331–8346 (1986).

99. Krystal, G. *et al.* Multiple Mechanisms for Transcriptional Regulation of the myc Gene Family in Small-Cell Lung Cancer. *Mol. Cell. Biol.* **8,** 3373–3381 (1988).

100. Xu, L., Morgenbesser, S. & Depinho, R. Complex Transcriptional Regulation of myc Family Gene Expression in the Developing Mouse Brain and Liver. *Mol. Cell. Biol.* **11,** 6007–6015 (1991).

101. Culjkovic, B., Topisirovic, I., Skrabanek, L., Ruiz-gutierrez, M. & Borden, K. L. B. eIF4E is a central node of an RNA regulon that governs cellular proliferation Biljana. *J. Cell Biol.* **175,** 415–426 (2006).

102. Dani, C. *et al.* Extreme instability of myc mRNA in normal and transformed human cells Biochemistry : *PNAS* **81,** 7046–7050 (1984).

103. Sampson, V. B. *et al.* MicroRNA Let-7a Down-regulates MYC and Reverts MYC-Induced Growth in Burkitt Lymphoma Cells. *Cancer Res.* **67,** 9762–9771 (2007).

104. Sachdeva, M. *et al.* p53 represses c-Myc through induction of the tumor suppressor miR-145. *PNAS* **106,** 3207–3212 (2009).

105. Bhatia, S., Kaul, D. & Varma, N. Potential tumor suppressive function of miR-196b in B-cell lineage acute lymphoblastic leukemia. *Mol. Cell. Biochem.* **340,** 97–106 (2010).

106. Abe, W. *et al.* miR-196b targets c-myc and Bcl-2 expression, inhibits proliferation and induces apoptosis in endometriotic stromal cells. *Hum. Reprod.* **28,** 750–761 (2013).

107. Miao, L. *et al.* MiR-449c targets c-Myc and inhibits NSCLC cell progression. *FEBS Lett.* **587,** 1359–1365 (2013).

108. Mazan-mamczarz, K., Lal, A., Martindale, J. L., Kawai, T. & Gorospe, M. Translational Repression by RNA-Binding Protein TIAR. *Mol. Cell. Biol.* **26,** 2716–2727 (2006).

109. Liao, B., Hu, Y. & Brewer, G. Competitive binding of AUF1 and TIAR to MYC mRNA controls its translation. *Nat. Struct. Mol. Biol.* **14,** 511–518 (2007).

110. Kim, H. H. *et al.* HuR recruits let-7 / RISC to repress c-Myc expression. *Genes Dev.* **23,** 1743–1748 (2009).

111. Kozak, M. The scanning model for translation:an update. *J. Cell Biol.* **108,** 229–241 (1989).

112. Darveau, A., Pelletier, J. & Sonenberg, N. Differential efficiencies of in vitro translation of mouse c-myc transcripts differing in the 5' untranslated region. *PNAS* **82,** 2315–2319 (1985).

113. Parkin, N., Darveau, A., Nicholson, R. & Sonenberg, N. cis-acting translational effects of the 5' noncoding region of c-myc mRNA. *Mol Cell Biol* **8,** 2875–2883 (1988).

114. Paulin, F. *et al.* Aberrant translational control of the c-myc gene in multiple myeloma. *Oncogene* **13,** 505–513 (1996).

115. Stoneley, M., Paulin, F. E., Quesne, J. P. Le, Chappell, S. A. & Willis, A. E. C-Myc 5' untranslated region contains an internal ribosome entry segment. *Oncogene* **16,** 423–428 (1998).

116. Paulin, F. E. M., Chappell, S. A. & Willis, A. E. A single nucleotide change in the c-myc internal ribosome entry segment leads to enhanced binding of a group of protein factors. *Nucleic Acids Res.* **26,** 3097–3103 (1998).

117. Chappell, S. a *et al.* A mutation in the c-myc-IRES leads to enhanced internal ribosome entry in multiple myeloma: a novel mechanism of oncogene de-regulation. *Oncogene* **19,** 4437–4440 (2000).

118. West, M., Sullivan, N. & AE, W. Translational upregulation of the c-myc oncogene in Bloom's syndrome cell lines. *Oncogene* **11,** 2515–2524 (1995).

119. Sears, R. *et al.* Multiple Ras-dependent phosphorylation pathways regulate Myc protein stability. *Genes Dev.* **14,** 2501–2514 (2000).

120. Popov, N., Schülein, C., Jaenicke, L. a & Eilers, M. Ubiquitylation of the amino terminus of Myc by SCF(β-TrCP) antagonizes SCF(Fbw7)-mediated turnover. *Nat. Cell Biol.* **12,** 973–981 (2010).

121. Vervoorts, J., Luscher-Firzlaff, J. & Luscher, B. The Ins and Outs of MYC Regulation by Posttranslational. *J. Biol. Chem.* **281,** 34725–34729 (2006).

122. Sabò, A., Doni, M. & Amati, B. SUMOylation of Myc-family proteins. *PLoS One* **9,** 1–11 (2014).

123. González-Prieto, R., Cuijpers, S. A. G., Kumar, R., Hendriks, I. A. & Vertegaal, A. C. O. c-Myc is targeted to the proteasome for degradation in a SUMOylation-dependent manner, regulated by PIAS1, SENP7 and RNF4. *Cell Cycle* **14,** 1859–1872 (2015).

124. Rabellino, A. *et al.* PIAS1 Promotes Lymphomagenesis through MYC Upregulation. *Cell Rep.* **15,** 2266–2278 (2016).

125. Hann, S. R. & Eisenman, R. N. Proteins Encoded by the Human c-myc Oncogene :

Differential Expression in Neoplastic Cells. *Mol. Cell. Biol.* **4,** 2486–2497 (1984).

126. Yeh, E. *et al.* A signalling pathway controlling c-Myc degradation that impacts oncogenic transformation of human cells. *Nat. Cell Biol.* **6,** 308–18 (2004).

127. Arnold, H. K. & Sears, R. C. Protein phosphatase 2A regulatory subunit B56α associates with c-Myc and negatively regulates c-Myc accumulation. *Mol. Cell. Biol.* **26,** 2832–2844 (2006).

128. Kim, S. Y. *et al.* Skp2 Regulates Myc Protein Stability and Activity The State University of New York at Stony Brook. *Mol. Cell* **11,** 1177–1188 (2003).

129. Von Der Lehr, N. *et al.* 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 (2003).

130. Vervoorts, J. *et al.* Stimulation of c-MYC transcriptional activity and acetylation by recruitment of the cofactor CBP. *EMBO Rep.* **4,** 484–490 (2003).

131. Patel, J. H. *et al.* The c-MYC Oncoprotein Is a Substrate of the Acetyltransferases hGCN5 / PCAF and TI60. *Mol. Cell. Biol.* **24,** 10826–10834 (2004).

132. Faiola, F. *et al.* 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 (2005).

133. Kalkat, M. *et al.* Identification of c-MYC SUMOylation by mass spectrometry. *PLoS One* **9,** 1–19 (2014).

134. Tansey WP. Mammalian MYC Proteins and Cancer. *New J. Sci.* **2014,** 1–27 (2014).

135. Kato, G. J., Barrett, J., Villa-Garcia, M. & Dang, C. V. An amino-terminal c-myc domain required for neoplastic transformation activates transcription. *Mol. Cell. Biol.* **10,** 5914–5920 (1990).

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

137. Li, L. H., Nerlov, C., Prendergast, G., MacGregor, D. & Ziff, E. B. c-Myc represses transcription in vivo by a novel mechanism dependent on the initiator element and Myc box II. *EMBO J.* **13,** 4070–9 (1994).

138. Zhang, X. Y., DeSalle, L. M. & McMahon, S. B. Identification of novel targets of MYC whose transcription requires the essential MbII domain. *Cell Cycle* **5,** 238–241 (2006).

139. Stone, J. *et al.* Definition of regions in human c-myc that are involved in transformation and nuclear localization. *Mol. Cell. Biol.* **7,** 1697–1709 (1987).

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

141. Frank, S. R., Schroeder, M., Fernandez, P., Taubert, S. & Amati, B. Binding of c-Myc to chromatin mediates mitogen-induced acetylation of histone H4 and gene activation. *Gene Dev.* 2069–2082 (2001). doi:10.1101/gad.906601.Recent

142. Koch, H. B. *et al.* Large-scale identification of c-MYC-associated proteins using a combined TAP/MudPIT approach. *Cell Cycle* **6,** 205–217 (2007).

143. Ponzielli, R., Katz, S., Barsyte-Lovejoy, D. & Penn, L. Z. Cancer therapeutics: Targeting the dark side of Myc. *Eur. J. Cancer* **41,** 2485–2501 (2005).

144. Tu, W. B. *et al.* Myc and its interactors take shape. *Biochim. Biophys. Acta* **5,** 469–483 (2015).

145. Dang, C. & Lee, W. Identification of the human c-myc protein nuclear translocation signal. *Mol. Cell. Biol.* **8,** 4048–54 (1988).

146. Conacci-Sorrel, M., Ngouenet, C. & Eisenman, R. N. Myc-nick: A cytoplasmic cleavage product of Myc that promotes α-tubulin acetylation and cell differentiation. *Cell* **142,** 480–493 (2010).

147. Herbst, A. *et al.* A conserved element in Myc that negatively regulates its proapoptotic activity. *EMBO Rep.* **6,** 177–83 (2005).

148. Herbst, A., Salghetti, S. E., Kim, S. Y. & Tansey, W. P. Multiple cell-type-specific elements regulate Myc protein stability. *Oncogene* **23,** 3863–3871 (2004).

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

150. Thomas, L. R. *et al.* Interaction with WDR5 Promotes Target Gene Recognition and Tumorigenesis by MYC. *Mol. Cell* **58,** 440–452 (2015).

151. Cowling, V. H., Chandriani, S., Whitfield, M. L. & Cole, M. D. A conserved Myc protein domain, MBIV, regulates DNA binding, apoptosis, transformation, and G2 arrest. *Mol. Cell. Biol.* **26,** 4226–39 (2006).

152. Thomas, L. R. *et al.* Interaction of MYC with host cell factor-1 is mediated by the evolutionarily conserved Myc box IV motif. *Oncogene* **35,** 3613–3618 (2016).

153. Murre, C. *et al.* Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* **58,** 537–544 (1989).

154. Smith, M. J., Prochownik, E. V & Charron-Prochownik, D. C. The leucine zipper of c-Myc is required for full inhibition of erythroleukemia differentiation. *Mol. Cell. Biol.* **10,** 5333–5339 (1990).

155. Dang, C. V *et al.* Intracellular leucine zipper interactions suggest c-Myc hetero-oligomerization. *Mol. Cell. Biol.* **11,** 954–62 (1991).

156. Littlewood, T. D., Amati, B., Land, H. & Evan, G. I. max and c-Myc/Max DNA-binding activities in cell extracts. *Oncogene* **7,** 1783–1792 (1992).

157. Blackwood, E. M. & Eisenman, R. N. Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. *Science* **251,** 1211–1217 (1991).

158. Prendergast, G. C., Lawe, D. & Ziff, E. B. Association of Myn, the murine homolog of Max, with c-Myc stimulates methylation-sensitive DNA binding and ras cotransformation. *Cell* **65,** 395–407 (1991).

159. Amati, B. *et al.* Transcriptional activation by the human c-Myc oncoprotein in yeast requires interaction with Max. *Nature* (1992).

160. Reddy, C. D. *et al.* Mutational analysis of Max: role of basic, helix-loop-helix/leucine zipper domains in DNA binding, dimerization and regulation of Myc-mediated transcriptional activation. *Oncogene* **7,** 2085–2092 (1992).

161. Kretzner, L., Blackwood, E. M. & Eisenman, R. N. Myc and Max proteins possess distinct transcriptional activities. *Nature* **355,** 242–244 (1992).

162. Crouch, D. H. *et al.* Gene-regulatory properties of Myc helix-loop-helix/leucine zipper mutants: Max-dependent DNA binding and transcriptional activation in yeast correlates with transforming capacity. *Oncogene* **8,** 1849–1855 (1993).

163. Amati, B. *et al.* Oncogenic activity of the c-Myc protein requires dimerization with Max. *Cell* **29,** 233–245 (1993).

164. Amati, B., Littlewood, T. D., Evan, G. I. & Land, H. The c-Myc protein induces cell cycle progression and apoptosis through dimerization with Max. *EMBO J.* **12,** 5083–7 (1993).

165. Ferré-D'Amaré, A. R., Prendergast, G. C., Ziff, E. B. & Burley, S. K. Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature* **363,** 38–45 (1993).

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

167. Kato, G. J., Lee, W. M. F., Chen, L. & Dang, C. V. Max: Functional domains and interaction with c-Myc. *Genes Dev.* **6,** 81–92 (1992).

168. O'Shea, E., Klemm, J., Kim, P. & Alber, T. X-ray Structure of the GCN4 Leucine Zipper , a Two-Stranded , Parallel Coiled Coil. *Science* **254,** 539–544 (1991).

169. Ellenberger, T., Brandl, C., Struhl, K. & Harrison, S. The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted α Helices: Crystal structure of the protein-DNA complex. *Cell* **71,** 1223–1237 (1992).

170. Ma, P., Rould, M., Weintraub, H. & Pabo, C. Crystal structure of MyoD bHLH domain-DNA complex: Perspectives on DNA recognition and implications for transcriptional activation. *Cell* **77,** 451–459 (1994).

171. Ferré-D'Amaré, A., Pognonec, P., Roeder, R. & Burley, S. Structure and function of the b / HLH / Z domain of USF. *EMBO J.* **13,** 180–189 (1994).

172. Shimizu, T. *et al.* Crystal structure of PHO4 bHLH domain – DNA complex : flanking base recognition. *EMBO J.* **16,** 4689–4697 (1997).

173. Peukert, K. *et al.* An alternative pathway for gene regulation by Myc. *EMBO J.* **16,** 5672–5686 (1997).

174. Herold, S. *et al.* Negative regulation of the mammalian UV response by Myc through association with Miz-1. *Mol. Cell* **10,** 509–521 (2002).

175. Kosan, C. *et al.* Transcription Factor Miz-1 Is Required to Regulate Interleukin-7 Receptor Signaling at Early Commitment Stages of B Cell Differentiation. *Immunity* **33,** 917–928

(2010).

176. Van Riggelen, J. *et al.* The interaction between Myc and Miz1 is required to antagonize TGFβ-dependent autocrine signaling during lymphoma formation and maintenance. *Genes Dev.* **24,** 1281–1294 (2010).

177. Wolf, E. *et al.* Miz1 is required to maintain autophagic flux. *Nat. Commun.* **4,** (2013).

178. Barrilleaux, B. L. *et al.* Miz-1 Activates Gene Expression via a Novel Consensus DNA Binding Motif. *PLoS One* **9,** 1–14 (2014).

179. Walz, S. *et al.* Activation and repression by oncogenic MYC shape tumour-specific gene expression profiles. *Nature* **511,** 483–487 (2014).

180. Lorenzin, F. *et al.* Different promoter affinities account for specificity in MYC-dependent gene regulation. *Elife* **5,** 1–35 (2016).

181. de Pretis, S. *et al.* Integrative analysis of RNA Polymerase II and transcriptional dynamics upon MYC activation. *Genome Res.* **in press,** (2017).

182. Bao, J. & Zervos, A. Isolation and characterization of Nmi, a novel partner of Myc proteins. *Oncogene* **16,** 2171–2176 (1996).

183. Li, H., Lee, T. & Avraham, H. A Novel Tricomplex of BRCA1 , Nmi , and c-Myc Inhibits c-Myc-induced Human Telomerase Reverse Transcriptase Gene (hTERT) Promoter Activity in Breast Cancer *. *Oncogene* **277,** 20965–20973 (2002).

184. Wang, Q., Zhang, H., Kajino, K. & Greene, M. I. BRCA1 binds c-Myc and inhibits its transcriptional and transforming activity in cells. *Oncogene* **1,** 1939–1948 (1998).

185. Kennedy, R. D. *et al.* BRCA1 and c-Myc Associate to Transcriptionally Repress Psoriasin , a DNA Damage – Inducible Gene. *Cancer Lett.* **65,** 10265–10273 (2005).

186. Gaubatz, S. *et al.* Transcriptional activation by Myc is under negative control by the transcription factor AP-2. *EMBO J.* **14,** 1508–1519 (1995).

187. Yu, L. *et al.* AP-2 α Inhibits c-MYC Induced Oxidative Stress and Apoptosis in HaCaT Human Keratinocytes. *J. Oncol.* **2009,** 1–9 (2009).

188. Blackwell, T. K. & Weintraub, H. Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. *Science* **250,** 1104–1110 (1990).

189. Hu, Y., Luscher, B., Adinon, A., Mermod, N. & Tjian, R. Transcription factor AP-4 contains multiple dimerization domains that regulate dimer specificity. *Genes Dev.* **4,** 1741–1752 (1990).

190. Dang, C. V, Dolde, C., Gillison, M. L. & Kato, J. Discrimination between related DNA sites by a single amino acid residue of Myc-related basic-helix-loop-helix proteins. *PNAS* **89,** 599–602 (1992).

191. Brownlie, P. *et al.* The crystal structure of an intact human Max-DNA complex: New insights into mechanisms of transcriptional control. *Structure* **5,** 509–520 (1997).

192. Lüscher, B. & Larsson, L. G. The basic region/helix-loop-helix/leucine zipper domain of Myc proto-oncoproteins: function and regulation. *Oncogene* **18,** 2955–2966 (1999).

193. Papoulas, O., Williams, N. G. & Kingston, R. E. DNA binding activities of c-Myc purified from eukaryotic cells. *J. Biol. Chem.* **267,** 10470–10480 (1992).

194. Ma, A. *et al.* DNA binding by N- and L-Myc proteins. *Oncogene* **8,** 1093–1098 (1993).

195. Blackwell, T. K. *et al.* Binding of myc proteins to canonical and noncanonical DNA sequences. *Mol. Cell. Biol.* **13,** 5216–5224 (1993).

196. Grandori, C., Mac, J., Siëbelt, F., Ayer, D. & Eisenman, R. Myc-Max heterodimers activate a DEAD box gene and interact with multiple E box-related sites in vivo. *EMBO J.* **15,** 4344–57 (1996).

197. Fisher, F. *et al.* Transcription activation by Myc and Max: flanking sequences target activation to a subset of CACGTG motifs in vivo. *EMBO J.* **12,** 5075–82 (1993).

198. Solomon, D. L., Amati, B. & Land, H. Distinct DNA binding preferences for the c-Myc/Max and Max/Max dimers. *Nucleic Acids Res.* **21,** 5372–5376 (1993).

199. Gordân, R. *et al.* Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. *Cell Rep.* **3,** 1093–1104 (2013).

200. Mordelet, F., Horton, J., Hartemink, A. J., Engelhardt, B. E. & Gordân, R. Stability selection for regression-based models of transcription factor-DNA binding specificity. *Bioinformatics* **29,** 117–125 (2013).

201. Afek, A., Schipper, J., Horton, J., Gordân, R. & Lukatsky, D. Protein−DNA binding in the

absence of specific base-pair recognition. *PNAS* **111,** 17140–17145 (2014).

202. Zhou, T. *et al.* Quantitative modeling of transcription factor binding specificities using DNA shape. *PNAS* **112,** 4654–4659 (2015).

203. John, S. *et al.* Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat. Genet.* **43,** 264–268 (2011).

204. Pique-Regi, R., Degner, J. & Pai, A. Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res.* **3,** 447–455 (2011).

205. Guertin, M. J., Martins, A. L., Siepel, A. & Lis, J. T. Accurate prediction of inducible transcription factor binding intensities in Vivo. *PLoS Genet.* **8,** (2012).

206. Arvey, A., Agius, P., Noble, W. W. S. & Leslie, C. Sequence and chromatin determinants of cell-type – specific transcription factor binding. *Genome Res.* **22,** 1723–1734 (2012).

207. Hebbes, T. R., Thorne, A. W., Clayton, A. L. & Crane-robinson, C. Histone acetylation and globin gene switching. *Nucleic Acids Res.* **20,** 1017–1022 (1992).

208. Hebbes, T. R., Clayton, A. L., Thorne, A. W. & Crane-Robinson, C. Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain. *EMBO J.* **13,** 1823–30 (1994).

209. Robertson, A. G. *et al.* Genome-wide relationship between histone H3 lysine 4 mono-and tri-methylation and transcription factor binding. *Genome Res.* **18,** 1906–1917 (2008).

210. Edelman, L. B. & Fraser, P. Transcription factories: Genetic programming in three dimensions. *Curr. Opin. Genet. Dev.* **22,** 110–114 (2012).

211. Jackson, D. A., Hassan, A. B., Errington, R. J. & Cook, P. R. Visualization of focal sites of transcription within human nuclei. *EMBO J.* **12,** 1059–65 (1993).

212. Wansink, D. G. *et al.* Polymerase II in Domains Scattered Throughout the Nucleus. *Cell* **122,** 283–293 (1993).

213. Osborne, C. S. *et al.* Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat. Genet.* **36,** 1065–1071 (2004).

214. Schoenfelder, S. *et al.* Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat. Genet.* **42,** 53–61 (2010).

215. Melnik, S. *et al.* The proteomes of transcription factories containing RNA polymerases I, II or III. *Nat. Methods* **8,** 963–968 (2011).

216. Guccione, E. *et al.* Myc-binding-site recognition in the human genome is determined by chromatin context. *Nat. Cell Biol.* **8,** 764–770 (2006).

217. Soufi, A. *et al.* Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161,** 555–568 (2015).

218. Osborne, C. S. *et al.* Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. *PLoS Biol.* **5,** 1763–1772 (2007).

219. Fernandez, P. C. *et al.* Genomic targets of the human c-Myc protein. *Genes Dev.* **17,** 1115–1129 (2003).

220. Zeller, K. I. *et al.* Global mapping of c-Myc binding sites and target gene networks in human B cells. *PNAS* **103,** 17834–9 (2006).

221. Lee, B. *et al.* Cell-type specific and combinatorial usage of diverse transcription factors revealed by genome-wide binding studies in multiple human cells. *Genome Res.* **22,** 9–24 (2012).

222. Nie, Z. *et al.* c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells. *Cell* **151,** 68–79 (2012).

223. Lin, C. Y. *et al.* Transcriptional amplification in tumor cells with elevated c-Myc. *Cell* **151,** 56–67 (2012).

224. Sabò, A. & Amati, B. Genome Recognition by MYC. *Cold Spring Harb. Perspect. Med.* **4,** (2014).

225. Richart, L. *et al.* BPTF is required for c-MYC transcriptional activity and in vivo tumorigenesis. *Nat. Commun.* **7,** 10153 (2016).

226. Shat, M., Ferré-D'Amarés, A., K, B. S. & Goss, D. J. Anti-cooperative biphasic equilibrium binding of transcription factori upstream stimulatory factor to its cognate DNA monitored by protein fluorescence changes. *J. Biol. Chem.* **270,** 19325–19329 (1995).

227. Cave, J. W., Kremer, W. & Wemmer, D. E. Backbone dynamics of sequence specific recognition and binding by the yeast Pho4 bHLH domain probed by NMR. *Protein Sci.* **9,** 2354–2365 (2000).

228. Sauvé, S., Tremblay, L. & Lavigne, P. The NMR solution structure of a mutant of the max

b/HLH/LZ free of DNA: Insights into the specific and reversible DNA binding mechanism of dimeric transcription factors. *J. Mol. Biol.* **342,** 813–832 (2004).

229. Sauvé, S., Naud, J. & Lavigne, P. The mechanism of discrimination between cognate and non-specific DNA by dimeric b/HLH/LZ transcription factors. *J. Mol. Biol.* **365,** 1163–75 (2007).

230. Walz, S. *et al.* Activation and repression by oncogenic MYC shape tumour-specific gene expression profiles. *Nature* **511,** 483–487 (2014).

231. Yap, C. S., Peterson, A. L., Castellani, G., Sedivy, J. M. & Neretti, N. Kinetic profiling of the c-Myc transcriptome and bioinformatic analysis of repressed gene promoters. *Cell Cycle* **10,** 2184–2196 (2011).

232. Wiese, K. E. *et al.* Repression of SRF target genes is critical for Myc-dependent apoptosis of epithelial cells. *EMBO J.* **34,** 1554–1571 (2015).

233. Wolf, E., Lin, C. Y., Eilers, M. & Levens, D. L. Taming of the beast: Shaping Myc-dependent amplification. *Trends Cell Biol.* **25,** 241–248 (2015).

234. Pires das Neves, R. *et al.* Connecting variability in global transcription rate to mitochondrial variability. *PLoS Biol.* **8,** (2010).

235. Marguerat, S. & Bähler, J. Coordinating genome expression with cell size. *Trends Genet.* **28,** 560–565 (2012).

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

237. Padovan-Merhar, O. *et al.* Single Mammalian Cells Compensate for Differences in Cellular Volume and DNA Copy Number through Independent Global Transcriptional Mechanisms. *Mol. Cell* **58,** 339–352 (2015).

238. Darzynkiewicz, Z., Traganos, F. & Melamed, M. New cell cycle compartments identified by multiparameter flow cytometry. *Cytometry* **1,** 98–108 (1980).

239. Felsher, D. W. & Bishop, J. M. Reversible tumorigenesis by MYC in hematopoietic lineages. *Mol. Cell* **4,** 199–207 (1999).

240. Jain, M. *et al.* Sustained Loss of a Neoplastic Phenotype by Brief Inactivation of MYC. *Science* **297,** 102–104 (2002).

241. Karlsson, Å. *et al.* Genomically complex lymphomas undergo sustained tumor regression upon MYC inactivation unless they acquire novel chromosomal translocations. *Blood* **101,** 2797–2803 (2003).

242. Wu, C. *et al.* Cellular senescence is an important mechanism of tumor regression upon c-Myc inactivation. *PNAS* **104,** 13028–13033 (2007).

243. Berg, T. Inhibition of transcription factors with small organic molecules. *Curr. Opin. Chem. Biol.* **12,** 464–471 (2008).

244. Prochownik, E. V & Vogt, P. K. Therapeutic Targeting of Myc. *Genes Cancer* **1,** 650–659 (2010).

245. Soucek, L. *et al.* Omomyc, a potential Myc dominant negative, enhances Myc-induced apoptosis. *Cancer Res.* **62,** 3507–3510 (2002).

246. Sodir, N. M. *et al.* Endogenous Myc maintains the tumor microenvironment. *Genes Dev.* **25,** 907–916 (2011).

247. Soucek, L. *et al.* Inhibition of Myc family proteins eradicates KRas-driven lung cancer in mice. *Genes Dev.* **27,** 504–513 (2013).

248. Annibali, D. *et al.* Myc inhibition is effective against glioma and reveals a role for Myc in proficient mitosis. *Nat. Commun.* **5,** 1–11 (2014).

249. Jung, L. A. *et al.* OmoMYC blunts promoter invasion by oncogenic MYC to inhibit gene expression characteristic of MYC-dependent tumors. *Oncogene* **36,** 1911–1924 (2017).

250. Soucek, L. *et al.* Modelling Myc inhibition as a cancer therapy. *Nature* **455,** 679–83 (2008).

251. Ran, F. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8,** 2281–2308 (2013).

252. Doudna, J. A. & Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346,** 1258096–1258096 (2014).

253. Gearing, M. *et al. CRISPR 101: A Desktop Resource*. (2017). doi:10.1128/genomeA.00232-16

254. Chen, F. *et al.* High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat. Methods* **8,** 753–755 (2011).

255. Bianchi, V. *et al.* Integrated systems for NGS data management and analysis: Open issues and available solutions. *Front. Genet.* **7,** 1–8 (2016).

256. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).

257. Kishore, K. *et al.* methylPipe and compEpiTools: a suite of R packages for the integrative analysis of epigenomics data. *BMC Bioinformatics* **16,** 313 (2015).

258. Ellenberger, T., Fass, D., Arnaud, M. & Harrison, S. C. Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. *Genes Dev.* **8,** 970–980 (1994).

259. Mateyak, M., Obaya, A., Adachi, S. & Sedivy, J. Phenotypes of c-Myc-deficient rat fibroblasts isolated by targeted homologous recombination. *Cell Growth Differ.* **8,** 1039–1048 (1997).

260. Cowling, V. & Cole, M. The Myc transactivation domain promotes global phosphorylation of the RNA polymerase II carboxy-terminal domain independently of direct DNA binding. *Mol. Cell. Biol.* **27,** 2059–73 (2007).

261. Penn, L. J. Z., Brooks, M. W., Laufer, E. M. & Land, H. Negative autoregulation of c-myc transcription. *EMBO J.* **9,** 1113–1121 (1990).

262. Chou, T., Dang, C. & GW, H. Glycosylation of the c_myc transactivation domain. *PNAS* **92,** 4417–4421 (1995).

263. Chipumuro, E. *et al.* CDK7 Inhibition Suppresses Super-Enhancer-Linked Oncogenic Transcription in MYCN-Driven Cancer. *Cell* **159,** 1126–1139 (2015).

264. Guo, J. *et al.* Sequence specificity incompletely defines the genome-wide occupancy of Myc. *Genome Biol.* **15,** (2014).

265. Allevato, M. *et al.* Sequence-specific DNA binding by MYC/MAX to low-affinity non-E-box motifs. *PLoS One* **12,** e0180147 (2017).

266. Littlewood, T. D., Hancock, D. C., Danielianl, P. S. & Parker, M. G. A modified oestrogen receptor ligand-binding domain as an improved switch for the regulation of heterologous proteins. *Methods* **23,** 1686–1690 (1995).

267. Kress, T. R. *et al.* Identification of MYC-dependent transcriptional programs in oncogene-addicted liver tumors. *Cancer Res.* **76,** 3463–3472 (2016).

268. Corsi, A. K., Brodigan, T. M., Jorgensen, E. M. & Krause, M. Characterization of a dominant negative C. elegans Twist mutant protein with implications for human Saethre-Chotzen syndrome. *Development* **129,** 2761–2772 (2002).

269. Marchegiani, S. *et al.* Recurrent Mutations in the Basic Domain of TWIST2 Cause Ablepharon Macrostomia and Barber-Say Syndromes. *Am. J. Hum. Genet.* **97,** 99–110 (2015).

270. Kim, S. *et al.* Localized TWIST1 and TWIST2 basic domain substitutions cause four distinct human diseases that can be modeled in Caenorhabditis elegans. *Hum. Mol. Genet.* **26,** 2118–2132 (2017).